arXiv:2312.09218v1 [quant-ph] 14 Dec 2023

Speed limits of two-qubit gates with qudits

Bora Basyildiz,[1,*] Casey Jameson,[2] and Zhexuan Gong[1,2,3,†]

[1]*Quantum Engineering Program, Colorado School of Mines, Golden, Colorado 80401, USA*
[2]*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*
[3]*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

The speed of elementary quantum gates ultimately sets the limit on the speed at which quantum circuits can operate. For a fixed physical interaction strength between two qubits, the speed of any two-qubit gate is limited even with arbitrarily fast single-qubit gates. In this work, we explore the possibilities of speeding up two-qubit gates beyond such a limit by expanding our computational space outside the qubit subspace, which is experimentally relevant for qubits encoded in multi-level atoms or anharmonic oscillators. We identify an optimal theoretical bound for the speed limit of a two-qubit gate achieved using two qudits with a bounded interaction strength and arbitrarily fast single-qudit gates. In addition, we find an experimentally feasible protocol using two parametrically coupled superconducting transmons that achieves this theoretical speed limit in a non-trivial way. We also consider practical scenarios with limited single-qudit drive strengths and off-resonant transitions. For such scenarios, we develop an open-source, machine learning assisted, quantum optimal control algorithm that can achieve a speedup close to the theoretical limit with near-perfect gate fidelity. This work opens up a new avenue to speed up two-qubit gates when the physical interaction strength between qubits cannot be easily increased while extra states outside the qubit subspace can be well controlled.

I. INTRODUCTION

As noisy intermediate-scale quantum computers [1, 2] have finite coherence times and imperfect gate fidelities, designing quantum gates with faster speed and higher fidelity has been a major focus in quantum computing [3–5]. Tremendous effort from both academia and industry has been made on improving gate fidelities. This improvement, when coupled with the recent development of low-cost error correction codes [6–8], means that we may no longer be far from fault-tolerant quantum computers. Therefore, increasing the speed of a quantum computer would become a more important challenge in the long term. The most direct way to address this challenge is to achieve faster quantum gates, especially two-qubit gates since they are usually the speed bottleneck for most hardware platforms [9–11]. However, the speed of two-qubit gates is fundamentally limited by the interaction strength between the physical qubits [12]. An upper limit of such interaction strength usually exists for a given hardware design, posing a strong constraint on the speed of quantum computers.

One possible way to alleviate this constraint is to speed up two-qubit gates using extra states present in physical systems that carry the qubits. For example, if qubits are encoded in atomic states, there often exist low-energy states other than the qubit states that can be well controlled and have long coherence times [13–15]. Alternatively, for qubits encoded in anharmonic oscillators such as superconducting qubits, one can often control and measure higher energy states above the qubit states [11, 16–18]. As a result, we can expand the computational space from qubits to qudits (where d stands for the dimension of the local Hilbert space). This allows us to go beyond the usual speed limit for two-qubit gates [12] without increasing the physical interaction strength. While we can also perform universal quantum computing using qudits as the native information carriers [13, 19–23], this approach would require significant effort in developing compatible quantum algorithms [24–26], efficient quantum circuit compilations [27–29], and quantum error correction codes [30–32]. Here we only intend to use qudits effectively as qubits for quantum computing, but with a potential speed advantage over the use of strictly qubits. More explicitly, we require the state of the system to be within the qubit subspace before and after each quantum gate, such that states beyond the qubit subspace can only be populated in the middle of the gate implementation. In fact, many existing quantum gate designs already involve states outside the qubit subspace in a similar way [33–35]. A simple example is a single-qubit gate achieved by a Raman transition. Gates designed this way do not require end users to consider states outside the qubit subspace, which greatly simplifies the design of quantum circuits.

The idea of using extra states beyond the qubit subspace to speed up two-qubit gates has been recently explored in a system made of two coupled weakly anharmonic superconducting qubits [36]. Here we ask a more general question: What is the fundamental speed limit of a two-qubit gate using two qudits with a bounded interaction strength? For $d = 2$, the answer to this question is known for any given two-qubit gate and the form of interaction, provided that arbitrarily fast single-qubit gates are available [12]. For $d > 2$, no similar answer exists to our best knowledge. Ref. [36] studies such a speed limit for a specific two-qudit Hamiltonian. But since the study relies on numerical optimization, no well-established speed limit was found. In this work, we provide more in-depth answers to the above question by first deriving a rigorous theoretical bound on the speed of certain two-qubit gates using a general two-qudit Hamiltonian. We show with explicit protocols that this bound cannot be generically improved and is thus optimal.

Next, we argue that for physical Hamiltonians, the speedup of two-qubit gates due to extra states in qudits comes from two distinct sources: (1) larger coupling strengths between higher energy states of each qudit and (2) constructive interference

* bbasyildiz@mines.edu
† gong@mines.edu

among multiple states beyond the qubit subspace. For example, the speedup observed in Ref. [36] is largely due to (1), and we provide an $O(d)$ upper bound of such speedup. Here we construct explicit protocols that can achieve speedup from either or both of the two sources. Importantly, we show that for $d = 3$, these protocols can be readily implemented using two parametrically coupled superconducting transmon qubits [37]. The use of just one additional excited state for each transmon qubit can already lead to a two-qubit iSWAP gate 3 times its maximum speed with only two qubits. We further show that this gate protocol is time optimal as no faster iSWAP gate can be achieved based on the same interaction Hamiltonian.

In practice, it is hard to achieve the theoretical speed limit mentioned above as it requires arbitrarily fast single-qudit gates and no errors due to off-resonant drives or couplings. To address this concern, we develop an open-source quantum optimal control algorithm combining the GRAPE algorithm [38] with techniques in machine learning to optimize the fidelity of a target two-qubit gate with two coupled qudits. We optimize the pulse shapes of the drives on both qudits and identify the shortest gate time for a gate fidelity above a desired threshold. As an example, we focus on designing a speed-optimized iSWAP gate using two coupled qutrits as discussed above. Our optimal control algorithm generates gates with $>99.99\%$ fidelity at a speed close to the theoretical speed limit, assuming the maximum Rabi frequency of the drives is about an order of magnitude larger than the interaction strength. If off-resonant transitions are taken into account, the gate fidelity will be lower (around 99%) for the gate close to the theoretical speed limit, but we can slow down the gate by a small amount to achieve $>99.9\%$ gate fidelity.

