Machine-learning-inspired quantum control in many-body dynamics

Meng-Yun Mao [1,2] Zheng Cheng,[1,2] Liangsheng Li [3] Ning Wu [4,5,*] and Wen-Long You [1,2,†]

[1]*College of Physics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[2]*Key Laboratory of Aerospace Information Materials and Physics (NUAA), MIIT, Nanjing 211106, China*
[3]*National Key Laboratory of Scattering and Radiation, Beijing 100854, China*
[4]*Center for Quantum Technology Research, School of Physics, Beijing Institute of Technology, Beijing 100081, China*
[5]*Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurements (MOE), School of Physics,*
Beijing Institute of Technology, Beijing 100081, China

Achieving precise preparation of quantum many-body states is crucial for the practical implementation of quantum computation and quantum simulation. However, the inherent challenges posed by unavoidable excitations at critical points during quench processes necessitate careful design of control fields. In this work, we introduce a promising and versatile dynamic control neural network tailored to optimize control fields. We address the problem of suppressing defect density and enhancing cat-state fidelity during the passage across the critical point in the quantum Ising model. Our method facilitates seamless transitions between different objective functions by adjusting the optimization strategy. In comparison to gradient-based power-law quench methods, our approach demonstrates significant advantages for both small system sizes and long-term evolutions. We provide a detailed analysis of the specific forms of control fields and summarize common features for experimental implementation. Furthermore, numerical simulations demonstrate the robustness of our proposal against random noise and spin number fluctuations. The optimized defect density and cat-state fidelity exhibit a transition at a critical ratio of the quench duration to the system size, coinciding with the quantum speed limit for quantum evolution.

I. INTRODUCTION

Nowadays it is imperative to achieve precise preparation of quantum many-body states due to the urgent demand for applications in various fields, including quantum computation [1,2] and quantum simulation [3–5]. In particular, quantum metrology, a rapidly advancing field of quantum information science, is currently experiencing a surge of theoretical developments [6–8] and experimental breakthroughs [9–12]. A pivotal focus of a quantum metrology scheme involves the preparation of optimal nonclassical states [13,14]. Through the utilization of inherent quantum properties, such as entanglement [15–17], coherence [18], and discord [19,20], quantum metrology can achieve the sensitivity for estimating unknown parameters that surpasses the constraints imposed by the standard quantum limit of classical strategies [21]. In fact, the enhanced sensitivity may even approach the limits set by the Heisenberg uncertainty principle [22]. To this end, various forms of entangled states have been generated in engineered many-body systems to increase the phase sensitivity [23,24]. However, a comprehensive understanding of the relationship between quantum features and ultimate scaling sensitivity beyond the standard quantum limit remains elusive.

A widely employed method for the theoretical [25–29] and experimental [30–34] preparation of quantum many-body states involves utilizing unitary evolution. This process transforms the initial state, typically the ground state of some simple Hamiltonian $\hat{H}_i$, into the ground state of the target Hamiltonian $\hat{H}_f$. According to the adiabatic theorem, the unitary transformation can be implemented with arbitrary accuracy by changing the Hamiltonian sufficiently slowly [35,36]. Nevertheless, the strict requirement of the adiabatic limit dictates infinitely long evolution time, inevitably resulting in system decoherence and the disappearance of entanglement [37]. In particular, crossing a quantum critical point (QCP) in finite time challenges the adiabatic condition due to the closing of the energy gap, which ultimately results in the formation of excitations [38]. Consequently, a trade-off must be considered between the evolution time and the excitations generated by the system. In this context, tremendous efforts have recently been dedicated to achieving an optimal passage through the nonadiabatic evolution, particularly across the QCP, aiming to minimize unwanted excitations [39,40] or to maximize fidelity [41,42]. The shortcut to adiabaticity (STA) provides a way of finding fast trajectories that connect the initial and final states by manipulating the system's parameters in a nonadiabatic fashion while still obtaining results akin to those of an adiabatic process [43–47]. Given the flexibility in selecting intermediate trajectories, the time-dependent control parameters of a system can be adjusted in various STA protocols [48]. One common method

---
*wunwyz@gmail.com
†wlyou@nuaa.edu.cn

of STA involves introducing counterdiabatic driving to the reference Hamiltonian that effectively compensates for the nonadiabatic behavior. However, implementing this technique in quantum many-body systems remains a challenging task for experimental execution.

Alternatively, a prevalent approach is quantum optimal control [49–53], where time-dependent control parameters of a system are fine-tuned using optimal control theory. Governed by time-energy uncertainty relations, the characteristic timescales during nonadiabatic evolution are encapsulated by the quantum speed limit [54–56], which delves into the minimum time required for quantum states to achieve specific predetermined objectives. This is particularly crucial in the field of quantum information, where rapid dynamics are often advantageous [56–58]. While the quantum optimal control finds widespread applications in various systems, it is of fundamental interest to formulate the control theory in a general framework. Over the past few years, machine-learning technology has become an integral part of the optimization theory and has been proved to be applicable to optimizing the parameters of variational states in a variety of interacting quantum many-body systems [59–64]. In this work, we propose a promising and generalizable dynamic control neural network (DCNN) to optimize the control field passing through two fixed points of the quantum critical systems that holds potential for testing and comparing approximate methods, as well as bearing implications for condensed matter physics. To demonstrate the power of our method, we focus on suppressing the defect density and improving the cat-state fidelity during the passage across the critical point in the quantum Ising model. Compared to the gradient-based power-law quench [65], the method offers significant advantages over small system size and long-term evolution. We also numerically demonstrate that our proposal is robust against the random noise and the spin number fluctuations. The optimized defect density and cat-state fidelity are uncovered to coincide with the quantum speed limit for quantum evolution.

The rest of the paper is organized as follows. In Sec. II we present the gradients of specific observables with respect to the time-dependent control parameter in the many-body systems that can be transformed into free fermions. Additionally, we introduce the DCNN method in this section. Section III is devoted to the application of the DCNN to the suppression of the defect density in the quantum Ising model. A thorough analysis of achieving the optimized control fields is also discussed. The effectiveness of the protocol, particularly under varying conditions such as random noise and fluctuations in the number of spins, is evaluated through comprehensive numerical simulations. Section IV delves into the improvement of the cat-state fidelity in the quantum Ising model. Finally, the conclusion and outlook are given in Sec. V.

II. DYNAMICAL PROTOCOL IN INTEGRABLE MANY-BODY MODELS

A. Quantum Ising-like models

The quantum Ising model has emerged as a prototypical model of quantum many-body systems due to its analytical solvability. It thus serves as a valuable tool for exploring emerging quantum phenomena [66–72] and as a benchmark for evaluating the effectiveness of state-of-the-art approaches and algorithms [73–75]. Furthermore, it is an especially compelling option as a versatile platform for quantum simulation [76–79]. The exact simulation of the Ising model has been implemented on quantum computers [80,81]. Actually, the steps to build quantum circuits at a scale that could provide an advantage for simulating the Ising Hamiltonian follow the same strategy as the analytical solution of the models. Therefore, the method can be extended to other integrable models. To this end, we focus on a class of integrable many-body spin systems that can be mapped into free fermion models. A variety of systems holding quantum phase transitions can be mapped into such kinds of models, including the one-dimensional (1D) quantum Ising model and XY model, which is regarded as one of two canonical quantum critical systems [82], as well as the 1D and two-dimensional (2D) Kitaev models [83,84], etc. These cover a variety of spin models that are of interest in both quantum information science and condensed matter physics.

Consider a closed quantum system which depends on time explicitly through a control field $g(t)$ that can be varied arbitrarily with certain constraints. The control field $g(t)$ is assumed to enter $\hat{H}[g(t)]$ via a term $g(t)\hat{Y}$ with $\hat{Y}$ some time-independent operator. We consider a quench starting from some fixed initial value $g_i = g(-T)$ and ending up with a fixed final value $g_f = g(T)$ within the time interval $t \in [-T, T]$. In this work, we mainly focus on quantum systems that themselves belong to, or can be mapped to, d-dimensional free-fermion models. In general, this family of arbitrary-dimensional time-dependent free-fermion Hamiltonians can be written as

$$\hat{H}(t) = \sum_{\vec{k}} \psi_{\vec{k}}^{\dagger} [\vec{d}_{\vec{k}}(g(t)) \cdot \vec{\sigma}_{\vec{k}}] \psi_{\vec{k}}, \tag{1}$$

where $\vec{\sigma}_{\vec{k}} = (\sigma_{\vec{k}}^x, \sigma_{\vec{k}}^y, \sigma_{\vec{k}}^z)$ are the Pauli matrices acting on the mode $\vec{k}$ and $\psi_{\vec{k}} = (a_{\vec{k}}, b_{\vec{k}})^T$ are certain fermionic operators. The function $\vec{d}_{\vec{k}}(g) = (d_{\vec{k}}^x(g), d_{\vec{k}}^y(g), d_{\vec{k}}^z(g))$ is determined by specific models, and the corresponding norm $\varepsilon_{\vec{k}} = |\vec{d}_{\vec{k}}(g)| = \sqrt{d_{\vec{k}}^x(g)^2 + d_{\vec{k}}^y(g)^2 + d_{\vec{k}}^z(g)^2}$ gives the single-particle dispersion. Equation (1) can be diagonalized directly within each $\vec{k}$ subspace as $\hat{H}(t) = \sum_{\vec{k}} \hat{H}_{\vec{k}}(t)$ with

$$\hat{H}_{\vec{k}}(t) = \varepsilon_{\vec{k}}(\Psi_{\vec{k}}^{\dagger} \Psi_{\vec{k}} - 1), \tag{2}$$

where $\Psi_{\vec{k}} = (A_{\vec{k}}, B_{\vec{k}})^T$ is associated with $\psi_{\vec{k}}$ by certain $\vec{k}$-dependent unitary transformation. Note that both the dispersion $\varepsilon_{\vec{k}}$ and the quasifermion operators $\Psi_{\vec{k}}$ are generally time dependent through the control field $g(t)$. The subground state $|G_{\vec{k}}[g(t)]\rangle$ of mode $\vec{k}$ satisfies

$$A_{\vec{k}}|G_{\vec{k}}[g(t)]\rangle = B_{\vec{k}}|G_{\vec{k}}[g(t)]\rangle = 0. \tag{3}$$

The subexcited state with the same fermion number parity as $G_{\vec{k}}[g(t)]$ is obviously

$$|\bar{G}_{\vec{k}}[g(t)]\rangle = A_{\vec{k}}^{\dagger}(t)B_{\vec{k}}^{\dagger}(t)|G_{\vec{k}}[g(t)]\rangle. \tag{4}$$

We assume that the ground state is always nondegenerate in the process of the quench, and the energy gap is closed

only at the QCP in the thermodynamic limit. Here both the global instantaneous ground state $|G[g(t)]\rangle$ and time evolution operator $U(t, -T)$ are separable,

$$|G[g(t)]\rangle = \prod_{\vec{k}} |G_{\vec{k}}[g(t)]\rangle, \tag{5}$$

$$U(t, -T) = \prod_{\vec{k}} U_{\vec{k}}(t, -T), \tag{6}$$

where $U_{\vec{k}}(t, -T) = \mathcal{T} \exp[i \int_{-T}^{t} d\tau H_{\vec{k}}(\tau)]$ with $\mathcal{T}$ the time-ordered operator. The time-evolved state is given by $|\phi(t)\rangle = U(t, -T)|G(g_i)\rangle$, which is also separable, i.e., $|\phi(t)\rangle = \prod_{\vec{k}} U_{\vec{k}}(t, -T)|G_{\vec{k}}(g_i)\rangle$.

