

Supplemental Information for: Preparation of Metrological States in Dipolar Interacting Spin Systems

Tian-Xing Zheng^{1,2}, Anran Li¹, Jude Rosen¹, Sisi Zhou^{1,3}, Martin Koppenhöfer¹, Ziqi Ma^{4,5}, Frederic T. Chong⁴, Aashish A. Clerk¹, Liang Jiang¹, and Peter C. Maurer¹✉

¹*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

²*Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*

³*Institute for Quantum Information and Matter,*

California Institute of Technology, Pasadena, California 91125, USA

⁴*Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA*

⁵*Microsoft, Redmond, Washington 98052, USA and*

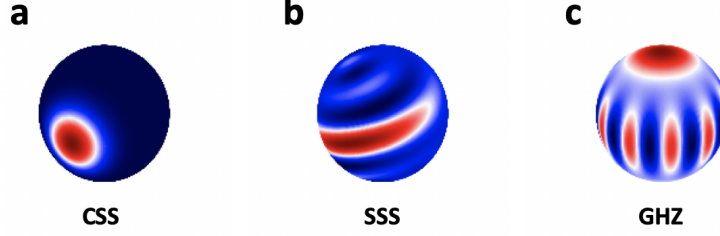
✉email: pmaurer@uchicago.edu

CONTENTS

Supplementary Figures	2
Supplementary Figure 1: Wigner function and standard metrological states	2
Supplementary Figure 2: Optimization results of different types of dipole-dipole interaction Hamiltonian	2
a. NV ensemble	3
b. Ising type interaction (large local disorder)	3
Supplementary Figure 3: Optimization results by using P_z^{tot} , P_z^π as measurement bases	4
Supplementary Figure 4: Mutual Information of the Metrological States	5
Supplementary Figure 5: Complete CFI data for Fig.2 in main text	6
Supplementary Figure 6: Required layers to reach given CFI for 2D square lattice	7
Supplementary Figure 7: Orders of interaction	7
Supplementary Figure 8: Non-Markovian noise sensing performance in 3D random spin configurations	8
Supplementary Figure 9: Optimized states with different readout fidelity	8
Supplementary Tables	9
Supplementary Table 1: Relative experimental parameter table (full)	9
Supplementary Notes	10
Supplementary Note 1: Designing the variational circuit	10
c. Entanglement generation gates from two-body interaction Hamiltonian and global rotations	10
d. Preservation of the collective spin direction	10
e. Choosing the most general gate set	11
Supplementary Note 2: CFI with respect to angle and frequency	11
Supplementary Note 3: Relation between CFI _{ω} and SNR in single qubit Ramsey experiment	13
Supplementary Note 4: Maximum Likelihood Estimator	14
Supplementary Note 5: Master equation for a non-Markovian environment	16
Supplementary Note 6: Performance of metrological states in a non-Markovian environment	19
Supplementary Note 7: Time Overhead	19
Supplementary Note 8: State preparation time comparing to adiabatic method	20
Supplementary Note 9: Numerically solving Schrödinger and Lindblad equations	21
Supplementary Discussion: Controllability	21
Controllability Test	22
Controllability of Dipolar Interacting Spin Systems	22
Finding Reachable States	23
Supplementary Methods	23
Large spin number ($N \rightarrow 100$) optimization: discrete truncated Wigner approximation (DTWA)	23
Supplementary References	25

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Wigner function and standard metrological states



Supplementary Figure 1. Three standard metrological states: Spin number $N = 10$. (a) Coherent Spin State (CSS). (b) Spin Squeezed States (SSS). Generated by one-axis twisting $H_{OAT} = \chi J_z^2$, $\chi t = 0.3$. (c) Greenberger-Horne-Zeilinger (GHZ) state.

The Wigner function of the spin system provides an intuitive visualization tool of the spin wavefunction. It maps the permutation symmetric subspace ($J = N/2$) of the N spin-1/2 system onto a quasiprobability distribution in the phase space by calculating the expectation value of the rotated parity operator:

$$W_\rho(\theta, \phi) = \text{Tr}[\rho R(\theta, \phi) M_0 R^\dagger(\theta, \phi)]. \quad (1)$$

The parity operator,

$$M_0 = \frac{2\pi}{J} \sum_{j=0}^{j=2J} \sqrt{\frac{2j+1}{4\pi}} T_{j0}, \quad (2)$$

Is the weighted sum of zeroth-order tensor operators $[T_{j0}]_{mm'} = \delta_{mm'} \sqrt{(2j+1)/(2J+1)} C_{Jm,j0}^{Jm}$. Here, $C_{Jm,j0}^{Jm}$ is a Clebsch-Gordan coefficient [1].

In Fig. 1, three standard metrological states are plotted. Coherent Spin State (CSS) is the classical state which reaches SQL and it's the suitable state for sensing when there is Markovian noise during the Ramsey time or when the experimental readout noise is high. Spin Squeezed States (SSS) beat the SQL by their squeezed variance of the spin angular momentum along the signal direction. Greenberger-Horne-Zeilinger (GHZ) states are able to reach the HL by utilizing the N fringes perpendicular to the equator to detect the external field simultaneously. All these standard metrological states lie in the permutation symmetric $J = N/2$ subspace, so the Wigner function plots in the main text are able to capture the essential features of the optimized metrological states.

Supplementary Figure 2: Optimization results of different types of dipole-dipole interaction Hamiltonian

The magnetic dipole-dipole interaction Hamiltonian under secular approximation has the general form [2, 3]:

$$H_{dd} = \sum_{i < j} V_{ij} (2S_{zi}S_{zj} - S_{xi}S_{xj} - S_{yi}S_{yj}) \quad (3)$$

with

$$V_{ij} = \sum_{i < j} \frac{\mu_0}{4\pi} \frac{\gamma_i \gamma_j \hbar^2}{|\mathbf{r}_i - \mathbf{r}_j|^3} \frac{(1 - 3 \cos \beta_{ij})}{2} \quad (4)$$

where μ_0 is the vacuum permeability, γ is the geomagnetic ratio of the spin, β_{ij} is the angle between the line segment connecting ($\mathbf{r}_i, \mathbf{r}_j$) and the direction of the bias external magnetic field (along z-direction in this case). Eq. (3) is able to describe the dipolar interaction for the spin systems with an arbitrary spin number as long as the spin angular momentum operators S_μ obey the commutation relation $[S_i, S_j] = i\epsilon_{ijk} S_k$. It applies to the spin-1/2 systems we discussed in the main text and Nitrogen-Vacancy (NV) centers which are spin-1 systems.

a. NV ensemble

Here we consider NV ensemble and only $|m_s = 1\rangle$ and $|m_s = 0\rangle$ are used as a 2-level system. The spin-1 operators are

$$S_x^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, S_y^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, S_z^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (5)$$

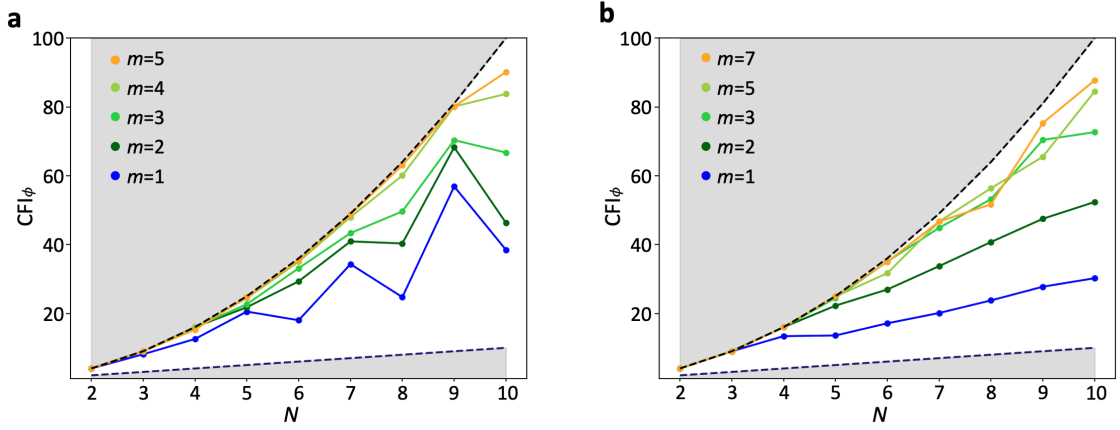
If we only take the $|m_s = 1\rangle, |m_s = 0\rangle$ subspace into consideration, the relations between the ‘truncated’ spin-1 operators and the spin-1/2 operators are:

$$S_y^{(1)} = \sqrt{2}S_y^{(\frac{1}{2})}, S_x^{(1)} = \sqrt{2}S_x^{(\frac{1}{2})}, S_z^{(1)} = \frac{I}{2} + S_z^{(\frac{1}{2})}. \quad (6)$$

Plugging Eq. (6) into Eq. (3), we get the effective dipole-dipole interaction Hamiltonian for NV ensemble $|m_s = 1\rangle, |m_s = 0\rangle$ subspace [2,4]:

$$H_{DD,NV} = \sum_{i < j} V_{ij} (S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{xi}^{(\frac{1}{2})} S_{xj}^{(\frac{1}{2})} - S_{yi}^{(\frac{1}{2})} S_{yj}^{(\frac{1}{2})}). \quad (7)$$

Fig. 2(a) shows the Classical Fisher Information (CFI) optimization results for 2D square lattice spin configuration. They are similar to the results we get in Fig.(2) of the main text for spin-1/2 systems.



Supplementary Figure 2. Different types of dipolar interactions: (a) Optimization results for NV-ensemble. The CFI saturates the theoretical upper bound Heisenberg Limit (HL) when the variational circuit layer number goes up from 1 to 7, and the CFI results are ‘oscillating’ for even/odd number of spins from shallow circuit. (b) Optimization results for Ising type spin interaction when there is large local disorder in the system.

b. Ising type interaction (large local disorder)

When the system has large local disorder, the flip-flop terms in the dipolar interaction Hamiltonian Eq. (3) are suppressed because of the large energy gap:

$$\begin{aligned} H_{DD,Ising} &= \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i < j} V_{ij} (2S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{xi}^{(\frac{1}{2})} S_{xj}^{(\frac{1}{2})} - S_{yi}^{(\frac{1}{2})} S_{yj}^{(\frac{1}{2})}) \\ &= \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i < j} 2V_{ij} (S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{+i}^{(\frac{1}{2})} S_{-j}^{(\frac{1}{2})} - S_{-i}^{(\frac{1}{2})} S_{+j}^{(\frac{1}{2})}) \\ &\approx \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i < j} 2V_{ij} S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})}. \end{aligned} \quad (8)$$

This location-dependent single-spin energy shift (δ_i) can be canceled by spin-echo pulse sequence where the interaction gate $D(\tau)$ needs to be applied:

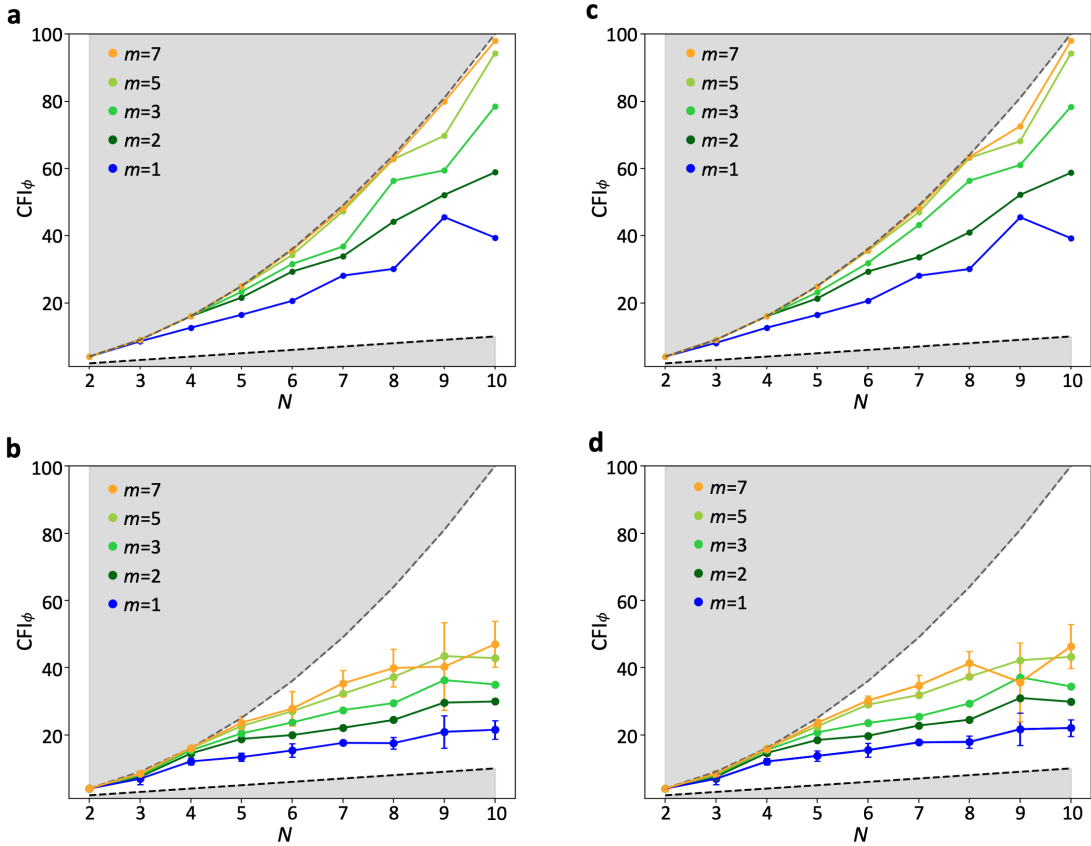
$$\begin{aligned} D(\tau) &= R_x(\pi) \exp[-i\tau H_{\text{DD,Ising}}] R_x(\pi) \exp[-i\tau H_{\text{DD,Ising}}] \\ &= \exp \left[-i\tau \sum_{i < j} 2V_{ij} S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} \right]. \end{aligned} \quad (9)$$

Eq. (9) is also valid when the local disorder δ_i is comparable with the interaction strength V_{ij} . If there is local disorder in the dipolar-interacting spin ensemble, applying spin-echo will generate the interaction gate $D(\tau)$ where the local disorder terms are canceled.

