# Quantum Flow Matching

Zidong Cui [1,2] Pan Zhang,[3,4] and Ying Tang [1,2,5,6,*]

[1]*Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China*

[2]*School of Physics, University of Electronic Science and Technology of China, Chengdu 611731, China*

[3]*Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

[4]*School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China*

[5]*Key Laboratory of Quantum Physics and Photonic Quantum Information, Ministry of Education, University of Electronic Science and Technology of China, Chengdu 611731, China*

[6]*Non-classical Information Science Basic Discipline Research Center of Sichuan Province, University of Electronic Science and Technology of China, Chengdu 611731, China*

(Received 30 August 2025; accepted 25 June 2026; published 6 August 2026)

The flow matching has rapidly become a dominant paradigm in classical generative modeling, offering an efficient way to interpolate between two complex distributions. We extend this idea to the quantum realm and introduce the quantum flow matching (QFM), a quantum-circuit realization that offers efficient interpolation between two density matrices. QFM offers systematic preparation of density matrices and generation of samples for accurately estimating observables, and can be realized on quantum computers without the need for costly circuit redesigns. We validate its versatility on a set of applications: (1) generating target states with prescribed magnetization and entanglement entropy, (2) estimating nonequilibrium free-energy differences to test the quantum Jarzynski equality, and (3) expediting the study on superdiffusion. These results position QFM as a unifying and promising framework for generative modeling across quantum systems.

DOI: 10.1103/s6zj-vzdp

## I. INTRODUCTION

The density matrix $\rho$ is key to describing finite-temperature quantum systems [1,2]. Preparing a target $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ as an ensemble of pure states and estimating observables from it [3] depend on two critical factors: determining the probability $p_i$ and preparing the pure state $|\psi_i\rangle$. Modern generative models [4] offer a route to capture both factors, including classical neural networks for ground-state search [5,6] and tracking $\rho$'s dynamics [7]; tensor networks for thermal-state simulation [8]; and quantum-classical hybrid models such as $\beta$-variational quantum eigensolver ($\beta$-VQE) [9] that integrate quantum circuits to estimate observables. Yet, these approaches are largely classical, leaving fully quantum protocols less explored. Typical quantum methods approximate $p_i$ via statistical sampling, such as the minimally entangled typical thermal-state (METTS) algorithm [10,11] that employs a Markov chain. However, preparing states in the ensemble remains nontrivial and costly, since each pure state demands a dedicated quantum circuit, leading to repeated circuit adjustments and significant overhead, especially in tasks such as estimating free-energy changes [12] and studying phase transitions [13]. These challenges call for more efficient fully quantum protocols to prepare and track $\rho$.

A quantum generative model that prepares $\rho$ with a single circuit would be an efficient framework aligned well with quantum platforms, potentially bypassing the bottleneck of circuit adjustments. The recent quantum denoising diffusion probabilistic model (QuDDPM) [14] generates state ensembles by evolving from Haar-random states [15,16], analogous to Gaussian noise classically. However, QuDDPM's inverse sampling prepares only the target ensemble, without capturing the time evolution of the target system, limiting its use in scenarios requiring flexible initialization or precise state control [17–19]. In contrast, the classical flow matching [20–22], which learns flows to map between distributions, provides an effective alternative. Extending this idea to the quantum realm may overcome these limitations and broaden applications in quantum dynamics.

In this work, we develop quantum flow matching (QFM). While classical flow matching learns a flow that transports probability distributions, QFM learns a density-matrix propagator implemented by a quantum circuit, under constraints of circuit depth, measurements, and sampling. The propagator can be obtained in a data-driven manner or analytically, depending on the task. As classical flow matching generalizes diffusion models, QFM generalizes QuDDPM by including it as the special case starting from a Haar-random ensemble, and enables the time evolution from any initial ensembles toward target ones. We benchmark QFM against QuDDPM on the topological-state evolution, the entanglement growth, and the transverse-field Ising model (TFIM) (Appendix F), and then apply it to more challenging tasks, including the thermal-state preparation for free-energy estimation [12] and the superdiffusive dynamics [19], both implemented with a

*Contact author: jamestang23@gmail.com

*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.*

TABLE I. Glossary of mathematical terms.

| Notion | Description |
|---|---|
| $\rho_0$ | The density matrix at step 0 |
| $n$ | The number of data qubits |
| $n_a$ | The number of ancilla qubits |
| $\sigma_z^i$ | Pauli operator $Z$ on qubit $i$ |
| $S_\tau$ | The state ensemble at step $\tau$ |
| $M$ | Number of states in one ensemble |
| $\|\psi_\tau^m\rangle$ | A sample state in $S_\tau$ |
| $\mathcal{T}$ | The number of steps |
| $\theta$ | One parameter in $\vec{\theta}_\tau$ |
| $\vec{\theta}_\tau$ | The parameter vector of step $\tau$ |
| $\vec{\theta}_\tau^{\text{op}}$ | The optimized parameter vector |
| $V(\vec{\theta}_\tau)$ | The circuit in step $\tau$ of QFM |
| $L$ | The layer number of EHA in each step |
| $U_{n_a}(\vec{\theta}_\tau)$ | The circuit with ancilla qubits |
| $U_n(\vec{\theta}_\tau)$ | The circuit only acts on data qubits |
| $U^{(1)}$ | Single-qubit gate in EHA |
| $U^{(2)}$ | Two-qubit gate in EHA |
| $U_{(r)}$ | The measurement-induced gate |
| $r_{n_a}$ | The measurement result of the $n_a$th qubit |
| $R_y$ | The rotation operator about the $y$ axis |
| $\|\tilde{\psi}_\tau^m\rangle$ | The generated state in step $\tau$ |
| $\tilde{S}_\tau$ | The generated ensemble in step $\tau$ |
| $M_z$ | Measurement along the $Z$ direction |
| $D(\vec{\theta}_\tau)$ | Loss function |
| $H(t)$ | Time-dependent Hamiltonian |
| $g(t)$ | Transverse field strength |
| $\beta_\tau$ | Inverse temperature |
| $J$ | 1D-interaction strength |
| $J_\perp$ | 2D-interaction strength |
| $K$ | Number of different interaction ratios $J_\perp/J$ |
| $C_{22}^{J_\perp/J}$ | Correlation function with prob qubit 2 |
| $a, b, c$ | The bonds of the Hamiltonian |

fixed circuit. We further provide scaling analysis and bounds on the generalization error of QFM, supporting its wider applicability. For clarity, all mathematical symbols used in this work are summarized in Table I.

## II. QUANTUM FLOW MATCHING

QFM is designed to learn the stepwise evolution of a density matrix, illustrated by the time evolution of ring-shaped state ensembles [Fig. 1(a)]. Starting from a density matrix $\rho_0$ approximated by an ensemble of $M$ sampled pure states $S_0 = \{|\psi_0^1\rangle, \ldots, |\psi_0^M\rangle\}$, QFM evolves each $|\psi_0^m\rangle$ independently over $\mathcal{T}$ steps with a fixed circuit, yielding a final ensemble $S_\mathcal{T}$ that approximates the target density matrix $\rho_\mathcal{T}$. At selected steps [Fig. 1(b)], ancilla qubits are inserted and measured, and the measurement outcomes condition the subsequent circuit evolution, thereby shaping the state distribution and realizing an explicit mapping between two ensembles. Such ensemble-to-ensemble transport, enabled by measurement-conditioned dynamics, is beyond conventional variational or ancilla-assisted circuits mainly used for single-state preparation or observable estimation. Quantum diffusion models such as QuDDPM [14] correspond to the QFM framework when the initial ensemble is a Haar-random ensemble, all evolution steps are implemented through ancilla-assisted measurements, and the circuits are optimized using a fidelity-based loss function.

These measurement-based dynamics (Appendix D) have attracted growing interest, including hybrid reconstruction of diffusion trajectories [23] and Petz-map recovery of ensemble averages [24]. QFM evolves arbitrary initial density matrices within a single reusable circuit architecture [Fig. 1(c)], enabling an ensemble-propagation framework for diverse applications [Fig. 1(d)], while avoiding sample-dependent circuit redesigns.

The circuit of QFM can be either analytically fixed or trainable, depending on target problems (Appendix C). For ensembles of Hamiltonians differing only in local terms, we construct circuits without training, whereas more general

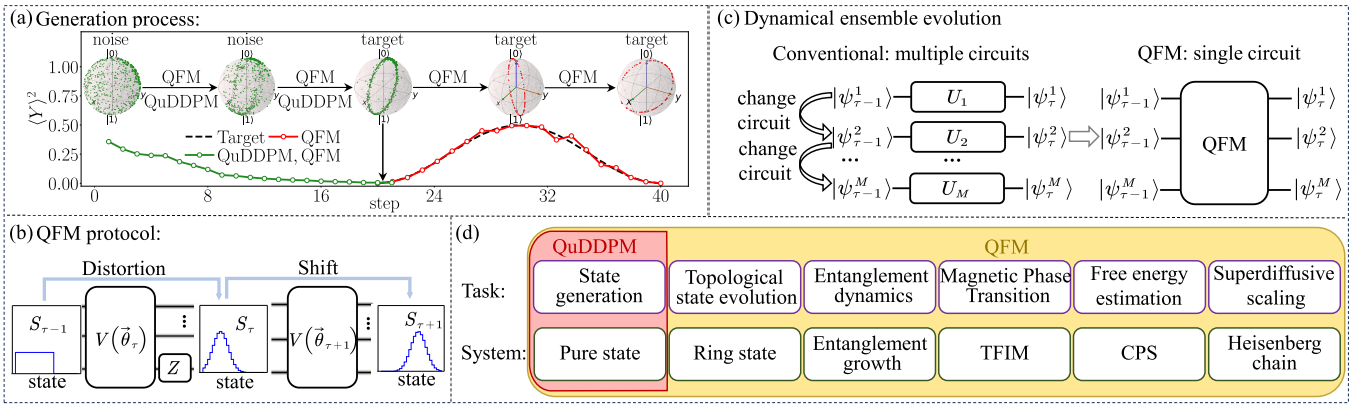


FIG. 1. Schematic of the quantum flow matching. (a) Inspired by classical flow matching [21], QFM learns the evolution between two density matrices. Unlike QuDDPM [14], it does not require starting from a Haar-random ensemble, allowing wider applications. (b) QFM alternates partly measured and unitary circuits to ensure convergence and stability. (c) A single circuit emulates density-matrix dynamics, reducing circuit adjustments; here, $i$ labels the time steps and $M$ denotes the ensemble size. (d) Applications include free-energy estimation and superdiffusive scaling (main text), as well as topological-state evolution, entanglement growth, and magnetic phase transitions (Appendix F).

cases require a training protocol. We denote the circuit at step $\tau$ as $V(\vec{\theta}\tau)$ with parameter vector $\vec{\theta}\tau$. Circuits with $n_a$ ancilla qubits are written as $V(\vec{\theta}\tau) = U_{n_a}(\vec{\theta}\tau)$, while those acting only on $n$ data qubits are $V(\vec{\theta}\tau) = U_n(\vec{\theta}\tau)$. For problems such as superdiffusion, $V(\vec{\theta}\tau)$ can be chosen with analytically structures. By measuring ancilla qubits, QFM generates an ensemble of distinct evolutions with a single circuit.

