Supplementary Materials for

Noise-Resilient Quantum Metrology with Quantum Computing

Xiangyu Wang, Chenrong Liu, Xue Lin, Yu Tian, Yishan Li, Xinfang Nie, Yufang Feng, Yuxuan Zheng, Ying Dong, Xinqing Wang, and Dawei Lu

Corresponding authors:
Ying Dong, yingdong@cjlu.edu.cn;
Xinqing Wang, wxqnano@cjlu.edu.cn;
Dawei Lu, ludw@sustech.edu.cn

This PDF file includes:

- Supplementary Materials
- Figs. S1 to S7
- References

Contents

1 SUPPLEMENTARY SECTION 1: THEORETICAL MODEL

The initial state $|\psi_0\rangle$ is evolved into $|\psi_t\rangle$ by a sensing process characterized by operator $\tilde{\mathcal{U}}$, and then a quantum noise channel with its operator $\tilde{N}$ act on state $|\psi_t\rangle$, and the final state thus becomes

$$\tilde{\rho}_t = P_0|\psi_t\rangle\langle\psi_t| + (1-P_0)\tilde{N}|\psi_t\rangle\langle\psi_t|\tilde{N}^\dagger \tag{S1}$$

as shown in Fig. S1.

$$|\psi_0\rangle \xrightarrow{\boxed{\tilde{\mathcal{U}}}} |\psi_t\rangle \xrightarrow{\boxed{\tilde{N}:(1-P_0)}} \tilde{\rho}_t$$

Figure S1: **The chart of the state changes by sensing process characterized by operator $\tilde{\mathcal{U}}$ and quantum noise channel $\tilde{N}$.**

1.1 Fidelity

The fidelity of the final state is

$$\tilde{F} = \langle\psi_t|\tilde{\rho}_t|\psi_t\rangle = P_0 + (1-P_0)|\langle\psi_t|\tilde{N}|\psi_t\rangle|^2. \tag{S2}$$

Our method requires identifying the quantum principal component of this state. We thus analyze the quantum principle component of the final state and compare the fidelity of it with that of the full state fidelity in Eq. (S2).

Firstly, we expand the final state $\tilde{N}|\psi_t\rangle$ by its original state $|\psi_t\rangle$ and its perpendicular state $|\psi_t^\perp\rangle$ as

$$\tilde{N}|\psi_t\rangle = \alpha|\psi_t\rangle + \beta|\psi_t^\perp\rangle, \tag{S3}$$

with $\langle\psi_t^\perp|\psi\rangle = \langle\psi|\psi_t^\perp\rangle = 0$, then the final state can be expressed in the basis of $|\psi\rangle$, $|\psi_t^\perp\rangle$ as

$$\begin{aligned}
\tilde{\rho}_t &= P_0|\psi\rangle\langle\psi_t| + (1-P_0)\,\tilde{N}\,|\psi_t\rangle\langle\psi_t|\,\tilde{N}^\dagger \\
&= \left(P_0 + |\alpha|^2(1-P_0)\right)|\psi_t\rangle\langle\psi_t| + |\beta|^2(1-P_0)\,|\psi_{t\perp}\rangle\langle\psi_{t\perp}| \\
&\quad + \alpha\beta^*(1-P_0)\,|\psi_t\rangle\langle\psi_{t\perp}| + \alpha^*\beta(1-P_0)\,|\psi_{t\perp}\rangle\langle\psi_t| \\
&= \begin{bmatrix} P_0 + |\alpha|^2(1-P_0) & \alpha\beta^*(1-P_0) \\ \alpha^*\beta(1-P_0) & |\beta|^2(1-P_0) \end{bmatrix} \\
&= \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}.
\end{aligned} \tag{S4}$$

The eigenvalues and its eigenvectors can be obtained by some mathematical calculation as

$$\begin{aligned}\lambda_+ &= \frac{1+\sqrt{1-4|\beta|^2(1-P_0)P_0}}{2},\\ \eta_+ &= \frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\sqrt{\lambda_+-\rho_{22}}\\ \frac{\rho_{12}^*}{\sqrt{\lambda_+-\rho_{22}}}\end{pmatrix},\end{aligned} \tag{S5}$$

and

$$\begin{aligned}\lambda_- &= \frac{1-\sqrt{1-4|\beta|^2(1-P_0)P_0}}{2},\\ \eta_- &= \frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\frac{\rho_{12}}{\sqrt{\rho_{11}-\lambda_-}}\\ \sqrt{\rho_{11}-\lambda_-}\end{pmatrix},\end{aligned} \tag{S6}$$

where

$$\begin{aligned}\Delta\lambda &= \lambda_+-\lambda_- = \sqrt{1-4|\beta|^2(1-P_0)P_0},\\ \lambda_+-\rho_{22} &= \frac{(\rho_{11}-\rho_{22})+\Delta\lambda}{2},\\ \lambda_--\rho_{11} &= -\frac{(\rho_{11}-\rho_{22})+\Delta\lambda}{2},\end{aligned}$$

with

$$\begin{aligned}(\lambda_+-\rho_{22})^2+|\rho_{12}|^2 &= (\lambda_+-\rho_{22})\Delta\lambda,\\ (\lambda_--\rho_{11})^2+|\rho_{12}|^2 &= -(\lambda_--\rho_{11})\Delta\lambda.\end{aligned}$$