B. Controlling observables and their gradients

Our aim is to minimize or maximize the final expectation value $O(T) \equiv \langle \phi(T)|\hat{O}|\phi(T)\rangle$ of a general observable $\hat{O}$. As will be discussed below, a key quantity utilized in the DCNN is the gradient of $O(T)$ with respect to the control field $g(t)$ [85] (see Appendix A),

$$\frac{\delta O(T)}{\delta g(t)} = 2\Im \langle G(g_i)|\hat{O}(T)\hat{Y}(t)|G(g_i)\rangle, \tag{7}$$

where $\hat{A}(t) = U^{\dagger}(t, -T)\hat{A}U(t, -T)$ is the Heisenberg-picture operator for any operator $\hat{A}$.

We further assume that the control term $\hat{Y}$ can be expressed as a summation over even operators of independent modes

$$\hat{Y} = \sum_{\vec{k}} \hat{Y}_{\vec{k}}. \tag{8}$$

Here we are interested in the following two types of observables, i.e., $\hat{O}_1 = \sum_{\vec{k}} \hat{O}_{1,\vec{k}}$ and $\hat{O}_2 = \prod_{\vec{k}} \hat{O}_{2,\vec{k}}$. The gradient of $O_1(T)$ has been calculated as (see Appendix A for detail)

$$\frac{\delta O_1(T)}{\delta g(t)} = 2\mathrm{Im} \sum_{\vec{k}} \langle \phi_{\vec{k}}(T)|\hat{O}_{1,\vec{k}}|\bar{\phi}_{\vec{k}}(T)\rangle \langle \bar{\phi}_{\vec{k}}(t)|\hat{Y}_{\vec{k}}|\phi_{\vec{k}}(t)\rangle, \tag{9}$$

where $|\bar{\phi}_{\vec{k}}(t)\rangle = U_{\vec{k}}(t, -T)|\bar{G}_{\vec{k}}(g_i)\rangle$ is the evolved subexcited state of mode $\vec{k}$. The gradient of $O_2(T)$ can be calculated in a similar way,

$$\frac{\delta O_2(T)}{\delta g(t)} = 2 \prod_{\vec{p}} \langle \phi_{\vec{p}}(T)|\hat{O}_{2,\vec{p}}|\phi_{\vec{p}}(T)\rangle$$

$$\times \sum_{\vec{k}} \frac{\mathrm{Im}\langle \phi_{\vec{k}}(T)|\hat{O}_{2,\vec{k}}|\bar{\phi}_{\vec{k}}(T)\rangle \langle \bar{\phi}_{\vec{k}}(t)|\hat{Y}_{\vec{k}}|\phi_{\vec{k}}(t)\rangle}{\langle \phi_{\vec{k}}(T)|\hat{O}_{2,\vec{k}}|\phi_{\vec{k}}(T)\rangle}. \tag{10}$$

C. Dynamic control neural network

We then introduce the DCNN method to optimize the observables in many-body dynamics. As an extension of the control neural network [86], the DCNN is a neural network with an adaptive dynamic structure and the flexible optimization strategy. This methodology not only allows us to design more diverse control fields but also facilitates the further minimization of specific observables, a capability not fully realized by other methods, e.g., the direct gradient algorithm based on the power-law quench [65]. To be concrete, the DCNN

possesses a similar architecture to the conventional neural network, which is composed of input, hidden, and output layers. Two adjacent layers are connected by the weights, biases, and activation function. The distinguishing features of the DCNN lie in its dynamic structure of the network and the learning optimization strategy. The dynamic structure of the network can be effectively executed by increasing the number of units in each hidden layer. We opt for two hidden layers and $\tanh(x)$ as the activation function.

The workflow of the DCNN is illustrated in Fig. 1. The neural network has two hidden layers with n_1 units in the first layer and n_2 in the second. The parameters of the neural network can be randomly initialized. For simplicity, the parameters can be initialized by training the neural network based on the empirical form of the control field. We then add δn_1 neurons to the first hidden layers and add δn_2 to the second. The corresponding parameters of the newly added neurons should be randomly initialized based on the average distribution. The input of the DCNN is t and the output is the following specific function:

$$f(t) = W_3^{\mathrm{T}} \tanh .[W_2 \tanh .(W_1 t + B_1) + B_2], \tag{11}$$

where W_i $(i = 1, 2, 3)$ are weight matrices, B_i $(i = 1, 2)$ are bias matrices, and $\tanh .(x)$ is the function acting on each elements of the matrix. The output $f(t)$ undergoes a linear transformation to produce the control fields $g(t)$. Interestingly, increasing the number of neurons within the neural-network-based approach exhibits a conceptual parallel in model space expansion, akin to the methods employed in traditional optimal control that augment the number of optimization parameters [87,88]. Nevertheless, the DCNN adaptively and flexibly explores model space, unlike the traditional optimal control method that enhances it by requiring prior knowledge and specific basis functions. The DCNN employs a pareto-optimization-based strategy, adeptly balancing multiple objectives without compromising any particular target. Due to the flexible optimization strategy of the DCNN, we can impose desired constraints on the form of the control field. In our work, the control field $g(t)$ has fixed values for initial moment $t = -T$ and final moment $t = T$, which can be evaluated using the loss function

$$\mathcal{L} = (g(-T) - r)^2 + (g(T) - s)^2, \tag{12}$$

where r and s are the fixed points of $g(t)$ for $t = -T$ and T, respectively. The other constraint involves the extremum of the final expectation value of a specific observable $O(T)$.

The DCNN is capable of conducting the optimization for multiple objectives, aligning with approaches in quantum optimal control [89]. To be specific, the classical approach to solve a multi-objective optimization problem is to assign a weight q_i to each normalized objective function so that the problem is converted to a single objective problem with a scalar objective function Q [90]. In our approach, the objective function is not necessarily normalized. We set Q to be $Q = q_1 O + q_2 \mathcal{L}$, where q_1 and q_2 are two arbitrary real numbers. Notably, the normalization of the individual objective functions is not imperative, given the adaptability provided by the adjustable coefficients q_i $(i = 1, 2)$. In order to obtain the minimum value of Q, we adopt the gradient descent algorithm
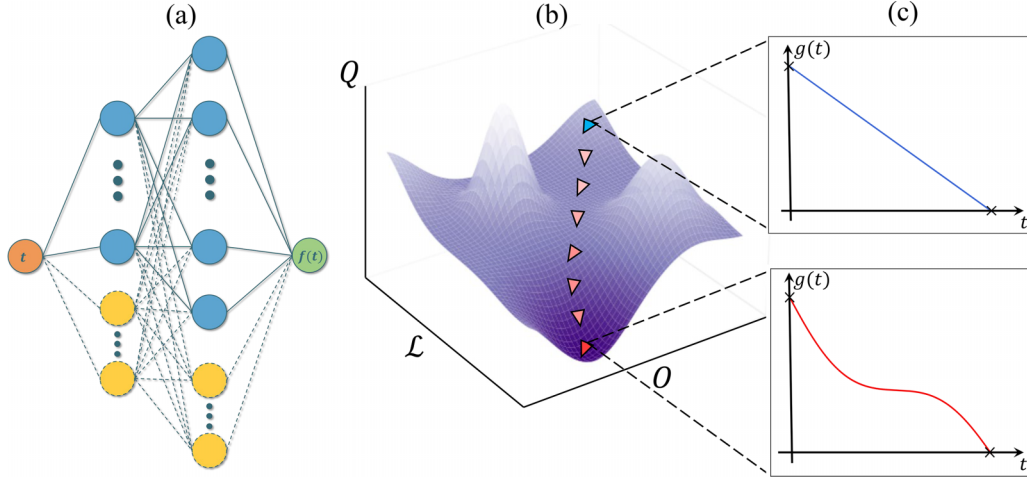
FIG. 1. Workflow of the DCNN and the visualization of the optimization landscape. (a) The structure of the DCNN. We choose two hidden layers with n_1 units in the first layer and n_2 in the second. t is the input of the neural network, and $f(t)$ is the output. Each neuron in the hidden and output layers has a corresponding weight and bias. The blue dots are the original units in the neural network, the yellow dots are the newly increased units, and the dashed lines represent the corresponding increased connection. (b) The optimization of the combined objective function Q for the observable O and the loss $\mathcal{L}$. An initial trial form of the control field is used as a starting point (blue triangle). Q moves downhill (darker red triangles) until the convergence is reached. To be noticed, the downhill movement signifies a simultaneous decline in both O and $\mathcal{L}$. (c) The initial and final forms of the control field $g(t)$ with two fixed points (black crosses).

to update all parameters,

$$W_i \leftarrow W_i + l_i^W \left(q_1 \frac{1}{N} \frac{\partial O(T)}{\partial W_i} + q_2 \frac{\partial \mathcal{L}}{\partial W_i} \right), \quad i = 1, 2, 3,$$

$$B_i \leftarrow B_i + l_i^B \left(q_1 \frac{1}{N} \frac{\partial O(T)}{\partial B_i} + q_2 \frac{\partial \mathcal{L}}{\partial B_i} \right), \quad i = 1, 2, \quad (13)$$

for the next variation. Here the learning rates l_i^W and l_i^B are adjustable parameters depending on the impact of the corresponding parameters on Q.

