

# Preparation of Metrological States in Dipolar-Interacting Spin Systems

Tian-Xing Zheng<sup>1,2</sup>, Anran Li<sup>1</sup>, Jude Rosen<sup>1</sup>, Sisi Zhou<sup>1,3</sup>, Martin Koppenhöfer<sup>1</sup>, Ziqi Ma<sup>4,5</sup>, Frederic T. Chong<sup>4</sup>, Aashish A. Clerk<sup>1</sup>, Liang Jiang<sup>1</sup>, and Peter C. Maurer<sup>1</sup>

<sup>1</sup>*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

<sup>2</sup>*Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*

<sup>3</sup>*Institute for Quantum Information and Matter,*

*California Institute of Technology, Pasadena, California 91125, USA*

<sup>4</sup>*Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA and*

<sup>5</sup>*Microsoft, Redmond, Washington 98052, USA*

Spin systems are an attractive candidate for quantum-enhanced metrology. Here we develop a variational method to generate metrological states in small dipolar-interacting ensembles with limited qubit controls and unknown spin locations. The generated states enable sensing beyond the standard quantum limit (SQL) and approaching the Heisenberg limit (HL). Depending on the circuit depth and the level of readout noise, the resulting states resemble Greenberger-Horne-Zeilinger (GHZ) states or Spin Squeezed States (SSS). Sensing beyond the SQL holds in the presence of finite spin polarization and a non-Markovian noise environment.

*Introduction.* Spin systems have emerged as a promising platform for quantum sensing [1–4] with applications ranging from tests of fundamental physics [5,6] to mapping fields and temperature profiles in condensed matter systems and life sciences [3]. Improving the sensitivity of these qubit sensors has so far largely relied on increasing the number of sensing spins and extending spin coherence through material engineering and coherent control. However, with increasing spin density, dipolar interactions between individual sensor spins cause single-qubit dephasing [7,8] and, in the absence of advanced dynamical decoupling [9–11], set a limit to the sensitivity.

Although dipolar interactions in dense spin ensembles lead to complex evolution, they can provide a resource for the creation of metrological states that enable sensing beyond the SQL. Current approaches to create such states (i.e., GHZ states and SSS) either require all-to-all interactions [12–14] or single-qubit addressability [15,16], which are challenging to implement experimentally. An alternative approach that relies on adiabatic state preparation requires less control but results in preparation times that increase exponentially with system size [17,18], leaving this method susceptible to dephasing.

Variational methods provide a powerful tool for controlling many-body quantum systems [19–22]. Such methods have been proposed for Rydberg-interacting atomic systems [23,24] and demonstrated in trapped ions [25]. However, these techniques rely on effective all-to-all interactions (e.g. almost constant interaction strength inside the Rydberg radius [23,26,27]) which are generally absent in solid-state spin ensembles. In this work, we develop a variational algorithm that drives dipolar-interacting spin systems [Fig. 1(a)] into highly entangled states. The resulting states can be subsequently used for Ramsey-interferometry-based single parameter estimation [1]. The required system control relies solely on uniform single-qubit rotations and free evolution under dipolar interactions. The optimization can be directly

performed on an experimental device using only its measurement outcomes without the need to know the spatial distribution of the spins (later referred as ‘spin configuration’). Potential experimental platforms include dipolar-interacting ensembles of NV centers, nitrogen defects in diamond (P1), rare-earth-doped crystals, and ultra-cold molecules.

*Variational Ansatz.* As shown in Fig. 1(b), the variational circuit  $\mathcal{S}(\theta) = \mathcal{U}_m \dots \mathcal{U}_2 \mathcal{U}_1$  is constructed by  $m$  layers of unitary operations. Each  $\mathcal{U}_i$  consists of the parameterized control gates

$$\mathcal{U}_i = R_y\left(\frac{\pi}{2}\right) D(\tau'_i) R_y\left(-\frac{\pi}{2}\right) R_x(\vartheta_i) D(\tau_i), \quad (1)$$

where  $R_\mu(\vartheta) = \exp(-i\vartheta \sum_{j=1}^N S_j^\mu)$  are single-qubit rotations and  $S_j^\mu$  ( $\mu \in \{x, y, z\}$ ) is the  $\mu$  component of the  $j$ -th spin operator.  $D(\tau) = \exp(-i\tau H_{\text{dd}}/\hbar)$  is the time evolution operator of the spin ensemble under dipolar-interaction Hamiltonian  $H_{\text{dd}} = \sum_{i < j} V_{ij}(2S_i^z S_j^z - S_i^x S_j^x - S_i^y S_j^y)$ . The coupling strength between two spins at positions  $\mathbf{r}_i$  and  $\mathbf{r}_j$  is

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

with  $\mu_0$  the vacuum permeability,  $\hbar$  the reduced Planck constant,  $\gamma$  the spin’s gyromagnetic ratio, and  $\beta_{ij}$  the angle between the line segment connecting  $(\mathbf{r}_i, \mathbf{r}_j)$  and the direction of bias magnetic field. An evolutionary algorithm [29,30] is applied on the  $m$ -layer circuit which contains  $3m$  free parameters constituting the vector  $\theta = (\tau_1, \vartheta_1, \tau'_1, \dots, \tau_i, \vartheta_i, \tau'_i, \dots, \tau_m, \vartheta_m, \tau'_m)$ . Each  $\tau_i$  is restricted to  $\tau_i \in [0, 1/\bar{f}_{\text{dd}}]$  where  $\bar{f}_{\text{dd}}$  is the average nearest-neighbor interaction strength for the considered spin configuration. The Ansatz in Eq. (1) is the most general set of global single-qubit gates that preserves the initial collective spin direction  $\langle \sum_i \mathbf{S}_i \rangle / |\langle \sum_i \mathbf{S}_i \rangle|$ , here chosen to be the  $x$ -direction [23,30]. Although this Ansatz

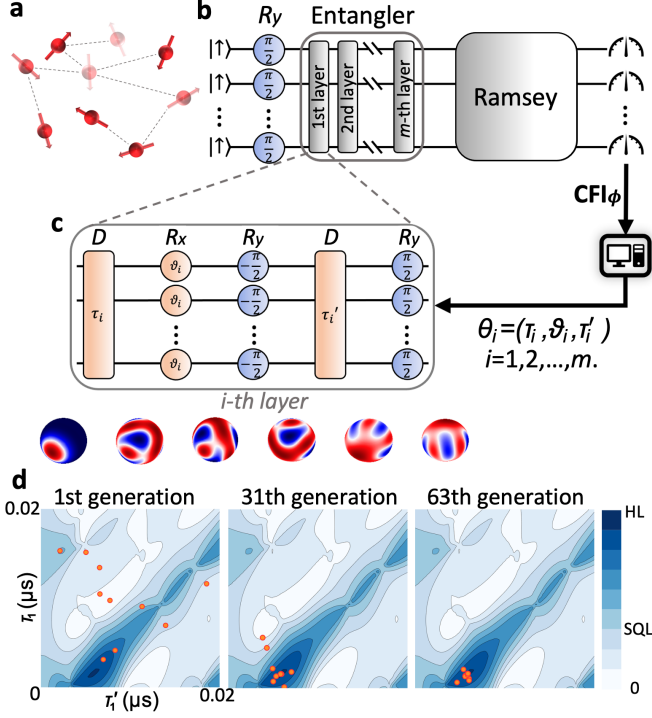


FIG. 1. (a) Schematic of a dipolar-interacting spin ensemble in a 3D-random configuration. (b) The quantum circuit consists of three parts: a sequence for generating entanglement (entangler), phase accumulation (Ramsey) and single-qubit readout in the  $S_z$  basis. Dipolar interactions during Ramsey interference are eliminated by dynamical decoupling [79,28]. The measurement outcome is processed on a classical computer and used to determine the next generation for  $\theta$ . (c) Gate sequence of each variational layer and the Wigner distributions for a 5-spin state after each gate. (d) Illustration of an optimization process on a 3-spin system with  $m = 1$ . The contour plots show the 2D projection of the multidimensional  $\theta$  space for fixed  $\vartheta_1$ . The orange points mark the sampling positions in the parameter space. Convergence to the global maximum is reached in the 63th generation.

does not enable universal system control [30–33], we show that with increasing circuit depth, sensing near the HL can be achieved.

**Metrological cost function.** The Ramsey protocol shown in Fig. 1(b) encodes the quantity of interest in the accumulated phase  $\phi = \omega t_R$ , with  $\omega$  the detuning frequency and  $t_R$  the Ramsey sensing time. The Classical Fisher Information (CFI) [1,30] quantifies how precisely one can estimate an unknown parameter  $\phi$  under a measurement basis. Our variational approach treats the spin systems as a black-box for which the algorithm finds a control sequence that maximizes the CFI associated with the parameter estimation problem

$$\text{CFI}_\phi = \sum_z \text{Tr}[P_z \rho_\phi] \left( \frac{\partial \log \text{Tr}[P_z \rho_\phi]}{\partial \phi} \right)^2. \quad (3)$$

The sum runs over the  $2^N$  basis states  $|z\rangle = \otimes_{i=1}^N |s_i^z\rangle$ , where  $s_i^z$  are the eigenvalues of  $S_i^z$ .  $P_z \equiv |z\rangle \langle z|$  denotes the corresponding measurement operator and  $\rho_\phi$  the density matrix. The CFI is chosen as cost function because it is a measure for the maximal achievable sensitivity for a given measurement basis [1,34]. Likewise,  $P_z$  provides the maximal information that can be gained from single-qubit measurements. However, measurement operators such as parity or total spin polarization result in a smaller outcome space and are therefore more efficient in experimental implementations. While this study optimizes the measurement for  $P_z$ , the obtained results likewise hold for parity and total spin polarization [30].

**Numerical results for regular and disordered spin configurations.** We start by testing our approach for three distinct regular spin configurations. Figure 2(a) shows the CFI after optimization for spins arranged on a linear chain (blue), a two-dimensional (2D) square lattice (orange), and a circle (green). All three configurations result in states with CFI above the SQL. When multiple circuit layers are added, the CFI further improves. Next, we simulate the case of disordered three-dimensional (3D) spin configurations (later referred as 3D-random). In our simulations the spins are randomly located in a box of length  $L \propto N^{1/3}$  (constant spin density). Compared to the regular spin array, the disordered case shows a noticeable saturation of the CFI as a function of  $N$ . With increased circuit depth, sensing precision beyond the SQL

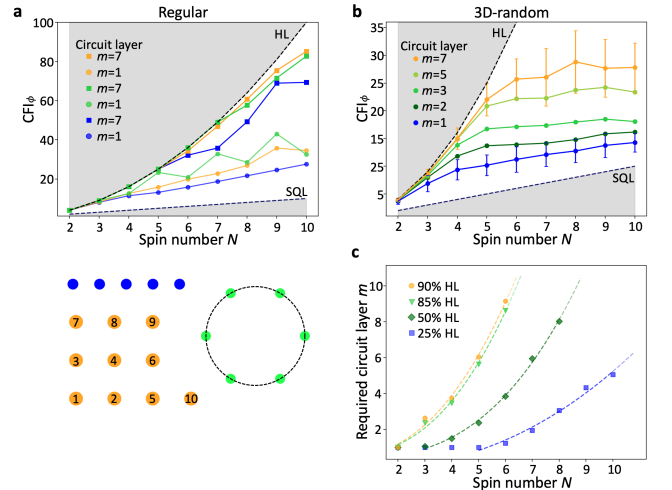


FIG. 2. (a) Top: CFI for  $m = 1$  (circles) and  $m = 7$  (squares) circuits. The colors correspond to the configurations shown on the left. Bottom: schematics of different spin configurations. The numbers in the 2D square lattice pattern label the order in which spins are added to form a lattice of size  $N$ . (b) Average CFI for 50 configurations of 3D-randomly distributed spins. (c) Average number of layers required to achieve a CFI within a given percentage of the HL in the case of 3D-random configuration. The fit  $m = aN^b + c$  with  $b = 2.45$  (goodness of fit  $R^2 = 0.996$ ) serve as a guide to the eye. The same data also fits to an exponential model with slightly lower  $R^2 = 0.995$ .

