Error-mitigated quantum metrology

Kaoru Yamamoto,¹ Suguru Endo,¹,² Hideaki Hakoshima,³,⁴ Yuichiro Matsuzaki,³ and Yuuki Tokunaga¹

¹NTT Computer and Data Science Laboratories, NTT Corporation, Musashino 180-8585, Japan
²JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan
³Research Center for Emerging Computing Technologies, National institute of Advanced Industrial Science and Technology (AIST), Central2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
⁴Center for Quantum Information and Quantum Biology, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

Quantum metrology with entangled resources aims to achieve sensitivity scaling beyond the standard quantum limit by harnessing quantum effects even in the presence of environmental noise. So far, scaling has been mainly discussed from the viewpoint of reducing statistical errors under the assumption of perfect knowledge of a noise model. However, we cannot always obtain complete information about a noise model due to coherence time fluctuations, which are frequently observed in experiments. Such unknown fluctuating noise leads to systematic errors and nullifies the quantum advantages. Here, we propose an error-mitigated quantum metrology that can filter out unknown fluctuating noise with the aid of purification-based quantum error mitigation. We demonstrate that our protocol mitigates systematic errors and recovers superclassical scaling in a practical situation with time-inhomogeneous bias-inducing noise.

Introduction.— Quantum metrology with entangled resources has been shown to reach the Heisenberg limit of sensitivity with respect to the number of qubits [1–6]. It may provide significant improvements for versatile applications such as atomic-frequency [7, 8] and electron-spin-resonance measurements [9, 10], magnetometry [12, 19], thermometry [20–22], and electrometer [23, 24].

The Heisenberg limit is susceptible to decoherence; for example, under the effect of Markovian dephasing, the sensitivity of entangled states scales to the standard quantum limit (SQL), as do separable states [7]. Several theoretical studies predict that scaling beyond the standard quantum limit is possible: they include the Zeno-scaling under time-inhomogeneous noise model [25–27], quantum scaling by quantum teleportation [27, 28], utilizing the collective effect of open quantum systems [28–30], and applying quantum error correction [31, 32]. So far, however, scaling has been mainly discussed from the viewpoint of statistical errors under the assumption that the noise model can be fully characterized [Fig. 1(a)].

In experiments, we cannot always obtain complete information of a noise model typically due to coherence time fluctuations [11, 33]; accordingly, noise characterization becomes intractable, leading to systematic errors. Throughout the present letter, we describe such a noise as “fluctuating noise”. Systematic errors usually result from a difference between the actual situation and the theoretical model used by experimentalists to estimate the target parameter [Fig. 1(b)]. Intractable noise characterization leads to inaccurate theoretical models and induces systematic errors in estimations. In practice, systematic errors are fatal to quantum metrology because they cannot be reduced even if the number of samples is increased, thus seriously limiting any sensitivity improvement [9, 34]. Despite that systematic errors are typically present in experiments, there is as yet no general approach to dealing them, although some studies have tackled specific scenarios [45, 49].

In the present letter, we propose a quantum-metrology protocol incorporating quantum error mitigation (QEM) to mitigate systematic errors, thereby improving scaling of sensitivity even in the presence of unknown fluctuating noise [Fig. 1(c)]. While conventional QEM methods have been designed for suppressing systematic errors in the expectation values produced by near-term quantum algorithms [44, 51, 60], they are not suitable for suppressing the systematic errors coming from unknown fluctuating noise. For example, probabilistic error cancellation cancels the effect of noise by inverting the noise map on the basis of characterization of the noise [52, 53], while error-extrapolation assumes the precise control of the noise model [52, 54]; thus, unknown fluctuating noise seriously degrades the performance of QEM. To deal with this problem, our error-mitigated quantum metrology, which is inspired by purification-based QEM [50, 61], filters out fluctuating noise that differs from one experimental run to another. In numerical simulations, we used it to suppress bias-inducing Markovian and time-inhomogeneous noise and thereby restore the scaling of sensitivity with respect to the number of qubits. In particular, we observed superclassical scaling with our method.

Quantum metrology with systematic errors.— Here, we describe a general theory for systematic errors in a Ramsey-type measurement of quantum metrology [62–64]. In a typical quantum-metrology setup, we prepare an initial state, expose this state to the target fields characterized by a parameter ω, and obtain a state ρ. Then,
we perform a measurement on this state that can be described by a projection operator $P$ producing a binary outcome. The measurement outcome, $m_j \in \{-1, 1\}$, is obtained from the measurement probability,

$$p = \text{Tr}[P\rho] = x + y\omega, \quad (1)$$

where $x$ and $y$ are some scalars. Here, we have assumed that $\omega$ is small and have ignored the higher order terms of $\omega$. Repeating the measurement $N_{\text{samp}}$ times yields the average value, $S_N = \sum_{j=1}^{N_{\text{samp}}} m_j/N_{\text{samp}}$. To estimate the parameter $\omega$, we need to fit this experimental data with a theoretical model. If we have imperfect knowledge about $\rho$, the theoretical model that we estimate would be different from the true one. The probability based on such an inaccurate theoretical model can be expressed as

$$p_e = \text{Tr}[P\rho_e] = x_e + y_e\omega, \quad (2)$$

where $\rho_e$, $x_e$, and $y_e$ denote the estimated values of $\rho$, $x$, and $y$, respectively. We estimate $\omega$ with $S_N$ by reference to Eq. (2) as $\omega_e = (S_N - x_e)/y_e$, which can be different from $\omega$, thereby leading to systematic errors.

Systematic errors require us to consider the estimation uncertainty of the target quantity $O_{\text{mit}}$. The estimation uncertainty of $\omega$ is defined as $\delta^2\omega = \langle (\omega - \omega_e)^2 \rangle$, with the brackets denoting the ensemble average, and is calculated as

$$\delta^2\omega = \frac{1}{y_e^2} [\text{Var}[p] + (x - x_e)^2], \quad (3)$$

where $\text{Var}[p]$ is the variance of $p$, which is typically $\text{Var}[p] = p(1-p)/N_{\text{samp}}$, and we have neglected the term $(y - y_e)^2\omega^2 - 2(x - x_e)(y - y_e)\omega$ because $\omega$ is small. Most of the previous theoretical studies focused on the first term in Eq. (3) by assuming $\rho = \rho_e$, which comes from the statistical error, as it decreases with increasing $N_{\text{samp}}$. The second term in Eq. (3) comes from the systematic error $x - x_e$ induced by incorrect estimation of the probability $p \neq p_e$. Since $x - x_e$ remains even when $N_{\text{samp}}$ increases, it spoils the scaling of $\delta^2\omega$. In the following, we focus on reducing the systematic error $x - x_e$ by using QEM.

**Error-mitigated quantum metrology.** — Here, we introduce our general framework of error-mitigated quantum metrology, which is inspired by purification-based QEM [50, 61]. We assume the noise fluctuates from one experimental run to another, i.e., a different quantum state for the $i$th measurement described by $\rho_i$. The key quantity of our framework is the following error-mitigated expectation value of the observable $O$ measured in the quantum metrology:

$$\langle O \rangle_{\text{mit}} = \frac{\text{Tr}[\overline{\rho}O]}{\text{Tr}[\overline{\rho}^n]}, \quad (4)$$

where $\overline{\rho} = \frac{1}{N_{\text{samp}}} \sum_i \rho_i^n$ is the unnormalized purified density matrix and $N_{\text{samp}}$ is the number of measurements; later, we will discuss its filtering effect on fluctuating noise.