The gradients of the $O(T)$ and $\mathcal{L}$ with respect to all the parameters are given by

$$\frac{\partial O(T)}{\partial W_i} = \int_{-T}^{T} dt \frac{\delta O(T)}{\delta g(t)} \frac{\partial g(t)}{\partial W_i},$$

$$\frac{\partial O(T)}{\partial B_i} = \int_{-T}^{T} dt \frac{\delta O(T)}{\delta g(t)} \frac{\partial g(t)}{\partial B_i}, \quad (14)$$

and

$$\frac{\partial \mathcal{L}}{\partial W_i} = 2(g(-T) - r) \frac{\partial g(t)}{\partial W_i}\bigg|_{t=-T} + 2(g(T) - r) \frac{\partial g(t)}{\partial W_i}\bigg|_{t=T},$$

$$\frac{\partial \mathcal{L}}{\partial B_i} = 2(g(-T) - r) \frac{\partial g(t)}{\partial B_i}\bigg|_{t=-T} + 2(g(T) - r) \frac{\partial g(t)}{\partial B_i}\bigg|_{t=T}.$$

The gradients of $g(t)$ with respect to its parameters are calculated using the chain rule and are then normalized to mitigate the risk of gradient explosion.

We update only the corresponding parameters of the newly increased neurons according to Eq. (13) while keep the other parameters unchanged. Next, if both the observable O and the loss $\mathcal{L}$, calculated using the updated parameters, are comparatively smaller than those obtained from the original parameters, we will accept the new neurons as part of the neural network, namely, $n_1 \leftarrow n_1 + \delta n_1$ and $n_2 \leftarrow n_2 + \delta n_2$.

Otherwise, the newly introduced units are reinitialized randomly. We can iterate through the aforementioned steps to obtain the desired control field. Adding more neurons into neural network architectures involves directly with the efficiency and scalability of the model, presenting numerical cost as an interesting aspect that warrants further study [91].

While it is challenging to provide a comprehensive overview due to the complicate interplay between model complexity and computational resources, the current insights suggest that computational time and memory requirements tend to scale nearly linearly with the number of parameters (see more details in Appendix B). It is also important to note that, in order to prevent the neural network from getting trapped in potential local minima, we can establish a criterion for the neural network to adjust its architecture after an extended period. Furthermore, we may approve the update even if the observable O or the loss $\mathcal{L}$, calculated using the updated parameters, is slightly higher than those computed with the original parameters.

III. OPTIMAL SUPPRESSION OF DEFECT GENERATION IN THE QUANTUM ISING MODEL

A. Quantum Ising chain and gradient of the observable

In this following, we will concretely apply the DCNN method to physical scenarios, specifically focusing on enhancing the suppression of defect generation and improving the preparation of cat states during the passage across the QCP of the 1D quantum Ising model. For the sake of numerical simplicity, we employ periodic boundary conditions for the spin chain [40,43]. While certain local observables, i.e., local longitudinal magnetization, may differ between periodic and open boundary conditions [92], it is expected that the global observables, such as defect density and cat-state fidelity, will demonstrate analogous behavior irrespective of the boundary

conditions. The Hamiltonian of a periodic quantum Ising chain with $\vec{\sigma}_{N+1} = \vec{\sigma}_1$, is given by

$$\hat{H}_{\text{QIM}} = -\sum_{j=1}^{N} \left[\sigma_j^x \sigma_{j+1}^x + g(t)\sigma_j^z\right], \qquad (15)$$

where $g(t)$ is a global transverse magnetic field that we aim to control, and for simplicity, we assume N to be even. This model exhibits a quantum phase transition at $g_c = 1$ between the ferromagnetic phase for $0 < g < 1$ and the paramagnetic phase for $g > 1$.

The Hamiltonian $\hat{H}_{\text{QIM}}$ can be diagonalized through a standard procedure, which involves applying the Jordan-Wigner transformation followed by the Fourier transformation on the fermionic operators. In this study, we are interested in dynamical protocols starting from the paramagnetic phase with $g_i > 1$, ensuring that only the subspace with an even fermion number parity is pertinent [93], yielding

$$\hat{H}_{\text{QIM}}^{(+)} = \sum_{k \in K} \hat{H}_k,$$

$$\hat{H}_k = 2(c_k^{\dagger}, c_{-k}) \begin{pmatrix} g + \cos k & \sin k \\ \sin k & -g - \cos k \end{pmatrix} \begin{pmatrix} c_k \\ c_{-k}^{\dagger} \end{pmatrix}, \quad (16)$$

where $K = \{\pi/N, 3\pi/N, \ldots, \pi - \pi/N\}$ and the c_k's are fermionic operators satisfying $\{c_k, c_{k'}\} = 0$ and $\{c_k, c_{k'}^{\dagger}\} = \delta_{kk'}$. Note that for open boundary conditions, the necessary adjustment to the momentum set K involves solving the corresponding transcendental equation [94]. The ground and excitation states of the mode Hamiltonian $\hat{H}_k$ are given by

$$|G_k\rangle = \cos \frac{\theta_k}{2} |\text{vac}\rangle_k + \sin \frac{\theta_k}{2} |k, -k\rangle,$$

$$|\bar{G}_k\rangle = \sin \frac{\theta_k}{2} |\text{vac}\rangle_k - \cos \frac{\theta_k}{2} |k, -k\rangle, \qquad (17)$$

where $|\text{vac}\rangle_k$ is the vacuum state of $c_{\pm k}$ and $|k, -k\rangle = c_k^{\dagger} c_{-k}^{\dagger} |\text{vac}\rangle_k$. The angle θ_k is determined by $\tan \theta_k = -\sin k/(g + \cos k)$, where $\Lambda_k(t) = 2\sqrt{g^2(t) + 2g(t)\cos k + 1}$ is the single-particle dispersion with which one has $\hat{H}_k|G_k\rangle = -\Lambda_k|G_k\rangle$ and $\hat{H}_k|\bar{G}_k\rangle = \Lambda_k|\bar{G}_k\rangle$. In the present problem, the operator $\hat{Y} = -\sum_j \sigma_j^z$ gives

$$\hat{Y}_k = -2\sum_{k \in K}(c_k^{\dagger} c_k + c_{-k}^{\dagger} c_{-k} - 1). \qquad (18)$$

Defects or domain walls will be inevitably produced during the transition from the paramagnetic phase to the ferromagnetic phase. The number of defects is measured by the operator

$$\hat{D} = \frac{1}{2}\sum_{j=1}^{N}\left(1 - \sigma_j^x \sigma_{j+1}^x\right) = \sum_{k>0} \hat{D}_k, \qquad (19)$$

where $\hat{D}_k$ is represented by the corresponding matrix form

$$\mathcal{D}_k = 1 + \begin{pmatrix} \cos k & -\sin k \\ -\sin k & -\cos k \end{pmatrix} \qquad (20)$$

in the 2D space spanned by $\{|\text{vac}\rangle_k, |k, -k\rangle\}$. According to Eq. (9), the gradient of $D(T)$ with respect to $g(t)$ is

$$\frac{\delta D(T)}{\delta g(t)} = 4\text{Im} \sum_{k>0, k \in K} \langle \phi_k(T)| \cos k\sigma_k^z - \sin k\sigma_k^x |\bar{\phi}_k(T)\rangle$$

$$\times \langle \bar{\phi}_k(t)|\sigma_z^k|\phi_k(t)\rangle. \qquad (21)$$

B. DCNN guided by the final defect density

We consider a quench from $g(-T) = 2$ to $g(T) = 0$ within the time interval $[-T, T]$. We focus on a power-law quench characterized by the function $g(t) = 1 - |t/T|^r \text{sgn}(t)$, which has been demonstrated to effectively reduce the final defect density, defined as $\rho(T) \equiv D(T)/N$ [65]. Notably, this power-law quench strategy surpasses the conventional linear quench [40], the assisted adiabatic passage [43], and the local adiabatic evolution approach [95] in minimizing $\rho(T)$. To test the DCNN method, we first consider the case of $T = 50$ and $N = 50$. In the initial neural network, there are two hidden layers with $n_1 = 10$ neurons in the first layer and $n_2 = 25$ neurons in the second layer. To improve the learning efficiency of the neural network, we employ pretraining based on a gradient-oriented power-law quench to establish the initial parameters. We then set $\delta n_1 = 5$ and $\delta n_2 = 10$. These values are adaptable and can be increased throughout the learning process. The parameters for the newly added neurons in these layers are randomly initialized, following a uniform distribution ranging from -0.1 to 0.1. We set the learning rates to be $l_i^W = l_i^B = 0.05$ for $i = 1, 2$ and $l_3^W = 0.5$. Due to the significant impact of W_3 on the Q, the learning rate of W_3 is set at a value tenfold higher than that of the other parameters. The weights assigned to Q are $q_1 = 5$ and $q_2 = 1$, and we ensure the normalization of gradients throughout the learning process.

The learning results of the DCNN for $T = 0.5, N = 50$ and $T = 50, N = 50$ are shown in Fig. 2. The number of neurons in the first layer n_1 exhibits a stepwise increase during the learning process. Each step signifies a transformation in the neural network's scale, leading to an enlargement of the model's representational space. During these steps, new nodes are introduced with different initializations. The primary objective of this process is to incrementally decrease the value of $\rho(T)$ through successive modifications of $g(t)$ facilitated by these additional nodes. The interplay between reducing the final defect density and the system's evolution through two fixed points presents an intriguing scenario. In certain cases, reducing $\rho(T)$ may result in the control field $g(t)$ diverging more substantially from these fixed points. To accelerate the convergence rate of the neural networks, we will selectively abandon the optimized form of the control field, although the trade-off could potentially become less pronounced as the neural network's learning process advances. In this context, the DCNN exhibits a pronounced ability to navigate multiple objectives and thereby effectively mitigates oscillatory behaviors between competing objectives, fostering a more rapid convergence. In Fig. 2 we observe an oscillatory decrease in $\rho(T)$ with increasing learning iteration number $\mathcal{N}$, indicating significant modifications to the neural network due to the introduction of additional modes. Note that outliers resulting
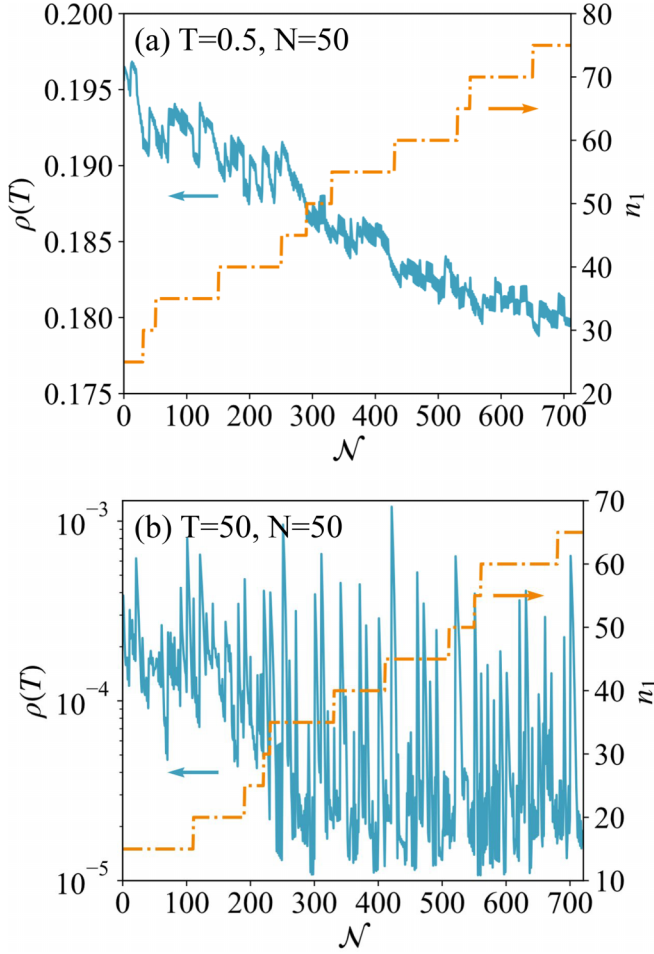
FIG. 2. Learning results of the DCNN for (a) $T = 0.5, N = 50$ and (b) $T = 50, N = 50$. The number of neurons in the first hidden layer n_1 (orange dash-dot line) illustrates the change of the structure of the neural network with the learning iterations $\mathcal{N}$. The blue line shows the specific optimization process of the final defect density $\rho(T)$.