We thus rewrite the $\tilde{\rho}_t$ as

$$\begin{aligned}\tilde{\rho}_t &= \lambda_+\eta_+\eta_+^\dagger+\lambda_-\eta_-\eta_-^\dagger\\ &= \lambda_+|+\rangle\langle+|+\lambda_-|-\rangle\langle-|\\ &= K\begin{pmatrix}\lambda_+ & 0\\ 0 & \lambda_-\end{pmatrix}K^\dagger,\end{aligned} \tag{S7}$$

and λ_+ part is indeed the principle component of the final state, where the diagonalization transformed matrix

$$K=(\eta_+,\eta_-)=\frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\sqrt{\lambda_+-\rho_{22}} & \frac{\rho_{12}}{\sqrt{\rho_{11}-\lambda_-}}\\ \frac{\rho_{12}^*}{\sqrt{\lambda_+-\rho_{22}}} & -\sqrt{\rho_{11}-\lambda_-}\end{pmatrix} \tag{S8}$$

with $\det K = -1$ as can be checked. Actually, the diagonaliztion is a basis change process, that is, in the new basis $|\pm\rangle$, the density matrix $\tilde{\rho}_t$ is diagonalized as Diag$\{\lambda_+,\ \lambda_-\}$. Then, we have the relation

$$\begin{aligned} |+\rangle &= K^\dagger|\psi_t\rangle \Rightarrow |\psi_t\rangle = K|+\rangle, \\ |-\rangle &= K^\dagger|\psi_t^\perp\rangle \Rightarrow |\psi_t^\perp\rangle = K|-\rangle. \end{aligned}$$

The fidelity of the principle component of the final state [state optimized by quantum computing (QC)] can be obtained as

$$F = \langle\psi_t|\rho_{f+}|\psi_t\rangle = |\langle+|\psi_t\rangle|^2 = |\langle+|K|+\rangle|^2, \tag{S9}$$

where

$$\begin{aligned} \langle+|K|+\rangle &= \frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\sqrt{\lambda_+ - \rho_{22}} & \frac{\rho_{12}}{\sqrt{\lambda_+ - \rho_{22}}}\end{pmatrix}\frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\sqrt{\lambda_+ - \rho_{22}} & \frac{\rho_{12}}{\sqrt{\rho_{11}-\lambda_-}} \\ \frac{\rho_{12}^*}{\sqrt{\lambda_+ - \rho_{22}}} & -\sqrt{\rho_{11}-\lambda_-}\end{pmatrix} \\ &\quad \frac{1}{\sqrt{\Delta\lambda}}\begin{pmatrix}\sqrt{\lambda_+ - \rho_{22}} \\ \frac{\rho_{12}^*}{\sqrt{\lambda_+ - \rho_{22}}}\end{pmatrix} \\ &= \frac{1}{(\Delta\lambda)^{\frac{3}{2}}}\left[(\lambda_+ - \rho_{22})^{\frac{3}{2}} + \frac{|\rho_{12}|^2}{\sqrt{\rho_{11}-\lambda_-}} + \frac{|\rho_{12}|^2}{\sqrt{\lambda_+ - \rho_{22}}} - \frac{|\rho_{12}|^2\sqrt{\rho_{11}-\lambda_-}}{(\lambda_+ - \rho_{22})}\right]. \end{aligned}$$

the differences between the quantum metrology (QM) state fidelity

$$\tilde{F} = P_0 + (1-P_0)\left|\langle\psi_t|\tilde{N}|\psi_t\rangle\right|^2 = \rho_{11} = 1 - \rho_{22} \leq 1$$

and its principle component fidelity $F - \tilde{F}$ can be calculated numerically in parameters setup as

$$\alpha = \cos\frac{\Theta}{2}, \beta = e^{i\phi}\sin\frac{\Theta}{2}$$

with

$$P_0 \in [0,1],\ \Theta \in [0, 2\pi],\ \phi \in [0, 2\pi].$$

Both of them are ϕ-independent, the numerical results as shown in Fig. S2 confirms that the fidelity of principle component in come special parameter zone of Θ, P_0 shall larger than the QM state.

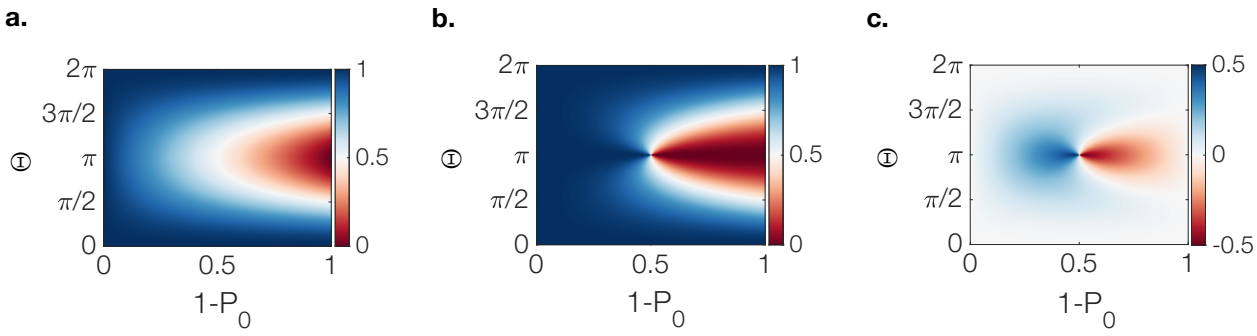

Figure S2: **Fidelity results based on the theoretical model. a,b,** State fidelity $\tilde{F}$ without QC optimization and F with QC optimization as functions of $1 - P_0$ and Θ. **c,** Fidelity enhancement $F - \tilde{F}$ achieved by the QM+QC method.