The CFI optimization results by using the effective Ising type interaction Hamiltonian $H_{\text{DD,Ising}} = \sum_{i < j} 2V_{ij} S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})}$ is shown in Fig. 2 (b).

From Fig. 2, the CFI results close to the Heisenberg Limit are observed, indicating that the variational method can be applied to different kinds of spins in solid state systems and generate highly entangled state for high-precision quantum metrology. We also observe that for shallow variational circuits, the CFI ‘oscillation’ between even and odd spin numbers only appears when there are flip-flop terms in the Hamiltonian. For Ising type interaction, the ‘oscillation’ disappears.

Supplementary Figure 3: Optimization results by using P_z^{tot} , P_z^π as measurement bases



Supplementary Figure 3. Optimized CFI with different measurement bases: (a) 2D square lattice using observable P_z^{tot} , (b) 3D-random configuration using observable P_z^{tot} (averaged over 5 cases), (c) 2D square lattice using observable P_z^π , (d) 3D-random configuration using observable P_z^π (averaged over 5 cases). All error bars indicate one standard deviation of the CFI from different spin configurations.

The optimization results shown in Fig.2 in the main text are obtained by using P_z as the measurement basis for the CFI (cost function) calculation. Although measuring all the diagonal elements in the density matrix of the resulting

states provides the maximum information one can get from single-qubit measurement and a large Hilbert space for the optimizer, it leads to an exponentially large (2^N) experimental repetition number when the CFI needs to be estimated from experimental data. Thus, we test the variational method on two other measurement bases which require less repetitions for readout.

The measurement basis on total spin polarization along z -direction is given by

$$P_z^{\text{tot}} \equiv |J = N/2, J_z\rangle \langle J = N/2, J_z|, \quad (10)$$

where J is the total spin angular momentum quantum number and J_z is the total spin angular momentum projection quantum number that runs from $N/2$ to $-N/2$. P_z^{tot} has $N + 1$ outcomes, so it scales linear with the system size.

The optimization results by using the CFI on P_z^{tot} as cost function are shown in Fig. 3 (a)(b). Surprisingly, compared to the results by using P_z , the optimization results from using P_z^{tot} are improved by about a factor of 1.3 for the 3D-random spin configuration. Since all the information one can extract from P_z^{tot} are contained in P_z , we attribute this improvement to the simpler parameter space structure that P_z^{tot} provides to the optimizer. Less local maximum points in the parameter space will help the optimizer to converge to a high CFI point, especially when the dimension of the parameter space ($3m$) is large.

Parity of the spin ensemble,

$$P_z^\pi \equiv \prod_i^N \sigma_{zi}, \quad (11)$$

provides a constant (2) dimensional outcome space for experimental readout. Improvements are also observed in 2D square lattice and 3D-random spin configurations (Fig. 3(c)(d)).

Supplementary Figure 4: Mutual Information of the Metrological States

The Rényi entropy of order n of a subsystem A is defined as

$$S_n(A) = \frac{1}{1-n} \log \text{Tr}(\rho_A^n) \quad (12)$$

which describes the general entanglement correlation between the subsystem A and the full system [5,6]. In the limit $n \rightarrow 1$, the Renyi entanglement entropy approaches the von Neumann entanglement entropy $E_{\text{vN}} = -\text{Tr}(\rho_A \log \rho_A)$, which we used to characterize the entanglement of the metrological states in Fig.3 of the main text. The 2-nd order Rényi entropy reflects the purity of the state $\text{Tr}(\rho_A^2) \leq 1$. It is an indication of entanglement when the purity of the subsystem is smaller than the purity of the full system [7], $\text{Tr}(\rho_A^2) < \text{Tr}(\rho_{\text{AB}}^2)$.

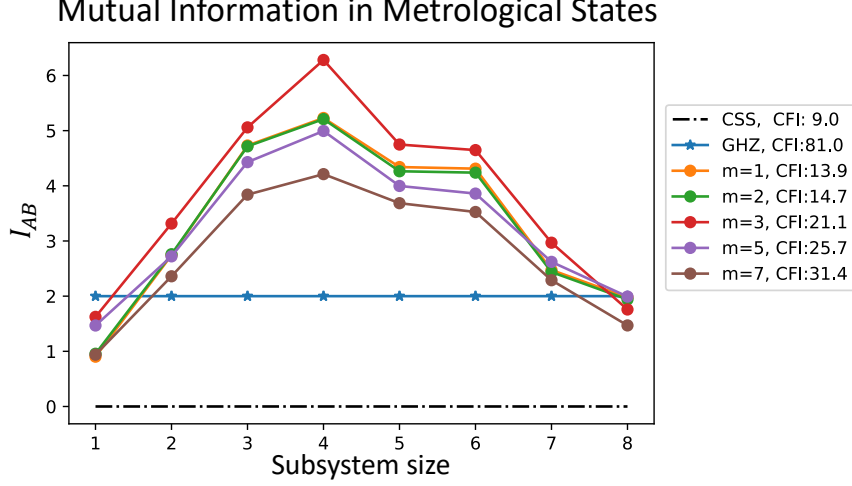
By measuring the 2-nd order Rényi entropy, we can calculate the mutual information between the two subsystems A and B.

$$I_{\text{AB}} = S_2(A) + S_2(B) - S_2(\text{AB}). \quad (13)$$

The scaling of the mutual information with respect to the subsystem size is one of the keys to study the entanglement properties of interacting quantum systems, e.g. area law and volume law [8,7,9]. Here, we numerically calculate the mutual information of the 9-spin metrological states generated by the variational circuits with different layer numbers.

As shown in Fig. 4, unlike CFI, the mutual information does not monotonically increase with the circuit layer number m . If entanglement is quantified by mutual information, all the metrological states we studied here have ‘more’ entanglement compared to the GHZ states, even though the CFI are lower. These results reflect that deep entanglement does not necessarily lead to better sensing performance (CFI), global and shallow entanglement is preferred for Heisenberg-limit quantum sensing.

We can understand the relation between CFI and entanglement by taking a 4-qubit GHZ state $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$ as an example. $|\text{GHZ}\rangle$ has the highest CFI. However, there is only 1 ebit of entanglement between the two qubits on the left and the two on the right. $S_2(A) = S_2(B) = 1, I_{\text{AB}} = 2$. Consider another state $|\psi\rangle = \frac{1}{2}(|0000\rangle + |1111\rangle + |0101\rangle + |1010\rangle)$. The amount of entanglement between the left and right two qubits are 2 ebits. $S_2(A) = S_2(B) = 2, I_{\text{AB}} = 4$. The last two terms in $|\psi\rangle$ contribute negatively to the CFI of the state, even though they provide more entanglement.



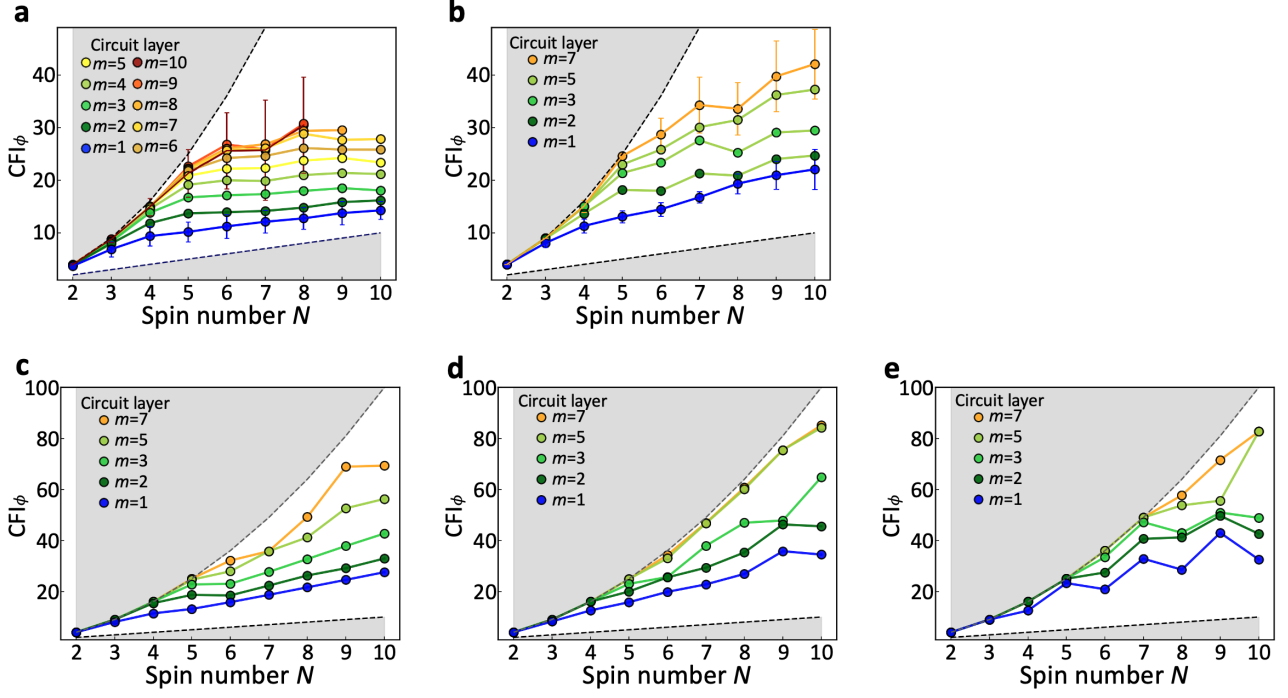
Supplementary Figure 4. Mutual information versus subsystem size of the 9-spin metrological states:

The spatial spin configuration is the same as Fig.3(b) (case No.2 in 50 random cases) and the $m = 1-7$ results are calculated (including the states that shown in Fig.3(b) left/right in the main text). The asymmetry of the mutual information with respect to the subsystem size comes from the random spin locations.

Supplementary Figure 5: Complete CFI data for Fig.2 in main text

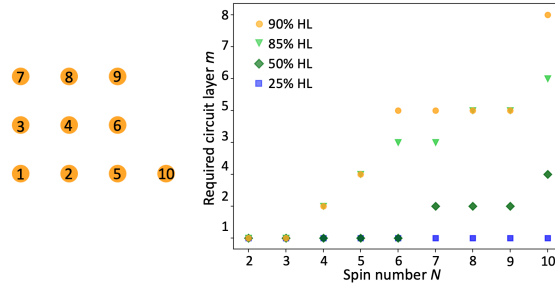
The complete data for dipolar-interacting spin systems' CFI optimization is shown in this section. Fig. 5(a) shows the 50-cases averaged optimization results for 3D-random spin configurations, the variational circuit layer number m is chosen from 1 to 10. The optimized CFI results are approaching to the Heisenberg Limit (HL) when more layers (m) are used. However, when $m > 7$, the CFI results stop increasing. This CFI 'saturation' effect might be caused by two reasons. First, when m is large, the number of the local maximum points in the high dimensional parameter space increases. This could potentially cause the optimizer to stuck in the local maximum point. Sometimes, take $N = 7, m = 10$ data in Fig. 5(a) as an example, adding more variational layers even leads to a lower CFI optimization result. The 'local maximum' problem could be solved by more advanced and powerful optimization algorithms, such as reinforcement learning [10–12], and more computational resources. Second, the 'saturation' effect reflects the global maximum CFI one can reach, no matter what kind of optimization algorithm is applied. It's still an open question what is the highest CFI the spin ensemble could reach for a given configuration.