Figure 1(c) is a schematic representation of the protocol to calculate $\langle O \rangle_{\text{mit}}$. The procedure is as follows:
(1) make 2n copies of the initial state \( \rho(0) \); (2) expose these 2n copies to the target field simultaneously (or almost at the same time) for an interaction time \( t \), so that we can obtain the 2n copies of \( \rho_i(t) \) with fluctuating noise; (3) divide these 2n copies of \( \rho_i(t) \) in half; (4) input n of 2n density matrices to the purification circuit and obtain a single shot measurement outcome to calculate the numerator in Eq. (4); (5) input the remaining n density matrices in a similar manner to (4) but with setting \( O = I \), to calculate the denominator in Eq. (4); (6) repeat (1)-(5) to compute \( \text{Tr}[\rho^n O] \) and \( \text{Tr}[\rho^n] \); then, compute Eq. (4); (7) estimate \( \omega \) by using the error-mitigated expectation value \( \langle O \rangle_{\text{mit}} \) instead of using \( p \) in Eq. (1). Note that the density matrix for the ensemble of \( N_{\text{samp}} \) states input to the purification circuit can be described as \( \rho = \sum_{i} n_i \rho_i^{\otimes n_i} / \sum_{i} n_i \); we can then obtain Eq. (1) from a simple calculation.

Now, we show that our protocol can filter out the noisy states and extract a dominant pure state even in the presence of fluctuating noise. Denoting the spectral decomposition of \( \rho_i \) as \( \rho = \sum_k p_k^{(i)} |\psi_k^{(i)}\rangle \langle \psi_k^{(i)}| \), we have

\[
\bar{\rho} = \frac{1}{N_{\text{samp}}} p_{\text{max}} \sum_{ik} \left( \frac{p_k^{(i)}}{p_{\text{max}}} \right)^n |\psi_k^{(i)}\rangle \langle \psi_k^{(i)}| = \frac{p_{\text{max}}}{N_{\text{samp}}} |\psi_{\text{max}}\rangle \langle \psi_{\text{max}}| \quad (n \to \infty),
\]

where \( p_{\text{max}} = \max_{ik} p_k^{(i)} \) and \( |\psi_{\text{max}}\rangle \) is the corresponding eigenstate. Thus, the contribution of states other than \( |\psi_{\text{max}}\rangle \) is exponentially suppressed as the number of copies \( n \) increases, which means that our method can filter out the fluctuating noise. In general, the dominant eigenvector of the mixed state is distorted by noise and differs from the ideal quantum state, which is called coherent mismatch. Nevertheless, our method clearly eliminates the systematic errors, to allow for a dramatic improvement in \( \delta^2 \omega \) in practical scenarios, as we will see later in the numerical simulations.

**Demonstration of error-mitigated quantum metrology.**—Here, we use error-mitigated quantum metrology to improve \( \delta^2 \omega \) under Markovian and time-inhomogeneous local amplitude damping. We choose the initial probe state \( \rho(0) = |\text{GHZ}\rangle \langle \text{GHZ}| \) as the L-qubit Greenberger–Horne–Zeilinger (GHZ) state \( |\text{GHZ}\rangle = (|0\ldots0\rangle + |1\ldots1\rangle) / \sqrt{2} \), where \( |0\rangle \) and \( |1\rangle \) are the eigenstates of the Pauli \( Z \) operator \( \sigma_z \), \( \sigma_z |1\rangle = |1\rangle \) and \( \sigma_z |0\rangle = -|0\rangle \). We consider the Zeeman Hamiltonian

\[
H = \sum_{j=1}^{L} \omega \sigma_j^{(j)} / 2
\]

with a parameter \( \omega \) determined by the target field. Throughout the present letter, we set \( \hbar = 1 \) and assume small \( \omega t \). We also assume that the time needed for state preparation, error mitigation, and readout is much shorter than the interaction time with the magnetic fields. Each state of the 2n copies after the time evolution in the magnetic field for the interaction time \( t \) with noise at the \( i \)th experimental run can be described as

\[
\rho_i(t) = A_i^{(L)} \cdots A_i^{(1)} [e^{-iHt \rho(0) e^{iHt}}],
\]

where \( A_i^{(j)} \) denotes the noise channel of the local amplitude damping acting on the \( j \)th qubit defined by

\[
A_i^{(j)}[\rho^{(j)}] = K_{i,1}^{(j)} [K_{i,1}^{(j)}]^{\dagger} + K_{i,2}^{(j)} [K_{i,2}^{(j)}]^{\dagger},
\]

\[
K_{i,1}^{(j)} = \begin{pmatrix} 0 & \sqrt{1 - \epsilon_i(t)} \\ 0 & 0 \end{pmatrix}, \quad K_{i,2}^{(j)} = \begin{pmatrix} 0 & \sqrt{\epsilon_i(t)} \\ 0 & 0 \end{pmatrix},
\]

with \( \epsilon_i(t) \) being the error rate at the \( i \)th experimental run and \( \rho^{(j)} \) being a single-qubit state of the \( j \)th qubit. Note that the noise channel of the local amplitude damping commutes with the unitary evolution in Eq. (1). The effect of fluctuating noise is included in \( \epsilon_i(t) = 1 - \exp(-t/T_i) \) for Markovian noise and in \( \epsilon_i(t) = 1 - \exp(-t/t_i)^2 \) for time-inhomogeneous noise, where \( T_i \) is the coherence time and fluctuates from one experimental run to another. From the 2n copies of \( \rho_i(t) \), we obtain \( \text{Tr} [\rho(t)^n] \) with \( O = I \) and \( \text{Tr} [\rho(t)^n Y] \) with \( O = Y \), as shown in Eq. (4), where \( Y = 2P_y - I \), \( P_y = |\text{GHZ}_y\rangle \langle \text{GHZ}_y| \), and \( |\text{GHZ}_y\rangle = ((0\ldots0) - i(1\ldots1)) / \sqrt{2} \).

These expectation values lead to an error-mitigated probability of \( \langle P_y \rangle_{\text{mit}} = \text{Tr} [\rho(t)^n P_y] / \text{Tr} [\rho(t)^n] \). These values in turn provide \( \delta^2 \omega \) in the error-mitigated quantum metrology.

In quantum metrology, we usually choose the interaction time that minimizes \( \delta^2 \omega \). However, since the actual error rate is not known, we instead use the following procedure to evaluate \( \delta^2 \omega \): (1) numerically find a ‘pseudo-optimal’ interaction time \( t_{\text{opt}}^{\text{est}} \) that minimizes the estimated uncertainty of omega without systematic errors, \( \delta^2 \omega_{\text{est}} = \text{Var}[\rho_e]/y_e^2 = p_e(1 - p_e)/(y_e^2 N_{\text{samp}}) \) for the total measurement time \( T = N_{\text{samp}} t_{\text{opt}}^{\text{est}} \) (2) calculate the estimation uncertainty using the ‘pseudo-optimal’ time as \( \delta^2 \omega(t = t_{\text{opt}}^{\text{est}}) \). Here, we should note that the estimated values with the subscript ‘\text{e}’ are calculated with the estimated coherence time \( T_{\text{e}} \) and that our protocol requires 2n times more samples than the conventional Ramsey-type protocol does, as we need 2n copies of \( \rho_i(t) \) for error mitigation.

