Featured in Physics

Generation of High-Resolution Handwritten Digits with an Ion-Trap Quantum Computer

Manuel S. Rudolph,[1] Ntwali Bashige Toussaint,[2] Amara Katabarwa,[2] Sonika Johri,[3] Borja Peropadre,[2] and Alejandro Perdomo-Ortiz[1,*]

[1]*Zapata Computing Canada Inc., 325 Front Street West, Toronto, Ontario, M5V 2Y1, Canada*
[2]*Zapata Computing Inc., 100 Federal Street, Boston, Massachusetts 02110, USA*
[3]*IonQ Inc., 4505 Campus Drive, College Park, Maryland 20740, USA*

(Received 26 November 2021; revised 19 April 2022; accepted 1 June 2022; published 15 July 2022)

Generating high-quality data (e.g., images or video) is one of the most exciting and challenging frontiers in unsupervised machine learning. Utilizing quantum computers in such tasks to potentially enhance conventional machine-learning algorithms has emerged as a promising application but poses big challenges due to the limited number of qubits and the level of gate noise in available devices. In this work, we provide the first practical and experimental implementation of a quantum-classical generative algorithm capable of generating high-resolution images of handwritten digits with state-of-the-art gate-based quantum computers. In our quantum-assisted machine-learning framework, we implement a quantum-circuit-based generative model to learn and sample the prior distribution of a generative adversarial network. We introduce a multibasis technique which leverages the unique possibility of measuring quantum states in different bases, hence enhancing the expressivity of the prior distribution. We train this hybrid algorithm on an ion-trap device based on $^{171}Yb^{+}$ ion qubits to generate high-quality images and quantitatively outperform comparable classical generative adversarial networks trained on the popular MNIST dataset for handwritten digits.

DOI: 10.1103/PhysRevX.12.031010 Subject Areas: Quantum Information

I. INTRODUCTION

In the last decades, machine-learning (ML) algorithms have significantly increased in importance and value due to the rapid progress in ML techniques and computational resources [1,2]. However, even state-of-the-art algorithms face significant challenges in learning and generalizing from an ever-increasing volume of unlabeled data [3–5]. With the advent of quantum computing, quantum algorithms for ML arise as natural candidates in the search of applications of noisy intermediate-scale quantum (NISQ) devices, with the potential to surpass classical ML capabilities [6]. Among the top candidates to achieve a quantum advantage in ML are generative models [7], i.e., probabilistic models aiming to capture the most essential features of complex data and to generate similar data by sampling from the trained model distribution. Although there has been promising progress toward demonstrating a quantum supremacy for specific quantum-computing tasks [8,9], and quantum generative models have been proven to learn distributions which are outside of classical reach [10–14], it is not clear whether these theoretical guarantees hold in practice, or whether such enhancements provided by a quantum generative model are limited to cases where one can prove a theoretical gap between classical and quantum algorithms.

In particular, quantum resources offer a divergent set of tools for tackling various challenges and could instead lead to a practical quantum advantage by avoiding pitfalls of conventional classical algorithms, for example, by improving training and consequently enhancing performance on generative tasks.

Despite all promises, applying and scaling quantum models on small quantum devices to tackle real-world datasets remains a big challenge for quantum ML algorithms. Reference [7] proposes to enable quantum models for practical application by exploiting the known dimensionality-reduction capabilities of deep neural networks [15] and compressing classical data before they are handed over to a small quantum device (see Ref. [16] for a framework utilizing tensor networks for data compression). Having a quantum model learn the so-called latent representation of data and take part in a joint quantum-classical training loop, opens up hybrid models to leverage quantum resources and potentially enhance performance when compared to purely classical algorithms. This synergistic interaction between a quantum model and classical deep

*alejandro@zapatacomputing.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

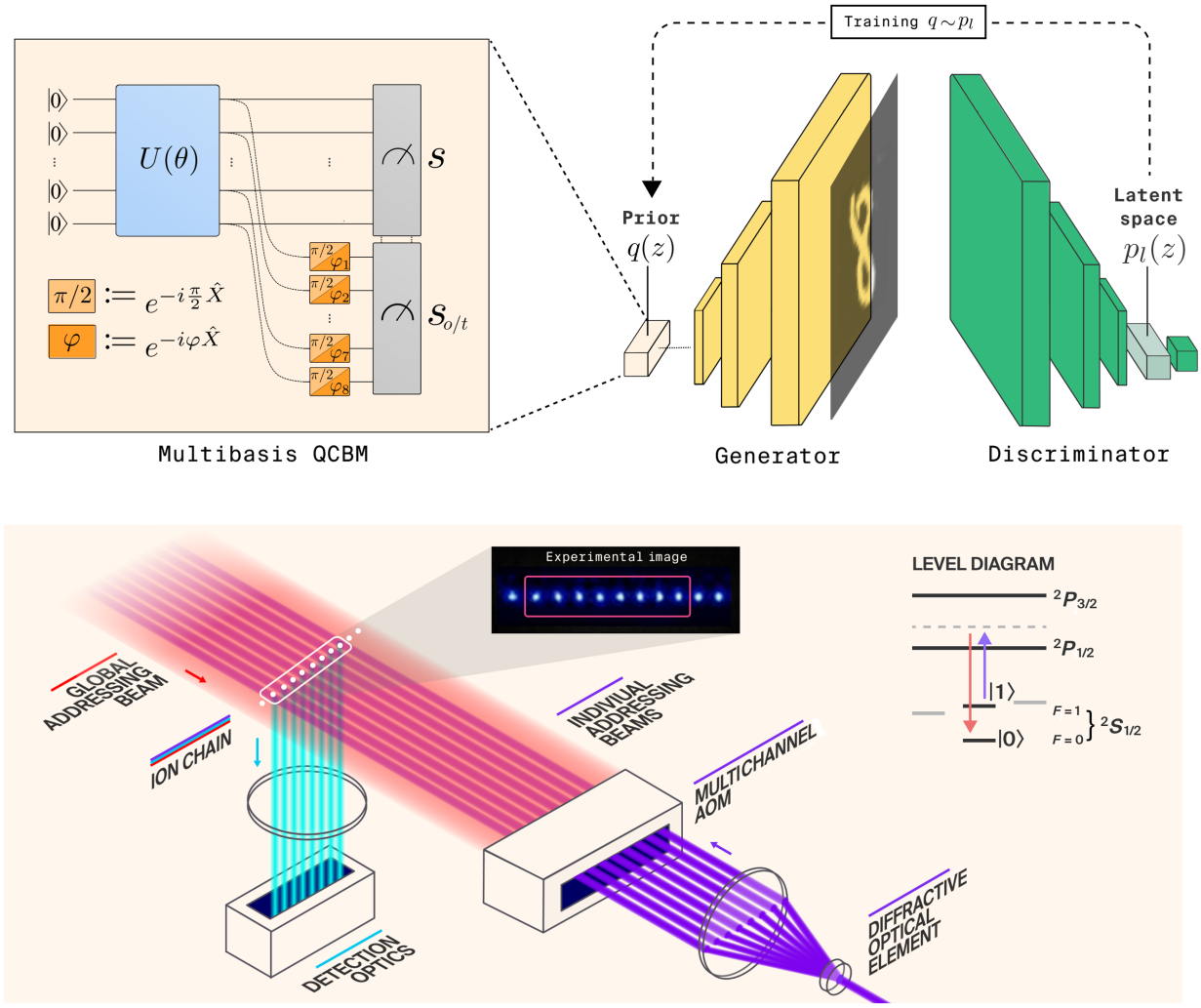

FIG. 1. Top: schematic description of our QC-AAN framework where the prior of a GAN is modeled by a multibasis QCBM. This quantum generative model with encoded distribution $q(z)$ is trained on activations z in the latent space of the discriminator, learning the feature distribution p_l. In the multibasis QCBM, trainable single-qubit rotations follow the parametrized quantum circuit $U(\boldsymbol{\theta})$ and allow for measuring the prepared wave function in additional bases. The angles for these postrotations can be fixed, e.g., at $\pi/2$ to measure s_o with all qubits in the orthogonal Y basis, or they can be trained along with other parameters in $U(\boldsymbol{\theta})$ to measure s_t. Measurements in computational basis s and $s_{o/t}$ are concatenated and forwarded as prior samples into the generator network. The GAN is otherwise trained conventionally. Bottom: illustration of the 11-qubit ion-trap quantum device by IonQ based on $^{171}\mathrm{Yb}^+$ ion qubits. The experimental implementation of the QC-AAN algorithm in this work is performed on eight qubits. The device is operated with automated loading of a linear chain of ions, which is then optically initialized with high fidelity. Computations are performed using a mode-locked 355-nm laser, which drives native single-qubit and two-qubit gates.

neural networks is at the heart of the proposed quantum-assisted Helmholtz machine [7,17] and more recent hybrid proposals [18,19] for enhancing associative adversarial networks (AANs) [20]. In the specific case of Ref. [19], the authors propose to use a quantum Boltzmann machine [21], while Refs. [17,18] experimentally demonstrated this concept with a D-Wave 2000Q annealing device. A similar adoption of this hybrid strategy with quantum annealers has been explored with variational autoencoders [22]. Despite these efforts, a definite demonstration utilizing truly quantum resources on NISQ devices and with full-size ML datasets, e.g., the MNIST dataset of handwritten digits [23], has remained elusive to date. Recent experimental results on gate-based quantum computers [24] illustrate that current proposals are far from generating high-quality MNIST digits.

