Designing High-Fidelity Single-Shot Three-Qubit Gates: A Machine Learning Approach

Ehsan Zahedinejad,[1] Joydip Ghosh,[1, 2, *] and Barry C. Sanders[1, 3, 4, 5, 6, †]

[1]*Institute for Quantum Science and Technology, University of Calgary, Alberta, Canada T2N 1N4*
[2]*Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*
[3]*Program in Quantum Information Science, Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada*
[4]*Hefei National Laboratory for Physical Sciences at the Microscale and Department of Modern Physics,*
University of Science and Technology of China, Anhui 230026, China
[5]*Shanghai Branch, CAS Center for Excellence and Synergetic*
Innovation Center in Quantum Information and Quantum Physics,
University of Science and Technology of China, Shanghai 201315, China
[6]*Institute for Quantum Information and Matter,*
California Institute of Technology, Pasadena, California 91125, USA
(Dated: October 25, 2016)

Three-qubit quantum gates are key ingredients for quantum error correction and quantum information processing. We generate quantum-control procedures to design three types of three-qubit gates, namely Toffoli, Controlled-Not-Not and Fredkin gates. The design procedures are applicable to a system comprising three nearest-neighbor-coupled superconducting artificial atoms. For each three-qubit gate, the numerical simulation of the proposed scheme achieves 99.9% fidelity, which is an accepted threshold fidelity for fault-tolerant quantum computing. We test our procedure in the presence of decoherence-induced noise as well as show its robustness against random external noise generated by the control electronics. The three-qubit gates are designed via the machine learning algorithm called Subspace-Selective Self-Adaptive Differential Evolution (SuSSADE).

PACS numbers: 42.50.Gy, 42.50.Ex

I. INTRODUCTION

Quantum computing requires a universal set of low-error quantum gates to enable fault-tolerant quantum computing [1], and characterization is typically performed using average fidelity from which gate error rate can be inferred [2]. A universal set of single- [3] and two-qubit [4] gates can be employed to decompose [5] multi-qubit gates into a series of single and two-qubit gates. In practice, this decomposition-based approach is undesirable because it leads to quantum circuits [6, 7] with long operation time. We employ the recently proposed machine-learning technique, which we named Subspace-Selective Self-Adaptive Differential Evolution (SuSSADE) [8], to generate *procedures* (i.e., set of instructions that determine the control parameters, and hence the effectiveness of the control scheme) for designing single-shot high-fidelity three-qubit gates without any need to resort to a decomposition. We test our procedure [9] in the presence of decoherence induced noise and demonstrates its robustness under the effect of random control noise. The three-qubit gates that we consider here are Toffoli [10–12] (which has already been discussed in Ref. [8] and we give a review here for completeness), Fredkin [13, 14] and Controlled-NOT-NOT or CXX [15], which are typical three-qubit gates employed for quantum information processing.

The three-qubit gates that we consider in this study are key ingredients for quantum algorithms and error correction. A quantum Toffoli gate is necessary for (non-topological) quantum error correction [6, 7] and a key component for reversible computing [16]. Toffoli gate together with single-qubit Hadamard gate comprise a universal set of quantum gates [17, 18]. The Fredkin gate enables reversible computing [16], and also forms a universal set along with Hadamard gate [19]. The CXX gate appears in the syndrome operator measurement circuit [20] for quantum error correction algorithms, such as the Steane code [21] and surface code [15]. Although the Fredkin and CXX gates can be decomposed into three and two CX (i.e., CNOT) gates respectively [4], we avoid the decomposition-based approach and generate a procedure to design these three-qubit gates that achieve the same gate action over a shorter timescale.

Thus far Toffoli and Fredkin gates are achieved by decomposition into single- and two-qubit gates [22–24] in various physical systems, and yet none of these efforts have achieved the threshold fidelity [4]. Recently we have proposed a quantum control scheme (called SuSSADE) for designing a single-shot high-fidelity ($> 99.9\%$) Toffoli gate for a system comprising three nearest-neighbor-coupled superconducting artificial atoms [8]. In this work we show that this machine learning technique enables the design of other three-qubit gates as well for the same physical model. For all the three-qubit gates considered in this work we show that the gates operate as fast as a two-qubit entangling Controlled-Z (CZ) gate under the same experimental constraints.

Recent progress in superconducting artificial atoms [25–27] has made them appealing for quantum information processing, especially for gate-based quantum computing [28]. An avoided-crossing-based CZ gate [29] has already been achieved for a system comprising coupled superconducting artificial atoms.' The idea of avoided-crossing-based gates is to vary the energy level of artificial atoms such that energy levels approach each other but avoid degeneracies. These avoided crossings mix population and dynamical phases at the corresponding levels such that the final evolution of the system gives the required phase and population for the target gates.

Our strategy to design the three-qubit gates is also based on the avoided level-crossings. We employ an open-loop quantum control approach to generate optimal external pulses for the frequencies of the superconducting artificial atoms. We employ machine learning as a quantum control tool to generate successful procedures and show that our procedure can also be implemented in an open quantum system where external noise is also acting on the system.

Other approaches exist to design quantum gates for a network of superconducting transmon systems, where one can couple each transmon with a microwave generator [30] or couple the transmons via tunable couplers [31] and control these external circuit elements to evolve the system toward a specific unitary operation. These approaches require more resources (additional circuit elements) compared to our approach [8], where we only control the transmon frequency via a quantum control scheme [32, 33] and evolve the system's dynamics toward the target gate.

Machine learning [34] is concerned with the construction of algorithms that can learn from data and make predictions on data. Typical machine learning algorithms tend to be greedy [35, 36], as they need less resource and computational time to complete the learning task as well as converge faster (in comparison to non-greedy approaches [37, 38]). However, we have observed that greedy machine-learning techniques failed to generate a successful procedure for designing high-fidelity three-qubit gates, which motivates us to employ the non-greedy machine-learning technique. Our learning algorithm is based on an enhanced version of Differential Evolution (DE) algorithm [39], hence the name we assigned: Subspace-Selective Self-adaptive DE or SuSSADE [8].

The rest of the paper is organized as follows. In Section II we explain the physical model that we use to design the three-qubit gates. In Section III we discuss the avoided-crossing-based gates for two- and three-qubit gates. In particular, we review the current theoretical framework for designing avoided-crossing-based CZ gate and also discuss that why taking the same theoretical approach is challenging for avoided-crossing-based three-qubit gates. In Section IV we discuss our quantum control scheme and show that how we translate the problem of designing a three-qubit gate into a learning algorithm.

In Section V we discuss the noise model. In Section VI we discuss each individual three-qubit gate and giving their effect on a quantum state. Section VII presents the results. The significance of the results is outlined in Section VIII and we conclude our work in Section IX.

II. PHYSICAL MODEL

We consider a system comprising three nearest-neighbor-coupled superconducting artificial atoms [4] with parameters appropriate for the transmon system [40, 41]. Each transmon is capacitively coupled to its nearest neighbor, where the location of each transmon is labeled by $k = 1, 2, 3$. The frequency $\varepsilon_k(t)$, in the rotating-frame, can be tuned via superconducting control electronics. The anharmonicities of the second and third energy levels are represented by η and η'.

We approximate $\eta' = 3\eta$, which is valid for the cubic approximation of the potential for the transmon system [29]. The transmons are coupled capacitively, which yields an XY interaction between adjacent transmons (in the rotating frame) with a coupling strength g. The Hamiltonian for three capacitively coupled transmons is thus [29]

$$\frac{\hat{H}(t)}{h} = \sum_{k=1}^{3} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \varepsilon_k(t) & 0 & 0 \\ 0 & 0 & 2\varepsilon_k(t) - \eta & 0 \\ 0 & 0 & 0 & 3\varepsilon_k(t) - \eta' \end{pmatrix}_k$$
$$+ \frac{g}{2} \sum_{k=1}^{2} (\hat{X}_k \hat{X}_{k+1} + \hat{Y}_k \hat{Y}_{k+1}), \quad (1)$$

where

$$\frac{\hat{Y}_k}{i} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & -\sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & -\sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix}_k, \hat{X}_k = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix}_k$$

are the generalized Pauli operators [29].

The experimental constrains for the transmons require specific values for each physical parameter in Eq. (1). The transmon frequencies $\varepsilon_k(t)$ are varied between 2.5 and -2.5 GHz. We consider

$$\eta = 200 \text{ MHz}, g = 30 \text{ MHz}. \quad (2)$$

Although the physical system considered for this work consists of superconducting circuits, our quantum control scheme is, however, not limited to a specific system.

The Hamiltonian (1) generating the three-qubit gates acts on a 4^3-dimensional Hilbert space $\mathscr{H}_4^{\otimes 3}$. Under the rotating-wave approximation, this Hamiltonian is a block-diagonal matrix with each block corresponding to a fixed number of excitations. This block diagonalization property permits us to reduce the 4^3-dimensional Hamiltonian to a subspace where at most 3 excitations

are present, which is the relevant subspace for three-qubit gates.

We define a projection operator $\mathscr{O}_m$ that truncates the Hamiltonian (1) up to the mth excitation subspace, and for $m = 3$ (that allows at most 3 excitations, which is all what we need for three-qubit gates) the projected Hamiltonian is given by,

$$\hat{H}_{\mathrm{p}}(t) = \mathscr{O}_3 \hat{H}(t) \mathscr{O}_3^{\dagger}. \tag{3}$$

We observe that the unitary evolution of the system is unaffected by this truncation, which is what we expect from a block-diagonal Hamiltonian.

We evolve $\hat{H}_{\mathrm{p}}(t)$ such that the resultant unitary operator is

$$U(\Theta) = \hat{\mathcal{T}} \exp\left\{-\mathrm{i} \int_0^{\Theta} \hat{H}_{\mathrm{p}}(\tau) \mathrm{d}\tau\right\} \tag{4}$$

with $\hat{\mathcal{T}}$ the time-ordering operator [42]. Note that $U(\Theta)$ is a 20×20 unitary operator whereas the three-qubit gates reside in 2^3-dimensional computational subspace. We therefore, define another projection operator $\mathscr{P}$, which projects $U(\Theta)$ into the computational subspace of the three-transmon system

$$U_{\mathrm{cb}}(\Theta) = \mathscr{P} U(\Theta) \mathscr{P}^{\dagger}, \tag{5}$$

where U_{cb} is the projected unitary operator.

Our goal is to achieve these specific three-qubit unitary operation (Toffoli, Fredkin or CXX) over the duration Θ, such that the distance between $U_{\mathrm{cb}}(\Theta)$ and the target three-qubit gate is minimal. We evolve the system Hamiltonian (4) such that the final time evolution operator approaches to the target three-qubit gate modulo some phases that can be compensated by local z-rotations on each transmon. This phase compensation [29, 43] is performed via the excursions of transmon frequencies and is trivial for superconducting circuits.