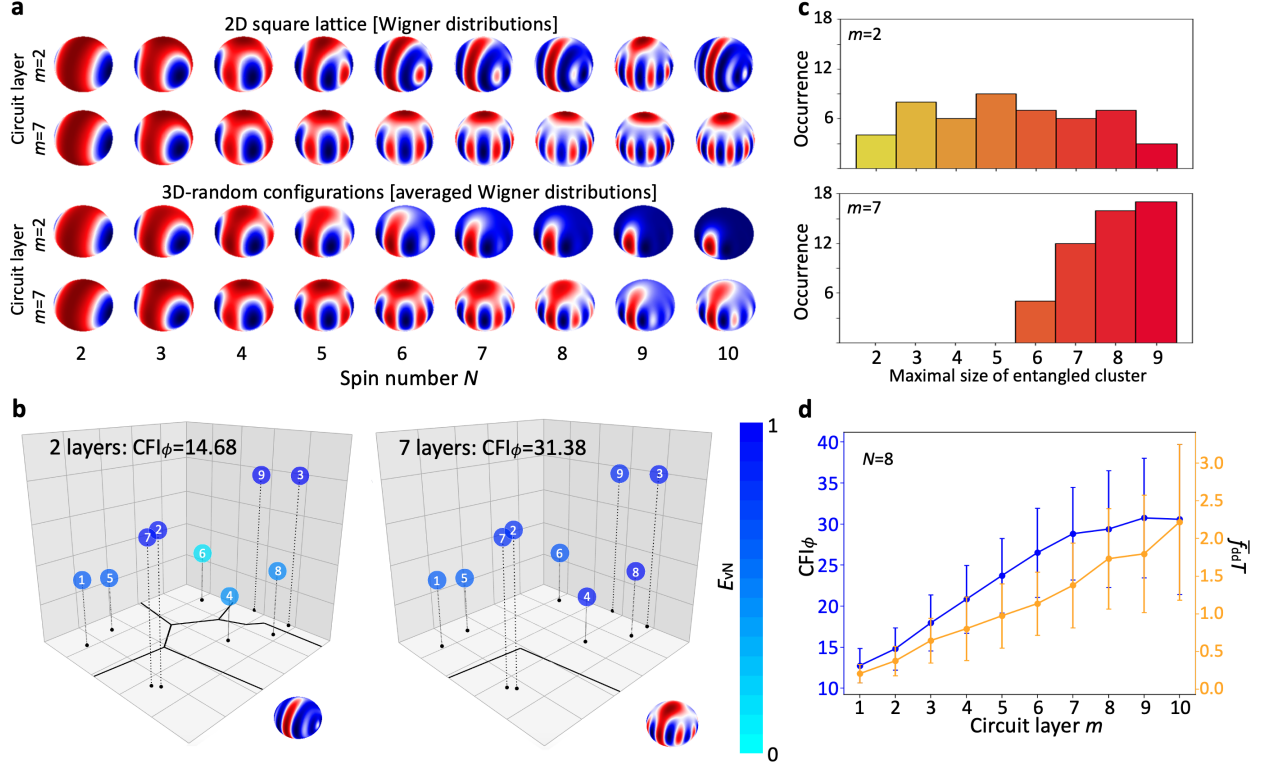


FIG. 3. (a) Wigner distributions versus spin number for  $m = 2$  and  $m = 7$  in the case of 2D square lattice and 3D-random configurations. (b) von-Neumann entanglement entropy for one specific 3D-random configuration of 9 spins for  $m = 2$  and  $m = 7$ . Individual spins are colored according to their von-Neumann entropies. Entangled clusters are marked by solid black lines. (c) Histograms depicting the maximal size of entangled clusters for 50 3D-random configurations. (d) Average CFI (blue) and state preparation time (orange) versus  $m$ . The state preparation time is given as a unitless quantity  $\bar{f}_{\text{dd}}T$  with  $T = \sum_{i=1}^m (\tau_i + \tau'_i)$ .

is maintained. The required circuit depth increases drastically with  $N$  [Fig. 2(c)].

*Characterization of entanglement.* We investigate the  $N$ -qubit entangled states created by our variational method. Figure 3(a) shows the corresponding Wigner distributions [35–37] for a regular 2D spin array (top) and the average Wigner distributions for 50 different 3D-random spin configurations (bottom). In both cases, the optimized states resemble GHZ states when  $N$  is small and  $m$  is large. For large  $N$  and small  $m$ , the states are close to SSS. Non-Gaussian states that provide sensitivity beyond the SSS but lower than GHZ states are also generated. Our algorithm tends to drive the systems into a GHZ state, as it has the unique property of attaining the HL in Ramsey spectroscopy [38].

For quantitatively analysing the buildup of entanglement, the von-Neumann entanglement entropy ( $E_{\text{vN}} = -\text{Tr}(\rho_s \log_2 \rho_s)$ ) [39] is used as a measure for the degree of entanglement between a spin subsystem ( $\rho_s = \text{Tr}_s \rho_{\text{tot}}$ ) and the remaining system. As an example, we explore one case of a 3D-random configuration of 9 spins. Figure 3(b) shows the von-Neumann entropy of each spin after employing a 2-layer circuit (left) and a 7-layer cir-

cuit (right). In the case of  $m = 2$ , the achieved degree of entanglement is modest with spin No.6 for example showing no substantial entanglement with the remaining spins. When the circuit depth is increased to 7, all spins display substantial entanglement. While the single-particle entropy detects spins unentangled with the remaining system, it does not determine whether all spins are entangled with each other or entanglement is local. We distinguish these two scenarios by identifying the smallest clusters with  $E_{\text{vN}} \leq 0.4$ . For  $m = 2$ , the spin ensemble segments into 5 clusters [Fig. 3(b)], while for  $m = 7$  only 2 clusters are found. The results verify that multiple layers are required to overcome the anisotropy of the dipolar interaction (Eq. (2)) when building up entanglement over the entire system. Finally, in Fig. 3(c) we analyze the size of the largest cluster for each of the 50 spin configurations and observe an overall increase of the largest cluster size and a decrease of the variance.

*State preparation time.* Minimizing the preparation time is central in practical applications, as it increases bandwidth, reduces decoherence, and enables more measurement repetitions [1]. Figure 3(d) shows the average state preparation time for 8 spins in 50 different 3D-

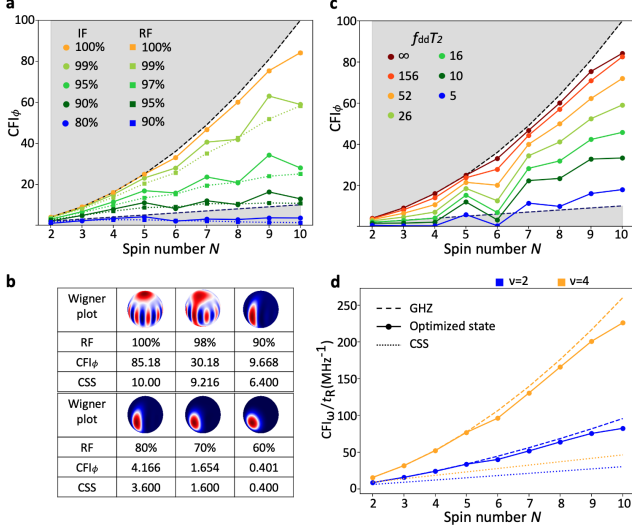


FIG. 4. (a)  $\text{CFI}_\phi$  under finite initialization fidelity (IF) and readout fidelity (RF). Circuit depth  $m = 5$ .  $\text{IF} = \frac{N_\uparrow - N_\downarrow}{N}$ , where  $N_\uparrow$  ( $N_\downarrow$ ) denotes the number of spins in the  $|\uparrow\rangle$  ( $|\downarrow\rangle$ ) state at the beginning of the sensing protocol in Fig. 1(b).  $\text{RF} = 1 - p(\downarrow | \uparrow) = 1 - p(\uparrow | \downarrow)$ . (b) Optimized Wigner functions and  $\text{CFI}_\phi$  of a 10-spin state versus different readout fidelities (RF). For comparison, the row ‘CSS’ represents the  $\text{CFI}_\phi$  for a coherent spin state given the same RF. (c)  $\text{CFI}_\phi$  in the presence of decoherence in the entangler. (d) Ramsey protocol’s results of the generated states when considering non-Markovian noise during signal accumulation. Data correspond to the optimized states from 2D square lattice configuration.

random configurations as a function of layer number. The preparation time increases with the layer number and is inversely proportional to the average dipole coupling strength of the nearest-neighbour spins  $f_{\text{dd}}$ . Compared to adiabatic methods [17], our approach results in an  $11\times$  reduction of the preparation time to reach the same CFI for identical spin number and density [30].

*State preparation under decoherence, initialization and readout errors* Until now our analysis assumed full coherence and perfect spin initialization and readout. However, dephasing, initialization, and readout errors will be limiting factors in experimental implementations. We next examine the impact of such imperfections on state preparation and sensing. Figure 4(a) shows the CFI in the presence of imperfect initialization and finite readout fidelity for spins on a 2D square lattice. For  $N \leq 10$ , beyond-SQL precision is reached with 90% initialization and 95% readout fidelity, respectively.

Next, we investigate the resulting states when readout errors are added into the optimizer. Figure 4(b) indicates that without readout errors, the Wigner distribution of the resulting state is close to a GHZ state. However, with a finite readout error rate, our algorithm drives the system into a state resembling a SSS. When the read-

out noise is further increased, the SSS transforms into a coherent spin state (CSS). The results agree with the fact that GHZ states are sensitive to single-spin readout errors while SSS are more robust [40].

During the state preparation, decoherence ( $T_2$ ) reduces entanglement. We assume independent, Markovian dephasing of each spin as described by a Lindblad master equation [39]. Figure 4(c) shows the CFI for various  $T_2$  times using the previously optimized gate parameters for 2D square lattice. While a finite  $T_2$  decreases the CFI, coherence times exceeding  $5/f_{\text{dd}}$  result in states with beyond-SQL sensitivity. Here,  $f_{\text{dd}}$  denotes the nearest-neighbor interaction strength for 2D square lattice. Since performing optimization with imperfections is numerically expensive, the results in Fig. 4(a), (c), (d) are obtained by optimizing the parameters in the absence of imperfections and using those parameters to compute the CFI under imperfect conditions. Thus, better results are expected if the optimization is directly run on experiments.