For problems requiring a trainable circuit, such as estimating the free-energy change, QFM employs at each step an $L$-layer entanglement-varied hardware-efficient ansatz (EHA) [25] with trainable parameters $\vec{\theta}\tau$. Each layer applies gates $U^{(1)} = \prod_{j\in\{x,y,z\}} \exp\{-i\theta_j\sigma_j/2\}$ on each qubit, followed by two-qubit gates $U^{(2)} = \prod_{j\in\{x,y,z\}} \exp\{-i\theta_j\sigma_j \otimes \sigma_j/2\}$ sequentially on all nearest-neighbor pairs, where $\theta_j \in \vec{\theta}\tau$ and $\sigma_j$ are Pauli operators. When $V(\vec{\theta}\tau) = U_n(\vec{\theta}\tau)$, the EHA acts directly on the data qubits. When $V(\vec{\theta}\tau) = U_{n_a}(\vec{\theta}\tau)$, measuring the $n_a$th ancilla qubit with outcome $r_{n_a} \in \{0, 1\}$ is equivalent to inserting, between $U^{(1)}$ and $U^{(2)}$ on its neighboring data qubit in each layer, a single-qubit gate $U_{(r_{n_a})} = \exp\{-i\theta(-1)^{r_{n_a}}\sigma_z/2\}$. If $U^{(2)}$ is allowed to couple each ancilla qubit with data qubits rather than only nearest neighbors, $U_{(r_{n_a})}$ depends on all ancilla measurement outcomes, increasing both expressivity and circuit depth.

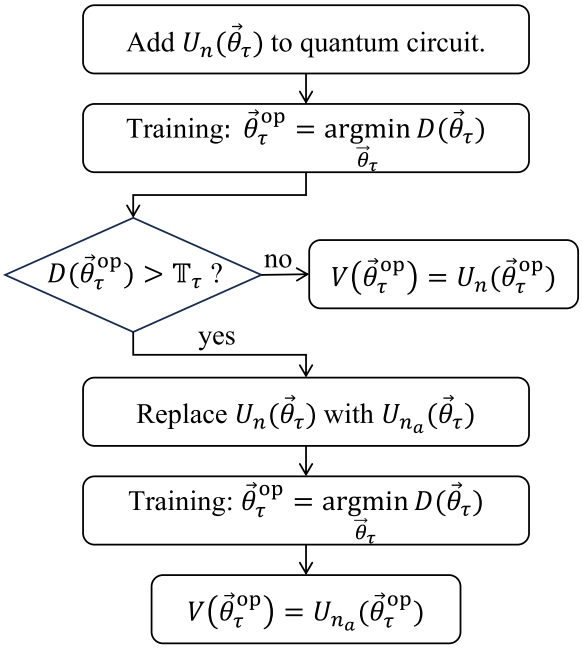


FIG. 2. The training procedure of QFM in step $\tau$. QFM first adopts an $n$-qubit unitary operator for each layer; if the loss for $U_n$ does not converge below the threshold $\mathbb{T}_\tau$ during training, $n_a$ ancilla qubits are introduced and the training is repeated again.

## III. GENERATION AND TRAINING

Starting from an initial state $|\psi_0^m\rangle \in S_0$, QFM sequentially applies $V(\vec{\theta}_1^{\,\mathrm{op}})\cdots V(\vec{\theta}_{\tau-1}^{\,\mathrm{op}})$ to generate $|\widetilde{\psi}_{\tau-1}^m\rangle \in \widetilde{S}_{\tau-1}$, where the tilde indicates results generated by QFM and the superscript of $\vec{\theta}_\tau^{\,\mathrm{op}}$ denotes "optimized." QFM uses a hybrid approach to construct $V(\vec{\theta}_\tau^{\,\mathrm{op}})$, which switches between two types of circuits. The first is a partially measured circuit, where $n_a$ ancilla qubits initialized by rotation operator $R_y(\theta) = \exp(-i\theta\sigma_y/2)$ to get $|\tau\rangle = R_y^{\otimes n_a}(\tau\pi/\mathcal{T})|0\rangle^{\otimes n_a}$ are projectively measured along the $z$ axis after evolving $|\widetilde{\psi}_\tau^m\rangle = U_{n_a}(\vec{\theta}_\tau^{\,\mathrm{op}})|\tau\rangle|\widetilde{\psi}_{\tau-1}^m\rangle \in \widetilde{S}_\tau$. The second is a unitary circuit, where we directly apply the circuit on $n$ data qubits.

For applications where the quantum circuit needs to be constructed through training, QFM begins by training from the first step $V(\vec{\theta}_1^{\,\mathrm{op}})$, and then proceeds step by step up to $V(\vec{\theta}_{\mathcal{T}}^{\,\mathrm{op}})$. At each step, the circuits that have already been trained are fixed. Specifically, after completing the training up to step $\tau-1$, we proceed to train $V(\vec{\theta}_\tau)$ at step $\tau$. We sequentially apply operators from $V(\vec{\theta}_1^{\,\mathrm{op}})$ to $V(\vec{\theta}_{\tau-1}^{\,\mathrm{op}})$, generating the state ensemble $\widetilde{S}_{\tau-1} = \{|\widetilde{\psi}_{\tau-1}^1\rangle, \ldots, |\widetilde{\psi}_{\tau-1}^M\rangle\}$. We then apply $U_n(\vec{\theta}_\tau)$ to states sampled from $\widetilde{S}_{\tau-1}$ and minimize the loss function $D(\vec{\theta}_\tau)$ to get $\vec{\theta}_\tau^{\,\mathrm{op}}$. If $D(\vec{\theta}_\tau^{\,\mathrm{op}})$ is lower than the threshold $\mathbb{T}_\tau$, $U_n(\vec{\theta}_l^{\,\mathrm{op}})$ will be the $\tau$th step of QFM; otherwise, we switch $U_n(\vec{\theta}_\tau^{\,\mathrm{op}})$ to $U_{n_a}(\vec{\theta}_\tau)$ and repeat the same training process with input state $S_\tau^{n_a} = \{|\tau\rangle|\widetilde{\psi}_{\tau-1}^1\rangle, \ldots, |\tau\rangle|\widetilde{\psi}_{\tau-1}^M\rangle\}$ as shown in Fig. 2. To avoid barren plateaus, we initialize parameters in each layer with a neighborhood of radius $\mathcal{O}(1/L)$ [28]. We determine the $\tau$th step parameter $\vec{\theta}_\tau$ by minimizing the loss function, which is generically constructed to quantify the discrepancy of the application-dependent observable statistics between the evolved ensemble at step $\tau$ and the target ensemble:

$$E_m = O_\tau\left(\left|\psi_\tau^m\right\rangle, V(\vec{\theta}_\tau)|\tau\rangle\left|\widetilde{\psi}_{\tau-1}^m\right\rangle\right), \tag{1}$$

$$\vec{\theta}_\tau^{\,\mathrm{op}} = \min_{\vec{\theta}_\tau} D(\vec{\theta}_\tau) = \min_{\vec{\theta}} f_\tau(E_1, E_2, \ldots, E_M), \tag{2}$$

where $O$ and $f$ are the operator and the function depending on the application. In general, the loss is chosen to compare a set of measurable ensemble observables whose expectation values sufficiently characterize the target distribution at each time step, so that matching these observables enforces the desired flow of the ensemble toward the target density matrix. For examples requiring the fidelity estimation, $O$ is the swap-test operator and $f$ is the average function. For the example of entanglement growth, $O$ estimates the entanglement entropy of $|\widetilde{\psi}_\tau^m\rangle$ and $f$ is the mean squared error (MSE). Training introduces an additional overhead in QFM through the optimization of circuit parameters. For applications requiring many ensemble samples, this cost can be distributed over repeated use of the trained circuit.

## IV. APPLICATIONS

We demonstrate the method on representative tasks. The benchmarks comparing QFM with QuDDPM are summarized in the main text, with further details provided in the Appendices A, B, and F. The main text focuses on free-energy estimation (variational circuit) and superdiffusive scaling (fixed circuit).

TABLE II. Hyperparameters and results in three benchmark examples. Here, $n$ is the number of data qubits, $n_a$ is the number of ancilla qubits, $M$ is the number of training states, and $L$ is the number of layers in each step of QFM.

| Task | $n$ | $n_a$ | $L$ | $O$ | $f$ | $\mathcal{T}$ | $M$ | Performace |
|---|---|---|---|---|---|---|---|---|
| Ring-state ensemble evolution | 1 | 0 | 5 | Swap test | Arithmetic mean | 20 | 100 | 0.037 |
| Entanglement growth | [2,3] | 1 | 40 | Entanglement entropy estimation | MSE | 10 | 100 | 0.012 |
| Magnetic phase transition | [2,11] | 1 | [5,32] | Energy estimation | Arithmetic mean | 15 | 100 | 0.035 |

### A. Benchmark tasks

To benchmark QFM against QuDDPM, we consider three representative examples. The ring-state evolution illustrates QFM's capability to map between two quantum state ensembles beyond the Haar-random initialization of QuDDPM. QFM also enables the generation of states with controlled entanglement entropy, the growth from separable to maximally entangled states, and the learning of the evolution from an ordered ferromagnetic phase to a disordered phase, capturing the phase transition. In each example, we evaluate the performance of QFM by the mean Kullback-Leibler (KL) divergence [29] between the generated and target distributions of an observable during the time evolution. The observable depends on the task: measurement results in the $Y$ direction for ring-state evolution, entanglement entropy for entanglement growth, and magnetization for the magnetic phase transition. We consider the KL divergence below 0.05 to indicate reliable distributional matching, which in practice also corresponds to a high state fidelity (99.5%) between the generated and target ensembles. Alternative metrics such as the Hellinger distance [30] or trace distance provide qualitatively consistent results. A summary of the performance is provided in Table II.

### B. QFM facilitates free-energy estimation

Having established the overall performance through benchmark comparisons, we now turn to two central tasks. We first consider QFM for facilitating free-energy estimation. The celebrated Jarzynski equality [31] enables us to calculate the free-energy change between two equilibrium states, from the work in the nonequilibrium process. In quantum simulations, the quantum Jarzynski equality [32–35] is typically tested by computing the average work $\bar{W}$ from evolving METTS [10–12] under a time-dependent Hamiltonian $H(t)$. METTS prepares a thermal density matrix by sampling classical product states (CPS), e.g., $|i\rangle$, via alternating $M_x$ and $M_z$ measurements and projecting each sampled CPS independently toward the thermal state by quantum imaginary time evolution (QITE). In conventional QITE, the circuit must be reconfigured for each sample to achieve thermalization [Fig. 3(a), left], and real-time evolution must be carried out for every thermal sample in free-energy estimation, resulting in an experimental overhead that scales with the number of samples.

QFM generates the METTS ensemble with a single circuit [Fig. 3(a), right]. Starting from $|\psi_0^m\rangle$, we apply $V(\vec{\theta}_1^{\,\mathrm{op}})\cdots V(\vec{\theta}_\tau^{\,\mathrm{op}})$ to obtain the target METTS $|\widetilde{\psi}_\tau^m\rangle$ at inverse temperature $\beta_\tau$, with the next input prepared by measurement. The parameters $\vec{\theta}_\tau^{\,\mathrm{op}}$ are obtained by minimizing the loss function

$$D(\vec{\theta}_\tau) = (1/M)\sum_{m=1}^{M}\left\langle\psi_\tau^m\middle|V(\vec{\theta}_\tau)\middle|\widetilde{\psi}_{\tau-1}^m\right\rangle, \tag{3}$$

where $S_\tau = \{|\psi_\tau^m\rangle = A^{-1}e^{-\beta_\tau H(0)/2}|m\rangle\}$ are training data prepared by QITE on CPS states $|m\rangle$, with $A = \sqrt{\langle\psi_\tau^m|e^{-\beta_\tau H(0)}|\psi_\tau^m\rangle}$, followed by real-time evolution to obtain $|\psi_\tau^m(t)\rangle$ (Appendix E).

For a six-qubit TFIM $H(t) = \sum_i \sigma_z^i\sigma_z^{i+1} + g(t)\sum_i \sigma_x^i$ with $g(t) = 1 + t/20$ ($t = 0 \to 10$), QFM generates METTS ensembles $\widetilde{S}\tau$ at inverse temperatures $\beta\tau = \tau/\mathcal{T}$ with $\mathcal{T} = 10$ and $L = 50$. In Fig. 3(b) ($\beta = 1$), once $\bar{W}$ converges, QFM reduces the total number of circuit adjustments by about 60% compared with Refs. [10,11], with larger savings for faster dynamics. This benefit comes at the cost of a one-time training stage (∼400 circuit updates), after which the same circuit can be reused across ensemble generation.