To steer the system dynamics towards a specific entangling gate operation, we define the equivalence class of a given three-qubit gate U_{target} under local z-rotations as [29]

$$U_{\mathrm{target}} \equiv U'_{\mathrm{target}} = U_{\mathrm{post}} U_{\mathrm{target}} U_{\mathrm{pre}}, \tag{6}$$

where

$$U_{\mathrm{pre,post}}(\beta_1, \beta_2, \beta_3) \equiv R_z(\beta_1) \otimes R_z(\beta_2) \otimes R_z(\beta_3). \tag{7}$$

R_z in (7) denotes a unitary single-qubit rotation about the z-axis. Equation (7) can be explicitly expressed in terms of $\{\beta_j\}$, which is the set of local phases acquired by the jth transmon:

$$U_{\mathrm{pre,post}} = \mathrm{diag}\left(1, \mathrm{e}^{-\mathrm{i}\beta_3}, \mathrm{e}^{-\mathrm{i}\beta_2}, \mathrm{e}^{-\mathrm{i}(\beta_2+\beta_3)}, \mathrm{e}^{-\mathrm{i}\beta_1},\right.$$
$$\left.\mathrm{e}^{-\mathrm{i}(\beta_1+\beta_3)}, \mathrm{e}^{-\mathrm{i}(\beta_1+\beta_2)}, \mathrm{e}^{-\mathrm{i}(\beta_1+\beta_2+\beta_3)}\right). \tag{8}$$

We use $U_{\mathrm{pre,post}}$, which are diagonal 8×8 matrices operating on 2^3 dimensional computational subspace of the three qubits, to perform phase compensation in the numerical simulation of each three-qubit gate.
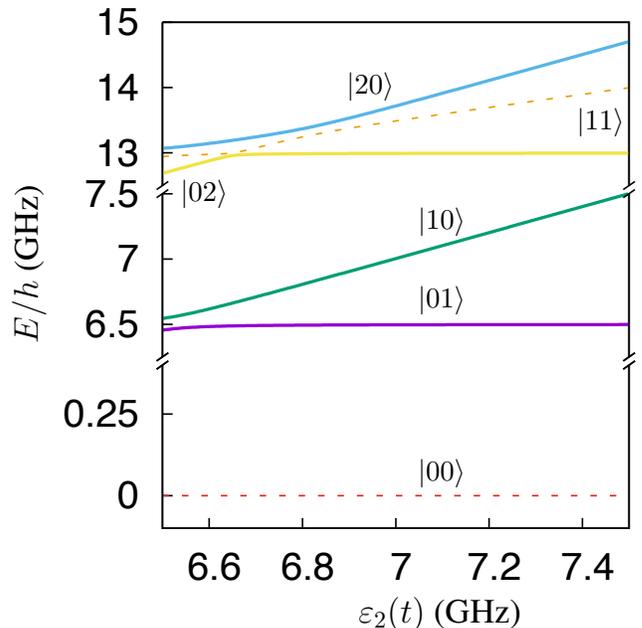

FIG. 1. (color online) The energy (E) spectrum of the system Hamiltonian of two capacitively coupled transmons. The frequency of the first transmon is fixed at 6.5 GHz. The frequency $\varepsilon_2(t)$ of the second transmon varies from 7.5 to 6.5 GHz

III. AVOIDED-CROSSING-BASED GATES

In this section, we first discuss the avoided-crossing-based technique in designing the two-qubit entangling CZ gate. Such a scheme was first proposed for a system of two coupled phase qubits [44], which was later adapted for a system of resonator-coupled superconducting qubits [29]. We describe the avoided-crossing-based CZ gate for the physical model of two capacitively coupled frequency-tunable transmons, with η and g ($\ll \eta$) being the anharmonicity and the coupling strength, respectively. This discussion is necessary to clarify why finding a theoretical solution for three-qubit gates is challenging, which is in fact the motivation for our quantum control approach.

A. CZ gate based on avoided level-crossing

CZ gate is a two-qubit entangling gate. It exerts a Pauli z-rotation on the second (target) qubit if and only if the first (control) qubit is $|1\rangle$. The CZ gate acts on the basis states according to

$$|00\rangle \mapsto |00\rangle, \quad |01\rangle \mapsto |01\rangle,$$
$$|10\rangle \mapsto |10\rangle, \quad |11\rangle \mapsto -|11\rangle, \tag{9}$$

which leaves three two-qubit basis states intact and imposes a sign change on one basis state.

The energy levels of two capacitively coupled transmons are shown in Fig. 1, where we fix the frequency of the first transmon at $\varepsilon_1 = 6.5$ GHz and allow the frequency of the second transmon $\varepsilon_2(t)$ to vary from $\omega_{\text{off}} = 7.5$ GHz (detuned frequency) to

$$\omega_{\text{on}} = \varepsilon_1 + \eta = 6.7 \text{ GHz}. \tag{10}$$

We vary the frequency of the second transmon using a time-dependent error function, with a switching time of t_{ramp} and the gate operation time of t_{gate}. This time-dependent frequency is [29],

$$\varepsilon_2(t) = \omega_{\text{off}} + \frac{\omega_{\text{on}} - \omega_{\text{off}}}{2} \left[\text{erf}\left(\frac{4t - 2t_{\text{ramp}}}{t_{\text{ramp}}}\right) \right.$$
$$\left. - \text{erf}\left(\frac{4t - 4t_{\text{gate}} + 2t_{\text{ramp}}}{t_{\text{ramp}}}\right) \right]. \tag{11}$$

The avoided-crossing-based CZ gate works as follows: Initially, we detune the transmon frequencies from each other by setting the frequency of the second transmon equal to $\varepsilon_2(t = 0) = \omega_{\text{off}}$. This makes all the eigenstates of the system non-degenerate. Then we tune the second transmon to ω_{on} for a time

$$t_{\text{on}} = t_{\text{gate}} - 2t_{\text{ramp}}, \tag{12}$$

and finally detune the second transmon again to the frequency ω_{off}. During the time t_{on}, the computational basis state $|11\rangle$ mixes with the two auxiliary levels $|02\rangle$ and $|20\rangle$, whereas all the other eigenstates in the computational basis ($|00\rangle$, $|01\rangle$, $|10\rangle$) are detuned from each other. The parameters of the control pulse are determined such that the mixing among $|11\rangle$, $|02\rangle$ and $|20\rangle$ states over the time interval t_{on} ensures the phase factors required for the CZ gate as shown in Eq. (9).

Depending on the timescale t_{on}, two distinct regimes exist in which a CZ gate can operate: the sudden-approximation regime and the adiabatic regime. In the sudden-approximation regime we vary the second qubit frequency fast enough, such that the switching time can be made sudden with respect to g (but still adiabatic with respect to η). For the sudden-approximation regime, the switching time has an inverse relation with the coupling factor as

$$t_{\text{on}} = \frac{\pi}{\sqrt{2}g}. \tag{13}$$

Under the sudden approximation, two parameters of the pulse ω_{on} and t_{on} must be optimized to obtain a high-fidelity CZ gate.

On the other hand, in the adiabatic regime, the switching is adiabatic with respect to g (and therefore, with respect to η as well). In this regime, stronger coupling is required between transmons to make the gate operate as fast as the gates in the sudden approximation regime. This increase of g leads to residual errors in a multiqubit device, whereas the advantage of operating the transmon in the adiabatic regime is that it suffices to optimize only one parameter. This is because the adiabatic

regime ensures that the population of each energy level is preserved. A CZ gate in the adiabatic regime is demonstrated in Ref. [45].

B. Avoided-crossing-based approach for three-qubit gates

Either in sudden or in adiabatic regime, the idea of engineering a pulse for the avoided-crossing-based gate in a superconducting system remains the same: Designing a control pulse for the qubit frequency, such that the $|11\rangle$ state mixes with the other states in the second excitation subspace, while the states in zero- and single-excitation subspaces remain detuned from each other. However, for practical implementations, both sudden and adiabatic regimes are unsuitable for obtaining the threshold fidelity required for fault-tolerance. Instead advanced machine-learning-based techniques can be employed to engineer optimal pulses, which is the motivation for our work.

One idea for designing three-qubit gates is to couple three transmons via a superconducting cavity, usually referred to as the circuit-quantum-electrodynamics (cQED) architecture [46], and tune the transmon frequencies in the dispersive regime such that the time-evolution operator gives rise to the target three-qubit at the end of operation. Such an approach has already been used to demonstrate a Toffoli gate [7]. We, however, do not consider the cQED hardware in our work, as the architecture can only contain a few transmons inside a superconducting cavity, and therefore, is not scalable.

Instead we consider a one-dimensional chain of three transmons with nearest-neighbor coupling. To see if such an approach is suitable for avoided-crossing-based three-qubit gates, we first plot the energy spectrum of such a three-transmon system in Fig. 2, where we fix the frequencies of the first and third transmons to 4.8 and 6.8 GHz, respectively, and vary the frequency of the second one from 4.5 to 7.5 GHz. We also set

$$g = 30 \text{ MHz}, \ \eta = 200 \text{ MHz}. \tag{14}$$

In contrast to the two-qubit case, the energy spectrum of a three-transmon system is crowded with many level-crossings (or anticrossings in the presence of interaction). For this system, therefore, finding optimal pulses theoretically for the transmon frequencies, such that the $|111\rangle$ state mixes strongly with the other states in the third excitation subspace, whereas all the other states are detuned from each other is a challenging task, which is why we employ the quantum control scheme and in particular the machine-learning technique to devise a procedure for designing such three-qubit gates [8].

IV. QUANTUM CONTROL

In this section we first give the application of quantum control and elaborate how a gate-design problem can be
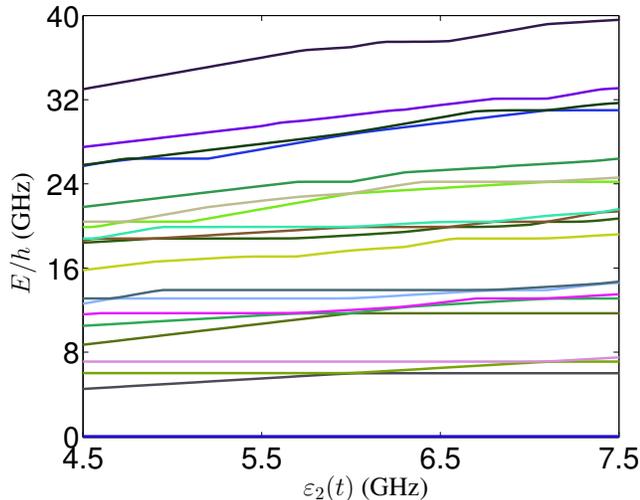
FIG. 2. (color online) The energy (E) spectrum of three nearest-neighbor-coupled transmons. The first and third transmon frequencies are fixed at 4.8 and 6.8 GHz respectively. The frequency of the second transmon varies from 4.5 to 7.5 GHz

transformed into a quantum-control problem. We introduce supervised machine learning as a tool to generate procedures for designing quantum gates. We explain the map from quantum-control problem to the learning problem in the second part of this section, and then show that generating a successful procedure via the learning algorithm becomes a feasibility problem for which optimization algorithms can be employed. We discuss some existing optimization algorithms that we employed to find feasible procedures, discuss how these algorithms remain inadequate to yield the required fidelity, and then introduce the enhanced version of differential evolution algorithm.

