





Machine-learning-inspired quantum optimal control of nonadiabatic geometric quantum computation via reverse engineering

Meng-Yun Mao ^{1,2} Zheng Cheng,^{1,2} Yan Xia ^{3,4,*} Andrzej M. Oleś ^{5,6,†} and Wen-Long You ^{1,2,‡}

¹College of Physics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

²Key Laboratory of Aerospace Information Materials and Physics (NUAA), MIIT, Nanjing 211106, China

³Fujian Key Laboratory of Quantum Information and Quantum Optics, Fuzhou University, Fuzhou 350116, China

⁴Department of Physics, Fuzhou University, Fuzhou 350116, China

⁵Max Planck Institute for Solid State Research, Heisenbergstrasse 1, D-70569 Stuttgart, Germany

⁶Institute of Theoretical Physics, Jagiellonian University, Prof. Stanisława Łojasiewicza 11, PL-30348 Kraków, Poland



(Received 23 February 2023; accepted 11 September 2023; published 27 September 2023)

Quantum control plays an irreplaceable role in practical use of quantum computers. However, some challenges have to be overcome to find more suitable and diverse control parameters. We propose a promising and generalizable average-fidelity-based machine-learning-inspired method to optimize the control parameters, in which a neural network with periodic feature enhancement is used as an ansatz. In the implementation of a single-qubit gate by cat-state nonadiabatic geometric quantum computation via reverse engineering, compared with the control parameters in the simple form of a trigonometric function, our approach can yield significantly higher-fidelity (>99.99%) phase gates, such as the $\pi/8$ gate (T gate). Single-qubit gates are robust against systematic noise, additive white Gaussian noise, and decoherence. We numerically demonstrate that the neural network possesses the ability to expand the model space. With the help of our optimization, we provide a feasible way to implement cascaded multiqubit gates with high quality in a bosonic system. Therefore, the machine-learning-inspired method may be feasible in quantum optimal control of nonadiabatic geometric quantum computation.

DOI: [10.1103/PhysRevA.108.032616](https://doi.org/10.1103/PhysRevA.108.032616)

I. INTRODUCTION

Multiqubit gates are widely used in quantum circuits [1–3], quantum error correction [4–6], and other fields [7–10]. Single-shot multiqubit gates [11–13] specify quantum circuits that evolve in a well-controlled, uninterrupted, and continuous-time way to implement the quantum computation [14]. Compared to the cascaded gates, single-shot multiqubit gates can greatly reduce the circuit depth and shorten the implementation time, thus suppressing decoherence [15,16]. However, the single-shot method is difficult to realize in experiments owing to the restricted conditions of simultaneously manipulating multiple physical systems and building complex couplings.

One of the ways to mitigate the above difficulty is the application of single- and two-qubit gates to equivalently implement the function of multiqubit gates [17–19], and the

decomposition is guaranteed by the Solovay-Kitaev theorem [20]. Although the decomposition method loses the upper hand in terms of the circuit depth and the implementation time, it has a wider scope of application due to direct execution on the quantum processing unit, namely, the brain of a quantum computer [21]. The realization of the synthetic gates depends on the universal single-qubit gates and a two-qubit entangling gate with high fidelity [16]. However, statistical imprecision in the experimental controls, interactions between the system and the environment, and random driving forces from the environment will cause a reduction in fidelity [22].

Universal single-qubit gates based on the geometric phase in quantum systems have recently shown robustness against control-parameter fluctuations [23]. An adiabatically evolving system driven by a nondegenerate Hamiltonian exhibits geometric phase under cyclic evolution [24–26]. The geometric phase arising from cyclic evolution of quantum systems is uniquely determined by the geometric structure of the enclosed path in the parameter space [27], which is the well-known analog of the rotational effect in differential geometry when a vector is parallel transported [28,29]. However, the strict condition of the adiabatic limit requires the evolution time to be infinitely long, which inevitably gives rise to decoherence of the system [30]. The nonadiabatic geometric phase gets rid of the bondage of the adiabatic condition, making it possible to shorten the evolution time of the system to a great extent. The nonadiabatic geometric phase lays a solid foundation for nonadiabatic geometric quantum

*xia-208@163.com

†a.m.oles@fkf.mpi.de

‡wlyou@nuaa.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

computation (NGQC). Recently, NGQC has been executed theoretically [31–34] and experimentally [35–39] in multiple quantum systems. Later, NGQC has been further promoted to NGQC+ [40]. The NGQC+ scheme loosens the conditions for the realization of NGQC to a certain extent, which becomes more compatible with optimal control methods. Several schemes have been developed, including counteradiabatic driving [41,42], dynamical decoupling [43,44], and machine-learning-based optimization techniques [45,46].

Recent researches have shown that logical qubit encoding is promising to protect quantum computation from errors [47–49]. However, in standard logical qubit systems based on multiple physical qubits [50–53], quantum error correction and logical operations are difficult to achieve because the number of error channels rapidly increases with the number of qubits [54]. For the realization of logical qubits, bosonic systems are promising candidates, because the number of error channels can dramatically drop [54,55] with taking advantage of the infinite-dimensional Hilbert space of the harmonic oscillator. The cat states of bosons have been widely used in quantum computation and quantum error correction [56–59]. Encoding in cat-state subspace via reverse engineering provides a feasible scheme for the realization of NGQC in a bosonic system [60]. The application of reverse engineering, which constructs the Hamiltonian based on the corresponding invariant, makes it easier to find more free parameters to control the evolution path [61,62]. Numerous studies have demonstrated that the tentative form of control parameters shapes the time evolution of quantum systems in a potentially useful way [63–65]. Typical forms of control parameters include polynomials of trigonometric functions [66,67], as well as the product form of the trigonometric and complex exponential functions [68,69]. In the system to be elucidated subsequently, the control parameters are limited to simple trigonometric functions. The adjustment of the evolution form of control parameters is of great importance in quantum computation.

Adopting the machine-learning technology and optimization theory has been proved to be applicable to optimizing the control parameters of variational states in a variety of interacting quantum many-body systems [70–74]. Although designing control parameters to acquire high-fidelity quantum gates by neural network has been extensively studied for a long time [14,75], it is still a flourishing and attractive research topic. Researchers designed dispersed and aperiodic control parameters by gradient ascent pulse engineering (GRAPE) under the guidance of state fidelity in nuclear magnetic resonance [76]. Here we introduce this method into the bosonic system, where the aperiodic discontinuous function is generalized to the periodic continuous function. We find that the incorporation of GRAPE enables the neural network to possess a powerful representation ability, which can expand the model space through the nonlinear activation function to fit any smooth periodic function and aperiodic function [77,78]. As a result, we optimize continuous and periodic control parameters, which are easier to physically implement, through the neural network with the enhancement of periodic characteristics.

The rest of the paper is organized as follows. In Sec. II we revisit the NGQC+ with cat states via reverse engineering.

Section III is devoted to the construction of the neural network guided by the average fidelity with periodic feature enhancement to improve the performance of single-qubit gates. In Sec. IV, we benchmark the optimization on the T gate ($\pi/8$ gate) and demonstrate that the neural network can effectively expand the model space. Furthermore, we assess the performance of the protocol under systematic noise, random noise, and decoherence effect via numerical simulations. Finally, the conclusions and outlook are given in Sec. VI.

II. NGQC+ WITH CAT STATES BASED ON REVERSE ENGINEERING

Applying reverse engineering to quantum computation not only permits the Hamiltonian to be more physically realizable, but also makes the implementation of quantum gates more flexible [61,79,80]. Consider a time-dependent Hamiltonian $H(t)$ and the corresponding dynamic invariant $I(t)$, which satisfies the following equation [81] ($\hbar = 1$):

$$i \frac{\partial}{\partial t} I(t) - [H(t), I(t)] = 0. \quad (1)$$

To realize NGQC+, we select a set of time-dependent eigenstates $|\phi_l(t)\rangle$ ($l = 1, 2, \dots, d$) of $I(t)$ to span a d -dimensional computational subspace \mathcal{S} , which are supposed to satisfy the three conditions below [40]. First, the computational basis should satisfy the boundary conditions at times $t = 0$ and L , i.e., $|\phi_l(0)\rangle = |\phi_l(L)\rangle$, to ensure that the evolution is cyclic. Here, L is the evolution period. Secondly, we can rewrite Eq. (1) based on eigenvectors of $I(t)$ as

$$\dot{\Xi}_l(t) = -i[H(t), \Xi_l(t)], \quad (2)$$

where $\Xi_l(t) = |\phi_l(t)\rangle\langle\phi_l(t)|$ is the projective operator of $|\phi_l(t)\rangle$. Finally, the cumulative dynamic phase of one cycle needs to vanish:

$$\Phi_l(L) = - \int_0^L dt \langle\phi_l(t)|H(t)|\phi_l(t)\rangle = 0. \quad (3)$$

This condition is the relaxation of parallel transportation $\langle\phi_l(t)|H(t)|\phi_k(t)\rangle = 0$ in NGQC.