from significant modifications in the learning process should not be a concern, as the optimization strategy automatically discards these values. As the learning process progresses in both scenarios, $\rho(T)$ eventually exhibits convergence with minimal fluctuations. Specifically, $\rho(T)$ converges to a finite value for $T = 0.5$, while for $T = 50$ it stabilizes around a minimal value.

C. Numerical results and discussions

In Fig. 3 we present the optimized control field $g(t)$ obtained by both the DCNN method and the gradient-based power-law quench. In contrast to the power-law control, the resulting $g(t)$ lacks symmetric structure and passes through the QCP at a specific time $t^*(< 0)$. Interestingly, we observe that t^* tends to increase as T increases. For the long-term evolution, it remains imperative to navigate through the QCP as slowly as feasible. Nevertheless, in both instances, the slowest rate of change in $g(t)$ occurs at $t = 0$. These characteristics of $g(t)$ could provide insights into the optimal timing and approach for traversing the critical
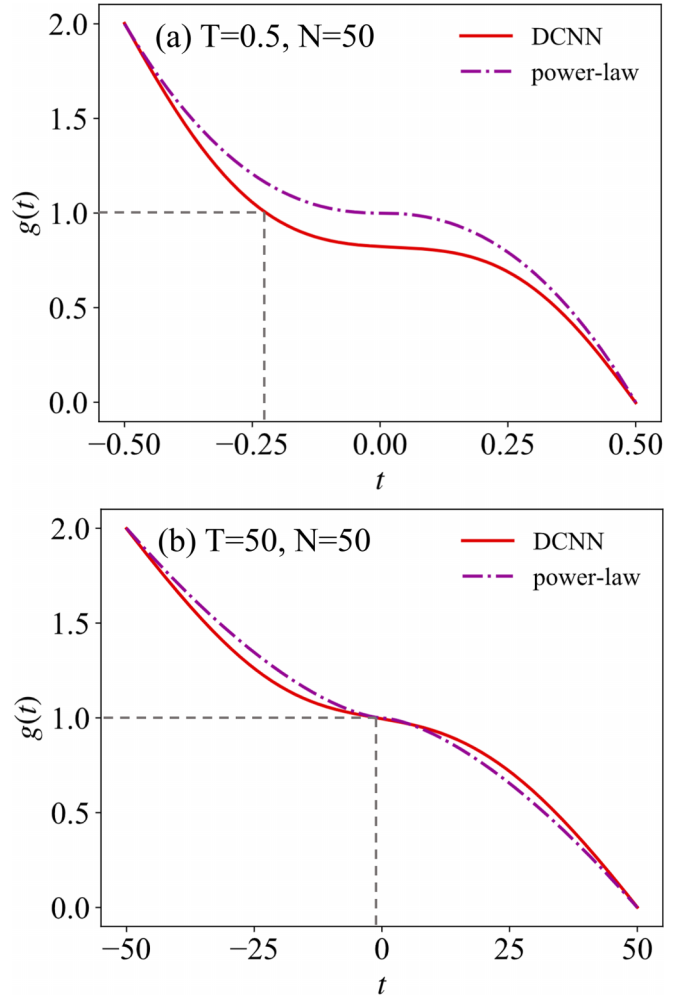

FIG. 3. The comparison between the optimized forms of the control field $g(t)$ based on the DCNN (red solid line) and the gradient-based power-law quench (purple dash-dot line) for (a) $T = 0.5, N = 50$ and (b) $T = 50, N = 50$. In both panels, the horizontal gray dashed lines indicate QCPs, while the gray dashed vertical lines identify t^*.

point, thereby minimizing defect generation in practical pulse design applications. Furthermore, such insights may hold the potential for generalization to state preparation protocols across various quantum critical systems.

The optimized final defect density $\rho(T)$ using the DCNN is plotted in Fig. 4(a) for various system sizes with T/N ranging from 0.01 to 1. To assess the relative effectiveness of different optimization methods, we plot in Fig. 4(b) the ratio R_ρ of the defect density optimized by the DCNN to that achieved through the gradient-based power-law quench. We see that the DCNN outperforms the optimal power-law quench across a broad range of parameters, particularly for small-sized systems and in the long-duration limit. Similar to the case of the optimal power-law profile, we observe a crossover in the scaled time T/N beyond which the optimized defect density undergoes a sharp decline. The simultaneous occurrence of this crossover for different values of T and N suggests that the crossover time T_c is proportional to N, contrasting with the Kibble-Zurek scale for the linear quench,
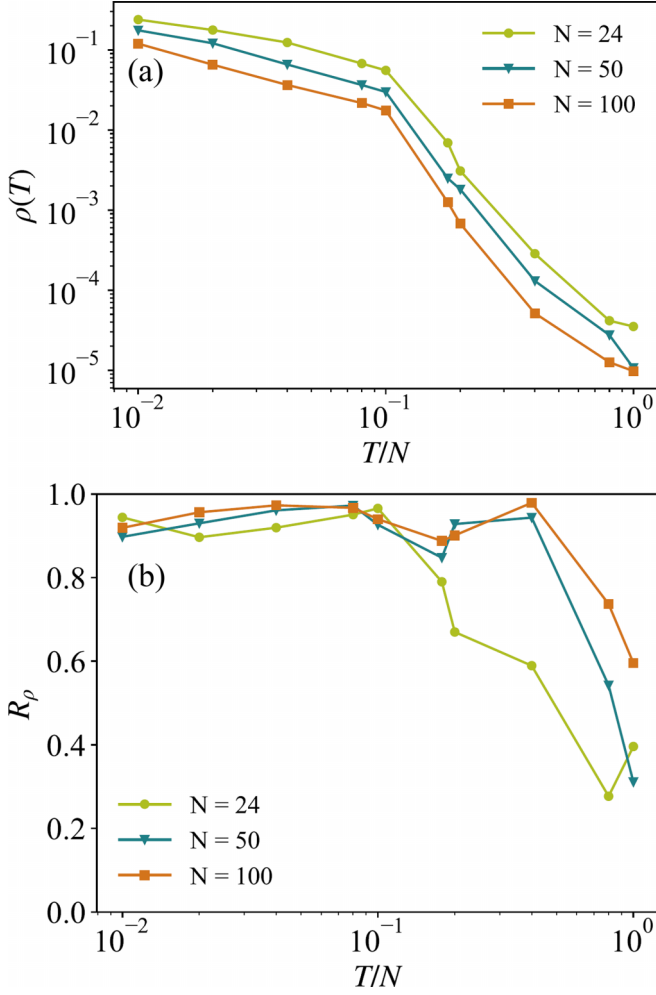
FIG. 4. (a) The final defect density $\rho(T)$ with respect to different T and N under the optimization by the DCNN. (b) The ratio R_ρ of the final defect density optimized by the DCNN relative to that obtained by the gradient-based power-law quench method.
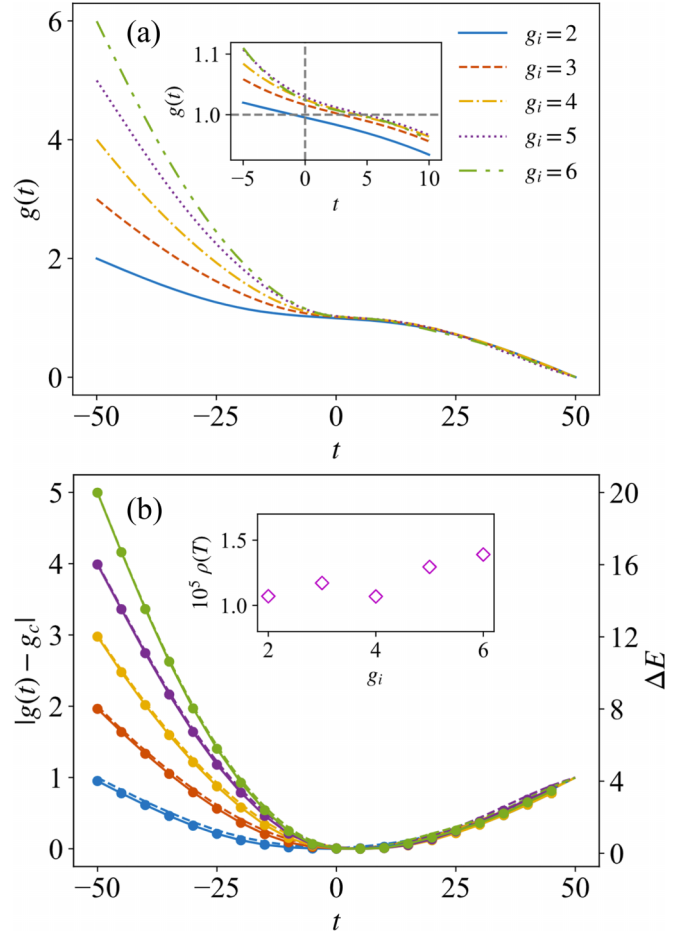