It is worth mentioning that the open-source repository we developed for our optimal control algorithm can be useful for a wide range of problems in quantum gate designs. Apart from its high efficiency due to the use of state-of-art machine learning techniques, it can handle a general time-dependent two-qudit Hamiltonian and multiple drives at different frequencies, including the off-resonant transitions they may induce. Our repository thus goes beyond several existing counterparts for GRAPE-based algorithms [39–43].

The paper is organized as follows. In Section II, we study the theoretical speed limit of two-qubit gates achieved with two coupled qudits. We discuss explicit protocols that can saturate the theoretical speed limit and explain the two different sources of speedup for physical Hamiltonians. In Section III, we introduce our quantum optimal control algorithm and apply it to find speed-optimized two-qubit gates using two parametrically coupled transmon qutrits. Section IV concludes the paper with an outlook.

II. THEORETICAL SPEED LIMIT

We first derive a theoretical speed limit for a given two-qubit gate assuming a bounded interaction strength between two qudits and arbitrarily fast single-qudit gates. The Hamiltonian we consider is written in the following general form:

$$H(t) = H_1(t) + H_2(t) + H_c(t) \tag{1}$$

where $H_1(t)$ and $H_2(t)$ are the Hamiltonians that act only on the qudit 1 and qudit 2 respectively. $H_c(t)$ denotes the interaction between the two qudits. We upper bound the interaction strength via $\|H_c(t)\| \leq J$, where $\|\cdot\|$ denotes the operator norm (largest singular value of the operator) and J is a constant that only depends on the dimension d of each qudit. We do not limit $\|H_{1,2}(t)\|$, which allows for arbitrarily fast single-qudit gates.

To find the speed limit for the above Hamiltonian $H(t)$ to generate a given two-qubit gate, we first recall results in the study of quantum speed limits [44] that bound the minimum time $t_\perp$ to evolve an arbitrary quantum state to an orthogonal state. For example, the Mandelstam-Tamm (MT) inequality [45] shows that for any time-independent Hamiltonian H,

$$t_\perp \geq \frac{\pi}{2\Delta E} \tag{2}$$

where $\Delta E = \sqrt{\langle H^2\rangle - \langle H\rangle^2}$ is the energy uncertainty of the initial state. We assume $\hbar = 1$ throughout this paper.

The MT inequality can be generalized to accommodate a time-dependent Hamiltonian $H(t)$ and an arbitrary overlap between an initial state $|\psi(0)\rangle$ and a final state $|\psi(T)\rangle$ [45], which reads:

$$\int_0^T \Delta E(t)dt \leq \arccos|\langle\psi(0)|\psi(T)\rangle| \tag{3}$$

where $\Delta E(t)$ is the uncertainty of $H(t)$ in the state $|\psi(t)\rangle$.

We cannot directly apply the MT inequality or its above generalization to bound the speed of a two-qubit gate. This is because we have allowed for arbitrarily fast single-qudit gates. Thus, it takes zero time to evolve from any state to an orthogonal state. To address this issue, we need to separate the single-qudit Hamiltonians $H_{1,2}(t)$ from the interaction Hamiltonian $H_c(t)$ in Eq. (1). This can be done by introducing the following interaction picture Hamiltonian for $H_c(t)$:

$$H_I(t) = U_{H_1}^\dagger(t)U_{H_2}^\dagger(t)H_c(t)U_{H_1}(t)U_{H_2}(t) \tag{4}$$

where we define the time evolution operator $U_G(t) \equiv \mathcal{T}e^{-i\int_0^t G(t')dt'}$ for a general time-dependent Hamiltonian $G(t)$. Since $U_G(t)$ is a unitary operator, it's easy to see $\|H_I(t)\| = \|H_c(t)\|$.

Our goal is to implement a given two-qubit target gate $\mathcal{U}$ in time T in the two-qubit subspace of the two qudits. This means we require $U_H(T) = \begin{pmatrix}\mathcal{U} & 0\\ 0 & \tilde{\mathcal{U}}\end{pmatrix}$, where $\tilde{\mathcal{U}}$ is an arbitrary unitary outside the two-qubit subspace. Such block-diagonal $U_H(T)$ ensures that as long as the initial state of the system belongs to the two-qubit subspace, the final state does so too.

We can use the generalized MT inequality in Eq. (3) to lower bound the time T to implement the target gate if we can show that the evolution operator $U_{H_I}(T)$ for H_I is capable of evolving some initial state to an orthogonal state. This is however complicated by the fact that $U_{H_1}(T)$ and $U_{H_2}(T)$ can be arbitrary due to the lack of constraints on $H_{1,2}(t)$. Fortunately, we can use the fact that $U_{H_1}(T)$ and $U_{H_2}(T)$ are only arbitrary single-qudit unitaries to move forward. For concreteness, we use the iSWAP gate as our target gate $\mathcal{U}$ below as an example.

Let us choose an initial state of the form $|\psi_0\rangle = (c_0|0\rangle + c_1|1\rangle)\otimes|0\rangle$ and denote $|\psi_T\rangle = U_{H_I}(T)|\psi_0\rangle$. Using the identity $U_{H_I}(T) = U_{H_1}^\dagger(T)U_{H_2}^\dagger(T)U_H(T)$, we find that

$$\begin{aligned}|\langle\psi_0|\psi_T\rangle| &= \left|\langle\psi_0|U_{H_1}^\dagger(T)U_{H_2}^\dagger(T)U_H(T)|\psi_0\rangle\right| \\ &\le \left|c_0\langle 0|U_{H_1}^\dagger(T)|0\rangle + c_1\langle 0|U_{H_1}^\dagger(T)|1\rangle\right|. \end{aligned} \quad (5)$$

where we used that fact that $U_H(T)$ is an iSWAP gate in the two-qubit subspace. For any $U_{H_1}(T)$, we can then always choose appropriate values of c_0 and c_1 to make $\langle\psi_0|\psi_T\rangle = 0$. In other words, there exists at least one initial state such that no matter how we choose $H_1(t)$, evolving under $H_I(t)$ for time T results in an orthogonal state, provided that the evolution under $H(t)$ generates an iSWAP gate in the two-qubit subspace.