Performing the training classically further lowers this cost. QFM also reproduces the average work and the analytical free-energy change over temperatures [Fig. 3(c)], and its circuit depth scales nearly linearly with the qubit number [Fig. 3(d)].

### C. Superdiffusion

We next consider superdiffusive scaling in fixed circuits, which provides a complementary setting to the variational scenario above. Quantum simulation was recently performed to study superdiffusion [19]. QFM extends this study to explore how superdiffusion depends on the strength and type of interactions in the 2D Heisenberg model, as a Floquet extension of a 1D system. In the limit of vanishing 2D-interaction strength $J_\perp$ [black box in Fig. 4(a)], the model reduces to an infinite-temperature 1D Heisenberg chain with interaction strength $J$. The first-order Trotterization with $\mathcal{T}$ steps is $e^{-iH_{2\mathrm{D}}(t)} = (e^{-iH_a t/\mathcal{T}}e^{-iH_c t/\mathcal{T}}e^{-iH_b t/\mathcal{T}})^{\mathcal{T}}$, where $H_k = \sum_{\langle i,j\rangle\in k} h_{i,j}$ for $k \in \{a, b, c\}$ denotes the Hamiltonian with bonds of type $k$ and $h_{i,j}(\vec{J}) = (J/4)\sum_{j=x,y,z}\lambda_j\sigma_j^i\sigma_j^j$ with $\vec{\lambda} = (\lambda_x, \lambda_y, \lambda_z)$. The 1D couplings are $\vec{J} = (1, 1, 1)\times J$, while 2D couplings are $\vec{J}_\perp = \vec{\lambda}\times J_\perp$. For the ten-qubit system [Fig. 4(a)], the 1D bonds of type $a$ are $\langle 2, 3\rangle$, $\langle 5, 6\rangle$, $\langle 10, 11\rangle$, of type $c$ are $\langle 3, 4\rangle$, $\langle 6, 7\rangle$, $\langle 8, 9\rangle$, of type $b$ are $\langle 4, 5\rangle$, $\langle 7, 8\rangle$, $\langle 9, 10\rangle$, and the 2D bond is $\langle 4, 9\rangle$ in type $a$.

The conventional simulations on superdiffusion scaling require $K$ circuits, each of which evaluates the correlation function $C_{22}^{J_\perp/J}$ at distinct $J_\perp/J$ values with $M$ random states $|\psi_{R,0}\rangle$. The correlation function can be expressed as a single-

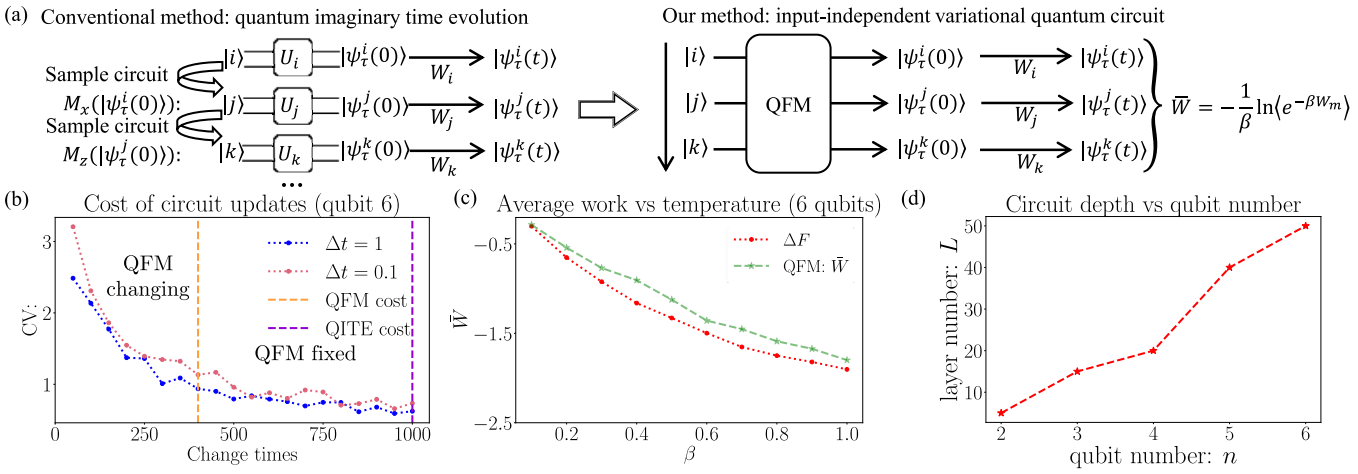


FIG. 3. QFM efficiently estimates nonequilibrium free-energy changes in a six-qubit TFIM. (a) Free energy is estimated via sampling pseudothermal states in METTS. Left: The quantum imaginary time evolution [12,26,27] requires repeated circuit adjustments for each state. Right: QFM generates the ensemble with a fixed circuit after training. (b) For $\beta = 1$, the coefficient of variation under time steps $\Delta t = 1$ (blue) and 0.1 (red) shows that QFM converges with ∼400 circuit updates (orange), compared with ∼1000 for QITE (purple) [12], with a larger gap for shorter $\Delta t$. (c) The averaged work agrees with the free-energy change vs inverse temperature $\beta$. (d) The QFM circuit depth scales nearly linearly with the qubit number.

point observable [36]:

$$C_{pp}(t) = \frac{1}{2M} \sum \langle \psi_{R,p} | \sigma_z^p(t) | \psi_{R,p} \rangle, \tag{4}$$

where $|\psi_{R,p}\rangle = |0\rangle_p |\psi_R\rangle$, $|\psi_R\rangle$ is a Haar-random state, and $p$ is the location of the probe qubit. In the example of Fig. 4, the correlation function is evaluated using 100 states per interaction strength, with $p = 2$ and $M = 100$ for each interaction strength.

In this example, we consider $J_\perp/J \in \{0, 1, 2, 3\}$ and $\mathcal{T} = 20$. The quantum circuit is $U(J_\perp/J) = (U_{1D}(a)U_{2D}(a)U_{1D}(c)U_{1D}(b))^{\mathcal{T}}$, where $U_{1D}(a) = \prod_{a\,\text{type}} \exp\{-ih_{i,j}(\vec{J})\}$ and $U_{2D}(a) = \exp\{-ih_{4,9}(\vec{J})\}$. It requires repeated circuit adjustments that increase the experimental overhead [Fig. 4(b), left]. QFM realizes the same task with a single fixed circuit [Fig. 4(b), right], without a substantial circuit depth increase. It introduces two ancilla qubits [gray nodes in Fig. 4(a)] and adjusts the 2D-interaction strength through measurements, enabling uniform $J_\perp/J$

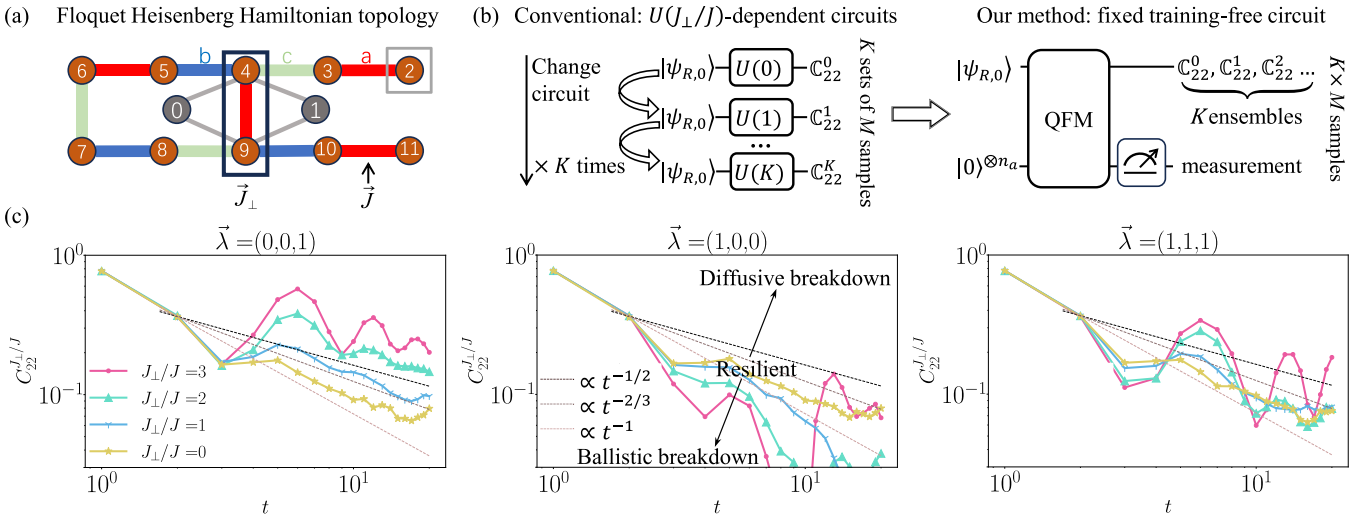


FIG. 4. QFM facilitates the estimation of superdiffusive scaling. (a) The conventional method [19] requires adjusting the quantum circuits for a Heisenberg model under various 2D-interaction strengths (black box). (b) QFM evolves the state ensemble with a fixed circuit and adjusts the 2D-interaction strength by measurements on ancilla qubits [gray node in panel (a)]. (c) The diffusive behaviors by QFM for 2D-interaction strengths $J_\perp/J \in \{0, 1, 2, 3\}$ in three interaction types, matching the conventional method [19]. Increasing $J_\perp/J$ accentuates both diffusive [$\vec{\lambda} = (0, 0, 1)$] and ballistic [$\vec{\lambda} = (1, 0, 0)$] breakdown. In the resilient case [$\vec{\lambda} = (1, 1, 1)$], stronger 2D interactions enlarge the variance of the superdiffusive scaling.

sampling without circuit redesigns. The QFM circuit over steps 1 to $\tau$ is

$$U_{n_a}(\tau) = \left[U_{1D}(a) \times \exp\left\{-i\left[\sigma_z^0 + 0.5\sigma_z^1\right] \otimes h_{4,9}(\vec{J}_\perp)\right\} \times \exp\{-i \times 1.5 h_{4,9}(\vec{J}_\perp)\} U_{1D}(c) U_{1D}(b)\right]^\tau H_{gate}^0 H_{gate}^1, \tag{5}$$

where $H_{gate}^0$ and $H_{gate}^1$ are Hadamard gates that act on qubits 0 and 1 to balance the probability of each measurement result. After measurements, the circuit becomes $U_{n_a}(\tau) = (U_{1D}(a)U_{2D}(a)U_{1D}(c)U_{1D}(b))^\tau$, where

$$U_{2D}(a) = \exp\{-i[(-1)^{r_0} + 0.5(-1)^{r_1}] \otimes h_{4,9}(\vec{J}_\perp)\} \times \exp\{-i \times 1.5 h_{4,9}(\vec{J}_\perp)\}, \tag{6}$$

and $r_i$ ($i \in \{0, 1\}$) is independently and uniformly sampled from $\{0, 1\}$, yielding circuits with $J_\perp/J = 0, 1, 2, 3$.

We show how interaction strengths affect superdiffusion in Fig. 4(c). For $\vec{\lambda} = (0, 0, 1)$, the system exhibits diffusive breakdown at $J_\perp/J \geqslant 2$. For $\vec{\lambda} = (1, 0, 0)$, ballistic transport breaks down at $J_\perp/J \geqslant 1$. The system remains resilient for $\vec{\lambda} = (1, 1, 1)$. Thus, QFM achieves consistent results by a single circuit, eliminating the multiple circuits required by the conventional method [19].