In this work, we introduce the quantum-circuit associative adversarial network (QC-AAN): a framework combining the capabilities of NISQ devices with classical deep-learning techniques for generative modeling to learn relevant full-scale datasets (see Fig. 1). The framework applies a quantum-circuit Born machine (QCBM) [25] to model and reparametrize the prior distribution of a generative adversarial network (GAN) [26]. Furthermore, we

introduce a multibasis technique for the QCBM and argue that the use of a quantum generative model could enhance deep generative algorithms by providing them with nonclassical distributions and quantum samples from a variety of measurement bases. Finally, to demonstrate the readiness of this framework, we train the QC-AAN with an experimental implementation of eight qubits to generate the first high-resolution handwritten digits with end-to-end training of the quantum component on an ion-trap quantum device. Our results imply that near-term quantum devices could effectively be employed for generative modeling and flexibly assist classical GANs in their learning task.

In Sec. II, we discuss GANs, which are the classical component of the QC-AAN, followed by the AAN framework in Sec. III, which aims to solve the common issues related to conventional GANs by providing them with an informed prior distribution. In Sec. IV, we introduce QCBMs and our multibasis technique, which can be used to possibly enhance the prior of a GAN with quantum measurements. Lastly, in Sec. V, we present simulation results and demonstrate the QC-AAN algorithm on hardware.

II. GENERATIVE ADVERSARIAL NETWORKS

GANs [26] are one of the most popular recent generative machine-learning algorithms that are able to generate remarkably realistic images and other data [27,28]. The algorithm consists of a generator G and a discriminator D, both of which are commonly implemented as deep artificial neural networks. The goal of a GAN is to train G until its generated data are of satisfactory quality. This is achieved by using D as a proxy for estimating the training loss and backpropagating the gradients to G. To that effect, D performs a binary classification task to distinguish between data x from the training dataset p_{data} and data $\tilde{x}$ which were generated data by G. The adversarial min-max loss function of a GAN can be written as

$$\mathcal{C}_{\text{GAN}} = \min_G \max_D [\mathbb{E}_{x\sim p_{\text{data}}}[\log\, D(x)] + \mathbb{E}_{\tilde{x}\sim G}[\log\,[1 - D(\tilde{x})]]], \tag{1}$$

where $\mathbb{E}_{x\sim p_{\text{data}}}$ and $\mathbb{E}_{\tilde{x}\sim G}$ denote the expectations over data sampled from the training dataset p_{data} and the generator G, respectively. The outputs of D are in the domain $D(x), D(\tilde{x}) \in (0,1)$, and it is trained to maximize $\mathcal{C}_{\text{GAN}}$ via $D(x) \to 1$ and $D(\tilde{x}) \to 0$. Conversely, G is trained to minimize $\mathcal{C}_{\text{GAN}}$ by generating data which D cannot classify with certainty $[D(x), D(\tilde{x}) \to 0.5]$. While GANs have been shown to be able to generate remarkable data, common challenges in training a GAN lie in mode collapse and nonconvergence [26,28], which are natural consequences of the delicately balanced adversarial game. One critical component of a GAN, which we argue is understudied and we aim to enhance with a quantum model, is the source of randomness in G, i.e., the prior distribution $q(z)$ in Fig. 1. G takes prior samples z as input to generate $\tilde{x} = G(z)$. In other words, the neural network of G learns a transformation between $q(z)$ and a high-quality data space. Because the model needs to be sampled efficiently on a classical computer, the N-dimensional prior is conventionally implemented as a continuous uniform distribution [i.e., $z \in (0,1)^{\otimes N}$] or a normal distribution [i.e., $z \in \mathcal{N}(\mu,\sigma)^{\otimes N}$ with mean $\mu = 0$ and width σ]. Discrete Bernoulli priors [i.e., $z \in \{0,1\}^{\otimes N}$] have also been shown empirically to be competitive [28] and are used throughout this work for the classical GANs. For such uninformed prior choices, a prior with large N requires a notably expressive neural-network architecture to be able to map the full prior space to high-quality outputs, whereas a small N could potentially lead to the algorithm not learning a good approximation of the full target data [29]. Consequently, ML practitioners often rely on sufficiently large N and scale the number of parameters in G for their purpose. Clearly, an uninformed prior distribution is unlikely to be ideal for any given dataset and GAN architecture, and the prior distribution $q(z)$ should, if possible, be adapted such that G can effectively map prior samples to a high-quality output space. The AAN framework [20] aims to address all of these challenges by implementing a nontrivial prior distribution for G.

III. ASSOCIATIVE ADVERSARIAL NETWORKS

A schematic overview of the AAN framework can be viewed in Fig. 1. In an AAN, the prior is modeled by a smaller generative model with distribution $q_\theta(z)$ which adapts to the training data, the GAN architecture, and the current stage of training. The parameters θ of the prior are tuned to model the node activations in a deep layer l of the discriminator D with matching size N. This layer l is called the latent-space layer, and it captures highly condensed features of the training data and generated data (see Fig. 1). The prior is trained to maximize the likelihood of its parametrized distribution $q_\theta(z)$ given latent data samples z from the latent data distribution $p_l(z)$:

$$\mathcal{C}_q = \max_q \mathbb{E}_{z\sim p_l(z)}[\log\, q_\theta(z)], \tag{2}$$

where $\mathbb{E}_{z\sim p_l(z)}$ indicates the expectation over the observed latent data samples. By training $q_\theta(z)$ to maximize $\mathcal{C}_q$, G receives explicit access to information which D deems to be important for its classification task.

Although the original AAN work proposed using restricted Boltzmann machines (RBMs) [30] to model the prior, RBMs have been shown to be outperformed by comparable QCBMs in learning and sampling probability distributions constructed from real-world data [31]. Therefore, in this work, we implement a QCBM in the prior of a GAN to tackle the common challenges of GANs and potentially enhancing them with measurements from

quantum distributions. As GANs typically employ 16 to 128 dimensional priors (see Appendix E in Ref. [28]), we expect our approach to be able to produce very competitive results on practical datasets by using near- to mid-term quantum devices.

IV. QUANTUM-CIRCUIT BORN MACHINES AND THE MULTIBASIS TECHNIQUE

A QCBM is a circuit-based generative model which encodes a data distribution in a quantum state. This approach allows for sampling of the QCBM by repeatedly preparing and measuring its corresponding wave function

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle. \tag{3}$$

$U(\boldsymbol{\theta})$ is a parametrized quantum circuit acting on an initial qubit state $|0\rangle$, with U chosen according to the capabilities and limitations of NISQ devices. The probabilities for observing any of the 2^n bit strings $\boldsymbol{s}$ in the n-bit (qubit) target probability distribution are modeled using the Born probabilities, such that

$$q_\theta(\boldsymbol{s}) = |\langle \boldsymbol{s}|\psi(\boldsymbol{\theta})\rangle|^2. \tag{4}$$

Importantly, QCBMs can be implemented on most NISQ devices (see, e.g., Refs. [32–36]) and additionally open our algorithm up to exploit unique features of quantum-circuit-based approaches, like measuring in a range of bases.

By training a QCBM on computational basis samples, families of sample distributions, i.e., projections of the wave function, become accessible in a range of other basis sets. This is information which is present in the QCBM but which is conventionally not used when measuring only in the computational basis. Thus, we propose a multibasis technique for the QCBM which provides the QC-AAN with a prior space consisting of quantum samples in flexible bases. The multibasis technique can potentially enhance the performance of the generator by providing it with quantum samples in different measurements bases which have no classical analog. Follow-up work has indeed indicated that quantum measurements in different bases can lead to a separation of quantum and classical generative models [37]. In principle, one could construct and measure many different basis sets, although this implies measuring increasingly redundant bases. The optimal number of basis sets likely depends on the given learning task and available quantum resources. In this work, we restrict ourselves to measurements in one additional basis set, as well as bases which can be reached by one layer of single-qubit rotations. This prevents adding significant depth to the QCBM circuit which is later implemented on quantum hardware. One could, however, measure in bases which are generated by more complex rotations.