We then bound the energy uncertainty $\Delta E(t)$ for the Hamiltonian $H_I(t)$ at any time t in any state $|\psi\rangle$ using

$$\Delta E(t) \le \sqrt{\langle H_I(t)^2\rangle} \le \sqrt{\|H_I(t)^2\|} = \|H_I(t)\| \le J \quad (6)$$

where we use the fact that $\|H_I(t)^2\|$ denotes the largest eigenvalue of $H_I(t)^2$, which cannot be smaller than $\langle H_I(t)^2\rangle$. Combining Eq. (6) with Eq. (3) above, we arrive at

$$t_{\text{iSWAP}} \ge \frac{\pi}{2J}. \quad (7)$$

Therefore, we obtain a theoretical speed limit for implementing a two-qubit iSWAP gate with a generic two-qudit Hamiltonian in Eq. (1) satisfying $\|H_c(t)\| \le J$.

A similar but possibly more complicated analysis can be applied to obtain the speed limit for other two-qubit gates. For the iSWAP gate, we can further show that the speed limit given by Eq. (7) is optimal. This can be seen first for the $d = 2$ case, where we can generate the iSWAP gate directly using $H = J(\sigma_1^+\sigma_2^- + \text{h.c.})$ that satisfies $\|H\| = J$. At time $T = \pi/(2J)$, we can easily check that $U_H(T) = e^{-iHT} = U_{\text{iSWAP}}$, exactly reaching the speed limit predicted by Eq. (7).

From Eq. (7), we also see that if $J \equiv \max\|H_c(t)\|$ is a constant independent of d, then no speedup can be obtained if we go from two qubits to two qudits. Fortunately, for many physical Hamiltonians, when we expand the computational space from qubits to qudits, the value of J will increase naturally without physically increasing the interaction strength (such as moving the two qudits spatially closer). For example, the interaction between two capacitively coupled superconducting transmons (which are anharmonic oscillators) can be described by

$$H_c = g(a_1 + a_1^\dagger)(a_2 + a_2^\dagger) \quad (8)$$

where g characterizes the physical interaction strength and a_i is the annihilation operator for the i^{th} anharmonic oscillator. We can truncate the Hilbert space of each anharmonic oscillator to d dimensions, making a system of two qudits. It is not hard to show that $\|H_c\| = gO(d)$. Intuitively, this is because the creation/annihilation operators have matrix elements of roughly $\sqrt{d}$ in the subspace formed by the two highest Fock states $|d-1\rangle$ and $|d-2\rangle$. Then Eq. (7) predicts a potential $O(d)$ speedup for the implementation of a two-qubit gate, consistent with the numerical findings in Ref. [36].

Here we construct an explicit protocol based on the Hamiltonian in Eq. (8) to speedup a two-qubit iSWAP gate by a factor of $d-1$. We can replace the constant coupling strength g in Eq. (8) by a parametric coupling $g\cos(\omega t)$ [37], where ω is set to match the energy difference between the $|d-1,d-2\rangle$ and $|d-2,d-1\rangle$ energy eigenstates. If we assume the anharmonicity and frequency difference of the two transmons are sufficiently large compared to g, we induce an effective Hamiltonian of the form

$$H_{\text{eff}} = g(d-1)(|d-1,d-2\rangle\langle d-2,d-1| + \text{h.c.}). \quad (9)$$

If we further assume that single-qudit gates are arbitrarily fast, we can swap the information stored in the states $|d-2\rangle$ and $|d-1\rangle$ to the computational basis states $|0\rangle$ and $|1\rangle$ in no time, thus achieving an iSWAP gate in time $T = \pi/[2g(d-1)]$.

A very different way to obtain the same amount of speedup is to couple many states in the qudit space collectively to some states in the two-qubit subspace. For example, we can consider the following effective coupling Hamiltonian

$$H'_{\text{eff}} = g(|0\rangle + |2\rangle + \cdots + |d-1\rangle)^{\otimes 2}\langle 11| + \text{h.c.} \quad (10)$$

Simply speaking, we couple the state $|11\rangle$ to $(d-1)^2$ orthogonal Fock states with a uniform coupling strength g. If we start the two qudits in the state $|00\rangle$ and perform a single-qudit gate that takes $|0\rangle$ to $\frac{1}{\sqrt{d-1}}(|0\rangle + |2\rangle + \cdots + |d-1\rangle)$ for each qudit before and after the evolution of H'_{eff} for time $T = \pi/[2g(d-1)]$, we end up in the $|11\rangle$ state. A more careful analysis reveals that this process implements an iSWAP gate if we further flip the second qubit before and after the process.

Both protocols above saturate the theoretical speed limit given by Eq. (7) in different ways and implement an iSWAP gate $O(d)$ faster than when the evolution is constrained to within the qubit subspace. Moreover, we can even combine the speedup obtained in the above two protocols in an experimentally realizable Hamiltonian for $d = 3$. By parametrically driving the coupling g in Eq. (8) with four different frequencies, corresponding to the energy differences between $|00\rangle$ and $|11\rangle$, $|02\rangle$ and $|11\rangle$, $|20\rangle$ and $|11\rangle$, as well as $|22\rangle$ and $|11\rangle$, we end up with the following effective coupling Hamiltonian in the rotating wave approximation:

$$H''_{\text{eff}} = g(|00\rangle + \sqrt{2}\,|02\rangle + \sqrt{2}\,|20\rangle + 2\,|22\rangle)\langle 11| + \text{h.c.} \quad (11)$$

It's not hard to show that an iSWAP gate can be achieved using the following gate sequence after projecting onto the two-qubit subspace:

$$U_{\text{iSWAP}} = X_2U_1^\dagger U_2^\dagger e^{-iH''_{\text{eff}}\frac{\pi}{6g}}U_2U_1X_2 \quad (12)$$

where U_1 and U_2 are rotations around the y axis by an angle $2\arctan(\sqrt{2})$ for the effective qubit formed by the $|0\rangle$ and $|2\rangle$ states for qutrit 1 and qutrit 2 respectively, and X_2 flips the states $|0\rangle$ and $|1\rangle$ for the second qutrit. If we again assume arbitrarily fast single-qudit gates, we can now achieve

an iSWAP gate in time $T = \pi/(6g)$, three times the speed for the same Hamiltonian restricted to the qubit subspace, where the speed limit is $\pi/(2g)$. We emphasize that this gate protocol also saturates the theoretical speed limit in Eq. (7) due to $\|H''_{\text{eff}}\| = 3g$. Interestingly, we note that for $d = 3$ the previous two protocols based on H_{eff} in Eq. (9) and H'_{eff} in Eq. (10) both give an iSWAP gate time $T = \pi/(4g)$ (assuming negligible single-qudit gate times). The protocol using H''_{eff} is faster as it uses both stronger coupling strengths for higher excited states and a collective (but not uniform) coupling between states in the qudit subspace and a state in the qubit subspace.