When the conditions of NGQC+ are all satisfied, the time evolution operator at the final time $t = L$ in subspace \mathcal{S} can be described as

$$U(L, 0) = \sum_l \exp[i\Theta_l(L)] \Xi_l(0), \quad (4)$$

where $\Theta_l(L)$ is the geometric phase, given by

$$\Theta_l(L) = \int_0^L dt \langle\phi_l(t)|i \frac{\partial}{\partial t} |\phi_l(t)\rangle. \quad (5)$$

Suppose a Hamiltonian can be represented as follows:

$$H(t) = \sum_{j=1}^g \lambda_j(t) G_j, \quad (6)$$

where g is the rank of the group and $\{G_j\}$ is a group of Hermitian generators of Lie algebra [61,62,82], obeying the following relations:

$$[G_i, G_j] = i \sum_k \mu_{ij}^k G_k, \quad (i, j, k \in \{1, 2, \dots, g\}), \quad (7)$$

where μ_{ij}^k is the corresponding structure constant. If an invariant can be written as

$$I(t) = \sum_{j=1}^g \xi_j(t) G_j. \quad (8)$$

According to Eq. (1), it yields

$$\dot{\xi}_k(t) = \sum_{i,j=1}^g \lambda_i(t) \xi_j(t) \mu_{ij}^k. \quad (9)$$

Once $\{\xi_j(t)\}$ are known, we can thus obtain $\{\lambda_j(t)\}$ according to Eq. (9).

We consider a system in which a resonant single-mode two-photon drive is applied to a Kerr nonlinear resonator. In the rotating frame, the system Hamiltonian [79,83] can be written by

$$H_{\text{cat}} = -K a^{\dagger 2} a^2 + \epsilon_2 (e^{2i\xi} a^{\dagger 2} + e^{-2i\xi} a^2), \quad (10)$$

where K is the Kerr nonlinearity, a^\dagger (a) is the creation (annihilation) operator of the cavity mode, ϵ_2 is the strength of the two-photon driving, and ξ is the phase of the driving. The coherent states $|\pm\alpha\rangle$ with $\alpha = \sqrt{\epsilon_2/K} \exp(i\xi)$ are the degenerate eigenstates of H_{cat} , whose superpositions

$$|\mathcal{C}_\pm\rangle = \frac{1}{\sqrt{\mathcal{N}_\pm}} (|\alpha\rangle \pm |-\alpha\rangle) \quad (11)$$

are referred to as even (odd) cat states with the normalization constants $\mathcal{N}_\pm = 2 \pm 2 \exp(-2|\alpha|^2)$. We apply an external single-photon drive [83]:

$$H_c(t) = \chi(t) a^\dagger a + \epsilon(t) a^\dagger + \epsilon^*(t) a, \quad (12)$$

where $\chi(t)$ and $\epsilon(t)$ are the detuning and strength of the driving, respectively. The total Hamiltonian is described by $H_{\text{tot}}(t) = H_{\text{cat}} + H_c(t)$. If the constraint that the energy gaps between cat states and other eigenstates are much larger than $\chi(t)$ and $\epsilon(t)$ is satisfied, the Hamiltonian can be reduced to two-dimensional subspace spanned by cat states $|\mathcal{C}_\pm\rangle$. The Pauli matrices defined by cat states can be chosen as the Hermitian generators of the Lie group. The driving Hamiltonian thus can be simplified as $H_c = \vec{\Omega}(t) \cdot \vec{\sigma}$, where $\vec{\Omega}(t) = [\Omega_x(t), \Omega_y(t), \Omega_z(t)]$ is a three-dimensional unit vector, and $\vec{\sigma} = [\sigma_x, \sigma_y, \sigma_z]$.

Consider a dynamic invariant $I(t) = \vec{\zeta}(t) \cdot \vec{\sigma}$, where $\vec{\zeta}(t) = [\zeta_x(t), \zeta_y(t), \zeta_z(t)]$. Based on Eq. (9), we can get that $\dot{\vec{\zeta}}(t) = 2\vec{\Omega}(t) \times \vec{\zeta}(t)$ and $|\zeta(t)|$ is constant. For convenience, we can let $\vec{\zeta}(t) = (\sin \eta \sin \mu, \cos \eta \sin \mu, \cos \mu)$, where μ and η are time-dependent control parameters. The eigenstates of $I(t)$ in the cat-state representation are

$$\begin{aligned} |\phi_+(t)\rangle &= \cos \frac{\mu}{2} |\mathcal{C}_+\rangle + i \exp(-i\eta) \sin \frac{\mu}{2} |\mathcal{C}_-\rangle, \\ |\phi_-(t)\rangle &= i \exp(i\eta) \sin \frac{\mu}{2} |\mathcal{C}_+\rangle + \cos \frac{\mu}{2} |\mathcal{C}_-\rangle. \end{aligned} \quad (13)$$

According to Eqs. (3)–(5), we can calculate the geometric phases

$$\Theta_\pm(L) = \pm \int_0^L dt \dot{\eta} \sin^2 \frac{\mu}{2}, \quad (14)$$

and the dynamic phases

$$\Phi_\pm(L) = \mp \int_0^L dt \left(\frac{1}{2} \dot{\eta} \sin^2 \mu + \Omega_z \right) \sec \mu. \quad (15)$$

In order to satisfy the conditions $\Phi_\pm(L) = 0$ and $\dot{\vec{\zeta}}(t) = 2\vec{\Omega}(t) \times \vec{\zeta}(t)$, we design $\vec{\Omega}(t)$ as

$$\begin{aligned} \Omega_x(t) &= \frac{1}{4} [\dot{\eta} \sin \eta \sin(2\mu) - 2\dot{\mu} \cos \eta], \\ \Omega_y(t) &= \frac{1}{4} [\dot{\eta} \cos \eta \sin(2\mu) + 2\dot{\mu} \sin \eta], \\ \Omega_z(t) &= -\frac{1}{2} \dot{\eta} \sin^2 \mu. \end{aligned} \quad (16)$$

Therefore, we set the parameters $\chi(t)$ and $\epsilon(t)$ as

$$\begin{aligned} \chi(t) &= \frac{\dot{\eta} \sin^2 \mu \mathcal{N}_+ \mathcal{N}_-}{|\alpha|^2 (\mathcal{N}_+^2 - \mathcal{N}_-^2)}, \\ \text{Re}[\epsilon(t)] &= \frac{\sqrt{\mathcal{N}_+ \mathcal{N}_-}}{4|\alpha|} (\Omega_x \cos \xi - e^{2|\alpha|^2} \Omega_y \sin \xi), \\ \text{Im}[\epsilon(t)] &= \frac{\sqrt{\mathcal{N}_+ \mathcal{N}_-}}{4|\alpha|} (\Omega_x \sin \xi + e^{2|\alpha|^2} \Omega_y \cos \xi), \end{aligned} \quad (17)$$

which are scarcely different from the forms presented in Ref. [60]. Based on Eq. (4), the time evolution operator can be represented as

$$\begin{aligned} U(L, 0) &= \begin{bmatrix} \cos \theta + i \cos \mu_0 \sin \theta & \exp(i\eta_0) \sin \mu_0 \sin \theta \\ -\exp(-i\eta_0) \sin \mu_0 \sin \theta & \cos \theta - i \cos \mu_0 \sin \theta \end{bmatrix}, \end{aligned} \quad (18)$$

where μ_0 and η_0 are the initial value of μ and η , respectively:

$$\theta = \int_0^L dt \dot{\eta} \sin^2 \frac{\mu}{2}. \quad (19)$$

If we choose different μ , η , and θ , we can implement an arbitrary unitary single-qubit gate [60].

III. CONSTRUCTION OF A NEURAL-NETWORK ANSATZ BASED ON THE AVERAGE FIDELITY

Recently a tentative scheme for the parameters is the usage of trigonometric functions as [60]

$$\mu = \mu_0 + \Lambda \sin^2 \left(\frac{\pi t}{L} \right), \quad \eta = \eta_0 + \pi \left[1 - \cos \left(\frac{\pi t}{L} \right) \right], \quad (20)$$

where Λ is an auxiliary parameter depending on the concrete form of the desired gate. To facilitate the subsequent discussion, the parameter selection scheme of Eq. (20) is referred to as the trigonometric-function-based protocol. We can numerically calculate the integral in Eq. (19) as

$$\begin{aligned} \theta &= \pi \left\{ 1 - \sqrt{\frac{\pi}{2\Lambda}} \left[\cos(\mu_0 + \Lambda) C \left(\sqrt{\frac{2\Lambda}{\pi}} \right) \right. \right. \\ &\quad \left. \left. + \sin(\mu_0 + \Lambda) S \left(\sqrt{\frac{2\Lambda}{\pi}} \right) \right] \right\}, \end{aligned} \quad (21)$$

