Global entangling gates for trapped ions: Numerical design and gate-efficient compilation for Hamiltonian simulation

Zhe-Yu Shen,[1] Lin Cheng,[2] Sheng-Chen Liu,[1] Jun-Yi Deng,[1] and Liang-You Peng[1,3,*]

[1]*State Key Laboratory for Mesoscopic Physics and Frontiers Science Center for Nano-optoelectronics, School of Physics, Peking University, Beijing 100871, China*

[2]*Future Research Lab, China Mobile Research Institute, Beijing 100053, China*

[3]*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China*

(Received 9 April 2026; revised 22 August 2026; accepted 14 October 2026; published 26 October 2026)

As quantum computing scales up, circuits relying on sequential two-qubit gates face severe challenges from excessive depth and decoherence. In trapped-ion systems, global entangling gates offer a highly parallel alternative but are hindered by complex numerical optimization. We propose an efficient numerical framework for piecewise-constant amplitude-modulated global gates. Based on the Mølmer-Sørensen protocol, we utilize null-space projection for linear constraints of spin-motion coupling and tangent-space gradient descent to minimize laser power, while achieving high-order robustness against frequency drifts. At the application level, we identify a structural correspondence between hardware-native global interactions and Pauli strings in Hamiltonian simulation and propose an efficient compilation scheme. By reducing common physical models to fundamental XX interactions and formulating interaction graph partitioning as a Set Cover problem, we leverage global gates to significantly optimize the compilation process. Explicit pulse-level benchmarks further show that, for the tested weighted interaction graphs, the reduction in entangling-gate count can translate into a substantial reduction in total entangling duration. These results demonstrate that integrating hardware-native operations with tailored compilation strategies can substantially reduce circuit overhead, providing a practical roadmap for the codesign of near-term quantum simulators.

DOI: 10.1103/rfgl-6q15

I. INTRODUCTION

In recent years, the field of quantum computing has made significant progress, demonstrating extraordinary computational potential beyond classical computers in areas such as molecular simulation, combinatorial optimization [1], and machine learning [2]. Thanks to unique physical properties like long coherence times [3,4] and inherent all-to-all connectivity, trapped-ion systems are widely recognized as one of the most promising platforms for realizing universal quantum computation [5–10].

As the fundamental building blocks of quantum computation, the performance of quantum logic gates directly determines the computational power of the system. Traditionally, trapped-ion quantum circuits rely heavily on sequential two-qubit entangling gates. To achieve high-fidelity and robust

*Contact author: liangyou.peng@pku.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

two-qubit gates, a mainstream approach builds upon the Mølmer-Sørensen (MS) protocol [11] by applying amplitude-modulated (AM) [12,13], frequency-modulated [14,15], or phase-modulated [16,17] laser pulses. Furthermore, recent advancements in two-dimensional (2D) trapped-ion arrays [18,19] and novel scalable architectures [20,21] have laid a solid foundation for scaling up the number of qubits.

However, as the problem size scales up, relying solely on sequential two-qubit gates leads to excessively deep quantum circuits, severely limiting performance due to system decoherence times. In this context, global entangling gates [22,23]—which operate in a single-instruction multiple-data fashion by coupling multiple qubits simultaneously—have emerged as a highly attractive alternative. They not only significantly enhance operational parallelism, compress circuit depth, and reduce total gate time [24], but also fully exploit the architectural advantage of the all-to-all connectivity in trapped-ion systems.

Despite their immense theoretical advantages, the physical implementation of global gates poses severe numerical challenges. Current design schemes typically model global pulses as complex nonlinear optimization problems, searching for optimal solutions in a high-dimensional parameter space. However, due to a massive number of nonlinear constraints, the objective landscape is notoriously ill-behaved, plagued by numerous local optima. This results in enormous computational overhead, making it difficult to scale to medium- or large-scale systems [22,24,25].

Recently, theoretical progress has been made to overcome this scalability barrier. For instance, a recently proposed scheme [26] employs a frequency-domain multi-tone representation and maps the pulse-design problem to an algebraic structure of linear and quadratic constraints, enabling fast construction for large ion crystals. Time-domain and frequency-domain descriptions are mathematically related, and their practical merits depend on the available control hardware and on implementation-specific bandwidth, sampling, and crosstalk constraints [27]. Here, we focus on ion-resolved, piecewise-constant AM pulses, which directly correspond to segmented envelope control and have been used in trapped-ion experiments [23]. Our numerical framework is tailored to this representation. Using the null-space projection of the linear motional constraints, we reduce the remaining quadratic phase equations to a lower-dimensional subspace. We then employ tangent-space gradient descent to reduce laser power and incorporate additional constraints to enhance robustness against motional-frequency drifts.

Beyond the physical implementation layer, effectively utilizing global gates in high-level quantum algorithms is currently a heavily researched topic. Leveraging global interactions can effectively reduce the total number of entangling gates in algorithms such as the quantum Fourier transform [28], constant-cost implementations of Clifford operations and multiply controlled gates [29], and quantum error correction codes like the toric code [30]. Moreover, recent theoretical works have demonstrated the algorithmic advantages of multiqubit gates in executing constant-depth Boolean functions [31] and general quantum circuits [32]. Recently, researchers have also developed universal compilation algorithms using ZX-calculus to automatically map arbitrary quantum circuits into sequences containing global gates [33,34]. While these universal algorithms have made strides in reducing gate counts, there remains significant room for improving optimization constants for domain-specific problems.

Quantum simulation is widely regarded as one of the most direct application scenarios for quantum computing [35], with numerous successful experiments already demonstrated [36–38]. In the current noisy intermediate-scale quantum (NISQ) era [39], the overhead of quantum simulation cannot be judged solely by asymptotic gate complexity; constant factors and specific implementation details often determine the viability of an algorithm [40]. We observe a natural structural similarity between hardware-native global interactions in trapped ions and Pauli strings in Hamiltonian simulation. For this specific task, the trapped-ion system ceases to be merely a general-purpose computing platform and acts more as a "domain-specific accelerator." Another core contribution of this work is the proposal of a gate-efficient compilation scheme tailored for Hamiltonian simulation circuits based on global gates. By directly mapping hardware-native global operations, we drastically reduce the number of pulses required to simulate specific physical models.

The remainder of this paper is organized as follows: In Sec. II, we introduce the theoretical framework and the numerical optimization algorithm for piecewise-constant AM global gates. Section III presents gate designs synthesized via the proposed algorithm, followed by a detailed evaluation of its power optimization performance. We further benchmark the numerical scheme's scalability, power efficiency, and robustness to validate its practical efficacy. In Sec. IV, we propose a compilation algorithm based on the Set Cover problem, demonstrate its workflow and performance through examples, and generalize it to broader scenarios. Finally, we conclude and discuss future applications in Sec. V.

II. EFFICIENT NUMERICAL SCHEME FOR GLOBAL ENTANGLEMENT GATE

In this section, we present a comprehensive numerical framework for implementing global multiqubit entanglement gates. We begin by establishing the theoretical foundation based on the MS interaction within an ion chain. For the AM scheme, we reformulate the pulse design as a combined problem of linear algebra and quadratic equations and search numerically for solutions satisfying the prescribed constraints. Building on this, we introduce a power optimization scheme that employs gradient descent on the solution manifold to minimize laser intensity. Furthermore, we extend the framework to incorporate robustness against motional-frequency drifts. Finally, we present the fidelity formula for n-qubit gates to quantitatively evaluate the impact of residual entanglement and phase deviations on gate performance.

A. Theoretical framework

In this work, we consider a chain of N ions confined in a linear trap, with the ions aligned along the z axis. We aim to implement an n-qubit global entanglement gate acting on a subset of ions $S = \{j_1, j_2, \ldots, j_n\}$. By addressing the ions in S with laser pulses, let $\Omega_j(t)$ denote the Rabi frequency for ion j, $\phi_j(t)$ the laser phase, and μ the two-photon detuning relative to the internal level transition. The resulting unitary evolution operator at the gate time τ is expressed as [22]

$$U(\tau) = \exp\left(\sum_{j\in S}\sum_{m}(\alpha_{j,m}(\tau)a_m^\dagger - \alpha_{j,m}^*(\tau)a_m)\sigma_{j,\vec{n}} + i\sum_{j,j'\in S, j<j'}\Theta_{j,j'}(\tau)\sigma_{j,\vec{n}}\sigma_{j',\vec{n}'}\right), \tag{1}$$

where $\sigma_{j,\vec{n}}$ represents the Pauli operator for ion j along the direction $\vec{n}$. For simplicity and without loss of generality, we set this to the x direction, i.e., $\sigma_{j,\vec{n}} = X_j$.

The summation over m encompasses all normal modes of the ion chain. Here, assuming the laser propagates along the x axis, we restrict the summation to the N radial x modes. This leverages their higher frequencies and lower heating rates to mitigate off-resonant excitations during entangling operations. Let ω_m be the frequency of the mth mode, b_j^m the corresponding normalized eigenvector, and $a_m^\dagger, a_m$ the creation and annihilation operators, respectively. We introduce the Lamb-Dicke parameter $\eta_{j,m} = \Delta k b_j^m \sqrt{\frac{\hbar}{2M_{\text{ion}}\omega_m}} \ll 1$, where M_{ion} is the mass of a single ion and Δk is the wave-vector difference of the two laser beams along the x direction.

In the above expression, $\alpha_{j,m}(\tau)$ characterizes the residual entanglement between the internal and motional states for ion j and mode m:

$$\alpha_{j,m}(\tau) = -i\eta_{j,m}\int_0^\tau \Omega_j(t)\sin(\mu t + \phi_j(t))e^{i\omega_m t}dt, \tag{2}$$

and $\Theta_{j,j'}(\tau)$ represents the coupling strength, or cumulative phase, between ions j and j':

$$\begin{aligned}\Theta_{j,j'}(\tau) = \sum_{m=1}^{N}\eta_{j,m}\eta_{j',m}\int_0^\tau dt_1\int_0^{t_1}dt_2\ \sin\left(\omega_m(t_1-t_2)\right)\\ \times\Big[\Omega_j(t_1)\Omega_{j'}(t_2)\sin\left(\mu t_1+\phi_j(t_1)\right)\sin\left(\mu t_2+\phi_{j'}(t_2)\right)\\ +\ \Omega_{j'}(t_1)\Omega_j(t_2)\sin\left(\mu t_1+\phi_{j'}(t_1)\right)\sin\left(\mu t_2+\phi_j(t_2)\right)\Big].\end{aligned} \tag{3}$$

Suppose the target n-qubit global gate we intend to implement is of the form

$$G(\theta) = \exp\left(i \sum_{j,j' \in S, j<j'} \theta_{j,j'} X_j X_{j'}\right), \tag{4}$$

where $\theta_{j,j'}$ is the desired distribution of cumulative phases. Note that while a uniform phase of $\theta_{j,j'} = \pi/4$ is a common choice, we consider a more general case here. To achieve this, the following constraints must be satisfied:

$$\begin{aligned} \alpha_{j,m}(\tau) &= 0, \quad \forall j \in S, m \in \{1, \ldots, N\}, \\ \Theta_{j,j'}(\tau) &= \theta_{j,j'}, \quad \forall j, j' \in S, j < j'. \end{aligned} \tag{5}$$

These constraints yield nN equations concerning the decoupling of internal and motional states and $n(n-1)/2$ equations for the cumulative phases, forming the mathematical foundation for our subsequent numerical algorithm design.

B. Amplitude-modulated solutions

In this section, we employ an AM scheme, where the laser phases are set to $\phi_j(t) \equiv 0$ for all $j \in S$. It is important to note that if the laser amplitude sequences applied to each ion were identical, the integral terms corresponding to each mode in the expression for $\Theta_{j,j'}(\tau)$ would become independent of the ion indices j and j'. In such a case, only the Lamb-Dicke parameters in the coefficients would vary across different ion pairs. Consequently, we would only have N tunable integral values to satisfy $n(n-1)/2$ constraints. Given that $2 \leqslant n \leqslant N$, the available degrees of freedom would generally be insufficient.

To address this, we apply distinct amplitude sequences to each ion. We divide the total gate time τ into N_s equidistant segments, each with a duration of $\tau_s = \tau/N_s$. The amplitude sequence for each ion is then characterized by an N_s-dimensional vector:

$$\mathbf{\Omega}_j = \left[\Omega_{j,1}, \ldots, \Omega_{j,N_s}\right]^T, \quad j \in S, \tag{6}$$

where $\mathbf{\Omega}_j$ represents the amplitude sequence vector for ion j, and the pth component denotes the Rabi frequency during the time interval $t \in [(p-1)\tau_s, p\tau_s]$.

The constraint $\alpha_{j,m}(\tau) = 0$ consists of one complex equation, which can be decomposed into two real linear equations. By substituting the discretized amplitude sequences, the decoupling constraints for all $\alpha_{j,m}$ can be expressed as

$$\mathbf{M}_j \mathbf{\Omega}_j = \mathbf{0}, \quad j \in S, \tag{7}$$

where $\mathbf{M}_j$ is a $2N \times N_s$ matrix. This yields a total of $2nN$ linear constraints (the explicit matrix elements are provided in Appendix A). Similarly, the constraints on the cumulative phases $\Theta_{j,j'}$ can be formulated as

$$\mathbf{\Omega}_j^T \mathbf{S}_{j,j'} \mathbf{\Omega}_{j'} = \theta_{j,j'}, \quad j, j' \in S, j < j', \tag{8}$$

where $\mathbf{S}_{j,j'}$ is an $N_s \times N_s$ matrix, representing a total of $n(n-1)/2$ quadratic constraints (see Appendix A for details).

