Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/TQE.2020.DOI

Quantum Generative Models for Small Molecule Drug Discovery

JUNDE LI[1], RASIT O. TOPALOGLU[2] (SM, IEEE), AND SWAROOP GHOSH[1] (SM, IEEE)
[1]Pennsylvania State University, PA 16802 USA
[2]IBM Inc., NY 12533 USA

Corresponding author: Junde Li (email: jul1512@psu.edu).

The work is supported in part by National Science Foundation (OIA-204066 and DGE-2113839) and seed grants from Penn State Institute for Computational and Data Sciences and Penn State Huck Institute of the Life Sciences. We acknowledge the use of IBM Quantum Services for this work. We also thank Mehrdad Mahdavi, Nikolay V. Dokholyan, Congzhou M Sha, and Jian Wang for helpful discussions.

ABSTRACT

Existing drug discovery pipelines take 5-10 years and cost billions of dollars. Computational approaches aim to sample from regions of the whole molecular and solid-state compounds called chemical space which could be on the order of 10^{60}. Deep generative models can model the underlying probability distribution of both the physical structures and property of drugs and relate them nonlinearly. By exploiting patterns in massive datasets, these models can distill salient features that characterize the molecules. Generative Adversarial Networks (GANs) discover drug candidates by generating molecular structures that obey chemical and physical properties and show affinity towards binding with the receptor for a target disease. However, classical GANs cannot explore certain regions of the chemical space and suffer from training instabilities. The practical utility of such models is limited due to the vast size of the search space, characterized by millions of parameters. A full quantum GAN may require more than 90 qubits even to generate small molecules with up to 9 heavy atoms. The proposed QGAN-HG model is composed of a hybrid quantum generator that supports various number of qubits and quantum circuit layers, and, a classical discriminator. QGAN-HG with less than 20% of the original parameters can learn molecular distributions as efficiently as its classical counterpart. Another extended version of the proposed QGAN-HG, that utilizes multiple quantum sub-circuits, considerably accelerates our standard QGAN-HG training process and avoids the potential gradient vanishing issue of deep neural networks.

INDEX TERMS

Algorithms, noisy intermediate-scale quantum algorithms and devices

I. INTRODUCTION

The drug development pipeline consists of stages of target discovery, molecular design, preclinical studies, and clinical trials, which makes the process of creating a marketable drug expensive and time consuming [1]. The majority of new drugs approved by US Food and Drug Administration are small-molecule drugs whose structural and functional diversity make their matching with biological binding sites possible [2]. Searching new drugs can be considered as navigating through the chemical space, which is an ensemble of all organic molecules. Navigation in unknown chemical space falls within the field of *de novo* drug design [3].

Machine learning techniques have been explored in all development stages, especially molecular design with desirable properties [1], [4], [5]. Generative models such as, variational autoencoders (VAEs) [6], generative adversarial networks (GANs) [7] and recurrent neural networks (RNNs) are specifically adopted for learning latent representations of molecules and generating large amount of drug candidates for further high-throughput screening. Deep generative models have been used for various representation types of molecules such as, string-based, graph-based and shape/structure-based [8]–[12] representations. Generative learning with graph-structured molecules is invariant to the orderings of atoms [10], [13] and automates the navigation to a chemical region to which the molecules in training set are close to. Quantum generative models have the relative advantage to cover the entire distribution because of the intrinsically probabilistic nature. The drug discovery process can be explained using a lock and key model where the receptor (a protein binding site associated with a disease) is considered as a lock and the drug is a key (Fig. 1(a)).

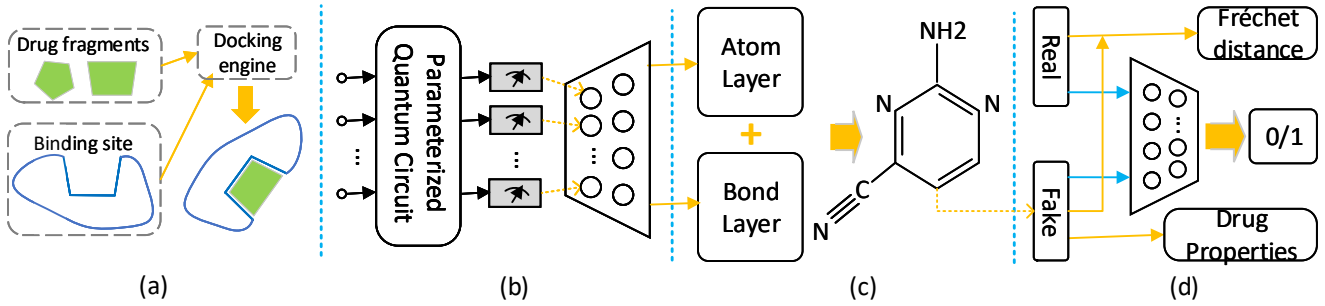

FIGURE 1. (a) Drug discovery process that shows protein binding site and drug fragments. Only generated molecules that have high affinity towards the receptor binding sites are considered as valid (evaluated by the docking engine); (b) quantum stage (which is a parameterized quantum circuit with last-layer N measuring the expectation values) and classical stage (neural network with last-layer out-feature dimension of 512) separated by blue dotted line; (c) application of atom layer and bond layer for generating atom vectors and bond matrices from which synthetic molecular graphs are reconstructed (one example generated molecule is shown); (d) a batch of real molecules sampled from training dataset (QM9 in this case) and a batch of synthetic (aka fake in GAN community) molecules generated from (c) are fed into classical discriminator for prediction of being real or fake and calculation of Frchet Distance (FD) score (which measures the divergence between real and fake molecules), and drug properties for synthetic molecules are evaluated using RDKit package. The prediction losses from discriminator are back-forwarded to two neural networks as well as quantum circuit for updating all parameters simultaneously in each training epoch.