1.2 Quantum Fisher Information

We begin by considering measurement precision in a sensing process. For example, using a ruler to measure an unknown length d, we perform N independent measurements to obtain a set of samples $\{d_i\}$. Due to inevitable measurement errors, each d_i slightly deviates from the true value, and the precision of the estimation $\hat{d}$ improves as this deviation decreases. In classical statistics, the Fisher information quantifies the lower bound of such estimation errors.

In QM, the sensing outcomes arise from quantum evolution rather than classical sampling. The corresponding limit of precision is given by the quantum Fisher information (QFI), which determines the minimum achievable variance in estimating a parameter ξ encoded in a quantum state $\rho(\xi)$:

$$\mathrm{Var}[\hat{\xi}] \geq \frac{1}{n\,\mathcal{F}[\rho(\xi)]}, \tag{S10}$$

where n denotes the number of independent measurements. Here, $\mathcal{F}[\rho(\xi)]$ measures the sensitivity of the quantum state $\rho(\xi)$ to changes in the parameter ξ, and is defined as

$$\mathcal{F}[\rho(\xi)] = \mathrm{Tr}\big[\rho(\xi)L^2\big], \tag{S11}$$

with L being the symmetric logarithmic derivative (SLD) operator satisfying

$$\frac{\partial\rho(\xi)}{\partial\xi} = \frac{1}{2}\big(L\rho(\xi) + \rho(\xi)L\big). \tag{S12}$$

The core of calculating the QFI lies in solving the SLD. Here we employ the spectral decomposition of the density matrix, since the quantum states under consideration can be

mixed in general. The density operator is expressed in terms of its eigenvalues and eigenvectors as $\rho = \sum_k s_k |k\rangle \langle k|$, where s_k are the eigenvalues and $|k\rangle$ the corresponding eigenvectors. Projecting the SLD equation onto the eigenbasis $\{|k\rangle\}$ gives

$$\langle i| \dot{\rho} |j\rangle = \tfrac{1}{2}\big(\langle i| L\rho |j\rangle + \langle i| \rho L |j\rangle \big) = \tfrac{s_i+s_j}{2} L_{ij}. \tag{S13}$$

This leads to the spectral decomposition form of the SLD:

$$L_{ij} = \begin{cases} \dfrac{2\langle i|\dot{\rho}|j\rangle}{s_i + s_j}, & s_i + s_j > 0, \\ 0, & s_i = s_j = 0. \end{cases} \tag{S14}$$

For the diagonal elements ($i = j$) with $s_i > 0$,

$$L_{ii} = \frac{\partial_\xi s_i}{s_i}. \tag{S15}$$

The QFI can then be evaluated as

$$\begin{aligned} \mathcal{F}[\rho(\xi)] = &\sum_{k:s_k>0} \frac{(\partial_\xi s_k)^2}{s_k} \\ &+ 2 \sum_{k,l:s_k+s_l>0} \frac{(s_k - s_l)^2}{s_k + s_l} \left|\langle \psi_k | \partial_\xi \psi_l \rangle\right|^2. \end{aligned} \tag{S16}$$

For a pure state $\rho = |\psi\rangle \langle\psi|$, this expression reduces to

$$\mathcal{F} = 4\big(\langle \partial_\xi \psi | \partial_\xi \psi \rangle - |\langle \psi | \partial_\xi \psi \rangle|^2\big). \tag{S17}$$

In practical calculations, the partial derivative $\dot{\rho}$ can be obtained numerically through a finite difference:

$$\partial_\xi \tilde{\rho}_t = \frac{\tilde{\rho}_t^+ - \tilde{\rho}_t^-}{2\delta}, \tag{S18}$$

where $\tilde{\rho}_t^\pm = \tilde{U}_{\phi\pm\delta}(|\psi_0\rangle \langle\psi_0|)$, and δ is a small detuning parameter.

The optimization of quantum states on a quantum computer can be viewed as an analogue of the principal component analysis process, realized through eigen-decomposition. By identifying the eigenstate corresponding to the largest eigenvalue, one obtains the optimized quantum state ρ_{NR}.

1.3 Noise model

Noise in quantum systems is inherently complex and diverse, and numerous theoretical models have been developed to describe its various forms. Following the approach outlined in the previous section, we have established a representative noise model to illustrate the optimization capability of the QM+QC method. In this section, we further extend our analysis by introducing additional types of quantum noise channels. As shown in Fig. S3, we consider three typical examples: the dephasing channel,

$$\rho \to E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger, \tag{S19}$$

where

$$E_0 = \sqrt{1-p} I, \; E_1 = \sqrt{p} Z, \tag{S20}$$

the amplitude damping channel,

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}, \tag{S21}$$

and the depolarizing channel,

$$\rho \to (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{S22}$$

Consistent with the main text, the enhancement in quantum Fisher information (QFI) is employed as a quantitative measure to characterize the optimization performance achieved by the QM+QC framework.

2 SUPPLEMENTARY SECTION 2: RANDOM NOISE IN THE EXPERIMENT

We next consider the influence of an unstable magnetic field on the Ramsey interferometry experiment. In Ramsey interferometry, the population signal exhibits cosine interference fringes that oscillate with the evolution time τ. When the magnetic field experiences Gaussian-distributed fluctuations, either naturally or through controlled perturbations, these fluctuations modify the phase accumulation of the quantum superposition. Averaged over many repetitions, the net effect of such instability manifests as decoherence of the interference fringes.