## V. GENERALIZATION ERROR

We analyze the bound on the generalization error [37,38] of QFM for training-based tasks. To learn a unitary $U_\tau \in \mathcal{U}$ at step $\tau$, we have a training set $S_\tau = \{(|\psi_{\tau-1}^m\rangle, U_\tau|\psi_{\tau-1}^m\rangle)\}_{m=1}^M$. The training cost is

$$D(\vec{\theta}_\tau^{op}) = \frac{1}{4M}\sum_{m=1}^{M} \left\| U_\tau |\psi_{\tau-1}^m\rangle\langle\psi_{\tau-1}^m| U_\tau^\dagger - V(\vec{\theta}_\tau^{op})|\widetilde{\psi}_{\tau-1}^m\rangle\langle\widetilde{\psi}_{\tau-1}^m| V(\vec{\theta}_\tau^{op})^\dagger \right\|_1^2. \tag{7}$$

The distance between $V(\vec{\theta}_\tau^{op})$ and target unitary $U_\tau$ gives the expected risk:

$$R_{P_\tau}(\vec{\theta}_\tau^{op}) = \tfrac{1}{4}\mathbb{E}_{|\Psi_{\tau-1}\rangle\sim P}\left[\left\| U_\tau |\Psi_{\tau-1}\rangle\langle\Psi_{\tau-1}| U_\tau^\dagger - V(\vec{\theta}_\tau^{op})|\Psi_{\tau-1}\rangle\langle\Psi_{\tau-1}| V(\vec{\theta}_\tau^{op})^\dagger \right\|_1^2\right], \tag{8}$$

where $P_\tau$ is the test set at step $\tau$. For an $\mathcal{N}$-gate unitary layer $U_n(\vec{\theta}_\tau^{op})$ in QFM, the expected risk $R_P^n(\vec{\theta}_\tau^{op})$ obeys

$$R_P^n(\widetilde{\theta}) \leqslant 2D_\tau(\vec{\theta}_\tau^{op}) + \mathcal{O}\left(\sqrt{\frac{\mathcal{N}\ln(\mathcal{N})}{M}}\right). \tag{9}$$

For partially measured layer $U_{n_a}(\vec{\theta}_\tau^{op})$, the expected risk $R_P^{n_a}(\vec{\theta}_\tau^{op})$ satisfies

$$\mathbb{E}_\tau = \frac{1}{M}\sum_{m=1}^{M} |\langle\widetilde{\psi}_\tau^m| U_\tau |\psi_{\tau-1}^m\rangle|^2, \tag{10}$$

$$D_\tau^{n_a}(\vec{\theta}_\tau^{op}) = 1 - \mathbb{E}_\tau, \tag{11}$$

$$R_P^{n_a}(\vec{\theta}_\tau^{op}) \leqslant 2D_\tau^{n_a}(\vec{\theta}_\tau^{op}) + \mathcal{O}\left(\sqrt{\frac{\mathcal{N}_{eff}\ln(\mathcal{N}_{eff})}{M}}\right), \tag{12}$$

where $\mathcal{N}_{eff}$ is the number of gates in $U_{n_a}$ without ancilla qubits.

## VI. CONCLUSION

We have presented a quantum generative model QFM for transforming between two density matrices and tracking the time evolution of state distributions. Compared with the classical neural-network approaches [4,7,39,40] and quantum-classical models [11,12,41], QFM employs a single quantum circuit and provides more efficient estimations of observables. We have demonstrated QFM's capabilities in learning topological-state evolution, magnetic phase-transition simulation, and entanglement-entropy growth. Especially, QFM can bypass costly circuit adjustments to estimate free-energy changes and superdiffusive scaling, offering a promising approach to track nonequilibrium quantum dynamics with less experimental overhead, where the circuit depth scales nearly linearly with the number of qubits (Appendix F). While current hardware may constrain fault-tolerant implementations on quantum processors, the advances in device performance [42–44] and quantum error correction [45–48] are expected to mitigate these constraints. Future directions include experimental implementations of QFM and its application to facilitate the simulation of a broader class of nonequilibrium quantum dynamics and quantum thermodynamics.

## ACKNOWLEDGMENTS

We acknowledge Lei Wang and Xiaopeng Li for helpful discussions. We thank Yi-Zhuang You for bringing the recent works [23,24] to our attention. This work was supported by Projects No. 12322501, No. 12575035, and No. 12325501 of the National Natural Science Foundation of China, and No. 2026NSFSCZY0124 of the Natural Science Foundation of Sichuan Province. The computational work was supported by the Center for HPC, University of Electronic Science and Technology of China.

## DATA AVAILABILITY

The data that support the findings of this article are openly available [55].

## APPENDIX A: FLOW MATCHING AND GENERATIVE DIFFUSION MODELS

We review two major paradigms in generative modeling—flow matching and diffusion models—and highlight their respective mechanisms and properties, which provide a basis for understanding the QFM and its distinction from quantum diffusion models. The flow matching [21] learns a continuous vector field $v(x, t)$ that transports samples $x_t$ from any source distribution $p_0(x)$ to any target distribution $p_1(x)$ by solving the ordinary differential equation (ODE):

$$\frac{dx_t}{dt} = v(x_t, t), \quad x_0 \sim p_0(x), \tag{A1}$$

with the objective of minimizing the mismatch between the pushed-forward prior and the target distribution along the trajectory.

Differently, generative diffusion models [49,50], described by a stochastic differential equation (SDE), define a forward stochastic process that gradually adds noise to the data and learns a reverse denoising process $\hat{p}(x_{t-1}|x_t)$ from Gaussian

TABLE III. Representative diffusion models and flow matching models.

| | Model/method | Property |
|---|---|---|
| Diffusion | DDPM [49] | Foundational model; noise addition and denoising |
| | Score-based generative models [50] | Uses SDEs + score matching |
| | Denoising diffusion implicit models [51] | Accelerated sampling via implicit reverse process |
| | Latent diffusion [52] | Diffusion in latent space; basis of stable diffusion |
| Flow matching | Flow matching for generative modeling [21] | Sampling via ODE; general FM framework |
| | Consistency models [53] | Refined training objective; faster convergence |
| | Rectified flow++/variants [54] | Systematic FM framework; relation to diffusion |

noise to the target, typically optimized via the variational objective:

$$\mathcal{L}_{\text{diffusion}} = \mathbb{E}_{x_0,\epsilon,t}[\|\epsilon - \epsilon_\theta(x_t, t)\|^2], \tag{A2}$$

where $\epsilon$ is the noise in this step. The key difference lies in the generation mechanism: Diffusion generates samples stochastically, mainly starting from Gaussian noise, and typically requires many denoising steps to handle complex or multimodal distributions. The flow matching is deterministic and can map between arbitrary distributions, enabling direct interpolation, exact likelihood evaluation, and transformation between diverse data types. The flow matching can efficiently transform samples between different image classes, interpolate between embeddings in text or speech models, or morph between distinct multimodal datasets—tasks that diffusion models cannot perform directly. We list in Table III the representative generative diffusion models and flow matching models.

Similarly, in the quantum regime, QFM is designed to offer efficient interpolation between state ensembles of two approximate density matrices, rather than requiring to start from a Haar-random ensemble as in quantum diffusion models, e.g., the quantum denoising diffusion probabilistic model [14].

## APPENDIX B: QUANTUM DENOISING DIFFUSION PROBABILISTIC MODEL

In this Appendix, we give a brief introduction to QuDDPM. Inspired by DDPM, QuDDPM operates entirely within the quantum information processing paradigm and consists of two coupled quantum processes: a forward noising diffusion and a reverse denoising generation. (1) The forward diffusion process starts from a set of $n$-qubit quantum states $\mathcal{S}_0 = \{|\psi_i^{(0)}\rangle\}$ sampled from the target distribution $\mathcal{E}_0$. QuDDPM progressively injects noise by applying random local unitary operations (i.e., quantum scrambling circuits): $|\psi_i^{(k)}\rangle = (\prod_{\ell=1}^{k} U_\ell^{(i)})|\psi_i^{(0)}\rangle, \quad k = 1, 2, \ldots, T$, where $U_\ell^{(i)}$ denotes a randomly chosen local unitary gate applied to the $i$th sample at step $\ell$. After $T \sim \mathcal{O}(n/\ln n)$ steps, the resulting ensemble $\mathcal{S}_T = \{|\psi_i^{(T)}\rangle\}$ becomes indistinguishable from a Haar-random (fully mixed) distribution. (2) The reverse denoising process starts from a noise ensemble $\tilde{\mathcal{S}}_T$ by sampling initial states $|\tilde{\psi}_i^{(T)}\rangle$ and iteratively applies a trainable parametrized quantum circuit (PQC) to reverse the diffusion: $|\tilde{\psi}_i^{(k-1)}\rangle \leftarrow \text{Measure}[\tilde{U}_k(|\tilde{\psi}_i^{(k)}\rangle \otimes |0\rangle^{\otimes n_A})]$, where $\tilde{U}_k$ is a learnable unitary acting on the system qubits and $n_A$ ancillary qubits. A computational-basis measurement on the ancillas projects the output back onto a pure state, enabling a contractive mapping that gradually recovers structure from noise.

### Layerwise training strategy

QuDDPM employs a hierarchical training scheme: At training stage $k$, the layers $\tilde{U}_{k+1}, \ldots, \tilde{U}_T$ are frozen, and only $\tilde{U}_k$ is optimized to minimize the discrepancy between the generated ensemble $\tilde{\mathcal{S}}_k$ and the corresponding noisy ensemble $\mathcal{S}_k$ obtained from the forward process. This decomposes the overall circuit—whose depth would otherwise scale as $\Omega(n)$—into $T$ shallow subcircuits of depth $\mathcal{O}(\ln n)$, thereby mitigating barren plateaus and enabling stable optimization. The loss function of QuDDPM relies on distance measures between the generated and target ensembles. Two primary metrics are used. (1) Maximum mean discrepancy (MMD) with fidelity kernel $F(|\phi\rangle, |\psi\rangle) = |\langle\phi|\psi\rangle|^2$: $D_{\text{MMD}}(\mathcal{E}_1, \mathcal{E}_2) = \bar{F}(\mathcal{E}_1, \mathcal{E}_1) + \bar{F}(\mathcal{E}_2, \mathcal{E}_2) - 2\bar{F}(\mathcal{E}_1, \mathcal{E}_2)$, where $\bar{F}(\mathcal{E}_1, \mathcal{E}_2) = \mathbb{E}_{|\phi\rangle\sim\mathcal{E}_1, |\psi\rangle\sim\mathcal{E}_2}[|\langle\phi|\psi\rangle|^2]$. This expectation can be estimated experimentally via the swap test. (2) Wasserstein distance: For data with nontrivial geometric or topological structure, we adopt the optimal transport-based Wasserstein distance of order $p$: $W_p(\mathcal{E}_1, \mathcal{E}_2) = (\inf_{\pi\in\Pi(\mathcal{E}_1,\mathcal{E}_2)} \int |\langle\phi|\psi\rangle|^2 \, d\pi(|\phi\rangle, |\psi\rangle))^{1/p}$, where $\Pi(\mathcal{E}_1, \mathcal{E}_2)$ denotes the set of all joint distributions with marginals $\mathcal{E}_1$ and $\mathcal{E}_2$.

## APPENDIX C: CIRCUIT STRUCTURE OF QUANTUM FLOW MATCHING

In this Appendix, we describe circuit designs of QFM, including the entanglement-varied hardware-efficient ansatz (EHA) [25] for training tasks and specialized training-free circuits for simulating the 2D Heisenberg model under superdiffusive scaling. We explain how ancilla measurements influence the circuit behavior and illustrate with the superdiffusive scaling how measurements on ancilla qubits can effectively modify the 2D-interaction strength of the simulated system. This provides a clear understanding of both training-based and training-free QFM circuit design principles.