If a molecule is structurally complementary to the binding site, the molecule plugs into the binding site and undergoes a chemical reaction.

Quantum computing can offer unique advantages over classical computing in many areas such as, chemistry simulation, machine learning, and optimization [14]–[16]. Quantum GAN is one of the main applications of near-term quantum computers due to its strong expressive power in learning data distributions even with much less parameters compared to classical GANs [17]. Quantum GANs can offer several opportunities e.g., (i) stronger expressibility and learning speedup making it possible to learn richer representation of molecules; (ii) ability to search exponentially growing chemical space with increasing qubit count and sample from distributions that may be difficult to model classically.

Quantum GAN is still at its nascent stage due to qubit constraints on noisy quantum computers. Huang et al. [18] proposed a quantum patch GAN mechanism to efficiently use limited qubits for generating hand-written digit images, however, the method only suits two digits of 0 and 1. QuGAN [17] aims to learn a data set of hand-written digits, but the original 784-dimension images were reduced to only 2 dimensions. No known existing quantum GAN mechanisms can essentially solve real-world complicated learning tasks.

Drug molecules can be represented as graphs where the nodes and edges correspond to atoms and bonds, respectively. Given the task complexity of learning molecule distribution, full quantum GAN can hardly encode all training data quantum mechanically. Take the small molecule dataset QM9 [19] for example. The total number of qubits required for reconstructing synthetic molecules is $\log_2 5^{36} + \log_2 5^9 > 90$ where 36 is the number of bonds, 9 is the number of atoms, and 5 is the number of bond types and atom types contained in QM9. At present, no commercially available gate-based quantum computers excluding quantum annealers support over 90 qubits for developing the variational quantum GAN algorithms. However, a hybrid GAN model, realized by connecting quantum measurement outcome with a following neural network (Fig. 1(b)), uses fewer qubits and still exploits the benefits of quantum computing.

We propose a qubit-efficient quantum GAN mechanism with hybrid generator and classical discriminator for efficiently learning molecule distributions based on classical MolGAN [10]. Since the proposed quantum GAN requires fewer qubits, simulation is still a viable option for training unlike full quantum GAN with large number qubits that cannot be simulated using classical computers. We also examine the patched circuit idea [18] by comparing to the original single large generator circuit implementation using metric of Frchet Distance and drug property scores. Fig. 1 shows the overall workflow of our qubit-efficient hybrid quantum GAN model for drug discovery. This work discovers a library of novel and valid molecules that can be screened by the docking engine in the next step (Fig. 1(a)). To the best of our knowledge, this is the first work on drug discovery using quantum generative models.

We, (1) propose a novel quantum GAN mechanism with hybrid generator to qubit-efficiently tackle any real-world learning tasks solvable on classical GANs; (2) generate drug molecular graphs with quality similar to classical methods in terms of Frchet Distance (FD) score and drug property scores, and achieve high training efficiency with less than 20% of the original parameters; (3) design a quantum GAN with patched sub-circuits which significantly decreases training time, making the quantum algorithm efficiently executable even in simulation environment; (4) provide a new quantum paradigm for generative models which bypass the thorny training instability issues; (5) validate the capability of generating small drug molecules on real IBM quantum computers by running inference stage of QGAN-HG model.

II. BACKGROUND

A. COMPUTATIONAL DRUG DISCOVERY

Classical generative models [7], [10], [20] have been explored for discovery of drug molecules with desired properties by learning drug molecule distribution based on given chemical dataset, and the potential of quantum machine learning for drug discovery has also been depicted in [21].

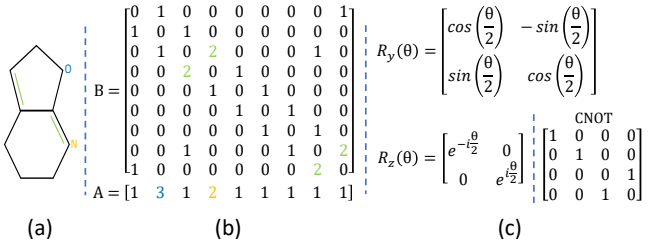

FIGURE 2. (a-b) A sample molecular graph from QM9 denoted by its corresponding atom vector A and bond matrix B; (c) all quantum gates used in this study.

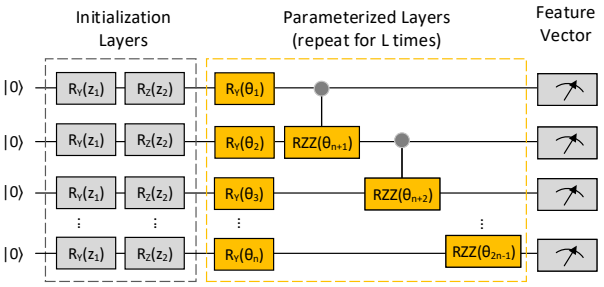

FIGURE 3. Parameterized quantum circuit to obtain feature vector of N dimensions. The circuit is composed of initialization layers, repeatable parameterized layers and measurement layer. Each RZZ gate is composed of two CNOT gates and one parametric RZ gate.