Fig. 5(b)-(e) show the CFI optimization result for 2D random (10-cases average), 1D chain, 2D square lattice and 2D symmetric cycle spin configurations. The results of regular patterns are better than those of 2D/3D-random patterns. Due to the angular dependent term $(1 - 3\cos^2\theta)$ in dipolar interaction, the results from 2D-random spin configuration (Fig. 5(b)) are better than the results from 3D-random patterns (Fig. 5(a)).



Supplementary Figure 5. The complete CFI data: (a) 3D random, (b) 2D random, (c) 1D chain, (d) 2D square lattice, and (e) 2D circle. All error bars indicate one standard deviation of the CFI from different spin configurations.

Supplementary Figure 6: Required layers to reach given CFI for 2D square lattice



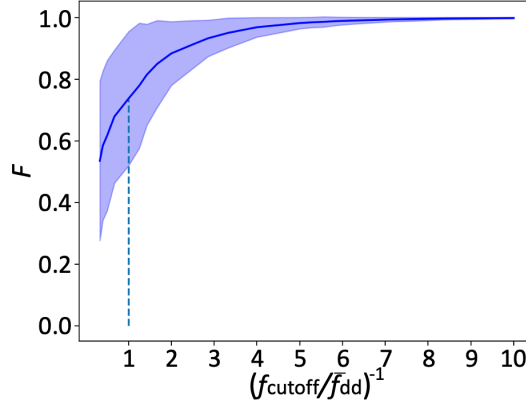
Supplementary Figure 6. Scalability of 2D square spin lattice: Left: Schematic of a 2D square lattice pattern. The numbers label the order in which spins are added to form a lattice of size N . Right: Number of layers required to achieve a CFI within a given percentage of the HL.

As shown by the schematic on the left, the distances between spin No.4 and spin No.5, 7, and 9 are the same, so the interaction strengths between each pair are the same. Similarly, the distance between spin No.4 and spin No.2, 3, 6, and 8 are the same (smaller). Therefore, the plateau features in Fig.(6) are likely due to this symmetry: adding one more spin to the lattice does not require an extra layer to reach a given percentage of the CFI.

Supplementary Figure 7: Orders of interaction

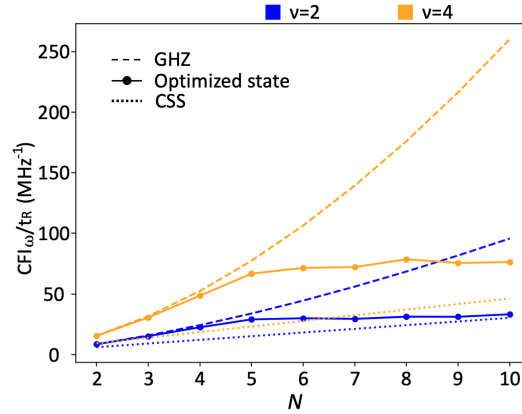
Due to the decaying feature ($\frac{1}{r^3}$) of dipolar interaction strength, the resulting states might be mainly generated by nearest-neighbor interaction. For studying ‘how much’ interaction is essential for generating the resulting entangled states, we calculate the overlap (state fidelity [13]) between the original state and the new state, which is generated

by using the cutoff Hamiltonian and optimized parameters. A cutoff interaction strength f_{cutoff} is chosen, and all the pairwise potential V_{ij} smaller than f_{cutoff} are set to zero in the cutoff Hamiltonian. Fig. 7 shows the relation between the state fidelity F versus f_{cutoff} . A state fidelity value less than 1 is observed when f_{cutoff} is set to be equal to the averaged nearest-neighbor interaction strength f_{dd} . This result reflects higher order interactions in the spins ensemble are utilized for the metrological states generation.



Supplementary Figure 7. Long range interaction for preparing metrological states: Average state fidelity vs. different cutoff strength in H_{dd} . The shaded area indicates the standard deviation. Data obtained from 3D-random $N = 10, m = 5$, 50-cases optimization results.

























Supplementary Figure 8: Non-Markovian noise sensing performance in 3D random spin configurations



Supplementary Figure 8. Sensing performance of 3D random spin configurations: Average Ramsey protocol's results of the generated entangled states in 3D random configurations when considering non-Markovian noise in the signal accumulation step. Blue and orange correspond to two different noise models ($\nu = 2$ and 4).

Supplementary Figure 9: Optimized states with different readout fidelity

We run the optimization with imperfect readout for $N = 4$ and $N = 10$ 2D square lattice spin configurations. The optimized states resemble GHZ states (high RF), SSS (low RF), CSS (RF close to 50%). For $N = 4$ case, the Gaussian state appears for RF lower than 92%, but for $N = 10$ case, the Gaussian states appears when RF is about 96%. We expected that for large spin system with finite RF, Gaussian states (e.g. SSS) are advantageous for quantum-enhanced metrology.

a	Wigner Plot								
	RF	1.00	0.99	0.98	0.97	0.96	0.95	0.94	0.93
	CFI	15.27	11.07	9.27	7.83	6.91	6.03	4.81	4.62
	CSS	4.00	3.84	3.67	3.53	3.39	3.24	3.10	2.96
	Spin number N = 4:								
	Wigner Plot								
	RF	0.92	0.91	0.90	0.80	0.70	0.60	0.50	
	CFI	3.83	3.54	3.28	1.56	0.65	0.16	10 ⁻³²	
	CSS	2.82	2.69	2.56	1.44	0.64	0.16	10 ⁻³³	
b	Wigner Plot								
	RF	1.00	0.98	0.96	0.94	0.92			
	CFI	85.18	30.18	18.18	14.34	11.66			
	CSS	10.00	9.216	8.464	7.744	7.056			
	Spin number N = 10:								
	Wigner Plot								
	RF	0.9	0.80	0.70	0.60				
	CFI	9.67	4.166	1.654	0.401				
	CSS	6.40	3.6	1.6	0.4				

Supplementary Figure 9. Highly entangled state requires high readout fidelity: Optimized states' Wigner distributions when finite readout fidelity (RF) is assumed in the optimization process. The 'CSS' row reports the CFI value when using the CSS for sensing under same RF.

SUPPLEMENTARY TABLES

Supplementary Table 1: Relative experimental parameter table (full)

Supplementary Table 1. Experimental platforms' data

System	$T_2^{(\text{best})}$	$T_2^{(\text{DD})}$	\bar{f}_{dd}	P_{ini}	F_{readout}	ν
NV ensemble	1.58(7)s ^a	7.9(2) μ s ^b	35kHz ^b	97.5% ^c	97.5% ^c	2 – 4 ^b
P1 centers	0.8ms ^e (DEER)	4.4 μ s ^f	0.7kHz ^e , 0.92MHz ^f	95% ^e	95% ^e	?
Rare-Earth crystals	23.2 \pm 0.5ms ^g	2.5 μ s ^h	1.96MHz ^h	97% ⁱ	94.6% ^j	2.4 \pm 0.1 ^g
Cold Molecules	1s ^k	80ms ^l	52Hz ^l , 1.5kHz ^m	97% ^m	97% ^m	?

^a T.H.Taminiau, NComm 2018, ^b H.Zhou, PRX 2020, ^c M.D.Lukin, PRL 2015, ^d L.Childress Science 2006

^e T.H.Taminiau, NComm 2021, ^f N.Yao, Nature 2021

^g P.Bertet, Science advances 2021, ^h A.Reiserer, PRL 2021, ⁱ J.Thompson, Science 2020, ^j J.Thompson, NComm 2020

^k M.R Tarbutt PRL 2020, ^l B.Yan, J.Ye, Nature 2013, ^m J Doyle, PRL 2020

Based on the simulation results shown in Fig.4(c) in main text, we need $\bar{f}_{\text{dd}}T_2 \geq 5$ to generate metrological states that beat the SQL. It's worth mentioning that the T_2 in this situation stands for the coherence time *without* the dipole-dipole interaction's influence. During the state preparation step, the dipolar interactions between the spins are included in the system Hamiltonian for the entanglement generation (D gate in Fig.1(c) in main text). Thus, the $T_2^{(\text{DD})}$ in Table 1 is a lower bound and $T_2^{(\text{best})}$ is a more precise estimation for the spin coherence time.

SUPPLEMENTARY NOTES

Supplementary Note 1: Designing the variational circuit

In this section, we discuss how to choose the experimentally realizable elementary gates in the variational sequence of the entangler based on limited quantum resource [14, 15].

c. Entanglement generation gates from two-body interaction Hamiltonian and global rotations

Consider a two-body interaction Hamiltonian:

$$H_{\text{int}} = \sum_{i < j} V_{ij} (J^I S_{zi} S_{zj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j). \quad (14)$$

In this Hamiltonian, $\mathbf{S} = (S_x, S_y, S_z)$ is the vector of spin-1/2 operators, V_{ij} is the interaction strength between spin i and j which depends on their locations, and $J^I (\neq 0)$, J^S are the Ising and symmetric coupling constant respectively.

The elementary gates in each layer of the variational circuit for preparing metrological states (Fig.1(c) main text) include two free evolutions under the interaction Hamiltonian $D(\tau), D(\tau')$, one global rotation along the x -axis $R_x(\vartheta)$ and two fixed $\pi/2$ rotations $R_y(-\frac{\pi}{2}), R_y(\frac{\pi}{2})$ along the y -axis. We define the interaction gate in the z -direction as

$$D_z(\tau) \equiv \exp(-i\tau H_{\text{int}}/\hbar) = \exp \left[-i\tau \sum_{i < j} V_{ij} (J^I S_{zi} S_{zj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j) / \hbar \right]. \quad (15)$$

The interaction gates in other directions can be obtained by $\pi/2$ rotations:

$$\begin{aligned} D_{x,y}(\tau) &= R_{y,x}(\pi/2) D_z(\tau) R_{y,x}(-\pi/2) \\ &= \exp \left[-i\tau \sum_{i < j} V_{ij} (J^I S_{x,yi} S_{x,yj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j) / \hbar \right]. \end{aligned} \quad (16)$$

In Eqs. (15) (16), the symmetric interaction term stay unchanged because inner product is conserved under global rotation and the ‘direction of interaction’ is only determined by the Ising term. Using these definitions, we simplify the gate set in each layer as

$$\begin{aligned} \mathcal{U}_i &= R_y \left(\frac{\pi}{2} \right) D(\tau'_i) R_y \left(-\frac{\pi}{2} \right) R_x(\vartheta_i) D(\tau_i) \\ &= D_x(\tau'_i) R_x(\vartheta_i) D_z(\tau_i). \end{aligned} \quad (17)$$

In the next two subsections, it will be shown that the sequence in Eq. (17) is the most general gate set that uses only global rotations and preserves the collective spin direction along x -direction.

d. Preservation of the collective spin direction

Define the x -parity operator $P_x \equiv \prod_i^N \sigma_{xi} = P_x^\dagger$, with $P_x^2 = I$. This operator describes the parity of a state in x -direction and is related to the global π rotation along x -axis up to a phase constant, $R_x(\pi) = \exp(-i\pi \sum_i S_{xi}) = (-i)^N \prod_i^N \sigma_{xi}$. Applying the x -parity operator onto individual spin’s angular momentum operator gives $P_x S_{\mu j} P_x = (\sigma_x S_\mu \sigma_x)_j = \pm S_{\mu j}$. Thus the interaction gates along x - and z -direction conserve the x -parity, $P_x D_{x,z} P_x = D_{x,z}$. Similarly, the only global rotation that conserves x -parity for arbitrary angles is $R_x(\vartheta)$. Then, based on Eq.(1) in the main text, the unitary operator of the whole control sequence conserves the x -parity

$$\begin{aligned} P_x \mathcal{S}(\boldsymbol{\theta}) P_x &= P_x \mathcal{U}_m \dots \mathcal{U}_2 \mathcal{U}_1 P_x \\ &= P_x \Pi_i [D_x(\tau'_i) R_x(\vartheta_i) D_z(\tau_i)] P_x \\ &= \mathcal{S}(\boldsymbol{\theta}). \end{aligned} \quad (18)$$

The initial spin state pointing to the $+x$ -direction is an eigenstate of P_x : $P_x |\uparrow_x\rangle^{\otimes N} = |\uparrow_x\rangle^{\otimes N}$. Thus, any state produced by this variational circuit remains an eigenstate of P_x :

$$\begin{aligned} P_x |\Psi(\theta)\rangle &= P_x \mathcal{S}(\theta) |\uparrow_x\rangle^{\otimes N} \\ &= P_x \mathcal{S}(\theta) P_x P_x |\uparrow_x\rangle^{\otimes N} \\ &= |\Psi(\theta)\rangle. \end{aligned} \quad (19)$$

Now consider the expectation value of the total spin angular momentum operator $J_\mu \equiv \sum_i S_{\mu i}$ ($\mu \in \{x, y, z\}$):

$$\begin{aligned} \langle J_{y,z} \rangle &= \langle \Psi(\theta) | J_{y,z} | \Psi(\theta) \rangle \\ &= \langle \Psi(\theta) | P_x P_x J_{y,z} P_x P_x | \Psi(\theta) \rangle \\ &= -\langle J_{y,z} \rangle = 0. \end{aligned} \quad (20)$$

Thus, the collective spin direction $\langle \mathbf{J} \rangle / |\langle \mathbf{J} \rangle|$ always points along the x -direction.

e. Choosing the most general gate set

To preserve the collective spin direction along x -axis, the global rotation and interaction gates that can be chosen are R_x , D_x , D_\perp where D_\perp stands for the interaction gates along any direction perpendicular to the x -direction. Combining R_x and D_z can generate any D_\perp , thus the simplest gate set fulfilling all the requirements is $D_x R_x D_z$, as described by Eq.(1) in the main text.