Figure 1 displays how the multibasis technique is applied in practice. For each measurement $\mathbf{s}$ in computational basis according to Eq. (4), a second measurement is prepared by applying parametrized single-qubit rotations $R_X(\varphi_i)$ to the wave function for each qubit i and with parameters φ. For $\varphi = \pi/2$, each qubit is rotated into the Y basis, which we refer to as the "orthogonal basis" and is denoted by o throughout this work. The more general case defines what we call the "trained-basis" approach, which we denote by t. Here, the parameters φ are trained for each qubit along with other circuit parameters to optimize the information extracted from the quantum state. When preparing the QCBM wave function in Eq. (3) and applying the corresponding single-qubit postrotations, we obtain a sample $\mathbf{s}_{o/t}$ in the orthogonal or trained basis, respectively. Both measurements $\mathbf{s}$ and $\mathbf{s}_{o/t}$ are then forwarded through a fully connected neural-network layer and into G to learn an effective utilization of the provided quantum resources. These multibasis variants of the QC-AAN are called QC_{+o}-AAN and QC_{+t}-AAN, respectively. For the implementation and training details of the multibasis QCBM, we refer to Appendixes B and G.

V. APPLYING THE QC-AAN FRAMEWORK

As a first step toward showcasing the QC-AAN and our multibasis technique, we numerically simulate training on the canonical MNIST dataset of handwritten digits [23], a standard dataset for benchmarking a variety of ML and deep-learning algorithms. All implementations in this work are performed using the ORQUESTRA™ platform. To isolate the effect of modeling the prior with a QCBM, we compare our quantum-classical models to purely classical deep convolutional GANs (DCGANs) with precisely the same neural-network architecture (see Appendix F for details) and with uniform prior distribution. The QCBM prior in the QC-AAN adds between 31 and 52 trainable parameters, which equates to an increase of approximately 0.02% in the total number of parameters relative to the reference DCGAN used in this work.

The QCBM is initiated with a warm start such that the prior distribution is uniform, and thus, QC-AANs and DCGANs are equivalent at the beginning of training. This initialization provides us with a good baseline for comparison and should additionally avoid complications related to barren plateaus [38]. We then employ a smooth transition training protocol where the QCBM slowly follows changes in the latent space of D to facilitate stable training of G (see Appendix G for details). As shown in Appendix H, the robust global sampling abilities of the QCBM have shown to be very useful in this training protocol where we observe instabilities with the RBM. To quantitatively assess performance, we calculate the Inception Score (IS) (see Appendix I) which evaluates the quality and diversity of generated images in GANs. The IS is high for a model which produces diverse images of high-quality handwritten digits. However, even the MNIST dataset does not achieve a perfect IS of 10, as some digits are not clearly identifiable.

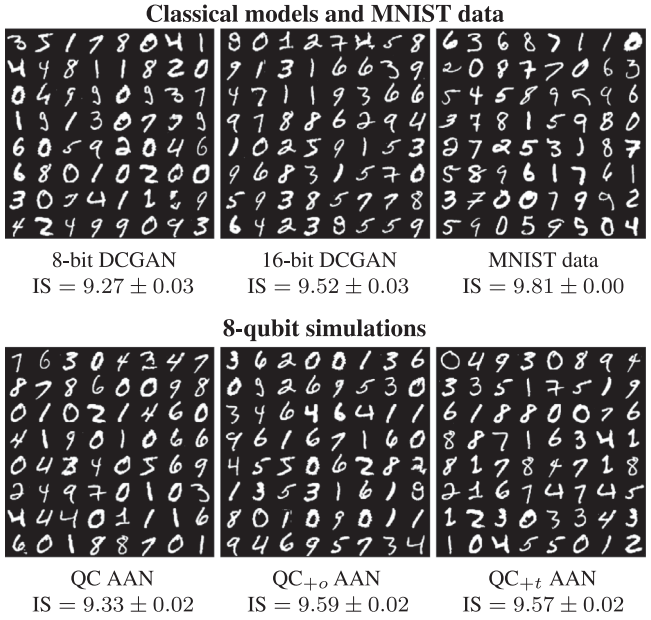

FIG. 2. Digits generated by our QC-AANs with eight simulated qubits and comparable classical DCGANs. The differences in performance are only due to replacing the uninformed random prior of the DCGANs with a multibasis QCBM. All neural-network architectures are equivalent. The models shown here are selected among several training repetitions to have a high IS for their respective model type. The multibasis technique in the $QC_{+o/t}$-AAN enhances the algorithm compared to the 8- and 16-bit DCGAN, and the "plain vanilla" QC-AAN (i.e., without measurements in a second basis).

Figure 2 shows the results of handwritten digits generated by our models. For each model type, we preselect the best-performing models in terms of the IS and, among those, we choose a single representative based on appearance of the images for a human observer. The generated digits themselves are random subsamples of the selected models. It is apparent that all models presented here can achieve good performance and output high-resolution handwritten digits. This is generally expected as the MNIST dataset is considered to be a simple dataset to learn for modern network architectures. In a quantitative evaluation of average model performance (see also Appendix E), we find that the 8-qubit QC-AAN without multibasis technique typically does not outperform comparable 8-bit DCGANs under any of the hyperparameters explored. For low-dimensional priors in general, we are not able to improve the performance of uniform prior distributions. In contrast to that, both multibasis QC-AAN models, the QC_{+o}-AAN and the QC_{+t}-AAN, generate visibly better images and achieve higher scores than the 8-bit and 8-qubit models without additional basis samples. In fact, Fig. 3 shows that, with an average IS of 9.28 and 9.36, respectively, both multibasis models outperform the 16-bit DCGAN with an average IS of 9.20. Considering that the values are expected to saturate toward the IS of the training data (IS = 9.81), these are remarkable results. Additionally, we observe that an 8-qubit multibasis QCBM does not require full access to a 16-qubit Hilbert space to outperform a 16-bit random prior, and that the trained-basis approach generally enhances the algorithm even more compared to the fixed orthogonal-basis approach. We emphasize that these differences in performance are achieved merely by implementing a multibasis QCBM as informed prior to the generator.

To provide final confirmation that the QC-AAN framework is fit for implementation on NISQ devices, we train both $QC_{+o/t}$-AAN algorithms on an 11-qubit quantum device from IonQ which is based on $^{171}Yb^{+}$ ion qubits. For details on the device, we refer to Fig. 1, Appendix A, and Ref. [39]. To achieve high fidelity and to attend to limitations of the quantum hardware, we consider the following adaptations as compared to the simulation setup. While this device allows for all-to-all connectivity, long-range gates are generally more noisy. We thus employ a next-neighbor entangling topology in our quantum circuits. Additionally, we utilize parametrized Mølmer Sørenson entangling gates [see Eq. (A1)] in the circuit ansatz of the QCBM which are native to the ion-trap device. This reduces gate overhead and thus noise introduced from gate recompilation. Finally, to reduce the overall number of calls to the quantum device, as well as wall time between evaluations of the same quantum circuit, we aggregate a large number of measurements between training steps of the quantum prior. To optimize the run-time of the algorithm, one can simultaneously perform and buffer measurements on the quantum device while the classical networks are training with samples from a constant prior distribution. Only during training of the quantum circuit do all other components need to halt. For more information on the training protocol of the QCBM, we refer to Appendix G.

The experimental results for the training on hardware can be viewed in Fig. 3. Every image generated by the GAN during and after training has been produced exclusively utilizing hardware measurements from multiple bases. Note that we demonstrate the results of a single run on hardware per model type. No averaging or postprocessing is performed. To the best of our knowledge, this is the first practical implementation of a quantum-classical algorithm capable of generating high-resolution digits on a NISQ device.