*Sensitivity in a non-Markovian environment.* In addition to impacts on state preparation, dephasing affects performance in Ramsey interferometry. In the presence of spatially uncorrelated Markovian noise, entanglement does not lead to a beyond-SQL scaling [48,49]. In a non-Markovian environment, such as that of a solid-state spin system, this limitation does not hold [50,51]. We examine the performance of our optimized states in a non-Markovian noise environment. We adopt a noise model [50] in which the amplitude of single-spin coherence reduces according to

$$\rho_{01}(t) = \rho_{01}(0)e^{-\left(\frac{t}{T_2}\right)^\nu} \quad (4)$$

where  $\nu$  is the stretch factor set by the noise properties. The time evolution under Ramsey propagation is simulated with a generalized Lindblad master equation [30,51]. The sensing performance of optimized states is characterized by the square of the signal-noise-ratio  $\text{SNR}^2 \propto \text{CFI}_w/t_R$  [30]. Figure 4(d) shows their performance compared to the CSS and the GHZ states for a  $\nu = 2$  and  $\nu = 4$  noise exponent [8]. The created entangled states provide an advantage over uncorrelated states. For small spin numbers, the SNR follows the HL scaling [50].

*Proposed experimental platforms.* Candidate systems for realizing the proposed variational approach need to possess long  $T_2$  coherence time, strong dipolar-interacting strength, and high initialization and readout fidelity. Recent developments in solid-state spin systems and ultracold molecules have demonstrated coherence times that exceed dipolar coupling times ( $1/f_{\text{dd}}$ ) as well as high-fidelity spin initialization and readout. Table S1 lists the experimentally observed parameters for different candidate systems, including Nitrogen Vacancy (NV) ensembles, P1 centers in diamond, rare-earth doped



TABLE I. Experimental platforms' relative parameters

| Systems             | $\tilde{f}_{dd}$ | $T_2^{(DD)}$        | $\tilde{f}_{dd}T_2^{(DD)}$ | $P_{ini}$  | $F_{readout}$ | $\nu$              |
|---------------------|------------------|---------------------|----------------------------|------------|---------------|--------------------|
| NV ensemble         | 35kHz [28]       | 7.9(2) $\mu$ s [28] | 0.28                       | 97.5% [41] | 97.5% [41]    | 2 – 4 [8]          |
| P1 centers          | 0.92MHz [42]     | 4.4 $\mu$ s [42]    | 4.0                        | 95% [43]   | 95% [43]      | ?                  |
| Rare-Earth crystals | 1.96MHz [44]     | 2.5 $\mu$ s [44]    | 4.9                        | 97% [15]   | 94.6% [45]    | 2.4 $\pm$ 0.1 [46] |
| Cold Molecules      | 52Hz [7]         | 80ms [7]            | 4.16                       | 97% [47]   | 97% [47]      | ?                  |

crystals, and ultracold molecule tweezer systems [30].

*Conclusion and Outlook.* This work introduces a variational circuit that generates entangled metrological states in a dipolar-interacting spin system without requiring knowledge of the actual spin configuration. The required system parameters are within the reach of several experimental platforms. While this study remains limited to small system sizes ( $N \leq 10$ , limited by computational resource), our results are of immediate interest to nanoscale quantum sensing where spatial resolution is paramount and the finite sensor size limits the number of spins that can be utilized. Extending our results to  $N > 10$  can either be achieved by utilizing symmetries in regular arrays or directly testing our optimization algorithms on an actual experimental platform. The developed method is also potentially applicable for preparing other relevant highly entangled states in quantum computing and quantum communication.

We thank D. DeMille, D. Freedmann, A. Bleszynski Jayich, S. Kolkowitz, T. Li, Z. Li, R. Kaubruegger, Y. Huang, Q. Xuan, Z. Zhang, S. von Kugelgen, C-J. Yu, Y. Bao, and P. Gokhale for helpful discussions. T-X.Z., M.K., F.T.C., A.C., L.J. and P.M. acknowledge support by Q-NEXT Grant No. DOE 1F-60579. T-X.Z. and P.M. acknowledge support by National Science Foundation (NSF) Grant No. OMA-1936118 and OIA-2040520, and NSF QuBBE QLCI (NSF OMA- 2121044). S.Z. acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907). Z.M. and F.T.C acknowledge support by EPiQC, an NSF Expedition in Computing, under grants CCF-1730082/1730449; in part by STAQ under grant NSF Phy-1818914; in part by the US DOE Office of Advanced Scientific Computing Research, Accelerated Research for Quantum Computing Program.; and in part by NSF OMA-2016136. The authors are also grateful for the support of the University of Chicago Research Computing Center for assistance with the numerical simulations carried out in this work.

Disclosure: F.T.C. is the Chief Scientist for Super.tech and an advisor to QCI.

[1] C. L. Degen, F. Reinhard, and P. Cappellaro, Reviews of modern physics **89**, 035002 (2017).

[2] J. F. Barry, J. M. Schloss, E. Bauch, M. J. Turner, C. A. Hart, L. M. Pham, and R. L. Walsworth, Reviews of Modern Physics **92**, 015004 (2020).

[3] R. Schirhagl, K. Chang, M. Loretz, and C. L. Degen, Annual review of physical chemistry **65**, 83 (2014).

[4] J.-P. Tetienne, Nature Physics **17**, 1074 (2021).

[5] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenber, R. F. Vermeulen, R. N. Schouten, C. Abellán, *et al.*, Nature **526**, 682 (2015).

[6] G. E. Marti, R. B. Hutson, A. Goban, S. L. Campbell, N. Poli, and J. Ye, Physical review letters **120**, 103201 (2018).

[7] B. Yan, S. A. Moses, B. Gadway, J. P. Covey, K. R. Hazzard, A. M. Rey, D. S. Jin, and J. Ye, Nature **501**, 521 (2013).

[8] L. Childress, M. G. Dutt, J. Taylor, A. Zibrov, F. Jelezko, J. Wrachtrup, P. Hemmer, and M. Lukin, Science **314**, 281 (2006).

[9] J. S. Waugh, L. M. Huber, and U. Haeberlen, Physical Review Letters **20**, 180 (1968).

[10] M. Mehring, *Principles of high resolution NMR in solids* (Springer Science & Business Media, 2012).

[11] J. Choi, H. Zhou, H. S. Knowles, R. Landig, S. Choi, and M. D. Lukin, Physical Review X **10**, 031002 (2020).

[12] E. Pedrozo-Peñañiel, S. Colombo, C. Shu, A. F. Adiyatullin, Z. Li, E. Mendez, B. Braverman, A. Kawasaki, D. Akamatsu, Y. Xiao, and V. Vuletić, Nature **588**, 414 (2020).

[13] M. Kitagawa and M. Ueda, Physical Review A **47**, 5138 (1993).

[14] T. Bilitewski, L. De Marco, J.-R. Li, K. Matsuda, W. G. Tobias, G. Valtolina, J. Ye, and A. M. Rey, Physical Review Letters **126**, 113401 (2021).

[15] S. Chen, M. Raha, C. M. Phenicie, S. Ourari, and J. D. Thompson, Science **370**, 592 (2020).

[16] P. Neumann, N. Mizuochi, F. Rempp, P. Hemmer, H. Watanabe, S. Yamasaki, V. Jacques, T. Gaebel, F. Jelezko, and J. Wrachtrup, science **320**, 1326 (2008).

[17] P. Cappellaro and M. D. Lukin, Physical Review A **80**, 032311 (2009).

[18] S. Choi, N. Y. Yao, and M. D. Lukin, arXiv:1801.00042 (2017).

[19] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Nature Reviews Physics , 1 (2021).

[20] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, *et al.*, Nature **569**, 355 (2019).

[21] Y. Li and S. C. Benjamin, Physical Review X **7**, 021050 (2017).

[22] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Physical Review X **10**, 021067 (2020).

[23] R. Kaubruegger, P. Silvi, C. Kokail, R. van Bijnen, A. M.

- Rey, J. Ye, A. M. Kaufman, and P. Zoller, *Physical review letters* **123**, 260505 (2019).
- [24] R. Kaubruegger, D. V. Vasilyev, M. Schulte, K. Hammerer, and P. Zoller, *Physical review X* **11**, 2160 (2021).
- [25] C. D. Marciniak, T. Feldker, I. Pogorelov, R. Kaubruegger, D. V. Vasilyev, R. van Bijnen, P. Schindler, P. Zoller, R. Blatt, and T. Monz, *arXiv:2107.01860* (2021).
- [26] V. Borish, O. Marković, J. A. Hines, S. V. Rajagopal, and M. Schleier-Smith, *Physical review letters* **124**, 063601 (2020).
- [27] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, *et al.*, *Nature* **551**, 579 (2017).
- [28] H. Zhou, J. Choi, S. Choi, R. Landig, A. M. Douglas, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, P. Cappelaro, H. S. Knowles, H. Park, and M. D. Lukin, *Physical review X* **10**, 031003 (2020).
- [29] N. Hansen, *arXiv:1604.00772* (2016).
- [30] See Supplemental Material for additional details.
- [31] S. G. Schirmer, H. Fu, and A. I. Solomon, *Physical Review A* **63**, 063410 (2001).
- [32] F. Albertini and D. D'Alessandro, *Linear algebra and its applications* **350**, 213 (2002).
- [33] F. Albertini and D. D'Alessandro, *ScienceDirect: Systems & Control Letters* **151**, 104913 (2021).
- [34] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, random processes, and statistical analysis* (Cambridge University Press, 2011).
- [35] M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, *Physics reports* **106**, 121 (1984).
- [36] J. P. Dowling, G. S. Agarwal, and W. P. Schleich, *Physical Review A* **49**, 4101 (1994).
- [37] B. Koczor, R. Zeier, and S. J. Glaser, *Physical Review A* **102**, 062421 (2020).
- [38] V. Giovannetti, S. Lloyd, and L. Maccone, *Physical review letters* **96**, 010401 (2006).
- [39] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
- [40] E. Davis, G. Bentsen, T. Li, and M. Schleier-Smith, *Advances in Photonics of Quantum Computing, Memory, and Communication X* **10118**, 101180Z (2017).
- [41] B. J. Shields, Q. P. Unterreithmeier, N. P. de Leon, H. Park, and M. D. Lukin, *Physical review letters* **114**, 136402 (2015).
- [42] C. Zu, F. Machado, B. Ye, S. Choi, B. Kobrin, T. Mittiga, S. Hsieh, P. Bhattacharyya, M. Markham, D. Twitchen, A. Jarmola, D. Budker, C. R. Laumann, J. E. Moore, and N. Y. Yao, *Nature* **597**, 45–50 (2021).
- [43] M. Degen, S. Loenen, H. Bartling, C. Bradley, A. Meisma, M. Markham, D. Twitchen, and T. Taminiau, *Nature Communications* **12**, 3470 (2021).
- [44] B. Merkel, P. C. Fariña, and A. Reiserer, *Physical Review Letters* **127**, 030501 (2021).
- [45] M. Raha, S. Chen, C. M. Phenicie, S. Ourari, A. M. Dibos, and J. D. Thompson, *Nature communications* **11**, 1605 (2020).
- [46] M. Le Dantec, M. Rančić, S. Lin, E. Billaud, V. Ranjan, D. Flanigan, S. Bertaina, T. Chanière, P. Goldner, A. Erb, *et al.*, *Science advances* **7**, eabj9786 (2021).
- [47] L. W. Cheuk, L. Anderegg, Y. Bao, S. Burchesky, S. S. Yu, W. Ketterle, K.-K. Ni, and J. M. Doyle, *Physical review letters* **125**, 043401 (2020).
- [48] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guță, *Nature communications* **3**, 1063 (2012).
- [49] B. Escher, R. de Matos Filho, and L. Davidovich, *Nature Physics* **7**, 406 (2011).
- [50] A. W. Chin, S. F. Huelga, and M. B. Plenio, *Physical review letters* **109**, 233601 (2012).
- [51] A. Smirne, J. Kołodyński, S. F. Huelga, and R. Demkowicz-Dobrzański, *Physical review letters* **116**, 120801 (2016).
- [52] G. Kucsko, S. Choi, J. Choi, P. C. Maurer, H. Zhou, R. Landig, H. Sumiya, S. Onoda, J. Isoya, F. Jelezko, *et al.*, *Physical review letters* **121**, 023601 (2018).
- [53] D. Barskiy, "Dipole-dipole interactions in NMR: explained." (unpublished).
- [54] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, *et al.*, *Nature* **543**, 221 (2017).
- [55] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* (MIT press, 2018).
- [56] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, *nature* **550**, 354 (2017).
- [57] P. Peng, X. Huang, C. Yin, L. Joseph, C. Ramanathan, and P. Cappelaro, *arXiv:2102.13161* (2021).
- [58] S. L. Braunstein and C. M. Caves, *Physical Review Letters* **72**, 3439 (1994).
- [59] H. Strobil, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi, and M. K. Oberthaler, *Science* **345**, 424 (2014).
- [60] J. J. Meyer, J. Borregaard, and J. Eisert, *NPJ Quantum Information* **7**, 1 (2021).
- [61] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Physical Review A* **99**, 032331 (2019).
- [62] D. Panchenko, "Properties of MLE: consistency, asymptotic normality. Fisher information." (unpublished).
- [63] A. S. Holevo, *Reports on mathematical physics* **32**, 211 (1993).
- [64] J. R. Johansson, P. D. Nation, and F. Nori, *Computer Physics Communications* **183**, 1760 (2012).
- [65] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. Heinzen, *Physical Review A* **50**, 67 (1994).
- [66] L. Pezzè and A. Smerzi, *Physical review letters* **102**, 100401 (2009).
- [67] P. Hyllus, O. Gühne, and A. Smerzi, *Physical Review A* **82**, 012337 (2010).
- [68] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, *Reviews of Modern Physics* **90**, 035005 (2018).
- [69] D. d'Alessandro, *Introduction to quantum control and dynamics* (Chapman and hall/CRC, 2021).
- [70] X. Wang, D. Burgarth, and S. Schirmer, *Physical Review A* **94**, 052319 (2016).
- [71] V. Ramakrishna and H. Rabitz, *Physical Review A* **54**, 1715 (1996).
- [72] T. Polack, H. Suchowski, and D. J. Tannor, *Physical Review A* **79**, 053403 (2009).
- [73] J. Chen, H. Zhou, C. Duan, and X. Peng, *Physical Review A* **95**, 032340 (2017).
- [74] F. Albertini and D. D'Alessandro, *Journal of Mathematical Physics* **59**, 052102 (2018).
- [75] F. Albertini and D. D'Alessandro, *Systems & Control Letters* **151**, 104913 (2021).
- [76] Y. Gao, H. Zhou, D. Zou, X. Peng, and J. Du, *Physical Review A* **87**, 032335 (2013).