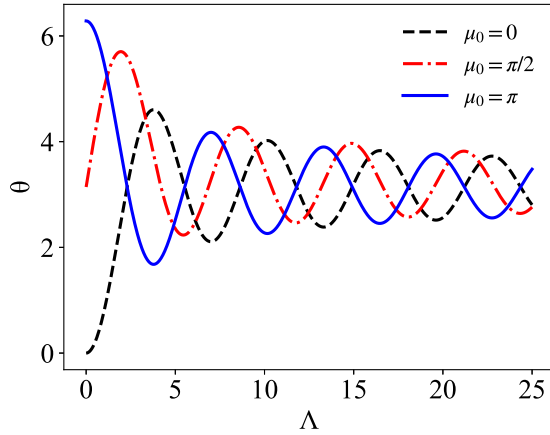


FIG. 1. The variation of θ with respect to Λ for $\mu_0=0, \pi/2$, and π .

where

$$S(x) = \int_0^x dt \sin(t^2), \quad C(x) = \int_0^x dt \cos(t^2)$$

are Fresnel integrals. It is obvious that θ is only dependent on μ_0 and Λ . We show the variation of θ with respect to Λ for a few typical values of μ_0 in Fig. 1. One observes that θ exhibits a decaying oscillation with respect to Λ and approaches π when Λ becomes sufficiently large. We find that θ cannot take the entire parameter range between 0 and 2π , which implies that the trigonometric-function-based protocol cannot implement arbitrary single-qubit gates. Especially it is difficult to accurately obtain Λ by solving complex nonlinear Eq. (19). Therefore, we improve the method of designing the variational parameters μ and η by machine-learning-inspired optimization based on GRAPE.

Subsequently, we employ the neural network under unsupervised machine learning as an ansatz. The neural network is composed of the input, hidden, and output layers. Two adjacent layers are connected by the weights, biases, and activation function. We choose one hidden layer and $\tanh(x)$ as the activation function. Because it is assumed that μ and η have nothing to do with each other, the neural network is not fully connected. If the control parameters are not independent of each other, the fully connected neural network will be adopted. The final outputs are the specific function expressions

$$\begin{aligned} \mu &= \sum_{i=1}^N W_i^{(2)} \tanh \left(W_i^{(1)} \tau^{(1)} + B_i^{(1)} \right) + B^{(2)}, \\ \eta &= \sum_{i=1}^N W_i^{(4)} \tanh \left(W_i^{(3)} \tau^{(2)} + B_i^{(3)} \right) + B^{(4)}, \end{aligned}$$

where N is the number of neurons in the hidden layer. Since the constructions of μ and η are similar, we take μ as an example. $\tau^{(1)}$ is the input of the neural network. The output of the neuron in the hidden layer is $\tanh \left(W_i^{(1)} \tau^{(1)} + B_i^{(1)} \right)$ with the weights $W_i^{(1)}$ and the biases $B_i^{(1)}$. Similarly, the output of the neural network is the specific function expression of μ with the weights $W_i^{(2)}$ and the bias $B_i^{(2)}$.

A. Feature enhancement

In parallel, we impose some restrictions on the variational parameters. To meet the cycle evolution condition $|\phi_{\pm}(0)\rangle = |\phi_{\pm}(L)\rangle$, the control parameters μ and η should be periodic and L is an integer multiple of the corresponding periods of μ and η . Considering the period of μ is T_{μ} and the period of η is T_{η} , it is supposed that $T_{\eta} = mT_{\mu}$ with m being any real number. To be noticed, the periodicity of μ and η is aimed at the real time t . For simplicity, we set $T_{\eta} = 2T_{\mu} = L$. The initial values of the control parameters μ_0 and η_0 can be determined by Eq. (18) for a target single-qubit quantum gate.

To summarize, the control parameters should meet three requirements below.

- (1) μ and η are periodic functions.
- (2) μ and η have initial value μ_0 and η_0 , respectively.
- (3) $T_{\eta} = 2T_{\mu} = L$.

The second condition can be satisfied easily. In particular, $B^{(2)}$ and $B^{(4)}$ can be set depending on $\mu(0) = \mu_0$ and $\eta(0) = \eta_0$. To achieve the goal that μ and η are periodic functions, we ought to make periodic feature enhancement.

Without loss of generality, we take the construction of a multilayer neural network as an example. Considering the lemma that if $\iota(x)$ is a given smooth periodic function with period L and $\Upsilon(\cdot)$ is a smooth function, then $\Upsilon[\iota(x)]$ is still a periodic function with period L [77]. To proceed, we apply the sinusoidal functions

$$\beta(x) = A \cos(\omega x + \phi) + c \quad (22)$$

in the first hidden layer with $\omega = 2\pi/L$. We choose a nonlinear activation function for the sake of guaranteeing the periodicity of the output and generating higher-frequency terms to expand the model space in training the neural network. For other hidden layers, the normal linear superposition of neurons in the former layer and nonlinear activation can be used. In this paper, we find that utilizing a small-scale neural-network ansatz with a single hidden layer is sufficient in optimizing the performance of target gates, which shows the superiority of our method. However, it is worth noting that increasing the number of hidden units or incorporating additional hidden layers may yield improved behavior at the cost of increased computational time and more difficult physical realization. In this respect, the final representations of μ and η of the neural network with the sole hidden layer are given by

$$\mu = \sum_{i=1}^N W_i^{(2)} \tanh \left[W_i^{(1)} \cos \left(\omega^{(1)} \tau^{(1)} + \phi_i^{(1)} \right) + B_i^{(1)} \right] + B^{(2)}, \quad (23)$$

$$\eta = \sum_{i=1}^N W_i^{(4)} \tanh \left[W_i^{(3)} \cos \left(\omega^{(2)} \tau^{(2)} + \phi_i^{(2)} \right) + B_i^{(3)} \right] + B^{(4)}. \quad (24)$$

Here, $\tau^{(1)} = 2\pi t/L$, $\tau^{(2)} = \pi t/L$, and $\omega^{(1)} = \omega^{(2)} = 1$. $\phi_i^{(1)}$ and $\phi_i^{(2)}$ are learnable parameters of the neural network, which will effectively expand the model space and satisfy the periodic relationship between μ and η .

B. Backpropagation guided by the average fidelity

The average fidelity is the benchmark to assess the performance of the quantum gates in the closed system and proves

to be more effective than assessing the fidelity of specific states, especially in improving the performance of synthetic multiqubit gates, given the uncertainties introduced by the preceding gate in the circuit. We thus choose the average fidelity as the objective function, which is defined as [84]

$$F(t) = \frac{1}{\mathcal{D}(\mathcal{D} + 1)} \{ \text{Tr}[M(t)M(t)^\dagger] + |\text{Tr}[M(t)]|^2 \}, \quad (25)$$

where \mathcal{D} is the dimension of the computational subspace, $M(t) = \mathcal{P}_c U_G^\dagger U_1(t) \mathcal{P}_c$, and \mathcal{P}_c is the projective operators of the subspace, while U_G and $U_1(t)$ respectively stand for the matrix representations of the ideal and actual gates. The application of Eq. (25) as the objective function instead of Eq. (19) offers two distinct advantages. First, Eq. (25) takes into consideration the leakage to unwanted levels, making it a more realistic measure of the performance of the scheme compared to θ , which is defined in the two-dimensional cat-state subspace. Secondly, while both the average fidelity and θ are nonconvex functions [85] due to the complex interactions among multiple parameters and the utilization of a nonlinear activation function, it is crucial to emphasize that there exists a clear global maximum for the average fidelity, which is unity. This allows for straightforward determination of when to finalize the neural network’s learning process. On the other hand, θ encompasses infinite ideal values, which may potentially confuse the network. Therefore, considering the aforementioned factors, employing the average fidelity as the objective function for neural network training is a more suitable choice compared to utilizing θ and the fidelity of specific states.

The workflow of the machine-learning-inspired optimization is illustrated in Fig. 2. In the neural-network ansatz, there are three layers with two input units, N hidden units, and two output units. The final-state average fidelity $\mathcal{F} \equiv F(L)$ measured at the final moment depends crucially on the specific evolution details of each previous moment. Considering the nonlinear relationship between the external single-photon drive in Eq. (17) and the control parameters, it is challenging to directly derive the variation of \mathcal{F} with respect to the neural-network parameters. Alternatively, we use the greedy algorithm, in which the temporal period L can be divided into n discrete time slices during an evolution cycle of realizing a single-qubit gate. Optimizing the average fidelity at each time slice can lead to a substantial reduction in complexity, ultimately resulting in a higher overall average fidelity \mathcal{F} for single-qubit gates. It is obvious that the evolution between two contiguous moments is described by the Schrödinger equation [76]. To this end, we calculate all the gradients of the average fidelity $F(t)$ with respect to parameters by the chain rule at each time slice. In order to obtain the maximum of average fidelity $F(t)$, we adopt the gradient ascent algorithm to update all the parameters

$$\begin{aligned} W^{(a)} &\leftarrow W^{(a)} + l_W^{(a)} \frac{\partial F(t)}{\partial W^{(a)}}, a = 1, 2, 3, 4, \\ B^{(b)} &\leftarrow B^{(b)} + l_B^{(b)} \frac{\partial F(t)}{\partial B^{(b)}}, b = 1, 3, \\ \phi^{(c)} &\leftarrow \phi^{(c)} + l_\phi^{(c)} \frac{\partial F(t)}{\partial \phi^{(c)}}, c = 1, 2, \end{aligned} \quad (26)$$