VI. DISCUSSION AND OUTLOOK

With as few as 8 qubits, we show signs of positively influencing the training of GANs and indicate general utility in modeling their prior with a multibasis QCBM on NISQ devices. Learning the choice of the measurement bases through the quantum-classical training loop, i.e., our QC_{+t}-AAN algorithm, appears to be the most successful approach in simulations and also in the experimental realization on the IonQ device. This is a great example of how quantum components in a hybrid quantum ML algorithm are capable of effectively utilizing feedback

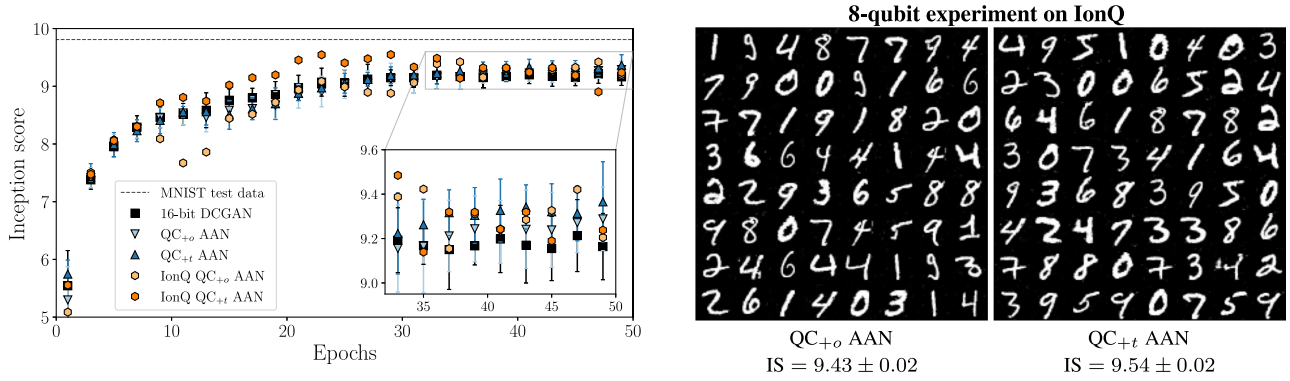

FIG. 3. Left: quantitative comparison between DCGANs with 16-bit random prior distribution and our 8-qubit QC$_{+o/t}$-AAN algorithm. The experimental realization on the IonQ device includes complete implementation of the multibasis QCBM on hardware. Where present, error bars indicate the standard deviation of ten independent training repetitions. The 8-qubit hybrid models generally outperform the classical DCGAN with uninformed prior distribution while the neural-network architectures are equivalent. Right: images of handwritten digits generated by the experimental implementation of the QC$_{+o/t}$-AAN models on the IonQ device. The QC$_{+o}$-AAN model achieves a maximal IS of over 9.4, while the QC$_{+t}$-AAN scores over 9.5 with overall better diversity.

coming from classical neural networks and a testament to the general ML approach of learning the best parameters rather than predetermining them. Unlike many other use-case implementations of quantum algorithms on NISQ devices, our models do not underperform compared to noise-free simulations. It is reasonable that significant reparametrization of the prior space paired with a modest noise floor provides GANs with an improved trade-off between exploration of the target space and convergence to high-quality data.

Our QC-AAN framework also extends flexibly to more complex datasets such as data with higher resolution and color, for which we expect refinement of the prior distribution to become more vital for performance of the algorithm. Besides extending to these more challenging datasets, we could adapt the learning strategy of the quantum prior to follow a different objective function as compared to the AAN framework, potentially one that directly ties into improving the generator's performance in the adversarial loss function in Eq. (1). Because of the distributed nature of the quantum-classical optimization loop, one may additionally adapt the training to follow a secure learning protocol [40–42] for the current scenario where quantum and classical components are owned by separate entities.

Our work has been driven by the challenge of finding a practical generative learning task which is more successfully accomplished with a quantum generative model [7]. This could, for example, be achieved by encoding classically out-of-reach quantum distributions [10–14]. In fact, recent work has used the idea introduced here of using measurements from multiple bases to prove one kind of such classical-quantum model separation [37]. However, our work also emphasizes another form of practical quantum advantage which is usually dismissed, but arguably equally important in reaching a high-performing generative model: the training trajectory in model space that needs to be traversed from the initial distribution to the target distribution. For example, we observe an improvement in robustness of the QCBM prior as compared to the RBM, which we showcase in Appendix H. Both quantum enhancement scenarios are schematically discussed and illustrated in Appendix J. Consequently, it is essential that we better understand the capabilities of our multibasis technique, which kind of quantum distributions can be built from families of basis measurements, and their practical impact on GAN training. Particularly, one needs to consider the trade-offs of the multibasis technique against doubling the number of qubits in the QCBM. While the multibasis model is less expressive, it has the advantage that it can be implemented on earlier quantum devices, has lower circuit depth, employs significantly fewer parameters, and may thus be more effectively trainable. Additionally, the multibasis QCBM exhibits natural synergy with the subsequent neural network that transforms the samples from different measurement bases. The seemingly rigid and structured bipartition of the effective prior space is significantly mitigated by a fully connected layer which has the potential to reorder and learn nontrivial correlations among the bits in the prior during training.

Moving forward, the use of a quantum generative model like a QCBM as a building block in the prior space of a larger classical architecture could unlock the path to reliably investigating the impact of certain prior shapes during training of large generative ML models. This is a field which is rather understudied in the context of classical deep-learning algorithms but has shown promising results [18–20,28]. Given that they are highly flexible and can provide true global samples from the encoded distribution, quantum generative models may offer the tools required to enable this research more effectively. Implementing prior distributions from quantum-inspired models such as the

tensor-network-based Born machines [43] is another exciting research direction that we will be exploring. In terms of generative applications outside of image generation, we will be studying the QC-AAN as a generator in the quantum-enhanced framework for combinatorial optimization problems presented in Ref. [44]. There, the QC-AAN allows quantum-enhanced optimizers to explore combinatorial problems with a larger number of variables than those considered with the quantum-inspired tensor network models, and it allows the framework to extend to non-discrete variables.

Finally, a full-fledged quantitative comparison between quantum and classical versions of machine-learning algorithms can be challenging. For instance, classical resources are currently much cheaper and more accessible than quantum resources. For the hardware experiments performed in this work, getting access to and training the quantum component takes significantly more time than sampling pseudo-random numbers for the conventional random GAN prior. Therefore, factors such as cost and training time, particularly for time-sensitive applications, need to be considered when deciding whether a model has reached practical quantum advantage. Our focus has been on showcasing our new quantum-assisted algorithm which leverages quantum circuits and their unique capabilities of using measurements in multiple bases. Although for this task one might be able to achieve better classical ML performance by choosing more sophisticated neural networks or training techniques, our QC-AAN framework is adaptive to any GAN architecture and could potentially enhance even the best classical GANs with NISQ devices on full-scale datasets. With the recent metrics for evaluating the generalization capabilities of generative models [45], this is a question that can be studied in a grounded and quantitative manner.

ACKNOWLEDGMENTS

The authors would like to thank Coleman Collins and Algert Sula for their support with the experimental images and the design of the hardware illustrations. The authors also thank Yudong Cao, Max Radin, Marta Mauri, Matthew Beach, Dax Enshan Koh, and Jérôme F. Gonthier for their feedback on an early version of this manuscript. The authors would like to acknowledge Zapata Computing's Platform Team for all the support with ORQUESTRA™: the software platform used during the execution of all the simulations and experiments shown here. M. S. R. would like to acknowledge Zapata Computing for hosting his Quantum Applications Internship. The authors acknowledge IonQ for sponsoring access to the quantum hardware through the ORQUESTRA™ platform.

APPENDIX A: DETAILS ON THE IONQ HARDWARE

The experimental circuits are implemented on an 11-qubit trapped-ion processor based on $^{171}Yb^{+}$ ion qubits. The hyperfine levels of the $^2S_{1/2}$ ground state are used as the states of the qubit with $|0\rangle \equiv |F = 0, m_F = 0\rangle$, and $|1\rangle \equiv |F = 1, m_F = 0\rangle$. Measurement of the entire qubit register is achieved through state-dependent fluorescence between $|1\rangle$ and $^2P_{1/2}$ states, with the scattered photons being collected through an aperture lens and passed through a dichroic mirror to an array of photon detectors.

The 11-qubit device is operated with automated loading of a linear chain of ions, which is then optically initialized with high fidelity. Computations are performed using a mode-locked 355-nm laser, which drives native single-qubit-gate (SQG) and two-qubit-gate (TQG) operations. TQG operations are done through the motional modes shared by all the ions; this allows for an all-to-all connectivity topology. The native entangling operation, the Mølmer Sørenson gate, written using Pauli operators is

$$\theta_{xx}^{i,j} = e^{-i\frac{\theta}{2}\sigma_x^i\sigma_x^j}. \tag{A1}$$

In order to maintain consistent gate performance, calibrations of the trapped-ion processor are automated. Additionally, phase calibrations are performed for the SQG and TQG sets, as required for implementing computations in queue and to ensure consistency of the gate performance.

The device is commercially available through IonQ's cloud service. On the cloud, the system has all-to-all connectivity, an average 1-qubit gate fidelity of 99.35%, an average 2-qubit gate fidelity of 96.02%, and SPAM fidelity of 99.3%. For more details, we refer to Ref. [39].