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

## CONTENTS

|  |    |
|--|----|
| Designing the variational circuit  | 2  |
| Entanglement generation gates from two-body interaction Hamiltonian and global rotations | 2  |
| Preservation of the collective spin direction  | 2  |
| Choosing the most general gate set   | 3  |
| Optimization algorithm: Covariance Matrix Adaptation Evolution Strategy                  | 3  |
| Optimization results of different types of dipole-dipole interaction Hamiltonian         | 3  |
| NV ensemble  | 4  |
| Ising type interaction (large local disorder)  | 5  |
| Optimization results by using $P_z^{\text{tot}}$ , $P_z^\pi$ as measurement bases        | 5  |
| Supplementary Data   | 6  |
| Complete CFI data for Fig.2 in main text   | 6  |
| Required layers to reach given CFI for 2D square lattice                                 | 7  |
| Orders of interaction  | 8  |
| Non-Markovian noise sensing performance  | 8  |
| Optimized states with different readout fidelity   | 9  |
| Relative experimental parameter table (full)   | 9  |
| Supplementary Derivations  | 10 |
| CFI with respect to angle and frequency  | 10 |
| Relation between $\text{CFI}_\omega$ and SNR in single qubit Ramsey experiment           | 11 |
| Maximum Likelihood Estimator   | 12 |
| Master equation for a non-Markovian environment  | 14 |
| Performance of metrological states in a non-Markovian environment                        | 16 |
| Time Overhead  | 17 |
| State preparation time comparing to adiabatic method                                     | 18 |
| Controllability  | 19 |
| Controllability Test   | 19 |
| Controllability of Dipolar Interacting Spin Systems                                      | 19 |
| Finding Reachable States   | 20 |

## DESIGNING THE VARIATIONAL CIRCUIT

In this section, we discuss how to choose the experimentally realizable elementary gates in the variational sequence in the entangler based on limited quantum resource [23, 24].

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

Consider a two-body interaction Hamiltonian:

$$H_{\text{int}} = \sum_{i < j} V_{ij} (J^I S_{zi} S_{zj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j). \quad (\text{S1})$$

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

$$D_z(\tau) \equiv \exp(-i\tau H_{\text{int}}/\hbar) = \exp \left[ -i\tau \sum_{i < j} V_{ij} (J^I S_{zi} S_{zj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j) / \hbar \right]. \quad (\text{S2})$$

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

$$\begin{aligned} D_{x,y}(\tau) &= R_{y,x}(\pi/2) D_z(\tau) R_{y,x}(-\pi/2) \\ &= \exp \left[ -i\tau \sum_{i < j} V_{ij} (J^I S_{x,yi} S_{x,yj} + J^S \mathbf{S}_i \cdot \mathbf{S}_j) / \hbar \right]. \end{aligned} \quad (\text{S3})$$

In Eqs. (S2) (S3), 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

$$\begin{aligned} \mathcal{U}_i &= R_y \left( \frac{\pi}{2} \right) D(\tau'_i) R_y \left( -\frac{\pi}{2} \right) R_x(\vartheta_i) D(\tau_i) \\ &= D_x(\tau'_i) R_x(\vartheta_i) D_z(\tau_i). \end{aligned} \quad (\text{S4})$$

In the next two subsections, it will be shown that the sequence in Eq. (S4) is the most general gate set that uses only global rotations and preserves the collective spin direction along  $x$ -direction.

### Preservation of the collective spin direction

Define the  $x$ -parity operator  $P_x \equiv \Pi_i^N \sigma_{xi} = P_x^\dagger$ , with  $P_x^2 = I$ . This operator describes the parity of a state in  $x$ -direction and is related to the global  $\pi$  rotation along  $x$ -axis up to a phase constant,  $R_x(\pi) = \exp(-i\pi \sum_i S_{xi}) = (-i)^N \Pi_i^N \sigma_{xi}$ . 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

$$\begin{aligned} P_x \mathcal{S}(\boldsymbol{\theta}) P_x &= P_x \mathcal{U}_m \dots \mathcal{U}_2 \mathcal{U}_1 P_x \\ &= P_x \Pi_i [D_x(\tau'_i) R_x(\vartheta_i) D_z(\tau_i)] P_x \\ &= \mathcal{S}(\boldsymbol{\theta}). \end{aligned} \quad (\text{S5})$$



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  $P_x$ :

$$\begin{aligned} P_x |\Psi(\theta)\rangle &= P_x \mathcal{S}(\theta) |\uparrow_x\rangle^{\otimes N} \\ &= P_x \mathcal{S}(\theta) P_x P_x |\uparrow_x\rangle^{\otimes N} \\ &= |\Psi(\theta)\rangle. \end{aligned} \quad (\text{S6})$$

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

$$\begin{aligned} \langle J_{y,z} \rangle &= \langle \Psi(\theta) | J_{y,z} | \Psi(\theta) \rangle \\ &= \langle \Psi(\theta) | P_x P_x J_{y,z} P_x P_x | \Psi(\theta) \rangle \\ &= -\langle J_{y,z} \rangle = 0. \end{aligned} \quad (\text{S7})$$

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

### 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.[23]. However, the interaction Hamiltonian we discuss here is more general. In Eq. (S1), when  $J^I = 1, J^S = 0$ , the interaction becomes Ising type interaction which is equivalent to the Rydberg interaction in ref.[23,24]. 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. (S1) 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 [7].

### OPTIMIZATION ALGORITHM: COVARIANCE MATRIX ADAPTATION EVOLUTION STRATEGY

The optimization in the  $3m$  dimensional parameter space is highly non-convex (Fig.1(d) in main text) due to the large inhomogeneity of the interaction strength. In our setting, the previously used Dividing Rectangles algorithm [23,20] cannot converge to a beyond-SQL result despite large number of iterations. We address this challenge by using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) as our optimization algorithm [29]. CMA-ES balances the exploration and exploitation process when searching in the parameter space so that convergence is reached after less than approximately 2,000 generations for  $N, m \leq 10$ . This corresponds to about  $10^8$  repetitions of the Ramsey experiment, which can be further reduced if collective measurement observables are measured.

We reduce the complexity of the optimization by restricting  $\tau_i$  within  $[0, 1/\bar{f}_{\text{dd}}]$  where  $\bar{f}_{\text{dd}}$  is the average nearest-neighbor interaction strength for the considered spin configuration. Setting a large parameter searching range for the interaction gates' time  $\tau_i$  would potentially ensure the global maximum CFI location is included in the parameter space. However, when the upper bound of  $\tau_i$  is much bigger than  $1/\bar{f}_{\text{dd}}$ , the evolution of neighboring spin pairs is fast when sweeping  $\tau_i$ . This would introduce a huge amount of local maximum points in the parameter searching so that it is impractical for the black-box optimization algorithm to converge to that global maximum point.

### OPTIMIZATION RESULTS OF DIFFERENT TYPES OF DIPOLE-DIPOLE INTERACTION HAMILTONIAN

The magnetic dipole-dipole interaction Hamiltonian under secular approximation has the general form [52, 53]:

$$H_{\text{dd}} = \sum_{i < j} V_{ij} (2S_{zi}S_{zj} - S_{xi}S_{xj} - S_{yi}S_{yj}) \quad (\text{S8})$$

with

$$V_{ij} = \sum_{i < j} \frac{\mu_0}{4\pi} \frac{\gamma_i \gamma_j \hbar^2}{|\mathbf{r}_i - \mathbf{r}_j|^3} \frac{(1 - 3 \cos \beta_{ij})}{2} \quad (\text{S9})$$

where  $\mu_0$  is the vacuum permeability,  $\gamma$  is the geomagnetic ratio of the spin,  $\beta_{ij}$  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. (S8) is able to describe the dipolar interaction for the spin systems with arbitrary spin number as long as the spin angular momentum operators  $S_\mu$  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.