We now turn our attention to the numerical results. First, we demonstrate that the systematic error spoils the scaling of \( \delta^2 \omega \) under the conventional Ramsey-type protocol. We set the estimated coherence time to a constant \( T_{\text{e}} = 1.0 \) and the total measurement time to \( T = 100 \). The ideal case is when the actual coherence time is constant and correctly estimated, \( T_i = T_{\text{e}} = 1.0 \) (the black dotted line in Fig. 2(a) and (b)). We can see that \( \delta^2 \omega \) follows the conventional scaling with the standard quantum limit, \( \delta^2 \omega \sim L^{-1} \), for Markovian noise and superclassical scaling, \( \delta^2 \omega \sim L^{-3/2} \), for time-inhomogeneous noise, which is also called Zeno scaling. However, the scaling is significantly spoiled by an incorrect estimation of the coherence time, \( T_i \neq T_{\text{e}} \). To simulate an incorrect estimation, we calculate \( \delta^2 \omega \) with the actual coherence time drifting from 1.0 to 0.5 as \( T_i = 1.0 - 0.5 t_i / N_{\text{samp}} \). As shown by the blue line with points in Fig. 2(a) and (b), the scaling of \( \delta^2 \omega \) is significantly spoiled by the incorrect estimation: \( \delta^2 \omega \sim L^{-5} \) for Markovian noise and
δ^2ω \sim L^{-1} (SQL) for time-inhomogeneous noise. This phenomenon can be explained as follows. The pseudo-optimal time $t_{\text{opt}}^c$ scales as $L^{-1} (L^{-1/2})$ for Markovian (time-inhomogeneous) noise, leading to $|x - x_e| \sim L^0$ and $y_e \sim L^0 (y_e \sim L^{-1})$; see the detailed expression in the Supplementary Materials [66]. Accordingly, the systematic error in Eq. (3), $(x - x_e)^2 / y_e^2$ scales as $L^0 (L^{-1})$, which is worse than the statistical error (the first term in Eq. (3)). Constant (SQL) scaling of $\delta^2\omega$ therefore appears when the systematic error is dominant, i.e., for a large $L$. In general, the scaling of $(x - x_e)^2 / y_e^2$ is worse than that of $\text{Var}[p]/(y_e^2 N_{\text{samp}})$ for a large $N_{\text{samp}}$, so the former scaling spoils the latter.

Next, we show that our scheme mitigates the systematic error and dramatically improves the scaling of $\delta^2\omega$. The red line with triangles (for $n = 2$) and the green line with squares (for $n = 3$) in Fig. 2(a) and (b) show $\delta^2\omega$ in our error-mitigated quantum metrology; they demonstrate that our protocol recovers the SQL scaling for Markovian noise and superclassical scaling for time-inhomogeneous noise even when the estimated coherence time is different from the actual one. In the usual purification-based QEM, increasing the number of copies $n$ suppresses the systematic error at the cost of a higher sampling cost [50]. The optimal $n$ therefore depends on $L$: increasing $n$ enlarges $\delta^2\omega$ when the statistical error is dominant and decreases it when the systematic error is dominant. The behavior of $\delta^2\omega$ in Fig. 2 is a good example [66]. In the present example in Fig. 2, however, $n = 2$ is optimal for the case with QEM ($n \geq 2$). This is explained as follows. The systematic error is suppressed by not only increasing $n$ but also increasing $L$ as $x_e - x \sim L^{-(n-1)}$ for both Markovian and time-inhomogeneous noise [66]. Thus, $(x - x_e)^2 / y_e^2$ in Eq. (3) scales as $L^{-(n-1)} (L^{-n})$ while $\text{Var}[p]/y_e^2$ in Eq. (3) scales as $L^{-1} (L^{-3/2})$ for Markovian (time-inhomogeneous) noise. This means that $(x - x_e)^2 / y_e^2$ decreases as fast as or faster than $\text{Var}[p]/y_e^2$ as $L$ increases for $n \geq 2$; in other words, $\text{Var}[p]/y_e^2$, which comes from statistical errors, is dominant for a large number of qubits for $n \geq 2$. As we increase the number of $n$, $\text{Var}[p]/y_e^2$ becomes larger due to a higher sampling cost of QEM when statistical errors are dominant. Therefore, the case $n = 2$ is optimal for the case with QEM. The $L$ dependence of $x - x_e$ is interpreted with respect to the Renyi entropy of the error states, which is related to the performance of purification-based QEM [50, 66]. In general, $L$ entangled qubits provide an exponentially increasing state space, and the Renyi entropy of such a state that is affected by decoherence typically increases. Thus, we expect that our protocol will work very well in quantum metrology using entangled states. The Supplementary Materials demonstrate the performance of our method when global depolarizing with the GHZ initial state is considered, where the Renyi entropy of the error states is almost maximum, and the separable state of a three-level system, where the Renyi entropy of the error states does not depend on $L$ [66].
method may be reduced by using other methods related to purification-based QEM [75, 77], and that coherent error may be further reduced by combining our method with generalized subspace expansion [78]. Our simulated results suggest that our scheme would be useful for quantum metrology affected by systematic errors, which typically occur due to fluctuating noise, and thus will pave the way to achieving high sensitivity in quantum metrology.