To solve this system, we first address the linear equations in Eq. (7). The null space of each $\mathbf{M}_j$ can be efficiently determined using singular-value decomposition or other standard numerical methods. Let $W_j = \text{span}\{\mathbf{e}_{j,1}, \ldots, \mathbf{e}_{j,d_j}\}$ denote this null space, where $\{\mathbf{e}_{j,i}\}_i$ is an orthonormal basis and $d_j = N_s - \text{rank}(\mathbf{M}_j)$. Under the generic full-row-rank assumption, $d_j = N_s - 2N$. We can expand $\mathbf{\Omega}_j$ in terms of this basis:

$$\mathbf{\Omega}_j = \sum_{i=1}^{d_j} a_{j,i} \mathbf{e}_{j,i} = \mathbf{V}_j \mathbf{a}_j, \quad j \in S, \tag{9}$$

where $\mathbf{V}_j = [\mathbf{e}_{j,1}, \ldots, \mathbf{e}_{j,d_j}]$ is the basis transformation matrix and $\mathbf{a}_j$ is the coefficient vector. By defining $\tilde{\mathbf{S}}_{j,j'} = \mathbf{V}_j^T \mathbf{S}_{j,j'} \mathbf{V}_{j'}$, the quadratic constraints in Eq. (8) are transformed into

$$\mathbf{a}_j^T \tilde{\mathbf{S}}_{j,j'} \mathbf{a}_{j'} = \theta_{j,j'}, \quad j, j' \in S, j < j'. \tag{10}$$

We then solve these $n(n-1)/2$ quadratic equations numerically for the reduced coefficients $\{\mathbf{a}_j\}$ and reconstruct the amplitude sequences through Eq. (9). A simple degree-counting argument gives the following necessary segment-number condition under the full-row-rank assumption:

$$N_s > 2N + \frac{n-1}{2}. \tag{11}$$

This condition ensures that the number of reduced variables exceeds the number of quadratic constraints, but it is not by itself sufficient to guarantee the existence of a real solution or the convergence of a local solver. In practice, we use Eq. (11) as a design guideline and determine feasibility from the numerical convergence and constraint residuals. The reduced quadratic equations are solved using Newton's method. This projection satisfies the linear constraints $\alpha_{j,m}(\tau) = 0$ by construction and substantially reduces the dimension of the quadratic equation system solved by Newton's method.

C. Power optimization

In the preceding section, we presented a method for obtaining numerically feasible solutions for $\mathbf{\Omega}_j$. In experimental practice, higher laser amplitudes directly translate to greater power requirements. Due to technical limitations and energy efficiency considerations, it is desirable to minimize the total power consumption. Consequently, our next step is to perform power optimization on the solutions obtained in the previous stage.

The solutions to Eq. (10) constitute a d'-dimensional manifold, where $d' = nd - n(n-1)/2$. We perform the gradient descent within the tangent space at a given solution $\mathbf{a}_0$. The Jacobian matrix $\mathbf{J}$, with elements defined as $J_{(j,j'),(k,i)} = \frac{\partial(\mathbf{a}_j^T \tilde{\mathbf{S}}_{j,j'} \mathbf{a}_{j'})}{\partial a_{k,i}}$, is an $\frac{n(n-1)}{2} \times nd$ matrix. Let $W' = \text{span}\{\mathbf{f}_1, \ldots, \mathbf{f}_{d'}\}$ denote the null space of $\mathbf{J}$, where $\{\mathbf{f}_i\}_i$ is an orthonormal basis that spans the tangent space. By introducing the $nd \times d'$ basis transformation matrix $\mathbf{U} = [\mathbf{f}_1, \ldots, \mathbf{f}_{d'}]$, a vector within the tangent space can be expressed as

$$\mathbf{a} = \mathbf{a}_0 + \mathbf{U}\mathbf{b}, \tag{12}$$

where $\mathbf{a} = [\mathbf{a}_{j_1}^T, \ldots, \mathbf{a}_{j_n}^T]^T$ is the concatenated nd-dimensional coefficient vector and $\mathbf{b}$ is a d'-dimensional vector of coefficients for the tangent-space basis expansion.

To minimize the total power, we define the objective function as the sum of squared amplitudes across all segments: $P = \sum_{j \in S} \mathbf{\Omega}_j^T \mathbf{\Omega}_j = \sum_{j \in S} \mathbf{a}_j^T (\mathbf{V}_j^T \mathbf{V}_j) \mathbf{a}_j$. This is a quadratic form in $\mathbf{a}$, denoted as

$$P = \mathbf{a}^T \begin{bmatrix} \mathbf{V}_1^T \mathbf{V}_1 & 0 & \ldots & 0 \\ 0 & \mathbf{V}_2^T \mathbf{V}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{V}_n^T \mathbf{V}_n \end{bmatrix} \mathbf{a} = \mathbf{a}^T \mathbf{V} \mathbf{a}. \tag{13}$$

Within the tangent space, the value of P is given by $P = (\mathbf{a}_0 + \mathbf{U}\mathbf{b})^T \mathbf{V} (\mathbf{a}_0 + \mathbf{U}\mathbf{b})$. The gradient with respect to $\mathbf{b}$ at the origin (i.e., at $\mathbf{a}_0$) is

$$\nabla_{\mathbf{b}} P = 2\mathbf{U}^T \mathbf{V} \mathbf{a}_0. \tag{14}$$

By choosing an appropriate step size h, we can move a small distance in the tangent space as $\mathbf{b} = -h \nabla_{\mathbf{b}} P$, resulting in an updated vector $\mathbf{a} = \mathbf{a}_0 + \mathbf{U}\mathbf{b}$.

Since the updated $\mathbf{a}$ may no longer strictly satisfy the quadratic constraints in Eq. (10), we apply Newton's method using the new $\mathbf{a}$ as the initial guess to pull the solution back onto the manifold, yielding a refined solution $\mathbf{a}_1$. This process is repeated iteratively until the value of P converges, ensuring a significant reduction in power consumption while restoring the gate constraints to the specified numerical tolerance after each update.

D. Robustness to frequency shifts

Due to the inherent complexity of experimental trap configurations and instability of trapping potentials [41], discrepancies often exist between theoretically calculated normal-mode frequencies and their actual experimental values. Such offsets lead to a degradation in gate fidelity, making the robustness of the pulse sequences a critical consideration.

In this work, we incorporate robustness against shifts in the normal-mode frequencies ω_m. We enforce w-order robustness by requiring the first w derivatives of $\alpha_{j,m}(\tau)$ with respect to ω_m to vanish:

$$\frac{\partial^k \alpha_{j,m}(\tau)}{\partial \omega_m^k} = 0, \quad k \in \{0, 1, \ldots, w\}, \; j \in S. \tag{15}$$

These conditions can be reformulated as a set of linear constraints:

$$\begin{bmatrix} \mathbf{M}_j^0 \\ \mathbf{M}_j^1 \\ \vdots \\ \mathbf{M}_j^w \end{bmatrix} \mathbf{\Omega}_j = \mathbf{0}, \quad j \in S, \tag{16}$$

where each $\mathbf{M}_j^k$ is a $2N \times N_s$ matrix. This system imposes a total of $2nN(w+1)$ linear constraints (see Appendix A for the explicit matrix elements).

It is important to note that, unlike the case for two-qubit gates, we must also explicitly consider the robustness of $\Theta_{j,j'}(\tau)$. In two-qubit gate design, there is only a single Θ constraint. If a frequency shift occurs, the constraint can often be resatisfied by simply scaling the entire amplitude sequence by a constant factor, rather than redesigning the pulse. However, in n-qubit global-gate design, scaling the amplitude sequences provides only limited degrees of freedom, which is generally insufficient to satisfy the $n(n-1)/2$ cumulative phase constraints.

Consequently, we require the derivatives of the phase to vanish as well:

$$\frac{\partial^k \Theta_{j,j'}(\tau)}{\partial \omega_m^k} = 0, \quad j, j' \in S, \; j < j', \quad m \in \{1, \ldots, N\}, \quad k \in \{1, \ldots, w\}. \tag{17}$$

These conditions can be expressed as quadratic forms:

$$\mathbf{\Omega}_j^T \mathbf{S}_{j,j'}^{k,m} \mathbf{\Omega}_{j'} = 0, \quad j, j' \in S, \; j < j', \quad m \in \{1, \ldots, N\}, \quad k \in \{1, \ldots, w\}. \tag{18}$$

Together with the primary constraints [Eq. (8)], this yields a total of $n(n-1)(1+wN)/2$ quadratic constraints (the explicit matrix forms are detailed in Appendix A).

The zeroth-order motional and accumulated-phase constraints are evaluated from the full expressions in Eqs. (2) and (3). To keep the derivative constraints analytically tractable, the first- and higher-order robustness matrices are constructed under the rotating-wave approximation (RWA), as detailed in Appendix A. The final frequency-drift performance is nevertheless evaluated with the full non-RWA expressions.

The resulting system can be solved using the numerical framework established in Secs. II A, II B, and II C. As the number of constraints increases significantly with the robustness order w, degree

counting gives the necessary segment-number estimate

$$N_s > 2N(w+1) + \frac{(n-1)(1+wN)}{2}. \tag{19}$$

Similarly to Eq. (11), this inequality is used as a parameter-selection guideline rather than a sufficient feasibility condition. The remainder of the numerical procedure is identical to the previously described methodology and is omitted here for brevity.

E. Fidelity analysis

If the constraint equations in Eq. (5) are not strictly satisfied (for instance, due to a drift in normal-mode frequencies), the fidelity will deviate from unity. To quantitatively estimate the performance, we must analyze the relationship between the fidelity F and the deviations in $\alpha_{j,m}(\tau)$ and $\Theta_{j,j'}(\tau)$.

A formal derivation shows that in the limit of small $\alpha_{j,m}(\tau)$ and small offsets $\Delta\Theta_{j,j'} = \Theta_{j,j'}(\tau) - \theta_{j,j'}$, the average fidelity $\bar{F}$ is given by [13,42,43]

$$\bar{F} \approx 1 - \frac{2^n}{2^n+1}\left(\sum_{j\in S}\sum_{m=1}^{N}(2\bar{n}_m+1)|\alpha_{j,m}(\tau)|^2 + \sum_{j,j'\in S, j<j'}(\Theta_{j,j'}(\tau)-\theta_{j,j'})^2\right). \tag{20}$$

The detailed derivation is provided in Appendix B. Here, $\bar{n}_m$ denotes the average phonon number of mode m. When cooling techniques are employed to initialize the system into its vibrational ground state, we can assume $\bar{n}_m \to 0$, rendering the coefficient of the $|\alpha_{j,m}(\tau)|^2$ term as unity.

This functional form is highly intuitive: The leading-order contributions of both the residual entanglement error and the phase error are second-order terms. Moreover, these two error sources are completely decoupled, and the total infidelity is simply their arithmetic sum. Notably, this expression is independent of whether the phase matrix elements $\theta_{j,j'}$ are uniformly set to $\pi/4$. Therefore, from the perspective of fidelity, it is unnecessary to impose additional phase symmetry constraints during the pulse design.

By examining the scaling of these two terms, we find that the residual entanglement error accumulates linearly with the number of ions involved in the gate, i.e., $O(n)$. In contrast, the cumulative phase error involves a summation over all participating ion pairs, scaling as $O(n^2)$. Consequently, as the scale of the entangling gate expands, the precision of phase control becomes increasingly critical and challenging. This scaling behavior provides further justification for introducing the robustness constraints for $\Theta_{j,j'}(\tau)$ discussed in Sec. II D.

III. RESULTS OF NUMERICAL EXPERIMENTS

In this section, we present the numerical results demonstrating the practical application of the proposed algorithm. Taking an ion chain as a benchmark system, we first illustrate the construction and power optimization of pulse sequences for a global entanglement gate. We then evaluate the performance of these pulses under robustness constraints when subjected to normal-mode frequency drifts. Finally, we discuss the scalability and performance of our approach, focusing on both the computational efficiency and the required laser power.

A. Pulse sequences and power optimization

We consider the 1D ion chain described in Sec. II A. To achieve a more uniform ion distribution, we adopt an axial potential of the form [44,45]

$$U = \sum_{i=1}^{N}\left(-\frac{1}{2}\alpha_2 z_i^2 + \frac{1}{4}\alpha_4 z_i^4\right) + \sum_{i<j}\frac{e^2}{4\pi\epsilon_0|z_i - z_j|}, \tag{21}$$

while the radial potential is simply taken as $U' = \sum_{i=1}^{N}(\frac{1}{2}M_{\text{ion}}\omega_x^2 x_i^2 + \frac{1}{2}M_{\text{ion}}\omega_y^2 y_i^2)$.