### 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

$$S_x^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, S_y^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, S_z^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (\text{S10})$$

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:

$$S_y^{(1)} = \sqrt{2}S_y^{(\frac{1}{2})}, S_x^{(1)} = \sqrt{2}S_x^{(\frac{1}{2})}, S_z^{(1)} = \frac{I}{2} + S_z^{(\frac{1}{2})} \quad (\text{S11})$$

Plugging Eq. (S11) into Eq. (S8), we get the effective dipole-dipole interaction Hamiltonian for NV ensemble  $|m_s = 1\rangle, |m_s = 0\rangle$  subspace [52,54]:

$$H_{\text{DD,NV}} = \sum_{i < j} V_{ij} (S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{xi}^{(\frac{1}{2})} S_{xj}^{(\frac{1}{2})} - S_{yi}^{(\frac{1}{2})} S_{yj}^{(\frac{1}{2})}) \quad (\text{S12})$$

Fig. S1(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.

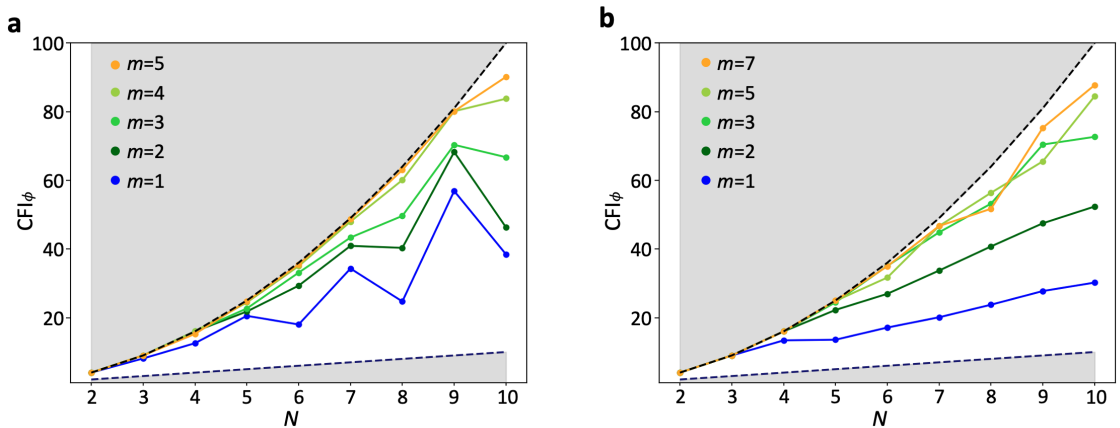


FIG. S1. (a) Optimization results for NV-ensemble. The CFI saturates the theoretical upper bound Heisenberg Limit (HL) when the variational circuit layer number goes up from 1 to 7, and the CFI results are ‘oscillating’ for even/odd number of spins from shallow circuit. (b) Optimization results for Ising type spin interaction when there is large local disorder in the system.

### Ising type interaction (large local disorder)

When the system has large local disorder, the flip-flop terms in the dipolar interaction Hamiltonian Eq. (S8) are suppressed because of the large energy gap:

$$\begin{aligned}
H_{\text{DD,Ising}} &= \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i<j} V_{ij} (2S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{xi}^{(\frac{1}{2})} S_{xj}^{(\frac{1}{2})} - S_{yi}^{(\frac{1}{2})} S_{yj}^{(\frac{1}{2})}) \\
&= \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i<j} 2V_{ij} (S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} - S_{+i}^{(\frac{1}{2})} S_{-j}^{(\frac{1}{2})} - S_{-i}^{(\frac{1}{2})} S_{+j}^{(\frac{1}{2})}) \\
&\approx \sum_i \delta_i S_{zi}^{(\frac{1}{2})} + \sum_{i<j} 2V_{ij} S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})}.
\end{aligned} \tag{S13}$$

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

$$\begin{aligned}
D(\tau) &= R_x(\pi) \exp[-i\tau H_{\text{DD,Ising}}] R_x(\pi) \exp[-i\tau H_{\text{DD,Ising}}] \\
&= \exp \left[ -i\tau \sum_{i<j} 2V_{ij} S_{zi}^{(\frac{1}{2})} S_{zj}^{(\frac{1}{2})} \right].
\end{aligned} \tag{S14}$$

Eq. (S14) is also valid when the local disorder  $\delta_i$  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. S1 (b).

From Fig. S1, 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.

### OPTIMIZATION RESULTS BY USING $P_z^{\text{tot}}$ , $P_z^\pi$ AS MEASUREMENT BASES

The optimization results shown in Fig.2 in the main text are obtain 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

$$P_z^{\text{tot}} \equiv |J = N/2, J_z\rangle \langle J = N/2, J_z| \tag{S15}$$

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  $-N/2$ .  $P_z^{\text{tot}}$  has  $N+1$  outcomes, so it scales linear with the system size.

The optimization results by using the CFI on  $P_z^{\text{tot}}$  as cost function are shown in Fig. S2 (a)(b). Surprisingly, compared to the results by using  $P_z$ , the optimization results from using  $P_z^{\text{tot}}$  are improved by about a factor of  $1.5 \sim 2$  for the 3D-random spin configuration. Since all the information one can extract from  $P_z^{\text{tot}}$  are contained in  $P_z$ , we attribute this improvement to the simpler parameter space structure that  $P_z^{\text{tot}}$  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,

$$P_z^\pi \equiv \prod_i^N \sigma_{zi}, \tag{S16}$$

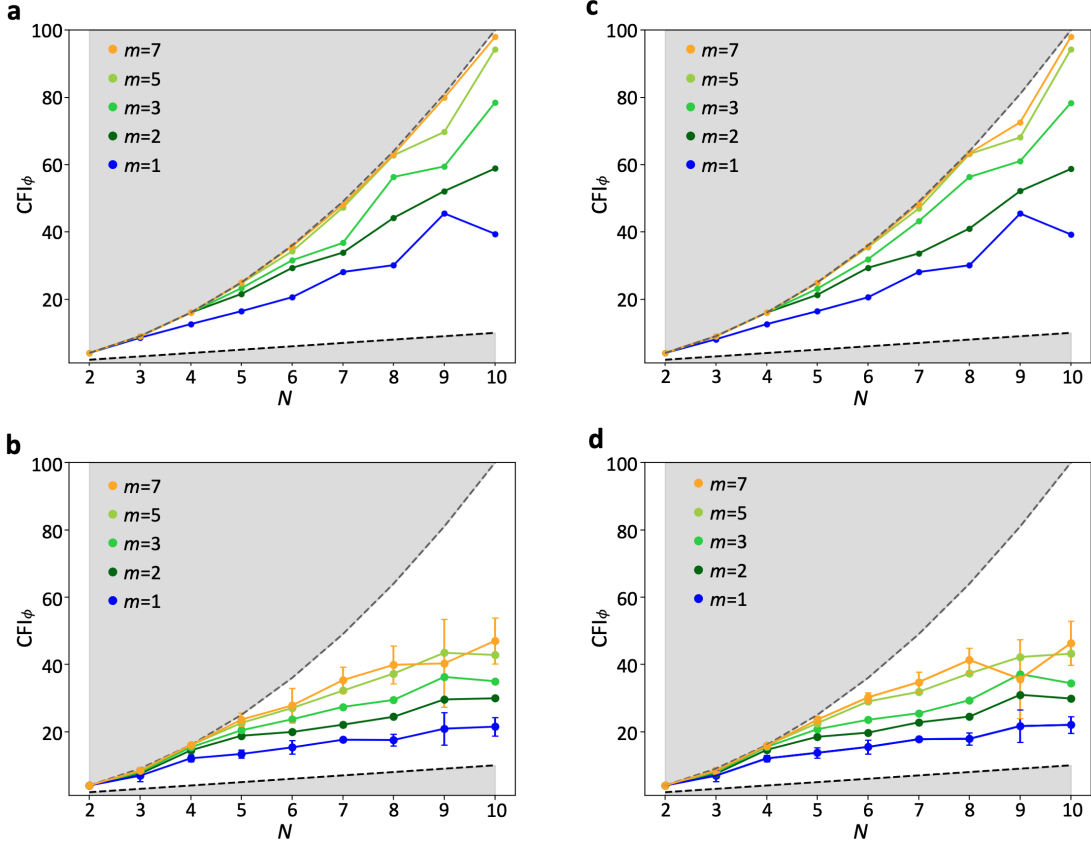


FIG. S2. CFI optimization data for (a) 2D square lattice using observable  $P_z^{\text{tot}}$ , (b) 3D-random configuration using observable  $P_z^{\text{tot}}$  (averaged over 5 cases), (c) 2D square lattice using observable  $P_z^{\pi}$ , (d) 3D-random configuration using observable  $P_z^{\pi}$  (averaged over 5 cases).

provides a constant (2) dimensional outcome space for experimental readout. Improvements are also observed in 2D square lattice and 3D-random spin configurations (Fig. S2(c)(d)).

## SUPPLEMENTARY DATA

### 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. S3(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 larger, 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. S3(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 [55-57], 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. S3(b)-(d) show the CFI optimization result for 1D chain, 2D square lattice and 2D symmetric cycle spin configurations. The results of regular patterns are better than those of 3D-random pattern.

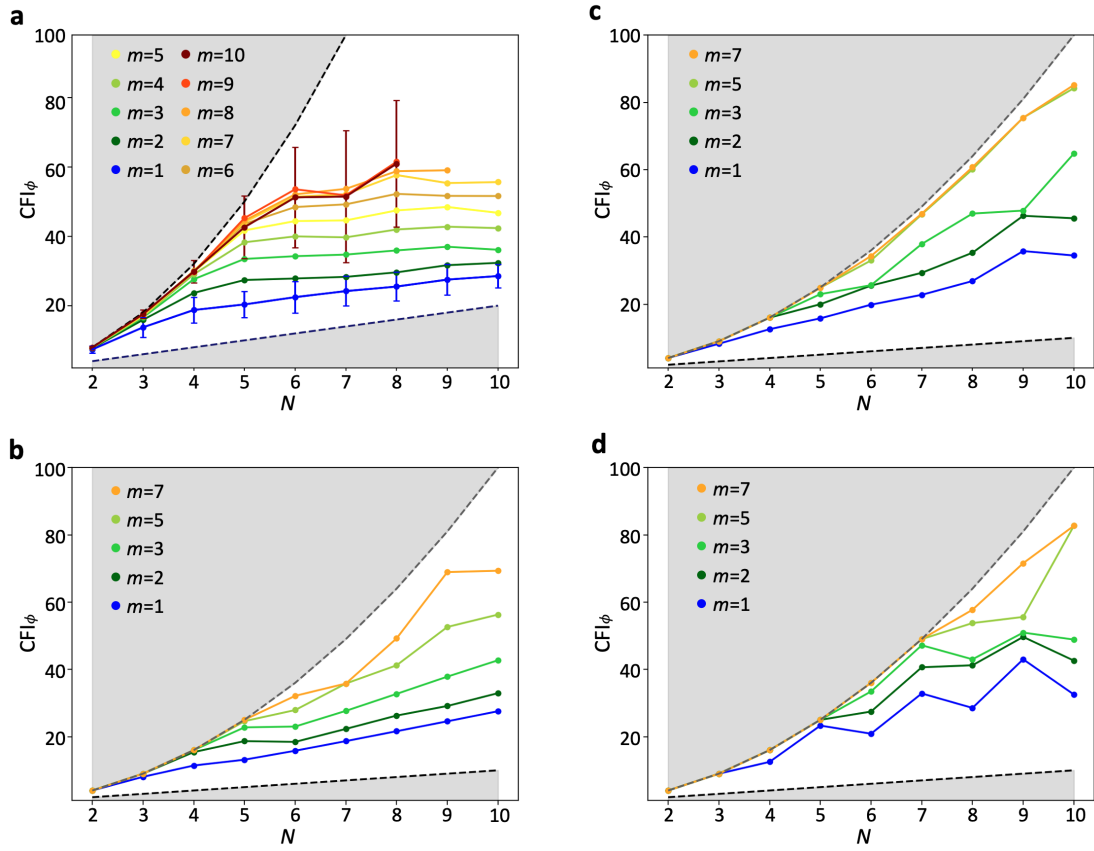


FIG. S3. The complete CFI data for (a) 3D random, (b) 1D chain, (c) 2D square lattice, and (d) 2D circle.