We thank Yang Wang for useful discussions. This work was supported by JST [Moonshot R&D][Grant No. JP-MJMS2061]; PRESTO, JST, Grant No. JPMPJR1916; MEXT Q-LEAP Grant Nos. JPMXS0120319794 and JPMXS0118068682; Leading Initiative for Excellent Young Researchers MEXT Japan; JST presto (Grant No. JP-MJPR1919) Japan. This work is supported by PRESTO, JST, Grants No. JPMJPR2114. This paper was (partly) based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Science 306, 1330 (2004).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. 96, 010401 (2006).
[3] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photonics 5, 222 (2011).
[4] G. Tóth and I. Apellaniz, J. Phys. A: Math. Theor. 47, 424006 (2014).
[5] C. E. Degen, F. Reinhard, and P. Cappellaro, Rev. Mod. Phys. 89, 035002 (2017).
[6] L. Pezzé, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Rev. Mod. Phys. 90, 035005 (2018).
[7] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett. 79, 3865 (1997).
[8] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, Science 304, 1476 (2004).
[9] R. P. Budoyo, K. Kakuyanagi, H. Toida, Y. Matsuzaki, and S. Saito, Appl. Phys. Lett. 116, 194001 (2020).
[10] Y. Kubo, I. Diniz, C. Grezes, T. Umeda, J. Isoya, H. Sumiya, T. Yamamoto, H. Abe, S. Onoda, T. Ohshima, V. Jacques, A. Dréau, J.-F. Roch, A. Auffeves, D. Vion, D. Esteve, and P. Bertet, Phys. Rev. B 86, 064514 (2012).
[11] H. Toida, Y. Matsuzaki, K. Kakuyanagi, X. Zhu, W. J. Munro, H. Yamaguchi, and S. Saito, Phys. Rev. Mater. 2, 33 (2019).
[12] J. M. Taylor, P. Cappellaro, L. Childress, L. Jiang, D. Budker, P. R. Hemmer, A. Yacoby, R. Walsworth, and M. D. Lukin, Nat. Phys. 4, 810 (2008).
[13] G. Balasubramanian, I. Y. Chan, R. Kolesov, M. Al-Hmoud, J. Tisler, C. Shin, C. Kim, A. Wojcik, P. R. Hemmer, A. Krueger, T. Hanke, A. Leitenstorfer, R. Bratschitsch, F. Jelezko, and J. Wrachtrup, Nature 455, 648 (2008).
[14] J. R. Maze, P. L. Stanwix, J. S. Hodges, S. Hong, J. M. Taylor, P. Cappellaro, L. Jiang, M. V. G. Dutt, E. Togan, A. S. Zibrov, A. Yacoby, R. L. Walsworth, and M. D. Lukin, Nature 455, 644 (2008).
[15] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, Phys. Rev. Lett. 104, 133601 (2010).
[16] F. Casola, T. van der Sar, and A. Yacoby, Nat. Rev. Mater. 3, 17088 (2018).
[17] J. F. Barry, J. M. Schloss, E. Bauch, M. J. Turner, C. A. Hart, L. M. Pham, and R. L. Walsworth, Rev. Mod. Phys. 92, 015004 (2020).
[18] S. Schmitt, T. Gelen, F. M. Stürner, T. Uden, G. Wolf, C. Müller, J. Scheuer, B. Naydenov, M. Markham, S. Pezzagna, J. Meijer, I. Schwarz, M. Plenio, A. Retzker, L. P. McGuinness, and F. Jelezko, Science 356, 832 (2017).
[19] J. M. Boss, K. Cuja, J. Zopes, and C. L. Degen, Science 356, 837 (2017).
[20] P. Neumann, I. Jakobi, F. Dolde, C. Burk, R. Reuter, G. Waldherr, J. Honert, T. Wolf, A. Brunner, J. H. Shim, D. Suter, H. Sumiya, J. Isoya, and J. Wrachtrup, Nano Lett. 13, 2738 (2013).
[21] M. Mehboudi, A. Sanpera, and L. A. Correa, J. Phys. A: Math. Theor. 52, 303001 (2019).
[22] K. V. Hovhannisyan, M. R. Järgensen, G. T. Landi, A. M. Alhambra, J. B. Brask, and M. Perarnau-Llobet, PRX Quantum 2, 020322 (2021).
[23] F. Dolde, H. Fedder, M. W. Doherty, T. Nöbauer, F. Rempp, G. Balasubramanian, T. Wolf, F. Reinhard, L. C. L. Hollenberg, F. Jelezko, and J. Wrachtrup, Nat. Phys. 7, 459 (2011).
[24] A. Facón, E.-K. Dietsche, D. Grosso, S. Haroche, J.-M. Raimond, M. Brune, and S. Gleyzes, Nature 535, 262 (2016).
[25] Y. Matsuzaki, S. C. Benjamin, and J. Fitzsimons, Phys. Rev. A 84, 012103 (2011).
[26] A. W. Chin, S. F. Huelga, and M. B. Plenio, Phys. Rev. Lett. 109, 233601 (2012).
[27] D. V. Averin, K. Xu, Y. P. Zhong, C. Song, H. Wang, and S. Han, Phys. Rev. Lett. 116, 010501 (2016).
[28] Y. Matsuzaki, S. Benjamin, S. Nakayama, S. Saito, and W. J. Munro, Phys. Rev. Lett. 120, 140501 (2018).
[29] S. Kukita, Y. Matsuzaki, and Y. Kondo, arXiv preprint arXiv:2103.11612 (2021).
[30] M. Beau and A. del Campo, Phys. Rev. Lett. 119, 010403 (2017).
[31] G. Arrad, Y. Vinkler, D. Aharonov, and A. Retzker, Phys. Rev. Lett. 112, 150801 (2014).
[32] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. 112, 150802 (2014).
[33] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus, Phys. Rev. Lett. 112, 080801 (2014).
[34] D. A. Herrera-Martí, T. Gelen, D. Aharonov, N. Katz, and A. Retzker, Phys. Rev. Lett. 115, 200501 (2015).
[35] T. Uden, P. Balasubramanian, D. Louzon, Y. Vinkler, M. B. Plenio, M. Markham, D. Twitchen, A. Stacey, I. Lovchinsky, A. O. Sushkov, M. D. Lukin, A. Retzker, B. Naydenov, L. P. McGuinness, and F. Jelezko, Phys. Rev. Lett. 116, 230502 (2016).
[36] Y. Matsuzaki and S. Benjamin, Phys. Rev. A 95, 032303 (2017).
[37] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, Nat. Commun. 9, 78 (2018).
[38] N. Shetttell, W. J. Munro, D. Markham, and K. Nemoto, New J. Phys. 23, 043038 (2021).
[39] S. Zhou and L. Jiang, PRX Quantum 2, 010343 (2021).
[40] I. Rojkov, D. Layden, P. Cappellaro, J. Home, and F. Reiter, arXiv:2101.05817 (2021).
[41] C. Müller, J. Lisenfeld, A. Shnirman, and S. Poletto, Phys. Rev. B 92, 035442 (2015).
[42] F. Yan, S. Gustavsson, A. Kamal, J. Birenbaum, A. P. Sears, D. Hover, T. J. Gudmundsen, D. Rosenberg, G. Samach, S. Weber, J. L. Yoder, T. P. Orlando, J. Clarke, A. J. Kerman, and W. D. Oliver, Nat. Commun. 7, 12964 (2016).
[43] L. V. Abdurakhimov, I. Mahboob, H. Toida, K. Kakuyanagi, and S. Saito, Appl. Phys. Lett. 115, 262601 (2019).
[44] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Nature 567, 491 (2019).
[45] T. Wolf, P. Neumann, K. Nakamura, H. Sumiya, T. Ohshima, J. Isoya, and J. Wrachtrup, Phys. Rev. X 5, 041001 (2015).
[46] G. Strubi and C. Bruder, Phys. Rev. Lett. 110, 083605 (2013).
[47] S. Pang, J. R. G. Alonso, T. A. Brun, and A. N. Jordan, Phys. Rev. A 94, 012329 (2016).
[48] I. Martínez-Rincón, W.-T. Liu, G. I. Viza, and J. C. Howell, Phys. Rev. Lett. 116, 100803 (2016).
[49] A. Shimada, H. Hakoshima, S. Endo, K. Yamamoto, and Y. Matsuzuki, arXiv:2103.14402 (2021).
[50] B. Koczor, Phys. Rev. X 11, 031057 (2021).
[51] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, J. Phys. Soc. Jpn. 90, 032001 (2021).
[52] K. Temme, S. Bravyi, and J. M. Gambetta, Phys. Rev. Lett. 119, 180509 (2017).
[53] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X 8, 031027 (2018).
[54] Y. Li and S. C. Benjamin, Phys. Rev. X 7, 021050 (2017).
[55] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, Sci. Adv. 5, eaaw5686 (2019).
[56] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, Nat. Commun. 11, 587 (2020).
[57] S. McArdle, X. Yuan, and S. Benjamin, Phys. Rev. Lett. 122, 180501 (2019).
[58] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O’Brien, Phys. Rev. A 98, 062339 (2018).
[59] J. Sun, X. Yuan, T. Tsumoda, V. Veidt, S. C. Benjamin, and S. Endo, Phys. Rev. Applied 15, 034026 (2021).
[60] R. LaRose, A. Mari, S. Kaiser, P. J. Kuralekas, A. A. Alves, P. Czarnik, M. E. Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. Thakre, N. Shammah, and W. J. Zeng, arXiv preprint arXiv:2009.04417 (2020).
[61] W. J. Huggins, S. McArdle, T. E. O’Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Phys. Rev. X 11, 041036 (2021).
[62] T. Sugiyama, Phys. Rev. A 91, 042126 (2015).
[63] Y. Takeuchi, Y. Matsuzuki, K. Miyamishi, T. Sugiyama, and W. J. Munro, Phys. Rev. A 99, 022325 (2019).
[64] H. Okane, H. Hakoshima, Y. Takeuchi, Y. Seki, and Y. Matsuzuki, arXiv preprint arXiv:2007.15903 (2020).
[65] Procedures (4) and (5) can be done in any order.
[66] See Supplementary Materials for more details.
[67] B. Koczor, arXiv preprint arXiv:2104.00608 (2021).
[68] I. Chiorescu, Y. Nakamura, C. M. Harmanos, and J. Mooij, Science 299, 1869 (2003).
[69] A. Lupaşcu, C. J. M. Verwijs, R. N. Schouten, C. J. P. M. Harmanos, and J. E. Mooij, Phys. Rev. Lett. 93, 177006 (2004).
[70] J. H. Plantenberg, P. C. de Groot, C. J. P. M. Harmanos, and J. E. Mooij, Nature 447, 836 (2007).
[71] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, Nat. Phys. 7, 565 (2011).
[72] M. Bai, C. Deng, J.-L. Orgiazzi, F. R. Ong, and A. Lupaşcu, Nat. Commun. 3, 1324 (2012).
[73] J. Clarke and F. K. Wilhelm, Nature 453, 1031 (2008).
[74] Y. Makhlkin, G. Schön, and A. Shnirman, Rev. Mod. Phys. 73, 357 (2001).
[75] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, arXiv preprint arXiv:2102.06056 (2021).
[76] M. Huo and Y. Li, arXiv preprint arXiv:2105.01239 (2021).
[77] Z. Cai, arXiv preprint arXiv:2107.07279 (2021).
[78] N. Yoshioka, H. Hakoshima, Y. Matsuzuki, Y. Tokunaga, Y. Suzuki, and S. Endo, arXiv preprint arXiv:2107.02611 (2021).
[79] A. Gruber, A. Dräbenstedt, C. Tietz, L. Fleury, J. Wrachtrup, and C. Von Borczyskowski, Science 276, 1202 (1997).
[80] F. Jelezko, T. Gaebel, I. Popa, A. Gruber, and J. Wrachtrup, Phys. Rev. Lett. 92, 076401 (2004).
[81] G. Balasubramanian, P. Neumann, D. Twitchen, M. Markham, R. Kolesov, N. Mizuochi, J. Isoya, J. Achard, J. Beck, J. Tissler, V. Jacques, P. R. Hommer, F. Jelezko, and J. Wrachtrup, Nat. Mater. 8, 383 (2009).
[82] E. D. Herbschleb, H. Kato, Y. Maruyama, T. Danjo, T. Makino, S. Yamashita, I. Ohki, K. Hayashi, H. Morishita, M. Fujiwara, and N. Mizuochi, Nat. Commun. 10, 3766 (2019).
[83] G. De Lange, Z. Wang, D. Riste, V. Dobrovitski, and R. Hanson, Science 330, 60 (2010).
[84] M. Hein, W. Dür, and H.-J. Briegel, Phys. Rev. A 71, 032350 (2005).
Supplementary materials for “Error-mitigated quantum metrology”