Under an ideal static magnetic field B, the signal measured by a Ramsey sequence follows

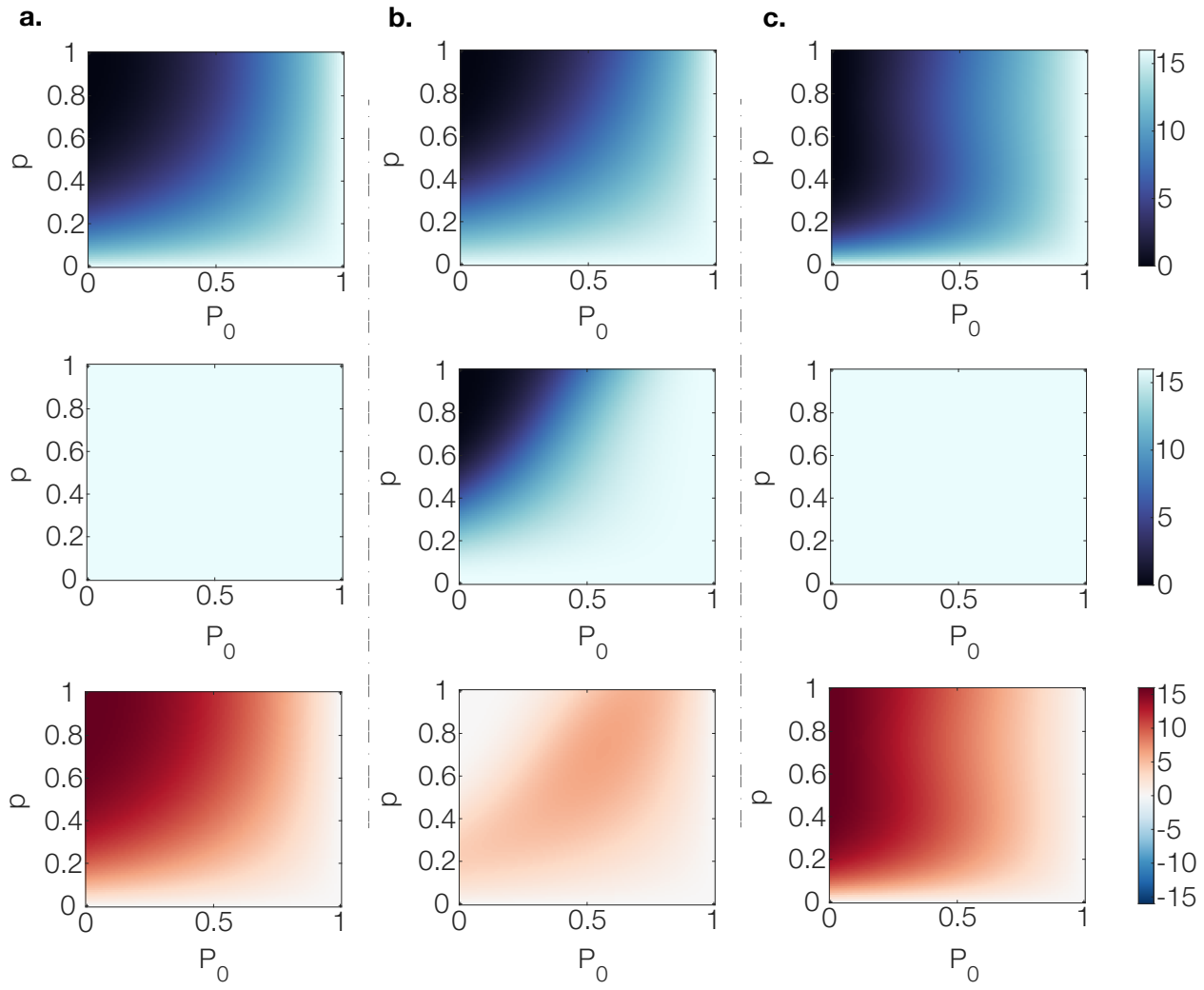

Figure S3: **QM+QC results under different noise conditions. a–c,** Quantum Fisher information (QFI) optimization results for the dephasing, amplitude-damping, and depolarizing channels, respectively. Each column, from top to bottom, shows: the QFI before QC optimization ($\mathcal{F}_{\mathrm{QM}}$), the QFI after QC optimization ($\mathcal{F}_{\mathrm{QM+QC}}$), and the corresponding QFI improvement ($\Delta\mathcal{F} = \mathcal{F}_{\mathrm{QM+QC}} - \mathcal{F}_{\mathrm{QM}}$).

a simple cosine oscillation with time τ:

$$\mathcal{P}(\tau) = \tfrac{1}{2}\big[1 + \cos(\Delta\omega\tau)\big], \tag{S23}$$

where $\Delta\omega = \gamma(B - B_{\mathrm{ref}})$ is the frequency shift induced by the magnetic field deviation from a reference value B_{ref}, and γ denotes the gyromagnetic ratio. The reference field B_{ref} represents the static bias field typically present in magnetic-field sensing experiments.