The derivations and results in this section about selecting the variational sequence agree with ref.[14]. However, the interaction Hamiltonian we discuss here is more general. In Eq. (14), when $J^I = 1, J^S = 0$, the interaction becomes Ising type interaction which is equivalent to the Rydberg interaction in ref.[14,15]. The Ising interaction can also describe spin systems with large local disorder. The optimization results are shown in the next section. When $J^I = 3, J^S = -1$, Eq. (14) becomes the dipolar interaction Hamiltonian between spin-1/2 particles. When $J^I = 2, J^S = -1$, it becomes the dipolar interaction Hamiltonian between spin-1 particles (such as NV centers). The simulation results for this case are shown in the next section. When $J^I = 1, J^S = -1$, the interaction can describe the dipolar interaction between cold molecules [16].

Supplementary Note 2: CFI with respect to angle and frequency

In general, the Classical Fisher Information (CFI) measures the sensitivity of a statistical model to small changes of a parameter θ [17, 18]. Let Z be a random variable and $P_z(\theta) \equiv P(z|\theta)$ be its probability distribution which depends on θ . Let Θ be an unbiased estimator of θ , i.e.

$$\theta = \langle \Theta \rangle = \sum_z \Theta \cdot P_z(\theta). \quad (21)$$

From Eq. (21) and the fact that the sum of probabilities of all outcomes is 1,

$$1 = \frac{\partial \langle \Theta \rangle}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_z \Theta P_z(\theta), \quad (22)$$

$$0 = \frac{\partial}{\partial \theta} \sum_z P_z(\theta). \quad (23)$$

Subtracting Eq. (23) multiplied by θ from Eq. (22), we get

$$\begin{aligned} 1 &= \sum_z (\Theta - \theta) \frac{\partial}{\partial \theta} P_z(\theta) \\ &= \sum_z P_z(\theta) (\Theta - \theta) \frac{1}{P_z(\theta)} \frac{\partial}{\partial \theta} P_z(\theta) \\ &= \langle (\Theta - \theta) \frac{1}{P_z(\theta)} \frac{\partial}{\partial \theta} P_z(\theta) \rangle \\ &= \langle (\Theta - \theta) \frac{\partial}{\partial \theta} \log P_z(\theta) \rangle. \end{aligned} \quad (24)$$

Letting $X = \Theta - \theta$ and $Y = \frac{\partial}{\partial \theta} \log P_z(\theta)$, by the Cauchy-Schwartz inequality for random variables: $\langle XY \rangle^2 \leq \langle X^2 \rangle \langle Y^2 \rangle$, we have

$$\langle (\Theta - \theta)^2 \rangle \left\langle \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2 \right\rangle \geq 1, \quad (25)$$

where

$$\begin{aligned} \langle (\Theta - \theta)^2 \rangle &= \langle \Theta^2 \rangle - (2\theta \langle \Theta \rangle - \langle \theta^2 \rangle) \\ &= \langle \Theta^2 \rangle - (2\theta^2 - \theta^2) \\ &= \langle \Theta^2 \rangle - \langle \Theta \rangle^2 \\ &= \Delta \Theta^2 \end{aligned} \quad (26)$$

is the variance of Θ . Define

$$\text{CFI} \equiv \sum_z P_z(\theta) \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2, \quad (27)$$

we have

$$\Delta \Theta^2 \geq \frac{1}{\text{CFI}}. \quad (28)$$

If the measurement is repeated M times, then by the additive property of CFI, we obtain the Cramér-Rao bound:

$$\Delta \Theta^2 \geq \frac{1}{M \cdot \text{CFI}}. \quad (29)$$

In our variational circuit, we use CFI with respect to an infinitesimal angle ϕ as the cost function to generate entangled states. In our program, we use a method similar to parameter shift to calculate the CFI_ϕ of our optimized states [19,20,21]. In the following notation,

1. z represents a multi-qubit state in the z -basis;
2. $\mathcal{U}(\phi) = e^{-i\phi J_y}$ is the rotation operator where ϕ is a small angle;
3. ψ is the state we create from the variational circuit;
4. $P_z(\theta)$ is the probability of measuring the state z with the state after rotation.

Then

$$\begin{aligned} \frac{\partial}{\partial \phi} P_z(\phi) \Big|_{\phi \rightarrow 0} &= \frac{\partial}{\partial \phi} |\langle z | \mathcal{U}(\phi) | \psi \rangle|^2 \Big|_{\phi \rightarrow 0} \\ &= \frac{\partial}{\partial \phi} \langle \psi | \mathcal{U}^\dagger(\phi) | z \rangle \langle z | \mathcal{U}(\phi) | \psi \rangle \Big|_{\phi \rightarrow 0} \\ &= \langle \psi | \mathcal{U}^\dagger(\phi) i J_y | z \rangle \langle z | \mathcal{U}(\phi) | \psi \rangle \Big|_{\phi \rightarrow 0} + \langle \psi | \mathcal{U}^\dagger(\phi) | z \rangle \langle z | \mathcal{U}(\phi) (-i) J_y | \psi \rangle \Big|_{\phi \rightarrow 0} \\ &= i \langle \psi | \left(J_y | z \rangle \langle z | - | z \rangle \langle z | J_y \right) | \psi \rangle. \end{aligned} \quad (30)$$

Note that assuming the rotation operator $\mathcal{U}(\phi) = e^{-i\phi J_y} \equiv \mathcal{U}_y(\phi)$ along y -axis is for calculation simplicity. In experiments, the signal (e.g. the external B-field) usually induces a rotation along z -axis, $\mathcal{U}_z(\phi) = e^{-i\phi J_z}$. It's equivalent to assume that the prepared state is firstly rotated by a $R_x(\pi/2)$ pulse and then accumulates a signal ϕ along y -axis, or firstly accumulates a signal along z -axis and then rotated by $R_x(-\pi/2)$ pulse [18]. In another word, $R_x(-\pi/2)\mathcal{U}_z(\phi) = \mathcal{U}_y(\phi)R_x(\pi/2)$, so the signal accumulation process we assumed in the calculation is able to simulate the experiments.

After creating the entangled states, we want to know how useful they are in a Ramsey spectroscopy, where the signal we want to detect is a frequency ω . By the same calculation as above except the difference that we take derivative with respect to $\omega = \frac{\phi}{t_R}$ where t_R is the Ramsey sensing time, we have

$$\text{CFI}_\omega = \text{CFI}_\phi \cdot t_R^2. \quad (31)$$

Supplementary Note 3: Relation between CFI_ω and SNR in single qubit Ramsey experiment

We illustrate the Ramsey protocol for a single qubit.

1. The qubit is initialized into the ground state $|0\rangle$.
2. A $\frac{\pi}{2}$ pulse along the y-direction is applied to transform it into a superposition state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Its matrix form is

$$\rho(t) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (32)$$

3. After evolving under noise and a signal with frequency ω for time t , its state becomes

$$\rho(t) = \frac{1}{2} \begin{pmatrix} 1 & e^{-i\omega t} e^{-2\gamma t} \\ e^{i\omega t} e^{-2\gamma t} & 1 \end{pmatrix} \quad (33)$$

where γ is the decoherence rate.

4. A second $\frac{\pi}{2}$ pulse along the x-direction is applied for readout. The qubit is then in the state

$$R_x\left(\frac{\pi}{2}\right) \rho(t) R_x^\dagger\left(\frac{\pi}{2}\right) \quad (34)$$

5. After the rotation, the probability of the qubit being in the ground state is

$$P_0 = \frac{1}{2} + \frac{1}{2} e^{-2\gamma t} \sin \omega t. \quad (35)$$

The CFI with respect to ω is

$$\text{CFI}_\omega = \frac{1}{P_0} \left(\frac{\partial P_0}{\partial \omega} \right)^2 + \frac{1}{P_1} \left(\frac{\partial P_1}{\partial \omega} \right)^2 = \frac{t^2 \cos^2 \omega t}{e^{4\gamma t} - \sin^2 \omega t}. \quad (36)$$

Assuming only quantum projection noise, the Signal-to-Noise Ratio (SNR) is $\frac{\delta P_0}{\sqrt{\frac{1}{M} P_0(1-P_0)}}$ where M is the total number of measurements. Then

$$\text{SNR}^2 = \frac{M t^2 \cos^2 \omega t \delta \omega^2}{e^{4\gamma t} - \sin^2 \omega t}. \quad (37)$$

Assuming no time overhead, i.e., $M = \frac{T_{\text{tot}}}{t_R}$ where T_{tot} is the total measurement time and t_R is the time between Ramsey pulses, we obtain the relationship

$$\text{CFI}_\omega \cdot \frac{T_{\text{tot}}}{t_R} \cdot \delta \omega^2 = \text{SNR}^2. \quad (38)$$

In unit time ($T_{\text{tot}} = 1\text{s}$), when $\text{SNR} = 1$, the smallest signal we can measure is

$$\delta \omega = \frac{1}{\sqrt{M \cdot \text{CFI}_\omega}}, \quad (39)$$

leading to the saturated Cramér-Rao bound.

Supplementary Note 4: Maximum Likelihood Estimator

Since a measurement collapses a quantum state to an eigenstate of the observable, it's impossible to directly measure $P(\theta)$. In experiments, we repeat the sequence to obtain the results for estimating the $P(\theta)$ and then get an estimate value of θ . To understand the relation between the variance of the estimation and CFI, we introduce the Maximum Likelihood Estimator (MLE), which has asymptotic properties to saturate the Cramér-Rao bound. An experimental implementation of MLE to estimate the phase shift of an entangled four-photon Dicke state is given in [22]. A proof that MLE is an unbiased estimator which asymptotically saturates the Cramér-Rao bound is given in [23]. We summarize the proof below.

Let $\mathbf{X} = \{X_1, X_2, \dots, X_M\}$ be a collection of independent and identically distributed (i.i.d.) random variables with a parametric family of probability distributions $\{P(X|\theta)|\theta \in \Theta\}$, where θ is an unknown parameter and Θ is the parameter space. Let $\mathbf{x} = \{x_1, x_2, \dots, x_M|x_i \in X_i\}$ be the experimental data set from M repetitions. The goal is to estimate θ (the signal we want to measure) from \mathbf{x} , i.e., find θ that is most likely to produce the outcome \mathbf{x} . Thus, the normalized log-likelihood function is defined as

$$L_M(\theta) = \frac{1}{M} \log P(\mathbf{X}|\theta) = \frac{1}{M} \log \prod_{i=1}^M P(X_i|\theta) = \frac{1}{M} \sum_{i=1}^M \log P(X_i|\theta). \quad (40)$$

A MLE maximizes the log-likelihood function

$$\Theta_{\text{MLE}} = \underset{\theta \in \Theta}{\operatorname{argmax}} L_M(\theta). \quad (41)$$

In the following, we first show that

1. Θ_{MLE} converges to the true parameter θ_0 ;
2. the distribution of $\sqrt{M}(\Theta_{\text{MLE}} - \theta_0)$ tends to a normal distribution $\mathcal{N}\left(0, \frac{1}{\text{CFI}_{\theta_0}}\right)$ as M increases.