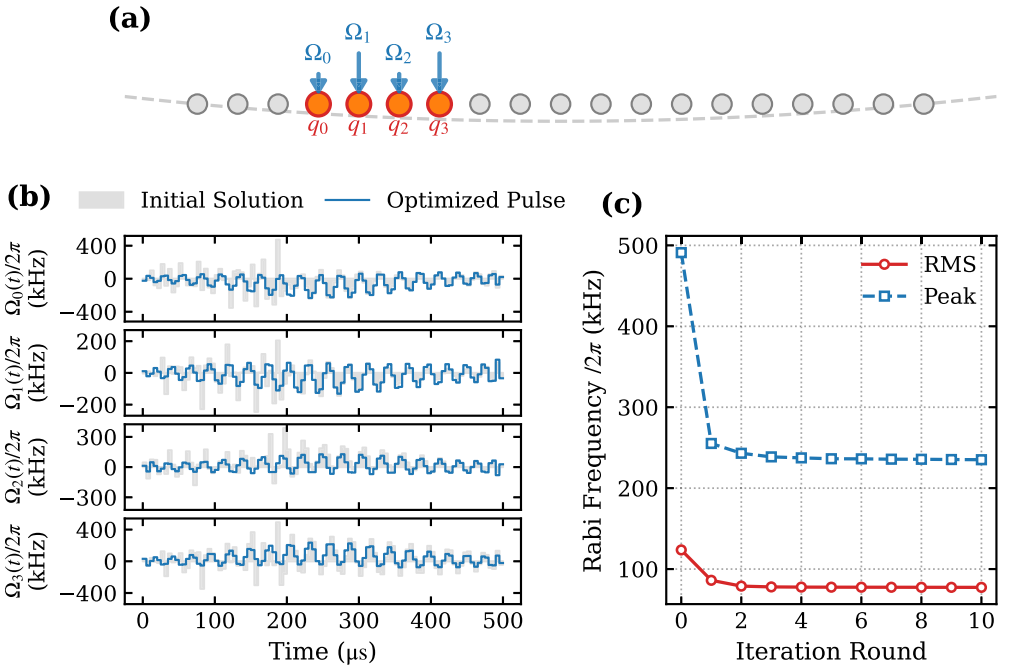


FIG. 1. A four-qubit AM global entanglement gate on a 19-ion chain. (a) Schematic of the ion chain, where an entanglement gate is applied to the fourth, fifth, sixth, and seventh ions (labeled as q_0 to q_3) using corresponding pulse sequences Ω_0 to Ω_3. (b) Pulse amplitude sequences before and after power optimization, with $N_s = 100$ segments, gate duration $\tau = 500\,\mu s$, and detuning $\mu = 1.002\omega_x$. (c) RMS and peak Rabi frequencies during ten rounds of tangent-space power optimization.

By appropriately nondimensionalizing the axial potential, we obtain

$$V = \sum_{i=1}^{N}\left(-\frac{1}{2}u_i^2 + \frac{1}{4}\gamma_4 u_i^4\right) + \sum_{i<j}\frac{1}{|u_i - u_j|}, \tag{22}$$

where $V = 4\pi\epsilon_0 l_0 U/e^2$ is the dimensionless potential, $u_i = z_i/l_0$ represents the dimensionless coordinates, and $l_0 = (e^2/4\pi\epsilon_0\alpha_2)^{1/3}$ is the characteristic length scale. The equilibrium configuration of the ions is then entirely determined by the dimensionless constant $\gamma_4 = \alpha_4 l_0^2/\alpha_2$. The normal modes can be determined by minimizing V.

Specifically, we consider a one-dimensional (1D) chain of $^{171}Yb^+$ ions with $N = 19$ as an example, as shown in Fig. 1(a). The characteristic length of the trap is set to $l_0 = 40\,\mu m$ with $\gamma_4 = 4.3$, and the radial trapping frequency (the center-of-mass mode) is taken as $\omega_x = 2\pi \times 3\,MHz$. Under these conditions, the ions are addressed by a laser with wavelength $\lambda = 355\,nm$. For this example, we set the laser detuning to $\mu = 1.002\omega_x$.

We employ the algorithm described in Secs. II B and II C to implement a four-qubit global entanglement gate on the fourth, fifth, sixth, and seventh ions in the chain, with a target phase of $\pi/4$ for all pairs, i.e., $\exp(i\sum_{j,j'\in\{4,5,6,7\},j<j'}\frac{\pi}{4}X_jX_{j'})$. We choose a total gate duration $\tau = 500\,\mu s$ and $N_s = 100$ pulse segments, which satisfies the necessary degree-counting condition in Eq. (11) for $N = 19$ and $n = 4$. Here, we initially omit the robustness constraints by setting the robustness order to $w = 0$.

The resulting pulse amplitude sequences are presented in Fig. 1(b), showing the waveforms before and after power optimization. As observed, the optimized waveforms exhibit a significant

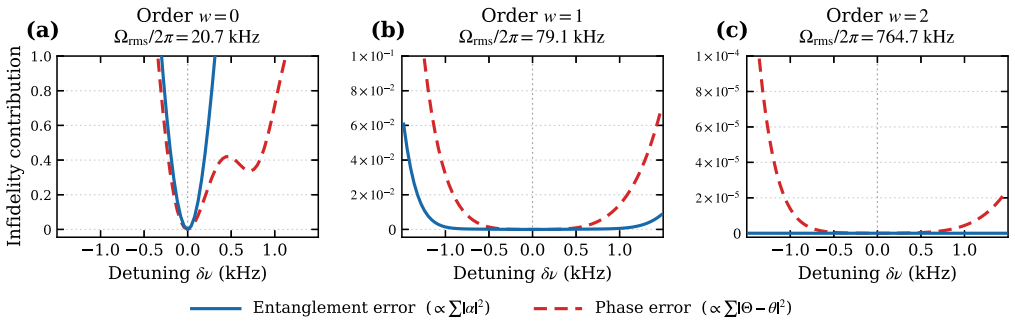


FIG. 2. Robustness of the four-qubit AM global entanglement gate on a 19-ion chain, with $N_s = 200$ segments, gate duration $\tau = 1000\,\mu$s, and detuning $\mu = 1.004\omega_x$. Panels (a)–(c) show the infidelity contributions as functions of a uniform normal-mode frequency drift $\delta\nu$ for robustness orders $w = 0$, 1, and 2, respectively. The blue solid lines show the contribution from residual motional displacement, and the red dashed lines show the accumulated-phase contribution. The higher-order robustness constraints are constructed within the RWA, whereas every plotted drift point is evaluated with the full non-RWA expressions.

reduction in amplitude. Figure 1(c) shows both the root-mean-square (RMS) Rabi frequency, defined as $\sqrt{P/(nN_s)}$ with P given by Eq. (13), and the peak Rabi frequency as functions of the iteration count. Both quantities decrease rapidly and stabilize within approximately ten iterations.

B. Robustness against frequency fluctuations

Following the methodology developed in Sec. II D, we can enhance the robustness of the generated pulse sequences by incorporating additional constraints into the optimization manifold. To demonstrate this, we revisit the four-qubit global entanglement gate described in the previous section and generate three distinct sets of pulse sequences corresponding to robustness orders $w = 0$, 1, and 2. For all three cases, we set $N_s = 200$, which is above the necessary degree-counting estimate in Eq. (19) for $N = 19$, $n = 4$, and $w_{\max} = 2$. Furthermore, the total gate duration is extended to $\tau = 1000\,\mu$s to ensure that the duration of individual pulse segments remains experimentally feasible.

To evaluate the robustness, we introduce a global shift $\delta\nu$ to the normal-mode frequencies, such that $\omega_m \to \omega_m + \delta\nu$ for all $m \in \{1, 2, \ldots, N\}$. For each pulse sequence, we calculate the resulting residual motional displacement $\alpha_{j,m}(\tau)$ and the accumulated phase $\Theta_{j,j'}(\tau)$ in the presence of this drift. The average infidelity $1 - \bar{F}$ is then computed as a function of $\delta\nu$ according to Eq. (20). The results are summarized in Fig. 2. Notably, to clearly distinguish the error contributions from residual entanglement and phase deviations [i.e., the two terms in Eq. (20)], we plot these components separately. The actual total infidelity is the sum of these two contributions.

Several key observations can be drawn from the results: First, the immunity of the gate to frequency drifts improves significantly as the robustness order w increases. Using an infidelity threshold of $1 - \bar{F} = 10^{-4}$ as a benchmark, the tolerable frequency shift increases from approximately ± 0.2 kHz for $w = 1$ to over ± 1.3 kHz for $w = 2$.

Second, as shown in Figs. 2(a)–2(c), the infidelity contribution from the accumulated phase error $\Theta_{j,j'}(\tau)$ is dominant. This aligns well with the discussion of error sources in Sec. II E.

Third, the required laser power increases with the robustness order. This arises because the robustness requirements introduce a large number of additional constraints, which make it more challenging to identify high-quality initial seeds and significantly hinder the efficacy of the power optimization process. Our numerical experiments indicate that this power penalty can be mitigated by extending the total gate duration to increase phase accumulation within each segment, and by

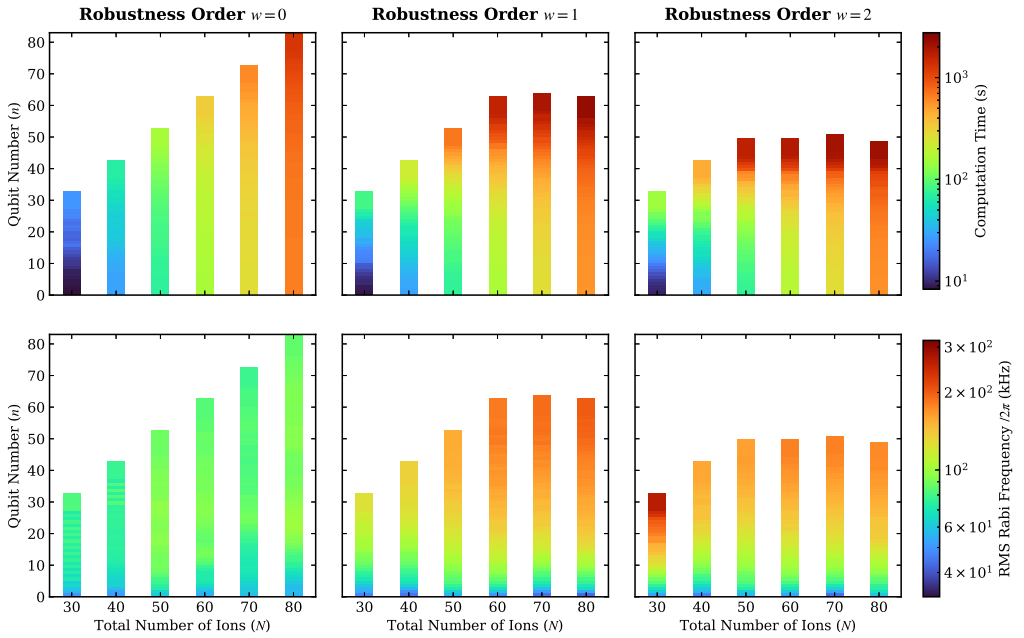


FIG. 3. Benchmarking results for the AM global entanglement gate scheme. The heatmaps show the computation time (top row) and RMS Rabi frequency (bottom row) as functions of the total number of ions N and the number of gate qubits n, for robustness orders $w = 0$, 1, and 2. For each point, N_s is set 50 above the necessary estimate in Eq. (19), and the segment duration is fixed at 5 μs. For each fixed (N, w), n is scanned in ascending order; the scan is terminated if the numerical solver fails or if the measured time of a completed point exceeds 30 min. All calculations were performed on a 28-core server node equipped with Intel® Xeon® Silver 4110 CPUs at 2.10 GHz.

increasing the number of pulse segments to provide more degrees of freedom for optimization, as detailed in Sec. III C.

C. Scalability and performance analysis

In practical applications, the computational performance of a gate-design algorithm—specifically the calculation time and power consumption as functions of the trap size N and the gate size n—is of paramount importance. Conventional approaches to global-gate design often struggle with the complexity of quadratic constraints, which typically limits their efficiency to relatively small-scale systems [22]. In contrast, our algorithm demonstrates a significant competitive advantage in terms of scalability.

We perform a series of benchmarks using the same trap and laser parameters as in Sec. III B. Figure 3 displays the computation time and RMS Rabi frequency required to implement global entanglement gates for robustness orders $w = 0$, 1, and 2, total ion numbers $N = 30, 40, \ldots, 80$, and varying gate sizes n. The gate is applied to the n centermost ions, with target phases $\theta_{j,j'} = \pi/4$ for all pairs. The number of pulse segments N_s is chosen 50 above the necessary estimate in Eq. (19), and the total duration is set to $\tau = N_s \times 5\,\mu s$. Figure 4 shows the corresponding cross section at $N = 40$ and includes the peak Rabi frequency in addition to the computation time and RMS value.