FIG. 5. (a) The optimized control fields $g(t)$ obtained by the DCNN protocol with different initial values g_i for $T = 50, N = 50$. Inset magnifies the control fields around t^*. (b) The absolute value of the relative offset of $g(t)$ to g_c (dashed lines) and the energy gap ΔE (solid lines with markers) with respect to t. Inset shows the corresponding final defect density $\rho(T)$ for different g_i.

where T_{KZ} scales as N^2 [40]. This occurrence can be ascribed to the implications of the quantum speed limit, which establishes a connection between the maximum speed of evolution and the system's energy uncertainty and mean energy. Prior studies suggest that the quantum evolution in a many-body model primarily involves two lowest states in the scenario of small excitations [96–98]. This enables effective mapping into two-level Hamiltonians [99–102]. In the even subspace of k mode in the 1D transverse Ising model, the mode Hamiltonian $\hat{H}_k$ in Eq. (16) can be reformulated in a Landau-Zener-type form as $\hat{H}_k = \gamma_k \sigma_k^z + \omega_k \sigma_k^x$ with $\gamma_k = 2(g + \cos k)$ and $\omega_k = -2 \cos k$. There exists an intrinsic quantum speed limit T_{QSL} to drive a general initial state to a final desired state [51], expressed by

$$\tan 2\omega_k T_{\mathrm{QSL}} \propto \frac{1}{\omega_k}.$$

For the quench with $t \in [-T, T]$ in the present scenario, the lowest mode k_N specified by $\omega_{k_N} \approx -2\pi/N$ yields $2\omega_{k_N} T_{\mathrm{QSL}} \approx -\pi/2$ or $T_{\mathrm{QSL}}/N \approx 1/8$ for sufficiently large N,

serving as a lower bound for T_{QSL} for the entire system, considering all modes k.

The gradient-based power-law quench method, which depends on a specific power-law scheme to pass through fixed points, faces limitations in adaptability across varied initial and final state scenarios. This limitation is particularly pronounced when these states do not exhibit symmetry relative to the critical point, thereby complicating the determination of a suitable power-law function. In contrast, the DCNN offers a more adaptable framework, enabling more straightforward modifications to its optimization strategy. This adaptability suggests that the DCNN approach may provide more nuanced and flexible solutions compared to the gradient-based power-law quench method, especially regarding its application across diverse quenching scenarios. We plot the optimized control fields with different initial values of g_i in Fig. 5(a) for $T = 50$ and $N = 50$. It is evident that $g(t)$ smoothly passes through the QCP at some t^* around $t = 0$ at a relatively slow pace. As g_i increases while keeping g_f constant, mitigating nonadiabatic effects within a finite duration becomes challenging. The optimized control fields strikes a balance between a deliberate

traversal of the QCP and avoiding excessive speed in other regions [39]. To investigate the role of the lowest mode in the optimization process, we plot the absolute value of the relative offset of $g(t)$ to g_c and the energy gap ΔE in Fig. 5(b). One finds a discernible correlation between the deviation of $g(t)$ from g_c and the energy gap associated with the lowest mode, encapsulated by the approximate relation $|g(t) - g_c| \approx 0.25\Delta E$.

To explore potential applications, we assess the robustness of our protocol to spin number fluctuations [50], considering cases with $T = 50, N = 50$ and $T = 0.5, N = 50$ as examples. In experimental implementations, it is inevitable to encounter discrepancies in the actual number of spins present in the system. We thus introduce a number fluctuation parameter, $\delta N = \pm 4$. The final defect density remains below 9.13×10^{-5} for $T = 50, N = 50$ and below 0.1798 for $T = 0.5$, $N = 50$, indicating the robustness of our proposal against spin number fluctuations. To delve deeper into the effects of random noise, we examine the additive white Gaussian noise model, denoted as $\mathcal{A}_G[g(t), \text{SNR}]$. We incorporate it into the control field using the relation

$$g_q(t) = g(t) + \mathcal{A}_G[g(t), \text{SNR}], \tag{22}$$

where q represents each random generator $\mathcal{A}_G[g(t), \text{SNR}]$, with $\text{SNR} = 10 \log_{10}(P_s/P_n)$, where P_s and P_n are the power of signal and noise, respectively. We evaluate our protocol's robustness against white Gaussian noise by analyzing the mean relative offsets, $\delta\rho$, from the ideal defect densities, conducting $\mathcal{R}$ numerical simulations with $\text{SNR} = 10$ for $T = 0.5$, $N = 50$ and $T = 50, N = 50$. Figure 6 depicts $\delta\rho$ versus $\mathcal{R} = 50p$, where p is the simulation index, showing convergence of defect densities to 3.57×10^{-5} and 0.1798 for the respective cases. These results highlight our scheme exhibits a high degree of robustness against random noise.

IV. OPTIMAL CAT-STATE PREPARATION IN THE QUANTUM ISING MODEL

In this section, we employ the DCNN to tackle a more complex task: optimally preparing a cat state in the quantum Ising chain. The Schrödinger cat states have attracted much attention from researchers due to their fundamental implications. The quantum superpositions of two macroscopically distinct states are not only interesting for testing the fundamentals of quantum mechanics such as the quantum-to-classical transition [103,104], but also a valuable resource for quantum information processing [105–107]. Noteworthy examples of maximally entangled cat states include the NOON state [108] and Greenberger-Horne-Zeilinger (GHZ) state [109], which can push the estimation precision to the Heisenberg limit [110–112]. Of particular interest are spin cat states for use in atomic clocks [113] and magnetometers [114] due to the potential applications in diverse fields from materials science to medical diagnostics [115]. However, the realization of cat states in practical experiments faces challenges due to their sensitivity to decoherence and the delicate entanglement properties [116].
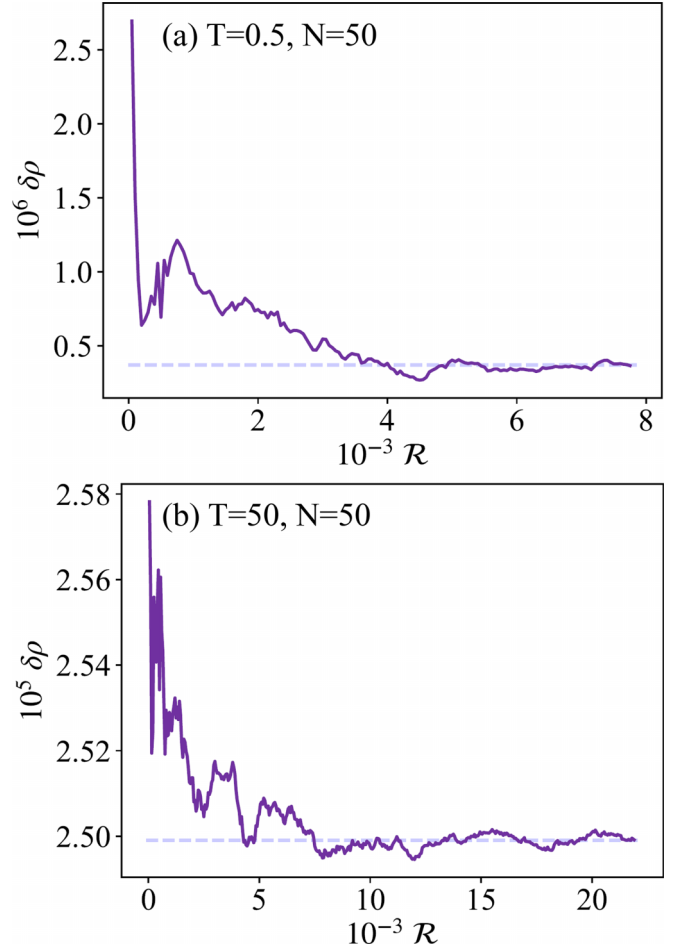

FIG. 6. Mean relative offsets $\delta\rho$ of the final defect density from the ideal value subjected to white Gaussian noise, as a function of the total number of numerical simulations $\mathcal{R}$, with $\text{SNR} = 10$. (a) $T = 0.5$, $N = 50$, and (b) $T = 50$, $N = 50$. Here we define $\mathcal{R} = 50p$, where $p = 155$ for (a) and $p = 438$ for (b). Dashed lines in both panels indicate the convergent values.

The target state we aim to achieve is

$$|\text{cat}\rangle = \frac{1}{\sqrt{2}}(| \rightarrow \cdots \rightarrow \rangle + | \leftarrow \cdots \leftarrow \rangle), \tag{23}$$

which is an equally weighted superposition of the two degenerate fully polarized states $| \rightarrow \cdots \rightarrow \rangle$ and $| \leftarrow \cdots \leftarrow \rangle$. Our goal is to identify an optimized function $g(t)$ that maximizes the final fidelity, defined as

$$F(T) = |\langle \phi(T)|\text{cat}\rangle|^2. \tag{24}$$

It is important to note that maximizing this fidelity poses a more stringent requirement than simply minimizing the final defect density. This is because any linear combination of $| \rightarrow \cdots \rightarrow \rangle$ and $| \leftarrow \cdots \leftarrow \rangle$ would result in vanishing defects.