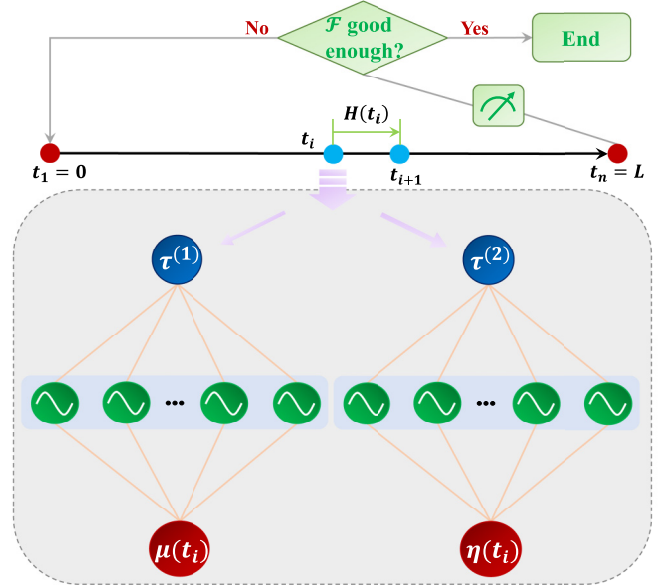


FIG. 2. The workflow of the machine-learning-inspired optimization based on the average fidelity. For a temporal cycle between $t_1 = 0$ and $t_n = L$ (red dots), which is divided into $n - 2$ slices $\{t_i\}_{i=2}^{n-1}$ (blue dots), we should perform the variation and update all parameters at each time slice to ensure that the neural network captures the information at each moment effectively. Two adjacent dots are connected by the Schrödinger equation, denoted as $U(t, 0)$. We choose the neural network with one hidden layer. $\tau^{(a)}$ ($a = 1, 2$) is the linear transformation of t_i as the input and the output μ and η are functions of $\tau^{(a)}$. Each neuron in the hidden and output layers has a corresponding weight and bias. The trigonometric functions $\cos(\omega\tau^{(a)} + \phi)$ with different phases can be used as the inputs of the neurons in the hidden layer to ensure the output of the neural network is a periodic function, and ω is determined by the period of μ and η . Adjusting the bias of the output is useful to make μ and η possess the fixed initial value. At the $t = L$ moment, the average fidelity is calculated according to the existing parameters to judge whether it is good enough to end the training.

for the next variation. Here, the learning rates $l_W^{(a)}$, $l_B^{(b)}$, and $l_\phi^{(c)}$ are the adjustable parameters, depending on the impact of the corresponding parameters on the average fidelity $F(t)$. We calculate the final-state average fidelity \mathcal{F} corresponding to the current parameters to judge whether the neural network has been well trained. The process of the above operations is defined as one variational process. After performing the variational processes N_{VP} times, we consider the training to be complete when \mathcal{F} approaches 1 with high precision. Note that the neural-network ansatz for the machine-learning-inspired method allows us to avoid solving Eq. (19), which has been assumed to be automatically met when the final-state average fidelity tends to unity. In this case, it is inevitable that we should verify whether Eq. (19) is valid or not according to the specific form of μ and η .

IV. NUMERICAL RESULTS AND DISCUSSION OF SINGLE-QUBIT GATES

The Gottesman-Knill theorem [86] tells us that a circuit using only Clifford gates and Pauli measurements [20] is

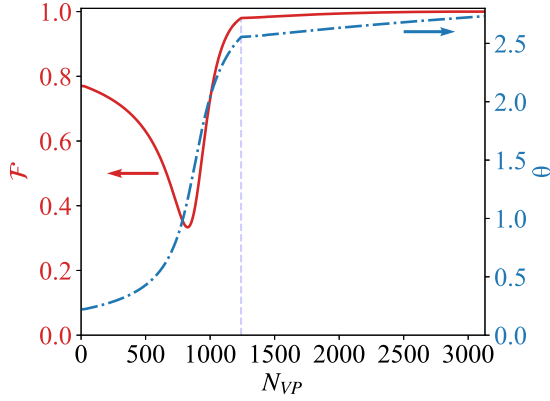


FIG. 3. The training results of the T gate. The final-state average fidelity \mathcal{F} (red solid line) and θ (blue dash-dot line) change with the number of variational processes N_{VP} .

insufficient for the universal quantum computation. T gate

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} \quad (27)$$

is the most natural and easiest single-qubit non-Clifford gate, which supplements the set of Clifford gates to achieve universal quantum computation [87,88]. The implementation of the T gate in the trigonometric-function-based protocol is not perfect. To realize the T gate, we should make the off-diagonal elements of the evolution $U(T, 0)$ in Eq. (18) vanish. Thus, we set $\mu_0 = 0$. In this case, $U(T, 0)$ has nothing to do with η_0 . We choose $\eta_0 = 0$ for simplicity. For the diagonal elements, it is readily yielded that $2k\pi - 2\theta = \pi/4$, namely, $\theta = k\pi - \pi/8$, where k is an arbitrary integer.

In the neural network, there are six hidden units. μ and η use half of the hidden units, respectively, as shown in Fig. 2. We pretrain the neural network according to the trigonometric-function-based protocol to obtain the initial parameters. Take a time series with $n = 1000$ data points evenly spaced between the time duration $t = 0$ and L . Then, we set $l_W^{(2)} = l_W^{(4)} = 10^{-4}$, $l_W^{(a)} = l_B^{(b)} = l_\phi^{(c)} = 10^{-5}$, $a = 1, 3$, $b = 1, 3$, $c = 1, 2$ for the first 1240 iterations of variational processes, and $l_W^{(2)} = l_W^{(4)} = 10^{-5}$, $l_W^{(a)} = l_B^{(b)} = l_\phi^{(c)} = 10^{-6}$, $a = 1, 3$, $b = 1, 3$, $c = 1, 2$ for the last 1890 iterations of variational processes. The learning rates of $W^{(2)}$ and $W^{(4)}$ are ten times more than those of other parameters, because the impact of $W^{(2)}$ and $W^{(4)}$ on the average fidelity $F(t)$ is much larger than other parameters. In the NGQC+, the amplitude of coherent states is $|\alpha| = 0.5$, the Kerr nonlinearity is $K = 2\pi \times 12.5$ MHz, the energy gap is $E_{\text{gap}} = 4K\alpha^2 = 78.5$ MHz [83], and the total interaction time is $T = 1$ μ s. For simplicity, we refer to our scheme as the machine-learning-inspired protocol.

The training results of the T gate through the code in Ref. [89] are shown in Fig. 3. The final-state average fidelity is 0.9999. According to Eq. (19), we get $\theta = 2.7332$, and the error from the theoretical value ($\theta = 7\pi/8$) is 0.0156. It can be seen that the trends of the final-state average fidelity and θ are different. During the initial several hundred iterations of the variational processes, the final-state average fidelity is moderately high, with the corresponding θ value already

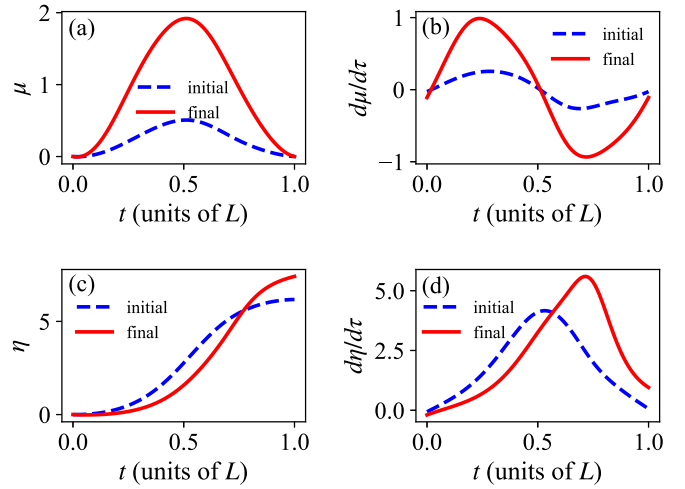


FIG. 4. The comparison between the initial and the final forms of (a) μ and (c) η . The initial forms take trigonometric functions, and the final forms are the outputs of the neural network. The first derivatives with respect to the inputs of the neural network (b) $d\mu/d\tau^{(1)}$ and (d) $d\eta/d\tau^{(2)}$ are also compared.

nearing another ideal value of $-\pi/8$. The final-state average fidelity will approach unity and simultaneously the actual value of θ converges to the ideal value as the variational iterations progress. More importantly, Fig. 3 illustrates the advantage of employing unsupervised learning in this paper. Unsupervised learning does not rely on a predefined set of labeled data for guidance, unlike supervised learning. Instead, it relies solely on gradients for learning, without the notion of right or wrong during the initial stages. As a result, it may initially go in the wrong direction. However, with proper initialization, it can eventually find the correct direction for learning. One can infer that the average fidelity is superior to the imposed constraint of θ in Eq. (19). Thus, it is wise to choose the average fidelity instead of θ as the objective function.