### 1. Circuit for variational quantum circuit

In this Appendix, we describe the $L$-layer EHA circuit in QFM for training tasks, highlighting its structure with and without ancilla qubits. The EHA circuit employs variational entanglers composed of parametrized $XX$, $YY$, and $ZZ$ gates. This design allows the circuit to rapidly adjust the

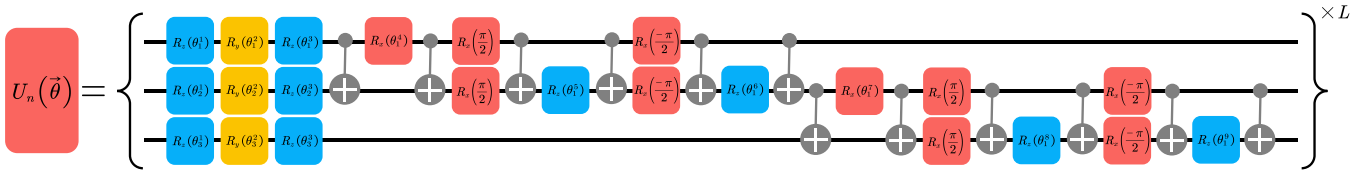

FIG. 5. The unitary circuit $U_n(\vec{\theta})$ in QFM. For steps without ancilla qubits in QFM, we apply the EHA circuit on data qubits. EHA includes layers of $X$ and $Y$ single-qubit rotations followed by nearest-neighbor gates $XX$, $YY$, and $ZZ$, which provide tunable entanglement in the $x$, $y$, and $z$ directions.

amount of entanglement to match the target ground state, thereby overcoming limitations such as barren plateaus and poor trainability. Compared with widely used ansätze, including hardware-efficient ansätze [56], Hamiltonian variational ansatz [57], hardware symmetry-preserving ansatz [58], and chemically inspired ansatzes such as unitary coupled cluster with all single and double excitations [59] and Givens rotations with all single and double excitations [60], EHA consistently achieves higher accuracy and greater robustness across both quantum many-body systems and molecular simulations. Importantly, EHA is problem-agnostic, making it broadly applicable without requiring specific prior knowledge of the Hamiltonian or tailored initial states.

Each layer of the EHA circuit includes arbitrary single-qubit rotations $U^{(1)} = \prod_{j\in\{x,y,z\}} \exp\{-i\theta_j\sigma_j/2\}$ and nearest-neighbor two-qubit gates $U^{(2)} = \prod_{j\in\{x,y,z\}} \exp\{-i\theta_j\sigma_j \otimes \sigma_j/2\}$, which we refer to as the $XX$, $YY$, and $ZZ$ gates depending on the choice of $j$. For intuitive understanding, we present in Fig. 5 the $L$-layer circuit structure of $U_n(\vec{\theta})$ for a three-qubit example, where each layer consists of arbitrary single-qubit gates and nearest-neighbor entangling gates along the $X$, $Y$, and $Z$ directions. All $\theta$ parameters in the figure are trainable.

In Fig. 6, we illustrate how two ancilla qubits are coupled to three data qubits. After applying $U_n(\vec{\theta})$ on the data qubits, we first apply the single-qubit gates from the EHA circuit to each ancilla qubit. Then, on the ancilla qubit adjacent to the data qubits, we sequentially apply the $XX$, $YY$, and $ZZ$ gates. The remaining ancilla qubits are similarly coupled sequentially to their nearest-neighbor ancilla qubits to construct $U_{n_a}$. At the end of $U_{n_a}$, each ancilla qubit is measured in the Pauli-$Z$ basis.

### 2. Circuit for 2D Heisenberg model

In this Appendix, we focus on the structure of QFM circuits in training-free tasks, describing how the 1D- and 2D-interaction components, along with ancilla qubits, are arranged to simulate the Heisenberg model and control interaction strengths without training. The 1D isotropic Heisenberg chain, an integrable model, displays superdiffusive spin transport governed by Kardar-Parisi-Zhang (KPZ) universality. In real materials [61] and simulators, however, integrability-breaking interactions suppress this behavior, leading to a crossover to conventional diffusion or ballistic transport. Resolving this breakdown is key to connecting idealized models with quantum materials.

We employ QFM to probe how 2D interactions, native to scalable hardware architectures like IBM's heavy-hex lattice, drive the suppression of superdiffusion. By tuning the strength and symmetry of these 2D couplings, we elucidate their role in destabilizing KPZ scaling. QFM employs a training-free circuit to simulate the 2D Heisenberg model with varying interaction strengths. By coupling ancilla qubits, the 2D interactions are adjusted based on their measurement outcomes. The circuit consists of a 1D-interaction part $U_{1D}$ and a 2D-interaction part $U_{2D}$.

We employ measurements on the ancilla qubits in QFM to modify the 2D-interaction strength of the simulated Heisenberg system on data qubits. For clarity, we decompose the evolution operator into interaction operators along different bonds and dimensions. Since these interactions are local and mutually commuting, the first-order Trotter circuit for simulating the 2D Heisenberg model is

$$U(J_\perp/J) = (U_{1D}(a)U_{2D}(a)U_{1D}(c)U_{1D}(b))^{\mathcal{T}}. \tag{C1}$$

In $U_{1D}$, the Hamiltonian consists of a series of two-local terms $h_{i,j}$, each acting on qubits $i$ and $j$. Since all these two-local terms mutually commute, the evolution under $U_{1D}$ can be factorized such that each term is implemented independently. Specifically, each two-local term $h_{i,j}$ can be represented by a corresponding evolution operator $e^{-ih_{i,j}t}$, allowing the full

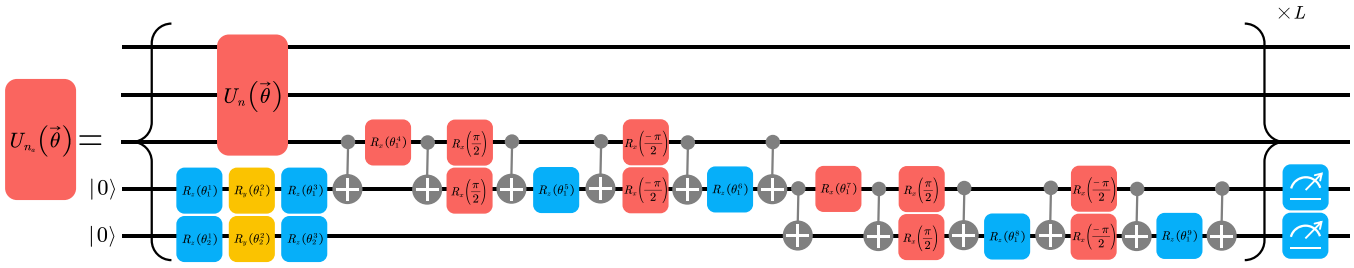

FIG. 6. The circuit structure of steps with $n_a$ ancilla qubits in QFM. For $U_{n_a}$, we add $n_a$ ancilla qubits, which are coupled to the data qubits through the same single- and two-qubit gates as in the EHA circuit, and measure the ancilla qubits with the Pauli-$Z$ operator at the end of each step.

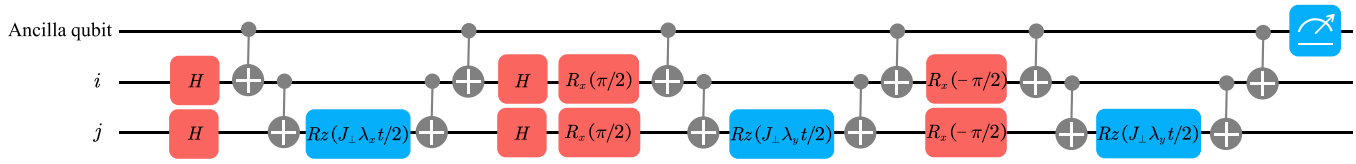


FIG. 7. The Trotter circuit for the time evolution of 2D term in $U_{n_a}$ for time $t$.

unitary $U_{\mathrm{1D}}$ to be expressed as the product of these operators. This decomposition facilitates efficient circuit implementation and ensures exact simulation of the 1D interactions:

$$e^{-ih_{i,j}(\vec{J})t} = \exp\left\{-iJ\left(\lambda_x\sigma_x^i\sigma_x^j + \lambda_y\sigma_y^i\sigma_y^j + \lambda_z\sigma_z^i\sigma_z^j\right)t/4\right\}. \tag{C2}$$

We implement the operator using the first-order Trotter-Suzuki decomposition, as shown in Fig. 8. For the 2D-interaction term $U_{\mathrm{2D}}$, QFM adjusts the interaction strength based on measurements of ancilla qubits. In the main-text example, the coupling strength between each ancilla qubit and the 2D system is parametrized by $\theta_i$, with an additional drift term $C$ applied to the 2D interaction:

$$\begin{aligned} U_{n_a}(\tau) = \Bigg[ & U_{\mathrm{1D}}(a) \times \exp\Bigg\{-i\frac{J_\perp}{4}[\theta_0(-1)^{r_0} + \theta_1(-1)^{r_1} \\ & + \theta_2(-1)^{r_2} + \theta_3(-1)^{r_3}] \otimes h_{4,9}\Bigg\} \\ & \times \exp\left\{-i\frac{C}{4}J_\perp \times h_{4,9}\right\} U_{\mathrm{1D}}(c)U_{\mathrm{1D}}(b)\Bigg]^\tau. \end{aligned} \tag{C3}$$

To implement different 2D-interaction strengths $J_\perp/J = 0, 1, 2, 3$ conditioned on outcomes of ancilla qubit measurements, we adjust the circuit parameters accordingly. Specifically, for the 1D interaction strength $J = 1$, the single-qubit rotation angles $\theta_i$ and the coupling constant $C$ are set to values that realize the desired interaction strength in the effective Hamiltonian. These parameter settings ensure that each measurement outcome on the ancilla qubits correctly tunes the two-dimensional coupling between data qubits. For the example in the main text, these parameters are set to

$$\begin{aligned} \theta_0 &= 1, \\ \theta_1 &= 0.5, \\ \theta_2 &= \theta_3 = 0, \\ C &= 1.5. \end{aligned} \tag{C4}$$

Furthermore, to avoid additional sampling overhead, it is necessary to generate a uniform distribution over $J_\perp/J$. This can be achieved by preparing each ancilla qubit in an equal superposition of its computational basis states, which is analogous to applying a Hadamard gate $\mathrm{H}_{\mathrm{gate}}$ to each ancilla qubit. In this way, all possible interaction strengths are sampled with equal probability without additional sampling or postselection:

$$\begin{aligned} U_{n_a}(\tau) = \big[ & U_{\mathrm{1D}}(a) \times \exp\left\{-i\left[\sigma_z^0 + 0.5\sigma_z^1\right] \otimes h_{4,9}(\vec{J}_\perp)\right\} \\ & \times \exp\{-i \times 1.5 h_{4,9}(\vec{J}_\perp)\} \\ & \times U_{\mathrm{1D}}(c)U_{\mathrm{1D}}(b)\big]^\tau \mathrm{H}_{\mathrm{gate}}^0 \mathrm{H}_{\mathrm{gate}}^1, \end{aligned} \tag{C5}$$

where the first-order Trotter-Suzuki decomposition of the 2D-interaction terms in $U_{n_a}$ is illustrated in Fig. 7, and interactions along different bonds and directions are implemented sequentially. After measurements, the circuit will be

$$U_{n_a}(\tau) = (U_{\mathrm{1D}}(a)U_{\mathrm{2D}}(a)U_{\mathrm{1D}}(c)U_{\mathrm{1D}}(b))^\tau, \tag{C6}$$

$$\begin{aligned} U_{\mathrm{2D}}(a) = & \exp\{-i[(-1)^{m_0} + 0.5(-1)^{m_1}] \otimes h_{4,9}(\vec{J}_\perp)\} \\ & \times \exp\{-i \times 1.5 h_{4,9}(\vec{J}_\perp)\}, \end{aligned} \tag{C7}$$

where each $r_i$ with $i \in \{0, 1\}$ is sampled independently with equal probability from $\{0, 1\}$. For the 2D interaction strength defined as $J_\perp = \theta_{a_0}(-1)^{r_0} + \theta_{a_1}(-1)^{r_1} + \theta_{a_2}(-1)^{r_2} + \theta_{r_3}(-1)^{r_3} + C$, the resulting distribution of $J_\perp$ is determined by the independent binary sampling of each $r_i$. This procedure ensures that all possible combinations of signs occur with equal probability, producing a discrete set of interaction strengths that can be used to implement a uniform or tailored ensemble of 2D couplings conditioned on the ancilla measurements:

$$p(r_0 = 0, r_1 = 0) = 0.25 : J_\perp = 3, \tag{C8}$$

$$p(r_0 = 0, r_1 = 1) = 0.25 : J_\perp = 2, \tag{C9}$$

$$p(r_0 = 1, r_1 = 0) = 0.25 : J_\perp = 1, \tag{C10}$$

$$p(r_0 = 1, r_1 = 1) = 0.25 : J_\perp = 0. \tag{C11}$$

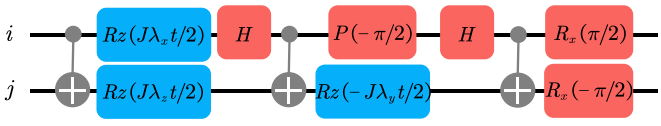


FIG. 8. The Trotter circuit for time evolution of $h_{i,j}$ over time $t$.