For each fixed (N, w), the gate size n is scanned in ascending order. If the numerical solver fails or the measured time of a completed calculation exceeds 30 min, the scan is terminated and the larger gate sizes are not attempted. Therefore, the boundaries of the colored regions in Fig. 3 indicate the computational range reached under this protocol. Without robustness constraints ($w = 0$), the

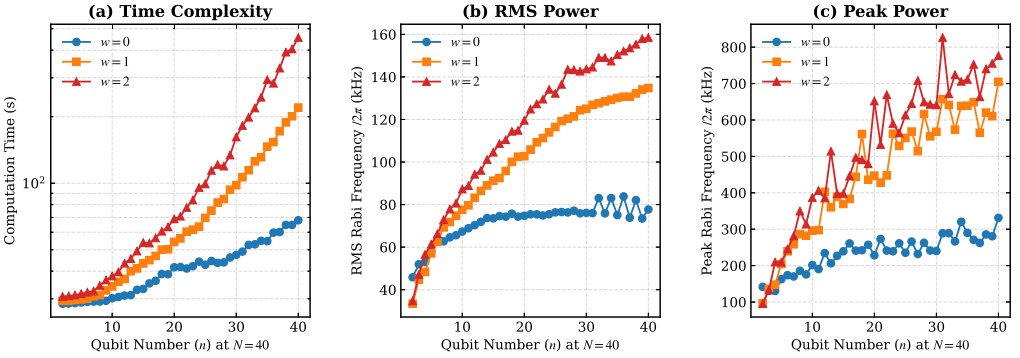


FIG. 4. Performance scaling for a fixed total ion number $N = 40$. (a) Computation time, (b) RMS Rabi frequency, and (c) peak Rabi frequency of the generated pulses as functions of the number of gate qubits n, for robustness orders $w = 0$, 1, and 2. The pulse-segment allocation and computing hardware are the same as in Fig. 3.

algorithm successfully constructs full-chain gates up to $N = n = 80$. Although the accessible range decreases as w increases, medium-scale gates with first- and second-order robustness remain well within reach. As shown in Fig. 4(a), the computation time grows rapidly with n, reflecting the increasing number of quadratic phase constraints for large all-to-all gates.

The lower panels of Figs. 3, 4(b), and 4(c) demonstrate the performance of the power-optimization scheme. The optimized RMS Rabi frequencies remain on the order of 10^2 kHz over a broad range of converged configurations, while the peak values are naturally higher and exhibit stronger instance-to-instance variation. Increasing the robustness order generally requires larger drive amplitudes because of the additional linear and quadratic constraints. Nevertheless, the results show that the proposed method can efficiently generate medium- and large-scale global gates with experimentally feasible pulse amplitudes, including gates with high-order robustness.

In summary, the numerical benchmarks reveal a trade-off between robustness, pulse power, and computation time, while confirming the strong scalability of the proposed construction for the medium-size global gates most relevant to circuit compilation.

IV. QUANTUM SIMULATION CIRCUIT COMPILATION VIA GLOBAL GATES

In this section, we focus on the problem of Hamiltonian simulation. Specifically, given a time-independent sparse Hamiltonian H and a simulation time t, the objective is to implement the evolution operator $U(t) = \exp(-iHt/\hbar)$. We show that the native global entangling gates in trapped-ion systems provide significant advantages in implementing such operators, allowing for a substantial constant-factor optimization in the gate count. We describe the specific compilation algorithms and the corresponding numerical results, proceeding from specific to general Hamiltonian forms.

A. Basic compilation algorithms

To elucidate the proposed compilation algorithm, we begin with the simplest scenario: simulating a 2-local Hamiltonian composed of Pauli strings aligned along the x direction, formulated as

$$H = \hbar \sum_{(j,j')\in E} J_{jj'} X_j X_{j'}, \tag{23}$$

where $G = (V, E)$ represents a simple undirected graph in which each edge defines a pair of interacting qubits, characterizing the interaction topology. This model serves as the

ALGORITHM 1. Basic global gate decomposition.

Input: Hamiltonian H from Eq. (23), time slice Δt, max gate size n
Output: Sequence of global gates $\mathcal{L} = \{G_i = \langle S_i, \theta_i \rangle\}$
Initialize set of uncovered edges $E' \leftarrow E$, $i \leftarrow 1$
while $E' \neq \emptyset$ **do**
  Find vertex $v \in V$ with the maximum residual degree in $G' = (V, E')$
  Initialize $S_i \leftarrow \{v\}$
  while $|S_i| < n$ and neighbors of S_i exist **do**
    Greedy expansion: Add $u \in \mathrm{adj}(S_i)$ to S_i that maximizes the number of newly covered edges in E'
  end while
  Pruning: If $|S_i|$ can be reduced without decreasing the number of covered edges in E', remove redundant vertices.
  Define phase matrix: $(\theta_i)_{j,j'} = -J_{j,j'}\Delta t$ for $(j, j') \in E' \cap \mathrm{edges}(S_i)$, else 0.
  Append $G_i = \langle S_i, \theta_i \rangle$ to $\mathcal{L}$
  Update $E' \leftarrow E' \setminus \mathrm{edges}(S_i)$, $i \leftarrow i + 1$
end while
return $\mathcal{L}$

foundation for subsequent extensions. Our objective is to implement the unitary evolution $U(t) = \exp(-i\sum_{(j,j')\in E}(J_{jj'}t)X_jX_{j'})$. We note that this operator perfectly matches the form of a global gate with arbitrary phases $\theta_{j,j'}$, defined as $G(\theta) = \exp(i\sum_{j,j'\in S, j<j'}\theta_{j,j'}X_jX_{j'})$, by identifying $\theta_{j,j'} \leftrightarrow -J_{jj'}t$. In principle, the simulation of such a Hamiltonian could be achieved using a single global gate. However, to facilitate numerical pulse optimization and ensure stability, it is preferable to keep the accumulated phases relatively small (e.g., $|\theta_{j,j'}| \leqslant \pi/4$). Consequently, we decompose the total evolution time t into several time slices Δt and implement $U(t)$ by repeatedly applying $U(\Delta t)$. Since all terms in the Hamiltonian commute, this decomposition introduces no digital error. The length of the time slice is typically chosen as

$$\Delta t \sim \frac{\pi}{4\max_{(j,j')\in E}|J_{jj'}|}. \tag{24}$$

A critical challenge, as discussed in Secs. II and III, is that the complexity of pulse sequence design scales rapidly with the number of qubits n involved in the global gate. Therefore, high-quality pulse designs for large-scale global gates cannot be efficiently generated within a reasonable time frame. We must assume that the maximum size of a global gate that can be reliably implemented is limited to n qubits. In practical scenarios where the system size $|V| > n$, the simulation is no longer trivial and requires partitioning the interactions.

Within each time slice Δt, the interactions must be partitioned into several groups, each implemented by a corresponding global gate. This task can be abstracted as follows: Given the graph $G = (V, E)$, let $\mathcal{Q} = \{W \subseteq V \mid 2 \leqslant |W| \leqslant n\}$ be the set of all possible vertex subsets with size at most n. Our goal is to find a collection of subsets $\mathcal{C} \subseteq \mathcal{Q}$ such that

(1) all edges in E are covered: $\bigcup_{S\in\mathcal{C}}\{(u, v) \mid u, v \in S\} \supseteq E$;
(2) the total number of gates $|\mathcal{C}|$ is minimized;
(3) secondary to minimizing $|\mathcal{C}|$, the sum of the subset sizes $\sum_{S\in\mathcal{C}}|S|$ is minimized.

By treating E as the universe and the edges within each subset S as the covering sets, this problem reduces to the well-known *Set Cover problem*. Since finding an optimal solution is NP-hard, we employ a modified greedy heuristic to achieve efficiency. The detailed procedure for the proposed decomposition is described in Algorithm 1.

An exhaustive search for the optimal vertex set would incur a prohibitive cost of $\mathcal{O}(|V|^n)$. Our heuristic instead initiates each subset from the vertex with the highest residual degree and expands greedily within its neighborhood. This approach remains nearly strictly greedy for random graphs.

The approximation ratio for the outer greedy loop is bounded by the harmonic series [46], summed up to the maximum size of the covering subsets:

$$\alpha = \sum_{k=1}^{n(n-1)/2} \frac{1}{k} = \mathcal{O}(\ln n). \tag{25}$$

Although the approximation ratio increases with n, it grows very slowly. Given that numerical pulse design is typically restricted to moderate n, this algorithm yields near-optimal results efficiently. Theoretically, since a single global gate can cover up to $n(n-1)/2$ pairwise interactions, we anticipate that this compilation strategy can offer a constant-factor speedup of approximately $\mathcal{O}(n^2)$ compared with traditional methods based on two-qubit gates. Numerical results are presented in the following section.

B. Application to typical spin models and random graphs

In this section, we demonstrate the performance of our compilation scheme by applying it to several typical Hamiltonian models. First, we demonstrate how the Hamiltonians of several common physical models can be reduced to the fundamental case described in Sec. IV A, namely, the implementation of evolution operators of the form $U = \exp(-i \sum_{(j,j')\in E} J_{jj'} t X_j X_{j'})$. To illustrate this, consider the Heisenberg model with a longitudinal external field:

$$H = \sum_{(i,j)\in E} (J_x X_i X_j + J_y Y_i Y_j + J_z Z_i Z_j) + \sum_i h_i Z_i. \tag{26}$$

A standard approach for quantum simulation is the Trotter-Suzuki decomposition [47,48]. For a Hamiltonian $H = \sum_j H_j$, the evolution operator e^{-iHt} can be approximated over a total time t by dividing it into m small time steps $\Delta t = t/m$. To first order, we have $e^{-iHt} \approx (\prod_j e^{-iH_j \Delta t})^m$, where the decomposition error is bounded by commutator-based estimates [49]. Consequently, Eq. (26) can be partitioned into four components: $H = H_X + H_Y + H_Z + H_1$, where $H_X = \sum J_x X_i X_j$, $H_Y = \sum J_y Y_i Y_j$, $H_Z = \sum J_z Z_i Z_j$, and $H_1 = \sum h_i Z_i$. By calculating the required Δt based on error tolerance, we implement the evolution of each part. Note that H_Y and H_Z can be transformed into the XX-type interaction basis via appropriate single-qubit rotations [see Eqs. (29) and (30)]. The interaction graph G corresponds to the lattice dimensionality of the physical system, while the term $e^{iH_1 \Delta t}$ is trivially implemented as single-qubit Z rotations.

We now present concrete examples to benchmark our compilation algorithm. Since a wide class of Hamiltonians can be reduced to the basic XX case, we consider the evolution $U(\Delta t) = \exp(-i \sum_{(j,j')\in E} J_{jj'} \Delta t X_j X_{j'})$ without loss of generality. We examine two interaction topologies: a random graph and a 2D square grid. Figure 5 illustrates the compilation results for a nine-vertex random graph (edge density $p = 1/3$, yielding an expected 12 edges) and a 3×3 grid. Figures 5(a.i) and 5(b.i) display the interaction weights, where the edge weight corresponds to $-J_{jj'} \Delta t$ in units of $\pi/4$. Following the requirements of Eq. (24), these weights are chosen as uniform random numbers in $[0, \pi/4]$.

Figures 5(a.ii) and 5(b.ii) show the results of Algorithm 1, assuming a hardware limit where a global gate can cover at most $n = 4$ qubits. Edges of the same color represent interactions covered by a single global entanglement gate. In both cases, the 12 interactions are efficiently covered by only four gates. For the physical implementation, following the ion-trap setup in Sec. III, we map these nine qubits to the central ions of a 19-ion chain.

In Figs. 5(a.iii) and 5(b.iii), we scan the laser detuning μ over $[0.97\omega_x, 1.03\omega_x]$ and construct pulses with robustness order $w = 2$ using the scheme in Sec. II. Using Eq. (20), we calculate the fidelity of each global gate with the full non-RWA expressions under a uniform mode-frequency drift $\delta\nu = 1.5\,\text{kHz}$ and take the product of the individual-gate fidelities to obtain F_{total}. The total infidelity exhibits a significant minimum near $\mu \approx \omega_x$, demonstrating the importance of detuning selection for the compiled pulse sequence.