I. TREATMENT OF FLUCTUATING NOISE

Noise typically fluctuates, leading to there being a different density matrix for each measurement. Here, we explain how our protocol treats a different density matrix for each measurement induced by fluctuating noise. In the following, we denote the density matrix for the \(i\)th measurement as \(\rho_i\). Let \(X_i\) denote a random variable describing the \(i\)th measurement outcome in the quantum circuit in Fig. 1(c). The key assumption is that we can define the average and variance of \(X_i\) using \(\rho_i\) as follows:

\[
E[X_i] = \frac{1 + \text{Tr}[\rho_i^2 O]}{2},
\]

\[
\text{Var}[X_i] = E[X_i] (1 - E[X_i]) = \frac{1 - \{\text{Tr}[\rho_i^2 O]\}^2}{4}. \tag{S2}
\]

Introducing the average of \(X_i\) as

\[
\bar{X} = \frac{1}{N_{\text{samp}}} \sum_{i=1}^{N_{\text{samp}}} X_i
\]

with \(N_{\text{samp}}\) being the number of samples, we can calculate the expectation value and variance of \(\bar{X}\) as

\[
E[\bar{X}] = \frac{1}{N_{\text{samp}}} \sum_{i=1}^{N_{\text{samp}}} E[X_i] = \frac{1}{N_{\text{samp}}} \sum_{i=1}^{N_{\text{samp}}} \frac{1 + \text{Tr}[\rho_i^2 O]}{2} = \frac{1 + \text{Tr}[\rho^2 O]}{2}, \tag{S4}
\]

and

\[
\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{N_{\text{samp}}} \sum_{i=1}^{N_{\text{samp}}} X_i\right] = \frac{1}{N_{\text{samp}}^2} \sum_{i=1}^{N_{\text{samp}}} \text{Var}[X_i] = \frac{1}{N_{\text{samp}}^2} \sum_{i=1}^{N_{\text{samp}}} \frac{1 - \{\text{Tr}[\rho_i^2 O]\}^2}{4} = \frac{1 - \{\text{Tr}[\rho^2 O]\}^2}{4}. \tag{S5}
\]

where \(\bar{X} = \sum_{i=1}^{N_{\text{samp}}} A_i / N_{\text{samp}}\) and we have assumed the independence of the \(X_i\)s.

The case of unitary \(O\).— The above \(E[\bar{X}]\) is indeed the expectation value of the measurement in the error-mitigation circuit in Fig. 1(c); we obtain \(\rho_{\text{num}} = (1 + \text{Tr}[\rho^2 O]) / 2\) and \(\text{Var}[\rho_{\text{num}}] = \left(1 - \{\text{Tr}[\rho^2 O]\}^2\right) / 4\) as well as \(\rho_{\text{denom}} = (1 + \text{Tr}[\rho^2 O]) / 2\) and \(\text{Var}[\rho_{\text{num}}] = \left(1 - \{\text{Tr}[\rho^2 O]\}^2\right) / 4\) with \(O = I\). The error-mitigated expectation value of \(O\) is calculated as

\[
\langle O \rangle_{\text{mit}} = \frac{2\rho_{\text{num}} - 1}{2\rho_{\text{denom}} - 1} = \frac{\text{Tr}[ho^2 O]}{\text{Tr}[ho^2 O]}.
\]

We can calculate the variance of \(\langle O \rangle_{\text{mit}}\) by using error propagation as follows. Suppose we would like to calculate the variance of \(z = f/g\). Then, we have

\[
z = z_0 + \Delta z = \frac{f_0 + \Delta f}{g_0 + \Delta g} \approx \frac{f_0}{g_0} + \frac{\Delta f}{g_0} - \frac{f_0 \Delta g}{g_0^2}, \tag{S7}
\]

where \(z_0, f_0,\) and \(g_0\) are the expectation values of \(z, f,\) and \(g,\) respectively, and \(\Delta z \equiv z - z_0, \Delta f \equiv f - f_0,\) and \(\Delta g \equiv g - g_0\). The variance of \(z\) is calculated as

\[
\text{Var}[z] = E[(\Delta z)^2] = \frac{\text{Var}[f]}{g_0^2} + \frac{f_0^2 \text{Var}[g]}{g_0^4} - 2 \frac{f_0 \text{E}[\Delta f \Delta g]}{g_0^3}. \tag{S8}
\]

Inserting \(f = 2\rho_{\text{num}} - 1\) and \(g = 2\rho_{\text{denom}} - 1\) into Eq. 8 gives the variance of \(\langle O \rangle_{\text{mit}}\) as

\[
\text{Var}[\langle O \rangle_{\text{mit}}] = \frac{1 - \{\text{Tr}[\rho^2 O]\}^2}{N_{\text{samp}} \{\text{Tr}[\rho^2 O]\}^2} + \frac{\text{Var}[\rho^2 O]}{N_{\text{samp}} \{\text{Tr}[\rho^2 O]\}^4}. \tag{S9}
\]
where we have used \( \text{Var}[f] = 4\text{Var}[p_{\text{num}}] = 1 - (\text{Tr}[\rho^n O])^2 \) and \( \text{Var}[g] = 4\text{Var}[p_{\text{denom}}] = 1 - (\text{Tr}[\rho^n])^2 \). Here, we have assumed that there is no correlation between \( p_{\text{num}} \) and \( p_{\text{denom}} \), which leads to no correlation between \( f = 2p_{\text{num}} - 1 \) and \( g = 2p_{\text{denom}} - 1 \) and \( E[\Delta f \Delta g] = 0 \) in Eq. (S8).