GAN consists of two networks, namely generator (G) and discriminator (D), competing with each other. The generator takes noise as input to generate a synthetic data sample whereas the discriminator flags if the sample is real or fake with a binary classifier. Generator $G(\boldsymbol{z};\theta_g)$ maps random input noise z to synthetic chemical data space p_g, while discriminator $D(\boldsymbol{x};\theta_d)$ outputs a single scalar indicating the probability that $\boldsymbol{x}$ come from real data rather than p_g. D is trained to maximize the probability to assign correct label and G is trained to minimize the difference between real and fake $\log(1 - D(G(\boldsymbol{z})))$. The two-player minimax game is trained based on the following value function:

$$\min_{\theta_g}\max_{\theta_d} V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Chemical compounds can be represented as graphs with nodes and edges designating various atoms and their bonds, respectively. For example, Fig. 2(a-b) shows a molecule and its graph representation where atom types of N and O are encoded as 2 and 3 in atom vector, and bond types of single and double are encoded as 1 and 2 in bond matrix. If the generated structure is chemically stable and exhibits high affinity towards the receptor binding sites then it can be treated as a valid drug molecule. The generator and discriminator are trained using example drugs/molecules until the synthetic chemical distribution is close to real chemical space. The quality of GAN outcome are measured by Frchet Distance and RDKit (for chemical properties) [22].

B. QUANTUM MACHINE LEARNING

Quantum systems have atypical patterns that classical computers cannot produce efficiently [14]. Machine learning tasks are sometimes hard to train on classical computers due to large-scale and high-dimensional data set. Quantum neural networks (QNNs) can represent a given dataset, either quantum or classical, and be trained using a series of parameter dependent unitary transformations. QNN architecture is dependent on qubit count, quantum circuit layer, and quantum gates applied because the architecture is essentially a variational quantum circuit. Therefore, the following quantum computing concepts are helpful in understanding a quantum neural network.

Quantum Circuit: Quantum circuits consist of gates that modulate the state of the qubits to perform computation. Quantum gates could be applied on 1 qubit (e.g., rotation gates R_y and R_z) or 2 qubits (e.g., CNOT gate), as shown in Fig. 2(c). Finally, measurement is applied to obtain the expectation value after certain number of shots. Fig. 3 displays the quantum circuit used in QGAN-HG. A *patched quantum circuit* [18] refers to multiple separate quantum sub-circuits whose measurement results are concatenated to construct an output vector of larger dimension. The patched quantum circuit creates no entanglement between qubits in different sub-circuits but allows us to simulate large quantum systems efficiently on classical machines.

Quantum Noise: Noisy Intermediate-Scale Quantum (NISQ) computers [23] suffer from noise sources such as, T1 relaxation time, T2 dephasing time, gate errors and readout errors. These are also called qubit quality metrics. Crosstalk, qubit-to-qubit variation and temporal variations in qubit quality also exist. Low quantum noise level is not detrimental for QNNs, rather it can even be beneficial for quantum machine learning applications as shown in [24], [25]. However, its potential benefits in QGAN-HG models are not specifically evaluated since it is not the focus of the present study and could be a separate work. Instead, noise resilience of quantum GANs is verified by running the inference for the patched version of QGAN-HG on a real IBM quantum computer.

III. QUANTUM GENERATIVE ADVERSARIAL NETWORKS

A. QUANTUM GAN FLAVORS

Quantum GAN has a few flavors of the generator and discriminator implementations depending on their execution environments, either on quantum computers, classical machines or quantum simulators. The flavor with quantum discriminator is not applicable here due to limited number of qubits on near-term quantum computers. Real data shown in Fig. 1(d) has to engage the state preparation stage, usually through amplitude encoding, for embedding the classical data in a quantum state which takes $N\log(M)$ qubits where N is the training set size and M is feature dimension [18], [26]. The flavor with a pure quantum generator is not directly applica-