It is shown in Ref. [93] that $|\text{cat}\rangle$ can be written in the momentum space as a product state

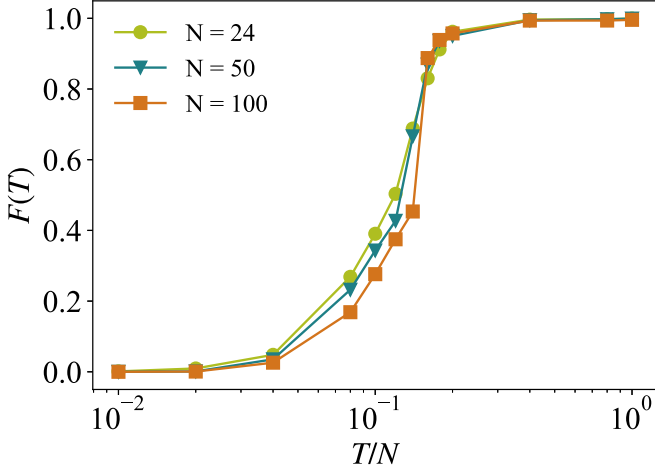$$|\text{cat}\rangle = \prod_{k \in K} |\chi_k\rangle \tag{25}$$

FIG. 7. The fidelity $F(T)$ of the cat state is evaluated under various values of T and N, optimized by the DCNN.

with $|\chi_k\rangle \equiv \sin\frac{k}{2}|\text{vac}\rangle_k + \cos\frac{k}{2}|k, -k\rangle$. Therefore, $F(T)$ is actually the expectation value of the operator $\hat{F} \equiv \prod_{k \in K} |\chi_k\rangle\langle\chi_k|$ in the final state $|\phi(T)\rangle$. From Eq. (10), the dependence of $F(T)$ with respect to $g(t)$ is delineated as

$$\frac{\delta F(T)}{\delta g(t)} = 2\text{Im} \prod_k |\langle\phi_k(T)|\chi_k\rangle|^2$$
$$\times \sum_{k'} \frac{\langle\chi_k|\bar{\phi}_{k'}(T)\rangle\langle\bar{\phi}_{k'}(t)|\hat{Y}_{k'}|\phi_{k'}(t)\rangle}{\langle\chi_k|\phi_{k'}(T)\rangle}. \quad (26)$$

Within the DCNN framework, we also examine the quench from $g(-T) = 2$ to $g(T) = 0$ over the time interval $[-T, T]$. The objective function is $Q = q_1 F(T) + q_2 \mathcal{L}$. The update of all parameters given by Eq. (13) will undergo corresponding adjustments to reflect changes in Q. The optimized fidelity results $F(T)$ obtained through the DCNN for sizes $N = 24, 50, 100$ and the ratio T/N ranging from 0.01 to 1 are presented in Fig. 7. As expected, the achieved fidelity increases as the duration T increases and approaches unity for larger T/N. In the intermediate quench time range, specifically between $0.14 < T/N < 0.18$, a notable surge in $F(T)$ is observed, which is accompanied by a marked decrease in final defect density, as illustrated in Fig. 4. It is noteworthy that this rise in $F(T)$ becomes more pronounced with the increase in system size. For shorter durations, the fidelity always remains extremely low, a consequence of the inherent constraints imposed by the quantum speed limit.

Figure 8 displays the optimized control fields $g(t)$, obtained either by maximizing cat-state fidelity $F(T)$ or by minimizing final defect density $\rho(T)$. In Fig. 8(a) a notable difference is observed between the two control fields for short durations. Specifically, the $g(t)$ that minimizes final defect density leads to considerably lower fidelity. Conversely, for longer durations, the two profiles are nearly identical in Fig. 8(b), with each leading to either very low final defect density or significantly high fidelity. This observation implies a complex relationship between final defect density and state fidelity, indicating that they are not straightforwardly correlated.
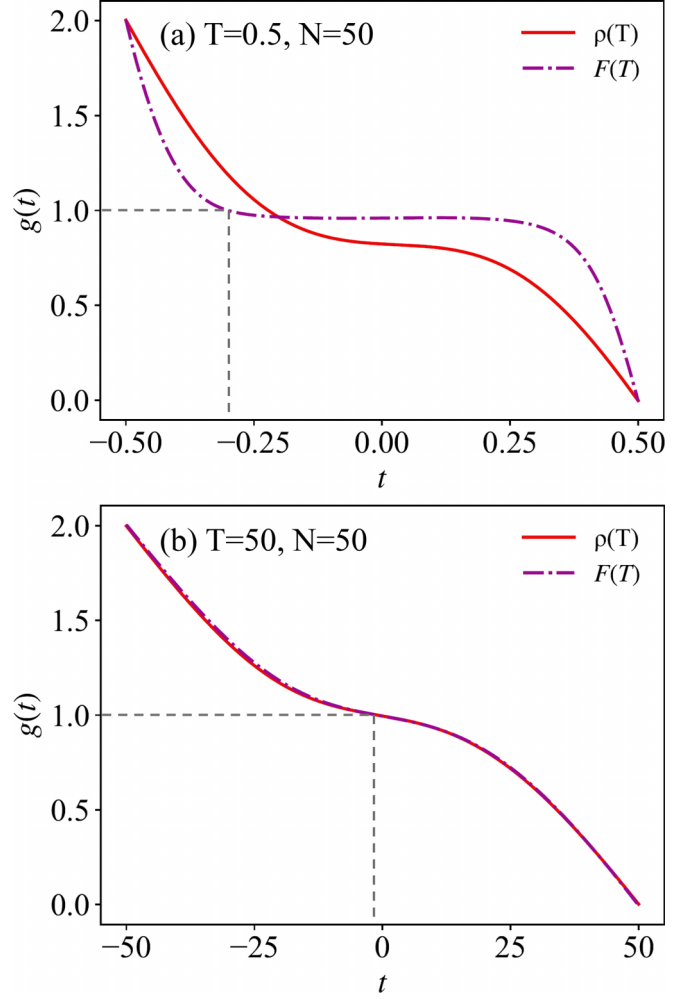

FIG. 8. Comparison of the control fields $g(t)$ optimized for the cat-state fidelity $F(T)$ (purple dash-dot line) and the final defect density $\rho(T)$ (red solid lines) with (a) $T = 0.5, N = 50$ and (b) $T = 50, N = 50$. In both panels, the horizontal gray dashed lines indicate QCPs, and the gray dashed vertical lines identify t^* of $g(t)$ guided by the cat-state fidelity.

V. CONCLUSION AND OUTLOOK

In this study, we introduce a dynamic control neural network (DCNN) approach to suppress the final defect density during a passage across a quantum critical point (QCP). Utilizing the dynamic architecture of the neural network combined with a strategic control mechanism, we successfully generate control fields that navigate through two fixed points while optimizing the desired objective. By considering various evolution time T and system sizes N, our method achieves a lower final defect density compared to traditional gradient-based power-law quench methods. Meanwhile, our study reveals that the optimized final defect density undergoes a crossover at a specific critical ratio of quench duration to system size. This transition closely aligns with the quantum speed limit for evolution, which is profoundly influenced by the lowest mode. We numerically demonstrate that our protocol is robust against the random noise and the spin number fluctuations. Additionally, we demonstrate the effectiveness

of our method using control fields that are optimized across a range of initial values. By analyzing variational control field forms and comparing them to power-law quench methods, we offer insights for experimental pulse design. The versatility of the DCNN is further highlighted by its application to enhancing cat-state fidelity, providing similar optimized solutions indicative of the quantum speed limit. Our findings reveal a non-one-to-one correspondence between final defect density and fidelity. To attain high-quality quantum many-body states, it is crucial to exceed the critical ratio of quench duration to system size.

Noting that the DCNN method introduced in this study extends beyond the preparation of spin cat states, holding promise for a broad range of applications in quantum technologies. With the rapid advancements in quantum computation [117–120] and quantum metrology [121–124], the DCNN's capability for optimizing complex quantum systems positions it as a valuable tool for enhancing precision in quantum information science. The application of the DCNN method to multi-objective optimization in quantum control demonstrates its potential applicability, especially in complex scenarios such as quantum many-body systems and open quantum systems.

ACKNOWLEDGMENTS

APPENDIX A: THE GRADIENTS OF $O(T)$ WITH RESPECT TO $G(T)$

The gradients of the final expectation value $O(T)$ with respect to the control field $g(t)$ can be written as

$$\frac{\delta O(T)}{\delta g(t)} = 2\text{Re}\langle G(g_i)|U^\dagger(T,-T)\hat{O}\frac{\delta U(T,-T)}{\delta g(t)}|G(g_i)\rangle,$$
(A1)

where

$$\frac{\delta U(T,-T)}{\delta g(t)} = -iU(T,-T)U^\dagger(t,-T)\frac{\partial H(t)}{\partial g(t)}U(t,-T)$$
$$= -iU(T,-T)U^\dagger(t,-T)\hat{Y}U(t,-T). \quad \text{(A2)}$$

Combining Eq. (A1) with Eq. (A2), we derive the following expression:

$$\frac{\delta O(T)}{\delta g(t)} = 2\text{Im}\langle G(g_i)|\hat{O}(T)\hat{Y}(t)|G(g_i)\rangle. \quad \text{(A3)}$$

For simplicity, the $|G(g_i)\rangle$ is abbreviated as $|G_i\rangle$. We have assumed that the control term $\hat{Y}$ can be expressed as a summation over even operators of independent mode $\hat{Y} = \sum_{\bar{k}} \hat{Y}_{\bar{k}}$. If $O(T)$ can be represented as a sum of even operators over independent modes, such that $\hat{O}(T) = \sum_{\bar{k}} \hat{O}_{\bar{k}}(T)$, then it follows that
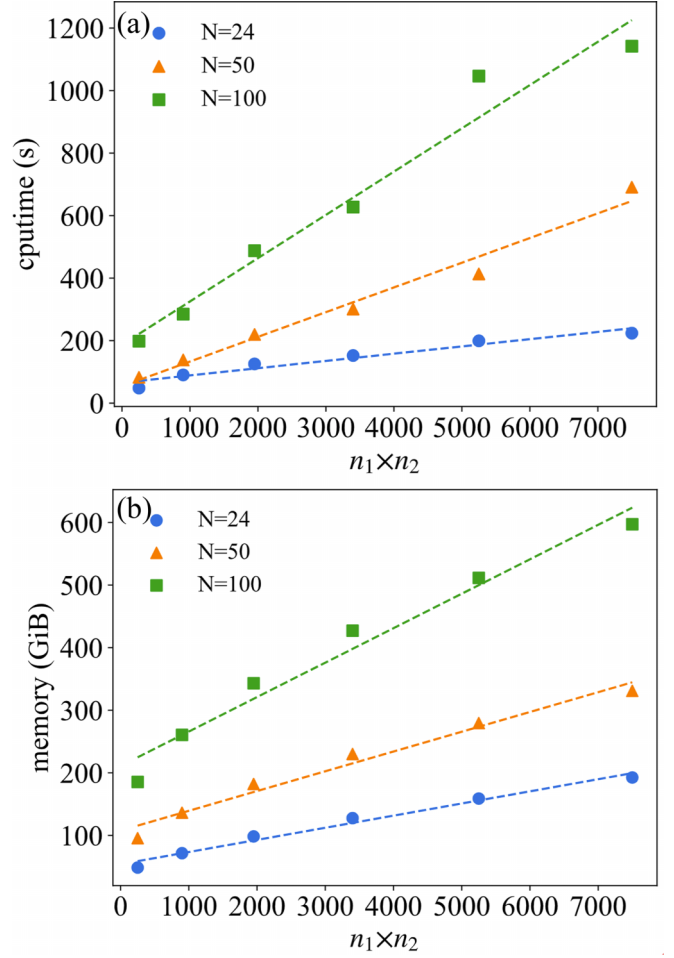