As such, we show the comparison between the initial and final points of $\mu(t)$ and $\eta(t)$ in Figs. 4(a) and 4(c). The initial forms take trigonometric functions, which are the outcomes of pretraining the neural network. The final forms are obtained by the outputs of the neural network. One finds that the final form has a clear deviation in the amplitude and the structure symmetry from the initial form after the entire training of the neural network. The final forms are no longer simple trigonometric functions, which can be clearly revealed by the derivatives of $\mu(t)$ with respect to $\tau^{(1)}$ and $\eta(t)$ with respect to $\tau^{(2)}$ shown in Figs. 4(b) and 4(d). The introduction of the neural network can broaden the model space, in which the control parameters can take more extensive and feasible trial forms.

To get more insights into the behaviors of $\mu(t)$ and $\eta(t)$ at the initial and final points of the training, we plot the trajectories of the eigenstates $|\phi_{\pm}(t)\rangle$ on the Bloch sphere in Fig. 5:

$$\vec{r}_{\pm}(t) = \sum_{k=x,y,z} \text{Tr}[|\phi_{\pm}\rangle\langle\phi_{\pm}| \sigma_k] \vec{e}_k, \quad (28)$$

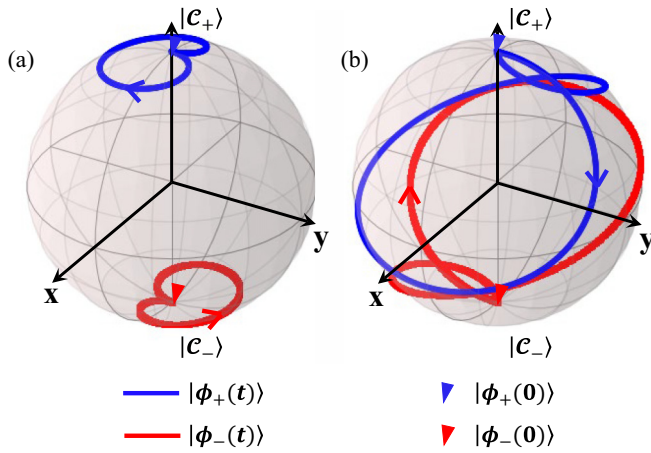


FIG. 5. Map of the changes of μ and η into the changes of the evolution path of the eigenstates $|\phi_{\pm}(t)\rangle$ on the Bloch sphere. The blue line is the evolution paths of $|\phi_{+}(t)\rangle$. The red line is the evolution path of $|\phi_{-}(t)\rangle$. $|c_{\pm}\rangle$ are the initial states of the evolution of $|\phi_{\pm}(t)\rangle$. (a) The evolution path corresponds to the initial μ and η . (b) The evolution path corresponds to the final μ and η .

where \bar{e}_k is the unit vector along the k axis. The differences between the initial and final $\mu(t)$ and $\eta(t)$ are magnified on the Bloch sphere. It can be seen that the evolution path varies a lot during the entire training. Thus, the neural-network ansatz shows unique advantages in quantum optimal control, which can obtain a more complex ansatz for possible control parameters.

Noise robustness

Next, we evaluate the performance of our scheme under different noisy circumstances. First, we consider the systematic noise effect, such as instrument defects and imperfection operations. Systematic errors can cause the average value of measured data to deviate significantly from the ideal value. The influence of systematic errors may be present in the parameters of the control Hamiltonian that can be written as $\Omega_k^e = (1 + \delta_k)\Omega_k$, $k = x, y, z$, where δ_k is the error coefficient. We plot the final-state average fidelity \mathcal{F} of the T gate with respect to the error coefficient δ_k in Fig. 6. We can find that when $\delta_x \in [-0.1, 0.1]$ ($\delta_y \in [-0.1, 0.1]$), the final-state average fidelity \mathcal{F} remains higher than 0.9986 (0.9984), while we can only obtain $\mathcal{F} \geq 0.9611$ when $\delta_z \in [-0.1, 0.1]$. It is obvious that the noise in the z -axis direction will cause more catastrophic decline in the final-state average fidelity than that in x and y axes. This effect can be understood because according to Eq. (3) the fluctuation in Ω_z will cause persistent adverse effects on the dynamic phase. The unidirectional offset of Ω_z will make the dynamic phase not vanish after a cycle, and then spoils the conditions of NGQC+.

We also consider the random noise effect, in which the amplitude, waveform, and phase are random at any time. Each random noise is still subject to certain statistical distribution. If the amplitude distribution of a noise follows Gaussian distribution and its power spectral density is uniformly distributed, this noise is called additive white Gaussian noise, $\mathcal{A}_G[\Omega_k(t), \text{SNR}]$ is one of the typical random noise models.

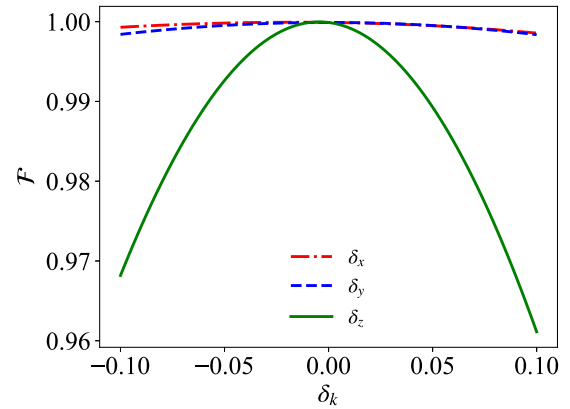


FIG. 6. The variation of final-state average fidelity \mathcal{F} of the T gate with respect to the systematic error coefficient δ_k , $k = x, y, z$.

Therefore, we take $\mathcal{A}_G[\Omega_k(t), \text{SNR}]$ as an example to analyze the robustness of our method to random processes and compare the robustness of the machine-learning-inspired and trigonometric-function based protocols. We add it to control parameters as

$$\Omega_k^q(t) = \Omega_k(t) + \mathcal{A}_G[\Omega_k(t), \text{SNR}], \quad (29)$$

where q represents each random generator $\mathcal{A}_G[\Omega_k(t), \text{SNR}]$. It is a function that generates $\mathcal{A}_G[\Omega_k(t), \text{SNR}]$ for the original signal $\Omega_k(t)$ with signal-to-noise ratio $\text{SNR} = 10 \log_{10}(P_{\text{signal}}/P_{\text{noise}})$, and P_{signal} and P_{noise} are the power of signal and noise, respectively. Due to the random generation of $\mathcal{A}_G[\Omega_k(t), \text{SNR}]$, we perform a large amount of numerical simulations to estimate the random noise effect. The logarithms of the deviations $\delta\mathcal{F}$ of the mean values of final-state average fidelities of the T gate from the ideal value of $50p$ ($p = 1, 2, 3, \dots$) iterations of numerical simulations are plotted in Fig. 7 with $\text{SNR} = 10$. When p tends to infinity, the simulation consequence is pretty close to the actual impact of the random noise. The ideal value of the final-state average

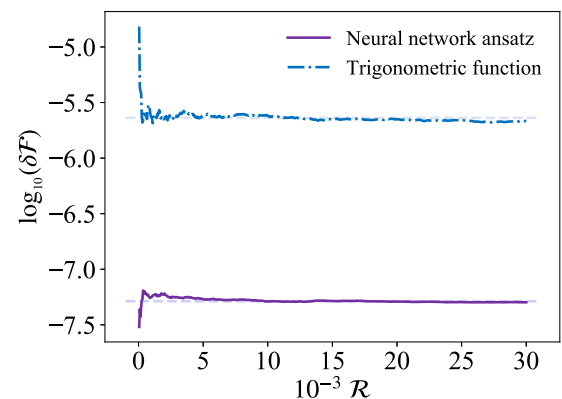


FIG. 7. The logarithms of the deviations $\delta\mathcal{F}$ of the mean values of final-state average fidelities of the T gate from the ideal value with respect to simulation times \mathcal{R} under the random noise effect with $\text{SNR} = 10$ in the machine-learning-inspired (purple solid line) and trigonometric-function based (blue dash-dot line) protocols. The dotted lines represent the convergence values of the two protocols. Here, $\mathcal{R} = 50p$, ($p = 1, 2, 3, \dots$).

TABLE I. Parameters and corresponding final-state average fidelities for the implementation of single-qubit gates. The rightmost three columns are to verify whether Eq. (19) is satisfied. We calculate θ_{actual} with the output of the neural network, and the relative error between θ_{actual} and θ_{ideal} .