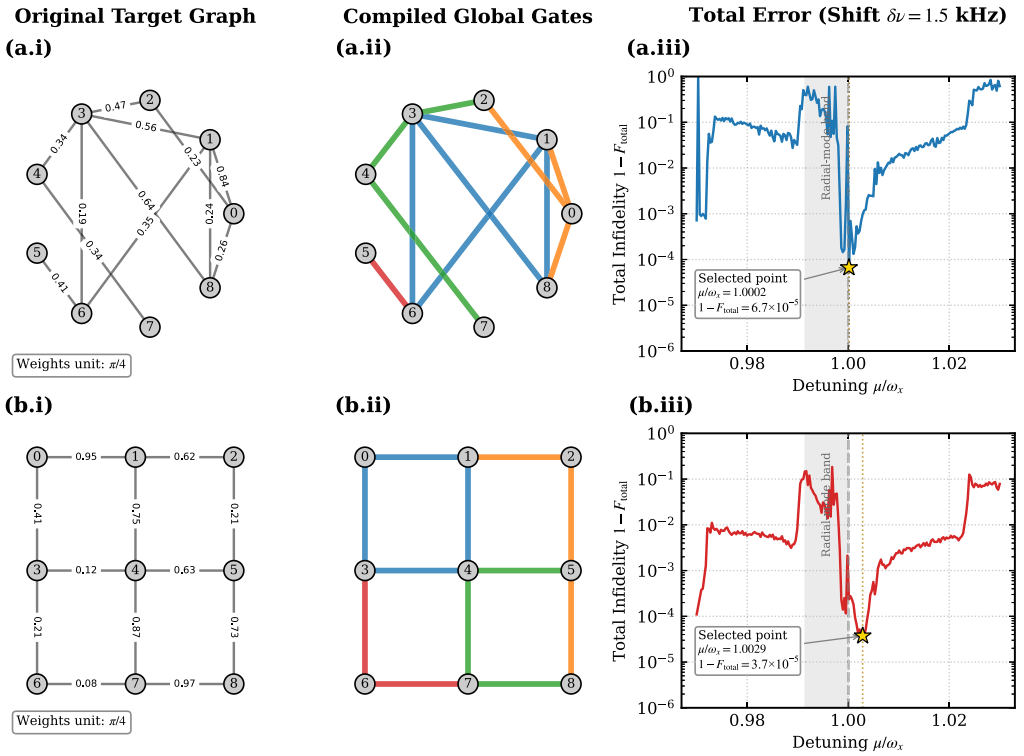


FIG. 5. Global-gate compilation schemes and infidelity scans for two interaction topologies. Row (a) represents a random interaction graph with nine vertices and density $p = 1/3$. Row (b) represents a 3×3 2D square grid. Column (i) shows the target Hamiltonian topology, where vertices represent qubits and edge values denote interaction weights (in units of $\pi/4$). Column (ii) displays the resulting global-gate sequence, where bold lines of the same color represent edges assigned to the same global entanglement gate. Column (iii) provides the total infidelity $1 - F_{\text{total}}$ as a function of the laser detuning μ/ω_x, assuming the qubits are mapped to the central 9 ions of a 19-ion chain with a frequency drift $\delta\nu = 1.5$ kHz. Pulse parameters are set to $\tau = 1000\,\mu$s, $N_s = 200$, and robustness order $w = 2$.

The efficiency of the compilation algorithm is first examined in Fig. 6(a). For random interaction graphs with $N_q = 80$, the number of required gates grows approximately linearly with the number of interaction edges. The fitted average numbers of covered edges per gate are 1.00, 8.37, and 19.55 for two-qubit gates and global gates with $n_{\max} = 5$ and 8, respectively. For $n_{\max} = 5$, the average coverage is close to the theoretical maximum $n_{\max}(n_{\max} - 1)/2 = 10$, indicating high utilization. Increasing the gate size to $n_{\max} = 8$ further improves the average coverage, although the value is farther from the corresponding maximum of 28. In general, denser interaction graphs permit a larger fraction of the available pairwise couplings within each global gate to be utilized. However, most physical Hamiltonians are sparse rather than fully connected, so the number of interaction edges available within a selected vertex subset does not grow as rapidly as $n_{\max}(n_{\max} - 1)/2$. Consequently, increasing $n_{\max}$ leads to diminishing improvements in gate-count reduction, while the computational cost of pulse design grows rapidly with the gate size, as discussed in Sec. III C. This results in a practical trade-off between circuit compression and pulse-design complexity.

Figures 6(b)–6(f) compare three implementations: a native two-qubit-gate decomposition ($n_{\max} = 2$), represented by an R_{XX}-based basis gate set [50], and global-gate compilations with $n_{\max} = 5$ and 8. We consider randomized connected 2D grids and random graphs with connection probability $p = 0.1$, with system sizes $N_q = 10, 20, \ldots, 80$. For each topology and system size, 20

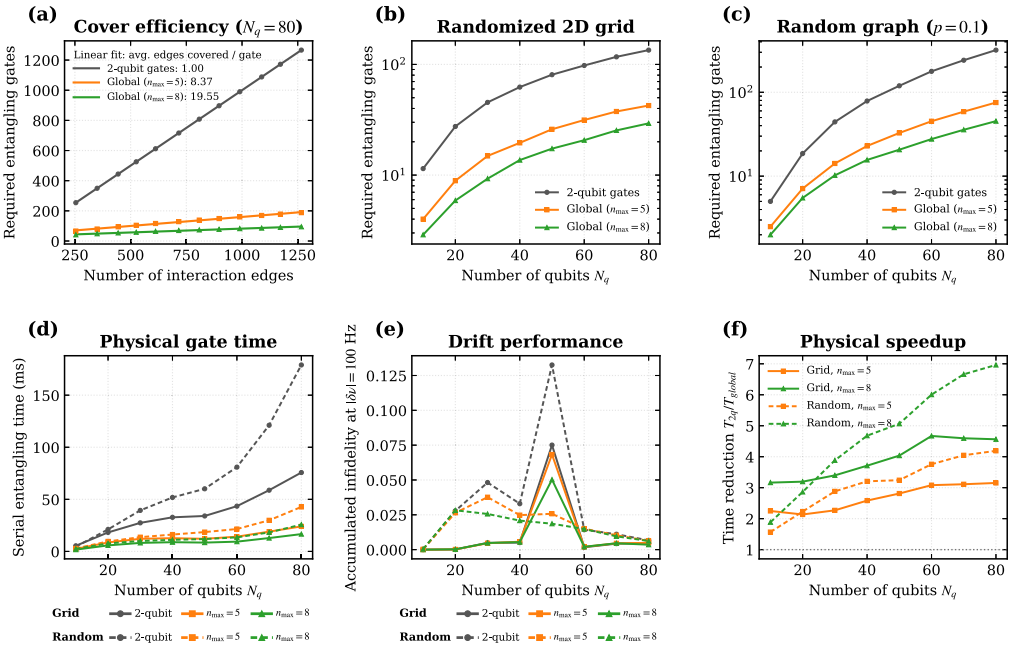


FIG. 6. Compilation efficiency and pulse-level performance. (a) Required entangling-gate count vs the number of interaction edges for random graphs with $N_q = 80$. The lines are linear fits; the fitted average numbers of covered edges per gate are 1.00, 8.37, and 19.55 for two-qubit gates and global gates with $n_{\max} = 5$ and 8, respectively. (b), (c) Entangling-gate counts for randomized 2D grids and random graphs with connection probability $p = 0.1$. (d) Total serial entangling time obtained by explicitly designing every compiled gate. (e) Accumulated infidelity under a uniform mode-frequency shift $|\delta\nu| = 100\,\mathrm{Hz}$. (f) Physical time reduction $T_{2q}/T_{\mathrm{global}}$, where T_{2q} and T_{global} denote the total serial entangling times of the two-qubit and global-gate compilations for the same interaction instance. Panels (b)–(f) show the average results using 20 independently generated weighted instances for each topology and N_q.

independent instances are generated and the corresponding results are averaged. The same instance is used for all three compilation choices. The nonzero interaction weights are independently sampled from $[0.2, 1] \times \pi/4$, corresponding to the interaction phases within a single Trotter step.

Figures 6(b) and 6(c) show that global gates substantially reduce the number of entangling operations for both interaction topologies. The reduction is more pronounced for the random graphs, where a selected set of qubits can contain many interaction edges. In contrast, the maximum degree Δ of the 2D grid is fixed at $\Delta = 4$, so its edge density decreases as N_q grows. The interactions covered by one global gate are therefore largely chainlike, and the improvement is closer to $\mathcal{O}(n_{\max})$ than to the dense-graph limit $\mathcal{O}(n_{\max}^2)$.

To evaluate the physical resources, we explicitly design an AM pulse for every compiled gate. We choose robustness order $w = 2$ and set the number of segments to 80 above the necessary estimate in Eq. (19). For each gate, the segment duration is progressively reduced using a coarse scan followed by a boundary refinement, and the shortest solution satisfying a residual tolerance of 10^{-8}, $\Omega_{\mathrm{RMS}}/(2\pi) \leqslant 500\,\mathrm{kHz}$, and $\Omega_{\mathrm{peak}}/(2\pi) \leqslant 1500\,\mathrm{kHz}$ is selected. The accumulated infidelity in Fig. 6(e) is evaluated with the full non-RWA expressions under a uniform mode-frequency shift $|\delta\nu| = 100\,\mathrm{Hz}$.

Figure 6(d) shows that the reduction in gate count leads to a substantial decrease in the total serial entangling time. For either compilation route, the total serial entangling time is defined as $T = \sum_{g=1}^{N_g} \tau_g$, where τ_g is the duration of the gth entangling pulse. Thus, T_{2q} and T_{global} both sum

the durations of all constituent entangling gates, without parallel scheduling of gates on disjoint qubits. This serial convention is appropriate for independently designed pulses on a common ion chain, since simultaneous drives generally couple through shared collective motional modes and can produce unwanted cross-phases unless the concurrent operations are jointly designed. In Fig. 6(f), T_{2q}/T_{global} is evaluated for each identical graph instance before averaging and reaches factors of several, with the largest gains obtained for random graphs and $n_{\max} = 8$. As shown in Fig. 6(e), the global-gate compilations maintain accumulated drift infidelities comparable to, and in many cases lower than, the two-qubit-gate baseline. The visible fluctuations with N_q originate from changes in the normal-mode spectrum and in the shortest feasible pulse duration for each system size.

Overall, the global-gate compilation provides a clear advantage in both entangling-gate count and physical gate time while preserving comparable robustness against frequency drifts.

C. Extension to more general cases

In this section, we provide a corresponding compilation approach for Hamiltonians, which are of a more general form.

1. 2-local terms with misaligned directions

In Secs. IV A and IV B, the proposed compilation strategy was restricted to cases where the two Pauli operators in each interaction term share the same direction. We now extend this framework to Hamiltonians with misaligned directions, such as

$$H = \sum_{(i,j)\in E_{XY}} J_{ij} X_i Y_j + \sum_{(i,j)\in E_{YZ}} K_{ij} Y_i Z_j + \sum_{(i,j)\in E_{ZX}} L_{ij} Z_i X_j, \tag{27}$$

where E_{XY}, E_{YZ}, E_{ZX} denote the sets of interaction edges for the respective misaligned categories. More formally, such a Hamiltonian can be expressed as

$$H = \sum_{P,Q\in\{X,Y,Z\}} \sum_{(i,j)\in E_{PQ}} c_{P_iQ_j} P_i Q_j. \tag{28}$$

Since terms in Eq. (28) may no longer commute, we employ the Trotter-Suzuki decomposition framework. The evolution time is partitioned into time slices Δt based on the desired simulation precision, and global gates are partitioned within each slice. Note that 1-local terms are omitted here, as the overhead for single-qubit rotations is negligible in trapped-ion systems.

To handle these terms, we apply pre- and postrotations to specific qubits to align the interactions with the x direction before applying the global gate. We define the rotation operator as $R_P(\theta) = \exp(i\theta P)$ for $P \in \{X, Y, Z\}$. To map the X operator onto other Pauli operators, we introduce the following transformation unitaries:

$$U_X = I, \quad U_Y = R_Z\left(\frac{\pi}{4}\right), \quad U_Z = R_Y\left(-\frac{\pi}{4}\right). \tag{29}$$

These definitions lead to the following identity:

$$P = U_P^\dagger X U_P, \quad P \in \{X, Y, Z\}. \tag{30}$$

This problem can be abstracted into a constrained graph covering problem on a tripartitelike layered graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, defined as

$$\mathcal{V} = \{i_P \mid i \in V, P \in \{X, Y, Z\}\}, \tag{31}$$

$$\mathcal{E} = \{(i_P, j_Q) \mid \text{the term } P_i Q_j \text{ exists in } H\}. \tag{32}$$

Our goal is to find a collection of subsets $\mathcal{C} \subseteq \mathcal{Q} = \{W \subseteq \mathcal{V} \mid 2 \leqslant |W| \leqslant n \text{ and } \forall i, |W \cap \{i_X, i_Y, i_Z\}| \leqslant 1\}$ that covers all edges in $\mathcal{E}$ while minimizing $|\mathcal{C}|$. The exclusivity constraint ($|W \cap \{i_X, i_Y, i_Z\}| \leqslant 1$) reflects the physical reality that only one local basis transformation can

ALGORITHM 2. Global-gate decomposition for misaligned 2-local terms.

Input: Hamiltonian H from Eq. (28), time slice Δt, max gate size n
Output: Sequence of global gates $\mathcal{L} = \{G_i = \langle S_i, \theta_i, \{U_{ij}\}_{j\in S_i}\rangle\}$

Construct the layered graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
Initialize set of uncovered edges $\mathcal{E}' \leftarrow \mathcal{E}$, $i \leftarrow 1$
while $\mathcal{E}' \neq \emptyset$ **do**
  Find node $v \in \mathcal{V}$ with the maximum residual degree in $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$
  Initialize virtual set $V_i \leftarrow \{v\}$
  while $|V_i| < n$ and valid neighbors exist **do**
    Greedy expansion: Add $u \in \mathrm{adj}(V_i)$ to V_i that maximizes newly covered edges in $\mathcal{E}'$, subject to the constraint: if $i_P \in V_i$, then $i_Q \notin V_i$ for $Q \neq P$.
  end while
  Pruning: Minimize V_i while maintaining the count of covered edges in $\mathcal{E}'$.
  Define active qubit set: $S_i = \{j \mid \exists P, j_P \in V_i\}$
  Phase matrix: $(\theta_i)_{j,j'} = -c_{P_j Q_{j'}}\Delta t$ for $(j_P, j'_Q) \in \mathcal{E}' \cap \mathrm{edges}(V_i)$, else 0.
  Basis transformations: $U_{ij} = U_P$ [from Eq. (30)] for each $j_P \in V_i$.
  Append $G_i = \langle S_i, \theta_i, \{U_{ij}\}\rangle$ to $\mathcal{L}$
  Update $\mathcal{E}' \leftarrow \mathcal{E}' \setminus \mathrm{edges}(V_i)$, $i \leftarrow i + 1$
end while
return $\mathcal{L}$

be applied to a single ion during one global-gate operation. The detailed procedure is outlined in Algorithm 2.