FIG. 9. (a) Computational time and (b) memory requirement for varying neural network scales with spin lengths of $N = 24$, $N = 50$, and $N = 100$. The notation $n_1 \times n_2$ denotes a network configuration with n_1 and n_2 neurons in the first and second hidden layers, respectively. Each point represents the average computational time or memory usage derived from ten iterations of neuron additions to networks of identical scale. The dashed lines indicate linear fits to the data points.

$$\langle G_i|\hat{O}(T)\hat{Y}(t)|G_i\rangle$$
$$= \sum_{\bar{k},\bar{k}'}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle$$
$$= \sum_{\bar{k}'\neq\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle + \sum_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)\hat{Y}_{\bar{k}}(t)|G_{i,\bar{k}}\rangle$$
$$= \sum_{\bar{k},\bar{k}'}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)|G_{i,\bar{k}}\rangle\langle G_{i,\bar{k}'}|\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle$$
$$\quad - \sum_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)|G_{i,\bar{k}}\rangle\langle G_{i,\bar{k}}|\hat{Y}_{\bar{k}}(t)|G_{i,\bar{k}}\rangle$$
$$\quad + \sum_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)\hat{Y}_{\bar{k}}(t)|G_{i,\bar{k}}\rangle$$
$$= \sum_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{\bar{k}}(T)|\bar{G}_{i,\bar{k}}\rangle\langle\bar{G}_{i,\bar{k}}|\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}}\rangle$$
$$= \sum_{\bar{k}}\langle\phi_{\bar{k}}(T)|\hat{O}_{1,\bar{k}}|\bar{\phi}_{\bar{k}}(T)\rangle\langle\bar{\phi}_{\bar{k}}(t)|\hat{Y}_{\bar{k}}|\phi_{\bar{k}}(t)\rangle. \quad \text{(A4)}$$

In deriving the penultimate line of Eq. (A4), we have used the identity

$$\mathbb{I}_{\bar{k}} = |G_{i,\bar{k}}\rangle\langle G_{i,\bar{k}}| + |\bar{G}_{i,\bar{k}}\rangle\langle \bar{G}_{i,\bar{k}}|. \tag{A5}$$

Then, we can obtain Eq. (9). If $O(T)$ can be expressed as a product over even operators of independent modes $\hat{O}(T) = \prod_{\bar{k}} \hat{O}_{\bar{k}}(T)$, it can be calculated as

$$\frac{\delta O_2(T)}{\delta g(t)}$$

$$= 2\mathrm{Im}\sum_{\bar{k}'}\langle G_{i,\bar{k}'}|\hat{O}_{2,\bar{k}'}(T)\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle \prod_{\bar{k}\neq\bar{k}'}\langle G_{i,\bar{k}}|\hat{O}_{2,\bar{k}}(T)|G_{i,\bar{k}}\rangle$$

$$= 2\mathrm{Im}\prod_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{2,\bar{k}}(T)|G_{i,\bar{k}}\rangle\sum_{\bar{k}'}\frac{\langle G_{i,\bar{k}'}|\hat{O}_{2,\bar{k}'}(T)\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle}{\langle G_{i,\bar{k}'}|\hat{O}_{2,\bar{k}'}(T)|G_{i,\bar{k}'}\rangle}$$

$$= 2\mathrm{Im}\prod_{\bar{k}}\langle G_{i,\bar{k}}|\hat{O}_{2,\bar{k}}(T)|G_{i,\bar{k}}\rangle$$

$$\times \sum_{\bar{k}'}\frac{\langle G_{i,\bar{k}'}|\hat{O}_{2,\bar{k}'}(T)|\bar{G}_{i,\bar{k}'}\rangle\langle \bar{G}_{i,\bar{k}'}|\hat{Y}_{\bar{k}'}(t)|G_{i,\bar{k}'}\rangle}{\langle G_{i,\bar{k}'}|\hat{O}_{2,\bar{k}'}(T)|G_{i,\bar{k}'}\rangle}$$

$$= 2\mathrm{Im}\prod_{\bar{k}}\langle \phi_{\bar{k}}(T)|\hat{O}_{2,\bar{k}}|\phi_{\bar{k}}(T)\rangle$$

$$\times \sum_{\bar{k}'}\frac{\langle \phi_{\bar{k}'}(T)|\hat{O}_{2,\bar{k}'}|\bar{\phi}_{\bar{k}'}(T)\rangle\langle \bar{\phi}_{\bar{k}'}(t)|\hat{Y}_{\bar{k}'}|\phi_{\bar{k}'}(t)\rangle}{\langle \phi_{\bar{k}'}(T)|\hat{O}_{2,\bar{k}'}|\phi_{\bar{k}'}(T)\rangle}. \tag{A6}$$

We also use the identity in Eq. (A5) in the above derivation.

APPENDIX B: THE NUMERICAL COST IN NEURAL NETWORK EXPANSION

The parameter scale of expanding a DCNN can be quantified as $O((n_1 + M\delta n_1)(n_2 + M\delta n_2))$, where M represents the number of expanding the neural network as the network approaches convergence. For an intuitive understanding, when the initial network size is substantially smaller than M, the parameter growth simplifies to $O(M^2)$. Conversely, for an initially larger network relative to M, the complexity remains closer to $O(n_1 n_2)$. For simplicity, we only consider the computational time and the memory consumption of increasing the neurons one time, transitioning the network from an $n_1 \times n_2$ to an $(n_1 + \delta n_1) \times (n_2 + \delta n_2)$ structure, where $n_1 \times n_2$ denotes a network with n_1 neurons in the first and n_2 in the second hidden layers, respectively. The overall impact on computational time and memory can be approximated by $\sum_i x_i y_i$, with x_i representing the number of iterations until successful neuron integration, and y_i capturing the associated time or memory costs. It is noteworthy that computational duration is subject to a variety of external factors, including CPU performance and specific elements of the matrix. Assuming a constant spin length, the variation in numerical costs linked to different network sizes is illustrated in Fig. 9. As the parameter scale increases, we observe an almost linear surge in computational time, indicative of the added complexity and data processing requirements. Memory usage similarly rises, though its growth trajectory is shallower compared to computational time, which suggests a higher data throughput and computational complexity for larger networks. An extension in spin length further accentuates these resource demands. However, a quantifiable relationship between network size expansion and computational resource consumption requires further investigation.

[1] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).

[2] X. Chen, B. Zeng, Z.-C. Gu, B. Yoshida, and I. L. Chuang, Phys. Rev. Lett. **102**, 220501 (2009).

[3] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).

[4] A. S. Sørensen, E. Altman, M. Gullans, J. V. Porto, M. D. Lukin, and E. Demler, Phys. Rev. A **81**, 061603(R) (2010).

[5] M. Ljubotina, B. Roos, D. A. Abanin, and M. Serbyn, PRX Quantum **3**, 030343 (2022).

[6] N. Shettell and D. Markham, Phys. Rev. Lett. **124**, 110502 (2020).

[7] I. Frérot and T. Roscilde, Phys. Rev. Lett. **121**, 020402 (2018).

[8] A. Altherr and Y. Yang, Phys. Rev. Lett. **127**, 060501 (2021).

[9] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. **112**, 150802 (2014).

[10] F. Hudelist, J. Kong, C. Liu, J. Jing, Z. Ou, and W. Zhang, Nat. Commun. **5**, 3049 (2014).

[11] C. D. Marciniak, T. Feldker, I. Pogorelov, R. Kaubruegger, D. V. Vasilyev, R. van Bijnen, P. Schindler, P. Zoller, R. Blatt, and T. Monz, Nature (London) **603**, 604 (2022).

[12] P. Yin, X. Zhao, Y. Yang, Y. Guo, W.-H. Zhang, G.-C. Li, Y.-J. Han, B.-H. Liu, J.-S. Xu, G. Chiribella *et al.*, Nat. Phys. **19**, 1122 (2023).

[13] G. Tóth and I. Apellaniz, J. Phys. A: Math. Theor. **47**, 424006 (2014).

[14] S. S. Szigeti, O. Hosten, and S. A. Haine, Appl. Phys. Lett. **118**, 140501 (2021).

[15] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[16] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).

[17] J. Huang, S. Wu, H. Zhong, and C. Lee, *Annual Review of Cold Atoms and Molecules* (World Scientific, Singapore, 2014), Chap. 7, pp. 365–415.

[18] C. Zhang, T. R. Bromley, Y.-F. Huang, H. Cao, W.-M. Lv, B.-H. Liu, C.-F. Li, G.-C. Guo, M. Cianciaruso, and G. Adesso, Phys. Rev. Lett. **123**, 180504 (2019).

[19] P. Giorda and M. G. A. Paris, Phys. Rev. Lett. **105**, 020503 (2010).

[20] H. Ollivier and W. H. Zurek, Phys. Rev. Lett. **88**, 017901 (2001).

[21] If there are N qubits in separable states, it is known that the uncertainty (that is, the inverse of the sensitivity) scales as $O(N^{-1/2})$, which is called the standard quantum limit, while quantum physics allows the ultimate scalings being $O(N^{-1})$, i.e., the Heisenberg scaling, in the absence of noise and $O(N^{-3/4})$ in the presence of realistic decoherence

[22] Y. Chu and J. Cai, Phys. Rev. Lett. **128**, 200501 (2022).

[23] B. Lücke, M. Scherer, J. Kruse, L. Pezzé, F. Deuretzbacher, P. Hyllus, O. Topic, J. Peise, W. Ertmer, J. Arlt *et al.*, Science **334**, 773 (2011).

[24] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzé, A. Smerzi, and M. K. Oberthaler, Science **345**, 424 (2014).

[25] A. Rahmani and C. Chamon, Phys. Rev. Lett. **107**, 016402 (2011).

[26] M. P. Zaletel, A. Kaufman, D. M. Stamper-Kurn, and N. Y. Yao, Phys. Rev. Lett. **126**, 103401 (2021).

[27] U. Farooq, A. Bayat, S. Mancini, and S. Bose, Phys. Rev. B **91**, 134303 (2015).

[28] F. Reiter, D. Reeb, and A. S. Sørensen, Phys. Rev. Lett. **117**, 040501 (2016).

[29] Z.-Y. Wei, D. Malz, and J. I. Cirac, Phys. Rev. Res. **5**, L022037 (2023).

[30] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Nat. Phys. **16**, 205 (2020).

[31] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale *et al.*, Nature (London) **613**, 468 (2023).

[32] J. Léonard, S. Kim, J. Kwan, P. Segura, F. Grusdt, C. Repellin, N. Goldman, and M. Greiner, Nature (London) **619**, 495 (2023).

[33] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, Nature (London) **551**, 579 (2017).

[34] N. S. Maslova, P. I. Arseyev, and V. N. Mantsevich, Sci. Rep. **9**, 3130 (2019).