A. quantum-gate design as control problem

In the context of optimal control theory, the main task of quantum control is to investigate how to steer quantum dynamics towards a specific quantum state or operation [33]. The emergence of new quantum technologies have realized new applications for quantum control in various fields, such as femtosecond lasers [47, 48], nuclear magnetic resonance and other resonators [49–53], laser-driven molecular reactions [54, 55], and quantum gate synthesis for quantum computing [56]. In particular, we employ quantum control scheme to design fast and high-fidelity three-qubit gates.

As we explained earlier, a chain of three capacitively coupled transmons constitute the physical model for the three-qubit gates. This physical system evolves according to Eq. (4) and our goal is to steer the evolution toward the target three-qubit gate operation. In order to turn

the problem into a quantum-control problem, we need to clarify what the control parameters are. In our control scheme, control parameters are the qubit frequencies which can be tuned via external pulses. The task of finding the optimal shape for the external pulses can be performed via quantum control schemes. Machine learning can be employed as a quantum control tool to perform this task.

B. Supervised machine learning: a quantum control tool

The task of machine learning [57] is to develop algorithms which can learn from system behaviour and predict the future behaviour of the system based on their past evolution. Machine learning algorithms have already been applied to various problems in quantum information science, such as phase estimation [9], asymptotic state estimation [58], and discriminating quantum measurement trajectories and improving readout [59]. One can classify the machine learning algorithms in three distinct categories namely, supervised learnings, unsupervised learnings and reinforcement learnings [57]. Our focus is on supervised machine learning algorithm as a quantum control tool.

A supervised learning task [60] is to infer a function (hypothesis) from the labeled data (training set). The training data comprises an input vector accompanied by its corresponding output vector. A supervised learning algorithm trains the hypothesis on the training data to construct an inferred hypothesis. This inferred hypothesis can be further used to label novel data. Examples of supervised learning problems are regression or classification problems [61, 62]. The idea of supervised learning can be generalized to develop quantum control schemes that deliver successful procedures for quantum-gate design. In what follows, we describe the procedure for turning a gate-design problem into a supervised machine learning problem.

A quantum logic gate is a map between an input and an output state. One can always represent the action of any quantum logical gate on the basis elements in terms of a truth table. There is a one-to-one correspondence between the input and output elements in this truth table, as the quantum logic gates are themselves reversible. In the context of supervised learning problem, we consider the truth table as the training set. Loosely speaking, we train our hypothesis on the truth table data as the training set.

Having clarified that the truth table represents the training set, we now discuss what the hypothesis is. In the context of quantum-gate design, the hypothesis is the external pulses. Therefore, we train the parameters of the external pulses on the truth table data to shape the external pulses such that the system evolution approximates the target gate. If the hypothesis is learnt successfully, it generates a procedure which determines the shape and

strength of the external pulses. Thus far, we have explained the training set and hypothesis in the context of quantum-gate design. However, we still need to know how to measure the success of our learning procedure as well as the explicit form of the hypothesis (external field), which we elaborate in the next two subsections.

C. Control pulses as learning parameters

In an avoided-crossing-based gate, the energy levels of the artificial atoms are varied using the external pulses to steer the dynamics of the system toward the desired operation. Our learning algorithm uses these external pulses as the learning parameters. Although the quantum control approach can generate any type of pulses for $\varepsilon_k(t)$, we only consider two types of pulses: piecewise-constant and piecewise-error-function.

In the case of piecewise-constant function we discretize each transmon frequency, $\varepsilon_k(t)$, with $k \in \{1,2,3\}$ and express it as sum of N orthogonal constant functions over the interval $[0, \Theta]$. Each ε_k then can be shown as

$$\varepsilon_k := \begin{pmatrix} \varepsilon_{k_1} \\ \varepsilon_{k_2} \\ \vdots \\ \varepsilon_{k_N} \end{pmatrix} \qquad (15)$$

with each ε_{k_l} the magnitude of the kth pulse at the lth time step. The time bins are chosen to be equally spaced over the interval Θ and is therefore, given by

$$\Delta t = \frac{\Theta}{N-1}. \qquad (16)$$

The pulse generators for superconducting devices can generate the piecewise-constant functions. However, such rectangular-shaped pulses get distorted by the Gaussian filters that bridge the control circuitry with the transmons. More realistic pulse shapes that can take such distortions into account should therefore, be considered.

In order to consider the distortion on the rectangular pulses, we connect each control parameter in (15) by including following error-function coefficient [29] in

$$\varepsilon_k(t) = \frac{\varepsilon_{k_l} + \varepsilon_{k_{l+1}}}{2} + \frac{\varepsilon_{k_{l+1}} - \varepsilon_{k_l}}{2}$$
$$\text{erf}\left[\frac{5}{\Delta t}\left(t - \frac{t_l + t_{l+1}}{2}\right)\right] \qquad (17)$$

for $t_l \leq t \leq t_{l+1}$, t_l being the lth time step. The resultant smooth pulse expressed in (17) accounts for the first-order distortion caused by Gaussian filters. In Sec. VIII C 3 we explain how one can implement an adaptive control loop to suppress the higher-order noise.

In general, designing smooth pulses are computationally more expensive than computing for piecewise-constant functions. We therefore, choose less expensive

piecewise-constant functions to analyze the gate fidelity against various parameters but could incorporate other shapes if the extra computational cost is warranted. We justify this choice by showing that the gate fidelity does not depend on the type of the pulse, but depends on the number of learning parameters.

D. Confidence or fitness functional

The standard method to measure the performance of a supervised learning algorithm is to define a confidence for the learnt hypothesis. The confidence is the ratio of the number of training data that are learnt successfully to the total number of training data. If the learning task is to generate a procedure for designing a quantum gate, one can define this confidence, $\mathcal{F}$, by the distance between target and approximated unitary operators:

$$\mathcal{F} = \|\mathscr{P}U(\Theta)\mathscr{P}^\dagger - U_{\text{target}}\| \qquad (18)$$

with $\|\bullet\|$ the operator norm so $\mathcal{F}$ is the trace distance [63] between the target and actual evolution operators projected to the computational subspace.

Our machine learning algorithm uses an explicit form of (18), where

$$\mathcal{F} = \frac{1}{8}\left|\text{Tr}\left(U_{\text{target}}^\dagger \, U_{\text{cb}}(\Theta)\right)\right|. \qquad (19)$$

Note that $\mathcal{F} = 1$ if $U_{\text{cb}}(\Theta)$ corresponds to the target three-qubit gate U_{target} (Toffoli, Fredkin or CXX) and $0 \leq \mathcal{F} < 1$ otherwise.

In the context of quantum-gate design, the confidence functional $\mathcal{F}$ is called the intrinsic fidelity which refers to the fidelity between the unitary evolution of the closed quantum system of (1) and the target operation when the decoherence noise is ignored. We follow the standard practice of gate design by first considering a closed quantum system and generating a successful procedure for the learning task, and then evaluate the performance of generated procedure in the presence of noise. We call the confidence of our learnt hypothesis in the presence of noise as average state fidelity $\bar{\mathcal{F}}$ [8], which is

$$\bar{\mathcal{F}} := \frac{1}{8}\sum_k \sqrt{|\langle\psi_k|\,\rho_k^{\text{final}}\,|\psi_k\rangle|}, \qquad (20)$$

where ρ_k^{final} is the final density matrix of the system and $|\psi_k\rangle$ is the kth basis state in the computational subspace.

E. Machine Learning and optimization algorithms

For a gate-design problem, one can turn the problem of finding a successful procedure into a feasibility problem by setting the fidelity between the obtained unitary operation and the target gate to a fixed value that is acceptable by fault-tolerant quantum computing. All procedures that result in the error within the threshold value

are called feasible (successful) procedures. Various optimization algorithms can be employed to tackle this feasibility problem.

Greedy algorithms [35, 36, 64] are at the heart of the machine-learning techniques. Given a quantum control system with no physical constraints, finding a feasible procedure is a trivial task for greedy algorithms [65]. However, we have already shown [38] that reducing either gate operation time Θ or number of control parameters raises the difficult of the quantum-control problem. One might need to try different optimization algorithms to generate procedures for designing quantum gates under constraints of operation time and experimental resource.

We have examined the existing quantum control schemes including the quasi-newton [66–70] and Nelder-Mead [71] algorithms (greedy) as well as DE [39] and Particle Swarm Optimization [72] algorithms (non-greedy). We observed that all these schemes fail to generate a fidelity better than 99.5% (summarized in Table V), with the best fidelity obtained by DE. We thus enhance DE by introducing modifications into its standard version to enable feasible procedures that lead to intrinsic fidelity beyond 99.99%. We first give the details of the standard version of DE, and then discuss our machine learning approach, which is called the Subspace-Selective Self-Adaptive Differential Evolution (SuSSADE).

F. Differential Evolution

Differential evolution is an evolutionary algorithm (EA). Similar to other EAs, DE is inspired by biological evolution. The robustness and effectiveness of DE have already been studied by researchers in various fields such as machine learning [57], optimization [73], and image classification [74].

DE is a population-based search heuristic algorithm. Each member of the initial population breeds with the three random members of the same generation to generate a "daughter". The fittest of the original member and daughter survive to the next generation. Three distinct operations exist: mutation, crossover and selection. These operations form the mathematical structure of DE. We now explain each of these operations in details.

For each initial population member D_i with $i \in \{1, \ldots, P\}$ and P the population size, the mutation operation generates a trail vector M_i as follow:

$$M_i = D_{r_{i_1}} + \mu \left(D_{r_{i_2}} - D_{r_{i_3}} \right), \qquad (21)$$

where

$$\{r_{i_1}, r_{i_2}, r_{i_3}\} \in \{1, \ldots, P\} \qquad (22)$$

are discrete random numbers and mutation rate μ is a random number uniformly sampled from [0,1]. The mutation rate determines the step-size on the control landscape. A higher value of μ means that DE tends to explore the un-searched region of the landscape than exploiting the current knowledge about the landscape.

For each initial population member C_i and trial vector M_i, the crossover operation generates a target vector C_i, such that

$$C_i(j) = \begin{cases} M_i(j) & \text{if } r_{i_j} < \xi \\ D_i(j) & \text{otherwise,} \end{cases} \qquad (23)$$

where j is the index denoting the dimension of each population member. In our quantum scheme, the maximum value of j is equal to the number of control parameters K. r_{i_j} is a uniform random number sampled from [0,1]. ξ is the crossover rate of the algorithm. A higher ξ means that DE exploits the current knowledge about the quantum control landscape without spending too much time in searching the unexplored area of the landscape.

The last operation of DE is the selection operation, where we construct

$$D_i' := \begin{cases} C_i & \text{if } f(C_i) > f(D_i) \\ D_i & \text{otherwise,} \end{cases} \qquad (24)$$

with $f(C_i)$ being any fitness function, which is the fidelity function (19) in our case. In an iterative process, the resultant new population at generation G replaces population in the previous generation $G - 1$, and DE continues with the new generation. The iterative process aborts when either the threshold fidelity reaches or a predefined number of generations is attained.

Similar to other EAs, a DE algorithm faces two obstacles when applied to problems with a large number of dimensions. First, finding the optimal algorithmic-parameters i.e., μ and ξ, which lead to the best performance of DE is computationally expensive. Second, DE converges slowly to the promising region of the landscape where an optimal solution exists. We address these two weakness of DE by proposing the enhanced version, called Subspace-Selective Self-Adaptive DE (SuSSADE).