In each experimental repetition, the magnetic field B is sampled from a Gaussian distribution $\mathcal{N}[B_0, (B_0\varsigma)^2]$, where B_0 is the average value of the magnetic field to be measured and

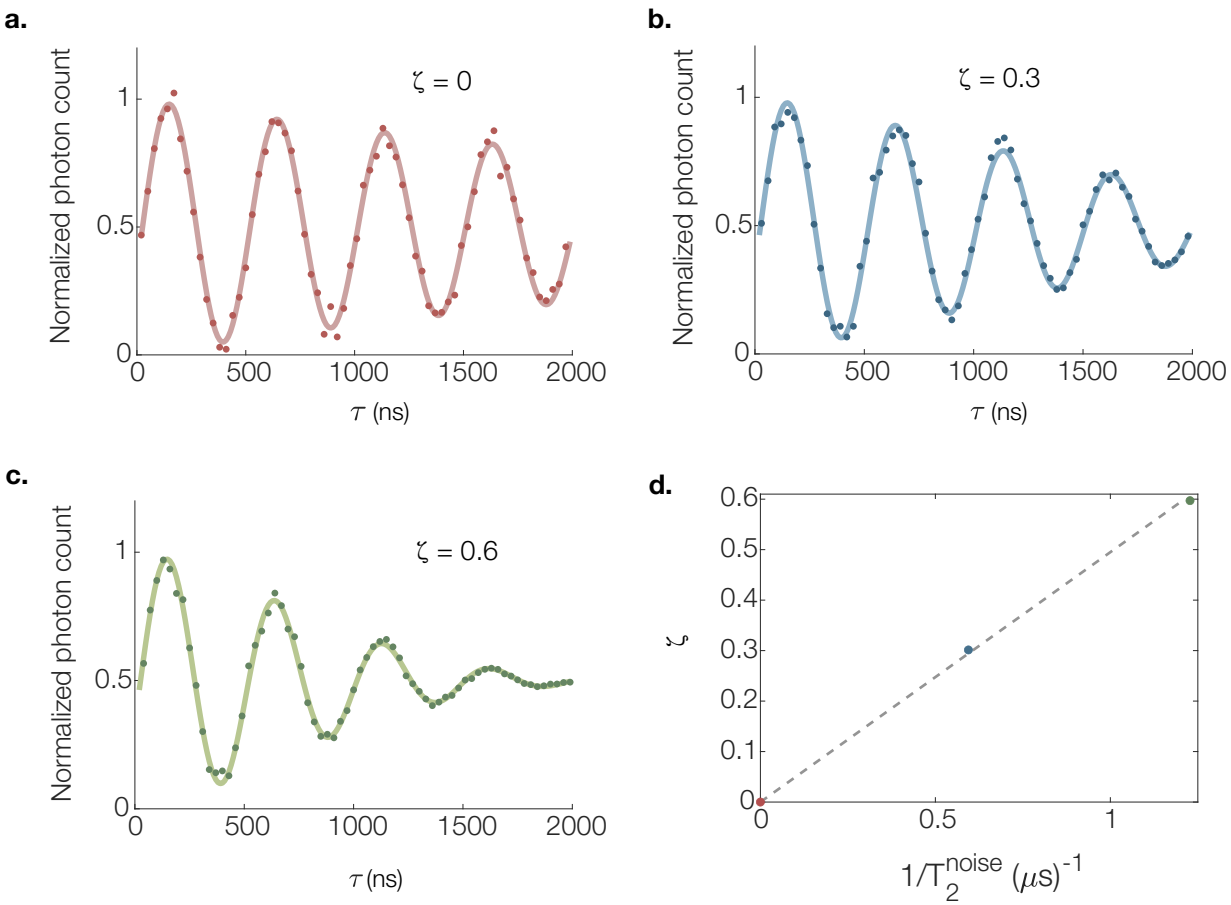

Figure S4: **Effects of different noise intensities on Ramsey signals. a-c,** The attenuation of the Ramsey signal under $\varsigma = 0, 0.3, 0.6$. We can obtain the equivalent decoherence time caused by noise by fitting it with a formula: $e^{-\tau/T_2^*}e^{-\left(\tau/T_2^{noise}\right)^2}$. **d,** Mapping relationship between equivalent decoherence time T_2^{noise} and noise intensity σ.

ς is standard deviation in percentage. This leading to a corresponding Ramsey signal

$$\mathcal{P}_i(\tau) = \tfrac{1}{2}\big[1 + \cos\big(\gamma(B_i - B_{\text{ref}})\tau\big)\big]. \tag{S24}$$

The ensemble-averaged signal can then be obtained by integrating over the Gaussian distribution of magnetic fields:

$$\bar{\mathcal{P}}(\tau) = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi}\, B_0\varsigma}\big[1 + \cos\big(\gamma(B - B_{\text{ref}})\tau\big)\big] e^{-\frac{1}{2}\left(\frac{B-B_0}{B_0\varsigma}\right)^2} dB. \tag{S25}$$

Neglecting the constant term and using the linearity of integration, the Fourier transform of

the Gaussian function yields

$$\int \cos\big(\gamma(B-B_{\mathrm{ref}})\tau\big)e^{-\frac{1}{2}\left(\frac{B-B_0}{B_0\varsigma}\right)^2}dB \propto \cos\big(\gamma(B_0-B_{\mathrm{ref}})\tau\big)e^{-\frac{1}{2}(\gamma B_0\varsigma\tau)^2}. \tag{S26}$$

Consequently, the ensemble-averaged Ramsey signal takes the form

$$\bar{\mathcal{P}}(\tau)=\frac{1}{2}\left[1+e^{-\frac{1}{2}(\gamma B_0\varsigma\tau)^2}\cos\big(\gamma(B_0-B_{\mathrm{ref}})\tau\big)\right]. \tag{S27}$$

The result shows that while the Ramsey oscillation frequency remains unchanged, its amplitude is modulated by a Gaussian envelope $e^{-\frac{1}{2}(\gamma B_0\varsigma\tau)^2}$. This behavior is equivalent to decoherence induced by quasi-static magnetic noise, and can thus be expressed as $e^{-\frac{1}{2}(\tau/T_2^{\mathrm{noise}})^2}$, where the corresponding dephasing time is $T_2^{\mathrm{noise}}=1/(\gamma B_0\varsigma)$.

