Response To Reviewer #2.

Running on real-world quantum hardware: We note that publicly available machines are less powerful for useful demonstrations mainly due to their size limit (# of gates and qubits). To include the real-world noise model in our simulations, in lines 271-278, we describe exactly the same type of *noisy* simulation from one ion-trap group.

Real-world applications of proposed quantum WGAN. In the revised version of the paper, we will add a real-world application of the quantum WGAN suggested by Reviewer 2. The specific task is to approximately implement large quantum circuits (denoted by U_0) by smaller ones (denoted by U_1). The connection is as follows: to approximate U_0 on, e.g., $|0\rangle$, quantum WGAN can find a more succinct generator U_1 s.t. $U_1 |0\rangle \approx U_0 |0\rangle$. To approximate on all inputs, we use the quantum state-channel isomorphism (i.e.,the Choi-Jamiołkowski state), which is $|\Psi_0\rangle = U_0 \otimes I |\Phi\rangle$ where $|\Phi\rangle$ is the maximally entangled state. It suffices to find a more succinct generator U_1 such that $|\Psi_1\rangle = U_1 \otimes I |\Phi\rangle \approx |\Psi_0\rangle$. The fidelity between $|\Psi_0\rangle$ and $|\Psi_1\rangle$ then becomes the average output fidelity over uniformly chosen inputs to U_0/U_1. Specifically, we studied the quantum Hamiltonian simulation circuit for 1-d 3-qubit Heisenberg model (in Eqn. (1) of arXiv:1711.10980v1). The best-known quantum circuit with the worst case error 10^{-3} (in operator norm) has over 11,900 gates. Using the above approach and our quantum WGAN (for 6-qubit), we discovered a circuit U_1 with 52 gates with an average output fidelity over 0.9999 and a worst-case error 0.15. The worst-case input is not realistic in current experiments and hence the high average fidelity implies very reasonable approximation in practice. This task could only be achieved using our quantum WGAN, rather than previous quantum GAN proposals, given its complexity.

Response To Reviewers #1 and #3.

Differences between classical and quantum data/sampling. We want to emphasize that the quantum extension of WGAN was *not* a straightforward extension of WGAN as suggested by Reviewer 3, due to the essential difference between quantum and classical data. Consider a classical random bit b with density $(0.4(b = 0), 0.6(b = 1))$. A classical readout (or sample) refers to a random variable with this distribution. In quantum mechanics, these are two *separate* concepts. An operator extension of density, called the *density* operator (semidefinite operators with trace 1, lines 135-150), represents an ensemble of *quantum data*, which includes information of both pure quantum states (as unit vectors) and their density. A classical readout on quantum states refers to a *quantum measurement* (lines 439-449). When measuring density operator Q using observable ψ, its outcome is a random variable with expectation $\mathrm{Tr}(Q\psi)$. Classical random bit $(0.4, 0.6)$ is simply a $\mathrm{diag}(0.4, 0.6)$ density operator and there is only one allowed measurement in classical mechanics. Hence, there is no distinction between these two concepts for classical data. A quantum bit (qubit) refers to a 2×2 density operator with potentially complicated off-diagonal terms. Moreover, one can have many measurements for one quantum data. This justifies why density operators represent the entity of quantum data.

The outcome of a quantum generator must hence be mathematically represented by a single density operator. A classical random bit can also be represented by a diagonal density operator, although it might not be very intuitive in the first use.

Cost function and the geometry of the sample space in qWGAN. The definition of cost function for quantum data must work with density operators. Let us first formulate the classical cost function (2.1) in the density operator form. Consider one random bit and choose c(0,0)=c(1,1)=0 and c(0,1)=c(1,0)=1. Then (2.1) becomes $\sum_{a,b\in\{0,1\}}$ $\pi(a, b)c(a, b)$ where π is the coupling of two random bits, which is mathematically the same as $\mathrm{Tr}(\pi C)$ where $\pi = \mathrm{diag}(\pi(0, 0), \pi(0, 1), \pi(1, 0), \pi(1, 1))$ and $C = \mathrm{diag}(c(0, 0), c(0, 1), c(1, 0), c(1, 1))$. (Note C is independent of π.) Our (3.1) is the quantum extension of the above with important distinctions. In (3.1), π is a density operator for the quantum coupling of P and Q, with potentially very complicated off-diagonal terms. The diagonal C in the classical case does not work for off-diagonal π. It is easy to find examples of P such that $\mathrm{qW}(P, P) > 0$ with the diagonal C. Our solution is to leverage the concept of *symmetric subspace* in quantum information. The projection onto any subspace V is a matrix with eigenspace V with eigenvalue 1, and eigenspace $V^\perp$ with eigenvalue 0. The projection onto the symmetric subspace, denoted Π_{sym}=(I+SWAP)/2, has the property that $\Pi_{\mathrm{sym}} P \otimes P$ = $P \otimes P$. By choosing C to be the projection of its orthogonal subspace, i.e., C=I- Π_{sym} =(I-SWAP)/2, we have $\mathrm{qW}(P, P) = 0$ for any P. It also encodes the geometry of the space of quantum states. Choose P=$\vec{v}\vec{v}^\dagger$ and Q=$\vec{u}\vec{u}^\dagger$ and $\mathrm{Tr}(\pi C)$ becomes 0.5 $(1 - |\vec{u}^\dagger\vec{v}|^2)$, where $|\vec{u}^\dagger\vec{v}|$ depends the angle between $\vec{u}$ and $\vec{v}$ which are unit vectors representing (pure) quantum states.

Evaluation of the loss function. The generator produces a density operator Q. The loss function is evaluated by approximating terms like $\mathrm{Tr}(Q\psi)$ (lines 221-246) via measuring multiple copies of Q (via multi-run of the generator).

Comments on the evaluation and experiments: Most existing literature is not explicit in architecture, with no publicly available code/data, and has only studied the 1-qubit case (except for Ref. [3] with 6-qubit). We are the only one with a thorough numerical study up to 8 qubits, with both large generator circuits and noisy simulation. Note that the sample space for 8-qubit is already of dimension $2^8 \times 2^8$. This exponential growth limits numerical evaluation by classical simulation in quantum computing and we did reach the limit of our computing resources. Our advantage to all existing literature (especially to Ref. [3]) is demonstrated in lines 279-295. Our to-add real-word application (in response to Reviewer #2) further demonstrates the ability of qWGAN to handle complicated tasks.

In the revised version of the paper, we will address all minor comments and also add a background section on quantum information to make our results further accessible to broader audience.