G. Subspace-Selective Self-Adaptive Differential Evolution

There are two approaches to find the optimal algorithmic-parameters for DE. One can run DE with many initial guesses to find the optimal parameters. This method is computationally expensive as it needs many trial runs of DE. This method also does not propose a general solution to the problem of finding the algorithmic-parameters because a new set of trial runs is needed if the learning problem is changed.

An alternative approach is to self-adaptively [75] change the parameters at each generation G as follows:

$$\mu_{i,G+1} = \begin{cases} \mu_l + r_1 . \mu_u & \text{if } r_2 < \kappa_1 \\ \mu_{i,G} & \text{otherwise} \end{cases} \qquad (25)$$

and

$$\xi_{i,G+1} = \begin{cases} r_3 & \text{if } r_4 < \kappa_2 \\ \xi_{i,G} & \text{otherwise,} \end{cases} \qquad (26)$$

where

$$r_j, j \in \{1, 2, 3, 4\} \tag{27}$$

are random numbers uniformly sampled from (0,1], and μ_l, μ_u, κ_1 and κ_2 are assigned to fixed values of 0.1, 0.1, 0.1 and 0.9, respectively, Using a self-adaptive version of DE improved the obtained fidelity up to 0.993, but the result is still sub-threshold because of the high dimensionality of the learning problem. Therefore, the next step is to enhance the self-adaptive DE using cooperative coevolution (CC) approach [76].

CC decomposes a high-dimensional problem into problems over several subspaces and evolve each subcomponent of the original high-dimensional problems using a choice of EA. CC then cooperatively combines the solution of each sub-component to form the final solution. There are several methods to decompose the high-dimensional problem into smaller sub-components [76–79]. The easiest method is to decompose a K-dimensional problem into K one-dimensional problems and evolve each component using EA.

This approach has drawbacks originating from ignoring the interdependencies between variables. One alternative method is to split the problem dimension in two halves and evolve each half over the course of learning process. However, should the K be high, the $K/2$ is also high, and therefore, the problem of a large number of dimensions is not addressed properly. Here we take a different approach in decomposing the high-dimensional problem. Our approach is inspired by Differential Evolution with Cooperative Coevolution (DECC-II) [80]. We first give a short description on DECC-II and then explain the enhanced version of DE.

DECC-II decomposes a K-dimensional problem into m-dimensional subspaces and run the learning algorithm on a subspace for a fixed number of cycles s while keeping the other subspaces unchanged. DECC-II combines the CC with Self-Adaptive DE with Neighborhood Search (NSDE) to address both the issues: the slow-convergence of DE and finding the optimal algorithmic parameters.

Using DECC-II in its original form did not improve the fidelity and the run-time and convergence of DE still remained slow in our case, as evaluating the fitness function is computationally expensive. We also found our result to be influenced by the choice of s and m and finding the optimal values of these parameters needed additional computational resources and trial runs, and even with this overhead the threshold fidelity was not achieved.

Inspired by the DECC-II algorithms, we set $s = 1$ and choose the dimension of subspace based on a random projection over the m-dimensional subspace with $m \in \{1, 2, 3, 4, 5\}$. Now, there is no need to look for the optimal values of s and m to perform the optimization. This new strategy, however, makes the convergence slower, as in each generation, only a small part of the candidate solutions are being selected for the optimization.

In order to enhance the convergence, our algorithm randomly switches breeding between the subspace and the whole space according to the value of an input switch parameter $S \in [0, 1]$, such that a uniformly distributed random number $r_j \in [0, 1]$ at generation j restricts breeding to the subspace, if $r_j < S$, and breeds in the whole space otherwise. As our algorithm selects an m-dimensional subspace at each generation and selfadaptively evolves the mutation and crossover rates, we call it Subspace-Selective Self-Adaptive DE (SuSSADE). For our purpose, we observe that choosing $m = 1$ suffices, which signifies that the selected subspace is trivial. We refer to this one-dimensional extreme case as 1DSuS-SADE.

V. NOISE MODEL

In order to incorporate decoherence into our system's evolution, we model each transmon as a harmonic oscillator suffering from the environmental effect [81]. The decoherence of each harmonic oscillator is represented by two damping rates: amplitude and phase damping with the corresponding amplitude relaxation time T_1 and dephasing time T_2. In order to make the noise model simpler, we assume $T = T_1 = T_2$, which is a valid assumption for transmons with tunable frequency [29].

Equation (4) expresses the system evolution in the absence of decoherence. In the presence of noise we model the system evolution by the time-dependent density matrix which is decomposed in terms of the sum of Kraus matrices as $\mathcal{L}_k$ [82]

$$\rho(t) = \sum_{k=0}^{n} \mathcal{L}_k(t)\rho(0)\mathcal{L}_k^\dagger(t), \tag{28}$$

with Kraus matrices satisfying the completeness relation

$$\sum_{k=0}^{n} \mathcal{L}_k^\dagger \mathcal{L}_k = \mathbb{1} \tag{29}$$

at each time step. We first discuss the construction of the Kraus matrices for each transmon.

Constructing the Kraus matrix representation for a system comprising three capacitively coupled transmons, will then be straightforward by performing all the possible tensor products of three transmons. We also assume that the decoherence only affect each individual transmon separately.

A. Amplitude damping

The Kraus matrix representation of amplitude damping of a single qubit (treated as a truncated harmonic oscillator) coupled to the environment can be modeled as multimode oscillators [81]. One can easily generalize this approach to represent the Kraus representation of amplitude damping of a single transmon (modeled as a

four-modal harmonic oscillator) coupled to environment modeled as multimode oscillators, where

$$A_l(t) = \sum_{j=l}^{3} \sqrt{\binom{j}{l}} \left(e^{-\frac{t}{T_1}}\right)^{\frac{j-l}{2}}$$

$$\left(1 - e^{-\frac{t}{T_1}}\right)^{\frac{l}{2}} |j-l\rangle \langle j| \qquad (30)$$

with $l \in \{0, 1, 2, 3\}$ labeling the Kraus matrices and $\exp(-t/T)$ represents the amplitude damping factor, which decays exponentially with the timescale T_1.

B. Phase damping

We express the Kraus matrix representation of the phase damping of a single transmon (with four energy levels) coupled to the environment [81] as

$$\mathbb{A}_l(t) = \sum_{j=0}^{3} \exp\left\{-\frac{j^2 t}{2T_2}\right\} \sqrt{\frac{(j^2 t / T_2)^l}{l!}} |j\rangle \langle j| \qquad (31)$$

with $l \in \{1, 2, 3, 4\}$. $\mathbb{A}_l$ is a Kraus matrix, and T_2 indicates the timescale for dephasing. In both equations (30) and (31), we need to put an upper bound on the Kraus-Matrix index l to enable numerical simulation of decoherence.

Such an upper bound would violate the completeness relation (29). However, if the evolution time of the quantum system t is much smaller than the coherence time T, the higher-order terms in (30) and (31) damp exponentially with respect to the t/T. As in our case $t \ll T$, we only consider l up to three for both amplitude and phase damping, and discard the higher-order terms in our numerical calculation as they have negligible effects.

VI. THREE-QUBIT LOGICAL GATES

In this section we give the details of the Toffoli, Fredkin and CXX gates. We discuss the circuit model of each gate as well as give the truth tables showing how the basis states transform under the actions of these gates.

A. Toffoli gate

A Toffoli gate is a three-qubit gate that applies a Pauli X operation on the third qubit if the quantum state of the first two control qubits are $|11\rangle$, and does nothing otherwise. A Toffoli gate is also called a Controlled-Controlled-NOT (CCX) gate, which is equivalent to a Controlled-Controlled-Z (CCZ) gate up to a local Hadamard transformation on the target qubit [8].

CCZ gate is a three qubit gate that applies a Pauli-Z operator on the third qubit if the quantum state of the first two qubits are $|11\rangle$ and, otherwise leaves the state

Input			Output		
C_1	C_2	T	C_1	C_2	T
$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$
$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$
$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$
$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$
$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$
$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$
$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$
$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$-|1\rangle$

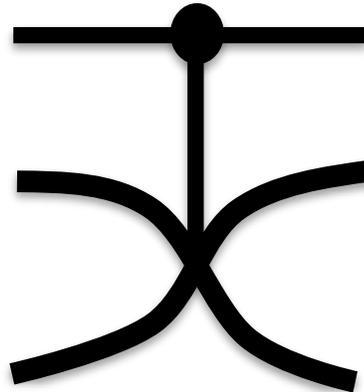TABLE I. The truth table for CCZ gate. C_1 and C_2 denote the control qubits and T represents the target qubit.

FIG. 3. The quantum circuit representation of Fredkin (Controlled-Swap) gate. The horizontal solid back line is the circuit wire, ● denotes the control qubit and the big cross sign shows the SWAP gate which acts on the target qubits (second and third qubits).

of the third qubit unchanged. In order to design the Toffoli gate, we only show how to implement the CCZ gate, since the Hadamard gates are trivial for superconducting circuits [3, 83].

In our supervised learning method, the truth table of the CCZ gate represents the training set. One can define the truth table of the CCZ gate based on its action on the three-qubit basis states (See Table I).

So far, the design of a high-fidelity Toffoli gate has been investigated in several physical systems with the achieved fidelities limited to 81% in a post-selected photonics circuit [84], 71% in an ion trap system [10], 68.5% in a three-qubit circuit QED [11] and 78% in a four-qubit circuit QED [7]. We recently proposed a quantum control approach to design a high-fidelity (>99.9%) Toffoli gate for a system comprising three nearest-neighbor-coupled transmons [8].

Input			Output		
C	T_1	T_2	C	T_1	T_2
$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$
$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$
$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$
$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$
$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$
$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$
$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$
$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$

TABLE II. The truth table representation of Fredkin gate. C denotes the control qubit and T_1 and T_2 represent the target qubit. The column under the Output and Input columns show the state of the three qubits before and after applying the Fredkin gate.

Input			Output		
C	T_1	T_2	C	T_1	T_2
$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$
$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$
$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$
$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$
$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$
$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$-|1\rangle$
$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$-|1\rangle$	$|0\rangle$
$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$

TABLE III. The truth table representation of CZZ gate. C denotes the control qubit and T_1 and T_2 represent the target qubits. The columns under the Output and Input show the states of the three qubits before and after applying CZZ.

B. Fredkin gate

A Fredkin gate (See Fig. 3) is a three-qubit gate that applies a SWAP operation between the second and the third qubits, if the state of the first qubit is $|1\rangle$, and leaves the state of the qubits unchanged otherwise. A Fredkin gate is an excitation-number-preserving operation, where the output state has the same number of excitations as that for the input states. It is also a self-inverse operation, which means that applying two consecutive Fredkin gates generates the Identity operation. The Fredkin gate is universal for reversible classical computation, as it can be used to construct any other reversible classical logic gates [85].

The truth table of Fredkin gate is shown in Table II. In this representation C denotes the control qubit and T_1 and T_2 represent the target qubits. This truth table provides the training set for Fredkin for supervised learning.