This exclusivity constraint does not fundamentally alter the performance of the algorithm; the underlying logic remains a greedy approach for the Set Cover problem. Although the set of valid candidate subsets is restricted compared to the uniform case in Sec. IV A, the approximation ratio remains $\mathcal{O}(\ln n)$. Thus, the same greedy Set Cover logic retains the potential $\mathcal{O}(n^2)$ dense-graph reduction in entangling-gate count relative to individual two-qubit interactions.

2. k-local terms

Finally, we address the most general and challenging case: Hamiltonians with arbitrary k-local Pauli strings, characterized by random connectivity, directions, and coefficients. Within each Trotter time slice Δt, 2-local terms are first partitioned and implemented using the strategy described in the previous section. The remaining k-local terms ($k > 2$) are then processed individually. We propose a decomposition method that achieves an $O(k)$ reduction in gate complexity compared to traditional methods.

Without loss of generality, consider the implementation of the unitary $U = \exp(i\theta X_1 X_2 \cdots X_k)$. For terms with different Pauli bases, we can apply local rotations to align them to the x direction, as demonstrated previously. The fundamental gates available are

$$\begin{aligned} &R_{XX}(i, j, \theta) = \exp(i\theta X_i X_j), \\ &R_P(i, \theta) = \exp(i\theta P_i), \quad P \in \{X, Y, Z\}. \end{aligned} \tag{33}$$

A key property of these gates is as follows: If a Pauli string P anticommutes with X_iX_j, i.e., $\{P, X_iX_j\} = 0$, then the following similarity transformation holds:

$$R_{XX}(i, j, -\pi/4)PR_{XX}(i, j, \pi/4) = R_{XX}(i, j, -\pi/2)P = -iX_iX_jP. \tag{34}$$

By introducing an ancilla qubit A and leveraging this property, we can iteratively transform the Z_A operator. For $k = 1$, we have

$$R_{XX}(1, A, -\pi/4)Z_AR_{XX}(1, A, \pi/4) = -iX_1X_AZ_A = -X_1Y_A. \tag{35}$$

For $k = 2$, repeating the process yields

$$R_{XX}(2, A, -\pi/4)(-X_1Y_A)R_{XX}(2, A, \pi/4) = -iX_2X_A(-X_1Y_A) = -X_1X_2Z_A. \tag{36}$$

By induction, we obtain the general form

$$\left[\prod_{j=1}^{k} R_{XX}(j, A, -\pi/4)\right] Z_A \left[\prod_{j=1}^{k} R_{XX}(j, A, \pi/4)\right] \tag{37}$$

$$= (-i)^k X_1 \cdots X_k (X_A)^k Z_A$$

$$= \begin{cases} (-1)^{k/2} X_1 \cdots X_k Z_A, & k \text{ is even}, \\ -(-1)^{(k-1)/2} X_1 \cdots X_k Y_A, & k \text{ is odd}. \end{cases} \tag{38}$$

Crucially, since all terms $\{X_jX_A\}_{j=1}^{k}$ commute with each other, the operator $V = \exp(i\frac{\pi}{4}\sum_{j=1}^{k} X_jX_A)$ can be implemented as a single global gate.

For even k, we implement the target unitary via

$$U = V^\dagger R_Z(A, \phi)V = \exp(i\phi V^\dagger Z_A V) = \exp(i\phi(-1)^{k/2} X_1 \cdots X_k Z_A). \tag{39}$$

By initializing the ancilla in the $|0\rangle$ state, the Z_A operator acts as identity on the computational subspace. Thus, setting $\phi = (-1)^{k/2}\theta$ correctly implements $e^{i\theta X_1 \cdots X_k}$.

For odd k, we have

$$U = V^\dagger R_Z(A, \phi)V = \exp\left(-i\phi(-1)^{(k-1)/2} X_1 \cdots X_k Y_A\right). \tag{40}$$

In this case, the ancilla is prepared in the $|+\rangle_y = R_X(A, \pi/2)|0\rangle$ state (the $+1$ eigenstate of Y_A). Setting $\phi = -(-1)^{(k-1)/2}\theta$ implements the desired evolution. The ancilla is then returned to $|0\rangle$ via $R_X(A, -\pi/2)$.

In summary, a k-local term is decomposed into just two global gates and one local rotation, resulting in $O(1)$ complexity in terms of global gates. In contrast, traditional compilation requires $O(k)$ two-qubit gates. Since physical Hamiltonians are typically dominated by local interactions, the qubit capacity of modern global gates is well suited for this optimization. Furthermore, the ancilla qubit is disentangled from the system after each k-local operation, allowing it to be recycled as a universal workspace qubit.

V. CONCLUSION

In this work, we have presented a comprehensive, efficient theoretical and numerical framework for implementing multiqubit global entanglement gates in trapped-ion systems. Furthermore, we applied this framework to optimize the compilation of quantum simulation circuits.

At the foundational level of pulse control, within the Mølmer-Sørensen multimode gate framework, we developed a numerical solution method for piecewise-constant AM pulses. The method uses null-space projection to satisfy the linear residual-entanglement constraints and transforms the remaining pulse-design problem into a set of quadratic equations in a lower-dimensional subspace. Combined with tangent-space gradient descent, it substantially reduces the RMS and peak laser amplitudes while maintaining the gate constraints to the prescribed numerical tolerance. Furthermore, we derived an analytical formula for the average fidelity of an n-qubit global gate and introduced high-order robustness constraints against normal-mode frequency drifts, generating high-fidelity and noise-resilient pulse sequences.

Notably, our algorithmic framework exhibits a high degree of physical compatibility and scalability. On the one hand, this framework can be easily extended to phase-modulated schemes [16,17]. By employing specific substitution techniques [22,51], the complex trigonometric integral constraints can be simplified into linear constraints accompanied by several additional quadratic constraints, which can then be solved using our algorithmic pipeline. On the other hand, based on the current

method, by simply making certain modifications to the solution of the normal modes to incorporate the effects of micromotion [52,53], this algorithm can be directly applied to 2D trapped-ion systems. However, this requires the experimental capability to simultaneously and independently address multiple ions, which currently faces several challenges [19] and will rely on future technological breakthroughs.

Moving to upper-level applications, we demonstrated the immense potential of global gates in compiling complex Hamiltonian evolutions. By abstracting the partitioning of the interaction graph into a Set Cover problem, we proposed a heuristic greedy compilation algorithm.

Numerical results indicate that, for direction-aligned 2-local physical models, the scheme achieves a significant constant-factor reduction in entangling-gate count compared with traditional two-qubit-gate decompositions. Explicit pulse-level benchmarks further show that this reduction is accompanied by a substantial decrease in total serial entangling time while maintaining comparable frequency-drift performance. Furthermore, we extended the compilation strategy to Hamiltonians with misaligned directions and general k-local interactions, preserving a stable resource advantage for a broad class of quantum-simulation tasks.

Synthesizing the numerical results from both low-level pulse design and upper-level compilation, we further revealed the law of diminishing marginal returns for global gates in sparse interaction systems. This implies that in current fault-tolerant quantum computing and NISQ experiments, the indiscriminate pursuit of ultra-large-scale global gates with all-to-all connectivity is not the optimal strategy. Conversely, utilizing our framework to design and apply medium-scale high-fidelity global gates is sufficient to strike the optimal balance between circuit depth compression and pulse control complexity. This establishes a practical roadmap for the hardware-software codesign of near-term quantum simulators.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grants No. 12474486 and No. 12234002) and by the National Key Research and Development Program of China (Grant No. 2024YFA1612101).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: EXPLICIT FORMS OF THE CONSTRAINT MATRICES

Based on the theoretical framework in Sec. II A and the settings in Sec. II B, by separating the real and imaginary parts of $\alpha_{j,m}(\tau)$, the elements of the matrix $\mathbf{M}_j$ can be straightforwardly derived as

$$\begin{aligned} M_j(p,q) &= \eta_{j,p} \int_{(q-1)\tau_s}^{q\tau_s} \sin(\mu t)\sin(\omega_p t)\,dt, \\ M_j(p+N,q) &= \eta_{j,p} \int_{(q-1)\tau_s}^{q\tau_s} \sin(\mu t)\cos(\omega_p t)\,dt, \end{aligned} \tag{A1}$$

where $p \in \{1, \ldots, N\}$, $q \in \{1, \ldots, N_s\}$, and $j \in S$. Similarly, the elements of the symmetric matrix $\mathbf{S}_{j,j'}$ are given by

$$S_{j,j'}(p,q) = \sum_{m=1}^{N} \eta_{j,m}\eta_{j',m} \begin{cases} \int_{(p-1)\tau_s}^{p\tau_s} dt_1 \int_{(q-1)\tau_s}^{q\tau_s} dt_2 \ \sin(\mu t_1)\sin(\mu t_2)\sin[\omega_m(t_1-t_2)], & p > q, \\ 2\int_{(p-1)\tau_s}^{p\tau_s} dt_1 \int_{(p-1)\tau_s}^{t_1} dt_2 \ \sin(\mu t_1)\sin(\mu t_2)\sin[\omega_m(t_1-t_2)], & p = q, \\ \int_{(q-1)\tau_s}^{q\tau_s} dt_1 \int_{(p-1)\tau_s}^{p\tau_s} dt_2 \ \sin(\mu t_1)\sin(\mu t_2)\sin[\omega_m(t_1-t_2)], & p < q. \end{cases} \tag{A2}$$

Based on the definition of robustness in Sec. II D, directly taking the kth-order derivatives of $\alpha_{j,m}(\tau)$ and $\Theta_{j,j'}(\tau)$ with respect to ω_m yields integral terms that, while analytically solvable, are exceedingly complex for numerical implementation. Therefore, for the robustness constraints, we adopt the RWA.

For the primary constraints $\Theta_{j,j'}(\tau) = \theta_{j,j'}$ and $\alpha_{j,m}(\tau) = 0$, we continue to use the full definitions in Eqs. (A1) and (A2); the approximation is applied only to the robustness conditions. Under the RWA and setting $\phi_j(t) \equiv 0$, the expressions become

$$\begin{aligned} \alpha_{j,m}(\tau) &\approx \frac{1}{2}\eta_{j,m}\int_0^\tau \Omega_j(t)e^{i(\omega_m-\mu)t}\,dt, \\ \Theta_{j,j'}(\tau) &\approx \frac{1}{4}\sum_{m=1}^{N}\eta_{j,m}\eta_{j',m}\int_0^\tau dt_1\int_0^{t_1} dt_2[\Omega_j(t_1)\Omega_{j'}(t_2)+\Omega_{j'}(t_1)\Omega_j(t_2)]\sin[(\omega_m-\mu)(t_1-t_2)]. \end{aligned} \tag{A3}$$

Taking the kth derivative of these approximated forms yields

$$\begin{aligned} \frac{\partial^k \alpha_{j,m}(\tau)}{\partial \omega_m^k} &= \frac{1}{2}\eta_{j,m} i^k \int_0^\tau t^k \Omega_j(t)e^{i(\omega_m-\mu)t}\,dt, \\ \frac{\partial^k \Theta_{j,j'}(\tau)}{\partial \omega_m^k} &= \frac{1}{4}\eta_{j,m}\eta_{j',m}\int_0^\tau dt_1\int_0^{t_1} dt_2(t_1-t_2)^k[\Omega_j(t_1)\Omega_{j'}(t_2)+\Omega_{j'}(t_1)\Omega_j(t_2)] \\ &\quad \times \sin\left((\omega_m-\mu)(t_1-t_2)+\frac{k\pi}{2}\right). \end{aligned} \tag{A4}$$

The robustness matrix elements $M_j^k(p,q)$ and $S_{j,j'}^{k,m}(p,q)$ are then obtained by separating the real and imaginary parts:

$$\begin{aligned} M_j^k(p,q) &= \frac{1}{2}\eta_{j,p}\int_{(q-1)\tau_s}^{q\tau_s} t^k \sin((\omega_p-\mu)t)\,dt, \\ M_j^k(p+N,q) &= \frac{1}{2}\eta_{j,p}\int_{(q-1)\tau_s}^{q\tau_s} t^k \cos((\omega_p-\mu)t)\,dt. \end{aligned} \tag{A5}$$

The elements of $\mathbf{S}_{j,j'}^{k,m}$ are defined piecewise for even and odd k as follows:

(1) For k even,

$$S_{j,j'}^{k,m}(p,q) = \begin{cases} \frac{1}{4}\eta_{j,m}\eta_{j',m}\int_{(p-1)\tau_s}^{p\tau_s} dt_1\int_{(q-1)\tau_s}^{q\tau_s} dt_2(t_1-t_2)^k\sin((\omega_m-\mu)(t_1-t_2)), & p > q, \\ \frac{1}{2}\eta_{j,m}\eta_{j',m}\int_{(p-1)\tau_s}^{p\tau_s} dt_1\int_{(p-1)\tau_s}^{t_1} dt_2(t_1-t_2)^k\sin((\omega_m-\mu)(t_1-t_2)), & p = q, \\ \frac{1}{4}\eta_{j,m}\eta_{j',m}\int_{(q-1)\tau_s}^{q\tau_s} dt_1\int_{(p-1)\tau_s}^{p\tau_s} dt_2(t_1-t_2)^k\sin((\omega_m-\mu)(t_1-t_2)), & p < q. \end{cases} \tag{A6}$$

(2) For k odd,

$$S^{k,m}_{j,j'}(p,q) = \begin{cases} \frac{1}{4}\eta_{j,m}\eta_{j',m}\int_{(p-1)\tau_s}^{p\tau_s} dt_1 \int_{(q-1)\tau_s}^{q\tau_s} dt_2 (t_1-t_2)^k \cos((\omega_m-\mu)(t_1-t_2)), & p>q, \\ \frac{1}{2}\eta_{j,m}\eta_{j',m}\int_{(p-1)\tau_s}^{p\tau_s} dt_1 \int_{(p-1)\tau_s}^{t_1} dt_2 (t_1-t_2)^k \cos((\omega_m-\mu)(t_1-t_2)), & p=q, \\ \frac{1}{4}\eta_{j,m}\eta_{j',m}\int_{(q-1)\tau_s}^{q\tau_s} dt_1 \int_{(p-1)\tau_s}^{p\tau_s} dt_2 (t_1-t_2)^k \cos((\omega_m-\mu)(t_1-t_2)), & p<q. \end{cases} \tag{A7}$$

All integrals in the above definitions can be computed analytically with ease, ensuring high computational efficiency for constructing the constraint matrices. In particular, if we only consider a uniform shift in the normal-mode frequencies, i.e., $\omega_m \to \omega_m + \delta\nu$ for all $m \in \{1, 2, \ldots, N\}$, the quadratic constraints on the S matrices can be simplified. By defining

$$\mathbf{S}^{k}_{j,j'} = \sum_{m=1}^{N} \mathbf{S}^{k,m}_{j,j'}, \tag{A8}$$

we only require the following condition to be satisfied:

$$\mathbf{\Omega}^T_j \mathbf{S}^{k}_{j,j'} \mathbf{\Omega}_{j'} = 0, \quad j, j' \in S,\ j < j', \quad k \in \{1, \ldots, w\}. \tag{A9}$$

This approach reduces the number of additional constraints to $1/N$ of the original requirement, thereby significantly facilitating the optimization process.

APPENDIX B: DERIVATION OF THE n-QUBIT GATE FIDELITY FORMULA

Here, we provide the detailed derivation of the relationship between the n-qubit gate fidelity F and the parameters $\alpha_{j,m}(\tau)$ and $\Theta_{j,j'}(\tau)$, as expressed in Eq. (20). While Ref. [13] provides a derivation specifically for two-qubit gates, we present a more general derivation following a similar methodology.

For the implementation of an n-qubit global entangling gate acting on a set of ions $S = \{j_1, j_2, \ldots, j_n\}$, the native global gate takes the form

$$U = \exp\left(\sum_{j\in S}\sum_{m} (\alpha_{j,m}(\tau) a^\dagger_m - \alpha^*_{j,m}(\tau) a_m) X_j + i \sum_{j,j'\in S, j<j'} \Theta_{j,j'}(\tau) X_j X_{j'} \right). \tag{B1}$$

We aim to approximate the ideal unitary transformation:

$$U_{\mathrm{id}} = \exp\left(i \sum_{j,j'\in S, j<j'} \theta_{j,j'} X_j X_{j'} \right). \tag{B2}$$

As noted in Refs. [42, 43], for a quantum channel $\mathcal{M}$, the average gate fidelity can be written as

$$\bar{F} = \frac{dF_e + 1}{d+1}, \tag{B3}$$

where $d = 2^n$ is the dimension of the Hilbert space and F_e is the entanglement fidelity of the maximally entangled state $|\Phi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle$ (where $\{|i\rangle\}$ is an orthonormal basis), given by

$$F_e = \langle\Phi|(I \otimes \mathcal{M})(|\Phi\rangle\langle\Phi|)|\Phi\rangle. \tag{B4}$$

For gate fidelity, the channel is defined as

$$\mathcal{M}(\rho) = U^\dagger_{\mathrm{id}} \mathcal{E}(\rho) U_{\mathrm{id}}, \tag{B5}$$

where $\mathcal{E}(\rho) = \mathrm{Tr}_{\mathrm{ext}}(U(\rho \otimes \rho_{\mathrm{th}})U^\dagger)$. Here, ρ_{th} represents the external motional state (thermal state of phonons). Since the qubits are carried by the internal states of the ions, we take the partial trace

over the external states. Further calculation of F_e yields

$$F_e = \frac{1}{d}\sum_{i,j}\langle\Phi|(|i\rangle\langle j|\otimes\mathcal{M}(|i\rangle\langle j|))|\Phi\rangle. \tag{B6}$$

Noting that for the maximally entangled state $|\Phi\rangle$:

$$\langle\Phi|(A\otimes B)|\Phi\rangle = \frac{1}{d}\mathrm{Tr}(A^T B), \tag{B7}$$

it follows that

$$F_e = \frac{1}{d^2}\sum_{i,j}\mathrm{Tr}(|j\rangle\langle i|\mathcal{M}(|i\rangle\langle j|)) = \frac{1}{d^2}\sum_{i,j}\langle i|\mathcal{M}(|i\rangle\langle j|)|j\rangle. \tag{B8}$$

We now calculate the specific expressions. We choose the basis $\{|i\rangle\}$ to be the eigenvectors of $\{X_j\}$, denoted as $\{|s\rangle\}$, which satisfy $X_j|s\rangle = s_j|s\rangle$, where $s_j \in \{-1, 1\}$ for all $j \in S$. It then suffices to calculate $\mathcal{M}(|s\rangle\langle s'|)$. First, calculating $\mathcal{E}(|s\rangle\langle s'|)$, we consider

$$U|s\rangle = e^{i\phi_s}\left(\prod_m D_m(\alpha_{s,m})\right)|s\rangle, \tag{B9}$$

where we introduce

$$D_m(\alpha) = \exp(\alpha a_m^\dagger - \alpha^* a_m),$$

$$\alpha_{s,m} = \sum_{j\in S} s_j\alpha_{j,m}(\tau), \tag{B10}$$

$$\phi_s = \sum_{j<j'}\Theta_{j,j'}(\tau)s_j s_{j'}. \tag{B11}$$

Then,

$$\mathcal{E}(|s\rangle\langle s'|) = e^{i(\phi_s-\phi_{s'})}\mathrm{Tr}(D_s\rho_{\mathrm{th}}D_{s'}^\dagger)|s\rangle\langle s'|, \tag{B12}$$

where $D_s = \prod_m D_m(\alpha_{s,m})$. Specifically, using the properties of displacement operators D from Ref. [13]:

$$D(\alpha)D(\beta) = e^{(\alpha\beta^*-\alpha^*\beta)/2}D(\alpha+\beta),$$

$$\mathrm{Tr}(D(\alpha)\rho_{\mathrm{th}}) = \exp\left(-\frac{|\alpha|^2}{2}\coth\left(\frac{\hbar\omega}{2k_BT}\right)\right), \tag{B13}$$

we have

$$\begin{aligned}
\mathrm{Tr}(D_s\rho_{\mathrm{th}}D_{s'}^\dagger) &= \prod_m \mathrm{Tr}_m(D_m(\alpha_{s,m})\rho_{\mathrm{th},m}D_m^\dagger(\alpha_{s',m}))\\
&= \prod_m \mathrm{Tr}_m(D_m^\dagger(\alpha_{s',m})D_m(\alpha_{s,m})\rho_{\mathrm{th},m})\\
&= \prod_m \mathrm{Tr}_m(D_m(-\alpha_{s',m})D_m(\alpha_{s,m})\rho_{\mathrm{th},m})\\
&= \prod_m e^{i\mathrm{Im}(\alpha_{s',m}^*\alpha_{s,m})}\mathrm{Tr}_m(D_m(\alpha_{s,m}-\alpha_{s',m})\rho_{\mathrm{th},m})\\
&= \exp\left(i\sum_m \mathrm{Im}(\alpha_{s',m}^*\alpha_{s,m}) - \frac{1}{2}\sum_m|\alpha_{s,m}-\alpha_{s',m}|^2(2\bar{n}_m+1)\right),
\end{aligned} \tag{B14}$$

where we introduce the average phonon number $\bar{n}_m$ satisfying $2\bar{n}_m + 1 = \coth(\frac{\hbar\omega_m}{2k_BT})$. In summary,

$$\mathcal{E}(|s\rangle\langle s'|) = \exp\left(i(\phi_s - \phi_{s'}) + i\sum_m \mathrm{Im}(\alpha^*_{s',m}\alpha_{s,m}) - \frac{1}{2}\sum_m |\alpha_{s,m} - \alpha_{s',m}|^2(2\bar{n}_m + 1)\right)|s\rangle\langle s'|$$
$$\triangleq \exp(i(\phi_s - \phi_{s'}) + iK_{s,s'} - L_{s,s'})|s\rangle\langle s'|. \tag{B15}$$

Substituting this into Eq. (B8) and using Eq. (B3), we obtain the explicit expression of the average fidelity:

$$\bar{F} = \frac{1}{d(d+1)}\left(\sum_{s,s'}\langle s|U_{\mathrm{id}}^\dagger\mathcal{E}(|s\rangle\langle s'|)U_{\mathrm{id}}|s'\rangle + d\right) \triangleq \frac{1}{d(d+1)}\left(\sum_{s,s'} f_{s,s'} + d\right), \tag{B16}$$

where

$$f_{s,s'} = \exp(i[(\phi_s - \Phi_s) - (\phi_{s'} - \Phi_{s'})] + iK_{s,s'} - L_{s,s'}), \quad \Phi_s = \sum_{j<j'}\theta_{j,j'}s_js_{j'}. \tag{B17}$$

Defining $\delta\phi_s = \phi_s - \Phi_s = \sum_{j<j'}(\Theta_{j,j'}(\tau) - \theta_{j,j'})s_js_{j'} \triangleq \sum_{j<j'}\delta\theta_{j,j'}s_js_{j'}$, we can expand $f_{s,s'}$ to the second order:

$$f_{s,s'} \approx 1 + i(\delta\phi_s - \delta\phi_{s'} + K_{s,s'}) - \tfrac{1}{2}(\delta\phi_s - \delta\phi_{s'} + K_{s,s'})^2 - L_{s,s'}. \tag{B18}$$

Clearly, when swapping s and s', the first-order terms change sign; thus, the first-order imaginary terms sum to zero. Next,

$$\begin{aligned}\sum_{s,s'} L_{s,s'} &= \frac{1}{2}\sum_m(2\bar{n}_m + 1)\sum_{s,s'}(|\alpha_{s,m} - \alpha_{s',m}|^2)\\ &= \frac{1}{2}\sum_m(2\bar{n}_m + 1)\sum_{s,s'}\left(\left|\sum_j(s_j - s'_j)\alpha_{j,m}(\tau)\right|^2\right)\\ &= \frac{1}{2}\sum_m(2\bar{n}_m + 1)\sum_j|\alpha_{j,m}(\tau)|^2\sum_{s,s'}(s_j - s'_j)^2.\end{aligned} \tag{B19}$$

Since $\sum_{s,s'}s_js'_j = (\sum_s s_j)^2 = 0$, we have $\sum_{s,s'}(s_j - s'_j)^2 = 2d^2$. Thus,

$$\sum_{s,s'} L_{s,s'} = d^2\sum_{j,m}(2\bar{n}_m + 1)|\alpha_{j,m}(\tau)|^2. \tag{B20}$$

Rewriting $K_{s,s'} = \sum_m \mathrm{Im}(\alpha^*_{s',m}\alpha_{s,m}) = \sum_{j,j'}s_js'_{j'}A_{j,j'}$, where $A_{j,j'} = \mathrm{Im}(\sum_m \alpha^*_{j',m}\alpha_{j,m})$, we then compute

$$\sum_{s,s'}(\delta\phi_s - \delta\phi_{s'} + K_{s,s'})^2 = \sum_{s,s'}\left(\sum_{j<j'}\delta\theta_{j,j'}(s_js_{j'} - s'_js'_{j'}) + \sum_{j,j'}A_{j,j'}s_js'_{j'}\right)^2. \tag{B21}$$