Gate	μ_0	η_0	Fidelity	θ_{actual}	θ_{ideal}	Error
T	0	0	0.9999	2.7332	$7\pi/8$	0.0157
X	$3\pi/2$	$\pi/2$	0.9999	1.5459	$\pi/2$	0.0249
H	$\pi/4$	$\pi/2$	0.9997	1.5738	$\pi/2$	0.0030
T^\dagger	0	0	0.9999	0.3569	$\pi/8$	0.0338
$R_x(\pi/4)$	$\pi/2$	$-\pi/2$	0.9992	3.6019	$9\pi/8$	0.0676

fidelity in the machine-learning-inspired protocol is 0.9999, while that in the trigonometric-function-based protocol is 0.8894. It is observed that our scheme performs significantly better in the presence of random noise. Compared to the trigonometric-function-based protocol, the mean value of the final-state average fidelities under random noise in the machine-learning-inspired scheme exhibits fewer fluctuations with respect to p and approaches $1-7.84 \times 10^{-4}$, with a smaller deviation from the ideal value as p becomes sufficiently large. It is thus acknowledged that our scheme is robust against the random noise, and enhancing the performance of the gate can improve the robustness to a certain degree.

As the system cannot be completely isolated from the environment, the inevitable interactions between the system and the environment will also lead to the decoherence. We mainly consider two dissipation factors, such as a single-photon loss and dephasing [47]. The evolution of the system can be described by the Lindblad master equation [47,90]:

$$\begin{aligned} \dot{\rho}(t) = & -i[H_{\text{cat}} + H_c(t), \rho(t)] \\ & + \Gamma \mathcal{L}[a]\rho(t) + \Gamma_\phi \mathcal{L}[a^\dagger a]\rho(t). \end{aligned} \quad (30)$$

Here, Γ and Γ_ϕ are the dissipation coefficients of a single-photon loss and dephasing, respectively, and the Lindblad superoperator \mathcal{L} acting on arbitrary operator o produces $\mathcal{L}[o]\rho(t) = o\rho(t)o^\dagger - o^\dagger o\rho(t)/2 - \rho(t)o^\dagger o/2$. In the presence of decoherence, the evolution is no more unitary. We can no longer use Eq. (25) to measure the performance of the quantum gates. Therefore, we take the evolution with initial state $|\mathcal{C}_+\rangle$ as an example and evaluate the fidelity of the T gate as

$$F_T = \langle \mathcal{C}_+ | U_T^\dagger \rho(T) U_T | \mathcal{C}_+ \rangle. \quad (31)$$

In our numerical simulation, we set $\Gamma = \Gamma_\phi = 0.05$ MHz, and we can obtain the fidelity of the T gate as $F_T = 0.9803$. This means the leakage to unwanted levels outside the subspace is still very small, and our scheme is insensitive to decoherence.

The implementations of the NOT gate (X gate), Hadamard gate (H gate), T^\dagger gate, and $R_x(\pi/4)$ gate are listed in Table I. Here, $R_x(\phi) = \exp(-\frac{i}{2}\phi\sigma_x)$ is a rotation gate around the x axis [21]. It can be seen that the machine-learning-inspired protocol excels for almost all kinds of single-qubit gates. Especially, our scheme shows superiority in phase gates, whose average fidelities can reach 0.9999, much higher than those in the trigonometric-function-based protocol. The performance

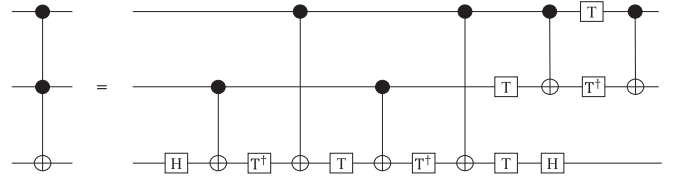


FIG. 8. The Toffoli gate (T) is composed of the CNOT gate, H gate, T gate ($\pi/8$), and $T^\dagger(-\pi/8)$ gate.

of the X gate in the two protocols is equally remarkable. Through the neural network, we can implement the rotation gates which are unrealizable in the trigonometric-function-based protocol. For the H gate, the obtained results are not very accurate, and more sophisticated neural networks are awaited. Furthermore, we realize the modified controlled-NOT (CNOT) gate with the final-state average fidelity 0.9996. Here, $U_{\text{CNOT}} = |\mathcal{C}_+\rangle\langle\mathcal{C}_+| \otimes \mathbb{I} + |\mathcal{C}_-\rangle\langle\mathcal{C}_-| \otimes (-i\sigma_x)$ and \mathbb{I} is the unit matrix acting on the cat-state subspace. The execution of a two-qubit controlled gate is shown in the Appendix. It is clear that for each gate, the higher the final-state average fidelity is, the smaller the error is.

To conclude, through the introduction of the neural network, we can lift the restrictions imposed on θ to a certain extent. We can realize arbitrary θ by adjusting the initial parameters and the structure of the neural network.

V. REALIZATION OF THE TOFFOLI GATE

In the trigonometric-function-based protocol, it is scarcely possible to execute the single-shot multiqubit gates, and the final-state average fidelity of synthetic multiqubit gates by combining high-fidelity single- and two-qubit gates will be rather low. The Toffoli gate [20]

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

is composed of the CNOT gate, H gate, T gate, and T^\dagger gate, as is shown in Fig. 8. The final-state average fidelity of Toffoli gate is 0.5169, when all gates in Fig. 8 are realized in the trigonometric-function-based protocol, and the main limitation is due to the bad performance of the T gate. In the machine-learning-inspired protocol, we can realize higher-fidelity multiqubit gates in the cascaded mode. The finally modified Toffoli gate can be synthesized by H gate and CNOT gate implemented in the trigonometric-function-based protocol and T gate and T^\dagger gate shown in Table I, and the final-state average fidelity of such a three-qubit entangling gate increases to 0.9976, with improved performance of the T and T^\dagger gates. However, we find that, although the final-state average fidelities of T and T^\dagger gates are up to 0.9999, it is still challenging to synthesize a high-fidelity multiqubit gate, which is seriously hindered by the lowest-fidelity quantum gates. When all gates in Fig. 8 are realized in the machine-learning-inspired pro-

tol, the final-state average fidelity can further increase to 0.9981. Thus, our scheme can provide a feasible routine to realize multiqubit gates in bosonic systems.

VI. CONCLUSION AND OUTLOOK

In this paper, we present a machine-learning-inspired method of optimizing the performance of the imperfect gates with cat-state NGQC+ via reverse engineering. By utilizing periodic feature enhancement and corresponding biases, we can obtain a periodic function as an output of a neural-network ansatz with fixed initial values. The machine-learning-inspired protocol allows us to not have to solve the difficult nonlinear equation [Eq. (19)], which can be automatically satisfied when the final-state average fidelity tends to be 1. Through analyzing the variational forms of the control parameters and comparing with the simple trigonometric functions, we prove that the neural network can greatly expand the model space and realize a more complex ansatz for possible control parameters. We find the final-state average fidelities of the phase gates and NOT gate can reach 0.9999, and those of the Hadamard gate and CNOT gate can be up to 0.9996. In order to improve the performance of the Hadamard gate, we can expand the scale of the neural network by increasing the number of hidden units and hidden layers. We can also adjust the periodic relationship between μ and η and the initial parameters in the hope of obtaining better results.

An alternative approach in the neural network is to use multiobjective optimization. Meanwhile, we show that we developed an approach for implementing high-fidelity rotation gates that are challenging to realize using trigonometric function-based protocol. Our scheme demonstrates robustness against various types of decoherence effects. Additionally, we observe that once the average fidelities of single- and two-qubit gates surpass a certain threshold the average fidelities of the synthetic gate may not be significantly compromised. Combining high-fidelity single- and two-qubit gates, we can implement the Toffoli gate with high fidelity, which cannot be simply realized in trigonometric-function-based protocol. In order to further improve the performance of the synthetic gate, we can use the average fidelity of the synthetic gate to guide the variational learning of the neural network, instead of only optimizing the single and two-qubit gates, and the improved scheme is left for a future study. We thus provide an alternative method of designing the control parameters. The machine-learning-inspired scheme paves the way for the optimization of continuous and periodic parameters in the quantum control, and can be generalized to more intricate neural networks featuring a substantial number of optimizable parameters, targeting increasingly complex quantum systems [66–69].

ACKNOWLEDGMENTS

The authors appreciate very insightful discussions with Yimin Wang, Ming Xue, and Meng-Jiao Lyu. This work is supported by College Students' Innovation and Entrepreneurship Training Program projects of Nanjing University of Aeronautics and Astronautics under Grant No. 202210287094Z. W.-L.Y. kindly acknowledges support by

the National Natural Science Foundation of China under Grant No. 12174194 and a startup fund of Nanjing University of Aeronautics and Astronautics under Grant No. 1008-YAH20006. A.M.O. kindly acknowledges Narodowe Centrum Nauki (Poland) Project No. 2021/43/B/ST3/02166 and is grateful for support via the Alexander von Humboldt Foundation Fellowship (Humboldt-Forschungspreis).