Proposals to implement the Fredkin gate are mainly restricted to the context of linear and nonlinear optical systems, and there has been no proposals yet for a single-shot Fredkin gate with superconducting circuits. Here we employ the quantum control scheme to design a fast single-shot high-fidelity Fredkin gate for a system comprising three nearest-neighbor-coupled superconducting transmon systems.

C. CXX

CXX is a three-qubit gate that applies Pauli-X operations on the second and third qubits when the first qubit is $|1\rangle$, and leaves the state of the qubits unaltered otherwise. CXX gate is equivalent to Controlled-Z-Z (CZZ) gate up to local Hadamard operations on both the second and the third qubits:

$$\text{CXX} = [\mathbb{1}\otimes H \otimes H]\,\text{CZZ}\,[\mathbb{1}\otimes H \otimes H]. \qquad (32)$$
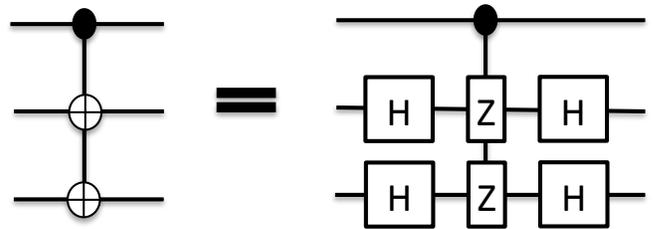

FIG. 4. The quantum circuit representation of the CXX gate (left), which is equivalent to the CZZ gate up to local Hadamard gates (right). The horizontal solid black lines are circuit wires, ● represents the control qubit and ⊕ denotes the Pauli-X operator acting on the target qubit. The boxes with Z and H denote the Pauli-Z and Hadamard operations, respectively.

The CZZ gate is a three-qubit gate that applies a Pauli-Z operation on the second and the third qubits, if and only if, the first qubit is $|1\rangle$. In order to design the CXX gate, we also assume that the implementation of fast and high-fidelity Hadamard gates are trivial for superconducting circuits [3, 83]; hence we employ the quantum control scheme to design the high-fidelity CZZ gate.

CZZ gate can easily be constructed by consecutive operations of two CZ gates. A high-fidelity CZ gate has already been proposed with the gate time about 26 ns [29]. Therefore, a high-fidelity CZZ gate decomposed into two consecutive CZ gates can be readily designed with a gate operation time no longer than 52 ns. However, a CZZ gate with a shorter operation time is more useful for quantum error correction, which motivates us to employ the quantum control scheme for this problem.

The truth table representation of a CZZ gate is shown in Table III. In this table, C denotes the control qubit and T_1 and T_2 represent the target qubits. We use the truth table data as the training set to train the qubit frequencies (our hypothesis).

VII.   RESULTS

In this section, we first review the results for designing a high-fidelity single-shot Toffoli gate from Ref. [8], and then present our analysis of performance of the Toffoli gate against the effect of random noise on the learning parameters (not given in Ref. [8]). Next we present the results for designing single-shot high-fidelity Fredkin and CXX gates. For each of these gates, we first give the optimal piecewise-constant and piecewise-error-function pulses that steer the system evolution towards the threshold fidelity. We then investigate the dependence of intrinsic fidelities on the coupling strength between transmons.

We test robustness of our optimal pulses for Fredkin and CXX gates by applying uniformly distributed random noise on each of the learning parameters and then calculating the intrinsic fidelity in the presence of such noise. Finally, we give the result for the average state fidelities in the presence of decoherence induced noise.

A.   Toffoli

A set of piecewise-constant and piecewise-error-function pulses are obtained via the SuSSADE algorithm for a single-shot high-fidelity Toffoli gate, which operates over 26 ns (See Fig. 1 in [8]) and has the same number of learning parameters for both pulses. The resultant fidelity for the Toffoli gate is higher than 0.999, even in the presence of decoherence (See Fig. 3 in Ref. [8]).

We have also explored the dependence of the intrinsic fidelity over the time of the system evolution by fixing the coupling strength g to various values and running the SuSSADE for the less-computationally-expensive piecewise-constant pulses. The control pulses are discretized into equally-spaced time intervals of 1 ns, which give rise to learning problems with different learning parameters. For each value of g we have also shown the dependence of intrinsic fidelity over the gate operation time (See Fig. 2A in Ref. [8]).

We study the robustness of the designed pulse against the random noise on the learning parameters. In order to test the robustness we choose a sample optimal pulse for the Toffoli and add random values ($\delta\varepsilon \times \mathrm{rand}(-1,1)$) to the learning parameters at each time-bin, with $\delta\varepsilon$ varies from 0 to 3000 KHz. We then use the distorted pulse to calculate the intrinsic fidelity for each value of $\delta\varepsilon$. Figure 5 shows the change in intrinsic fidelity originated from such random noise on learning parameters.

B.   Fredkin

We employ SuSSADE to design a high-fidelity Fredkin gate for a system comprising three nearest-neighbor-coupled superconducting transmons. The system Hamiltonian evolves over 26 ns under the piecewise-constant
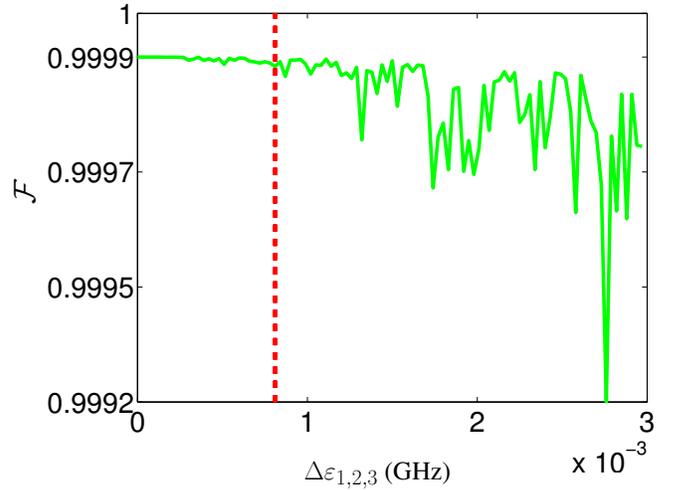

FIG. 5. (color online) Intrinsic fidelity of the Toffoli gate $\mathcal{F}$ as a function of $\delta\varepsilon$ for the CCZ gate. The vertical red dotted-line denotes the threshold, such that $\mathcal{F} > 0.9999$ on the left of the dotted line.

pulse (See Fig. 6A), where the final evolution approximates a Fredkin gate with an intrinsic fidelity of $\mathcal{F} = 0.9999$. We also show a more realistic piecewise-error-function pulse in Fig. 6B. The learning algorithm uses the same number of learning parameters to shape the transmon frequencies through the learning procedure.

In order to show that the efficacy of the quantum control approach does not depend on the type of the gate, we conduct the same analyses on the Fredkin gate as we did for Toffoli. In Fig. 7A we analyze the change in intrinsic fidelity with the gate operation time for different values of coupling strengths. In Fig. 7B we fix the intrinsic fidelity at $\mathcal{F} = 0.999$ and compute the relation between coupling strength and the inverse of the gate operation time, where the points on the curve show the actual numerical data, and the solid line is the cubic-fitted interpolation plot.

Figure 8 shows the evolution of the system under the decoherence. We set $g = 30$ MHz and evolve the system towards the Fredkin gate with an intrinsic fidelity higher than $\mathcal{F} = 0.9999$. Then we apply the noise model on each transmon to analyze how the fidelity changes over the coherence time, T, of each transmon. Under our noise model each transmon goes under amplitude- and phase-damping, and we assume $T = T_1 = T_2$ for tunable transmons.

Figure 9 shows the effect of random noise on the learning parameters of the Fredkin gate. We choose the optimal pulse shown in Fig.6 and apply random noise up to 3000 KHz on the learning parameters. We then use the distorted pulse to investigate the change in the intrinsic fidelity as a function of the applied random noise.
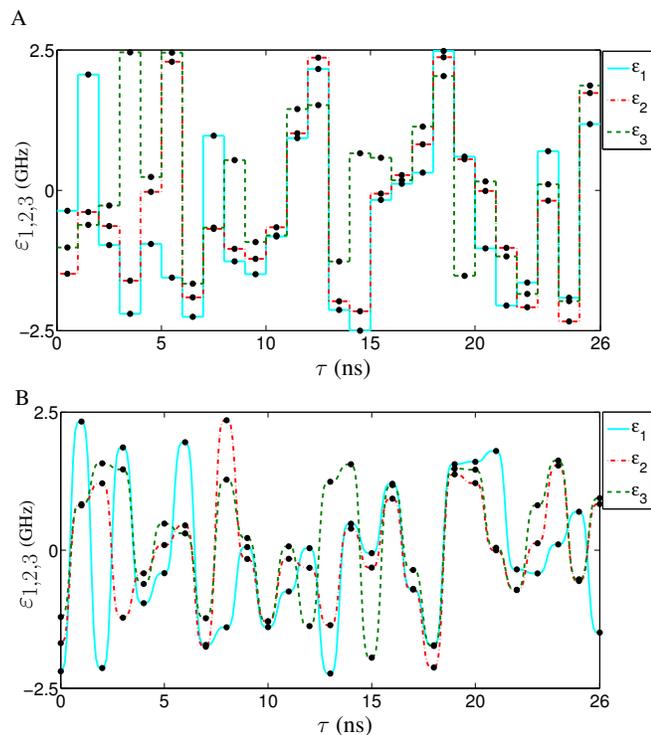
A

B

FIG. 6. (color online) Optimal pulses for designing Fredkin gate with the resultant fidelity better than 0.999 and the gate operation time of 26 ns. System frequencies, ε_i, are varied from -2.5 to 2.5 GHz, which are within the experimental requirements of transmon implementation. The black dots denote the learning parameters for SuSSADE. A) The piecewise-constant pulses for each transmon frequency. B) The piecewise-error-function pulses for each transmon frequency.

C. Controlled-Z-Z (CZZ)

Fig. 10A shows the piecewise-constant pulse generating a high-fidelity CZZ gate in 31 ns with an intrinsic fidelity higher than 0.9999. We have 31 learning parameters for each pulse (93 in total) to design the pulse shapes for the transmon frequencies. Under the piecewise-error-function pulse (shown in Fig. 10B), the system evolution approximate CZZ gate with a fidelity higher than 0.9999 in the same gate operation time as the piecewise-constant pulse. In designing the optimal pulses in Fig. 10 we set $g = 30$ MHz.