3 SUPPLEMENTARY SECTION 3: LOSS FUNCTION

As discussed in the main text, the objective of the loss function is to obtain a target operator by minimizing its value through iterative optimization. Here, we design an appropriate loss function to achieve the diagonalization of a density matrix. Any density matrix can be expressed in its spectral decomposition as $\rho=\sum_{j=0}^{\mathcal{M}-1}\lambda_j\left|\psi_j\right\rangle\left\langle\psi_j\right|$, where $\{|\psi_j\rangle\}$ form an orthogonal basis and $\mathcal{M}=2^m$ denotes the Hilbert space dimension of an m-qubit system. Without loss of generality, the eigenvalues are arranged in descending order, $\lambda_0\geq\lambda_1\geq\cdots\geq\lambda_{\mathcal{M}-1}$. If a unitary transformation U can be found such that it maps $|\psi_j\rangle$ to the computational basis state $|j\rangle$, i.e.

$$\tilde{\rho}_t=U\rho U^\dagger=\sum_{j=0}^{\mathcal{M}-1}\lambda_j\left|j\right\rangle\left\langle j\right|, \tag{S28}$$

then measurements in the computational basis $\{|j\rangle\}$ directly yield the eigenvalues λ_j as outcome probabilities. For generic states, however, efficiently finding such a unitary U is generally intractable.

To formulate a practical loss function, we introduce a semidefinite, positive, and non-degenerate Hermitian operator P, whose eigenstates are the computational basis states $|j\rangle$ with corresponding eigenvalues s_j, such that $P=\sum_{j=0}^{\mathcal{M}-1}s_j\left|j\right\rangle\left\langle j\right|$. Unlike the eigenvalues of ρ, we arrange the set $\{s_j\}$ in ascending order, $0\leq s_0<s_1<\cdots<s_{\mathcal{M}-1}$, and impose the normalization condition $\sum_j s_j=1$. Such an operator is straightforward to construct. A convenient choice is a diagonal matrix with elements $\{0,\ 1,\ 2,\cdots,\ \mathcal{M}-1\}$, normalized by

the factor $\mathcal{M}(\mathcal{M}-1)/2$. In the Pauli basis, it can be written as

$$P = \frac{1}{\mathcal{M}(\mathcal{M}-1)} \sum_{j=1}^{n} 2^{j-1}\left(\sigma_z^j + 1\right), \tag{S29}$$

where σ_z^j denotes the Pauli-Z operator acting on the jth qubit [1].

We assert that a unitary operator U that minimizes the expectation value $\mathrm{Tr}\left[U\rho U^\dagger P\right]$ will diagonalize the density matrix ρ in the computational basis. The operator P can be any Hermitian matrix satisfying the conditions described above. The task therefore reduces to finding the optimal unitary U. In practice, we input the state ρ into a PQC represented by $U(\boldsymbol{\theta})$, and measure the expectation value of P at the circuit output. For instance, evaluating the operator P defined in Eq. S29 requires only a simultaneous Pauli-Z measurement on all qubits, which is experimentally straightforward. The corresponding objective function is defined as

$$L(\boldsymbol{\theta}) = \mathrm{Tr}\left[U(\boldsymbol{\theta})\,\rho\,U^\dagger(\boldsymbol{\theta})\cdot P\right]. \tag{S30}$$

By minimizing this objective function over the parameter space using a suitable classical optimization algorithm, one can obtain a good approximation to the target unitary U.

4 SUPPLEMENTARY SECTION 4: PQC MODEL

To realize QC optimization, we employ a PQC approach. For small systems, the PQC structure can be relatively simple. For instance, in the single-qubit case, only two rotation gates along distinct axes are sufficient, rendering the optimization process rather straightforward. However, as the number of qubits increases, the choice of PQC architecture becomes crucial, as it can substantially influence the convergence and final performance of the QC optimization. This aspect is further analyzed in the following section.

We first outline the complete PQC workflow. After the noisy sensing stage, the resulting quantum state $\tilde{\rho}_t$ is transferred into a quantum processor (some transmission losses may occur in practice, although they do not affect the subsequent PQC procedure). The state then evolves under a parameterized unitary transformation $U(\boldsymbol{\theta})$, which forms the variational component of the hybrid quantum–classical optimization loop.