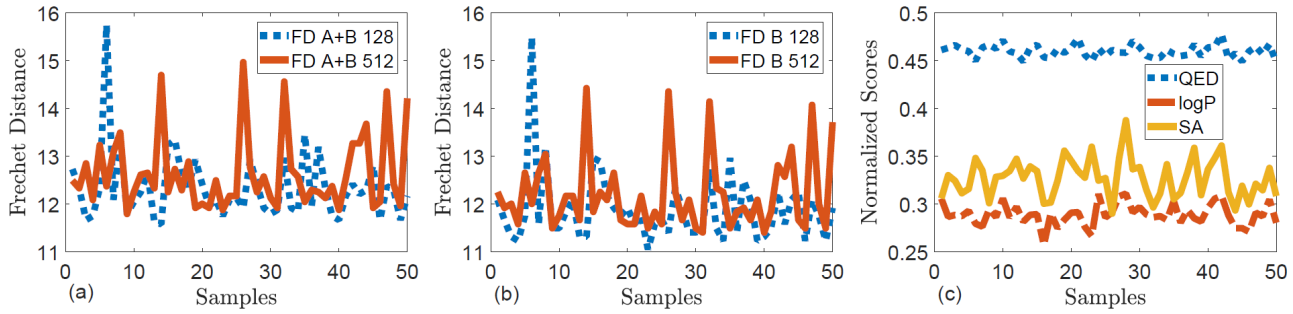

FIGURE 4. Metrics of Frchet Distance and molecule properties for real data points: (a) Frchet Distances (with lower bound of zero) calculated based on 90 dimensional samples with both Atom and Bond (A+B) for batch sizes of 128 and 512; (b) Frchet Distances calculated based on 81 dimensional samples with only Bond (B) for batch sizes of 128 and 512; (c) three major drug properties (QED - druglikeness, logP - solubility, SA - synthesizability) evaluated on sample batches of 128 molecules from (a). Note that these 50 samples are independently sampled from QM9.

ble either considering the complicated task of drug discovery. As noted in Section I, more than 90 qubits are needed to discover QM9-like molecules (not suitable for near-term quantum computers). Thus we propose a new quantum GAN with hybrid generator and classical discriminator to exploit the quantum benefits.

B. QUANTUM GAN WITH HYBRID GENERATOR

Quantum GAN with hybrid generator (QGAN-HG) is composed of a parameterized quantum circuit to get a feature vector of qubit size dimension, and a classical deep neural network to output an atom vector and a bond matrix for the graph representation of drug molecules. Another patched quantum GAN with hybrid generator (P-QGAN-HG) is considered as the variation of QGAN-HG where the quantum circuit is formed by concatenating few quantum sub-circuits.

QGAN-HG Quantum Circuit: In this variant, a quantum layer is added for exploiting the strong expressive power of variational quantum circuits which perform low-rank matrix computations in $\mathcal{O}(poly(\log(M)))$ time [18], [27]. The variational quantum circuit (Fig. 3) consists of 3 stages, namely initialization, parameterized (repeatable for L layers with $L(2N-1)$ parameter count) and measurement stages. In the parameterized layers, each single-qubit RY gate contains an angle parameter, and each two-qubit RZZ gate, following the naming convention of IBM [28], has one angle parameter. Two parameters z_1 and z_2 are uniformly sampled from $[-\pi, \pi]$, which essentially substitute the random Gaussian noise input for classical GANs. After applying the initialization layers, the input state in mathematical form $|\psi_0(z_1, z_2)\rangle = (R_Z(z_2)R_Y(z_1)|0\rangle)^{\otimes N}$ is prepared. Let us denote the parameterized layers repeated for L times as unitary matrix $U(\boldsymbol{\theta})$. The final quantum state is of the form $|\Psi\rangle = U(\boldsymbol{\theta})|\psi_0(z_1, z_2)\rangle$. A series of measurement operators are applied to obtain the expectation value for each qubit and further form the feature vector to be fed to classical neural network.

QGAN-HG Neural Network: The classical stage of hybrid generator is a standard neural network with input layer receiving the feature vector of expectation values. The final layer consists of the separate atom and bond layers for creating atom vectors and bond matrices, respectively. Like MolGAN [10], a categorical re-parameterization step with Gumbel-Softmax [29], which supports gradient calculation in the backward pass, is taken to obtain discrete fake molecular graphs. Note that, 85.07% and 98.03% of generator parameters are dropped by reducing major linear layers from classical GAN [10] to demonstrate the strong expressive power of quantum circuits. Due to the necessity of reconstructing QM9-like molecules (with structure of $\boldsymbol{X} \in \mathbb{R}^{9X5}$ for atoms and $\boldsymbol{A} \in \mathbb{R}^{9X9X5}$ for bonds), neural network architecture can hardly be further reduced.

Patched QGAN-HG: The patched quantum GAN with hybrid generator consists of the same two ingredients as above. The patched quantum circuit splits the circuit into independent multiple quantum sub-circuits by removing the entangled RZZ gates among sub-circuits. Then the patched circuit is integrated by concatenating the expectation vales from each sub-circuit [18]. Theoretically, P-QGAN-HG has its pros and cons relative to QGAN-HG with an integral quantum circuit. P-QGAN-HG requires less quantum resources because multiple sub-circuits can be executed sequentially or in parallel. Another benefit is that each circuit can be simulated more efficiently, which speeds up the learning process accordingly. However, one of the obvious drawbacks is reduced expressive power since quantum state dimension is reduced from 2^N to $2^{N/2}$ (say two circuits with half size) in Hilbert space. The performance of patched QGAN-HG is compared with QGAN-HG in the following section.