#### Required layers to reach given CFI for 2D square lattice

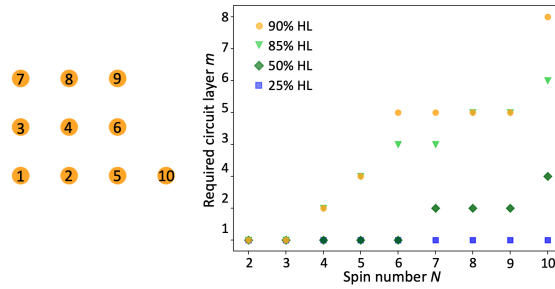


FIG. S4. Left: Schematic of a 2D square lattice pattern. The numbers label the order in which spins are added to form a lattice of size N. Right: Number of layers required to achieve a CFI within a given percentage of the HL.

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. S4 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.



### 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 [39]) between the original state and the new state, which is generated by using the cutoff Hamiltonian and optimized parameters. A cutoff interaction strength  $f_{\text{cutoff}}$  is chosen, and all the pairwise potential  $V_{ij}$  smaller than  $f_{\text{cutoff}}$  are set to zero in the cutoff Hamiltonian. Fig. S5 shows the relation between the state fidelity  $F$  versus  $f_{\text{cutoff}}$ . A state fidelity value less than 1 is observed when  $f_{\text{cutoff}}$  is set to be equal to the averaged nearest-neighbor interaction strength  $f_{\text{dd}}$ . This result reflects higher order interactions in the spins ensemble are utilized for the entangled state generation.

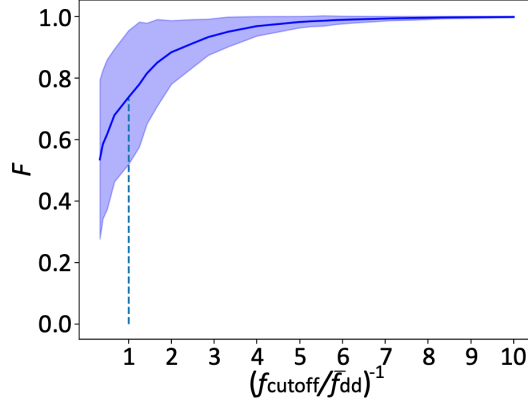


FIG. S5. Average state fidelity vs. different cutoff strength in  $H_{\text{dd}}$ . The shaded area corresponds to the error range. Data obtained from 3D-random  $N = 10, m = 5$ , 50-cases optimization results.

### Non-Markovian noise sensing performance

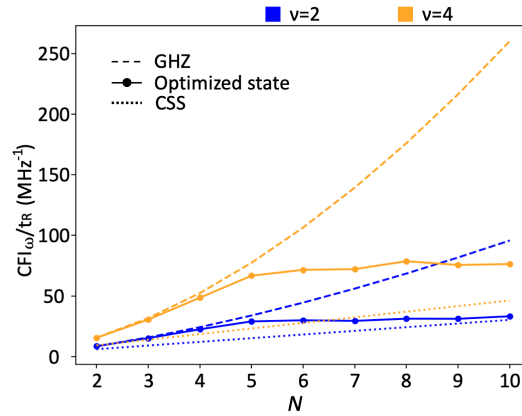


FIG. S6. Average Ramsey protocol's results of the generated entangled states in 3D random configurations when considering non-Markovian noise in the signal accumulation step. Blue and orange correspond to two different noise models ( $\nu = 2$  and 4).

### 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.

|          |             |       |       |       |       |       |      |            |      |
|----------|-------------|-------|-------|-------|-------|-------|------|------------|------|
| <b>a</b> | 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 |
|          | 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 | $10^{-33}$ |      |
|          | Wigner Plot |       |       |       |       |       |      |            |      |
|          | RF          | 1.00  | 0.98  | 0.96  | 0.94  | 0.92  |      |            |      |
| <b>b</b> | CFI         | 85.18 | 30.18 | 18.18 | 14.34 | 11.66 |      |            |      |
|          | CSS         | 10.00 | 9.216 | 8.464 | 7.744 | 7.056 |      |            |      |
|          | 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   |       |      |            |      |
|          | 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   |       |      |            |      |

FIG. S7. Optimized states' Wigner distributions when finite RF is assumed in the CFI calculation.

### Relative experimental parameter table (full)

TABLE S1. Experimental platforms' data

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

<sup>a</sup> T.H.Taminiau, NComm 2018, <sup>b</sup> H.Zhou, PRX 2020, B.J.Shields, <sup>c</sup> M.D.Lukin, PRL 2015, <sup>d</sup> L.Childress Science 2006

<sup>e</sup> T.H.Taminiau, NComm 2021, <sup>f</sup> N.Yao, Nature 2021

<sup>g</sup> P.Bertet, Science advances 2021, <sup>h</sup> A.Reiserer, PRL 2021, <sup>i</sup> J.Thompson, Science 2020, <sup>j</sup> J.Thompson, NComm 2020

<sup>k</sup> M.R.Tarbutt PRL 2020, <sup>l</sup> B.Yan, J.Ye, Nature 2013, <sup>m</sup> 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 [S1](#) is a lower bound and  $T_2^{(\text{best})}$  is a more precise estimation for the spin coherence time.

## SUPPLEMENTARY DERIVATIONS

### 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  $\theta$  [58, 59]. Let  $Z$  be a random variable and  $P_z(\theta) \equiv P(z|\theta)$  be its probability distribution which depends on  $\theta$ . Let  $\Theta$  be an unbiased estimator of  $\theta$ , i.e.

$$\theta = \langle \Theta \rangle = \sum_z \Theta \cdot P_z(\theta). \quad (\text{S17})$$

From Eq. (S17) and the fact that the sum of probabilities of all outcomes is 1,

$$1 = \frac{\partial \langle \Theta \rangle}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_z \Theta P_z(\theta), \quad (\text{S18})$$

$$0 = \frac{\partial}{\partial \theta} \sum_z P_z(\theta). \quad (\text{S19})$$

Subtracting Eq. (S19) multiplied by  $\theta$  from Eq. (S18), we get

$$\begin{aligned} 1 &= \sum_z (\Theta - \theta) \frac{\partial}{\partial \theta} P_z(\theta) \\ &= \sum_z P_z(\theta) (\Theta - \theta) \frac{1}{P_z(\theta)} \frac{\partial}{\partial \theta} P_z(\theta) \\ &= \langle (\Theta - \theta) \frac{1}{P_z(\theta)} \frac{\partial}{\partial \theta} P_z(\theta) \rangle \\ &= \langle (\Theta - \theta) \frac{\partial}{\partial \theta} \log P_z(\theta) \rangle. \end{aligned} \quad (\text{S20})$$

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

$$\langle (\Theta - \theta)^2 \rangle \left\langle \left( \frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2 \right\rangle \geq 1, \quad (\text{S21})$$

where

$$\begin{aligned} \langle (\Theta - \theta)^2 \rangle &= \langle \Theta^2 \rangle - (2\theta \langle \Theta \rangle - \langle \theta^2 \rangle) \\ &= \langle \Theta^2 \rangle - (2\theta^2 - \theta^2) \\ &= \langle \Theta^2 \rangle - \langle \Theta \rangle^2 \\ &= \Delta \Theta^2 \end{aligned} \quad (\text{S22})$$

is the variance of  $\Theta$ . Defining

$$\text{CFI} = \sum_z P_z(\theta) \left( \frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2, \quad (\text{S23})$$

we have

$$\Delta \Theta^2 \geq \frac{1}{\text{CFI}}. \quad (\text{S24})$$

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

$$\Delta \Theta^2 \geq \frac{1}{M \cdot \text{CFI}}. \quad (\text{S25})$$

In our variational circuit, we use CFI with respect to an infinitesimal angle  $\phi$  as the cost function to generate entangled states. In our program, we use a method similar to parameter shift to calculate the  $\text{CFI}_\phi$  of our optimized states [19, 60, 61]. 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  $\phi$  is a small angle;
3.  $\psi$  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

$$\begin{aligned}
\left. \frac{\partial}{\partial \phi} P_z(\phi) \right|_{\phi \rightarrow 0} &= \left. \frac{\partial}{\partial \phi} |\langle z | \mathcal{U}(\phi) | \psi \rangle|^2 \right|_{\phi \rightarrow 0} \\
&= \left. \frac{\partial}{\partial \phi} \langle \psi | \mathcal{U}^\dagger(\phi) | z \rangle \langle z | \mathcal{U}(\phi) | \psi \rangle \right|_{\phi \rightarrow 0} \\
&= \left. \langle \psi | \mathcal{U}^\dagger(\phi) i J_y | z \rangle \langle z | \mathcal{U}(\phi) | \psi \rangle \right|_{\phi \rightarrow 0} + \left. \langle \psi | \mathcal{U}^\dagger(\phi) | z \rangle \langle z | \mathcal{U}(\phi) (-i) J_y | \psi \rangle \right|_{\phi \rightarrow 0} \\
&= i \langle \psi | \left( J_y | z \rangle \langle z | - | z \rangle \langle z | J_y \right) | \psi \rangle.
\end{aligned} \tag{S26}$$

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  $\phi$  along  $y$ -axis, or firstly accumulates a signal along  $z$ -axis and then rotated by  $R_x(-\pi/2)$  pulse [59]. 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  $\omega$ . 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

$$\text{CFI}_\omega = \text{CFI}_\phi \cdot t_R^2. \tag{S27}$$

### Relation between $\text{CFI}_\omega$ 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

$$\rho(t) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{S28}$$

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

$$\rho(t) = \frac{1}{2} \begin{pmatrix} 1 & e^{-i\omega t} e^{-2\gamma t} \\ e^{i\omega t} e^{-2\gamma t} & 1 \end{pmatrix} \tag{S29}$$

where  $\gamma$  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

$$R_x\left(\frac{\pi}{2}\right) \rho(t) R_x^\dagger\left(\frac{\pi}{2}\right) \tag{S30}$$

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

$$P_0 = \frac{1}{2} + \frac{1}{2} e^{-2\gamma t} \sin \omega t. \tag{S31}$$

The CFI with respect to  $\omega$  is

$$\text{CFI}_\omega = \frac{1}{P_0} \left( \frac{\partial P_0}{\partial \omega} \right)^2 + \frac{1}{P_1} \left( \frac{\partial P_1}{\partial \omega} \right)^2 = \frac{t^2 \cos^2 \omega t}{e^{4\gamma t} - \sin^2 \omega t}. \quad (\text{S32})$$

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

$$\text{SNR}^2 = \frac{Mt^2 \cos^2 \omega t \delta \omega^2}{e^{4\gamma t} - \sin^2 \omega t}. \quad (\text{S33})$$

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

$$\text{CFI}_\omega \cdot \frac{T_{\text{tot}}}{t_R} \cdot \delta \omega^2 = \text{SNR}^2. \quad (\text{S34})$$

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

$$\delta \omega = \frac{1}{\sqrt{M \cdot \text{CFI}_\omega}}, \quad (\text{S35})$$

leading to the saturated Cramér-Rao bound.