### 3. Performance of nonadaptive circuits in entanglement growth

In this Appendix, we describe how the adaptive selection between $U_n$ and $U_{n_a}$ in QFM prevents disruptions to accurate and stable mapping of state distributions across steps. At step $\tau$, we adaptively select $U_n$ and $U_{n_a}$ to ensure convergency. $U_n$ handles shifts in the state distribution from step $\tau - 1$, but cannot account for distortions, while $U_{n_a}$ can approximate distorted distributions. However, using $U_{n_a}$ on purely shifted distributions introduces fluctuations and hinders convergency. By appropriately choosing between $U_n$ and $U_{n_a}$, we accurately map the state distribution from step $\tau - 1$ to $\tau$.

We numerically compare circuits using only $U_n$ with those using only $U_{n_a}$, e.g., QuDDPM, in generating entangled states. A circuit composed solely of $U_n$ cannot distort the distribution of quantum states; therefore, the average entanglement

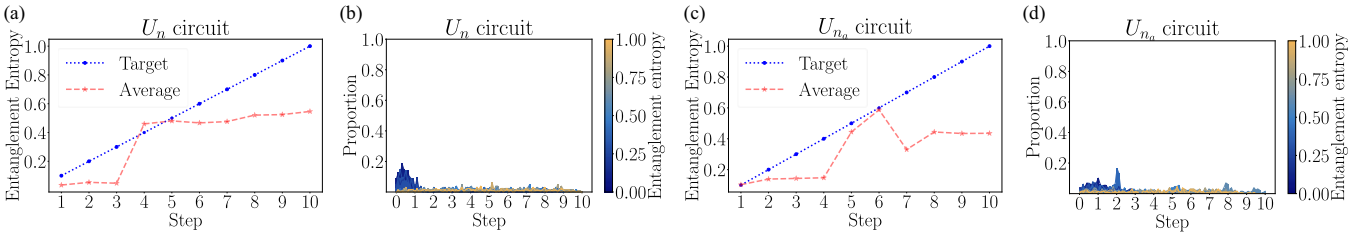


FIG. 9. The learning performance in entanglement growth for each step of QFM with only $U_n$ and for each step of QFM with only $U_{n_a}$. Both circuits fail to learn the expected evolution process: The average entanglement entropy (red dashed line) does not match the target entanglement entropy (blue dashed line), and the entanglement entropy distribution of generated states at each step does not concentrate around the target entanglement entropy.

entropy of the generated states fails to match the growth of the target entanglement entropy [Fig. 9(a)]. This is reflected in the distributions at each step: In the early steps, the entanglement entropy of the generated states exhibits a peak concentrated near the target entanglement entropy, but in later steps the distribution gradually becomes uniform [Fig. 9(b)]. We speculate that this is because, at the beginning, the growth of entanglement entropy is relatively small, so the target state distribution is not far from the initial state distribution, allowing some degree of fitting. However, for highly entangled states, the shape of the target distribution differs significantly from the initial distribution, leading to increasingly poor learning performance in subsequent steps. In contrast, a circuit composed solely of $U_{n_a}$ tends to alter the shape of the state distribution at every step. This increases the training difficulty in steps where the distribution shape should ideally remain unchanged [Fig. 9(c)]. Consequently, the generated states exhibit a large variance in entanglement entropy from the very beginning [Fig. 9(d)]. This illustrates the limitation of using a single type of circuit and highlights the need for carefully choosing the circuit structure at each step.

## APPENDIX D: MEASUREMENT-BASED DYNAMICS

In this Appendix, we discuss and compare QFM with the measurement-based quantum diffusion models (MBQDMs) [23] under measurement-based dynamics [62]. The measurement-based dynamics can be described by an SDE. For a pure state $|\psi_t\rangle$ at time $t$, continuously weakly measured by operator $X$ with measurement strength $k$, the stochastic evolution under weak measurements is governed by the stochastic Schrödinger equation:

$$\mathrm{d}|\psi_t\rangle = [-k\,(X_t - \langle X_t\rangle)^2\,\mathrm{d}t + \sqrt{2k}\,(X_t - \langle X_t\rangle)\,\mathrm{d}W_t]|\psi_t\rangle, \tag{D1}$$

where $\langle X_t\rangle = \langle\psi_t|X_t|\psi_t\rangle$ and $\mathrm{d}W_t$ describes a standard Wiener process. In QFM, ancilla qubits are evolved and measured in the $\sigma_z$ basis to implement weak measurements in different bases, enabling states sampled from the initial ensemble to be mapped onto that of target states through an appropriate choice of measurement operators, realized via suitable evolution of the ancilla qubits. Meanwhile, MBQDM uniformly spreads states sampled from the initial ensemble across the Hilbert space by choosing $X_t$ from an appropriate distribution, forming the basis of quantum diffusion models. To further illustrate the difference between QFM and MBQDM in state recovery, we represent the initial state as the projection operator $\rho_t = |\psi_t\rangle\langle\psi_t|$, and the stochastic partial differential equation takes the form

$$\mathrm{d}\rho_t = -k[X_t, [X_t, \rho_t]]\,\mathrm{d}t + \sqrt{2k}\,(X_t\rho_t + \rho_t X_t - 2\langle X_t\rangle\rho_t)\,\mathrm{d}W_t. \tag{D2}$$

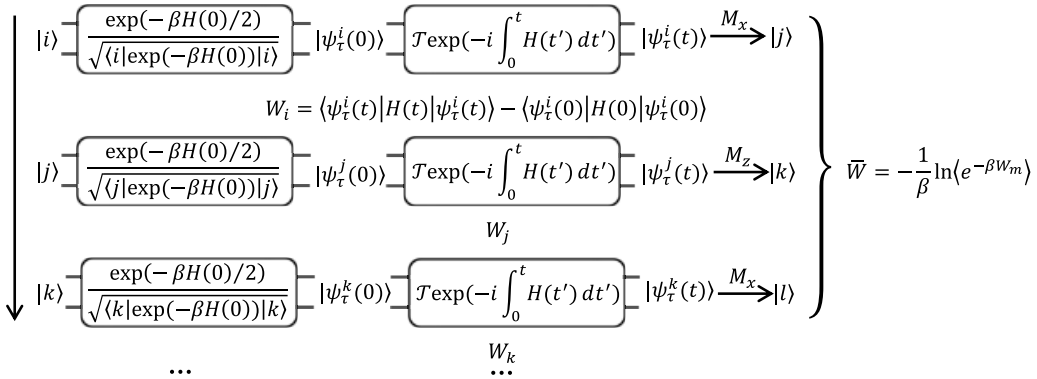


FIG. 10. The procedure of the METTS algorithm at $\beta_\tau$. The procedure starts from CPS $|i\rangle$. Each CPS is evolved in imaginary time to obtain the thermal state $|\psi_\tau^i(0)\rangle$. Alternating projective measurements in the $z$ and $x$ bases generate the next CPS $|j\rangle$. The state $|\psi_\tau^i(0)\rangle$ is then evolved under the real-time dynamics to $|\psi_\tau^i(t)\rangle$, from which a work sample $W_i$ is obtained. The free-energy difference is finally estimated by averaging over the work samples.

MBQDM expands $\rho_t = \boldsymbol{z_t}\boldsymbol{P}/2^n$ and $X_t = \boldsymbol{x_t}\cdot\boldsymbol{P}$ in the Pauli-operator basis $\boldsymbol{P} = \{P_i\}_{i=1}^{4^n}$, with $z_{t,i} = \mathrm{Tr}(\rho_t P_i) = \langle\psi_t|P_i|\psi_t\rangle$ and $x_{t,i} = \mathrm{Tr}(X_t P_i)/2^n$. It adopts different strategies for state trajectory-level and ensemble-average recovery. For trajectory-level recovery, MBQDM adopts a classical-quantum hybrid scheme, where a classical decoder is trained to infer the Pauli-basis coefficient vector $\{\boldsymbol{z}_t\}$ from measurement operators and outcomes, and generates the parameter $\theta$ of the denoising Hamiltonian $H_\theta(\boldsymbol{z}_t, t)$ for each state. For ensemble-average recovery, MBQDM applies the Petz recovery map to provide a principled way to invert quantum channels. Here, QFM generates an ensemble of distinct Hamiltonian evolutions from a single circuit by measuring the ancilla qubits. The evolution corresponding to each initial state is ensured, either through training or through analytical methods, to evolve each state to its corresponding target. As a result, QFM achieves ensemble-average recovery within a single quantum circuit.

## APPENDIX E: ESTIMATION ON FREE-ENERGY CHANGE

In this Appendix, we demonstrate how QFM can be used to estimate free-energy differences and verify the quantum Jarzynski equality. The Jarzynski equality [31] provides a remarkable connection between nonequilibrium processes and equilibrium free-energy differences. It states that the exponential average of the work $W$ performed on a system during a finite-time, nonequilibrium process satisfies

$$\langle e^{-\beta W}\rangle = e^{-\beta\Delta F}, \tag{E1}$$

where $\beta$ is the inverse temperature, $\Delta F$ is the free-energy difference between the initial and final equilibrium states, and the average is taken over many realizations of the process. This equality implies that equilibrium free-energy differences can be obtained from measurements of work in arbitrarily far-from-equilibrium transformations, making it a powerful tool in both classical and quantum thermodynamics. It also underpins fluctuation theorems and has been widely applied in experimental and computational studies of small systems.

ALGORITHM 1. Estimation of thermal average work $\bar{W}$ via QFM.

**Input:** Inverse temperature $\beta_\tau$, Hamiltonian $H(t)$
**Output:** Work samples $\{W_m\}$ and estimated $\Delta F$
**for** each sample index $m = 1$ to $M$ **do**
  (1) QFM evolves a CPS state $|i\rangle$ into METTS $|\widetilde{\psi}_\tau^m\rangle = |\widetilde{\psi}_\tau^m(0)\rangle$;
  (2) Compute initial energy: $E_i = \langle\widetilde{\psi}_\tau^m(0)|H(0)|\widetilde{\psi}_\tau^m(0)\rangle$;
  (3) Apply real time-evolution to get: $|\widetilde{\psi}_\tau^m(t)\rangle$;
  (4) Compute final energy: $E_f = \langle\widetilde{\psi}_\tau^m(t)|H(t)|\widetilde{\psi}_\tau^m(t)\rangle$;
  (5) Calculate work sample: $W_m = E_f - E_i$;
  (6) Measure $|\widetilde{\psi}_\tau^m\rangle$ to obtain next CPS state;
**end**
Compute thermal average: $e^{-\beta\Delta F} \approx e^{-\beta\bar{W}} = \langle e^{-\beta W_m}\rangle$.

In the quantum computing [32–34], METTS [10,11] provide a conceptually simple approach to preparing finite-temperature quantum states by representing the thermal density matrix $\rho_\beta$ as an ensemble of pure states. In practice, each METTS sample is generated by starting from a CPS $|i\rangle$, which is obtained through a sequence of projective measurements (typically alternating between the $M_z$ and $M_x$ bases).