Discriminator and Optimizer: The discriminator is kept the same as MolGAN [10] since its parameter size is on par with the hybrid generator. However, the reward network is discarded since the reward value is too minuscule to noticeably contribute to training the model. Generated molecules are evaluated using RDKit together with Frchet Distance based metric. Quantum gate parameters and weights in neural network are updated simultaneously using a single optimizer while discriminator employs a separate one to be updated alternatively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TQE.2021.3104804, IEEE Transactions on Quantum Engineering

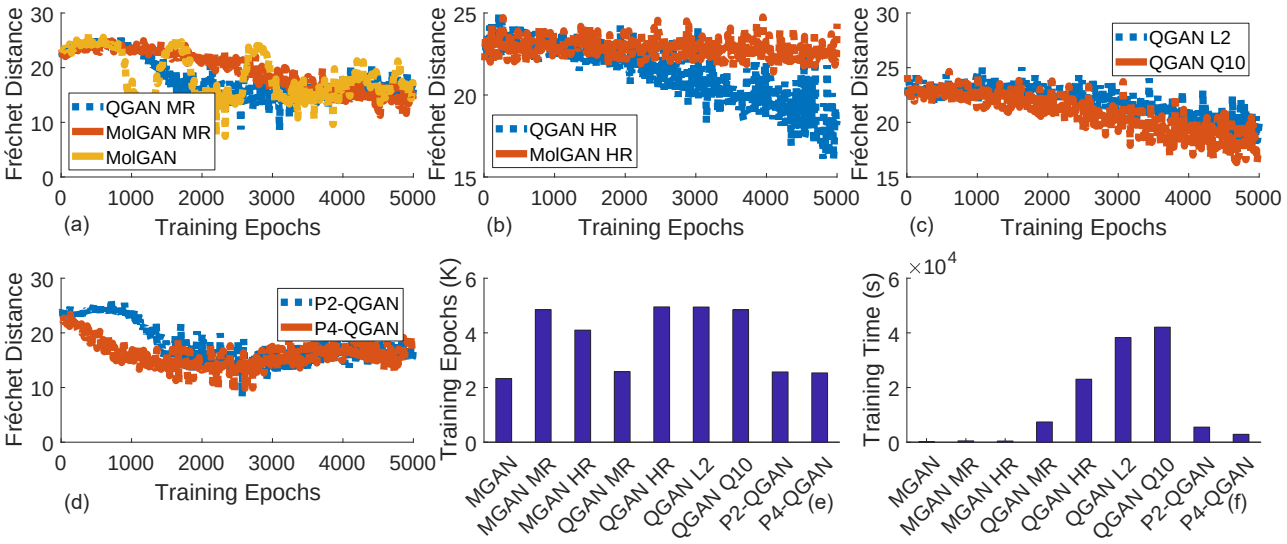

FIGURE 5. Training comparison among various GAN flavors: (a) Frchet Distances for MolGAN, moderately reduced (14.93%) MolGAN and QGAN-HG; (b) Frchet Distances for highly reduced (1.97%) MolGAN and QGAN-HG; (c) learning curves for highly reduced QGAN-HG with quantum circuit level $L = 2$ and $N = 10$, respectively; (d) learning curves for patched QGAN-HG with two sub-circuits and four sub-circuits; (e) training epochs with lowest Frchet Distances for all GAN flavors; (f) training times elapsed (early stopping at epochs from (e)) for all GAN flavors. All QGAN variants in (a, b, d) are composed of 8 qubits and a single quantum layer. Abbreviation: QGAN/MolGAN MR - QGAN/MolGAN with moderately reduced generator; QGAN/MolGAN HR - QGAN/MolGAN with highly reduced generator; QGAN L2 - QGAN with quantum circuit depth of 2; QGAN Q10 - QGAN with 10-qubit quantum circuit; P2/4-QGAN - Patched QGAN with 2/4 sub-circuits; MGAN - MolGAN.

IV. EXPERIMENTAL SETUP

Dataset, metrics and implementation details that support the findings of this study are available in the GitHub repository https://github.com/jundeli/quantum-gan.

A. DATASET AND METRICS

Dataset: All the experiments are conducted with quantum machine learning benchmarking QM9 [19] dataset which contains 133,885 molecules with up to 9 heavy atoms of types of carbon, nitrogen, oxygen, and fluorine.

Frchet Distance: Learning results of the proposed GANs are evaluated with Frchet Distance metric which measures the similarity between real and generated molecule distributions. Generated molecule distribution is approximately created by generating a batch of molecules, and real one is approximately formed by randomly sampling the same number of molecules from QM9 dataset. Each sample batch of molecules is concatenated and considered as a multi-dimensional point in the distribution, then Frchet Distance is calculated using 50 of these points (sampled for 50 batches from both distributions). Fig. 4 shows Frchet Distances calculated for two batch sizes of 128 and 512, and molecule batches are independently sampled for 50 times. FD A+B (see Fig. 4(a)) is calculated based on 90 dimensional samples with both atom vectors (9 entries) and bond matrices (81 entries); while FD B (see Fig. 4(b)) based on 81 dimensional samples with only bond matrices. FD A+B includes more random noises due to extra 9 dimensional atom vectors and is projected to severely disturb the similarity between real molecule batches. Interestingly, FD A+B correlates well with the FD B, indicating the strong inherent connection between atoms and bonds. The means (12.3342, 12.6387) and variations (0.7057, 0.7849) between FD calculated for 128 and 512 batch sizes are close. Generated molecules are considered realistic enough once the preset FD cut-off point of 12.5 is reached during the learning process. Thus, all following experiments are evaluated with FD A+B metric and 128 batch size.