APPENDIX: REALIZATION OF THE TWO-QUBIT CONTROLLED GATE

The Hamiltonian of two cavity modes driven by two Kerr-nonlinear resonators can be described as

$$H_{\text{cat},2} = \sum_{n=1,2} \left[-K a_n^{\dagger 2} a_n^2 + \epsilon_2 (e^{2i\xi} a_n^{\dagger 2} + e^{-2i\xi} a_n^2) \right]. \quad (\text{A1})$$

Here, a_n (a_n^{\dagger}) is the annihilation (creation) operator of the n th mode. The product states of two-mode coherent states $\{|\alpha\rangle_1 \otimes |\alpha\rangle_2\}$, with $\alpha = \pm\sqrt{\epsilon_2/K} \exp(i\xi)$, are fourfold degenerate eigenstates of $H_{\text{cat},2}$. $\{|\mathcal{C}_{\pm}\rangle_1 \otimes |\mathcal{C}_{\pm}\rangle_2\}$ can span the four-dimensional subspace \mathcal{S}_2 to implement the two-qubit gates. The control Hamiltonian [90–92] is given by

$$H_{c2}(t) = \chi_{12}(t) a_1^{\dagger} a_1 a_2^{\dagger} a_2 + a_1^{\dagger} a_1 [\lambda^*(t) a_2 + \lambda(t) a_2^{\dagger}] + \epsilon^*(t) a_2 + \epsilon(t) a_2^{\dagger} + \sum_{n=1,2} \chi_n(t) a_n^{\dagger} a_n. \quad (\text{A2})$$

Here, $\chi_{12}(t)$ is the cross-Kerr parameter, $\lambda(t)$ is the longitudinal interaction strength between modes 1 and 2, $\epsilon(t)$ is the strength of the extra driving of mode 2, and χ_n ($n = 1, 2$) is the detuning of the n th mode. Similarly, it is assumed that the parameters of H_{c2} should be much smaller than the energy gaps between cat states and other eigenstates of $H_{\text{cat},2}$. To realize the two-qubit controlled gate $U_2(T, 0) = |\mathcal{C}_+\rangle_1 \langle \mathcal{C}_+| \otimes \mathbb{I}_2 + |\mathcal{C}_-\rangle_1 \langle \mathcal{C}_-| \otimes U_s(T, 0)$, the parameters of H_{c2} are set as follows:

$$\begin{aligned} \chi_{12}(t) &= \frac{-2\Omega_z \mathcal{N}_+^2 \mathcal{N}_-^2}{|\alpha|^4 (\mathcal{N}_+^2 - \mathcal{N}_-^2)^2}, \\ \chi_1(t) &= -\frac{|\alpha|^2 (\mathcal{N}_+^2 + \mathcal{N}_-^2)}{2\mathcal{N}_+ \mathcal{N}_-} \chi_{12}(t), \\ \chi_2(t) &= -|\alpha|^2 \frac{\mathcal{N}_-}{\mathcal{N}_+} \chi_{12}(t), \\ \text{Re}[\lambda(t)] &= \frac{(\mathcal{N}_+ \mathcal{N}_-)^{\frac{3}{2}}}{4(\mathcal{N}_+^2 - \mathcal{N}_-^2) |\alpha|^3} (\Omega_x \cos \xi - \Omega_y e^{2|\alpha|^2} \sin \xi), \\ \text{Im}[\lambda(t)] &= \frac{(\mathcal{N}_+ \mathcal{N}_-)^{\frac{3}{2}}}{4(\mathcal{N}_+^2 - \mathcal{N}_-^2) |\alpha|^3} (\Omega_x \sin \xi + \Omega_y e^{2|\alpha|^2} \cos \xi), \\ \text{Re}[\epsilon(t)] &= -|\alpha|^2 \frac{\mathcal{N}_-}{\mathcal{N}_+} \text{Re}[\lambda(t)], \\ \text{Im}[\epsilon(t)] &= -|\alpha|^2 \frac{\mathcal{N}_-}{\mathcal{N}_+} \text{Im}[\lambda(t), \end{aligned} \quad (\text{A3})$$

which are slightly different from the parameters chosen in Ref. [60]. The optimization of two-qubit controlled gates in the neural networks is similar to that of single-qubit gates.

- [1] T. Häner, M. Roetteler, and K. M. Svore, *Quantum Inf. Comput.* **17**, 673 (2017).
- [2] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, *Phys. Rev. Lett.* **93**, 130502 (2004).
- [3] W.-Q. Liu, H.-R. Wei, and L.-C. Kwek, *Adv. Quantum Technol.* **5**, 2100136 (2022).
- [4] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Nature (London)* **482**, 382 (2012).
- [5] T. H. Taminiau, J. Cramer, T. van der Sar, V. V. Dobrovitski, and R. Hanson, *Nat. Nanotechnol.* **9**, 171 (2014).
- [6] A. S. Darmawan, B. J. Brown, A. L. Grimsmo, D. K. Tuckett, and S. Puri, *PRX Quantum* **2**, 030345 (2021).
- [7] A. Peres, *Phys. Rev. A* **32**, 3266 (1985).
- [8] R. P. Feynman, *Opt. News* **11**, 11 (1985).
- [9] D. Chatterjee and A. Roy, *Prog. Theor. Exp. Phys.* **2015**, 093A02 (2015).
- [10] C.-R. Fan, B. Lu, X.-T. Feng, W.-C. Gao, and C. Wang, *Quantum Eng.* **3**, e67 (2021).
- [11] A. J. Baker, G. B. Huber, N. J. Glaser, F. Roy, I. Tsitsilin, S. Filipp, and M. J. Hartmann, *Appl. Phys. Lett.* **120**, 054002 (2022).
- [12] B. P. Lanyon, M. Barbieri, M. P. Almeida, T. Jennewein, T. C. Ralph, K. J. Resch, G. J. Pryde, J. L. O'Brien, A. Gilchrist, and A. G. White, *Nat. Phys.* **5**, 134 (2009).
- [13] T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, *Phys. Rev. Lett.* **102**, 040501 (2009).
- [14] R. J. Spiteri, M. Schmidt, J. Ghosh, E. Zahedinejad, and B. C. Sanders, *New J. Phys.* **20**, 113009 (2018).
- [15] S. Daraeizadeh, S. P. Premaratne, N. Khammassi, X. Song, M. Perkowski, and A. Y. Matsuura, *Phys. Rev. A* **102**, 012601 (2020).
- [16] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, *Nat. Phys.* **18**, 783 (2022).
- [17] Y. He, M.-X. Luo, E. Zhang, H.-K. Wang, and X.-F. Wang, *Int. J. Theor. Phys.* **56**, 2350 (2017).
- [18] A. J. da Silva and D. K. Park, *Phys. Rev. A* **106**, 042602 (2022).
- [19] X. Gu, J. Fernández-Pendás, P. Vikstål, T. Abad, C. Warren, A. Bengtsson, G. Tancredi, V. Shumeiko, J. Bylander, G. Johansson, and A. F. Kockum, *PRX Quantum* **2**, 040348 (2021).
- [20] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010).
- [21] K. M. Nakanishi, T. Satoh, and S. Todo, [arXiv:2109.13223](https://arxiv.org/abs/2109.13223).
- [22] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, New York, 2013).
- [23] E. Cohen, H. Larocque, F. Bouchard, F. Nejdassattari, Y. Gefen, and E. Karimi, *Nat. Rev. Phys.* **1**, 437 (2019).
- [24] J. Anandan, *Nature (London)* **360**, 307 (1992).
- [25] M. Berry, *Sci. Am.* **259**, 46 (1988).
- [26] M. V. Berry, *Proc. R. Soc. A* **392**, 45 (1984).
- [27] J. Zhang, T. H. Kyaw, S. Filipp, L.-C. Kwek, E. Sjöqvist, and D. Tong, *Phys. Rep.* **1027**, 1 (2023).
- [28] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems: Foundations, Mathematical Concepts, and Applications in Molecular and Condensed Matter Physics* (Springer, New York, 2003).
- [29] A. Shapere and F. Wilczek, *Geometric Phases in Physics* (World Scientific, Singapore, 1989), Vol. 5.
- [30] W. Xiang-Bin and M. Keiji, *Phys. Rev. Lett.* **87**, 097901 (2001).
- [31] P. Solinas, P. Zanardi, N. Zanghì, and F. Rossi, *Phys. Rev. A* **67**, 052309 (2003).
- [32] E. Sjöqvist, D. M. Tong, L. M. Andersson, B. Hessmo, M. Johansson, and K. Singh, *New J. Phys.* **14**, 103035 (2012).
- [33] P. Z. Zhao, X.-D. Cui, G. F. Xu, E. Sjöqvist, and D. M. Tong, *Phys. Rev. A* **96**, 052316 (2017).
- [34] T. Chen and Z.-Y. Xue, *Phys. Rev. Appl.* **10**, 054051 (2018).
- [35] G. Feng, G. Xu, and G. Long, *Phys. Rev. Lett.* **110**, 190501 (2013).
- [36] Y. Xu, Z. Hua, T. Chen, X. Pan, X. Li, J. Han, W. Cai, Y. Ma, H. Wang, Y. P. Song, Z.-Y. Xue, and L. Sun, *Phys. Rev. Lett.* **124**, 230503 (2020).
- [37] S.-L. Zhu and Z. D. Wang, *Phys. Rev. Lett.* **89**, 097902 (2002).
- [38] Y.-Y. Huang, Y.-K. Wu, F. Wang, P.-Y. Hou, W.-B. Wang, W.-G. Zhang, W.-Q. Lian, Y.-Q. Liu, H.-Y. Wang, H.-Y. Zhang, L. He, X.-Y. Chang, Y. Xu, and L.-M. Duan, *Phys. Rev. Lett.* **122**, 010503 (2019).
- [39] J. W. Zhang, L.-L. Yan, J. C. Li, G. Y. Ding, J. T. Bu, L. Chen, S.-L. Su, F. Zhou, and M. Feng, *Phys. Rev. Lett.* **127**, 030502 (2021).
- [40] B.-J. Liu, X.-K. Song, Z.-Y. Xue, X. Wang, and M.-H. Yung, *Phys. Rev. Lett.* **123**, 100501 (2019).
- [41] X. Chen, I. Lizuain, A. Ruschhaupt, D. Guéry-Odelin, and J. G. Muga, *Phys. Rev. Lett.* **105**, 123003 (2010).
- [42] J. Zhang, T. H. Kyaw, D. M. Tong, E. Sjöqvist, and L.-C. Kwek, *Sci. Rep.* **5**, 18414 (2015).
- [43] K. Khodjasteh and D. A. Lidar, *Phys. Rev. Lett.* **95**, 180501 (2005).
- [44] A. M. Souza, G. A. Álvarez, and D. Suter, *Phys. Rev. Lett.* **106**, 240501 (2011).
- [45] X.-C. Yang, M.-H. Yung, and X. Wang, *Phys. Rev. A* **97**, 042324 (2018).
- [46] X.-M. Zhang, Z.-W. Cui, X. Wang, and M.-H. Yung, *Phys. Rev. A* **97**, 052333 (2018).
- [47] P. M. Harrington, E. J. Mueller, and K. W. Murch, *Nat. Rev. Phys.* **4**, 660 (2022).
- [48] E. T. Campbell, B. M. Terhal, and C. Vuillot, *Nature (London)* **549**, 172 (2017).
- [49] A. Blais, S. M. Girvin, and W. D. Oliver, *Nat. Phys.* **16**, 247 (2020).
- [50] D. G. Cory, M. D. Price, W. Maas, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo, *Phys. Rev. Lett.* **81**, 2152 (1998).
- [51] P. Schindler, J. T. Barreiro, T. Monz, V. Nebendahl, D. Nigg, M. Chwalla, M. Hennrich, and R. Blatt, *Science* **332**, 1059 (2011).
- [52] J. Chiaverini, D. Leibfried, T. Schaetz, M. D. Barrett, R. B. Blakestad, J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, R. Ozeri, and D. J. Wineland, *Nature (London)* **432**, 602 (2004).
- [53] K. Noh and C. Chamberland, *Phys. Rev. A* **101**, 012316 (2020).
- [54] L. Hu, Y. Ma, W. Cai, X. Mu, Y. Xu, W. Wang, Y. Wu, H. Wang, Y. P. Song, C.-L. Zou, S. M. Girvin, L.-M. Duan, and L. Sun, *Nat. Phys.* **15**, 503 (2019).
- [55] D. S. Schlegel, F. Minganti, and V. Savona, *Phys. Rev. A* **106**, 022431 (2022).
- [56] Y. Ma, Y. Xu, X. Mu, W. Cai, L. Hu, W. Wang, X. Pan, H. Wang, Y. P. Song, C.-L. Zou, and L. Sun, *Nat. Phys.* **16**, 827 (2020).