We perform the same analysis on the dependence of intrinsic fidelity on the gate operation time, as we did for Toffoli and Fredkin gates. The analysis is performed for various values of coupling strengths and is shown in Fig. 11A. The discrete points on the plot show the actual data and the curves are the cubic-fit to the data. Fig. 11B represents the relation between the inverse of the gate operation time and coupling strength. The discrete points on the plot show the actual numerical data and the solid line is the linear fit to data.
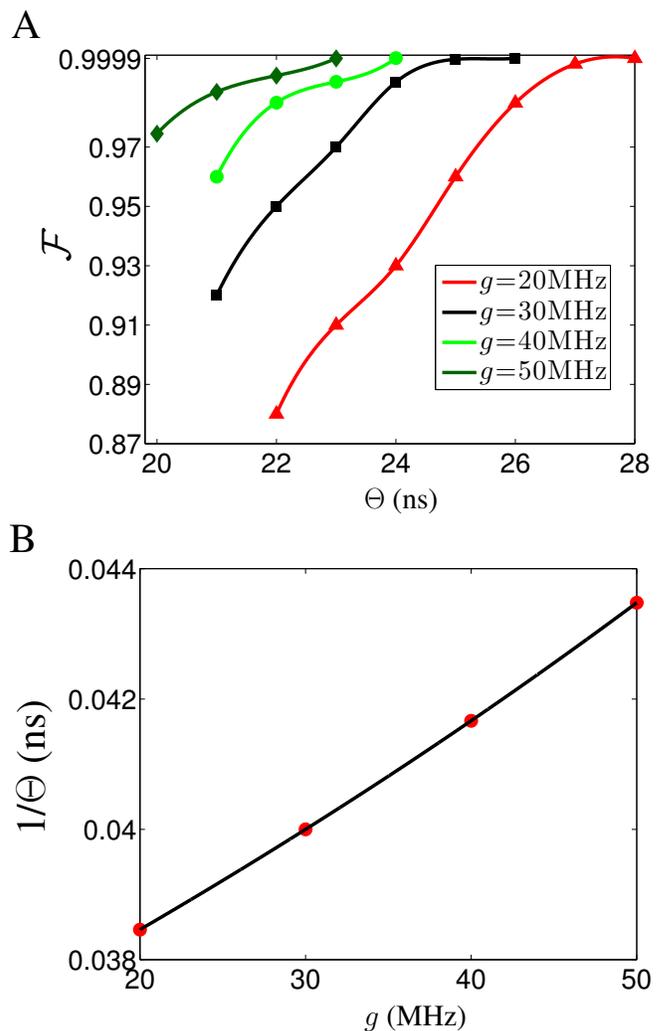
Figure 12 shows the decoherence induced noise for the

A

B

FIG. 7. (color online) A) The dependence of intrinsic fidelity of the Fredkin gate on the evolution time τ of the system for various values of g. The discretized values show the actual numerical data with $\triangle$, $\square$, $\circ$, $\diamond$ corresponding to the values of g to be 20, 30, 40, 50 MHz, respectively. A cubic interpolation fits the curves to the data. B) The relation between the inverse of the gate operation time and coupling strength between transmons where the dots denote the actual numerical results for various values of $g \in \{20, 30, 40, 50\}$. A linear-fit interpolates the points to the actual data.

CZZ gate. We plot fidelity vs transmon coherence time. The actual discrete points are connected via a linear interpolation. Similar to the Toffoli and Fredkin gates, the decoherence appears in terms of the amplitude- and phase-damping on the transmons. Here the coupling strength is $g = 30$ MHz.

We follow the same procedure as for the Toffoli and Fredkin gates to test the robustness of our designed pulse for the CZZ gate. We employ the optimal pulse in Fig. 10 and apply random noise on each learning parameter. Figure 13 shows the effect of such random noise on the intrinsic fidelity of the CZZ gate which operates in 31 ns.
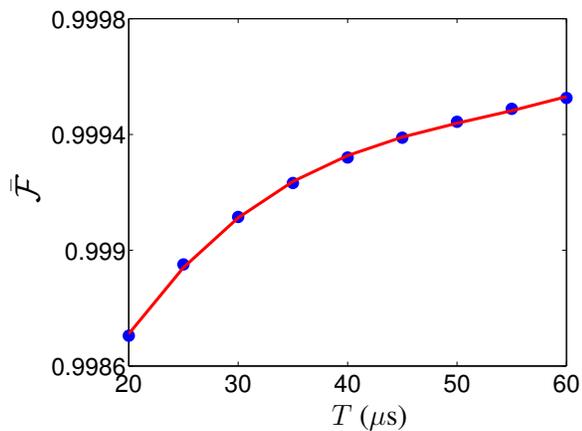
FIG. 8. (color online) Fidelity vs coherence time for the Fredkin gate. The dots denote the actual numerical data and the red solid line shows a cubic-fit interpolation on the actual data.
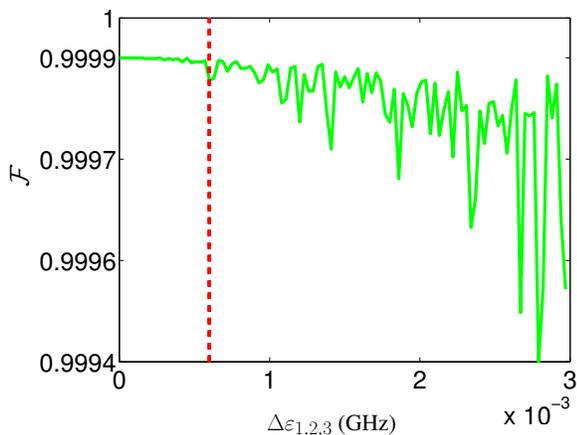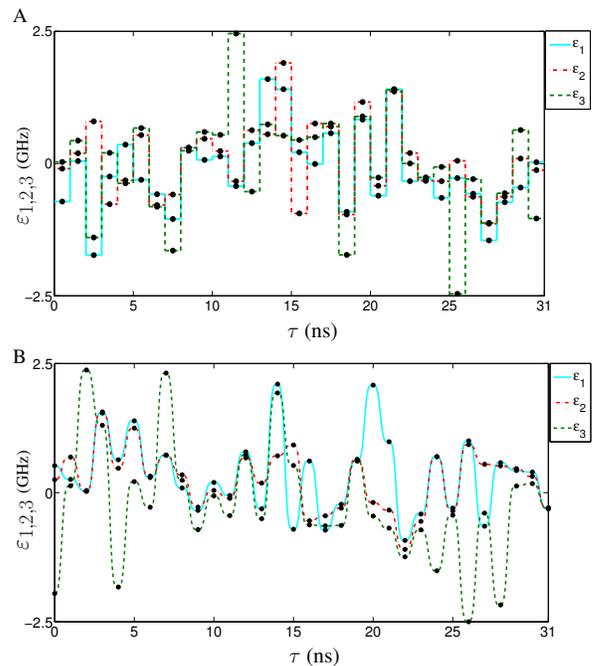

FIG. 9. Intrinsic fidelity $\mathcal{F}$ vs random noise applied to the optimal pulse of the Fredkin gate. The vertical red dotted-line denotes the threshold, such that on the left side of the line $\mathcal{F} > 0.9999$.

FIG. 10. (color online) Optimal pulses for designing CZZ gate with the resultant fidelity better than 0.999 and the gate operation time of 31 ns. System frequencies, ε_i, vary from -2.5 to 2.5 GHz which are within the experimental constraints of transmon implementation. The black dots denote the learning parameter for SuSSADE. A) The piecewise-constant pulses for each transmon frequency. B) The piecewise-error-function pulses for each transmon frequency.
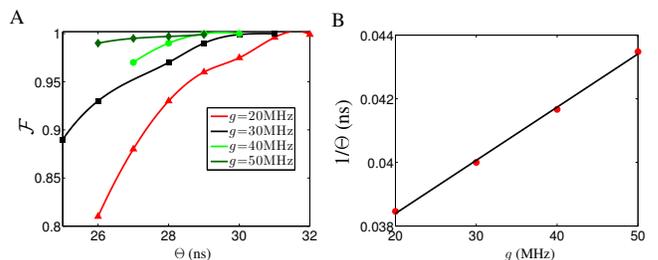

FIG. 11. (color online) A) The dependence of the intrinsic fidelity of the CZZ gate on the evolution time τ of the system for various values of g. The discretized values show the actual numerical data with $\diamond$, $\triangle$, $\circ$, and $\square$ corresponding to the values of g to be 20, 30, 40, 50 MHz, respectively. A cubic interpolation fits the curves to the data. B) The relation between the inverse of the gate operation time and coupling strength between transmons where the dots denote the actual numerical results for various values of $g \in \{20, 30, 40, 50\}$. A linear-fit interpolates the points to the actual data.

VIII. DISCUSSION

Gates	Θ(ns)	$\mathcal{F}$	$T(\mu s)$	$\overline{\mathcal{F}}$
Toffoli	26	0.9999	30	0.9992
Fredkin	26	0.9999	30	0.9991
CXX	31	0.9999	30	0.9990

TABLE IV. Comparison of fidelities among various three-qubit gates for $g = 30$ MHz. Θ, $\mathcal{F}$, T and $\overline{\mathcal{F}}$ are total gate time (in nanoseconds), intrinsic fidelity (defined by Eq. (19)), coherence time (in microseconds) and average state fidelity respectively.

We employed a one-dimensional system comprising three coupled superconducting artificial atoms to design single-shot high-fidelity three-qubit gates, such as Tof-foli, Fredkin and CXX. Our results include the optimal pulses for each three-qubit gate, analysis of dependence of the intrinsic fidelity on the physical model parameters, and analysis of performance of these gates under decoherence induced noise. Table IV shows a comparison among the numerical simulation of optimal Toffoli,
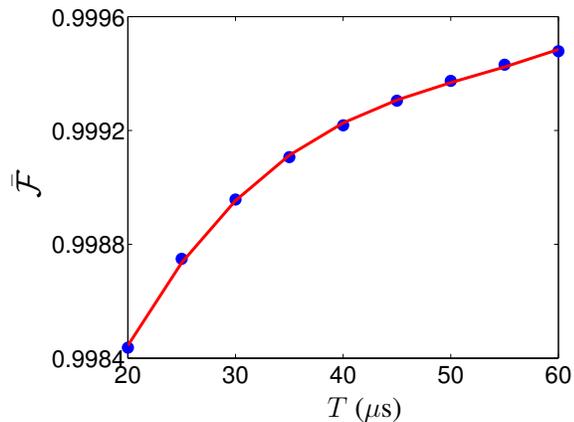
FIG. 12. (color online) Fidelity vs coherence time for the CZZ gate. The dots denote the actual numerical data and the blue solid line shows a cubic-fit interpolation on the actual data.
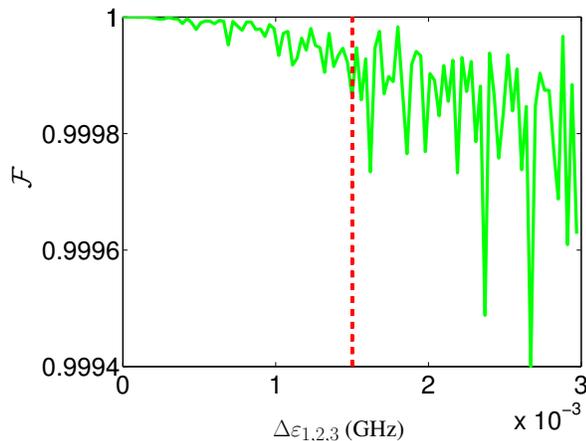

FIG. 13. Intrinsic fidelity $\mathcal{F}$ vs $\delta\varepsilon$ for the CZZ gate. The vertical red dotted-line denotes the threshold, such that on the left side of this line $\mathcal{F} > 0.9999$.

Fredkin and CXX gates for the same coupling strength $g = 30$ MHz. The results indicate that the efficacy of SuSSADE is independent of the desired quantum operations. In this section, we discuss the results for all three gates. We first elaborate on our choice of the control pulses. Then we discuss the dependence of fidelities for three-qubit gates on the physical model parameters, discuss about the effects of noise, and finally compare the SuSSADE scheme against alternative approaches.