Drug Properties: Molecule properties are the metrics for drug quality evaluation during inference stage. Three primary properties include, (i) quantitative estimate of druglikeness (QED) which measures the likelihood of compound being a drug; (ii) log octanol-water partition coefficient (logP) which measures the solubility of a compound; and (iii) and synthetic accessibility (SA) which quantifies the ease of a compound being synthesized in pharmaceutical factory. Fig. 4 (c) shows the property scores normalized to range [0, 1]. The three curves display the possible ranges of real drug properties in terms of QED, logP, and SA which are also indicators for the learning quality of all GAN variants discussed in this study. Together with other properties, they are measured using RDKit.

B. IMPLEMENTATION DETAILS

The quantum circuits can be executed either on a simulator or real quantum machine. The simulator supports customized setting of noise levels and sources (noiseless environment is set in this paper), while real quantum devices have different noise characteristics across different machines.

Training: We modify the classical MolGAN [10] to implement our QGAN-HG and P-QGAN-HG algorithms. As mentioned in Section III, some linear layers and reward

network are dropped in our experiments to evaluate the expressive power of quantum circuit and to use drug property metrics fairly. Our QGAN variations are trained with a mini-batch of 128 molecules using the Adam optimizer on a single RTX 2080 Ti GPU for the classical part and the PennyLane platform [30] with the default qubit plugin for the quantum stage. As explained in Section I, the real quantum machine is not utilized during the training stage due to the long queuing time on IBM Q machines. The learning rate is initially set to 0.0001 for both generator and discriminator and starts decaying uniformly at a factor of $1/2000$ after 3000 epochs. Total training epoch is set with 5000, and early stopping based on Frchet Distance is applied if model collapse happens.

Inference: Only hybrid generator is involved during inference stage. Since QGAN is well trained, the quantum circuit in the generator is executed on both PennyLane simulator and real quantum device of *ibmq_quito* [28] for comparison. Drug quality for generated molecules are evaluated by properties such as, QED, logP, and SA, among others, all of which are normalized to be within $[0, 1]$. Finally, GAN variations are compared by taking 1000 generated molecules.

V. EVALUATION RESULTS

A. QGAN-HG RESULTS

Fig. 3 shows QGAN-HG performance may rely on both qubit count N and repeatable layer count L. Higher qubit count and layer count presumably correspond to stronger expressive power of hybrid generator. We reduce the neural network parameter count to two levels, namely, 14.93% (MR-moderately reduced) and 1.97% (HR-highly reduced) of generator parameters of original MolGAN to demonstrate the expressive power of the hybrid generator. Fig. 5(a-b) show the training performance comparison between MolGAN and QGAN-HG for moderately and highly reduced architectures, respectively. One can observe from Fig. 5(a) that all mechanisms can achieve a reasonably good training point (see Fig. 4 for benchmark) within 5000 epochs. However, moderately reduced MolGAN takes around 4000 iterations while baseline MolGAN and QGAN-HG take only 2500 iterations or so. Also note that, MolGAN and QGAN-HG both reach a slightly lower Frchet Distance than the reduced classical counterpart. As shown in Fig. 5(b), MolGAN with highly reduced architecture can hardly be trained though a slight downward trend is observed. The weak learning ability of MolGAN-HR is attributed to mainly two reasons: (1) the features of QM9 drug molecules cannot be well represented using a light-weight neural network; (2) parameter count of generator is not at par with that of discriminator. Intriguingly, a sharp downward learning curve for QGAN-HG is still observed. It is worth mentioning that only 15 gate parameters are used in the quantum circuit. These are clear evidences of strong expressive power of variational quantum circuits.

Model collapsing occurs for GAN variations if Frchet Distances (approximate indicator of learning quality) start increasing after a certain training point. We adopt early stopping technique (the lowest reached point of Frchet Distance) to somewhat prevent the training instability issue in GANs. To measure the effects of circuit layer and qubit count, we also implement QGAN-HG with $L = 2$ and $N = 10$ separately. All other quantum variants are configured with default single layer $L = 1$ and 8 qubits $N = 8$, if not specified otherwise. However, the enhanced QGAN-HG variants with more circuit layer and qubits (see Fig. 5(c, e)) do not help accelerate learning process much relative to the QGAN-HG in Fig. 5(b).

B. PATCHED QGAN-HG (P-QGAN-HG) RESULTS

The patched QGAN-HG mechanism is developed on the basis of [18]. However, the proposed P-QGAN-HG uses all qubits for creating feature vector and has no specific qubits for non-linear mapping because of the following classical neural network. We demonstrate the expressive power of two patched QGAN-HG variants, i.e., P2-QGAN with two sub-circuits (each has 4 qubits and 7 gate parameters) and P4-QGAN with four sub-circuits (each has 2 qubits and 3 gate parameters). These gate parameters refer to the theta angles shown in the parameterized layers of Fig. 3. Surprisingly, the learning quality of these patched QGANs are comparable to QGAN with an integral circuit, as shown in Fig. 5(a, d), though patched QGANs have even fewer gate parameters. Further, the simulation time (see Fig. 5(f)) for patched quantum circuits are significantly reduced because of smaller qubit count and early convergence. Therefore, we consider patched QGAN-HG with multiple sub-circuits is an alternative to classical GAN since GAN training issues such as instability and vanishing gradients can be mitigated by shortening neural network depth. When a larger variational quantum circuit is adopted, the quantum algorithms would be further accelerated if executed on a dedicated real quantum computer.