Analyzing the terms, cross terms vanish upon summation over s or s'. Noting $\sum_s s_ks_l = d\delta_{kl}$, the first quadratic part contributes $2d^2\sum_{j<j'}\delta\theta^2_{j,j'}$, and the second quadratic part contributes $d^2\sum_{j,j'}A^2_{j,j'}$. Thus,

$$\sum_{s,s'} f_{s,s'} = d^2 - d^2\sum_{j,m}(2\bar{n}_m + 1)|\alpha_{j,m}(\tau)|^2 - \frac{1}{2}\left(2d^2\sum_{j<j'}\delta\theta^2_{j,j'} + d^2\sum_{j,j'}A^2_{j,j'}\right). \tag{B22}$$

Substituting back into the original formula:

$$\begin{aligned}\bar{F} &\approx 1-\frac{d}{d+1}\left(\sum_{j,m}(2\bar{n}_m+1)|\alpha_{j,m}(\tau)|^2+\sum_{j<j'}\delta\theta_{j,j'}^2+\frac{1}{2}\sum_{j,j'}A_{j,j'}^2\right)\\ &=1-\frac{2^n}{2^n+1}\left(\sum_{j,m}(2\bar{n}_m+1)|\alpha_{j,m}(\tau)|^2+\sum_{j<j'}(\Theta_{j,j'}(\tau)-\theta_{j,j'})^2\right.\\ &\quad\left.+\sum_{j<j'}\mathrm{Im}\left(\sum_m\alpha_{j',m}^*(\tau)\alpha_{j,m}(\tau)\right)^2\right),\end{aligned}\tag{B23}$$

where we have used $A_{j,j'}=-A_{j',j}$. Moreover, for $n=2$, the above expression correctly reduces to the original 4/5 coefficient as reported in Ref. [13]. Note that the last term is of order $O(\alpha^4)$. As an approximation, we aim to retain only the leading-order terms for all relevant physical quantities (namely, the residual entanglement and cumulative phase errors). Therefore, we can neglect this last term to arrive at the final result:

$$\bar{F}\approx 1-\frac{2^n}{2^n+1}\left(\sum_{j,m}(2\bar{n}_m+1)|\alpha_{j,m}(\tau)|^2+\sum_{j<j'}(\Theta_{j,j'}(\tau)-\theta_{j,j'})^2\right).\tag{B24}$$

[1] A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, B. G. Elmegreen, *et al.*, Challenges and opportunities in quantum optimization, Nat. Rev. Phys. **6**, 718 (2024).

[2] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[3] Y. Wang, M. Um, J. Zhang, S. An, M. Lyu, J.-N. Zhang, L.-M. Duan, D. Yum, and K. Kim, Single-qubit quantum memory exceeding ten-minute coherence time, Nat. Photon. **11**, 646 (2017).

[4] P. Wang, C.-Y. Luan, M. Qiao, M. Um, J. Zhang, Y. Wang, X. Yuan, M. Gu, J. Zhang, and K. Kim, Single ion qubit with estimated coherence time exceeding one hour, Nat. Commun. **12**, 233 (2021).

[5] H. Haffner, C. Roos, and R. Blatt, Quantum computing with trapped ions, Phys. Rep. **469**, 155 (2008).

[6] R. Blatt and C. F. Roos, Quantum simulations with trapped ions, Nat. Phys. **8**, 277 (2012).

[7] C. Monroe and J. Kim, Scaling the ion trap quantum processor, Science **339**, 1164 (2013).

[8] I. Pogorelov, T. Feldker, C. D. Marciniak, L. Postler, G. Jacob, O. Krieglsteiner, V. Podlesnic, M. Meth, V. Negnevitsky, M. Stadler, B. Höfer, C. Wächter, K. Lakhmanskiy, R. Blatt, P. Schindler, and T. Monz, Compact ion-trap quantum computing demonstrator, PRX Quantum **2**, 020343 (2021).

[9] R. Yao, W.-Q. Lian, Y.-K. Wu, G.-X. Wang, B.-W. Li, Q.-X. Mei, B.-X. Qi, L. Yao, Z.-C. Zhou, L. He, and L.-M. Duan, Experimental realization of a multiqubit quantum memory in a 218-ion chain, Phys. Rev. A **106**, 062617 (2022).

[10] S.-C. Liu, L. Cheng, L.-Y. Peng, and Q. Gong, Logical quantum circuits protected by the Steane code for specific noises in trapped ions, Phys. Rev. Appl. **22**, 034020 (2024).

[11] A. Sørensen and K. Mølmer, Quantum computation with ions in thermal motion, Phys. Rev. Lett. **82**, 1971 (1999).

[12] T. Choi, S. Debnath, T. A. Manning, C. Figgatt, Z.-X. Gong, L.-M. Duan, and C. Monroe, Optimal quantum control of multimode couplings between trapped ion qubits for scalable entanglement, Phys. Rev. Lett. **112**, 190502 (2014).

[13] Y. Wu, S.-T. Wang, and L.-M. Duan, Noise analysis for high-fidelity quantum entangling gates in an anharmonic linear Paul trap, Phys. Rev. A **97**, 062325 (2018).

[14] P. H. Leung, K. A. Landsman, C. Figgatt, N. M. Linke, C. Monroe, and K. R. Brown, Robust 2-qubit gates in a linear ion crystal using a frequency-modulated driving force, Phys. Rev. Lett. **120**, 020501 (2018).

[15] M. Kang, Q. Liang, B. Zhang, S. Huang, Y. Wang, C. Fang, J. Kim, and K. R. Brown, Batch optimization of frequency-modulated pulses for robust two-qubit gates in ion chains, Phys. Rev. Appl. **16**, 024039 (2021).

[16] T. J. Green and M. J. Biercuk, Phase-modulated decoupling and error suppression in qubit-oscillator systems, Phys. Rev. Lett. **114**, 120502 (2015).

[17] A. R. Milne, C. L. Edmunds, C. Hempel, F. Roy, S. Mavadia, and M. J. Biercuk, Phase-modulated entangling gates robust to static and time-varying errors, Phys. Rev. Appl. **13**, 024022 (2020).

[18] S.-A. Guo, Y.-K. Wu, J. Ye, L. Zhang, W.-Q. Lian, R. Yao, Y. Wang, R.-Y. Yan, Y.-J. Yi, Y.-L. Xu, B.-W. Li, Y.-H. Hou, Y.-Z. Xu, W.-X. Guo, C. Zhang, B.-X. Qi, Z.-C. Zhou, L. He, and L.-M. Duan, A site-resolved two-dimensional quantum simulator with hundreds of trapped ions, Nature (London) **630**, 613 (2024).

[19] Y.-H. Hou, Y.-J. Yi, Y.-K. Wu, Y.-Y. Chen, L. Zhang, Y. Wang, Y.-L. Xu, C. Zhang, Q.-X. Mei, H.-X. Yang, J.-Y. Ma, S.-A. Guo, J. Ye, B.-X. Qi, Z.-C. Zhou, P.-Y. Hou, and L.-M. Duan, Individually addressed entangling gates in a two-dimensional ion crystal, Nat. Commun. **15**, 9710 (2024).

[20] D. Schwerdt, L. Peleg, Y. Shapira, N. Priel, Y. Florshaim, A. Gross, A. Zalic, G. Afek, N. Akerman, A. Stern, A. B. Kish, and R. Ozeri, Scalable architecture for trapped-ion quantum computing using rf traps and dynamic optical potentials, Phys. Rev. X **14**, 041017 (2024).

[21] L. Cheng, S.-C. Liu, L.-Y. Peng, and Q. Gong, Crosstalk suppression of parallel gates for fault-tolerant quantum computation with trapped ions via optical tweezers, Phys. Rev. Appl. **22**, 034021 (2024).

[22] Y. Lu, S. Zhang, K. Zhang, W. Chen, Y. Shen, J. Zhang, J.-N. Zhang, and K. Kim, Global entangling gates on arbitrary ion qubits, Nature (London) **572**, 363 (2019).

[23] N. Grzesiak, R. Blümel, K. Wright, K. M. Beck, N. C. Pisenti, M. Li, V. Chaplin, J. M. Amini, S. Debnath, J.-S. Chen, and Y. Nam, Efficient arbitrary simultaneously entangling gates on a trapped-ion quantum computer, Nat. Commun. **11**, 2963 (2020).

[24] P. Baßler, M. Zipper, C. Cedzich, M. Heinrich, P. H. Huber, M. Johanning, and M. Kliesch, Synthesis of and compilation with time-optimal multi-qubit gates, Quantum **7**, 984 (2023).

[25] Y. Shapira, R. Shaniv, T. Manovitz, N. Akerman, L. Peleg, L. Gazit, R. Ozeri, and A. Stern, Theory of robust multiqubit nonadiabatic gates for trapped ions, Phys. Rev. A **101**, 032330 (2020).

[26] L. Peleg, D. Schwerdt, J. Nemirovsky, Y. Shapira, N. Akerman, A. Stern, A. B. Kish, and R. Ozeri, Fast design and scaling of multiqubit gates in large-scale trapped-ion quantum computers, PRX Quantum **7**, 033021 (2026).

[27] F. Liu, G. Tang, L. Duan, and Y. Wu, Performance analysis for crosstalk errors between parallel entangling gates in trapped-ion quantum error correction, Phys. Rev. Appl. **24**, 014032 (2025).

[28] D. Maslov and Y. Nam, Use of global interactions in efficient quantum circuit constructions, New J. Phys. **20**, 033018 (2018).

[29] S. Bravyi, D. Maslov, and Y. Nam, Constant-cost implementations of Clifford operations and multiply-controlled gates using global interactions, Phys. Rev. Lett. **129**, 230501 (2022).

[30] D. Schwerdt, Y. Shapira, T. Manovitz, and R. Ozeri, Comparing two-qubit and multiqubit gates within the toric code, Phys. Rev. A **105**, 022612 (2022).

[31] J. Allcock, J. Bao, J. F. Doriguello, A. Luongo, and M. Santha, Constant-depth circuits for Boolean functions and quantum memory devices using multi-qubit gates, Quantum **8**, 1530 (2024).

[32] J. Nemirovsky, M. Chuchem, and Y. Shapira, Efficient compilation of quantum circuits using multi-qubit gates, Quantum Sci. Technol. **11**, 01LT02 (2026).

[33] J. van de Wetering, Constructing quantum circuits with global gates, New J. Phys. **23**, 043015 (2021).

[34] A. Villoria, H. Basold, and A. Laarman, Optimization and synthesis of quantum circuits with global gates, Quantum Sci. Technol. **11**, 015040 (2026).

[35] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[36] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).

[37] L. Feng, O. Katz, C. Haack, M. Maghrebi, A. V. Gorshkov, Z. Gong, M. Cetina, and C. Monroe, Continuous symmetry breaking in a trapped-ion spin chain, Nature (London) **623**, 713 (2023).

[38] Y.-T. Lee, K. Kumaran, B. Pokharel, A. Scheie, C. L. Sarkis, D. A. Tennant, T. Humble, A. Schleife, A. Kandala, and A. Banerjee, Benchmarking quantum simulation with neutron-scattering experiments, arXiv:2603.15608.

[39] A. D. Corcoles, A. Kandala, A. Javadi-Abhari, D. T. McClure, A. W. Cross, K. Temme, P. D. Nation, M. Steffen, and J. M. Gambetta, Challenges and opportunities of near-term quantum computing systems, Proc. IEEE **108**, 1338 (2020).

[40] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, Proc. Natl. Acad. Sci. USA **115**, 9456 (2018).

[41] K. A. Landsman, Y. Wu, P. H. Leung, D. Zhu, N. M. Linke, K. R. Brown, L. Duan, and C. Monroe, Two-qubit entangling gates within arbitrarily long chains of trapped ions, Phys. Rev. A **100**, 022332 (2019).

[42] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).

[43] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).

[44] M.-S. Li, X.-X. Rao, Z. Wang, P.-F. Lu, F. Zhu, Y. Liu, and L. Luo, Homogeneous linear ion crystal in a hybrid potential, Quantum Inf. Proc. **21**, 65 (2022).

[45] G.-D. Lin, S.-L. Zhu, R. Islam, K. Kim, M.-S. Chang, S. Korenblit, C. Monroe, and L.-M. Duan, Large-scale quantum computation in an anharmonic linear ion trap, Europhys. Lett. **86**, 60004 (2009).

[46] D. S. Johnson, Approximation algorithms for combinatorial problems, J. Comput. Syst. Sci. **9**, 256 (1974).

[47] H. F. Trotter, On the product of semi-groups of operators, Proc. Am. Math. Soc. **10**, 545 (1959).

[48] M. Suzuki, Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, Commun. Math. Phys. **51**, 183 (1976).

[49] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter error with commutator scaling, Phys. Rev. X **11**, 011020 (2021).

[50] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, arXiv:2405.08810.

[51] S.-C. Liu, L. Cheng, G.-Z. Yao, Y.-X. Wang, and L.-Y. Peng, Efficient numerical approach to high-fidelity phase-modulated gates in long chains of trapped ions, Phys. Rev. E **107**, 035304 (2023).

[52] S.-T. Wang, C. Shen, and L.-M. Duan, Quantum computation under micromotion in a planar ion crystal, Sci. Rep. **5**, 8555 (2015).

[53] Y.-K. Wu, Z.-D. Liu, W.-D. Zhao, and L.-M. Duan, High-fidelity entangling gates in a three-dimensional ion crystal under micromotion, Phys. Rev. A **103**, 022419 (2021).