A. Control pulses

To formulate the problem of designing the high-fidelity quantum gates into a learning algorithm we represent the qubit frequency in terms of external control functions (hypothesis). The choice of the control function is ubiquitous and we only choose two pulse profiles, piecewise-constant and error-piecewise-constant, which are relevant for superconducting control electronics. The learning algorithm shapes the external pulses to obtain high-fidelity quantum gates. Each of these control pulses has its own advantage and drawback in terms of the computational resource and practical implementation.

The piecewise-constant control function are computationally less expensive, and on average, a single run of the learning procedure using the square pulse takes an order of magnitude less run-time in contrast to the piecewise-error-function pulses. Piecewise-constant pulses are easy to generate using the current superconducting control electronics. However, the Gaussian filters connecting the control electronics and the physical qubits cause distortion on the square pulses. Transmons thus receive a distorted pulse. A smooth pulse must, therefore, be generated to account for the first order of distortion numerically.

The piecewise-error-function connects the control parameters smoothly such that the function approximates the realistic control pulses for transmon system. In this way we overcome the problem of infinite bandwidth of the square pulses as well as the first order distortion. We have numerically shown that (See [8] and Figs. 6 and 10) the learning procedure does not depend on the shape of the control pulse but rather depends on the number of learning parameters.

For designing each high-fidelity three-qubit gate, we used the same number of learning parameters either for the piecewise-constant or piecewise-error-function pulses. For Toffoli and Fredkin gate, we used 3×26 parameters to design the gates, which operate on 26 ns using either piecewise-constant or piecewise-error-function pulses (See [8] and Fig. 6). The number of learning parameter to design a CZZ gate (Fig. 10), which operates over 31 ns, is 31×3 for both piecewise-constant and piecewise-error-function pulses.

B. Intrinsic fidelity

In [8] and Figs. 7A, and 11A, we plot the intrinsic fidelity as a function of the gate operation time for various values of $g \in \{20, 30, 40, 50\}$MHz. Keeping the value of g fixed, the fidelity is a monotonically increasing function of the gate operation time. This is consistent with the notion that increasing the evolution time of the system increases the fidelity between the unitary evolution and the target gate [8]. As the coupling strengths become stronger, the gate operation time to reach to a threshold fidelity becomes shorter. This is consistent with the idea of avoided-crossing-based gates (13). For smaller values of g, the gate operation time becomes longer, and fixing the time-bin to 1 ns, the number of learning parameters increase. Our learning algorithm still delivers the high-fidelity gates despite the increase of the learning parameters.

In [8] and Figs. 7B, and 11B, we have shown that a

linear relation between the inverse of the gate operation time and coupling strength exists. Providing a theoretical framework to explain this relation is a challenging task for three-qubit gates. However, from our discussion for avoided-crossing-based CZ gate (Sec. III A), one can intuitively expect a linear relation between the coupling strength and the gate operation time for a fixed intrinsic fidelity, even for three-qubit avoided-crossing-based gates. This linear relation also emphasizes the efficacy of the learning algorithm to discover the correct relation between the coupling strength and gate operation time.

C. Noise

In this section, we explain the effect of noise on the system evolution of each three-qubit gate in three parts. In the first part, we discuss the effect of random noise on the system and discuss the robustness of our procedure. In the second part, we explain the effects of decoherence induced noise, which are caused by the environment. In the last part of this section, we discuss how the higher orders of noise on the generated procedure can be suppressed.

1. Robustness

We test the robustness of our procedure for designing the three-qubit gates by applying random noise on the learning parameters. Our analysis shows that the devised procedures to design Toffoli, Fredkin and CZZ gates are robust against the external random noise (Figs. 5, 9, 13), if the magnitude of the random noise is lower than 800, 600 and 1500 KHz respectively, which are within the limit of current state-of-the-art superconducting control-electronics [86].

2. decoherence induced noise

The amplitude-damping and dephasing rates (T_1^{-1} and T_2^{-1}) determine the decoherence rate of our three-transmon system. Assuming $T = T_1 = T_2$ for all the tunable transmon devices, we can plot $\bar{\mathcal{F}}$ versus the coherence time of the transmons (See Ref.[8] and Figs. 8, 12). For a gate operation time much faster than the coherence time of the transmons, i.e., $T \gg \Theta$, if the intrinsic fidelity is much smaller than the fidelity, with this reduction caused by decoherence induced errors, we can approximate

$$\bar{\mathcal{F}} \sim 1 - \frac{\Theta}{T}. \tag{33}$$

Equation (33) matches our numerical simulation of decoherence in [8] and Figs. 8 12.

When the coherence time of transmon is significantly higher than the gate operation time i.e., $T \gg \Theta$, the error

from the intrinsic fidelity is the main source of noise. Under this condition, the intrinsic fidelity must be so high, such that the resultant gate fidelity meets the threshold fidelity for fault-tolerant quantum computing. With the long coherence time ($20 \sim 60\mu$s) of the state-or-the-art superconducting artificial atoms accompanied with the machine learning approach that delivers a gate with $\bar{\mathcal{F}} > 0.999$, our proposal enables the implementation of high-fidelity three-qubit gates under current experimental conditions.

3. Distortion of control pulses

The learning algorithm can shape any type of external pluses to design high-fidelity three-qubit gates. We employed the piecewise-error-function pulse to resolve the infinite bandwidth problem of the square pulses and to account for the first-order distortion on the pulse. However, higher order distortions can change the optimal shape of the designed pulse and can lead to a sub-threshold fidelity. For example, we ignored the weak dependence of g and η on the transmon frequency.

This frequency dependence of physical parameters can introduce small perturbations into the system Hamiltonian, thereby distorting the optimal pulses. One viable option to suppress the higher degree of distortions on the learning parameters is closed-loop learning control [87, 88], which can be used in conjunction with our control scheme.

D. Comparison against alternative approaches

Method	$\mathcal{F}_{\text{best}}$
Quasi-Newton	0.9912
Simplex	0.9221
DE	0.9931
SuSSADE	0.9999

TABLE V. Comparison against alternative approaches of designing Toffoli gates for $g = 30$ MHz and $\Theta = 26$ ns.

Table V shows a comparison among various approaches of generating the optimal pulse shapes for the three-qubit Toffoli gate. $\mathcal{F}_{\text{best}}$ denotes the best intrinsic fidelity, defined by Eq. (19), obtained for the corresponding method. Note that we have compared SuSSADE against both the greedy optimization algorithms (Quasi-Newton and Simplex) as well as against the global optimization algorithms (DE). The comparison data shown in Table V indicates the efficacy of SuSSADE over the alternative approaches.

IX. CONCLUSION

In conclusion, we have transformed a problem of designing three-qubit gates into a quantum-control problem. The control problem is then mapped into a supervised machine learning algorithm. In the context of supervised machine learning, we used the truth-table data for each quantum gate as the training data. The transmon frequencies represent the hypothesis. The supervised learning algorithm then trains the qubit frequencies on the truth table data to generate procedures for designing high-fidelity three-qubit gates. Our approach to defining a gate-design problem as a supervised machine learning can inspire the application of other supervised machine learning methods, such as Support Vector Machine [89] and Neural Network [90] for designing quantum gates which acts on more than three qubits.

We have already introduced the quantum control scheme, named Subspace-Selective Self-Adaptive Differential Evolution (SuSSADE), in designing high-fidelity Toffoli gate [8]. Here, we employ the SuSSADE algorithm to generate procedures for other three-qubit gates, such as Fredkin and CXX gates. The two three-qubit gates considered here operate as fast as the two-qubit entangling CZ gate under the same experimental constraints. The robust performance of these gates against decoherence induced as well as random errors signifies the efficacy and robustness of the SuSSADE scheme for optimizing promising superconducting architectures. The

system considered here comprises three nearest-neighbor-coupled transmons that can serve as a module for any 1D or 2D architecture, and, in fact, our three-qubit gates can be realized in such multi-qubit systems, if the undesired couplings are switched off. Our results here establish the efficacy of SuSSADE as a machine learning approach to designing high-fidelity quantum operations.

[1] D. Gottesman, Phys. Rev. A **57**, 127 (1998), URL http://link.aps.org/doi/10.1103/PhysRevA.57.127.
[2] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, New J. Phys. **18**, 012002 (2016), URL http://stacks.iop.org/1367-2630/18/i=1/a=012002.
[3] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Phys. Rev. Lett. **103**, 110501 (2009), URL http://link.aps.org/doi/10.1103/PhysRevLett.103.110501.
[4] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, et al., Nature **508**, 500 (2014), ISSN 0028-0836, URL http://dx.doi.org/10.1038/nature13171.
[5] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, Phys. Rev. Lett. **93**, 130502 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.93.130502.
[6] 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), URL http://link.aps.org/doi/10.1103/PhysRevLett.81.2152.
[7] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature **482**, 382 (2012), ISSN 0028-0836, URL http://dx.doi.org/10.1038/nature10786.
[8] E. Zahedinejad, J. Ghosh, and B. C. Sanders, Phys. Rev. Lett. **114**, 200502 (2015), URL http://link.aps.org/doi/10.1103/PhysRevLett.114.200502.
[9] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **104**, 063603 (2010), URL http://link.aps.org/doi/10.1103/PhysRevLett.104.063603.
[10] 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), URL http://link.aps.org/doi/10.1103/PhysRevLett.102.040501.
[11] A. Fedorov, L. Steffen, M. Baur, M. P. da Silva, and A. Wallraff, Nature **481**, 170 (2012), ISSN 0028-0836, URL http://dx.doi.org/10.1038/nature10713.
[12] V. M. Stojanović, A. Fedorov, A. Wallraff, and C. Bruder, Phys. Rev. B **85**, 054504 (2012), URL http://link.aps.org/doi/10.1103/PhysRevB.85.054504.
[13] G. J. Milburn, Phys. Rev. Lett. **62**, 2124 (1989), URL http://link.aps.org/doi/10.1103/PhysRevLett.62.2124.
[14] J. Shamir, H. J. Caulfield, W. Micelli, and R. J. Seymour, Appl. Opt. **25**, 1604 (1986), URL http://ao.osa.org/abstract.cfm?URI=ao-25-10-1604.
[15] J. Ghosh, A. G. Fowler, and M. R. Geller, Phys. Rev. A **86**, 062318 (2012), URL http://link.aps.org/doi/10.1103/PhysRevA.86.062318.
[16] A. Peres, Phys. Rev. A **32**, 3266 (1985), URL http://link.aps.org/doi/10.1103/PhysRevA.32.3266.
[17] D. Aharonov, in *IN QUANT-PH/0301040* (2003).
[18] Y. Shi, Quantum Info. Comput. **3**, 84 (2003), ISSN 1533-7146, URL http://dl.acm.org/citation.cfm?id=2011508.2011515.

[19] T. Sleator and H. Weinfurter, Phys. Rev. Lett. **74**, 4087 (1995), URL http://link.aps.org/doi/10.1103/PhysRevLett.74.4087.