III. OPTIMAL CONTROL RESULTS

In practice the single-qudit gates, such as the X_2, U_1, U_2 in Eq. (12), cannot be arbitrarily fast. Driving the transitions in a qudit too strongly will off-resonantly drive other transitions in the qudit. It may also drive transitions outside the qudit subspace and transitions in the other qudit if individual addressing is not achieved by spatial separation, such as in most superconducting qubit platforms. Admittedly, these problems occur even if we are only driving the qubit transitions in usual gate designs, but the fact that we have to drive transitions outside the qubit subspace resonantly and strongly makes it much harder to control the errors caused by these off-resonant transitions. While we can control such errors using sufficiently weak drives, the long time we spend on single-qudit gates can easily negate any speedup we obtained for the gate protocols in Section II.

To overcome this challenge, we develop a quantum optimal control method based on the GRAPE algorithm [38] to find optimal single-qudit Hamiltonians [$H_{1,2}(t)$ in Eq. (1)] that lead to a two-qubit gate in a coupled two-qudit system close to the theoretical limit studied in Section II with limited single-qudit drive strengths and the presence of off-resonant transitions. For simplicity, throughout the following discussion we assume the coupling Hamiltonian $H_c(t)$ in Eq. (1) is time-independent, but we can also optimize $H_c(t)$ if one can engineer the interaction Hamiltonian temporally.

The goal of our optimal control algorithm is to maximize the average fidelity F between the evolution operator $U_H(T) = \mathcal{T} e^{-i\int_0^T [H_1(t)+H_2(t)+H_c]dt}$ projected to the two-qubit subspace (denoted by $\mathcal{U}_H$ below) and the target two-qubit gate $\mathcal{U}$, defined by [46]

$$F = \frac{1}{5} + \frac{1}{80}\sum_j \text{tr}\left(\mathcal{U}U_j^\dagger \mathcal{U}^\dagger \mathcal{U}_H U_j \mathcal{U}_H^\dagger\right) \tag{13}$$

where $U_j \in \{I, X, Y, Z\} \otimes \{I, X, Y, Z\}$ is one of the 16 two-qubit Pauli operators. To be specific, we focus on finding the fastest possible two-qubit iSWAP gate with a sufficiently high fidelity F using two coupled qutrits with the interaction Hamiltonian given by Eq. (11). As we introduced in Section II, this Hamiltonian can be experimentally realized using parametrically coupled superconducting transmons [37]. To access the full space of two qutrits, we will drive resonantly the $|0\rangle - |1\rangle$ and the $|1\rangle - |2\rangle$ transitions for each qutrit and optimize the pulse envelopes of each drive as our control.

We first consider the practical scenario where single-qudit drives have finite strengths but off-resonant transitions can be ignored. This can be a good approximation if the anharmonicity and frequency difference of the two transmon qubits are much larger than the drive strengths and the interaction strength g. In this case, the drive Hamiltonians take the form

$$\begin{aligned} H_1(t) &= (\Omega_{1,1}(t)|0\rangle\langle 1| + \Omega_{1,2}(t)|1\rangle\langle 2| + \text{h.c.}) \otimes I \\ H_2(t) &= I \otimes (\Omega_{2,1}(t)|0\rangle\langle 1| + \Omega_{2,2}(t)|1\rangle\langle 2| + \text{h.c.}) \end{aligned} \tag{14}$$

Here $\Omega_{i,k}(t)$ $(i, k = 1, 2)$ represents the complex Rabi frequency of the drive that is resonant with the transition $|k-1\rangle - |k\rangle$ in the qutrit i. Note that here we do not include the diagonal Hamiltonian that defines the energies of the basis states $|0\rangle, |1\rangle, |2\rangle$ for each qutrit, as we are in the rotating frame of such a Hamiltonian.

To perform numerical optimization efficiently, we need to discretize the continuous functions $\Omega_{i,k}(t)$ above. As in the standard GRAPE algorithm, we divide $\Omega_{i,k}(t)$ from $t = 0$ to $t = T$ into M segments each with a duration $\tau = T/M$. Within each segment, we assume $\Omega_{i,k}(t)$ is a slowly varying function that ensures $\Omega_{i,k}(t)$ is continuous across different segments. An example is:

$$\Omega_{i,k}(t) = A_{i,k}^{(m)} \sin^2[\pi(m - 1 - t/\tau)],\ t \in [(m-1)\tau, m\tau] \tag{15}$$

with $m = 1, 2, \cdots, M$. An illustration of $\Omega_{i,k}(t)$ in this form is shown in Fig. 1. We set $M = 40$ in our calculations, which can be reached experimentally using commercial arbitrary waveform generators assuming a two-qubit gate time on the order of 10-100ns, typical for transmon qubits. Further increasing M does not lead to any noticeable improvement in our results. We then optimize the $4M$ complex parameters given by $\{A_{i,k}^{(m)}\}$ via a gradient descent method with random initial values. We also run the optimization with multiple random initializations and choose the highest fidelity obtained from all runs.

To address the practical concern that single-qudit drives are of limited strengths, we apply a constraint $|A_{i,k}^{(m)}| \le \Omega_{\text{max}}$ for any i, k, m in our optimization, implying the Ω_{max} is the maximum Rabi frequency of each drive. In Fig. 2, we show the minimum gate time we obtained for a gate fidelity $F > 99.99\%$ for different values of Ω_{max} (measured in units of the interaction strength g). We choose this fidelity threshold as it is higher than any experimental two-qubit gate achieved up to date. From Fig. 2, one finds that if $\Omega_{\text{max}}/g \gtrsim 10$, a near-perfect iSWAP gate can be obtained for $T = 0.4T_{\text{min}}$, which is close to the theoretical speed limit $T = T_{\text{min}}/3$. Here $T_{\text{min}} = \pi/(2g)$ is the minimum iSWAP gate time for the coupling Hamiltonian H_c in Eq. (11) restricted to the two-qubit subspace (i.e. $g(|01\rangle\langle 10| + \text{h.c.})$. Even for Ω_{max} as small as $3g$, we can still obtain a faster gate than the two-qubit case.