The case in the main text. — However, the above calculation cannot be directly applied to the case in the main text, where we would like to calculate \( \langle p_y \rangle_{\text{mit}} \) but \( P_y \) is not a unitary operation. In the main text, therefore, we set \( O = Y \) defined by \( P_y = (I + Y)/2 \), which is unitary. Accordingly, the error-mitigation circuit in Fig. 1(c) provides \( p_{\text{num}} = (1 + \text{Tr}[\rho^n Y])/2 \) and \( p_{\text{denom}} = (1 + \text{Tr}[\rho^n])/2 \). The error-mitigated expectation value of \( P_y \) is then calculated to be

\[
\langle p_y \rangle_{\text{mit}} = \frac{\text{Tr}[\rho^n] + \text{Tr}[\rho^n Y]}{2} = \frac{p_{\text{denom}} + p_{\text{num}} - 1}{2p_{\text{denom}} - 1}.
\] (S10)

Inserting \( f = p_{\text{denom}} + p_{\text{num}} - 1 \) and \( g = 2p_{\text{denom}} - 1 \) into Eq. (S8) gives the variance of \( \langle p_y \rangle_{\text{mit}} \) as

\[
\text{Var}[\langle p_y \rangle_{\text{mit}}] = \frac{2 - (\text{Tr}[\rho^n])^2 - (\text{Tr}[\rho^n Y])^2 + (\text{Tr}[\rho^n] + \text{Tr}[\rho^n Y])^2 \left[ 1 - (\text{Tr}[\rho^n])^2 \right]}{4 (\text{Tr}[\rho^n])^2 N_{\text{samp}}} + \frac{4 (\text{Tr}[\rho^n])^4 N_{\text{samp}}}{4 (\text{Tr}[\rho^n])^4 N_{\text{samp}}}
- \frac{2 (\text{Tr}[\rho^n])^3 N_{\text{samp}}}{}.
\] (S11)

where we have used \( \text{Var}[f] = \text{Var}[p_{\text{num}}] + \text{Var}[p_{\text{denom}}] = \left[ 2 - (\text{Tr}[\rho^n])^2 - (\text{Tr}[\rho^n Y])^2 \right]/4 \) and \( \text{Var}[g] = 4\text{Var}[p_{\text{denom}}] = 1 - (\text{Tr}[\rho^n])^2 \). Here, since \( f = p_{\text{denom}} + p_{\text{num}} - 1 \) and \( g = 2p_{\text{denom}} - 1 \) provides \( \Delta f = \Delta p_{\text{denom}} + \Delta p_{\text{num}} \) and \( \Delta g = 2\Delta p_{\text{denom}} \), we can calculate \( E[\Delta f \Delta g] \) as follows:

\[
E[\Delta f \Delta g] = E \left[ 2(\Delta p_{\text{denom}})^2 + 2\Delta p_{\text{num}} \Delta p_{\text{denom}} \right] = 2E \left[ (\Delta p_{\text{denom}})^2 \right] = 2\text{Var}[p_{\text{denom}}] = \frac{1 - (\text{Tr}[\rho^n])^2}{2},
\] (S12)

where we have used \( E[\Delta p_{\text{num}} \Delta p_{\text{denom}}] = 0 \), which comes from the above assumption of no correlation between \( p_{\text{num}} \) and \( p_{\text{denom}} \).

II. RELATIONSHIP BETWEEN SYSTEMATIC ERRORS AND RÉNYI ENTROPY

In this section, we discuss the relation between the bias term \( x - x_e \) and the Rényi entropy of the error states under a protocol similar to the one in the main text but for general incoherent noise. After the time evolution in the magnetic field with noise for the interaction time \( t \) at the \( i \)th experimental run, the state is described as

\[
\rho_i(t) = \mathcal{E}_i \left[ e^{-iHt} \rho(0) e^{iHt} \right] = [1 - \epsilon_i(t)] e^{-iHt} \rho(0) e^{iHt} + \epsilon_i(t) \sum_{k=1}^{2^L-1} p_k(t) |\psi_k(t)\rangle \langle \psi_k(t)|
\] (S13)

where \( \epsilon_i(t) \) is the error rate at the \( i \)th experimental run and \( |\psi_k(t)\rangle \) is an error state that forms an orthonormal basis set with \( e^{-iHt} |\text{GHZ}\rangle \) with \( p_k(t) \) being the corresponding probability satisfying \( \sum_{k=1}^{2^L-1} p_k(t) = 1 \). For error-mitigated quantum metrology, we calculate \( \text{Tr} [\rho(t^n)] \) and \( \text{Tr} [\rho(t)^n P_y] \) as follows:

\[
\text{Tr} [\rho(t)^n] = [1 - \epsilon(t)]^n + \epsilon(t)^n \sum_{k=1}^{2^L-1} p_k(t)^n \equiv [1 - \epsilon(t)]^n + \epsilon(t)^n ||p(t)||_n^n,
\] (S14)

\[
\text{Tr} [\rho(t)^n P_y] \approx \frac{1}{2} \left[ (1 - \epsilon(t))^n + \epsilon(t)^n ||p_\perp(t)||_n^n \right] + \frac{1}{2} \left[ (1 - \epsilon(t))^n - \epsilon(t)^n ||p_\perp(t)||_n^n \right] L \omega t,
\] (S15)

where \( ||p(t)||_n = \sum_{k=1}^{2^L-1} |p_k(t)|^n \) with \( ||p(t)||_n \) being the n-norm of the probability vector \( p(t) = [p_1(t), ..., p_{2^L-1}(t)] \), and \( p_\perp(t) \) is one of the probabilities in \{\( p_k(t) \)\}_{k=1}^{2^L-1} \) that corresponds to the state \( e^{-iHt} |\text{GHZ}_\perp\rangle \) with \( |\text{GHZ}_\perp\rangle =
\[
\frac{\text{Tr} \left[ \rho(t)^n P_y \right]}{\text{Tr} \left[ \rho(t)^n \right]} = x + y \omega, \quad (S16)
\]

where \( E(t) = \frac{\epsilon(t)^n}{1 - \epsilon(t)^n} \). Let \( x \) be the \( x \) calculated from the estimated error rate \( \epsilon_e \). Then, the difference between the actual and estimated values, \( |x - x_e| \), is

\[
|x - x_e| = \frac{1}{2} \frac{1 + E(t)[p_{\perp}(t)]^n}{1 + E(t)\|p(t)\|_n^2} - \frac{1}{2} \frac{1 + E_e(t)[p_{\perp}(t)]^n}{1 + E_e(t)\|p(t)\|_n^2}
\]

\[
= \|p(t)\|_n^2 \left| E_e(t) - E(t) \right| \left\{1 - \left( \frac{p_{\perp}(t)}{\|p(t)\|_n} \right)^n \right\}^{1/n}, \quad (S19)
\]

where \( E_e(t) = \frac{\epsilon_e(t)^n}{1 - \epsilon_e(t)^n} \). This expression clarifies the condition for \( x - x_e = 0 \) i.e. no systematic errors: \( E = E_e \), which means a correct estimation of error, or \( p_{\perp}(t) = \|p(t)\|_n \), which means that the error state is only \( e^{-iHt} |\text{GHZ}_k\rangle \). For example, the latter condition, \( p_{\perp}(t) = \|p(t)\|_n \), is satisfied with separable initial states of a two-state system and global or local dephasing noise with initial GHZ states. We also find that the \( L \) dependence of \( |x - x_e| \) is determined by that of \( \|p(t)\|_n \left| E_e(t) - E(t) \right| \), so it is roughly determined by that of \( \epsilon(t)^n \frac{\|p(t)\|_n}{n} \). Here, \( \epsilon(t)^n \frac{\|p(t)\|_n}{n} \) is related to the Rényi entropy of the error states [50],

\[
\epsilon(t)^n \frac{\|p(t)\|_n}{n} = e^{-(n-1)H_n}, \quad (S20)
\]

where

\[
H_n = \frac{1}{1 - n} \ln \left( \epsilon(t)^n \sum_{k=1}^{2^L-1} [p_k(t)]^n \right) = \frac{n}{1 - n} \ln \left\{ \left( \epsilon(t)^n \right)^{1/n} \|p(t)\|_n \right\}
\]

is the Rényi entropy of the error states. Therefore, the performance of our measurement in the presence of incoherent noise depends on the \( L \) dependence of the Rényi entropy of the error states. We should remark here that the local amplitude damping in the main text is coherent noise and the above discussion cannot be applied. Nevertheless, the \( L \) dependence of \( |x - x_e| \) can be roughly interpreted as the Rényi entropy of the error states, as indicated in the next section.