To project such a CPS toward a thermal state, QITE is applied to approximate the nonunitary operator $\exp -\beta H/2$ and to obtain a normalized thermal-like state:

$$|\psi^i\rangle = \frac{e^{-\beta H/2}|i\rangle}{\|e^{-\beta H/2}|i\rangle\|}. \tag{E2}$$

Importantly, this QITE must be carried out independently for each sampled CPS, since different measurement outcomes lead to distinct initial states. As a result, the QITE circuit needs to be repeatedly adapted or reoptimized for every METTS sample in order to achieve proper thermaliza-

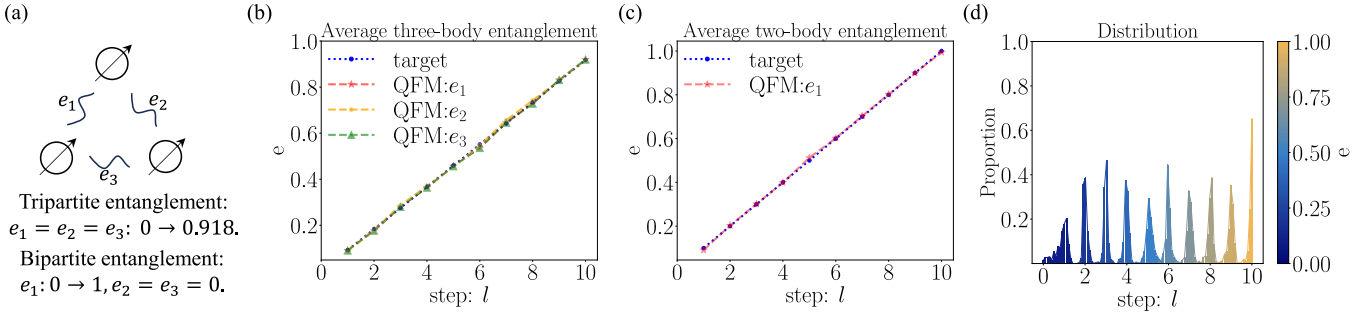


FIG. 11. Generation of entangled state ensembles in the tripartite entanglement system and the bipartite entanglement system. (a) Experiments on tripartite and bipartite entangled systems. In the tripartite case, the entanglement entropy between any two qubits increases simultaneously at each step, reaching the tripartite upper bound of $e_1 = e_2 = e_3 = 0.918$. In the bipartite case, the entanglement between two selected qubits rises to the maximal value of $e = 1$. (b) Red, orange, and green dots represent the average entanglement entropy between each pair of qubits in the generated three-qubit state ensemble at various steps, while blue dots indicate the target entanglement entropy at each step. (c) Red dots are the average entanglement entropy of the generated two-qubit state ensemble at various steps, and blue dots are the target entanglement entropy in each step. (d) The entanglement entropy distribution of the two-qubit state ensemble generated at each step is denoted by different colors.

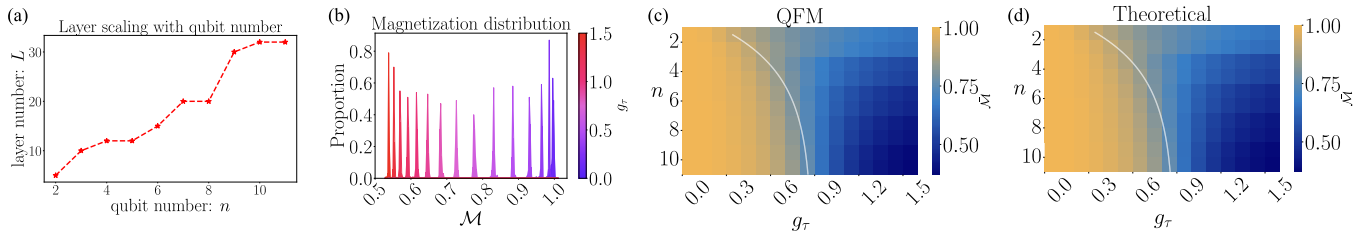


FIG. 12. Generation of magnetized state ensemble. QFM is trained to generate state ensembles of different magnetization $\mathcal{M}$ under various magnetic field strengths $g$. (a) The number of QFM layers scales with the number of qubits. (b) Magnetization distribution of the ensemble generated at each step in a four-qubit system, denoted by color. (c) The transition line (white) and the average magnetization $\bar{\mathcal{M}}$ for various system sizes $n$ and $g$, which closely matches the theoretical result in panel (d).

tion (Fig. 10). Consequently, the total experimental overhead scales with the number of samples required to represent the thermal ensemble, making METTS costly to implement on near-term quantum devices where frequent circuit recompilation or parameter updates are challenging.

To estimate free-energy differences using the Jarzynski equality, each METTS $|\psi^m\rangle$ is evolved under a finite-time, nonequilibrium protocol corresponding to a Hamiltonian change, and the work $W_m$ performed is recorded. The exponential average over the ensemble,

$$e^{-\beta\bar{W}} = \langle e^{-\beta W_m}\rangle = \frac{1}{M}\sum_{m=1}^{M} e^{-\beta W_m}, \tag{E3}$$

then provides an estimate of the free-energy difference $\Delta F$. This approach efficiently combines thermal-state sampling with quantum simulation of nonequilibrium dynamics, enabling accurate free-energy estimation on quantum devices without requiring full thermal density-matrix preparation. The thermal average work $\bar{W}$ is estimated by evolving sufficiently METTS. The detailed estimation procedure by QFM is provided in Algorithm 1.

For the example in the main text, we set $M = 1000$ for each $\beta_\tau$ in a six-qubit system.

## APPENDIX F: MORE BENCHMARKS FOR THE QUANTUM FLOW MATCHING

In this Appendix, we apply QFM to three quantum tasks—topological-state evolution, entanglement growth, and the magnetic phase transition in the TFIM—to demonstrate its versatility. Unlike QuDDPM, which generates specific target ensembles from a Haar-random ensemble, QFM can interpolate between different ensembles and capture entanglement growth, a task unattainable with the $U_{n_a}$ circuit alone.

### 1. Ring-state evolution

As the first application, we implement QFM to simulate the time evolution of quantum states with nontrivial topology, which appears when the classical dataset is encoded in quantum states [63–65]. Specifically, we focus on the rotational dynamics of the $Z$ axis when encoding classical ring-shaped data distributions into quantum states, preserving their topological connectivity through rotations of $\mathcal{T} = 20$ steps [Fig. 13(a)]. The evolution starts from $S_0 = \{|\psi_0^m\rangle||\psi_0^m\rangle = e^{-i\sigma_x G_m}|0\rangle, G_m \sim U(0, 2\pi)\}$ to $S_\tau = \{|\psi_\tau^m\rangle||\psi_\tau^m\rangle = e^{-i\sigma_z G_\tau}|\psi_{\tau-1}^m\rangle, G_\tau = \pi\tau/\mathcal{T}\}$. The loss function of $l$th step is chosen as

$$D(\vec{\theta}_l) = 1/M\sum_{m=1}^{M}\left\langle\psi_\tau^m\middle|V(\vec{\theta}_\tau)\middle|\widetilde{\psi}_{\tau-1}^m\right\rangle, \tag{F1}$$

leading to the generated ensembles in Fig. 13(a). The performance of QFM is validated by the match of the deviation from the theoretical and generated results [Fig. 13(b)].

### 2. Entanglement growth

Through dynamically tuned multibody interactions within a layered quantum-circuit architecture, QFM enables entanglement growth from separable to maximally entangled states, a key capability for scalable quantum-network node preparation and fault-tolerant logic-gate implementation [66–69]. We apply QFM to grow the entanglement entropy in two- and three-qubit quantum systems [70] over ten steps. The loss function at the $\tau$th step is defined as

$$D(\vec{\theta}_\tau) = \frac{1}{M}\sum m = 1^M\sum_{i=1}^{3}\left(e_i - \widetilde{e}_i^m(\vec{\theta}_\tau)\right)^2, \tag{F2}$$

where $\widetilde{e}_i^m(\vec{\theta}_\tau)$ denotes the entanglement entropy of the dynamically generated state $|\widetilde{\psi}_\tau^m\rangle = V(\vec{\theta}_\tau^{\rm op})|\widetilde{\psi}_{\tau-1}^m\rangle$ in the ensemble $\widetilde{S}_\tau$, and the target value $e_i$ increases linearly as $\tau/\mathcal{T}$. As shown in Figs. 11(b) and 11(c), the ensemble-averaged entanglement entropy closely follows the target linear growth. Figure 11(d) shows the distribution of states with given en-

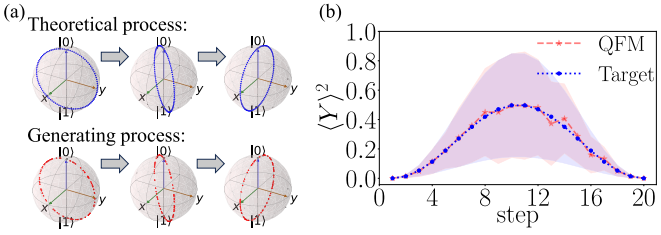


FIG. 13. Generation of ring-state ensemble. (a) Blue dots are sampled states used to train the model, and red dots are generated states. (b) The generated states (red) have matched deviation $\langle Y^2\rangle$ with the theoretical values (blue).

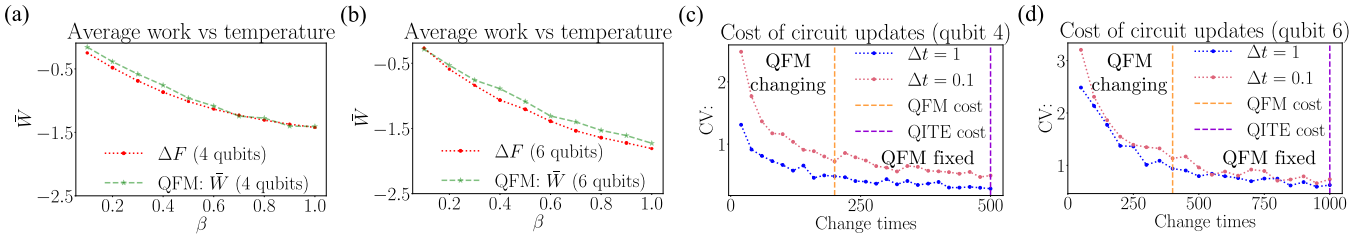


FIG. 14. Free-energy estimation by QFM for four- and six-qubit systems. (a), (b) For the four- and six-qubit cases, the averaged work agrees with the free-energy change as a function of the inverse temperature $\beta$. (c), (d) The CV (coefficient of variation) at $\beta = 1$ for time steps $\Delta t = 1$ (blue) and $\Delta t = 0.1$ (red) shows that QFM converges with about 200 (four qubits) and 400 (six qubits) circuit updates (orange), whereas QITE requires about 500 and 1000 updates, respectively (purple).

tanglement entropy at each step. In terms of circuit depth, QFM maintains the same depth for both the two-qubit and three-qubit systems. Therefore, for larger systems, if the target entanglement can be decomposed into multiple two- or three-body correlations, the required circuit depth scales linearly with the number of entangled units. When such a decomposition is not possible and QFM is trained directly to increase the entanglement of the entire system, the circuit depth still increases moderately.