### 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  $\theta$ . 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. Below we summarize the proof given in [62].

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  $\theta$  is an unknown parameter and  $\Theta$  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  $\theta$  (the signal we want to measure) from  $\mathbf{x}$ , i.e., find  $\theta$  that is most likely to produce the outcome  $\mathbf{x}$ . Thus, the normalized log-likelihood function is defined as

$$L_M(\theta) = \frac{1}{M} \log P(\mathbf{X}|\theta) = \frac{1}{M} \log \prod_{i=1}^M P(X_i|\theta) = \frac{1}{M} \sum_{i=1}^M \log P(X_i|\theta). \quad (\text{S36})$$

A MLE maximizes the log-likelihood function

$$\Theta_{\text{MLE}} = \underset{\theta \in \Theta}{\text{argmax}} L_M(\theta). \quad (\text{S37})$$

In the following, we first show that

1.  $\Theta_{\text{MLE}}$  converges to the true parameter  $\theta_0$ ;
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

$$L(\theta) = \langle \log P(\mathbf{X}|\theta) \rangle_{\theta_0} \quad (\text{S38})$$

which denotes the expected log-likelihood function with respect to  $\theta_0$ , then by the Weak Law of Large Numbers (WLLN), the average outcomes from a large number of trials should approach the expected value:

$$\forall \theta, L_M(\theta) \xrightarrow{M \rightarrow \infty} L(\theta). \quad (\text{S39})$$



In fact,  $\theta_0$  maximizes  $L(\theta)$ :

$$\begin{aligned}
\forall \theta, L(\theta) - L(\theta_0) &= \langle \log P(\mathbf{X}|\theta) \rangle_{\theta_0} - \langle \log P(\mathbf{X}|\theta_0) \rangle_{\theta_0} \\
&= \left\langle \log \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} \\
&\leq \left\langle \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} - 1 \right\rangle_{\theta_0} \\
&= \sum_{\mathbf{x} \in \mathbf{X}} \left( \frac{P(\mathbf{x}|\theta)}{P(\mathbf{x}|\theta_0)} - 1 \right) P(\mathbf{x}|\theta_0) \\
&= 1 - 1 = 0.
\end{aligned} \tag{S40}$$

Moreover, we show that  $\theta_0$  is the unique maximizer. Jensen's inequality states that for a strictly convex function  $f$  and a random variable  $Y$ ,

$$\langle f(Y) \rangle > f(\langle Y \rangle). \tag{S41}$$

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

$$\left\langle -\log \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} > -\log \left\langle \frac{P(\mathbf{X}|\theta)}{P(\mathbf{X}|\theta_0)} \right\rangle_{\theta_0} = 0, \tag{S42}$$

or

$$L(\theta_0) > L(\theta). \tag{S43}$$

Therefore, since

1.  $\Theta_{\text{MLE}}$  maximizes  $L_M(\theta)$ ,
2.  $\theta_0$  maximizes  $L(\theta)$ , and
3.  $L_M(\theta) \xrightarrow{M \rightarrow \infty} L(\theta)$ ,

$\Theta_{\text{MLE}}$  converges to  $\theta_0$ .

Now we use this property to prove that the distribution of  $\Theta_{\text{MLE}}$  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  $[\Theta_{\text{MLE}}, \theta_0]$ :

$$\begin{aligned}
0 &= L'_M(\Theta_{\text{MLE}}) = L'_M(\theta_0) + L''_M(\theta_1)(\theta_0 - \Theta_{\text{MLE}}) \\
\implies \theta_0 - \Theta_{\text{MLE}} &= -\frac{L'_M(\theta_0)}{L''_M(\theta_1)} \\
\implies \sqrt{M}(\theta_0 - \Theta_{\text{MLE}}) &= -\sqrt{M} \frac{L'_M(\theta_0)}{L''_M(\theta_1)}
\end{aligned} \tag{S44}$$

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

$$\begin{aligned}
L'_M(\theta_0) &= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' \\
&= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' - L'(\theta_0) \\
&= \frac{1}{M} \sum_{i=1}^M (\log P(X_i|\theta_0))' - \langle (\log P(\mathbf{X}|\theta_0))' \rangle_{\theta_0} \\
&= \frac{1}{M} \left( \sum_{i=1}^M \log P(X_i|\theta_0) \right)' - \langle (\log P(\mathbf{X}|\theta_0))' \rangle_{\theta_0}
\end{aligned} \tag{S45}$$

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

$$\mathcal{N}\left(0, \text{Var}_{\theta_0}(\log P(X_i|\theta_0))'\right) \quad (\text{S46})$$

where the variance

$$\begin{aligned} \text{Var}_{\theta_0}(\log P(X_i|\theta_0))' &= \langle [(\log P(X_i|\theta_0))']^2 \rangle_{\theta_0} - \langle (\log P(X_i|\theta_0))' \rangle_{\theta_0}^2 \\ &= \sum_{x \in X_1} P(x|\theta_0) \left( \frac{P'(x|\theta_0)}{P(x|\theta_0)} \right)^2 - (L'(\theta_0))^2 \\ &= \text{CFI}_{\theta_0} \end{aligned} \quad (\text{S47})$$

by the definition of CFI and that  $\theta_0$  maximizes  $L(\theta)$ . By the consistency property,  $\Theta_{\text{MLE}}$  converges to  $\theta_0$ , and thus  $\theta_1$  converges to  $\theta_0$ . The denominator

$$L''_M(\theta_1) \rightarrow L''_M(\theta_0) = \frac{1}{M} \sum_{i=1}^M [\log P(X_i|\theta_0)]'' \rightarrow \langle [\log P(X_1|\theta_0)]'' \rangle_{\theta_0} \quad (\text{S48})$$

by the WLLN. We further show that Eq. (S48) is in fact the additive inverse of CFI:

$$\begin{aligned} \langle [\log P(X_1|\theta_0)]'' \rangle_{\theta_0} &= \left\langle \frac{\partial^2}{\partial \theta^2} \log P(X_1|\theta_0) \right\rangle_{\theta_0} \\ &= \sum_{x \in X_1} [\log P(x|\theta_0)]'' P(x|\theta_0) \\ &= \sum_{x \in X_1} \left( \frac{P''(x|\theta_0)}{P(x|\theta_0)} - \left( \frac{P'(x|\theta_0)}{P(x|\theta_0)} \right)^2 \right) P(x|\theta_0) \\ &= \sum_{x \in X_1} P''(x|\theta_0) - \sum_{x \in X_1} \frac{(P'(x|\theta_0))^2}{P(x|\theta_0)} \\ &= 0 - \text{CFI}_{\theta_0} = -\text{CFI}_{\theta_0}. \end{aligned} \quad (\text{S49})$$

Finally, Eq. (S44) becomes

$$\begin{aligned} \sqrt{M}(\theta_0 - \Theta_{\text{MLE}}) &\xrightarrow{p} \mathcal{N}\left(0, \frac{\text{CFI}_{\theta_0}}{\text{CFI}_{\theta_0}^2}\right) = \mathcal{N}\left(0, \frac{1}{\text{CFI}_{\theta_0}}\right) \\ \implies \Theta_{\text{MLE}} &\xrightarrow{p} \mathcal{N}\left(\theta_0, \frac{1}{M \cdot \text{CFI}_{\theta_0}}\right) \end{aligned} \quad (\text{S50})$$

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

### 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 [51]. A brief summary of the derivation is given below.

1. 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).
2. 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  $\mathcal{LL}(\mathbb{C}^d)$ . Then the action of  $\Lambda \in \mathcal{LL}(\mathbb{C}^d)$  on  $\tau \in \mathcal{L}(\mathbb{C}^d)$  can be expressed as

$$\Lambda[\tau] = \sum_{ij=1}^{d^2} \langle l_i, \Lambda[l_j] \rangle \langle l_j, \tau \rangle l_i. \quad (\text{S51})$$

Thus,  $\Lambda$  has a unique correspondence with the matrix  $\Lambda$  with entries

$$\Lambda_{ij} \equiv \langle l_i, \Lambda[l_j] \rangle. \quad (\text{S52})$$

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

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{m} & \mathbf{M} \end{pmatrix}, \quad (\text{S53})$$

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  $\rho$  on  $\mathbb{C}^2$  can be written as

$$\rho = \frac{1}{2}(\mathbf{I} + \mathbf{v} \cdot \boldsymbol{\sigma}) \quad (\text{S54})$$

where  $\mathbf{v}$  is a three-dimensional real vector and  $\boldsymbol{\sigma}$  is the vector of Pauli matrices. Then a map  $\Lambda$  whose matrix representation is given by Eq. (S53) acting on  $\rho$  gives

$$\Lambda[\rho] = \frac{1}{2}(\mathbf{I} + (\mathbf{m} + \mathbf{M}\mathbf{v}) \cdot \boldsymbol{\sigma}). \quad (\text{S55})$$

5. The noisy evolution of a state  $\rho$  is described by

$$\rho(t) = \Lambda(t)[\rho(0)]. \quad (\text{S56})$$

The time local master equation satisfies

$$\frac{d}{dt}\rho(t) = \Xi(t)[\rho(t)]. \quad (\text{S57})$$

So

$$\Xi(t) = \frac{d\Lambda(t)}{dt} \circ \Lambda(t)^{-1} \quad (\text{S58})$$

with the corresponding matrix representation

$$\Xi(t) = \frac{d\Lambda(t)}{dt} \Lambda(t)^{-1}. \quad (\text{S59})$$

6. Consider the evolution of one qubit described by  $\Lambda(t) = \mathcal{U}(t) \circ \Gamma(t)$ .  $\mathcal{U}(t)$  is defined as

$$\mathcal{U}(t)[\rho(0)] \equiv U(t)\rho(0)U^\dagger(t) \quad (\text{S60})$$

where  $U(t) = e^{-i\frac{\omega t}{2}\sigma_z}$  represents the signal accumulation. By Eq. (S51) and Eq. (S60), the matrix representation of  $\mathcal{U}(t)$  is

$$\mathcal{U}(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \omega t & -\sin \omega t & 0 \\ 0 & \sin \omega t & \cos \omega t & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (\text{S61})$$

$\Gamma(t)$  represents the noise which is trace- and hermicity- preserving, i.e., has the form in Eq. (S53).