During the optimization process, the parameterized unitary operation is applied to the input quantum state, after which the corresponding loss function is evaluated (the form and physical meaning of this function were introduced in the previous section). To obtain the gradient required for parameter updates, one of the circuit parameters is perturbed by a fixed amount when implementing the parameterized operation. Unlike conventional differentiation

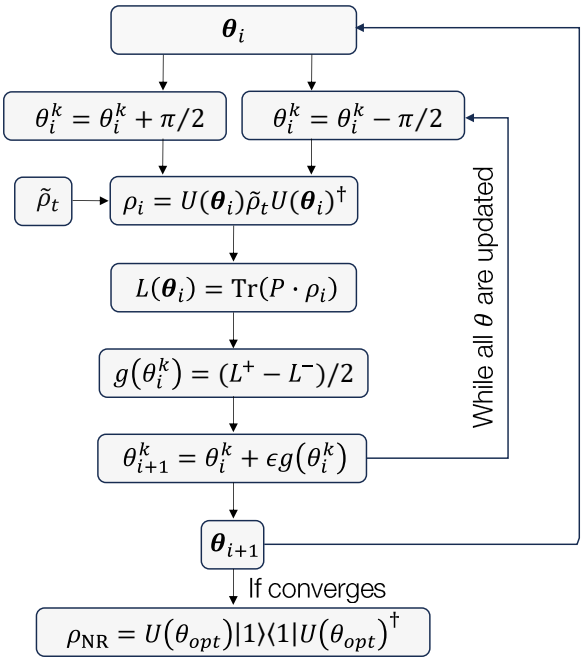

Figure S5: **PQC Flowchart.**

methods that rely on infinitesimal variations, it has been proven that for rotation operations, shifting a parameter by $\pm\pi/2$ can yield an exact analytical gradient. The computed gradients are then used to iteratively update the circuit parameters toward the loss minimum.

As the optimization proceeds, the parameters are progressively refined, and for a well-structured PQC model, the iterative process converges to a physically meaningful solution. The key objective of this section is therefore to explore the characteristics of an effective PQC design. In practice, the design considerations primarily involve the number of trainable parameters, the arrangement of two-qubit entangling gates, the overall circuit depth, and the specific choice of gate set.

We next introduce the circuit models explored in this work. To illustrate the importance of PQC architecture design more intuitively, we first present a set of suboptimal circuit configurations. As shown in Fig. S6a, although the fidelity increases markedly when the circuit depth grows from one to two layers, further increasing the number of layers to four yields almost no improvement. This observation indicates that the ansatz suffers from an expressibility or train-ability limitation, and therefore does not constitute a suitable PQC model for our purpose. Fig. S6b then presents a properly designed PQC architecture, which exhibits distinct convergence behavior and achieves superior final optimization performance.