APPENDIX B: THE QUANTUM-CIRCUIT BORN MACHINE

Figure 4 shows the quantum-circuit ansatz used throughout this work to implement the QCBM state preparation unitary U such that

$$|\psi(\boldsymbol{\theta})\rangle = \mathrm{U}(\boldsymbol{\theta})|0\rangle. \tag{B1}$$

The ansatz is inspired by capabilities of current ion-trap quantum devices and is structured in layers were expressivity of the model increases as layers are added. Although the QCBM equipped with this ansatz can become a powerful generative model, one needs to consider important trade-offs in the ansatz hyperparameter choice. For NISQ quantum devices, shallow quantum circuits are generally desired as deeper circuits can significantly decrease fidelity of the quantum states. Additionally, deep circuits oftentimes come with an excess number of parametrized quantum gates that enhance expressivity but can compromise trainability as well as create a model that strongly overfits training data. For our work, we limit ourselves to two layers to minimize the number of gates used while introducing entanglement into the quantum state. In order to increase state fidelity for the experimental implementation, we additionally reduce the all-to-all

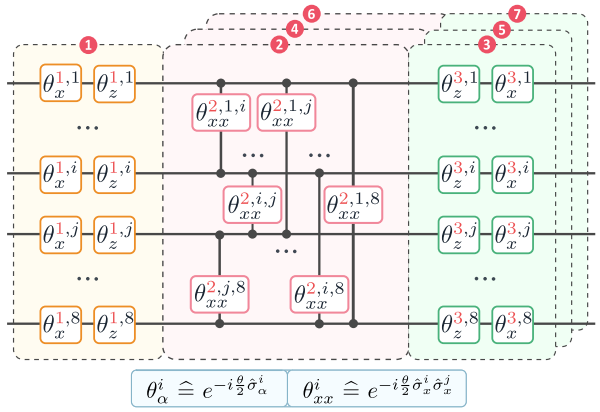

FIG. 4. Quantum-circuit ansatz for the QCBM. The ansatz is structured in layers to control expressivity of the model and fidelity of the prepared state. Throughout this work, we use a layer depth of 2 in order to maximize fidelity in the experimental implementation and enforce some interpolation between learned samples. The red numbers indicate the layer-counting convention. For simulations, we use the all-to-all entangling layer as shown here, whereas for the experimental implementation, we instead adopt linear nearest-neighbor connectivity.

connectivity of the XX gates shown in Fig. 4 to a linear chain of entangling operations for the experiment on the IonQ quantum device. The circuit parameters are initialized with a warm start such that the QCBM encodes a uniform distribution in the computational basis as well as the o/t bases discussed in the main text and Appendix C. In our application, the QCBM is trained by minimizing the clipped negative log-likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = -\sum_x p(\mathbf{x}) \log \ \max[q_{\boldsymbol{\theta}}(\mathbf{x}), \epsilon], \tag{B2}$$

where $p(\mathbf{x})$ is the probability over all training data samples $\mathbf{x}$, and $q_{\boldsymbol{\theta}}(\mathbf{x}) = |\langle \mathbf{x}|\psi(\boldsymbol{\theta})\rangle|^2$ is the QCBM model distribution. Minimizing this function is equivalent to maximizing the expression in Eq. (2). A regularization constant ϵ prevents singularity of the logarithm for samples with zero probability. The probability distribution $q_{\boldsymbol{\theta}}(\mathbf{x})$ of samples $\mathbf{x}$ is estimated by sampling the prepared state and accumulating the measurements.

APPENDIX C: THE MULTIBASIS TECHNIQUE

In this work, we introduce a multibasis technique for quantum-circuit-based models to expand the repertoire of quantum machine-learning researchers.

Commonly, when referring to sampling a generative model, one means generating instances of data that follow the encoded probability distribution. For classical models, one is limited to one basis—the computational basis. In quantum models, this is not the case. When encoding a probability distribution into a qubit wave function, as is the case with a QCBM, the learned wave function contains a potential family of sample distributions which are accessible by measuring in different bases. These additional distributions, or more specifically, projections of the wave function, can be evaluated by applying arbitrary postrotations to the quantum registers before measurement. In this work, we explore the questions of whether we can enhance a generative model by including measurements in additional bases and how we can maximize the benefit of measuring additional basis sets. For this purpose, we quantitatively compare performance inside the QC-AAN framework using samples from an orthogonal measurement basis, i.e., the Y basis, and flexible bases where the postrotation angles are trained together with the ansatz parameters. This is done by doubling the latent space in the discriminator and training the samples of the multibasis QCBM on the respective latent activations. While we do not explore a variety of possible basis choices, we argue that the trainable multibasis model should converge to a learned constellation of measurement bases, which is close to optimal for this specific training instance. Future work will need to study a wider range of variations of the multibasis technique to uncover how to most efficiently utilize near-term quantum devices.

As an explicit example of the sample concatenation, assume samples $\mathbf{s} = 1010$ and $\mathbf{s}_{o/t} = 1100$, which sample in the computational and o/t basis, respectively. Together, they then define $\mathbf{s}^* = 10101100$. For a series of measurements in the computational and o/t bases, the assignment of which pair of measurements is forwarded to the neural network is arbitrary as there is no direct correlation between the computational basis and o/t basis distributions other than that they obey the normalization constraint of the QCBM wave function. This technique generalizes to the measurement and concatenation of samples of any observable able to be measured on a quantum device.

A more subtle advantage of this multibasis technique in the context of generative modeling with a QCBM is that the prior size N (see Sec. II) is doubled from n to $2n$ where n is the number of qubits. The effective sample space for the binary samples of the QCBM scales like 2^N which thus increases from 2^n to 2^{2n}. Although the sample $\mathbf{s}^*$ lives in the $\{0,1\}^{2n}$ space, the multibasis QCBM does not have access to the full $2n$ qubit space because of the normalization constraint of the wave function. However, increasing the sample dimension and additionally the amount of information encoded and utilized from small near-term quantum computers, proves of immense value here by enhancing the expressivity and the robustness of the hybrid QC-AAN model considered.

APPENDIX D: ASSOCIATIVE ADVERSARIAL NETWORKS

AANs were first proposed in Ref. [20] as an extension of the popular GANs. Their purpose was to improve trainability and consequently general performance of GANs by

modeling and reparametrizing the generator G's prior distribution with a RBM. The objective functional for the AAN can be written as

$$\mathcal{C}_{\mathrm{AAN}} = \mathcal{C}_{\mathrm{GAN}} \circ \mathcal{C}_q, \tag{D1}$$

and consists of the traditional min-max game objective $\mathcal{C}_{\mathrm{GAN}}$ of a GAN in Eq. (1), as well as the prior loss $\mathcal{C}_q$ in Eq. (2) which is optimized in tandem with the GAN. Optimizing $\mathcal{C}_q$ maximizes the likelihood of the RBM distribution

$$q(\mathbf{z}) = \sum_{\mathbf{h}} \frac{e^{-E(\mathbf{z},\mathbf{h})}}{Z} \tag{D2}$$

given the latent distribution p_l in the discriminator D. $E(\mathbf{z},\mathbf{h})$ is the energy functional for the RBM with visible units $\mathbf{z}$ and hidden units $\mathbf{h}$, and $Z = \sum_{\mathbf{z},\mathbf{h}} e^{-E(\mathbf{z},\mathbf{h})}$ is the partition function. For more details, we refer the reader to the original AAN paper in Ref. [20].

APPENDIX E: SIMULATED QC-AAN RESULTS FOR 6 AND 8 QUBITS

To benchmark the performance of a QC-AAN with few qubits, we compare the average IS of our QC-AANs, with 6- and 8-qubit multibasis QCBM in the model prior. For this work, we compare only models with the same prior dimension to isolate the effect that reparametrization of the prior distribution has on GAN training. In Fig. 5, we show that for 6- and 8-qubit QC-AANs, we do not achieve an advantage in learning a nontrivial prior. For a 6- (8-) qubit QCBM, there are only 64 (256) distinct samples available for the neural network to map to high-quality images. Since the IS is very sensitive to class imbalance of the generated images, we see that modeling those priors does not lead to meaningful improvements. In fact, the results shown in Fig. 5 for the QC-AAN with 6 and 8 qubits are obtained by minimally disturbing only the uniform distribution. The $\mathrm{QC}_{+o/t}$-AANs with the multibasis technique show interesting results, where for 6 qubits, we almost reach the performance of a 12-bit DCGAN, whereas with 8 qubits, we outperform a 16-bit DCGAN (see Fig. 3 in main text). This is despite the fact that the $\mathrm{QC}_{+o/t}$-AAN models are restricted to an effective subspace compared to a Hilbert space with double the number of qubits. Note that these results are specific to our neural-network architecture. It is possible that more expressive and computationally intensive neural networks face fewer challenges in learning a great model with a uniform prior distribution. For a general learning task and network architecture, the QC-AAN algorithm allows us to consider the hyperparameters of the trainable prior as hyperparameters toward a more successful overall generative algorithm.