In other words, not only does the MLE converge to the true parameter, it converges at a rate $\frac{1}{\sqrt{M}}$.

Define

$$L(\theta) = \langle \log P(\mathbf{X}|\theta) \rangle_{\theta_0} \quad (42)$$

which denotes the expected log-likelihood function with respect to θ_0 , then by the Weak Law of Large Numbers (WLLN), the average outcomes from a large number of trials should approach the expected value:

$$\forall \theta, L_M(\theta) \xrightarrow{M \rightarrow \infty} L(\theta). \quad (43)$$

In fact, θ_0 maximizes $L(\theta)$:

$$\begin{aligned} \forall \theta, L(\theta) - L(\theta_0) &= \langle \log P(\mathbf{X}|\theta) \rangle_{\theta_0} - \langle \log P(\mathbf{X}|\theta_0) \rangle_{\theta_0} \\ &= \left\langle \log \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} \\ &\leq \left\langle \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} - 1 \right\rangle_{\theta_0} \\ &= \sum_{\mathbf{x} \in \mathbf{X}} \left(\frac{P(\mathbf{x}|\theta)}{P(\mathbf{x}|\theta_0)} - 1 \right) P(\mathbf{x}|\theta_0) \\ &= 1 - 1 = 0. \end{aligned} \quad (44)$$

Moreover, we show that θ_0 is the unique maximizer. Jensen's inequality states that for a strictly convex function f and a random variable Y ,

$$\langle f(Y) \rangle > f(\langle Y \rangle). \quad (45)$$

Taking $f(y) = -\log y$ and $P(\mathbf{X}|\theta) \neq P(\mathbf{X}|\theta_0)$, we have

$$\left\langle -\log \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} > -\log \left\langle \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} = 0, \quad (46)$$

or

$$L(\theta_0) > L(\theta). \quad (47)$$

Therefore, since

1. Θ_{MLE} maximizes $L_M(\theta)$,
2. θ_0 maximizes $L(\theta)$, and
3. $L_M(\theta) \xrightarrow{M \rightarrow \infty} L(\theta)$,

Θ_{MLE} converges to θ_0 .

Now we use this property to prove that the distribution of Θ_{MLE} tends to the desired normal distribution, where we will apply the Central Limit Theorem (CLT): Suppose $\mathbf{X} = \{X_1, \dots, X_M\}$ is a sequence of i.i.d. random variables with $\langle X_i \rangle = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$. Then as $M \rightarrow \infty$, the random variable $\sqrt{M}(\bar{\mathbf{X}} - \mu)$ converges in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

We start with the Mean Value Theorem for the function L'_M , the derivative of L_M (continuous by assumption), on the interval $[\Theta_{\text{MLE}}, \theta_0]$:

$$\begin{aligned} 0 &= L'_M(\Theta_{\text{MLE}}) = L'_M(\theta_0) + L''_M(\theta_1)(\theta_0 - \Theta_{\text{MLE}}) \\ \implies \theta_0 - \Theta_{\text{MLE}} &= -\frac{L'_M(\theta_0)}{L''_M(\theta_1)} \\ \implies \sqrt{M}(\theta_0 - \Theta_{\text{MLE}}) &= -\sqrt{M} \frac{L'_M(\theta_0)}{L''_M(\theta_1)} \end{aligned} \quad (48)$$

for some $\theta_1 \in [\Theta_{\text{MLE}}, \theta_0]$. We analyze the numerator and denominator respectively. The numerator

$$\begin{aligned} L'_M(\theta_0) &= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' \\ &= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' - L'(\theta_0) \\ &= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' - \langle (\log P(\mathbf{X}|\theta_0))' \rangle_{\theta_0} \\ &= \frac{1}{M} \left(\sum_{i=1}^M \log P(X_i|\theta_0) \right)' - \langle (\log P(X_i|\theta_0))' \rangle_{\theta_0} \end{aligned} \quad (49)$$

where the last equality is obtained from the linearity of expected value and derivative. By the CLT, the distribution of $\sqrt{M}L'_M(\theta_0)$ converges to

$$\mathcal{N}\left(0, \text{Var}_{\theta_0}(\log P(X_i|\theta_0))'\right) \quad (50)$$

where the variance

$$\begin{aligned} \text{Var}_{\theta_0}(\log P(X_i|\theta_0))' &= \langle [(\log P(X_i|\theta_0))']^2 \rangle_{\theta_0} - \langle (\log P(X_i|\theta_0))' \rangle_{\theta_0}^2 \\ &= \sum_{x \in X_1} P(x|\theta_0) \left(\frac{P'(x|\theta_0)}{P(x|\theta_0)} \right)^2 - (L'(\theta_0))^2 \\ &= \text{CFI}_{\theta_0} \end{aligned} \quad (51)$$

by the definition of CFI and that θ_0 maximizes $L(\theta)$. By the consistency property, Θ_{MLE} converges to θ_0 , and thus θ_1 converges to θ_0 . The denominator

$$L''_M(\theta_1) \rightarrow L''_M(\theta_0) = \frac{1}{M} \sum_{i=1}^M [\log P(X_i|\theta_0)]'' \rightarrow \langle [\log P(X_i|\theta_0)]'' \rangle_{\theta_0} \quad (52)$$

by the WLLN. We further show that Eq. (52) is in fact the additive inverse of CFI:

$$\begin{aligned}
\langle [\log P(X_1|\theta_0)]'' \rangle_{\theta_0} &= \left\langle \frac{\partial^2}{\partial \theta^2} \log P(X_1|\theta_0) \right\rangle_{\theta_0} \\
&= \sum_{x \in X_1} [\log P(x|\theta_0)]'' P(x|\theta_0) \\
&= \sum_{x \in X_1} \left(\frac{P''(x|\theta_0)}{P(x|\theta_0)} - \left(\frac{P'(x|\theta_0)}{P(x|\theta_0)} \right)^2 \right) P(x|\theta_0) \\
&= \sum_{x \in X_1} P''(x|\theta_0) - \sum_{x \in X_1} \frac{(P'(x|\theta_0))^2}{P(x|\theta_0)} \\
&= 0 - \text{CFI}_{\theta_0} = -\text{CFI}_{\theta_0}.
\end{aligned} \tag{53}$$

Finally, Eq. (48) becomes

$$\begin{aligned}
\sqrt{M}(\theta_0 - \Theta_{\text{MLE}}) &\xrightarrow{p} \mathcal{N}\left(0, \frac{\text{CFI}_{\theta_0}}{\text{CFI}_{\theta_0}^2}\right) = \mathcal{N}\left(0, \frac{1}{\text{CFI}_{\theta_0}}\right) \\
\implies \Theta_{\text{MLE}} &\xrightarrow{p} \mathcal{N}\left(\theta_0, \frac{1}{M \cdot \text{CFI}_{\theta_0}}\right)
\end{aligned} \tag{54}$$

Thus, the MLE is unbiased and asymptotically saturates the Cramér-Rao bound.

Supplementary Note 5: Master equation for a non-Markovian environment

To simulate the performance of our optimized states during the Ramsey measurement with non-Markovian noise, we use a time-local master equation given by [24]. A brief summary of the derivation is given below.

1. The time evolution of a state ρ is described by

$$\rho(t) = \Lambda(t)[\rho(0)], \tag{55}$$

where $\Lambda(t)$ includes both the dissipative part and the unitary encoding part. Since $\Lambda(t)$ is a physical operator that maps a density matrix to a density matrix, it must preserve Hermiticity, unit trace, and positivity. Moreover, any quantum state ρ can be extended by an ancilla to a state $\rho \otimes \sigma$ of a larger system. The corresponding map $\Lambda \otimes 1$ must also be positive. This property is known as complete positivity [25]. Thus, $\Lambda(t)$ is a completely positive and trace preserving (CPTP) linear map.

2. Define the action of Λ . Let $\mathcal{L}(\mathbb{C}^d)$ be the Hilbert space of linear operators acting on \mathbb{C}^d , where the inner product is defined as $\langle \sigma, \tau \rangle = \text{Tr}(\sigma^\dagger \tau)$ (the Hilbert-Schmidt inner product). Let $\mathcal{LL}(\mathbb{C}^d)$ be the Hilbert space of linear operators acting on $\mathcal{L}(\mathbb{C}^d)$ which has dimension $d^2 \times d^2$. Let $\{l_i\}_{i=1, \dots, d^2}$ be an orthonormal basis of $\mathcal{L}(\mathbb{C}^d)$. Then the action of $\Lambda \in \mathcal{LL}(\mathbb{C}^d)$ on $\tau \in \mathcal{L}(\mathbb{C}^d)$ can be expressed as

$$\begin{aligned}
\Lambda[\tau] &= \sum_{i,j=1}^{d^2} \langle l_i, \Lambda[l_j] \rangle \langle l_j, \tau \rangle l_i \\
&\equiv (\Lambda \mathbf{b})^T l
\end{aligned} \tag{56}$$

where the matrix Λ has elements

$$\Lambda_{ij} \equiv \langle l_i, \Lambda[l_j] \rangle \tag{57}$$

and \mathbf{b} is a vector with entries $\mathbf{b}_j = \langle l_j, \tau \rangle$. Thus, Λ has a unique correspondence with the matrix Λ .

3. $\Lambda \in \mathcal{LL}(\mathbb{C}^d)$ is trace- and hermiticity-preserving if and only if its matrix representation Λ can be written as

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{m} & \mathbf{M} \end{pmatrix}, \tag{58}$$

where $\mathbf{0}$ is the zero row vector of length $d^2 - 1$, \mathbf{m} is a real column vector of length $d^2 - 1$, and \mathbf{M} is a $(d^2 - 1)(d^2 - 1)$ real matrix.

4. For a single qubit, any operator ρ on \mathbb{C}^2 can be written as

$$\rho = \frac{1}{2}(\mathbf{I} + \mathbf{v} \cdot \boldsymbol{\sigma}) \quad (59)$$

where \mathbf{v} is a three-dimensional real vector and $\boldsymbol{\sigma}$ is the vector of Pauli matrices. Then a map Λ whose matrix representation is given by Eq. (58) acting on ρ gives

$$\Lambda[\rho] = \frac{1}{2}(\mathbf{I} + (\mathbf{m} + \mathbf{M}\mathbf{v}) \cdot \boldsymbol{\sigma}). \quad (60)$$

5. The Nakajima—Zwanzig projection operator technique can be used to describe non-Markovian behavior [26]. The basic idea of the technique is that the operation of tracing over the environment is regarded as a projection $\rho_C \rightarrow P\rho_C$ where ρ is the composite state and P is an projection operator. The technique makes use of some fairly general assumptions and gives

$$\frac{d}{dt}\rho(t) = \int_0^t K(t, s)\rho(s) ds \quad (61)$$

where $K(t, s)$ is a memory kernel.

6. Applying Eq. (55) to Eq. (61) gives [27]

$$\begin{aligned} \frac{d}{dt}\rho(t) &= \int_0^t (K(t, s) \circ \Lambda(s))\rho(0) ds \\ &= \int_0^t (K(t, s) \circ \Lambda(s) \circ \Lambda(t)^{-1})\Lambda(t)[\rho(0)] ds \\ &\equiv \Xi(t)\rho(t), \end{aligned} \quad (62)$$

which is of time-local form.

7. Both $\Lambda(\rho)$ and $\Xi(\rho)$ can be expanded in the form of Eq. (56):

$$\rho(t) = \Lambda(t)[\rho(0)] = [\Lambda(\mathbf{t})\mathbf{b}(0)]^T l, \quad (63)$$

$$\frac{d}{dt}\rho(t) = \Xi(t)[\rho(t)] = [\Xi(\mathbf{t})\mathbf{b}(\mathbf{t})]^T l. \quad (64)$$

Then

$$\frac{d}{dt}\rho(t) = \left[\frac{d}{dt}\Lambda(\mathbf{t})\mathbf{b}(0)\right]^T l = [\Xi(\mathbf{t})\Lambda(\mathbf{t})\mathbf{b}(0)]^T l. \quad (65)$$

The corresponding matrix relation is

$$\frac{d}{dt}\Lambda(\mathbf{t}) = \Xi(\mathbf{t})\Lambda(\mathbf{t}). \quad (66)$$

Equivalently,

$$\Xi(t) = \frac{d\Lambda(t)}{dt}\Lambda(t)^{-1}. \quad (67)$$

8. Assume the noise to be independent, identical, and phase covariant, so that each qubit is described by a map $\Lambda(t) = \mathcal{U}(t) \circ \Gamma(t)$ and that the uncorrelated noise commutes with signal accumulation [24,28]. Consider the evolution of one qubit described by $\Lambda(t) = \mathcal{U}(t) \circ \Gamma(t)$. $\mathcal{U}(t)$ is defined as

$$\mathcal{U}(t)[\rho(0)] \equiv U(t)\rho(0)U^\dagger(t) \quad (68)$$

where $U(t) = e^{-i\frac{\omega t}{2}\sigma_z}$ represents the signal accumulation. By Eq. (56) and Eq. (68), the matrix representation of $\mathcal{U}(t)$ is

$$\mathcal{U}(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \omega t & -\sin \omega t & 0 \\ 0 & \sin \omega t & \cos \omega t & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (69)$$

$\Gamma(t)$ represents the noise which is trace- and hermicity- preserving, i.e., has the form in Eq. (58).