C. DRUG PROPERTIES

The training of GAN variants is evaluated by Frchet Distance, whereas the quality of drug molecules generated from GANs is specifically evaluated using a series of molecule property metrics, three of which are visualized in Fig. 4(c). The drug property evaluation is performed by a specific inference stage. All GAN variations pick a point with lowest Frchet Distance within 5000 epochs for inference. We run the inference stage for QGAN-HG on IBM Q quantum machines as well. Drug properties are calculated using 1000 sampled molecules. Table 1 displays the drug property results which are generally consistent with Frchet distance results. Note that diversity and novel scores for all models are high, whereas synthetic accessibility is low relative to benchmark shown in Fig. 4(c). P2-QGAN-HG results executed on the simulator and the real quantum computer are quite close, indicating the noise resilience of QGAN algorithms. Though the benchmark MolGAN shows higher valid score, the performances of the following GAN variants, MolGAN, MolGAN MR, QGAN-HG MR, QGAN-HG HR Q10, P2-QGAN-HG MR, P4-QGAN-HG MR, and P2-QGAN-HG

TABLE 1. Drug properties evaluated from 1000 generated molecules for all GAN variations with complexity. Best results are shown in bold. Zeros indicate all generated molecules are either reconstructed with NONE type by RDKit, or not drug-like, soluble and synthesizable. Values in parenthesis denote the standard deviations for corresponding properties. Novelty scores for all GAN variants are 1, thereby not showing in an extra column. Note that, the last column displays the model complexities for all GAN variants. (n QGP) denotes the corresponding model has n extra quantum gate parameters.

Method	QED	logP	SA	Diversity	Validity	Uniqueness	Model Complexity
MolGAN* [10]	**0.51** (0.05)	0.66 (0.15)	0.08 (0.04)	**1.0** (0.00)	**0.86**	0.18	1.0 (0 QGP)
MolGAN MR	0.48 (0.05)	**0.70** (0.17)	**0.12** (0.11)	**1.0** (0.01)	0.41	0.61	0.15 (0 QGP)
MolGAN HR	0	0	0	**1.0** (0.00)	0.10	**1.0**	**0.02** (0 QGP)
QGAN-HG MR (proposed)	**0.51** (0.06)	0.49 (0.15)	0.11 (0.09)	**1.0** (0.00)	0.44	0.54	0.15 (15 QGP)
QGAN-HG HR (proposed)	0	0	0	0.99 (0.02)	0.29	**1.0**	**0.02** (15 QGP)
QGAN-HG HR L2 (proposed)	0	0	0	**1.0** (0.00)	0.04	**1.0**	**0.02** (30 QGP)
QGAN-HG HR Q10 (proposed)	0.48 (0.03)	0.46 (0.09)	0.07 (0.01)	0.97 (0.09)	0.04	**1.0**	**0.02** (19 QGP)
P2-QGAN-HG MR (proposed)	0.49 (0.05)	0.62 (0.19)	**0.12** (0.10)	**1.0** (0.02)	0.52	0.41	0.15 (14 QGP)
P4-QGAN-HG MR (proposed)	0.49 (0.05)	0.50 (0.19)	**0.12** (0.09)	**1.0** (0.00)	0.57	0.46	0.15 (12 QGP)
P2-QGAN-HG MR (on *ibmq_quito*)	0.49 (0.05)	0.61 (0.19)	0.11 (0.08)	**1.0** (0.01)	0.48	0.43	0.15 (14 QGP)

MolGAN* refers to MolGAN [10] trained in the present study.

MR (on *ibmq_quito*), are similar in terms of the drug properties for generated molecules because of the tradeoff between validity and uniqueness, indicating the lowest unique score for MolGAN. Drug quality is evaluated by considering all metrics rather than a single metric.

D. SUMMARY OF OUR FINDINGS

Our findings can be summarized as follows:

- Classical MolGAN is cumbersome and requires millions of parameters to learn molecular distributions.
- Classical MolGAN with 85.07% reduced parameters (i.e., moderately reduced MolGAN) cannot learn properly, however our QGAN-HG with only 15 extra quantum gate parameters combined with reduced MolGAN can achieve the same accuracy.
- Proposed QGAN-HG with 98.03% reduced parameters (i.e., highly reduced MolGAN) is significantly efficient than the highly reduced MolGAN.
- The patched quantum GAN achieves comparable learning accuracy in terms of drug properties with only 12 extra gate parameters, and considerably accelerates the quantum GAN with an integral circuit.
- Moderately reduced MolGAN takes more than 5000 epochs for convergence while its quantum counterpart takes around 2500 epochs (i.e., quantum version converges faster). Highly reduced MolGAN hardly learns but its quantum version shows good learning quality.
- Inferencing in real quantum hardware shows similar performance as simulations.
- Our study provides a new quantum paradigm for GANs to alleviate possible gradient vanishing problem in neural networks due to short chain of gradients.

VI. CONCLUSION

We propose a novel quantum GAN with a hybrid generator for discovery of new drug molecules. Our hybrid GAN with patched quantum circuits concatenates feature vectors from different patches. We propose several variants of hybrid GAN namely, moderately reduced QGAN, highly reduced QGAN, advanced QGAN with 2 layers and 10 qubits, highly reduced QGAN with 2 quantum patches and 4 patches, and compare them with benchmark classical MolGAN. Hybrid quantum generative models can learn complex data distributions on near-term quantum computers since they are noise resilient. Phenomenal representation power of quantum circuit speeds up the progress of learning molecular distribution in terms of training epochs. Light QGAN models with shallow depth are achieved by reducing up to 98.03% of generator parameters, which also helps preventing possible training issue of vanishing gradients in classical neural networks.