7. Solving the commutation relation that gives phase covariant qubit map [51, 63]

$$[\mathcal{U}(t), \Gamma(t)] = 0 \iff [\mathcal{U}(t), \Gamma(t)] = 0, \quad (\text{S62})$$

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

$$\Lambda(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \eta_\perp(t) \cos \omega t & -\eta_\perp(t) \sin \omega t & 0 \\ 0 & \eta_\perp(t) \sin \omega t & \eta_\perp(t) \cos \omega t & 0 \\ \kappa(t) & 0 & 0 & \eta_\parallel(t) \end{pmatrix}, \quad (\text{S63})$$

where  $\mathbf{m} = (0, 0, \kappa(t))^T$  describes a translation along the  $z$ -axis, and  $\mathbf{M} = \begin{pmatrix} \eta_\perp(t) \cos \omega t & -\eta_\perp(t) \sin \omega t & 0 \\ \eta_\perp(t) \sin \omega t & \eta_\perp(t) \cos \omega t & 0 \\ 0 & 0 & \eta_\parallel(t) \end{pmatrix}$  describes a rotation along the  $z$ -axis and a contraction characterized by  $\eta_\perp$  and  $\eta_\parallel$ .

8. By Eq. (S59), we obtain the time-local master equation for a single qubit:

$$\begin{aligned}\Xi(t)[\rho(t)] = & -\frac{i}{2}\omega[\sigma_z, \rho(t)] \\ & + \gamma_+(t)(\sigma_+\rho(t)\sigma_- - \frac{1}{2}\{\sigma_-\sigma_+, \rho(t)\}) \\ & + \gamma_-(t)(\sigma_-\rho(t)\sigma_+ - \frac{1}{2}\{\sigma_+\sigma_-, \rho(t)\}) \\ & + \gamma_z(t)(\sigma_z\rho(t)\sigma_z - \rho(t)),\end{aligned}\tag{S64}$$

where

$$\gamma_+(t) = \frac{1}{2} \left( \kappa'(t) - \frac{\eta_{\parallel}'(t)}{\eta_{\parallel}(t)}(\kappa(t) + 1) \right),\tag{S65}$$

$$\gamma_-(t) = -\frac{1}{2} \left( \kappa'(t) + \frac{\eta_{\parallel}'(t)}{\eta_{\parallel}(t)}(1 - \kappa(t)) \right),\tag{S66}$$

$$\gamma_z(t) = \frac{1}{4} \left( \frac{\eta_{\parallel}'(t)}{\eta_{\parallel}(t)} - 2\frac{\eta_{\perp}'(t)}{\eta_{\perp}(t)} \right).\tag{S67}$$

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

$$\eta_{\perp}(t) = e^{-(\frac{t}{T_2})^{\nu}},\tag{S68}$$

where  $\nu$  is the stretch character which equals 1 for Markovian noise. Then

$$\gamma_z(t) = \frac{\nu}{2} \frac{t^{\nu-1}}{T_2^{\nu}}.\tag{S69}$$

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

$$\rho = \sum_{i,j} \rho_{ij} |i\rangle \langle j| \mapsto |\rho\rangle = \sum_{i,j} \rho_{ij} |j\rangle \otimes |i\rangle.\tag{S70}$$

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

$$\Xi(t) = -\frac{i}{2}\omega(I \otimes \sigma_z - \sigma_z \otimes I) + \gamma_z(t)(\sigma_z \otimes \sigma_z - I \otimes I).\tag{S71}$$

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 [64].

### Performance of metrological states in a non-Markovian environment

To calculate the derivative of probability with respect to  $\omega$  in the calculation for  $\text{CFI}_{\omega}$ , we use a method similar to parameter shift that utilizes the property that the signal accumulation operator ( $\mathcal{U}(\omega) = e^{-i\omega t J_z}$ ) 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.  $\rho$  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 [59]);

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

$$\begin{aligned}
 \frac{\partial}{\partial \omega} P(z|\omega) \Big|_{\omega \rightarrow 0} &= \frac{\partial}{\partial \omega} \text{Tr} [|z\rangle \langle z| \mathcal{U}(\omega) \rho \mathcal{U}^\dagger(\omega)] \Big|_{\omega \rightarrow 0} \\
 &= \frac{\partial}{\partial \omega} \text{Tr} [\mathcal{U}^\dagger(\omega) |z\rangle \langle z| \mathcal{U}(\omega) \rho] \Big|_{\omega \rightarrow 0} \\
 &= \text{Tr} \left[ \frac{\partial}{\partial \omega} \mathcal{U}^\dagger(\omega) |z\rangle \langle z| \mathcal{U}(\omega) \rho \right] \Big|_{\omega \rightarrow 0} + \text{Tr} [\mathcal{U}^\dagger(\omega) |z\rangle \langle z| \frac{\partial}{\partial \omega} \mathcal{U}(\omega) \rho] \Big|_{\omega \rightarrow 0} \\
 &= it \text{Tr} [(J_y |z\rangle \langle z| - |z\rangle \langle z| J_y) \rho].
 \end{aligned} \tag{S72}$$

From Eq. (S34), since  $T_{\text{tot}}$  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.

### Time Overhead

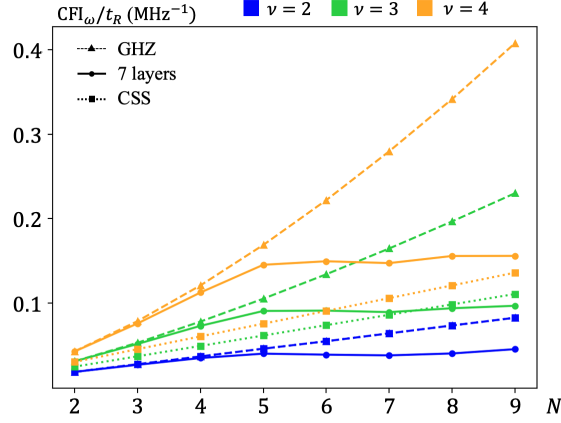


FIG. S8. 50 cases average sensing performance of the optimized states when using 7-layer circuit on 3D random spin configuration.

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

$$\text{SNR}^2 = \frac{T_{\text{tot}} t_R^2 \cos^2(\omega t_R) \delta\omega^2}{(t_R + t_{\text{oh}}) \left( e^{2\left(\frac{t_R}{T_2}\right)^\nu} - \sin^2(\omega t_R) \right)}. \tag{S73}$$

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

$$\text{SNR}^2 \propto \frac{t_R^2}{e^{2\left(\frac{t_R}{T_2}\right)^\nu}}. \tag{S74}$$

Taking the derivative of Eq. (S74) with respect to  $t_R$  gives us the best  $t_R$  if the time overhead is significantly larger:

$$t_R = \frac{T_2}{\nu^{\frac{1}{\nu}}}. \tag{S75}$$



Similarly, the same calculations for a GHZ state where the decay term in Eq. (S73) becomes  $e^{2n\left(\frac{t_R}{T_2}\right)^\nu}$  show that the best Ramsey sensing time is

$$t_R = \frac{T_2}{(n\nu)^{\frac{2}{\nu}}}. \quad (\text{S76})$$

Plugging Eq. (S75) and Eq. (S76) into Eq. (S74), we find that the ratio of the  $\text{SNR}^2$  of a GHZ state to that of an uncorrelated spin state is  $n^{1-\frac{2}{\nu}}$ . Thus, only when

$$\nu > 2 \quad (\text{S77})$$

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. S8 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.

### 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 [17] 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 [65]) and CFI. 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. S7(b)). In this case, the squeezing parameter is

$$\xi^2 = N \frac{(\Delta J_y)^2}{|\langle J_x \rangle|^2}, \quad (\text{S78})$$

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,

$$(\Delta J_y)^2 (\Delta J_z)^2 \geq \frac{1}{4} |\langle J_x \rangle|^2. \quad (\text{S79})$$

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

$$4(\Delta J_z)^2 \geq \frac{|\langle J_x \rangle|^2}{(\Delta J_y)^2} = N/\xi^2. \quad (\text{S80})$$

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 [58,66,67]:

$$\text{QFI} = 4(\Delta J_z)^2. \quad (\text{S81})$$

Combining Eq. (S80) and Eq. (S81), we obtain the relation between CFI and squeezing parameter of a SSS:

$$\text{CFI} \leq \text{QFI} \geq N/\xi^2. \quad (\text{S82})$$

The first inequality in Eq. (S82) 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 [12]). The second inequality originates from the uncertainty principle (Eq. (S79)). Since the optimal SSS saturate the Heisenberg uncertainty relation [68] and the SSS generated by the adiabatic method [17] belongs to these states, we obtain the relation between the squeezing parameter and CFI

$$\text{CFI} = N/\xi^2. \quad (\text{S83})$$

Based on the data shown in Fig.3 from ref. [17], 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  $\bar{f}_{\text{dd}}$ , 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.

## 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

$$H(t) = H_0 + \sum_{k=1}^K u_k(t)H_k, \quad (\text{S84})$$

governed by the Schrödinger equation,  $i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle$ , have been studied extensively [69,31].  $H_0$  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) [70,71].

### Controllability Test

The way we investigate the controllability of a generic system (S84) 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 [69,31].

A quantum system of the form (S84) is completely controllable if either  $\mathcal{L}_0 \cong u(\mathcal{N})$  or  $\mathcal{L}_0 \cong su(\mathcal{N})$  [69], 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. [69] and ref. [31] 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 (S84) by defining the free evolution Hamiltonian to be the dipolar interaction  $H_{\text{dd}}$  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  $\theta_y(t)$ :

$$H(t) = H_{\text{dd}} + \theta_x(t)J_x + \theta_y(t)J_y. \quad (\text{S85})$$

Ref. [31,72] 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. [73] 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. [74] extend their result for symmetric Ising spin

| <b>Algorithm.</b> Generating $\mathcal{L}_0$ and finding $\dim \mathcal{L}_0$ .                              |
|--|
| <i>Input:</i> 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 $r = N^2$ then $O \equiv B$ else $O \equiv \{\}$   |
| 4. If $r = N^2$ or $ B  = 0$ then terminate  |
| 5. $C \equiv [O, B] \cup [B, B]$ , where<br>$[S_1, S_2] \equiv \{[s_1, s_2] \mid s_1 \in S_1, s_2 \in S_2\}$ |
| 6. $O = O \cup B$  |
| 7. $B =$ maximal linearly independent extension of $O$ with<br>elements from $C$                             |
| 8. $r = r +  B $ ; Go to 4   |
| <i>Output:</i> basis $O$ of $\mathcal{L}_0$ and $\dim \mathcal{L}_0 = r$                                     |

TABLE S2. Implementation of [31]’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    |
| $H_{\text{dd}}$                                | 9       | 39      | 225     |         |
| Symmetric Ising: $\binom{N+3}{N} - 1$          | 9       | 19      | 34      | 55      |

TABLE S3. 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 [31]’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.

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 shown to be  $\binom{N+3}{N} - 1$ . This is relevant to our system because [75] show that if we replace the Ising interaction with a more general two body interaction—which includes  $H_{\text{dd}}$ —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 [69]. 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,

$$\mathcal{S}(\theta) = e^{-i\frac{\pi}{2}J_y} \prod_{i=1}^m e^{-i\tau_i H_{\text{dd}}} e^{-i\vartheta_i J_x} e^{i\frac{\pi}{2}J_y} e^{-i\tau'_i H_{\text{dd}}} e^{-i\frac{\pi}{2}J_y}, \quad (\text{S86})$$

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

$$\mathcal{F}(|\text{GHZ}\rangle, \mathcal{S}(\theta) |0\rangle^{\otimes N}) = \left| \langle \text{GHZ} | \mathcal{S}(\theta) |0\rangle^{\otimes N} \right|^2, \quad (\text{S87})$$

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. [73, 76] 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.