The speedup we obtained from our optimal control algorithm for a finite single-qudit drive strength is in fact nontrivial. If we instead follow the exact theoretical protocol in Eq. (12), we need to perform 4 single-qudit gates in serial

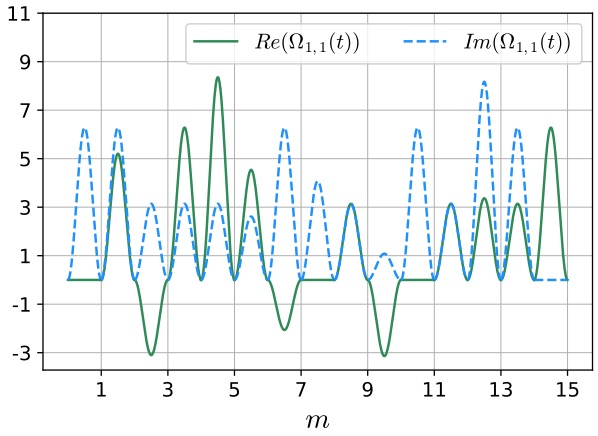

Figure 1. An example of the pulse shape $\Omega_{1,1}(t)$ (in units of g) in Eq. (15) for the drive resonant with the $|0\rangle$-$|1\rangle$ transition of the first qudit. This pulse shape is obtained from our optimal control algorithm. For a better presentation, only the first 15 segments of the full 40-segment pulse are shown here.

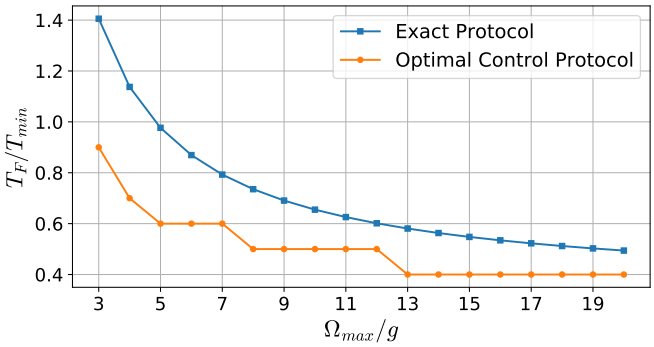

Figure 2. The minimum gate times T_F for $F \geq 99.99\%$ achieved by the exact protocol in Eq. (12) and the optimal control protocol without considering off-resonant transitions as a function of the maximum Rabi frequency Ω_{max} for the single-qudit drives.

and also spend a time $T_{\text{min}}/3$ for evolving H_c. Moreover, the single-qudit gates $U_{1,2}$ in Eq. (12) require us to drive the transition $|0\rangle - |2\rangle$, which is a two-photon transition whose dipole moment is typically much smaller than that for the $|0\rangle - |1\rangle$ or the $|1\rangle - |2\rangle$ transitions in a transmon. Yet even if we assume that the maximum Rabi frequency of the $|0\rangle - |2\rangle$ drive is also Ω_{max}, we still need to spend a minimum total gate time of

$$T_{\text{exact}} = \frac{\pi + 2\arctan\left(\sqrt{2}\right)}{\Omega_{\text{max}}} + \frac{\pi}{6g} \tag{16}$$

to obtain an exact iSWAP gate in the two-qubit subspace following Eq. (12). As shown in Fig. 2, we see that T_{exact} is significantly larger than the gate time we obtained using the optimal control algorithm, despite that both gate times approach the theoretical limit $\pi/(6g)$ in the $\Omega_{\text{max}} \to \infty$ limit. The speed advantage brought by the optimal control method is likely due to the fact that single-qudit drives are applied simultaneously with the two-qubit interaction [47], while for the exact protocol the drives and the interaction are applied in serial.

The optimal control algorithm can also give us much higher gate fidelity if we slow down the gate a bit. As shown in Fig. 3, we can achieve a gate infidelity below 10^{-13} for a gate time $T = 0.45T_{\text{min}}$, which is still more than twice the speed of an iSWAP gate using only two qubits.

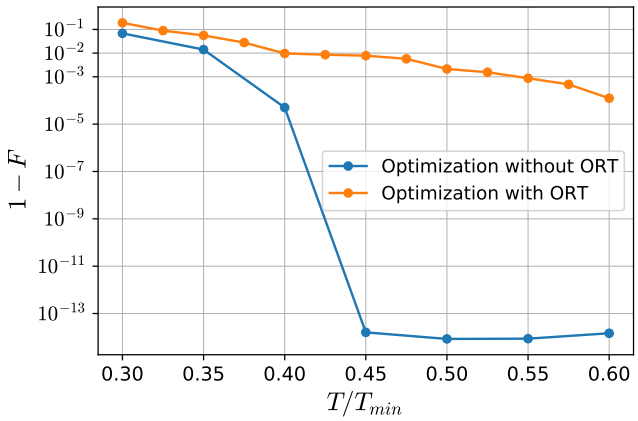

Figure 3. The infidelity $1 - F$ of the iSWAP gate achieved by our optimal control algorithm with and without considering off-resonant transitions (ORT) as a function of the gate time T using two coupled qutrits. The theoretical speed limit is given by $T = T_{\text{min}}/3$, which can only be achieved with infinitely fast single-qudit gates. Here we assume a maximum single-qudit Rabi frequency $\Omega_{\text{max}} = 20g$ ($\Omega_{\text{max}} = 40g$) for the optimization without (with) ORT.

The results shown so far assume that the four drives in Eq. (14) do not drive any other transitions off-resonantly. This assumption may not be valid if the two transmons are already strongly coupled, which is preferred for a fast two-qubit gate. For example, for a coupling strength g in Eq. (8) on the order of 10MHz, we need to drive each transmon with a maximum Rabi frequency on the order of 100MHz in order to approach the theoretical speed limit as indicated by Fig. 2. This is not much smaller than the anharmonicity of the transmon or the frequency difference between two transmon qubits (both typically a few hundred MHz [37, 47]). Therefore, our next step is perform the optimization with the off-resonant transitions induced by the drives modelled by the Hamiltonians $H_{1,2}(t)$.