9. Solving the commutation relation that gives phase covariant qubit map [24,28]

$$[\mathcal{U}(t), \Gamma(t)] = 0 \iff [\mathbf{U}(t), \Gamma(t)] = 0, \quad (70)$$

we obtain the matrix representation of $\Gamma(t)$:

$$\Gamma(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \eta_{\perp}(t) \cos \theta & -\eta_{\perp}(t) \sin \theta & 0 \\ 0 & \eta_{\perp}(t) \sin \theta & \eta_{\perp}(t) \cos \theta & 0 \\ \kappa(t) & 0 & 0 & \eta_{\parallel}(t) \end{pmatrix}, \quad (71)$$

where $\mathbf{m} = (0, 0, \kappa(t))^T$ describes a translation along the z -axis, and $\mathbf{M} = \begin{pmatrix} \eta_{\perp}(t) \cos \theta & -\eta_{\perp}(t) \sin \theta & 0 \\ \eta_{\perp}(t) \sin \theta & \eta_{\perp}(t) \cos \theta & 0 \\ 0 & 0 & \eta_{\parallel}(t) \end{pmatrix}$ describes a rotation along the z -axis and a contraction characterized by η_{\perp} and η_{\parallel} . The product of $\mathbf{U}(t)$ and $\Gamma(t)$ then gives $\Lambda(t)$:

$$\Lambda(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \eta_{\perp}(t) \cos \phi & -\eta_{\perp}(t) \sin \phi & 0 \\ 0 & \eta_{\perp}(t) \sin \phi & \eta_{\perp}(t) \cos \phi & 0 \\ \kappa(t) & 0 & 0 & \eta_{\parallel}(t) \end{pmatrix} \quad (72)$$

where $\phi = \omega t + \theta$.

10. By Eq. (67), we obtain the time-local master equation for a single qubit:

$$\begin{aligned} \Xi(t)[\rho(t)] = & -\frac{i}{2}\omega[\sigma_z, \rho(t)] \\ & + \gamma_+(t)(\sigma_+\rho(t)\sigma_- - \frac{1}{2}\{\sigma_-\sigma_+, \rho(t)\}) \\ & + \gamma_-(t)(\sigma_-\rho(t)\sigma_+ - \frac{1}{2}\{\sigma_+\sigma_-, \rho(t)\}) \\ & + \gamma_z(t)(\sigma_z\rho(t)\sigma_z - \rho(t)), \end{aligned} \quad (73)$$

where

$$\gamma_+(t) = \frac{1}{2} \left(\kappa'(t) - \frac{\eta'_{\parallel}(t)}{\eta_{\parallel}(t)}(\kappa(t) + 1) \right), \quad (74)$$

$$\gamma_-(t) = -\frac{1}{2} \left(\kappa'(t) + \frac{\eta'_{\parallel}(t)}{\eta_{\parallel}(t)}(1 - \kappa(t)) \right), \quad (75)$$

$$\gamma_z(t) = \frac{1}{4} \left(\frac{\eta'_{\parallel}(t)}{\eta_{\parallel}(t)} - 2\frac{\eta'_{\perp}(t)}{\eta_{\perp}(t)} \right). \quad (76)$$

Considering only T_2 noise, $\gamma_-(t) = \gamma_+(t) = 0$, η_{\parallel} is constant, and

$$\eta_{\perp}(t) = e^{-(\frac{t}{T_2})^{\nu}}, \quad (77)$$

where ν is the stretch character which equals 1 for Markovian noise. Then

$$\gamma_z(t) = \frac{\nu}{2} \frac{t^{\nu-1}}{T_2^{\nu}}. \quad (78)$$

We further need to express $\Xi(t)$ as a superoperator acting on the vectorization of $\rho(t)$. Defining the vectorization of a matrix as the map

$$\rho = \sum_{i,j} \rho_{ij} |i\rangle \langle j| \mapsto |\rho\rangle = \sum_{i,j} \rho_{ij} |j\rangle \otimes |i\rangle. \quad (79)$$

Define the left and right multiplication superoperators by $\mathcal{L}(A)[\rho] = A\rho$ and $\mathcal{R}(A)[\rho] = \rho A$ so that $[A, \rho] = \mathcal{L}(A)[\rho] - \mathcal{R}(A)[\rho]$. By this definition, we can calculate the matrix representation $\mathcal{L}(A) = I \otimes A$ and $\mathcal{R}(A) = A^T \otimes I$. Using the superoperator notation, we can express $\Xi(t)$ as

$$\Xi(t) = -\frac{i}{2}\omega(I \otimes \sigma_z - \sigma_z \otimes I) + \gamma_z(t)(\sigma_z \otimes \sigma_z - I \otimes I). \quad (80)$$

With this expression, we numerically simulate the evolution of our entangled states under non-Markovian noise by using the Time-Dependent Master Equation Solver in QuTip [29].

Supplementary Note 6: Performance of metrological states in a non-Markovian environment

To calculate the derivative of probability with respect to ω in the calculation for CFI_ω , we use a method similar to parameter shift that utilizes the property that the signal accumulation operator ($\mathcal{U}(\omega) = e^{-i\omega t J_y}$) and the noisy operator commutes. In the following notation,

1. z represents a multi-qubit state in the z basis;
2. $\mathcal{U}(\omega)$ is the effective signal accumulation operator: $\mathcal{U}(\omega) = e^{-i\omega t J_y}$;
3. ρ is the state density matrix of our optimized state after the noisy evolution without signal for some Ramsey time and a $\frac{\pi}{2}$ pulse along the x direction (here we switch the order of the signal accumulation and the second pulse of the Ramsey protocol [18]);
4. $P(z|\omega)$ is the probability of measuring the state z with our rotated optimized state after the noisy evolution and signal accumulation.

Then

$$\begin{aligned} \frac{\partial}{\partial \omega} P(z|\omega) \Big|_{\omega \rightarrow 0} &= \frac{\partial}{\partial \omega} \text{Tr}[|z\rangle \langle z| \mathcal{U}(\omega) \rho \mathcal{U}^\dagger(\omega)] \Big|_{\omega \rightarrow 0} \\ &= \frac{\partial}{\partial \omega} \text{Tr}[\mathcal{U}^\dagger(\omega) |z\rangle \langle z| \mathcal{U}(\omega) \rho] \Big|_{\omega \rightarrow 0} \\ &= \text{Tr}[\frac{\partial}{\partial \omega} \mathcal{U}^\dagger(\omega) |z\rangle \langle z| \mathcal{U}(\omega) \rho] \Big|_{\omega \rightarrow 0} + \text{Tr}[\mathcal{U}^\dagger(\omega) |z\rangle \langle z| \frac{\partial}{\partial \omega} \mathcal{U}(\omega) \rho] \Big|_{\omega \rightarrow 0} \\ &= it \text{Tr}[(J_y |z\rangle \langle z| - |z\rangle \langle z| J_y) \rho]. \end{aligned} \quad (81)$$

From Eq. (38), since T_{tot} and $\delta\omega^2$ are constants, SNR is proportional to $\sqrt{\frac{\text{CFI}_\omega}{t_R}}$. Thus we choose $\frac{\text{CFI}_\omega}{t_R}$ as the result we show in Fig.4(d) in the main text.

Supplementary Note 7: Time Overhead

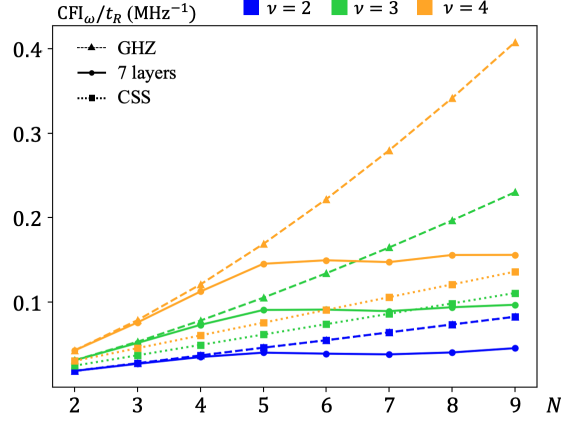
In experiments, the time overhead, including the state preparation and readout time, reduces the repetition number of the sensing sequence and thus decreases the sensitivity. For all the proposed experimental platforms the readout time exceeds the typical T_2 and state preparation time, which means that even for protocols with CSS the time overhead is significant [30].

If we consider a nonzero time overhead, i.e., $M = \frac{T_{\text{tot}}}{t_R + t_{\text{oh}}}$, the expression for SNR^2 for an uncorrelated spin state becomes

$$\text{SNR}^2 = \frac{T_{\text{tot}} t_R^2 \cos^2(\omega t_R) \delta\omega^2}{(t_R + t_{\text{oh}}) \left(e^{2\left(\frac{t_R}{T_2}\right)^\nu} - \sin^2(\omega t_R) \right)}. \quad (82)$$

If $t_{\text{oh}} \gg t_R$, we ignore the term t_R in the denominator and

$$\text{SNR}^2 \propto \frac{t_R^2}{e^{2\left(\frac{t_R}{T_2}\right)^\nu}}. \quad (83)$$



Supplementary Figure 10. Sensing performance under large time overhead: 50 cases average sensing performance of the optimized states when using 7-layer circuit on 3D random spin configuration.

Taking the derivative of Eq. (83) with respect to t_R gives us the best t_R if the time overhead is significantly larger:

$$t_R = \frac{T_2}{\nu^{\frac{1}{\nu}}}. \quad (84)$$

Similarly, the same calculations for a GHZ state where the decay term in Eq. (82) becomes $e^{2n\left(\frac{t_R}{T_2}\right)^\nu}$ show that the best Ramsey sensing time is

$$t_R = \frac{T_2}{(n\nu)^{\frac{2}{\nu}}}. \quad (85)$$

Plugging Eq. (84) and Eq. (85) into Eq. (83), we find that the ratio of the SNR^2 of a GHZ state to that of an uncorrelated spin state is $n^{1-\frac{2}{\nu}}$. Thus, only when

$$\nu > 2 \quad (86)$$

do GHZ states provide an advantage in SNR over uncorrelated spin states when $t_{\text{oh}} \gg t_R$. We compare the SNR of the states generated by the optimizer with that of the CSS and GHZ states when $\nu = 2, 3, 4$. Fig. 10 shows that when we assume a long time overhead, the generated entangled states are less sensitive than CSS when $\nu = 2$ and $\nu = 3$ for large spin numbers.

Supplementary Note 8: State preparation time comparing to adiabatic method

State preparation time is one of the major components of the time overhead in the generalized Ramsey sensing sequence which influences the sensitivity. The state preparation time of the variational method depends on the circuit layer number m , system size N and is proportional to the inverse of average interaction strength $1/\bar{f}_{\text{dd}}$. The adiabatic method [31] is an alternative approach to generate entangled states for quantum metrology in dipolar-interacting spin systems by only using single-qubit rotations (global pulses).