### 3. Magnetic phase transition

Magnetic phase transitions, driven by quantum-fluctuation-induced symmetry breaking, reveal universal scaling in collective spins and guide quantum material design with controlled quantum phases [71,72]. For a TFIM described by the Hamiltonian,

$$H(g_\tau) = -\sum_i \sigma_z^i \sigma_z^{i+1} - g_\tau \sum_i \sigma_x^i, \tag{F3}$$

which undergoes a phase transition from the ordered ferromagnetic phase to the disordered phase with $g = 1$ [Fig. 12(a)]. We apply QFM to generate the ground-state ensemble $\widetilde{S}_\tau = \{|\widetilde{\psi}_\tau^m\rangle | |\widetilde{\psi}_\tau^m\rangle = U(\vec{\theta}_\tau^{\,\mathrm{op}})|\widetilde{\psi}_{\tau-1}^m\rangle\}$ of $H(1.5\tau/\mathcal{T})$, where $\tau = 1 \to 15$ and $\mathcal{T} = 15$. The loss function of step $\tau$ is

$$D(\vec{\theta}_\tau^{\,\mathrm{op}}) = (1/M)\sum_{m=1}^{M} \langle \widetilde{\psi}_{\tau-1}^m V^\dagger(\vec{\theta}_\tau) | H(1.5\tau/\mathcal{T}) | V(\vec{\theta}_\tau) \widetilde{\psi}_{\tau-1}^m \rangle. \tag{F4}$$

In Fig. 12(b), we demonstrate the performance in a four-qubit system, generating ground state-ensembles with QFM and measuring magnetization $\mathcal{M} = (\sum_i \sigma_z^i)/n$. In contrast to QuDDPM, which only generates the state ensemble with specific magnetization, QFM can track varying average magnetization $\bar{\mathcal{M}}$ of state ensembles, as shown in Fig. 12(c) with various qubits $n = 2 \to 11$ and field strengths $g_\tau = 0.0 \to 1.5$, which matches the theoretical result [Fig. 12(d)].

## APPENDIX G: ERROR AND ADVANCEMENT IN LARGER SYSTEMS

To ensure that QFM continues to generate high-fidelity results and maintains a reduced number of circuit adjustments for different system sizes, we performed free-energy estimation for systems of varying sizes. In both the four-qubit [Fig. 14(a)] and six-qubit cases [Fig. 14(b)], QFM accurately reproduces the free-energy values while requiring only about 40% of the circuit adjustments needed by conventional methods [Figs. 14(c) and 14(d)]. This confirms that the efficiency advantage of QFM is intrinsic to the method and scales favorably with system complexity, and that the error does not grow rapidly with system size.

[1] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Colloquium: Nonequilibrium dynamics of closed interacting quantum systems, Rev. Mod. Phys. **83**, 863 (2011).
[2] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics, Adv. Phys. **65**, 239 (2016).
[3] M. C. Banuls, R. Blatt, J. Catani, A. Celi, J. I. Cirac, M. Dalmonte, L. Fallani, K. Jansen, M. Lewenstein, and S. Montangero, Simulating lattice gauge theories within quantum technologies, Eur. Phys. J. D **74**, 165 (2020).
[4] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).
[5] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and many-body excitations with neural-network quantum states, Phys. Rev. Lett. **121**, 167204 (2018).
[6] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[7] M. Schmitt and M. Heyl, Simulating dynamics of correlated matter with neural quantum states, arXiv:2506.03124.
[8] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix product density operators: Simulation of finite-temperature and dissipative systems, Phys. Rev. Lett. **93**, 207204 (2004).
[9] J.-G. Liu, L. Mao, P. Zhang, and L. Wang, Solving quantum statistical mechanics with variational autoregressive networks

and quantum circuits, Mach. Learn.: Sci. Technol. **2**, 025011 (2021).
[10] E. Stoudenmire and S. R. White, Minimally entangled typical thermal state algorithms, New J. Phys. **12**, 055026 (2010).
[11] S. R. White, Minimally entangled typical quantum states at finite temperature, Phys. Rev. Lett. **102**, 190601 (2009).
[12] L. Bassman Oftelie, K. Klymko, D. Liu, N. M. Tubman, and W. A. de Jong, Computing free energies with fluctuation relations on quantum computers, Phys. Rev. Lett. **129**, 130603 (2022).
[13] J. C. Getelina, N. Gomes, T. Iadecola, P. P. Orth, and Y.-X. Yao, Adaptive variational quantum minimally entangled typical thermal states for finite temperature simulations, SciPost Phys. **15**, 102 (2023).
[14] B. Zhang, P. Xu, X. Chen, and Q. Zhuang, Generative quantum machine learning via denoising diffusion probabilistic models, Phys. Rev. Lett. **132**, 100602 (2024).
[15] F. G. Brandao, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).
[16] P. Ćwikliński, M. Horodecki, M. Mozrzymas, Ł. Pankowski, and M. Studziński, Local random quantum circuits are approximate polynomial-designs: Numerical results, J. Phys. A: Math. Theor. **46**, 305301 (2013).
[17] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).
[18] A. Schuckert, O. Katz, L. Feng, E. Crane, A. De, M. Hafezi, A. V. Gorshkov, and C. Monroe, Observation of a finite-energy phase transition in a one-dimensional quantum simulator, Nat. Phys. **21**, 374 (2025).
[19] K. Kumaran, M. Sajjan, B. Pokharel, J. Gibbs, J. Cohn, B. Jones, S. Mostame, S. Kais, and A. Banerjee, Superdiffusion resilience in Heisenberg chains with 2D interactions on a quantum processor, Phys. Rev. B **113**, 174408 (2026).
[20] N. Ma, M. Goldstein, M. S. Albergo, N. M. Boffi, E. Vanden-Eijnden, and S. Xie, SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers, arXiv:2401.08740.
[21] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, Flow matching for generative modeling, arXiv:2210.02747.
[22] D. Rezende and S. Mohamed, Variational inference with normalizing flows, in *International Conference on Machine Learning* (JMLR.org, Lille, France, 2015), p. 1530.
[23] X. Liu, J. Zhuang, W. Hou, and Y.-Z. You, Measurement-based quantum diffusion models, arXiv:2508.08799.
[24] F. Hu, G. Liu, Y. Zhang, and X. Gao, Local diffusion models and phases of data distributions, arXiv:2508.06614.
[25] X. Wang, B. Qi, Y. Wang, and D. Dong, Entanglement-variational hardware-efficient ansatz for eigensolvers, Phys. Rev. Appl. **21**, 034059 (2024).
[26] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandao, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).
[27] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, npj Quantum Inf. **5**, 75 (2019).
[28] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).
[29] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).
[30] W. Roga, D. Spehner, and F. Illuminati, Geometric measures of quantum correlations: Characterization, quantification, and comparison by distances and operations, J. Phys. A: Math. Theor. **49**, 235301 (2016).
[31] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys. Rev. Lett. **78**, 2690 (1997).
[32] J. Kurchan, A quantum fluctuation theorem, arXiv:cond-mat/0007360.
[33] H. Tasaki, Jarzynski relations for quantum systems and some applications, arXiv:cond-mat/0009244.
[34] S. Mukamel, Quantum extension of the Jarzynski relation: Analogy with stochastic dephasing, Phys. Rev. Lett. **90**, 170604 (2003).
[35] P. Talkner, E. Lutz, and P. Hänggi, Fluctuation theorems: Work is not an observable, Phys. Rev. E **75**, 050102(R) (2007).
[36] J. Richter and A. Pal, Simulating hydrodynamics on noisy intermediate-scale quantum devices with random circuits, Phys. Rev. Lett. **126**, 230501 (2021).
[37] J. Benton, G. Deligiannidis, and A. Doucet, Error bounds for flow matching methods, Trans. Machine Learn. Res. (2024), https://openreview.net/forum?id=uqQPyWFDhY .
[38] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, Nat. Commun. **14**, 3751 (2023).
[39] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, Autoregressive neural network for simulating open quantum systems via a probabilistic formulation, Phys. Rev. Lett. **128**, 090501 (2022).
[40] Y. Tang, J. Liu, J. Zhang, and P. Zhang, Learning nonequilibrium statistical mechanics and dynamical phase transitions, Nat. Commun. **15**, 1117 (2024).
[41] E. Zapusek, K. Kirova, W. Hahn, M. Marthaler, and F. Reiter, Variational quantum thermalizers based on weakly-symmetric nonunitary multi-qubit operations, Quantum Sci. Technol. **11**, 025006 (2026).
[42] A. D. King, A. Nocera, M. M. Rams, J. Dziarmaga, R. Wiersema, W. Bernoudy, J. Raymond, N. Kaushal, N. Heinsdorf, and R. Harris, Beyond-classical computation in quantum simulation, Science **388**, 199 (2025).
[43] X. Wu, C. Zhu, J. Wang, and X. Wang, BOSS: Blocking algorithm for optimizing shuttling scheduling in ion trap, *2025 IEEE International Symposium on High Performance Computer Architecture (HPCA)* (IEEE Computer Society, Los Alamitos, CA, USA, 2025), pp. 290–303.
[44] D. Gao *et al.*, Establishing a new benchmark in quantum computational advantage with 105-qubit Zuchongzhi 3.0 processor, Phys. Rev. Lett. **134**, 090601 (2025).
[45] R. Acharya, D. A. Abanin, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, and N. Astrakhantsev, Quantum error correction below the surface code threshold, Nature (London) **638**, 920 (2025).
[46] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, Nature (London) **627**, 778 (2024).
[47] H. Cao, F. Pan, D. Feng, Y. Wang, and P. Zhang, Generative decoding for quantum error-correcting codes, arXiv:2503.21374.

[48] P. Zhang, Correcting a noisy quantum computer using a quantum computer, arXiv:2506.08331.

[49] J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, *Advances in Neural Information Processing Systems (NeurIPS 2020)*, Vol. 33 (Curran Associates, Inc., 2020), pp. 6840–6851.

[50] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, Score-based generative modeling through stochastic differential equations, in *International Conference on Learning Representations (ICLR)* (Curran Associates, Inc., New Orleans, LA, USA, 2021), pp. 37799–37812.

[51] J. Song, C. Meng, and S. Ermon, Denoising diffusion implicit models, *International Conference on Learning Representations* (ICLR, 2021).

[52] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, High-resolution image synthesis with latent diffusion models, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2022), pp. 10684–10695.

[53] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, Consistency models, arXiv:2303.01469.

[54] P. Guo and A. G. Schwing, Variational rectified flow matching, *Proceedings of the 42nd International Conference on Machine Learning* (PMLR, 2025), Vol. 267, pp. 20921–20940.

[55] Z. Cui, The code for "Quantum flow matching", GitHub (2026), https://github.com/Machine-learning-and-complex-systems/QFM.git.

[56] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[57] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the Hamiltonian variational ansatz, PRX Quantum **1**, 020319 (2020).

[58] C. Lyu, X. Xu, M.-H. Yung, and A. Bayat, Symmetry enhanced variational quantum spin eigensolver, Quantum **7**, 899 (2023).

[59] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Sci. Technol. **4**, 014008 (2019).

[60] J. M. Arrazola, O. Di Matteo, N. Quesada, S. Jahangiri, A. Delgado, and N. Killoran, Universal quantum circuits for quantum chemistry, Quantum **6**, 742 (2022).

[61] A. J. McRoberts and R. Moessner, Parametrically long lifetime of superdiffusion in nonintegrable spin chains, Phys. Rev. Lett. **133**, 256301 (2024).

[62] K. Jacobs and D. A. Steck, A straightforward introduction to continuous quantum measurement, Contemp. Phys. **47**, 279 (2006).

[63] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[64] S. Lloyd, S. Garnerone, and P. Zanardi, Quantum algorithms for topological and geometric analysis of data, Nat. Commun. **7**, 10138 (2016).

[65] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, Nat. Commun. **12**, 2631 (2021).

[66] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[67] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, and D. A. Buell, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[68] H. J. Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).

[69] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[70] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, Quantum entanglement growth under random unitary dynamics, Phys. Rev. X **7**, 031016 (2017).

[71] M. Z. Hasan and C. L. Kane, Colloquium: Topological insulators, Rev. Mod. Phys. **82**, 3045 (2010).

[72] Z.-Y. Wang, X.-C. Cheng, B.-Z. Wang, J.-Y. Zhang, Y.-H. Lu, C.-R. Yi, S. Niu, Y. Deng, X.-J. Liu, and S. Chen, Realization of an ideal Weyl semimetal band in a quantum gas with 3D spin-orbit coupling, Science **372**, 271 (2021).