Since the qubit frequency of most transmons is much larger than their anharmonicity, we can still safely ignore any off-resonant two-photon transitions. For two qutrits, there are four single-photon transitions: the $|0\rangle - |1\rangle$ and $|1\rangle - |2\rangle$ transitions for each qutrit. Each of the four drives in Eq. (14) is resonant with one of the four transitions, but can drive the remaining three transitions off-resonantly. In addition, we should also consider the fact that each of the four drives may also drive the transition $|2\rangle - |3\rangle$ off-resonantly for each transmon and excite the system outside the Hilbert space of two qutrits. To address this issue, we will expand our Hilbert space to $d = 4$ to explicitly take into account the $|3\rangle$ state for each transmon. We then add a penalty proportional to the maximum and the average population outside the two-qutrit subspace to the loss function, i.e. the infidelity $1 - F$. As we discuss in the Appendix, this penalty ensures that the leakage of the

state outside the two-qutrit subspace causes a negligible (less than 0.1%) drop of the gate fidelity. Therefore, we can safely ignore off-resonant transitions involving even higher energy states, such as $|4\rangle$, $|5\rangle$, etc.

Note that we can now set $\Omega_{\max}/g$ to a larger value (40 from now on) as we have modelled all off-resonant transitions (5 for each drive tone). We set the anharmonicity to be $10g$ and the frequency difference between the two transmon qubits to be $15g$, which are within the experimental reach for g as large as 30MHz [37, 47]. As shown in Fig. 3, our optimal control algorithm managed to find pulse shapes that lead to $>99\%$ gate fidelity at $T = 0.4T_{\min}$ and $>99.9\%$ gate fidelity at $T = 0.55T_{\min}$. While both the gate fidelity and gate speed are lower than the case without considering off-resonant transitions, we still obtain a notable speedup with a sufficiently high gate fidelity. For $g = 30$MHz (in frequency not angular frequency), this means we can achieve an iSWAP gate in time $T \approx 4.6$ns with $>99.9\%$ fidelity, which is much faster than any two-qubit gate experimentally achieved with transmon qubits. In addition, we expect even higher fidelity and faster gate if the anharmonicity and the frequency difference between the two qubits are bigger.

Before we close this section, we point out a number of technical achievements we made in our optimal control algorithm. Our algorithm improves upon existing GRAPE based algorithms for quantum gate designs such as QuTiP [39], GRAPE-Tensorflow [40], and Juqbox [42]. First, we use the state-of-art machine learning library Pytorch to perform auto-differentiation of the loss function over the parameters in the drive Hamiltonians. This auto-differentiation technique was also used in Ref. [40]. Inspired by the state-of-art machine learning practice, we further employ a stochastic gradient descent algorithm with the Nesterov momentum method [48] to improve the efficiency of the optimization process. In a recent work [47], we have shown that this machine learning inspired approach outperforms the standard GRAPE implementation (such as the one in the QuTiP software [39] which uses the L-BFGS-B method) significantly when the parameter space is large and a high fidelity (e.g. $>99.9\%$) is needed. Second, we adopt a symplectic second-order Runge-Kutta (SRK2) method [49] in finding the evolution operator $U_H(T)$ for the time-dependent Hamiltonian $H(t)$. A symplectic ordinary differential equation (ODE) solver conserves the unitarity of the calculated evolution operator, which allows us to use a larger step size to achieve a similar numerical accuracy, thus making the optimization more efficient. A similar symplectic method was used in Refs. [42, 43] also for GRAPE based algorithms.

We have made an open-source repository [50] that implements our optimal control algorithm in Python. This repository may benefit a wide range of researchers working on optimal quantum gate designs not only due to its high efficiency, but also due to its general framework that allows for multiple drives within or outside the two-qubit subspace and accounts for many off-resonant transitions. We have not seen such capabilities in any existing open-source repositories for optimal quantum gate designs [39–43].

IV. CONCLUSION AND OUTLOOK

In this work, we have shown two primary results. First, we derived a theoretical speed limit for certain two-qubit gates implemented with two coupled qudits under the assumption of arbitrarily fast single-qudit gates. This speed limit is optimal and it can be exactly achieved with physical interactions among two qudits, providing an $O(d)$ speedup in the two-qubit gate time. In addition, we find that this speedup can come from naturally stronger coupling between higher energy states of the qudits, collective couplings between states in and outside the qubit subspace, or a combination of both. Also, we have constructed explicit protocols to demonstrate the speedup due to each of the above three cases.

Second, we addressed the concern that single-qudit gates cannot be arbitrarily fast in practice by developing a GRAPE based quantum optimal control algorithm and applying it to obtain an iSWAP gate in the two-qubit subspace close to the aforementioned theoretical speed limit for two coupled qutrits. We develop an open-source repository for our algorithm that implements state-of-art numerical optimization techniques, allows for multiple drive frequencies, and takes into account off-resonant transitions that cannot be ignored. We expect this repository to be widely useful for researchers working on optimal quantum gate designs, even if no states outside the qubit subspace is used intentionally.

An alternative way to speedup a two-qubit gate is to use one or more ancilla qubits, which is similar to the approach here in that both are using extra states to go beyond the two-qubit speed limit. However, it remains unclear where this approach can actually lead to a speedup. Finding out the speed limit of a two-qubit gate in a many-qubit system is also fundamental interesting as it relates to the study of Lieb-Robinson bounds and emergent locality in quantum many-body systems [51].

A natural follow-up to this work is to experimentally demonstrate the speedup we obtained for the iSWAP gate using two parametrically coupled transmons. In addition, we are also interested in designing specific protocols for other hardware platforms such as trapped ions to accelerate two-qubit gates there by using extra low-energy states that can be well controlled. Admittedly, these protocols will bring an extra level of complication in designing high-fidelity gates. But we believe such complication can be tackled with a combination of improvements in hardware and software, such as better individual addressing, more precise controls, longer coherence times, and more powerful optimal control methods. Our work thus creates a new pathway in further speeding up two-qubit gates when conventional approaches such as increasing the interaction strength between qubits have been exhausted.