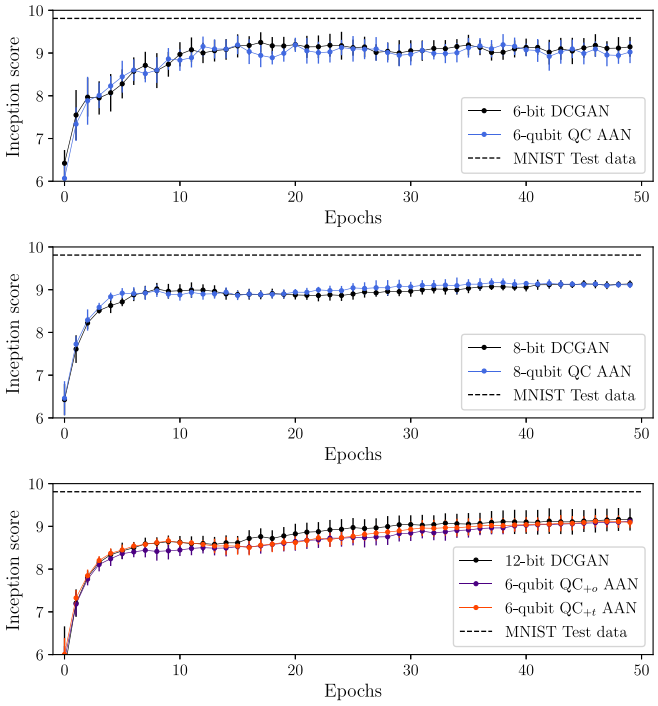

FIG. 5. Simulation results of QC-AANs with 6 and 8 qubits relative to comparable DCGANs with uniform prior distribution. Depicted are average ISs and standard deviations of ten independent training repetitions per model. We observe no significant improvement for the 6- and 8-qubit QC-AANs, as well as for the $\mathrm{QC}_{+o/t}$-AANs with 6 qubits. The first improvements are observed for the 8-qubit $\mathrm{QC}_{+o/t}$-AAN, for which we refer to Fig. 3 in the main text.

The 8-qubit $\mathrm{QC}_{+o/t}$-AANs operate on a space with $2^{16} = 65536$ potential samples. Although the multibasis QCBM is a binary model, with this many different images, an occasional IS of over 9.5 can be considered comparable to state-of-the-art GANs with continuous priors. In fact, Ref. [28] argues that binary units may perform at least as good as continuous uniform or normal distributions.

APPENDIX F: NEURAL-NETWORK ARCHITECTURES AND TRAINING

Figure 6 shows a schematic overview of the network architectures of the generator G and discriminator D neural networks used throughout this work. We emphasize that these architectures are employed for both our hybrid QC-AANs, and the classical DCGANs. The networks have approximately inverse structure with three convolutional layers, although it is not generally required for stable GAN training. The second to last layer in D (latent space) has the same size as the first layer in G (prior space) to be able to train the quantum model in the QC-AAN on the latent activations of D. The total number of parameters in each network amounts to approximately 2.77×10^5. The multibasis QCBM in the QC-AAN adds between 31 and 52

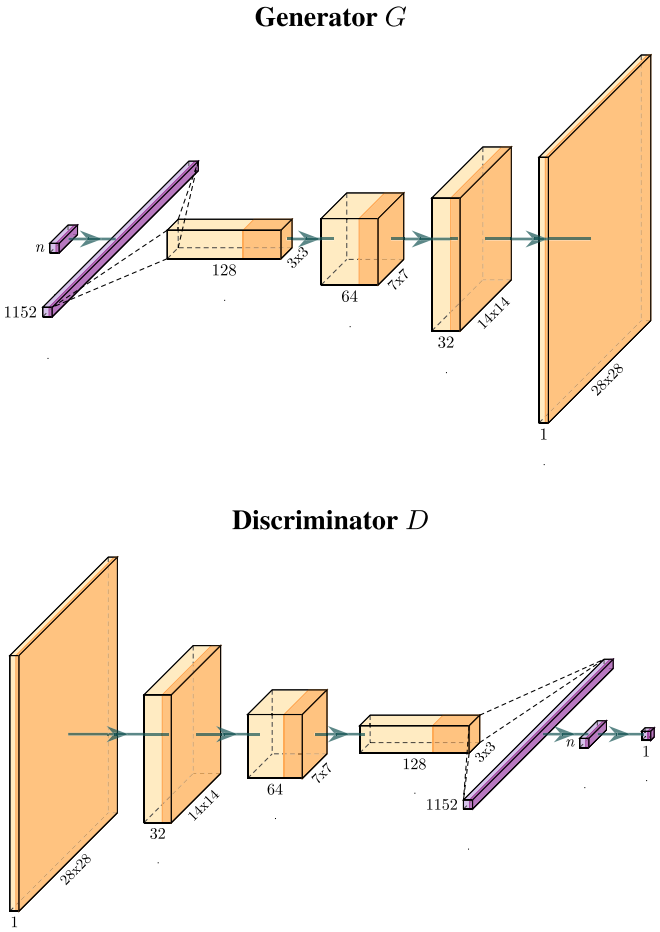

FIG. 6. Schematic neural-network architectures of the generator G and discriminator D used throughout this work for the MNIST dataset of handwritten digits. Purple color indicates a 1D layer of nodes, whereas the orange blocks represent 2D convolutional layers with 128, 64, and 32 channels. The dark color change inside each block indicates the application of a nonlinear activation function after every linear transformation. G and D are approximately inverse, although this is not strictly required. Note that the second to last layer in D represents the latent space which contains n bits, the same size chosen for the prior of G.

trainable parameters for the 8-qubit model, equating to an increase of 0.02% in the total number of parameters.

All convolutional layers have batch normalization [46] and leaky rectified linear unit activation functions [47]. In training D, a small percentage (3%) of training samples have their label flipped, and label smoothing is applied to the training images. The optimizer for both networks is the ADAM optimizer with parameters $[\beta_1, \beta_2] = [0.5, 0.9]$. Generally, our neural-network architectures and hyperparameter choices are inspired by successful practices in modern DCGANs [27].

APPENDIX G: TRAINING DETAILS OF QCBM AS MODEL PRIOR

The QCBM in this work implements a hardware-efficient ansatz inspired by capabilities of ion-trap quantum computers (see Fig. 4). The layer depth of the ansatz is chosen to be shallow with only one layer of single-qubit and entangling gates, respectively. For the numerical simulations, we choose an all-to-all connectivity between qubits, whereas in the hardware implementation, we use linear connectivity to improve the state fidelity. For the case of the MNIST training set, we do not observe on average significant negative effects in reducing the ansatz connectivity. We expect that for a more challenging generative modeling tasks, the circuit ansatz will play a more crucial role.

Over one QC-AAN training epoch on the entire MNIST dataset with $N = 60\,000$ images, we perform an update of the QCBM parameters every 100 batches for the simulation and every 600 batches for the experimental implementation. The latter implies one training step per training epoch. We use the simultaneous perturbation stochastic approximation algorithm [48] for training the parametrized quantum circuit to adapt the model distribution of the QCBM while minimizing calls of the quantum device. The gradients are evaluated with 1000 readout measurements (shots). For the experimental implementation on the quantum device, we sample the 8-qubit distributions with 10^4 shots per measured basis and are able to construct multibasis samples appropriately by resampling those measurements until the next training step. For the numerical simulations, this resampling is not performed, and the circuits are evaluated with as many shots as required for the GAN, i.e., for each image generated by the generator.

One technique that is shown to stabilize training for the QC-AAN is to freeze the prior, i.e., to fix the QCBM parameters, after a certain number of training epochs. Altering the prior distribution significantly in the latter stages of training has shown to destabilize training and lead to visibly worse images. This is likely due to the converging generator network where parameters are slowly settling. Altering the prior at these stages introduces a new impulse for training, which is not matched by the decreasing step sizes of the ADAM optimizer of the networks (see Appendix F). Throughout this work, we freeze the prior after ten epochs.