[20] Y. S. Weinstein, Quantum Inf. Process. **14**, 1841 (2015), ISSN 1570-0755, URL http://dx.doi.org/10.1007/s11128-015-0988-y.

[21] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996), URL http://link.aps.org/doi/10.1103/PhysRevLett.77.793.

[22] H. F. Chau and F. Wilczek, Phys. Rev. Lett. **75**, 748 (1995), URL http://link.aps.org/doi/10.1103/PhysRevLett.75.748.

[23] J. A. Smolin and D. P. DiVincenzo, Phys. Rev. A **53**, 2855 (1996), URL http://link.aps.org/doi/10.1103/PhysRevA.53.2855.

[24] V. V. Shende and I. L. Markov, Quantum Info. Comput. **9**, 461 (2009), ISSN 1533-7146, URL http://dl.acm.org/citation.cfm?id=2011791.2011799.

[25] M. Geller, E. Pritchett, A. Sornborger, and F. Wilhelm, in *Manipulating Quantum Coherence in Solid State Systems*, edited by M. Flatté and I. Ţifrea (Springer Netherlands, 2007), vol. 244 of *NATO Science Series II: Mathematics, Physics and Chemistry*, pp. 171–194, ISBN 978-1-4020-6135-6, URL http://dx.doi.org/10.1007/978-1-4020-6137-0_5.

[26] M. H. Devoret, A. Wallraff, and J. Martinis, arXiv preprint cond-mat/0411174 (2004).

[27] J. Clarke and F. K. Wilhelm, Nature **453**, 1031 (2008), URL http://dx.doi.org/10.1038/nature07128.

[28] C. T. Rigetti, *Quantum Gates for Superconducting Qubits* (BiblioBazaar, Charleston, South Carolina, 2011).

[29] J. Ghosh, A. Galiautdinov, Z. Zhou, A. N. Korotkov, J. M. Martinis, and M. R. Geller, Phys. Rev. A **87**, 022309 (2013), URL http://link.aps.org/doi/10.1103/PhysRevA.87.022309.

[30] C. Rigetti and M. Devoret, Phys. Rev. B **81**, 134507 (2010), URL http://link.aps.org/doi/10.1103/PhysRevB.81.134507.

[31] Y. Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, et al., Phys. Rev. Lett. **113**, 220502 (2014), URL http://link.aps.org/doi/10.1103/PhysRevLett.113.220502.

[32] M. Shapiro and P. Brumer, Principles of the Quantum Control of Molecular Processes, by Moshe Shapiro, Paul Brumer, pp. 250. ISBN 0-471-24184-9. Wiley-VCH, February 2003. **1** (2003).

[33] D. Dong and I. R. Petersen, IET, Contr. Theor. Ap. **4**, 2651 (2010).

[34] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1 (MIT press Cambridge, Cambridge, 1998).

[35] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Phys. Rev. A **84**, 022305 (2011), URL http://link.aps.org/doi/10.1103/PhysRevA.84.022305.

[36] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011), URL http://link.aps.org/doi/10.1103/PhysRevLett.106.190501.

[37] C. J. Bardeen, V. V. Yakovlev, K. R. Wilson, S. D. Carpenter, P. M. Weber, and W. S. Warren, Chem. Phys. Lett. **280**, 151 (1997).

[38] E. Zahedinejad, S. Schirmer, and B. C. Sanders, Phys. Rev. A **90**, 032310 (2014), URL http://link.aps.org/doi/10.1103/PhysRevA.90.032310.

[39] R. Storn and K. Price, J. Global Optim. **11**, 341 (1997).

[40] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **76**, 042319 (2007), URL http://link.aps.org/doi/10.1103/PhysRevA.76.042319.

[41] A. Houck, J. Koch, M. Devoret, S. Girvin, and R. Schoelkopf, Quantum Inf. Process. **8**, 105 (2009), ISSN 1570-0755, URL http://dx.doi.org/10.1007/s11128-009-0100-6.

[42] G. Dattoli, J. Gallardo, and A. Torre, J. Math. Phys. **27**, 772 (1986).

[43] A. Galiautdinov, A. N. Korotkov, and J. M. Martinis, Phys. Rev. A **85**, 042321 (2012), URL http://link.aps.org/doi/10.1103/PhysRevA.85.042321.

[44] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Phys. Rev. Lett. **91**, 167005 (2003), URL http://link.aps.org/doi/10.1103/PhysRevLett.91.167005.

[45] L. DiCarlo, J. Chow, J. Gambetta, L. S. Bishop, B. Johnson, D. Schuster, J. Majer, A. Blais, L. Frunzio, S. Girvin, et al., Nature **460**, 240 (2009).

[46] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R. S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, Nature **431**, 162 (2004), URL http://dx.doi.org/10.1038/nature02851.

[47] A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, and G. Gerber, Science **282**, 919 (1998).

[48] D. Meshulach and Y. Silberberg, Nature **396**, 239 (1998).

[49] C. A. Ryan, M. Laforest, J. C. Boileau, and R. Laflamme, Phys. Rev. A **72**, 062317 (2005).

[50] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson **172**, 296 (2005).

[51] R. Ruskov, K. Schwab, and A. N. Korotkov, Phys. Rev. B **71**, 235407 (2005).

[52] S. Mancini, D. Vitali, and P. Tombesi, arXiv preprint quant-ph/9802034 (1998).

[53] A. Hopkins, K. Jacobs, S. Habib, and K. Schwab, Phys. Rev. B **68**, 235328 (2003).

[54] P. Brumer and M. Shapiro, Annu. Rev. Phys. Chem. **43**, 257 (1992).

[55] D. J. Tannor and S. A. Rice, J. Chem. Phys. **83**, 5013 (1985).

[56] T. Schulte-Herbrüggen, A. Spörl, N. Khaneja, and S. J. Glaser, Phys. Rev. A **72**, 042331 (2005), URL http://link.aps.org/doi/10.1103/PhysRevA.72.042331.

[57] C. M. Bishop, *Pattern recognition and machine learning* (springer, Singapore, 2006).

[58] M. Guta and W. Kotłowski, New J. Phys. **12**, 123032 (2010), URL http://stacks.iop.org/1367-2630/12/i=12/a=123032.

[59] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, Phys. Rev. Lett. **114**, 200501 (2015), URL http://link.aps.org/doi/10.1103/PhysRevLett.114.200501.

[60] S. B. Kotsiantis, in *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies* (IOS Press, Amsterdam, The Netherlands, The Netherlands, 2007), pp. 3–24, ISBN 978-1-58603-780-2, URL http://dl.acm.org/citation.cfm?id=1566770.1566773.

[61] R. Caruana and A. Niculescu-Mizil, in *Proceedings of the 23rd International Conference on Machine Learning* (ACM, New York, NY, USA, 2006), ICML '06, pp. 161–168, ISBN 1-59593-383-2, URL `http://doi.acm.org/10.1145/1143844.1143865`.

[62] L. Banchi, N. Pancotti, and S. Bose, arXiv preprint arXiv:1509.04298 (2015).

[63] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 4 of *Johns Hopkins Studies in the Mathematical Sciences (Book 3)* (JHU Press, Baltimore, 2012).

[64] S. G. Schirmer and P. de Fouquieres, New J. Phys. **13**, 073029 (2011).

[65] H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal, Science **303**, 1998 (2004), http://www.sciencemag.org/content/303/5666/1998.full.pdf, URL `http://www.sciencemag.org/content/303/5666/1998.abstract`.

[66] C. G. Broyden, IMA J. Appl. Math. **6**, 76 (1970).

[67] R. Fletcher, Comput. J. **13**, 317 (1970).

[68] R. Fletcher, *Practical Methods of Optimization*, vol. 2 (Wiley, Padstow, 2013).

[69] D. Goldfarb, Math. Comput. **24**, 23 (1970).

[70] D. F. Shanno, Math. Comput. **24**, 647 (1970).

[71] D. M. Olsson and L. S. Nelson, Technometrics **17**, 45 (1975).

[72] M. Clerc and J. Kennedy, Evol. Comput., IEEE Trans. on **6**, 58 (2002).

[73] J. Vesterstrom and R. Thomsen, in *Congress on Evol. Comput., 2004. CEC2004.* (IEEE, 2004), vol. 2, pp. 1980–1987.

[74] M. Omran, A. Engelbrecht, and A. Salman, in *The 2005 IEEE Congress on Evolutionary Computation, 2005.* (2005), vol. 2, pp. 966–973 Vol. 2.

[75] J. Brest, V. Zumer, and M. S. Maucec, in *IEEE C. Evol. Computat., 2006. CEC 2006* (IEEE, 2006), pp. 215–222.

[76] M. A. Potter and K. A. D. Jong, in *Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: Parallel Problem Solving from Nature* (Springer-Verlag, London, UK, UK, 1994), PPSN III, pp. 249–257, ISBN 3-540-58484-6, URL `http://dl.acm.org/citation.cfm?id=645822.670374`.

[77] Y. Liu, X. Yao, Q. Zhao, and T. Higuchi, in *Proceedings of the 2001 Congress on Evolutionary Computation, 2001.* (2001), vol. 2, pp. 1101–1108 vol. 2.

[78] Y.-j. Shi, H.-f. Teng, and Z.-q. Li, in *Advances in Natural Computation*, edited by L. Wang, K. Chen, and Y. Ong (Springer Berlin Heidelberg, 2005), vol. 3611 of *Lecture Notes in Computer Science*, pp. 1080–1088, ISBN 978-3-540-28325-6, URL `http://dx.doi.org/10.1007/11539117_147`.

[79] D. Sofge, K. De Jong, and A. Schultz, in *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on* (2002), vol. 1, pp. 413–418.

[80] Z. Yang, K. Tang, and X. Yao, in *IEEE C. Evol. Computat., 2007. CEC 2007* (IEEE, 2007), pp. 3523–3530.

[81] Y.-x. Liu, S. K. Özdemir, A. Miranowicz, and N. Imoto, Phys. Rev. A **70**, 042308 (2004), URL `http://link.aps.org/doi/10.1103/PhysRevA.70.042308`.

[82] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[83] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I. C. Hoi, E. Jeffrey, et al., Phys. Rev. Lett. **112**, 240504 (2014), URL `http://link.aps.org/doi/10.1103/PhysRevLett.112.240504`.

[84] 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), ISSN 1745-2473, URL `http://dx.doi.org/10.1038/nphys1150`.

[85] E. Fredkin and T. Toffoli, Int. J. Theor. Phys. **21**, 219 (1982), ISSN 0020-7748, URL `http://dx.doi.org/10.1007/BF01857727`.

[86] Private communication with Pedram Roushan.

[87] H. Rabitz, R. d. Vivie-Riedle, M. Motzkus, and K. Kompa, Science **288**, 824 (2000), ISSN 00368075, URL `http://www.jstor.org/stable/3075194`.

[88] R. S. Judson and H. Rabitz, Phys. Rev. Lett. **68**, 1500 (1992), URL `http://link.aps.org/doi/10.1103/PhysRevLett.68.1500`.

[89] S. Haykin, *Neural Networks: A Comprehensive Foundation* (Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998), 2nd ed., ISBN 0132733501.

[90] C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995), ISSN 0885-6125, URL `http://dx.doi.org/10.1023/A:1022627411411`.