### III. DETAILED CALCULATION FOR THE MAIN TEXT

Here, we provide a detailed calculation of the error-mitigated probability for the case in the main text. The density matrix at the \( i \)th experimental run after interacting with the target magnetic field with noise is described as Eq. (6) in the main text, which provides the following expression for the density matrix:

\[
\rho_i(t) = \frac{1}{2} \left[ 1 - \epsilon_i(t) \right] L |1...1 \rangle \langle 1...1| + \frac{1}{2} \left( 1 + [\epsilon_i(t)]^L \right) |0...0 \rangle \langle 0...0|
\]

\[
+ \frac{1}{2} \left[ 1 - \epsilon_i(t) \right] L/2 (e^{-iLt} |1...1 \rangle \langle 0...0| + e^{iLt} |0...0 \rangle \langle 1...1|) + \frac{1}{2} \sum_{k=1}^{2^L-2} p_{k,i}(t) |\psi_k(t)\rangle \langle \psi_k(t)|
\]

\[
= \frac{1}{2} \left( \lambda_{+,i}(t) |\lambda_{+,i}(t)\rangle \langle \lambda_{+,i}(t)| + \lambda_{-,i}(t) |\lambda_{-,i}(t)\rangle \langle \lambda_{-,i}(t)| \right) + \frac{1}{2} \sum_{k=1}^{2^L-2} p_{k,i}(t) |\psi_k(t)\rangle \langle \psi_k(t)|, \quad (S23)
\]

where

\[
\lambda_{\pm,i}(t) = \frac{1 + \epsilon_i(t)L + [1 - \epsilon_i(t)]^L}{2} \pm \sqrt{\left( 1 + \epsilon_i(t) \right)^L - [1 - \epsilon_i(t)]^L}^2 + 4(1 - \epsilon_i(t))^L \quad (S25)
\]
are eigenvalues and

\[ |\lambda_{+,\ell}(t)\rangle = \cos \theta(t)e^{-iL\omega t/2}|1\rangle + \sin \theta(t)e^{iL\omega t/2}|0\rangle, \]

\[ |\lambda_{-,\ell}(t)\rangle = -\sin \theta(t)e^{-iL\omega t/2}|1\rangle + \cos \theta(t)e^{iL\omega t/2}|0\rangle \]

are the corresponding eigenstates with

\[ \cos 2\theta(t) = \frac{[1 - \epsilon(t)]^L - 1 - [\epsilon(t)]^L}{\sqrt{[1 + [\epsilon(t)]^L - [1 - \epsilon(t)]^L]^2 + 4[1 - \epsilon(t)]^L}}, \]

\[ \sin 2\theta(t) = \frac{2[1 - \epsilon(t)]^{L/2}}{\sqrt{[1 + [\epsilon(t)]^L - [1 - \epsilon(t)]^L]^2 + 4[1 - \epsilon(t)]^L}}. \]  

Here, |ψ_k(t)⟩ = e^{-iHt}|ψ_k⟩ with |ψ_k⟩ being a computational basis that is orthogonal to |0⟩ and |1⟩ (e.g., |010...0⟩, |011...0⟩), and p_{k,i}(t) = [\epsilon(t)]^k[1 - \epsilon(t)]^{L-k} is the corresponding probability with k being the number of 0s in |ψ_k⟩.

To calculate the error-mitigated expectation value, we calculate \( \tr{[\rho(t)]^n} \), \( \tr{[\rho(t)]^nP_y} \), and \( \tr{[\rho(t)]^nY} \) as follows:

\[ \tr{[\rho(t)]^n} = \frac{1}{2^n} \left[ \lambda_{+,\ell}(t)^n + \lambda_{-,\ell}(t)^n + \sum_{k=1}^{L-1} \binom{L}{k} \epsilon(t)^n[1 - \epsilon(t)]^{n(L-k)} \right], \]  

\[ \tr{[\rho(t)]^nP_y} \approx \frac{1}{2} \lambda_{+,\ell}(t)^n + \lambda_{-,\ell}(t)^n + \sum_{k=1}^{L-1} \binom{L}{k} \epsilon(t)^n \left[ 2\theta(t) - \lambda_{-,\ell}(t)^n \sin 2\theta(t) \frac{L\omega_t}{2^n} \right], \]

\[ \tr{[\rho(t)]^nY} = 2\tr{[\rho(t)]^nP_y} - \tr{[\rho(t)]^n}. \]

These quantities provide an error-mitigated probability as

\[ \langle P_y \rangle_{\text{mit}} = \frac{\tr{[\rho(t)]^nP_y}}{\tr{[\rho(t)]^n}} = x + y\omega, \]

where

\[ x = \frac{1}{2} \frac{\lambda_{+,\ell}(t)^n + \lambda_{-,\ell}(t)^n}{\lambda_{+,\ell}(t)^n + \lambda_{-,\ell}(t)^n + \sum_{k=1}^{L-1} \binom{L}{k} \epsilon(t)^n[1 - \epsilon(t)]^{n(L-k)}} \]

\[ y = \frac{1}{2} \frac{\lambda_{+,\ell}(t)^n \sin 2\theta(t) - \lambda_{-,\ell}(t)^n \sin 2\theta(t)}{\lambda_{+,\ell}(t)^n + \lambda_{-,\ell}(t)^n + \sum_{k=1}^{L-1} \binom{L}{k} \epsilon(t)^n[1 - \epsilon(t)]^{n(L-k)}} \]

The estimation uncertainty in our protocol is calculated to be \( \langle \sum_{k=1}^{L-1} \frac{|\epsilon_k|}{\sqrt{[1 + [\epsilon_k]^L - [1 - \epsilon_k]^L]^2 + 4[1 - \epsilon_k]^L}} \rangle_{\text{mit}}, \) where \( \langle \sum \rangle_{\text{mit}} \) is the variance of \( \langle P_y \rangle_{\text{mit}} \) calculated with Eq. (S31). Here, \( x_{\text{e}}, y_{\text{e}} \) are the estimated values of x and y, respectively, and they are calculated using Eqs. (S34), (S35) with the error rate at the ith experimental run, \( \epsilon_{i,e}(t) = 1 - \exp(-t/T_{i,e}) \) for Markovian noise and \( \epsilon_{i,e}(t) = 1 - \exp(-t/T_{i,e})^2 \) for time-inhomogeneous noise with \( T_{i,e} \) being the estimated coherence time, which fluctuates from one experimental run to another.