APPENDIX H: QCBM AND RBM IN THE AAN FRAMEWORK

QCBMs are promising quantum generative models that offer global sampling capabilities of the encoded distribution and additionally provide access to quantum measurements that may be beneficial in learning a strong model. One of those properties, which we leverage in this work, is the possibility to measure in additional bases. Still, one needs to weigh the costs and benefits for such a quantum model when RBMs are lightweight generative models with an efficient but local sampling algorithm. Reference [19] shows that the AAN framework with a RBM modeling the generator's prior could be improved by instead implementing a quantum Boltzmann machine. Finding the best

hyperparameters for their respective models is a notoriously difficult task, and comparing models across all possible hyperparameter combinations is in general unfeasible. In this work, we argue that, given a poor choice of hyperparameters and the same number of model parameters, RBMs can easily become unstable in our smooth learning protocol while QCBMs in the QC-AAN retain their stability. Figure 7 shows the average training performance of ten individual QC-AANs and AANs, both with an eight-dimensional prior on the MNIST training set. The priors are trained in a smooth transition protocol, where every 100 batches, their distributions are altered by one or five training steps of stochastic gradient descent with a step size of 0.01. *A priori*, this step size seems small enough that both of the hyperparemter options of one or five training steps do not seem unreasonable. However, Fig. 7 shows that five training steps is in fact a suboptimal hyperparameter choice. Notably, the RBM suffers from its native local sampling technique and can become very unstable. We also provide the sampled prior distributions of the worst-performing models across ten training runs. It is apparent that the RBM has reached a distribution which it cannot effectively sample. If this happens at any stage of our smooth transition training protocol, further training of the AAN becomes impossible. For the QCBM, the change in performance between two hyperparameter choices is more continuous. This example indicates a potentially improved robustness of the QC-AAN framework as compared to the AAN framework.

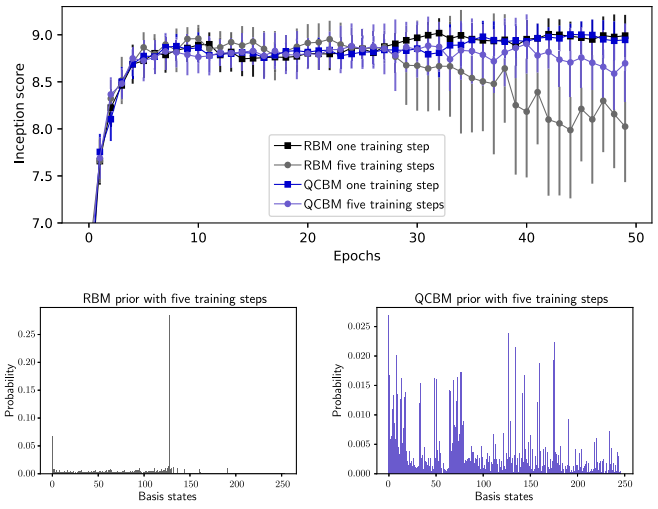

FIG. 7. Example comparison of two hyperparameter configurations for an AAN and QC-AAN, which implement a RBM and QCBM as the model prior, respectively. A training protocol of five training steps per training instance appears to be too much variation in the GAN prior for stable training, but unlike a QCBM, a RBM is prone to becoming unstable under suboptimal configurations and sampling only one distribution mode. When that happens, training of the algorithm fails.

APPENDIX I: INCEPTION SCORE DEFINITION AND DISCUSSION

The IS

$$\mathrm{IS}(G) = \exp\left(\mathbb{E}_{\tilde{x}\sim G}[\mathrm{KL}[p(y|\tilde{x})||p(y)]]\right) \tag{I1}$$

is a popular metric for evaluating GANs. For a given generator G, it measures the quality $p(y|\tilde{x})$ of generated images $\tilde{x}$ and also their diversity $p(y)$ across all possible classes y of the original dataset. Quantitatively, this is done by calculating the KL divergence between the posterior probability distribution $p(y|\tilde{x})$ and the prior probabilities $p(y)$ for each class label. For details, we refer to Ref. [49]. The IS is a human-readable metric with values between 1 and the number of total classes in the dataset, i.e., 10 for the MNIST dataset of handwritten digits. Although it has been proven to be very useful [49], one of the main criticisms of the IS is that it does not depict the realism of generated images for a human observer because it is calculated with classifiers which are trained to search and find exactly the class labels that they have been trained for. Images can either be warped or noisy and still achieve very high classification certainty [50]. In this work specifically, we achieve a surprisingly high IS with models that implement six- and eight-dimensional priors, resulting in only 64 and 256 distinct images in total. Although the IS is high for those models, a human observer does not judge them as being particularly clear images. For such limited models, there arises an interesting effect where the discriminator can remember all images generated by the generator, constantly pushing it away, preventing further convergence and thus introducing noisy artifacts. Another classifier will clearly identify the digits as a member of their particular class, regardless of the noise. Nevertheless, IS is a straightforward quantitative performance measure for GANs that commonly correlates well with human perception.

The IS is commonly calculated with use of the pretrained INCEPTION-V3 network [51] as a proxy to calculate the probabilities in Eq. (I1). Since the INCEPTION-V3 networks can be applied only to colored data, we instead utilize a convolutional classifier with approximately 99.3% accuracy on the MNIST dataset to calculate the ISs.

APPENDIX J: PRACTICAL QUANTUM ADVANTAGE

Although theoretical gaps between classical and quantum generative modeling have been provided (see, e.g., Refs. [12–14,37]), these are not sufficient to achieve a practical quantum advantage since such a gap is not ensured to manifest in real-world datasets. Also, it is not evident that quantum algorithms can outperform only classical methods in cases where a theoretical gap exists for the desired target distribution. For example, the results presented in Ref. [13] imply that for a certain family of

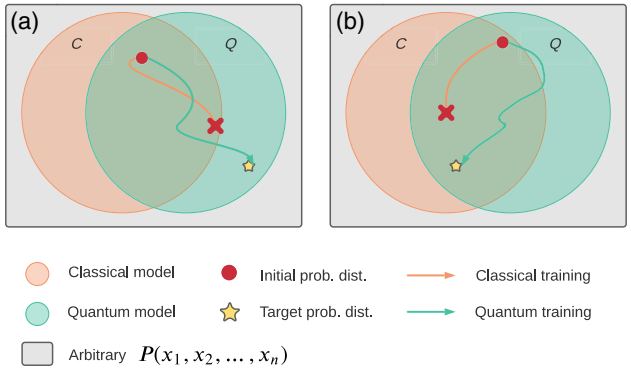

FIG. 8. Sketch for two potential scenarios for practical quantum advantage in generative modeling. The large overlapping circles indicate the set of probability distributions which could be expressed within two given parametrized classical or quantum generative models. Note that this does not represent all distributions which could possibly be expressed by any classical or quantum model. (a) The target distribution lies outside the space that can be expressed by the classical model, similar to Fig. 5 in Ref. [13]. This might be due to the distribution not being tractable classically or because this specific parametrized model cannot reach it. (b) The target distribution lies in a region which both models can reach. However, both models do not traverse the space equally, and the classical model might be hindered, for example, by the presence of more local minima or due to pitfalls in conventional heuristics. As indicated in Fig. 7, training success could be significantly different for the classical and quantum algorithms.

classically intractable target distributions, model distributions during training may also be intractable. As schematically illustrated in Fig. 8(b), quantum resources offer a different toolset for tackling problems such that quantum algorithms could exhibit a practical quantum advantage by other means, even in cases where the generative task is within reach of the classical parametrized model, as is the case for the GANs considered in this work. For example, it is oftentimes overlooked that the training of an algorithm, here specifically a generative algorithm, is essential to its final performance and is arguably more important than tractability of the final outcome. If a quantum model provides more stable training or can navigate the cost function landscape more effectively, as indicated in Appendix H and Fig. 7, this opens up another path toward a practical quantum advantage.