REFERENCES

[1] C. Grow, K. Gao, D. D. Nguyen, and G.-W. Wei, "Generative network complex (gnc) for drug discovery," arXiv preprint arXiv:1910.14650, 2019.

[2] R. F. Beall, T. J. Hwang, and A. S. Kesselheim, "Pre-market development times for biologic versus small-molecule drugs," Nature Biotechnology, vol. 37, no. 7, pp. 708–711, 2019.

[3] J.-L. Reymond and M. Awale, "Exploring chemical space for drug discovery using the chemical universe database," ACS chemical neuroscience, vol. 3, no. 9, pp. 649–657, 2012.

[4] S. Ekins, A. C. Puhl, K. M. Zorn, T. R. Lane, D. P. Russo, J. J. Klein, A. J. Hickey, and A. M. Clark, "Exploiting machine learning for end-to-end drug discovery and development," Nature materials, vol. 18, no. 5, p. 435, 2019.

[5] M. Batool, B. Ahmad, and S. Choi, "A structure-based drug discovery paradigm," International journal of molecular sciences, vol. 20, 2019.

[6] D. P. Kingma and M. Welling, "Auto-encoding variational bayes." in Proceedings of the International Conference on Learning Representations (ICLR), 2014.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014.

[8] G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, and A. Aspuru-Guzik, "Objective-reinforced generative adversarial networks (organ) for sequence generation models," arXiv preprint arXiv:1705.10843, 2017.

[9] M. J. Kusner, B. Paige, and S. Gelly, "Jm herná ndez-lobato. grammar variational autoencoder," in Proceedings of the 34th International Conference on Machine Learning, ICML, vol. 17.

[10] N. De Cao and T. Kipf, "Molgan: An implicit generative model for small molecular graphs," arXiv preprint arXiv:1805.11973, 2018.

[11] M. Skalic, D. Sabbadin, B. Sattarov, S. Sciabola, and G. De Fabritiis, "From target to drug: Generative modeling for the multimodal structure-based ligand design," Molecular pharmaceutics, vol. 16, no. 10, pp. 4282–4291, 2019.

[12] M. Skalic, J. Jiménez, D. Sabbadin, and G. De Fabritiis, "Shape-based generative modeling for de novo drug design," Journal of chemical information and modeling, vol. 59, no. 3, pp. 1205–1214, 2019.
[13] M. Simonovsky and N. Komodakis, "Graphvae: Towards generation of small graphs using variational autoencoders," in International Conference on Artificial Neural Networks. Springer, 2018, pp. 412–422.
[14] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," Nature, vol. 549, no. 7671, pp. 195–202, 2017.
[15] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry," Reviews of Modern Physics, vol. 92, no. 1, p. 015003, 2020.
[16] G. Nannicini, "Performance of hybrid quantum-classical variational heuristics for combinatorial optimization," Physical Review E, vol. 99, no. 1, p. 013304, 2019.
[17] S. A. Stein, B. Baheri, R. M. Tischio, Y. Mao, Q. Guan, A. Li, B. Fang, and S. Xu, "Qugan: A generative adversarial network through quantum states," arXiv preprint arXiv:2010.09036, 2020.
[18] 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," arXiv preprint arXiv:2010.06201, 2020.
[19] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," Scientific Data, vol. 1, 2014.
[20] M. Ragoza, T. Masuda, and D. R. Koes, "Learning a continuous representation of 3d molecular structures with deep generative models," arXiv preprint arXiv:2010.08687, 2020.
[21] Y. Cao, J. Romero, and A. Aspuru-Guzik, "Potential of quantum computing for drug discovery," IBM Journal of Research and Development, vol. 62, no. 6, pp. 6–1, 2018.
[22] "Rdkit: Open-source cheminformatics software," https://www.rdkit.org/.
[23] J. Preskill, "Quantum computing in the nisq era and beyond," Quantum, vol. 2, p. 79, 2018.
[24] Y. Du, M.-H. Hsieh, T. Liu, D. Tao, and N. Liu, "Quantum noise protects quantum classifiers against adversaries," Physical Review Research, vol. 3, no. 2, p. 023153, 2021.
[25] N. H. Nguyen, E. C. Behrman, and J. E. Steck, "Quantum learning with noise and decoherence: a robust quantum neural network," Quantum Machine Intelligence, vol. 2, no. 1, pp. 1–15, 2020.
[26] M. Plesch and Č. Brukner, "Quantum-state preparation with universal gate decompositions," Physical Review A, vol. 83, no. 3, p. 032302.
[27] S. Lloyd and C. Weedbrook, "Quantum generative adversarial learning," Physical review letters, vol. 121, no. 4, p. 040502, 2018.
[28] "IBM Quantum." 2021. [Online]. Available: https://quantum-computing.ibm.com/
[29] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," arXiv preprint arXiv:1611.01144, 2016.
[30] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, C. Blank, K. McKiernan, and N. Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," arXiv preprint arXiv:1811.04968, 2018.