- [57] J. M. Gertler, B. Baker, J. Li, S. Shirol, J. Koch, and C. Wang, *Nature (London)* **590**, 243 (2021).
- [58] V. V. Albert, S. O. Mundhada, A. Grimm, S. Touzard, M. H. Devoret, and L. Jiang, *Quantum Sci. Technol.* **4**, 035007 (2019).
- [59] W. Cai, Y. Ma, W. Wang, C.-L. Zou, and L. Sun, *Fundam. Res.* **1**, 50 (2021).
- [60] Y.-H. Kang, Y.-H. Chen, X. Wang, J. Song, Y. Xia, A. Miranowicz, S.-B. Zheng, and F. Nori, *Phys. Rev. Res.* **4**, 013233 (2022).
- [61] Y.-H. Kang, Z.-C. Shi, B.-H. Huang, J. Song, and Y. Xia, *Phys. Rev. A* **101**, 032322 (2020).
- [62] E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, *Phys. Rev. A* **89**, 043408 (2014).
- [63] D. J. Egger and F. K. Wilhelm, *Phys. Rev. Lett.* **112**, 240503 (2014).
- [64] S. A. Rice and M. Zhao, *Optical Control of Molecular Dynamics* (Wiley, New York, 2000).
- [65] M. Shapiro and P. Brumer, *Principles of the Quantum Control of Molecular Processes* (Wiley, New York, 2003).
- [66] W. Li, *New J. Phys.* **23**, 073039 (2021).
- [67] S. Wang, W.-S. Ni, Z.-W. Chen, B. Feng, Y. Ning, Y.-H. Kang, and Y. Xia, *Laser Phys. Lett.* **20**, 015203 (2023).
- [68] Y.-H. Kang, Z.-C. Shi, J. Song, and Y. Xia, *Phys. Rev. A* **102**, 022617 (2020).
- [69] S. Liu, J.-H. Shen, R.-H. Zheng, Y.-H. Kang, Z.-C. Shi, J. Song, and Y. Xia, *Front. Phys.* **17**, 21502 (2022).
- [70] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [71] X. Gao and L.-M. Duan, *Nat. Commun.* **8**, 662 (2017).
- [72] M. J. Hartmann and G. Carleo, *Phys. Rev. Lett.* **122**, 250502 (2019).
- [73] M. Schmitt and M. Heyl, *Phys. Rev. Lett.* **125**, 100503 (2020).
- [74] J. Dborin, V. Wimalaweera, F. Barratt, E. Ostby, T. E. O'Brien, and A. G. Green, *Nat. Commun.* **13**, 5977 (2022).
- [75] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, *Phys. Rev. A* **84**, 022305 (2011).
- [76] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, *J. Magn. Reson.* **172**, 296 (2005).
- [77] S. Dong and N. Ni, *J. Comput. Phys.* **435**, 110242 (2021).
- [78] L. Ziyin, T. Hartwig, and M. Ueda, in *Advances in Neural Information Processing Systems*, NeurIPS Proceedings Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran, New York, 2020), pp. 1583–1594.
- [79] S. Puri, S. Boutin, and A. Blais, *npj Quantum Inf.* **3**, 18 (2017).
- [80] Y.-H. Kang, Y.-H. Chen, Z.-C. Shi, B.-H. Huang, J. Song, and Y. Xia, *Phys. Rev. A* **97**, 033407 (2018).
- [81] H. R. Lewis and W. B. Riesenfeld, *J. Math. Phys.* **10**, 1458 (1969).
- [82] R. S. Kaushal and H. J. Korsch, *J. Math. Phys.* **22**, 1904 (1981).
- [83] A. Grimm, N. E. Frattini, S. Puri, S. O. Mundhada, S. Touzard, M. Mirrahimi, S. M. Girvin, S. Shankar, and M. H. Devoret, *Nature (London)* **584**, 205 (2020).
- [84] L. H. Pedersen, N. M. Møller, and K. Mølmer, *Phys. Lett. A* **367**, 47 (2007).
- [85] L. Banchi, N. Pancotti, and S. Bose, *npj Quantum Inf.* **2**, 16019 (2016).
- [86] D. Gottesman, *Phys. Rev. A* **57**, 127 (1998).
- [87] P. O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, *Inf. Process. Lett.* **75**, 101 (2000).
- [88] M. Howard and J. Vala, *Phys. Rev. A* **86**, 022316 (2012).
- [89] M.-Y. Mao, Z. Cheng, Y. Xia, A. M. Oleś, and W.-L. You, Computer code for the study “Machine-learning-inspired optimal quantum control of nonadiabatic geometric quantum computation via reverse engineering (2023),” doi:10.5281/zenodo.8085591.
- [90] M. Mirrahimi, Z. Leghtas, V. V. Albert, S. Touzard, R. J. Schoelkopf, L. Jiang, and M. H. Devoret, *New J. Phys.* **16**, 045014 (2014).
- [91] J. Guillaud and M. Mirrahimi, *Phys. Rev. X* **9**, 041053 (2019).
- [92] S. Touzard, A. Kou, N. E. Frattini, V. V. Sivak, S. Puri, A. Grimm, L. Frunzio, S. Shankar, and M. H. Devoret, *Phys. Rev. Lett.* **122**, 080502 (2019).