[1] Y. LeCun, Y. Bengio, and G. Hinton, *Deep Learning*, Nature (London) **521**, 436 (2015).
[2] J. Schmidhuber, *Deep Learning in Neural Networks: An Overview*, Neural Netw. **61**, 85 (2015).
[3] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, *Model Compression and Acceleration for Deep Neural Networks: The Principles, Progress, and Challenges*, IEEE Signal Process. Mag. **35**, 126 (2018).
[4] R. Novak, Y. Bahri, D. A. Abolafia, J. Pennington, and J. Sohl-Dickstein, *Sensitivity and Generalization in Neural Networks: An Empirical Study*, arXiv:1802.08760.
[5] B. Neyshabur, Z. Li, S. Bhojanapalli, Y. LeCun, and N. Srebro, *Towards Understanding the Role of Over-Parametrization in Generalization of Neural Networks*, arXiv:1805.12076.
[6] J. Preskill, *Quantum Computing in the NISQ Era and Beyond*, Quantum **2**, 79 (2018).
[7] A. Perdomo-Ortiz, M. Benedetti, J. Realpe-Gómez, and R. Biswas, *Opportunities and Challenges for Quantum-Assisted Machine Learning in Near-Term Quantum Computers*, Quantum Sci. Technol. **3**, 030502 (2018).
[8] 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.*, *Quantum Supremacy Using a Programmable Superconducting Processor*, Nature (London) **574**, 505 (2019).
[9] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, *Quantum Computational Advantage Using Photons*, Science **370**, 1460 (2020).
[10] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, *Expressive Power of Parametrized Quantum Circuits*, Phys. Rev. Research **2**, 033125 (2020).
[11] I. Glasser, R. Sweke, N. Pancotti, J. Eisert, and I. Cirac, *Expressive Power of Tensor-Network Factorizations for Probabilistic Modeling*, Adv. Neural Inf. Process. Syst. **32**, 1498 (2019), https://proceedings.neurips.cc/paper/2019/hash/b86e8d03fe992d1b0e19656875ee557c-Abstract.html.
[12] R. Sweke, J.-P. Seifert, D. Hangleiter, and J. Eisert, *On the Quantum versus Classical Learnability of Discrete Distributions*, Quantum **5**, 417 (2021).
[13] B. Coyle, D. Mills, V. Danos, and E. Kashefi, *The Born Supremacy: Quantum Advantage and Training of an Ising Born Machine*, npj Quantum Inf. **6**, 60 (2020).
[14] M. Hinsche, M. Ioannou, A. Nietner, J. Haferkamp, Y. Quek, D. Hangleiter, J.-P. Seifert, J. Eisert, and R. Sweke, *Learnability of the Output Distributions of Local Quantum Circuits*, arXiv:2110.05517.
[15] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science **313**, 504 (2006).
[16] J. Qi, C.-H. H. Yang, and P.-Y. Chen, *QTN-VQC: An End-to-End Learning Framework for Quantum Neural Networks*, arXiv:2110.03861.
[17] M. Benedetti, J. Realpe-Gómez, and A. Perdomo-Ortiz, *Quantum-Assisted Helmholtz Machines: A Quantum–Classical Deep Learning Framework for Industrial Datasets in Near-Term Devices*, Quantum Sci. Technol. **3**, 034007 (2018).
[18] M. Wilson, T. Vandal, T. Hogg, and E. G. Rieffel, *Quantum-Assisted Associative Adversarial Network: Applying Quantum Annealing in Deep Learning*, Quantum Mach. Intell. **3**, 1 (2021).
[19] E. R. Anschuetz and C. Zanoci, *Near-Term Quantum-Classical Associative Adversarial Networks*, Phys. Rev. A **100**, 052327 (2019).
[20] T. Arici and A. Celikyilmaz, *Associative Adversarial Networks*, arXiv:1611.06953.

[21] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, *Quantum Boltzmann Machine*, Phys. Rev. X **8**, 021050 (2018).
[22] W. Winci, L. Buffoni, H. Sadeghi, A. Khoshaman, E. Andriyash, and M. H. Amin, *A Path towards Quantum Advantage in Training Deep Generative Models with Quantum Annealers*, Mach. Learn. **1**, 045028 (2020).
[23] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-Based Learning Applied to Document Recognition*, Proc. IEEE **86**, 2278 (1998).
[24] H.-L. Huang, Y. Du, M. Gong, Y. Zhao, Y. Wu, C. Wang, S. Li, F. Liang, J. Lin, Y. Xu *et al.*, *Experimental Quantum Generative Adversarial Networks for Image Generation*, Phys. Rev. Applied **16**, 024051 (2021).
[25] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, *A Generative Modeling Approach for Benchmarking and Training Shallow Quantum Circuits*, npj Quantum Inf. **5**, 45 (2019).
[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative Adversarial Nets*, Adv. Neural Inf. Process. Syst. **27**, 2672 (2014), https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.
[27] A. Radford, L. Metz, and S. Chintala, *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, arXiv:1511.06434.
[28] A. Brock, J. Donahue, and K. Simonyan, *Large Scale GAN Training for High Fidelity Natural Image Synthesis*, arXiv:1809.11096.
[29] M. Padala, D. Das, and S. Gujar, *Effect of Input Noise Dimension in GANs*, arXiv:2004.06882.
[30] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, *A Learning Algorithm for Boltzmann Machines*, Cogn. Sci. **9**, 147 (1985).
[31] J. Alcazar, V. Leyton-Ortega, and A. Perdomo-Ortiz, *Classical versus Quantum Models in Machine Learning: Insights from a Finance Application*, Mach. Learn. **1**, 035003 (2020).
[32] K. E. Hamilton, E. F. Dumitrescu, and R. C. Pooser, *Generative Model Benchmarks for Superconducting Qubits*, Phys. Rev. A **99**, 062323 (2019).
[33] V. Leyton-Ortega, A. Perdomo-Ortiz, and O. Perdomo, *Robust Implementation of Generative Modeling with Parametrized Quantum Circuits*, Quantum Mach. Intell. **3**, 1 (2021).
[34] B. Coyle, M. Henderson, J. C. J. Le, N. Kumar, M. Paini, and E. Kashefi, *Quantum versus Classical Generative Modelling in Finance*, Quantum Sci. Technol. **6**, 024013 (2021).
[35] D. Zhu, N. M. Linke, M. Benedetti, K. A. Landsman, N. H. Nguyen, C. H. Alderete, A. Perdomo-Ortiz, N. Korda, A. Garfoot, C. Brecque *et al.*, *Training of Quantum Circuits on a Hybrid Quantum Computer*, Sci. Adv. **5** (2019).
[36] K. E. Hamilton and R. C. Pooser, *Error-Mitigated Data-Driven Circuit Learning on Noisy Quantum Hardware*, Quantum Mach. Intell. **2**, 1 (2020).
[37] X. Gao, E. R. Anschuetz, S.-T. Wang, J. I. Cirac, and M. D. Lukin, *Enhancing Generative Models via Quantum Correlations*, Phys. Rev. X **12**, 021037 (2022).
[38] J. Mcclean, S. Boixo, V. Smelyanskiy, R. Babbush, and H. Neven, *Barren Plateaus in Quantum Neural Network Training Landscapes*, Nat. Commun. **9**, 4812 (2018).
[39] K. Wright, K. M. Beck, S. Debnath, J. Amini, Y. Nam, N. Grzesiak, J.-S. Chen, N. Pisenti, M. Chmielewski, C. Collins *et al.*, *Benchmarking an 11-Qubit Quantum Computer*, Nat. Commun. **10**, 1 (2019).
[40] Y.-B. Sheng and L. Zhou, *Distributed Secure Quantum Machine Learning*, Sci. Bull. **62**, 1025 (2017).
[41] W. Song, Y. Lim, H. Kwon, G. Adesso, M. Wieśniak, M. Pawłowski, J. Kim, and J. Bang, *Quantum Secure Learning with Classical Samples*, Phys. Rev. A **103**, 042409 (2021).
[42] W. Li, S. Lu, and D.-L. Deng, *Quantum Federated Learning through Blind Quantum Computing*, Sci. China Phys. Mech. Astron. **64**, 1 (2021).
[43] S. Cheng, J. Chen, and L. Wang, *Information Perspective to Probabilistic Modeling: Boltzmann Machines versus Born Machines*, Entropy **20**, 583 (2017).
[44] J. Alcazar and A. Perdomo-Ortiz, *Enhancing Combinatorial Optimization with Quantum Generative Models*, arXiv:2101.06250.
[45] K. Gili, M. Mauri, and A. Perdomo-Ortiz, *Evaluating Generalization in Classical and Quantum Generative Models*, arXiv:2201.08770.
[46] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, in *International Conference on Machine Learning* (2015), pp. 448–456, https://proceedings.mlr.press/v37/ioffe15.html.
[47] B. Xu, N. Wang, T. Chen, and M. Li, *Empirical Evaluation of Rectified Activations in Convolutional Network*, arXiv:1505.00853.
[48] J. C. Spall *et al.*, *Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation*, IEEE Trans. Autom. Control **37**, 332 (1992).
[49] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, *Improved Techniques for Training GANs*, Adv. Neural Inf. Process. Syst. **29**, 2234 (2016), https://papers.nips.cc/paper/2016/hash/8a3363abe792db2d8761d6403605aeb7-Abstract.html.
[50] A. Borji, *Pros and Cons of GAN Evaluation Measures*, Computer Vision and Image Understanding **179**, 41 (2019).
[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 2818–2826, 10.1109/CVPR.2016.308.