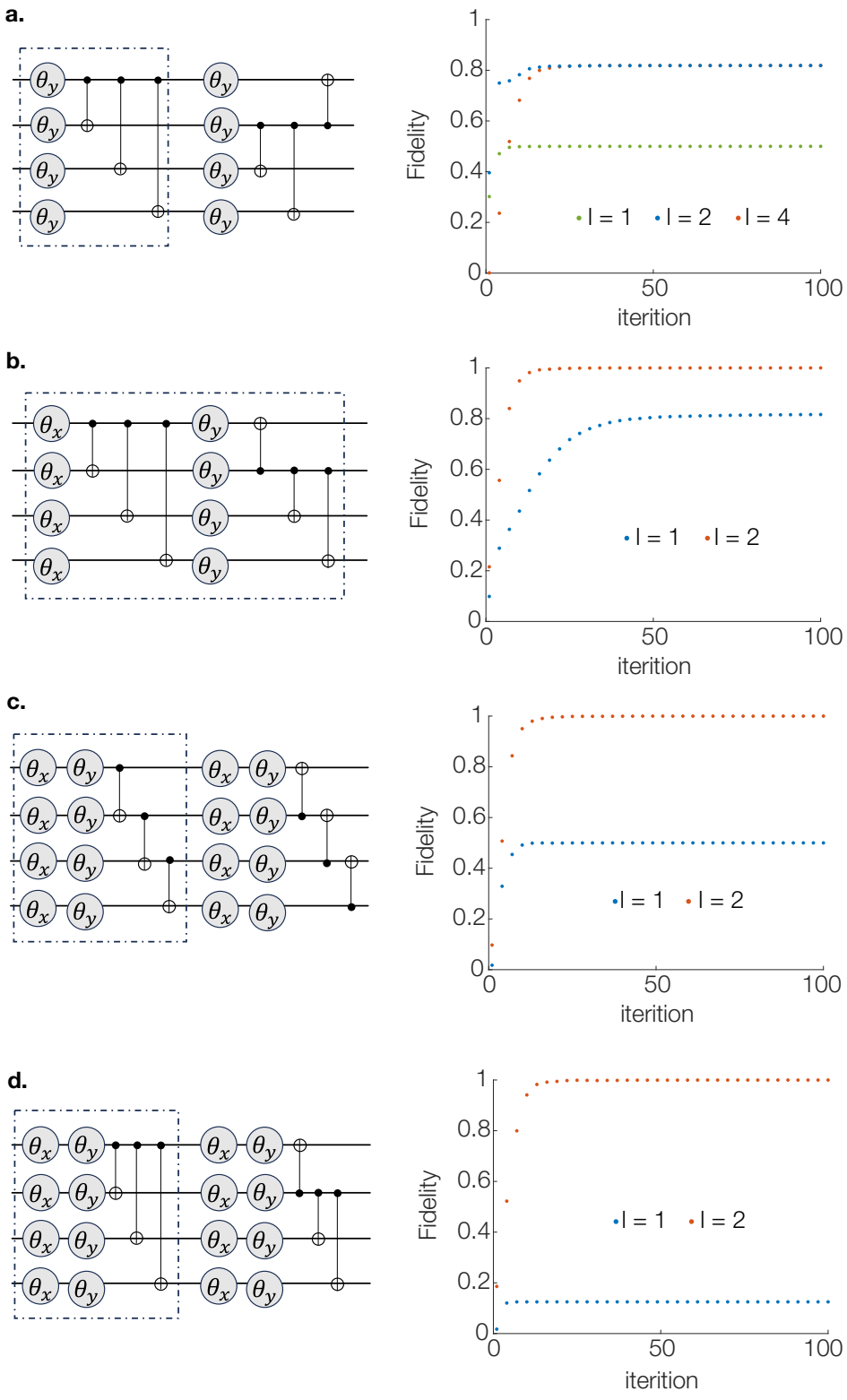

Figure S6: **The impact of different PQC on the optimization process. a-d,** The left side is a schematic diagram of PQC. The area within the dashed box is called a layer. The right side shows the optimization results for different number of layers l.

These results highlight the crucial role of circuit structure in determining both convergence speed and optimization accuracy.

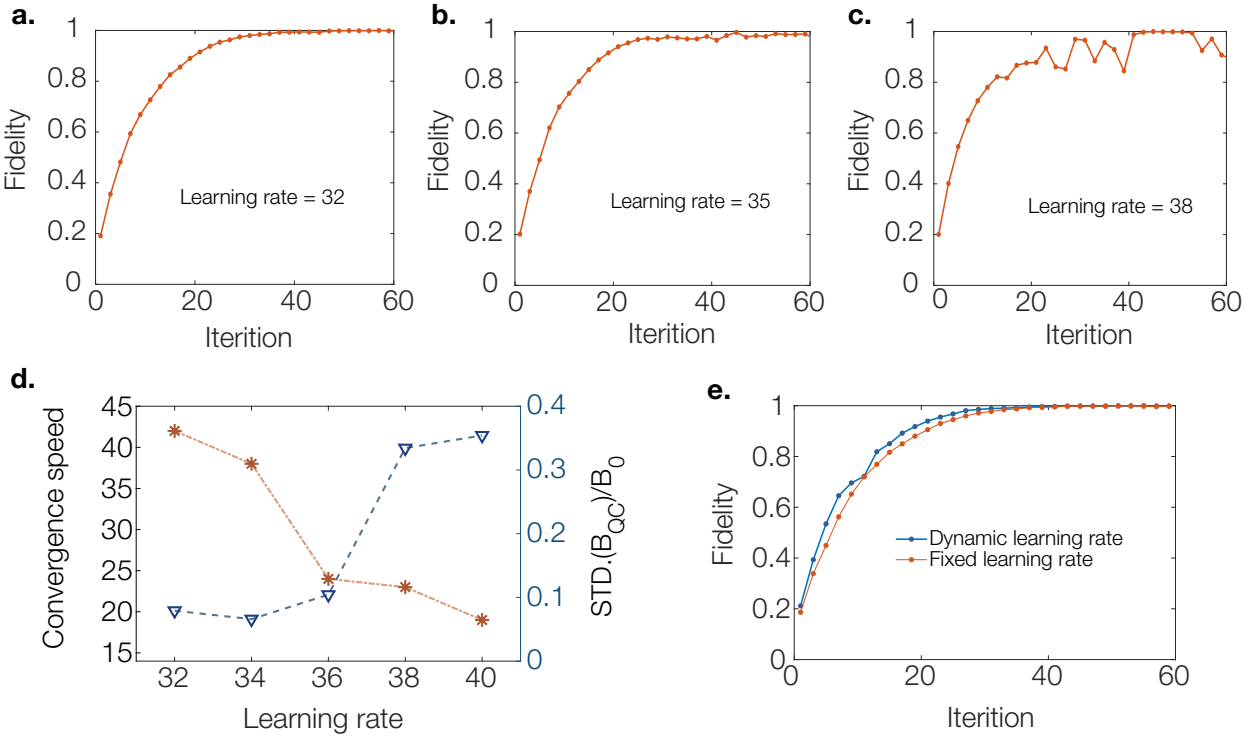

Figure S7: **Impact of different learning rates on PQC performance. a-c,** Evolution of fidelity during the optimization process for different learning rates. **d,** Relationship between convergence speed and the standard deviation of the optimized results as a function of learning rate. Here, convergence speed denotes the number of iterations required to reach convergence. **e,** Comparison between dynamic and fixed learning rates, showing that a dynamic learning-rate schedule achieves faster convergence without increasing the standard deviation.

As demonstrated above, different PQC architectures exhibit distinct optimization behaviors. Even for an identical circuit structure, however, the choice of optimization hyperparameters can strongly influence the convergence process. As shown in Fig. S7, we examine iteration step sizes of 32, 35, and 38, and observe that the corresponding convergence speeds differ noticeably. Although a larger step size can accelerate convergence, it cannot be increased arbitrarily. As illustrated in Fig. S7d, an excessively large step size causes pronounced oscillations after apparent convergence, leading to a higher standard deviation in the optimized results. This issue can be mitigated by adopting a dynamic learning rate. As shown in Fig. S7e, employing a properly designed learning-rate schedule that gradually decays with the number of iterations enables faster convergence while maintaining a low

standard deviation.

References

[1] Tao Xin, Liangyu Che, Cheng Xi, Amandeep Singh, Xinfang Nie, Jun Li, Ying Dong, and Dawei Lu. Experimental quantum principal component analysis via parametrized quantum circuits. *Phys. Rev. Lett.*, 126(11):110502, 2021.