APPENDIX: LEAKAGE ERRORS

In our optimal control algorithm discussed in Section II, we have modelled the two-qutrit system with four states $(|0\rangle, |1\rangle, |2\rangle, |3\rangle)$ to account for errors induced by possible leakage of the state outside the two-qutrit subspace for two transmons. We further add a penalty to our loss function in

the optimization to suppress the population outside the two-qutrit subspace. To see how much leakage we still have, we calculate the probability P_k ($k = 2, 3, \cdots$) for at least one transmon to be in the state $|k\rangle$ as well as the probability $P_{0,1}$ for both transmons to be in the two-qubit subspace. We start the system in one of the four computational basis states ($|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$) and plot in Fig. 4 the probabilities $P_{0,1}$, P_2, P_3, and P_4 each averaged over the four initial states.

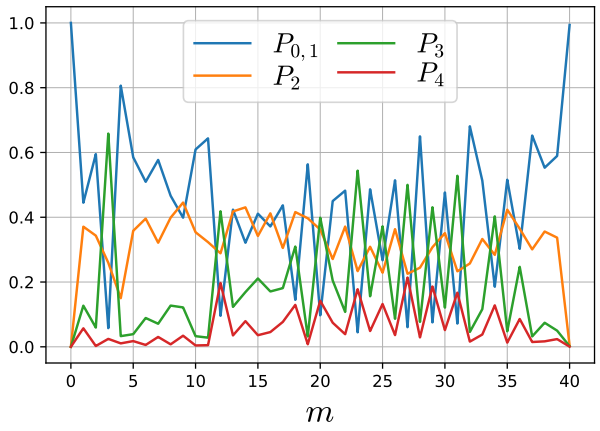

Figure 4. Probabilities P_k for at least one transmon to be in the state $|k\rangle$ during the gate protocol obtained from the $T = 0.5T_{\text{min}}$ point of the optimization with ORT curve in Fig. 3. $P_{0,1}$ denotes the probability in the two-qubit subspace.

As we expected, over the application of $M = 40$ pulse segments, $P_{0,1}$ starts at exactly 1 and ends at 0.9978, showing that we have achieved a high-fidelity two-qubit gate in the subspace of two qudits with little leakage outside the qubit subspace. Due to the penalty we add to minimize excitations to states outside the two-qutrit subspace, we see that the population in the subspace involving state $|k\rangle$, denoted by P_k decreases as we increase k. While the maximum value of P_k is not negligible even for $k = 4$, we note that the value of P_k gets very close to zero at the end of the gate implementation.

Fig. 5 shows that the leakage to higher excited states during the gate implementation indeed has negligible effects on the gate fidelity obtained. As an example, we take the pulse shapes from the optimal control algorithm for a gate time $T = 0.5T_{\text{min}}$ in the optimization with off-resonant transitions (see Section II and Fig. 3). We then expand our Hilbert space to consider all basis states from $|0\rangle$ to $|n_{\text{max}}\rangle$ for each transmon, and calculate the gate fidelity with off-resonant transitions to all higher excited states up to $|n_{\text{max}}\rangle$ taken into account. We see that the gate fidelity only decreases by less than 0.1% as we increase n_{max} from 3 (used in the optimization algorithm) to 9. In fact, the gate fidelities reported in Fig. 3 for the optimization with ORT are already the fidelities calculated using $n_{\text{max}} = 9$ (although the optimization is always done with $n_{\text{max}} = 3$ for efficiency reasons). Thus we have properly taken into account the leakage errors.

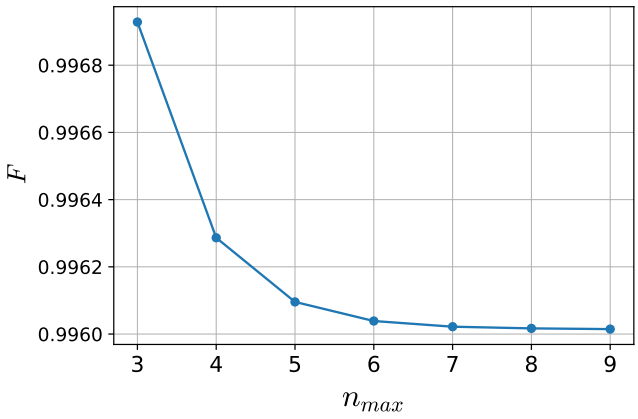

Figure 5. Gate fidelity F to the ideal iSWAP gate when considering leakage to higher energy states due to off-resonant transitions. We consider all energy eigenstates of each transmon from $|0\rangle$ to $|n_{\text{max}}\rangle$. The particular gate calculated here is from the $T = 0.5T_{\text{min}}$ point of the optimization with ORT curve in Fig. 3

ACKNOWLEDGMENTS

We thank Dr. Raymond Simmonds, Tongyu Zhao, and Dr. Sahel Ashhab for enlightening discussions and the HPC center at the Colorado School of Mines for providing computational resources. We acknowledge funding support from the NSF NRT program under Grant. No. 2125899, the NSF RAISE program under Grant No. 1839232, and the W. M. Keck Foundation.

[1] J. Preskill, Quantum **2**, 79 (2018).
[2] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Nature **618**, 500 (2023).
[3] L. Ding, M. Hays, Y. Sung, B. Kannan, J. An, A. D. Paolo, A. H. Karamlou, T. M. Hazard, K. Azar, D. K. Kim, B. M. Niedzielski, A. Melville, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, J. A. Grover, K. Serniak, and W. D. Oliver, arXiv:2304.06087.
[4] T. Noh *et al.*, Nature Physics **19**, 1445 (2023).
[5] R. Srinivas, S. C. Burd, H. M. Knaack, R. T. Sutherland, A. Kwiatkowski, S. Glancy, E. Knill, D. J. Wineland, D. Leibfried, A. C. Wilson, D. T. C. Allcock, and D. H. Slichter, Nature **597**, 209 (2021).
[6] N. P. Breuckmann and J. N. Eberhardt, PRX Quantum **2**, 040101 (2021).
[7] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, arXiv:2308.07915.
[8] Q. Xu, J. P. B. Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasic, M. D. Lukin, L. Jiang, and H. Zhou, arXiv:2308.08648.