[35] T. Kato, J. Phys. Soc. Jpn. **5**, 435 (1950).

[36] S. Tabatabaei, H. Haas, W. Rose, B. Yager, M. Piscitelli, P. Sahafi, A. Jordan, P. J. Poole, D. Dalacu, and R. Budakian, Phys. Rev. Appl. **15**, 044043 (2021).

[37] W. Xiang-Bin and M. Keiji, Phys. Rev. Lett. **87**, 097901 (2001).

[38] J. Dziarmaga, Adv. Phys. **59**, 1063 (2010).

[39] R. Barankov and A. Polkovnikov, Phys. Rev. Lett. **101**, 076801 (2008).

[40] J. Dziarmaga, Phys. Rev. Lett. **95**, 245701 (2005).

[41] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[42] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[43] A. del Campo, M. M. Rams, and W. H. Zurek, Phys. Rev. Lett. **109**, 115703 (2012).

[44] J. D. Sau and K. Sengupta, Phys. Rev. B **90**, 104306 (2014).

[45] B. Damski, J. Stat. Mech. (2014) P12019.

[46] S. Deffner, C. Jarzynski, and A. del Campo, Phys. Rev. X **4**, 021013 (2014).

[47] S. Campbell, G. De Chiara, M. Paternostro, G. M. Palma, and R. Fazio, Phys. Rev. Lett. **114**, 177206 (2015).

[48] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Rev. Mod. Phys. **91**, 045001 (2019).

[49] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, EPJ Quantum Tech. **9**, 19 (2022).

[50] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011).

[51] G. C. Hegerfeldt, Phys. Rev. Lett. **111**, 260501 (2013).

[52] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009).

[53] L. Giannelli, S. Sgroi, J. Brown, G. S. Paraoanu, M. Paternostro, E. Paladino, and G. Falci, Phys. Lett. A **434**, 128054 (2022).

[54] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. A **67**, 052109 (2003).

[55] L. B. Levitin and T. Toffoli, Phys. Rev. Lett. **103**, 160502 (2009).

[56] S. Deffner and S. Campbell, J. Phys. A: Math. Theor. **50**, 453001 (2017).

[57] M. Bukov, D. Sels, and A. Polkovnikov, Phys. Rev. X **9**, 011034 (2019).

[58] M. Zhang, H.-M. Yu, and J. Liu, npj Quantum Inf. **9**, 97 (2023).

[59] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[60] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Phys. Rev. Lett. **121**, 167204 (2018).

[61] K. Choo, A. Mezzacapo, and G. Carleo, Nat. Commun. **11**, 2368 (2020).

[62] Y. Zhu, Y.-D. Wu, G. Bai, D.-S. Wang, Y. Wang, and G. Chiribella, Nat. Commun. **13**, 6222 (2022).

[63] G. Carleo, Y. Nomura, and M. Imada, Nat. Commun. **9**, 5322 (2018).

[64] A. G. R. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, Phys. Rev. Lett. **122**, 020601 (2019).

[65] N. Wu, A. Nanduri, and H. Rabitz, Phys. Rev. B **91**, 041115(R) (2015).

[66] X. Mi, M. Ippoliti, C. Quintana, A. Greene, Z. Chen, J. Gross, F. Arute, K. Arya, J. Atalaya *et al.*, Nature (London) **601**, 531 (2022).

[67] P. Frey and S. Rachel, Sci. Adv. **8**, eabm7652 (2022).

[68] I.-C. Chen, B. Burdick, Y. Yao, P. P. Orth, and T. Iadecola, Phys. Rev. Res. **4**, 043027 (2022).

[69] X. Mi, M. Sonner, M. Y. Niu, K. W. Lee, B. Foxen, R. Acharya, I. Aleiner, T. I. Andersen, F. Arute *et al.*, Science **378**, 785 (2022).

[70] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Phys. Rev. B **98**, 155134 (2018).

[71] D. S. Wild, D. Sels, H. Pichler, C. Zanoci, and M. D. Lukin, Phys. Rev. Lett. **127**, 100504 (2021).

[72] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Phys. Rev. Lett. **125**, 260505 (2020).

[73] P. Kairys, A. D. King, I. Ozfidan, K. Boothby, J. Raymond, A. Banerjee, and T. S. Humble, PRX Quantum **1**, 020320 (2020).

[74] B. F. Schiffer, J. Tura, and J. I. Cirac, PRX Quantum **3**, 020347 (2022).

[75] D.-B. Zhang, G.-Q. Zhang, Z.-Y. Xue, S.-L. Zhu, and Z. D. Wang, Phys. Rev. Lett. **127**, 020502 (2021).

[76] B. Kraus, Phys. Rev. Lett. **107**, 250503 (2011).

[77] P. Schauss, Quantum Sci. Technol. **3**, 023001 (2018).

[78] K. Kim, M.-S. Chang, S. Korenblit, R. Islam, E. E. Edwards, J. K. Freericks, G.-D. Lin, L.-M. Duan, and C. Monroe, Nature (London) **465**, 590 (2010).

[79] S. Mostame and R. Schützhold, Phys. Rev. Lett. **101**, 220501 (2008).

[80] A. Cervera-Lierta, Quantum **2**, 114 (2018).

[81] 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 (London) **618**, 500 (2023).

[82] S. Sachdev, Phys. World **12**, 33 (1999).

[83] A. Kitaev, Ann. Phys. **303**, 2 (2003).

[84] N. Wu, Phys. Lett. A **376**, 3530 (2012).

[85] C. Brif, R. Chakrabarti, and H. Rabitz, New J. Phys. **12**, 075008 (2010).

[86] Z. Cheng, M.-J. Lyu, T. Myo, H. Horiuchi, H. Toki, Z.-Z. Ren, M. Isaka, M.-Y. Mao, H. T. Niu Wan *et al.* (unpublished).

[87] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Phys. Rev. A **92**, 062343 (2015).

[88] R. E. Goetz, M. Merkel, A. Karamatskou, R. Santra, and C. P. Koch, Phys. Rev. A **94**, 023420 (2016).

[89] R. Chakrabarti, R. Wu, and H. Rabitz, Phys. Rev. A **78**, 033414 (2008).

[90] I. Giagkiozis and P. Fleming, Info. Sci. **293**, 338 (2015).

[91] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Science **377**, eabk3333 (2022).

[92] M. Białończyk and B. Damski, J. Stat. Mech. (2020) 013108.

[93] N. Wu, Phys. Rev. E **101**, 042108 (2020).

[94] G. G. Cabrera and R. Jullien, Phys. Rev. B **35**, 7062 (1987).

[95] J. Roland and N. J. Cerf, Phys. Rev. A **65**, 042308 (2002).

[96] Y. Wang, Y. Su, X. Chen, and C. Wu, Phys. Rev. Appl. **14**, 044043 (2020).

[97] A. K. Bera, S. M. Yusuf, S. K. Saha, M. Kumar, D. Voneshen, Y. Skourski, and S. A. Zvyagin, Nat. Commun. **13**, 6888 (2022).

[98] Y. Wang, J. Zhang, C. Wu, J. Q. You, and G. Romero, Phys. Rev. A **94**, 012328 (2016).

[99] M. G. Bason, M. Viteau, N. Malossi, P. Huillery, E. Arimondo, D. Ciampini, R. Fazio, V. Giovannetti, R. Mannella, and O. Morsch, Nat. Phys. **8**, 147 (2012).

[100] A. D. Cimmarusti, Z. Yan, B. D. Patterson, L. P. Corcos, L. A. Orozco, and S. Deffner, Phys. Rev. Lett. **114**, 233602 (2015).

[101] X. Li, Sci. Rep. **13**, 14734 (2023).

[102] A. Vepsäläinen, S. Danilin, and G. S. Paraoanu, Sci. Adv. **5**, eaau5999 (2019).

[103] E. Schrödinger, Naturwissenschaften **23**, 807 (1935).

[104] D. J. Wineland, Rev. Mod. Phys. **85**, 1103 (2013).

[105] W. H. Zurek, Nature (London) **412**, 712 (2001).

[106] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang *et al.*, Nature (London) **536**, 441 (2016).

[107] F.-X. Sun, S.-S. Zheng, Y. Xiao, Q. Gong, Q. He, and K. Xia, Phys. Rev. Lett. **127**, 087203 (2021).

[108] J. P. Dowling, Contemp. Phys. **49**, 125 (2008).

[109] J. Huang, H. Huo, M. Zhuang, and C. Lee, Phys. Rev. A **105**, 062456 (2022).

[110] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[111] C. Lee, Phys. Rev. Lett. **97**, 150402 (2006).

[112] L. Pezzé and A. Smerzi, Phys. Rev. Lett. **102**, 100401 (2009).

[113] M. J. Martin, M. Bishof, M. D. Swallows, X. Zhang, C. Benko, J. von Stecher, A. V. Gorshkov, A. M. Rey, and J. Ye, Science **341**, 632 (2013).

[114] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, Phys. Rev. Lett. **104**, 133601 (2010).

[115] D. Le Sage, K. Arai, D. R. Glenn, S. J. DeVience, L. M. Pham, L. Rahn-Lee, M. D. Lukin, A. Yacoby, A. Komeili, and R. L. Walsworth, Nature (London) **496**, 486 (2013).

[116] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer *et al.*, Nature (London) **438**, 639 (2005).

[117] F. Cesa and H. Pichler, Phys. Rev. Lett. **131**, 170601 (2023).

[118] J. Xia, X. Zhang, X. Liu, Y. Zhou, and M. Ezawa, Phys. Rev. Lett. **130**, 106701 (2023).

[119] H. Bombín, C. Dawson, R. V. Mishmash, N. Nickerson, F. Pastawski, and S. Roberts, PRX Quantum **4**, 020303 (2023).

[120] N. Mueller, J. A. Carolan, A. Connelly, Z. Davoudi, E. F. Dumitrescu, and K. Yeter-Aydeniz, PRX Quantum **4**, 030323 (2023).

[121] R. Salvia, M. Mehboudi, and M. Perarnau-Llobet, Phys. Rev. Lett. **130**, 240803 (2023).

[122] S. Zhou, S. Michalakis, and T. Gefen, PRX Quantum **4**, 040305 (2023).

[123] P. Faist, M. P. Woods, V. V. Albert, J. M. Renes, J. Eisert, and J. Preskill, PRX Quantum **4**, 040336 (2023).

[124] Q. Liu, Z. Hu, H. Yuan, and Y. Yang, Phys. Rev. Lett. **130**, 070803 (2023).