To compare the performance of our variational method with the adiabatic method, we derive the relation between the squeezing parameter (Wineland parameter [32]) and CFL. Without loss of generality, we consider a SSS with collective spin direction $+x$ and is squeezed along the y -axis (such as the 3rd Wigner distribution shown in Fig. 9(b)). In this case, the squeezing parameter is

$$\xi^2 = N \frac{(\Delta J_y)^2}{|\langle J_x \rangle|^2}, \quad (87)$$

where $(\Delta O)^2 \equiv \langle O^2 \rangle - \langle O \rangle^2$ and N is the number of spins. According to the uncertainty principle,

$$(\Delta J_y)^2(\Delta J_z)^2 \geq \frac{1}{4}|\langle J_x \rangle|^2. \quad (88)$$

The relation between the squeezing parameter and total spin angular momentum uncertainty projection in z -direction is

$$4(\Delta J_z)^2 \geq \frac{|\langle J_x \rangle|^2}{(\Delta J_y)^2} = N/\xi^2. \quad (89)$$

It's been proven that for a pure Gaussian state, the quantum Fisher information (QFI) is directly related to the variance of the projected spin angular momentum [17,33,34]:

$$\text{QFI} = 4(\Delta J_z)^2. \quad (90)$$

Combining Eq. (89) and Eq. (90), we obtain the relation between CFI and squeezing parameter of a SSS:

$$\text{CFI} \leq \text{QFI} \geq N/\xi^2. \quad (91)$$

The first inequality in Eq. (91) is saturated by measuring the SSS along the direction where it is squeezed (y -axis, or equivalently measuring it in z -basis after applying a $R_x(\frac{\pi}{2})$ pulse [35]). The second inequality originates from the uncertainty principle (Eq. (88)). Since the optimal SSS saturate the Heisenberg uncertainty relation [36] and the SSS generated by the adiabatic method [31] belongs to these states, we obtain the relation between the squeezing parameter and CFI

$$\text{CFI} = N/\xi^2. \quad (92)$$

Based on the data shown in Fig.3 from ref. (31), it takes about $200\mu\text{s}$ for the adiabatic method to prepare an 8-spin SSS with $\xi^2 = 0.4$ which corresponds to $\text{CFI} = 20$. The 2D spin density $8/(30\text{nm} \times 30\text{nm})$ corresponds to $\bar{f}_{\text{dd}} = 43.5\text{kHz}$. According to Fig.3(d) in the main text, the variational method is able to prepare an 8-spin entangled state with $\text{CFI} \approx 20$ by a 4-layer circuit with $\bar{f}_{\text{dd}}T = 0.8$. Plugging in the same average nearest neighbor dipolar interaction strength \bar{f}_{dd} , we finally calculated the state preparation time of the variational method is $T = 18.4\mu\text{s}$, which is about 11 times faster than the adiabatic method under the same condition.

Supplementary Note 9: Numerically solving Schrödinger and Lindblad equations

In this work, both the Schrodinger equations for unitary evolution and Lindblad master equations for noisy evolution are not solved via diagonalization, since the matrix dimension increases exponentially with the spin ensemble size and makes the computation ineffective and memory-consuming. In particular, for noisy cases, using the superoperator formalism would lead to a dimension growth of 4^N . For 10 spins, the dimension would exceed one million. Instead, the equations are solved with linear multistep methods, which are well-suited numerical integrators to solve ordinary differential equations in the form of $\frac{dy}{dt} = f(y, t)$, $y(t_0) = y_0$. Depending on the stiffness of the equation, implicit Adams method (for nonstiff) and backward differentiation formulas (for stiff) are applied. In simple words, the approximated values for y at previous time steps initialize slopes f , which are then interpolated by a polynomial. Replacing the original integrand by the interpolation polynomial gives a new approximated y , solved via an implicit equation. These methods are provided by `scipy.integrate.ode` of the `scipy` package of Python, setting the integrator to be “`zvode`”, meaning complex-valued variable-coefficient ODE solver. Details of the algorithm are described in [37].

SUPPLEMENTARY DISCUSSION: CONTROLLABILITY

Since all the black-box optimization algorithms cannot ensure that the optimized result is the global maximum/minimum point of in the parameter space, it is still an open question that if the variational method is able to find the 'best' metrological state for a given spin configuration or not. In this section, we're interested in the theoretically achievable controllability of dipolar interacting spin systems. The question is, given any (possibly infinite) arbitrary sequence of evolution under each Hamiltonian governing the dynamics of our system, can we drive any arbitrary unitary operator? Quantum control systems of the general form

$$H(t) = H_0 + \sum_{k=1}^K u_k(t)H_k, \quad (93)$$

governed by the Schrödinger equation, $i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle$, have been studied extensively [38,39,40]. H_0 is the unperturbed or free evolution Hamiltonian, H_k are the control interactions, and $u_k(t)$ are the piecewise continuous control fields. There are several distinct but related notions of controllability that have different conditions for ‘full’ controllability. The notion of ‘operator’ or ‘complete’ controllability is the strictest condition and is defined as above. For generic interacting spin systems, all of these notions are equivalent. Complete controllability is equivalent to universal quantum computation (UQC) in quantum information processing (QIP) [41,42].

Controllability Test

The way we investigate the controllability of a generic system (Eq. 93) is by examining the so-called ‘dynamical Lie algebra’ $\mathcal{L}_0 \subseteq u(\mathcal{N})$ or $su(\mathcal{N})$ generated by the operators $\{-iH_0, -iH_1, \dots, -iH_K\}$, which are represented by $\mathcal{N} \times \mathcal{N}$ matrices in a basis we choose [38,39].

A quantum system of the form (Eq. 93) is completely controllable if either $\mathcal{L}_0 \cong u(\mathcal{N})$ or $\mathcal{L}_0 \cong su(\mathcal{N})$ [38], where $u(\mathcal{N})$ is the unitary Lie algebra represented by the set of skew-Hermitian $\mathcal{N} \times \mathcal{N}$ matrices and $su(\mathcal{N})$ is the special unitary Lie algebra represented by the same set of matrices with the extra condition that they are traceless. Note that $\dim u(\mathcal{N}) = \mathcal{N}^2$ and $\dim su(\mathcal{N}) = \mathcal{N}^2 - 1$, and the difference of 1 comes from counting identity operation (I) as a dimension or not. We must find a basis for \mathcal{L}_0 by iteratively taking the Lie bracket $[\cdot, \cdot]$ of H_0, H_1, \dots, H_K until we have a set of $\dim \mathcal{L}_0$ linearly independent matrices, where the Lie bracket is the commutator $[A, B] = AB - BA$ for matrices A and B . Ref.[38] and ref.[39] present an algorithm for generating this basis. Thus, if $\dim \mathcal{L}_0 = \mathcal{N}^2$ or $\mathcal{N}^2 - 1$ we can say that the system is completely controllable. Note that for generic spin systems $\mathcal{N} = 2^N$ for N spins.

Controllability of Dipolar Interacting Spin Systems

We write our system in the form (Eq. 93) by defining the free evolution Hamiltonian to be the dipolar interaction H_{dd} and two control interactions J_x and J_y , as these operators are generators of rotation, with respective independent control fields $\theta_x(t)$ and $\theta_y(t)$:

$$H(t) = H_{\text{dd}} + \theta_x(t)J_x + \theta_y(t)J_y. \quad (94)$$

Ref.[39,43] demonstrate that we cannot achieve complete controllability with global controls due to inherent symmetries, so we know that $\dim \mathcal{L}_0 < 4^N - 1$.

However, complete controllability is a rather strict condition. Not being able to drive any arbitrary unitary does not mean we cannot drive unitaries that produce metrological states.

In fact, ref.[44] demonstrate for a long-range Ising spin model (all-to-all interactions) with global controls that metrological states, such as the GHZ and W states are reachable. Ref.[45] extend their result for symmetric Ising spin networks with global controls and demonstrate that one can reach any state that preserves spin permutation invariance. This is known as subspace controllability. The dimension of their dynamical Lie algebra, $\mathcal{L}^{\text{Ising}} \equiv \mathcal{L}^{\text{PI}} \cap su(2^N)$, is

Algorithm. Generating \mathcal{L}_0 and finding $\dim \mathcal{L}_0$.

Input: Hamiltonians $I \equiv \{H_0, H_1, \dots, H_K\}$

1. $B \equiv$ maximal linearly independent subset of I
2. $r \equiv |B|$
3. If $r = \mathcal{N}^2$ then $O \equiv B$ else $O \equiv \{\}$
4. If $r = \mathcal{N}^2$ or $|B| = 0$ then terminate
5. $C \equiv [O, B] \cup [B, B]$, where
 $[S_1, S_2] \equiv \{[s_1, s_2] \mid s_1 \in S_1, s_2 \in S_2\}$
6. $O = O \cup B$
7. $B =$ maximal linearly independent extension of O with elements from C
8. $r = r + |B|$; Go to 4

Output: basis O of \mathcal{L}_0 and $\dim \mathcal{L}_0 = r$

Supplementary Table 2. Implementation of [39]’s algorithm with a few physically motivated modifications. Note $|S|$ indicates the cardinality of set S .

Lie algebra dimension	$N = 2$	$N = 3$	$N = 4$	$N = 5$
Completely controllable: 4^N (or $4^N - 1$)	16	64	256	1024
H_{dd}	9	39	225	
Symmetric Ising: $\binom{N+3}{N} - 1$	9	19	34	55

Supplementary Table 3. Lie algebra dimensions for the complete controllable system, dipolar interacting system and symmetric Ising system (lower bound for subspace controllability). Dipolar interacting spin systems' $\dim \mathcal{L}_0$ is calculated using an implementation of [39]'s algorithm, and is necessarily bounded by the complete and subspace controllability dimensions. Lie algebra dimensions for dipolar interacting systems are only calculated up to $N = 4$ due to stability issues stemming from numerical errors in how matrix rank is calculated.

shown to be $\binom{N+3}{N} - 1$. This is relevant to our system because [46] show that if we replace the Ising interaction with a more general two body interaction—which includes H_{dd} —the dimension of the dynamical Lie algebra is necessarily greater than or equal to that of the symmetric Ising case, and it is therefore subspace controllable. This means that we can write $\binom{N+3}{N} - 1 \leq \dim \mathcal{L}_0 < 4^N - 1$ and say that \mathcal{L}_0 is subspace controllable but not completely controllable. Therefore, we can achieve arbitrary permutation invariant states, including metrological states such as a GHZ state.

Finding Reachable States

\mathcal{L}_0 is associated with a Lie group $e^{\mathcal{L}_0}$ by the Lie group–Lie algebra correspondence [38]. The Lie algebra $u(\mathcal{N})$ corresponds to the Lie group $U(\mathcal{N})$, and $su(\mathcal{N})$ corresponds to $SU(\mathcal{N})$. We can define $\mathcal{R} \equiv e^{\mathcal{L}_0}$ as the reachable set of unitaries we can drive under $\{H_k\}_{k=0,\dots,K}$, and so starting from an initial state $|\psi_0\rangle$, $\mathcal{R}_{|\psi_0\rangle}$ is the set of states we can reach.

As demonstrated in the previous section, our dynamical Lie algebra is a superset of $\mathcal{L}^{\text{Ising}}$ and a strict subset of $su(2^N)$, so we can write $e^{\mathcal{L}^{\text{Ising}}} \subseteq e^{\mathcal{L}_0} \subset SU(2^N)$. Because $|\text{GHZ}\rangle \in \mathcal{R}_{|0\rangle^{\otimes N}}^{\text{Ising}}$ we can write $|\text{GHZ}\rangle \in \mathcal{R}_{|0\rangle^{\otimes N}}^{\text{dipolar}}$. In fact, this is true for any permutation invariant state, which includes all metrological states we're interested in.