[9] V. M. Schäfer, C. J. Ballance, K. Thirumalai, L. J. Stephenson, T. G. Ballance, A. M. Steane, and D. M. Lucas, Nature **555**, 75 (2018).

[10] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **622**, 268 (2023).

[11] I. N. Moskalenko, I. A. Simakov, N. N. Abramov, A. A. Grigorev, D. O. Moskalev, A. A. Pishchimova, N. S. Smirnov, E. V. Zikiy, I. A. Rodionov, and I. S. Besedin, npj Quantum Information **8**, 130 (2022).

[12] G. Vidal, K. Hammerer, and J. I. Cirac, Phys. Rev. Lett. **88**, 237902 (2002).

[13] P. J. Low, B. M. White, A. A. Cox, M. L. Day, and C. Senko, Phys. Rev. Res. **2**, 033128 (2020).

[14] P. J. Low, B. White, and C. Senko, arXiv:2306.03340.

[15] P. Hrmo *et al.*, Nature Communications **14**, 2242 (2023).

[16] J. Long, T. Zhao, M. Bal, R. Zhao, G. S. Barron, H. sheng Ku, J. A. Howard, X. Wu, C. R. H. McRae, X.-H. Deng, G. J. Ribeill, M. Singh, T. A. Ohki, E. Barnes, S. E. Economou, and D. P. Pappas, arXiv:2103.12305.

[17] L. E. Fischer, A. Chiesa, F. Tacchino, D. J. Egger, S. Carretta, and I. Tavernelli, PRX Quantum **4**, 030327 (2023).

[18] S. Cao, M. Bakr, G. Campanaro, S. D. Fasciati, J. Wills, D. Lall, B. Shteynas, V. Chidambaram, I. Rungger, and P. Leek, "Emulating two qubits with a four-level transmon qudit for variational quantum algorithms," (2023), arXiv:2303.04796 [quant-ph].

[19] Y. Wang, Z. Hu, B. C. Sanders, and S. Kais, Frontiers in Physics **8**, 589504 (2020).

[20] M. Luo and X. Wang, Science China Physics, Mechanics & Astronomy **57**, 1712 (2014).

[21] Y. Chi, J. Huang, Z. Zhang, J. Mao, Z. Zhou, X. Chen, C. Zhai, J. Bao, T. Dai, H. Yuan, M. Zhang, D. Dai, B. Tang, Y. Yang, Z. Li, Y. Ding, L. K. Oxenløwe, M. G. Thompson, J. L. O'Brien, Y. Li, Q. Gong, and J. Wang, Nature Communications **13**, 1166 (2022).

[22] L. Yeh, arXiv:2304.12504.

[23] P. Roy, J. van de Wetering, and L. Yeh, Electronic Proceedings in Theoretical Computer Science **384**, 142–170 (2023).

[24] P. W. Shor, SIAM Journal on Computing **26**, 1484 (1997).

[25] L. K. Grover, (1996), arXiv:quant-ph/9605043.

[26] A. W. Harrow, A. Hassidim, and S. Lloyd, Physical Review Letters **103** (2009).

[27] R. Wille, S. Hillmich, and L. Burgholzer, in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (2020) pp. 1–5.

[28] Z. Shan, Y. Zhu, and B. Zhao, Scientific Reports **12**, 7132 (2022).

[29] N. Quetschlich, L. Burgholzer, and R. Wille, arXiv:2212.04508.

[30] A. Steane, Proceedings: Mathematical, Physical and Engineering Sciences **452**, 2551 (1996).

[31] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).

[32] J. Roffe, Contemporary Physics **60**, 226–245 (2019).

[33] K. N. Nesterov, I. V. Pechenezhskiy, C. Wang, V. E. Manucharyan, and M. G. Vavilov, Phys. Rev. A **98**, 030301 (2018).

[34] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature **460**, 240 (2009).

[35] C. J. Wood and J. M. Gambetta, Phys. Rev. A **97**, 032306 (2018).

[36] S. Ashhab, F. Yoshihara, T. Fuse, N. Yamamoto, A. Lupascu, and K. Semba, Phys. Rev. A **105**, 042614 (2022).

[37] X. Y. Jin, K. Cicak, Z. Parrott, S. Kotler, F. Lecocq, J. Teufel, J. Aumentado, E. Kapit, and R. W. Simmonds, arXiv:2305.02907.

[38] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296–305 (2005).

[39] J. Johansson, P. Nation, and F. Nori, Computer Physics Communications **183**, 1760 (2012).

[40] N. Leung, M. Abdelhafez, J. Koch, and D. Schuster, Phys. Rev. A **95**, 042318 (2017).

[41] T. Propson, B. E. Jackson, J. Koch, Z. Manchester, and D. I. Schuster, Phys. Rev. Appl. **17**, 014036 (2022).

[42] N. A. Petersson, F. M. Garcia, A. E. Copeland, Y. L. Rydin, and J. L. DuBois, arXiv:2001.01013.

[43] N. Anders Petersson and F. Garcia, SIAM Journal on Scientific Computing **44**, A3592–A3616 (2022).

[44] M. R. Frey, Quantum Information Processing **15**, 3919 (2016).

[45] L. Mandelstam and I. Tamm, "The uncertainty relation between energy and time in non-relativistic quantum mechanics," in *Selected Papers*, edited by B. M. Bolotovskii, V. Y. Frenkel, and R. Peierls (Springer Berlin Heidelberg, 1991) pp. 115–123.

[46] M. A. Nielsen, Physics Letters A **303**, 249–252 (2002).

[47] J. Howard, A. Lidiak, C. Jameson, B. Basyildiz, K. Clark, T. Zhao, M. Bal, J. Long, D. P. Pappas, M. Singh, and Z. Gong, Phys. Rev. Res. **5**, 043194 (2023).

[48] Y. Nesterov, Proceedings of the USSR Academy of Sciences **269**, 543 (1983).

[49] K. Feng and M. Qin, "Symplectic Runge-Kutta methods," (Springer Berlin Heidelberg, 2010) pp. 277–364.

[50] https://github.com/b-basyildiz/QuOpt.

[51] C.-F. Chen, A. Lucas, and C. Yin, Reports on Progress in Physics **86**, 116001 (2023).