While we know that metrological states are in the reachable set, determining the parameters that drive the unitaries to produce those states is a highly convex optimization problem equivalent to our variational circuit, using state fidelity between the ideal state and the current state instead of CFI as the cost function. That is we optimize the output unitary of the variational circuit,

$$\mathcal{S}(\boldsymbol{\theta}) = e^{-i\frac{\pi}{2}J_y} \prod_{i=1}^m e^{-i\tau_i H_{\text{dd}}} e^{-i\vartheta_i J_x} e^{i\frac{\pi}{2}J_y} e^{-i\tau'_i H_{\text{dd}}} e^{-i\frac{\pi}{2}J_y}, \quad (95)$$

where m is the (possibly infinite) number of layers, for state fidelity,

$$\mathcal{F}(|\text{GHZ}\rangle, \mathcal{S}(\boldsymbol{\theta})|0\rangle^{\otimes N}) = \left| \langle \text{GHZ} | \mathcal{S}(\boldsymbol{\theta}) | 0 \rangle^{\otimes N} \right|^2, \quad (96)$$

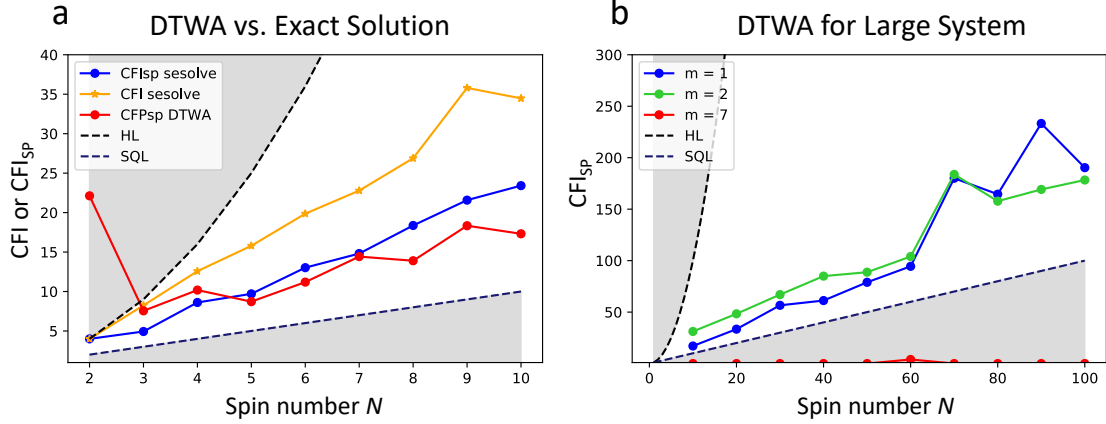
for pure states. If there exists some $\boldsymbol{\theta}$ such that $\mathcal{F}(|\text{GHZ}\rangle, \mathcal{S}(\boldsymbol{\theta})|0\rangle^{\otimes N}) = 1$, then we can say that $|\text{GHZ}\rangle \in \mathcal{R}_{|0\rangle^{\otimes N}}^{\text{dipolar}}$. From the previous section, we know such a $\boldsymbol{\theta}$ must exist, but it may be the case that $m \rightarrow \infty$, in which case it is not possible to find this exactly. This is the method employed in ref.[44,47] to demonstrate the reachability of GHZ and W states for Ising spin models. Our variational circuit method represents an improvement in the efficiency of searching for such metrological states.

SUPPLEMENTARY METHODS

Large spin number ($N \rightarrow 100$) optimization: discrete truncated Wigner approximation (DTWA)

The optimization results for small spin number ($N \leq 10$) are of particular interest to nanoscale quantum sensing, where the sensor size limits N . On the other hand, the advantage of Heisenberg limit quantum sensing ($\frac{1}{N}$) becomes more dominant compared to standard quantum limit sensing ($\frac{1}{\sqrt{N}}$) when the spin number N is large. Thus, it is worth to discuss the performance of the variational method in large N case. However, numerically simulate large N system is challenging because the Hilbert space grows exponentially.

We employ discrete truncated Wigner approximation (DTWA) [48,49], Monte Carlo algorithm, to semi-classically simulate the time evolution of large interacting spin system. Instead of using a 2^N dimensional density matrix (or state vector) to describe the spin system, DTWA represent each spin by a 3-dimensional vector (s_i^x, s_i^y, s_i^z) , where $s_i^{x,y,z}$ are real numbers that stand for the spin components for number i 'th spin. The initial values of $s_i^{x,y,z}$ are randomly sampled following the probability distribution of the initial quantum state and the time evolution of $s_i^{x,y,z}$ is described by a classical Hamiltonian equation [48]. The final time evolution results are obtained by averaging the results from different initial values of $s_i^{x,y,z}$.



Supplementary Figure 11. DTWA optimization results: In 2D square lattice spin configuration. (a) Sanity check for $N = 2 - 10$, circuit layer $m = 1$. Results from the optimizer using the exact Schrödinger equation solver (sesolve) and DTWA are shown. (b) Optimization up to $N = 100$ by using DTWA, circuit layer $m = 1, 2, 7$.

Due to the statistical nature of the Monte Carlo simulation, DTWA is not compatible with parameter shift, the way we previously computed the cost function [19,20,21]. Thus, we use the CFI_{sp} of a Gaussian states which related to the squeezing parameter as the cost function for DTWA optimization (derived in Eq. 92). Figure 11(a) shows the optimized data by using DTWA and exact numerical simulation (sesolve function in QuTip) for the 2D square lattice spin configuration when $N \leq 10$. We observed that the optimized results from DTWA are lower than those from the exact simulation when $N > 5$. This can be attributed to the statistical fluctuation induced by the DTWA algorithm and its influence to the landscape of the parameter space which might be harder to optimize. Note, the unphysical beyond-HL sensitivity for $N = 2$ is an example where DTWA fails in describing a highly entangled state. Based on the $\text{CFI} = \text{CFI}_{\text{sp}} = 4$ result by using the exact numerical simulation, we know that a GHZ state (Bell pair) is generated and DTWA is not able to correctly capture the properties of such highly entangled state.

In Fig. 11(b), the optimized CFI_{sp} of 1, 2 and 7 layer circuit for spin number up to 100 are shown. Beyond-SQL sensitivity is still reached in 1 and 2 layer circuits' results. Note, the circuits for $m=1$ (and $m=2$) are a sub-set of $m=7$, yet the shallower circuits outperform $m=7$ in DTWA. We attribute this to CMA-ES failing to optimize the cost function in a large dimensional space, which is particularly difficult given the statistical nature of DTWA.

The DTWA optimization results indicate that our approach is promising to realize beyond-SQL quantum sensing for 2D spin systems, even when the spin number is large and the circuit shallow. However, further investigations regarding scalability of our variational method on 2D random/3D random spin configurations is needed. More robust optimization algorithms that are compatible with large N and m would likely further the improvements. On the other hand, running the optimization directly on the experimental platform will potentially get around the difficulty of simulating large spin ensembles precisely and optimize for suitable quantum states under all kinds of experimental imperfections.

SUPPLEMENTARY REFERENCES

- [1] B. Koczor, R. Zeier, and S. J. Glaser, Fast computation of spherical phase-space functions of quantum many-body states, *Phys. Rev. A* **102**, 062421 (2020).
- [2] G. Kucsko *et al.*, Critical thermalization of a disordered dipolar spin system in diamond, *Phys. Rev. Lett.s* **121**, 023601 (2018).
- [3] C. Slichter, Principles of magnet resonance, Spring Series in Solid (1980).
- [4] S. Choi *et al.*, Observation of discrete time-crystalline order in a disordered dipolar many-body system, *Nature* **543**, 221–225 (2017).
- [5] A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 4 (University of California Press, 1961) p. 547–562.
- [6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [7] R. o. Islam, Measuring entanglement entropy in a quantum many-body system, *Nature* **528**, 77–83 (2015).
- [8] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Area laws in quantum systems: mutual information and correlations, *Phys. Rev. Lett.s* **100**, 070502 (2008).
- [9] D. Bluvstein *et al.*, A quantum processor based on coherent transport of entangled atom arrays, *Nature* **604**, 451–456 (2022).
- [10] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* (MIT press, 2018).
- [11] D. Silver *et al.*, Mastering the game of Go without human knowledge, *Nature* **550**, 354–359 (2017).
- [12] P. Peng, X. Huang, C. Yin, L. Joseph, C. Ramanathan, and P. Cappellaro, Deep reinforcement learning for quantum Hamiltonian engineering, Preprint at <https://arxiv.org/abs/2102.13161> (2021).
- [13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
- [14] R. Kaubruegger *et al.*, Variational spin-squeezing algorithms on programmable quantum sensors, *Phys. Rev. Lett.s* **123**, 260505 (2019).
- [15] R. Kaubruegger, D. V. Vasilyev, M. Schulte, K. Hammerer, and P. Zoller, Quantum variational optimization of Ramsey interferometry and atomic clocks, *Phys. Rev. X* **11**, 2160 (2021).
- [16] B. Yan *et al.*, Observation of dipolar spin-exchange interactions with lattice-confined polar molecules, *Nature* **501**, 521–525 (2013).
- [17] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, *Phys. Rev. Lett.s* **72**, 3439 (1994).
- [18] H. Strobel *et al.*, Fisher information and entanglement of non-gaussian spin states, *Science* **345**, 424–427 (2014).
- [19] M. Cerezo *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* , 1–20 (2021).
- [20] J. J. Meyer, J. Borregaard, and J. Eisert, A variational toolbox for quantum multi-parameter estimation, *NPJ Quantum Inf.* **7**, 1–5 (2021).
- [21] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [22] R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, P. Hyllus, L. Pezzé, and A. Smerzi, Useful multiparticle entanglement and sub-shot-noise sensitivity in experimental phase estimation, *Phys. Rev. Lett.* **107**, 080504 (2011).
- [23] D. Panchenko, Properties of MLE: consistency, asymptotic normality. Fisher information, (unpublished).
- [24] A. Smirne, J. Kołodyński, S. F. Huelga, and R. Demkowicz-Dobrzański, Ultimate precision limits for noisy frequency estimation, *Phys. Rev. Lett.s* **116**, 120801 (2016).
- [25] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States* (Cambridge University Pres, 2006).
- [26] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Pres, 2002).
- [27] E. Andersson, J. D. Cresser, and M. J. W. Hall, Finding the kraus decomposition from a master equation and vice versa, *J. Mod. Opt.* **54**, 1695 (2007).
- [28] A. S. Holevo, A note on covariant dynamical semigroups, *Rep. Math. Phys.* **32**, 211–216 (1993).
- [29] J. R. Johansson, P. D. Nation, and F. Nori, Qutip: An open-source python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **183**, 1760 (2012).
- [30] L. Jiang *et al.*, Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae, *Science* **326**, 267–272 (2009).
- [31] P. Cappellaro and M. D. Lukin, Quantum correlation in disordered spin systems: Applications to magnetic sensing, *Phys. Rev. A* **80**, 032311 (2009).
- [32] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. Heinzen, Squeezed atomic states and projection noise in spectroscopy, *Phys. Rev. A* **50**, 67 (1994).
- [33] L. Pezzé and A. Smerzi, Entanglement, Nonlinear Dynamics, and the Heisenberg Limit, *Phys. Rev. Lett.s* **102**, 100401 (2009).
- [34] P. Hyllus, O. Gühne, and A. Smerzi, Not all pure entangled states are useful for sub-shot-noise interferometry, *Phys. Rev. A* **82**, 012337 (2010).

- [35] E. Pedrozo-Peñafiel *et al.*, Entanglement on an optical atomic-clock transition, *Nature* **588**, 414–418 (2020).
- [36] L. Pezze, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, *Rev. Mod. Phys.* **90**, 035005 (2018).
- [37] G. D. Byrne and A. C. Hindmarsh, A polyalgorithm for the numerical solution of ordinary differential equations, *ACM Trans. Math. Softw.* **1**, 71–96 (1975).
- [38] D. d’Alessandro, *Introduction to quantum control and dynamics* (Chapman and hall/CRC, 2021).
- [39] S. G. Schirmer, H. Fu, and A. I. Solomon, Complete controllability of quantum systems, *Phys. Rev. A* **63**, 063410 (2001).
- [40] N. Khaneja, R. Brockett, and S. J. Glaser, Time optimal control in spin systems, *Phys. Rev. A* **63**, 032308 (2001).
- [41] X. Wang, D. Burgarth, and S. Schirmer, Subspace controllability of spin-1/2 chains with symmetries, *Phys. Rev. A* **94**, 052319 (2016).
- [42] V. Ramakrishna and H. Rabitz, Relation between quantum computing and quantum controllability, *Phys. Rev. A* **54**, 1715 (1996).
- [43] T. Polack, H. Suchowski, and D. J. Tannor, Uncontrollable quantum systems: A classification scheme based on Lie subalgebras, *Phys. Rev. A* **79**, 053403 (2009).
- [44] J. Chen, H. Zhou, C. Duan, and X. Peng, Preparing Greenberger-Horne-Zeilinger and W states on a long-range Ising spin model by global controls, *Phys. Rev. A* **95**, 032340 (2017).
- [45] F. Albertini and D. D’Alessandro, Controllability of symmetric spin networks, *J. Math. Phys.* **59**, 052102 (2018).
- [46] F. Albertini and D. D’Alessandro, Subspace controllability of multi-partite spin networks, *Syst. Control. Lett.* **151**, 104913 (2021).
- [47] Y. Gao, H. Zhou, D. Zou, X. Peng, and J. Du, Preparation of Greenberger-Horne-Zeilinger and w states on a one-dimensional Ising chain by global control, *Phys. Rev. A* **87**, 032335 (2013).
- [48] J. Schachenmayer, A. Pikovski, and A. M. Rey, Many-body quantum spin dynamics with monte carlo trajectories on a discrete phase space, *Phys. Rev. X* **5**, 011022 (2015).
- [49] M. A. Perlin, C. Qu, and A. M. Rey, Spin squeezing with short-range spin-exchange interactions, *Phys. Rev. Lett.* **125**, 223401 (2020).