Finally, we discuss the \( L \) dependence of the systematic error by considering \( x - x_{\text{e}} \). Since the error rate at the 'optimal' time \( \epsilon(t_{\text{opt}}) \) is the order of \( L^{-1} \), we will only consider the term of unity, \( [1 - \epsilon(t_{\text{opt}})]^L \), and \( \epsilon(t_{\text{opt}})[1 - \epsilon(t_{\text{opt}})]^L \) and neglect higher order terms of \( \epsilon(t_{\text{opt}}) \). Accordingly, we obtain \( \lambda_{+,\ell}(t_{\text{opt}}) \approx 1 + [1 - \epsilon(t_{\text{opt}})]^L \), \( \lambda_{-,\ell}(t_{\text{opt}}) \approx 0 \) and \( \sum_{k=1}^{L-1} \binom{L}{k} \epsilon_{e}(t_{\text{opt}})^n[1 - \epsilon_{e}(t_{\text{opt}})]^{n(L-k)} \approx L\epsilon_{e}(t_{\text{opt}})[1 - \epsilon_{e}(t_{\text{opt}})]^{n(L-1)} \). Inserting these quantities into Eq. (S34) yields

\[ x \approx \frac{1}{2} \frac{[1 - \epsilon(t_{\text{opt}})^L]^{nL}}{[1 + [1 - \epsilon(t_{\text{opt}})]^L]^n + L\epsilon_{e}(t_{\text{opt}})[1 - \epsilon_{e}(t_{\text{opt}})]^{n(L-1)}}. \]

With the approximation above, \( x - x_{\text{e}} \) becomes

\[ |x - x_{\text{e}}| \sim \frac{1}{2} L \left\{ \frac{[1 + [1 - \epsilon(t_{\text{opt}})]^L]^n}{[1 + [1 - \epsilon(t_{\text{opt}})]^L]^n} - \frac{\epsilon_{e}(t_{\text{opt}})^n[1 - \epsilon_{e}(t_{\text{opt}})]^{n(L-1)}}{[1 + [1 - \epsilon(t_{\text{opt}})]^L]^n} \right\} \approx L^{-1}. \]
To derive this scaling, we used the following idea. Since $t_{e}^{\text{opt}} \sim L^{-1}$ and $t_{c}^{\text{opt}} \sim L^{-1/2}$ for Markovian and time-inhomogeneous noise, respectively, as noted in the main text, $[1 - \epsilon_{e}(t_{e}^{\text{opt}})]^{L} = \exp[-L(t_{e}^{\text{opt}} / T_{i})] \sim \operatorname{const.}$ for Markovian noise and $[1 - \epsilon_{e}(t_{e}^{\text{opt}})]^{L} = \exp[-L(t_{c}^{\text{opt}} / T_{i})^{2}] \sim \operatorname{const.}$ for time-inhomogeneous noise. Here, for large $L$, $\epsilon_{e}(t_{e}^{\text{opt}}) = 1 - \exp(-t_{e}^{\text{opt}} / T_{i}) \sim L^{-1}$ for Markovian noise and $\epsilon_{e}(t_{c}^{\text{opt}}) = 1 - \exp(-(t_{c}^{\text{opt}} / T_{i})^{2}) \sim (t_{c}^{\text{opt}} / T_{i})^{2} \sim L^{-1}$ for time-inhomogeneous noise. Therefore, $|x - x_{e}|$ scales as $L^{-(n-1)}$ and decreases as $L$ increases. We can understand the scaling in Eq. (S37) in terms of the Rényi entropy of the error states, although the discussion in Sec. II cannot be directly applied because the local amplitude damping involves coherent errors. The Rényi entropy of the error states $|\psi_{k}(t)\rangle$ with the approximation above is roughly calculated using Eqs. (S22) and (S24) as

$$H_{n} = \frac{1}{1 - n} \ln \left( Le(t_{e}^{\text{opt}})^{n} [1 - \epsilon(t_{e}^{\text{opt}})^{n}] L^{-1} \right) \sim \frac{1}{1 - n} \ln [(L^{n+1})] = \ln (L), \quad (S38)$$

Thus, it increases as $L$ increases. This rough approximation reproduces the above exact result of the scaling in Eq. (S37).

IV. GLOBAL DEPOLARIZING NOISE ON GHZ INITIAL STATES FOR A LARGE RÉNYI ENTROPY

In Sec. II we showed that error-mitigated quantum metrology works well for a large Rényi entropy. Here, we demonstrate an improvement of $\delta^{2} \omega$ for the case involving almost the largest Rényi entropy: the global depolarizing noise on GHZ initial states. This noise model is rather artificial but it is expected to provide the highest performance in our scheme. The protocol similar to the one in the main text but uses global depolarizing noise on the GHZ initial states as follows: after the time evolution in the magnetic field with noise for the interaction time $t$, the output state at the $i$th experimental run is described as

$$\rho_{i}(t) = \mathcal{E} \left[ e^{-iHt} \rho(0) e^{iHt} \right] = [1 - \epsilon_{i}(t)] e^{-iHt} \rho(0) e^{iHt} + \frac{\epsilon_{i}(t)}{2L} P_{0}$$

$$= [1 - \epsilon_{i}'(t)] e^{-iHt} \rho(0) e^{iHt} + \epsilon_{i}'(t) \sum_{k=1}^{2^{L-1}} p_{\psi_{k}(t)} |\psi_{k}(t)\rangle \langle \psi_{k}(t)|, \quad (S39)$$

where $\epsilon_{i}(t)$ is the error rate at the $i$th experimental run, $\epsilon_{i}'(t) = \epsilon_{i}(t) \left( 1 - 2^{-L} \right)$, $p = (2^{L} - 1)^{-1}$, and $|\psi_{k}(t)\rangle = e^{-iHt} |\psi_{k}\rangle$ with $|\psi_{k}\rangle$ being the state orthogonal to $|\text{GHZ}\rangle$. Note that we have used the fact that the channel of depolarizing noise commutes with the unitary evolution induced by the Hamiltonian $H$. The effect of fluctuating noise is included in $\epsilon_{i}(t) = 1 - \exp(-t/T_{i})$ for Markovian noise and $\epsilon_{i}(t) = 1 - \exp(-t/T_{i})^{2}$ for time-inhomogeneous noise, where $T_{i}$ is the coherence time at the $i$th experimental run and fluctuates from one experimental run to another. From the input $2n$ copies of $\rho_{i}(t)$, we can obtain the error-mitigated expectation value in Eq. (4) by computing the average of the denominators $\operatorname{Tr} \left[ \rho(t)^{n} \right]$ and numerators $\operatorname{Tr} \left[ \rho(t)^{n} Y \right]$, where

$$\operatorname{Tr} \left[ \rho(t)^{n} \right] = \frac{[1 - \epsilon'(t)]^{n} + \epsilon'(t)^{n} p^{n-1}}{L \omega t},$$

$$\operatorname{Tr} \left[ \rho(t)^{n} P_{y} \right] \approx \frac{1}{2} \left\{ [1 - \epsilon'(t)]^{n} + \epsilon'(t)^{n} p^{n} \right\} + \frac{1}{2} \left\{ [1 - \epsilon'(t)]^{n} - \epsilon'(t)^{n} p^{n} \right\} L \omega t,$$

$$\operatorname{Tr} \left[ \rho(t)^{n} Y \right] = 2 \operatorname{Tr} \left[ \rho(t)^{n} P_{y} \right] - \operatorname{Tr} \left[ \rho(t)^{n} \right]. \quad (S42)$$

These expectation values lead to the error-mitigated probability,

$$\langle P_{y} \rangle_{\text{mit}} = \frac{\operatorname{Tr} \left[ \rho(t)^{n} P_{y} \right]}{\operatorname{Tr} \left[ \rho(t)^{n} \right]} = x + y \omega, \quad (S43)$$

where

$$x = 1 \frac{[1 - \epsilon'(t)]^{n} + \epsilon'(t)^{n} p^{n}}{2 [1 - \epsilon'(t)]^{n} + \epsilon'(t)^{n} p^{n-1}},$$

$$y = 1 \frac{[1 - \epsilon'(t)]^{n} - \epsilon'(t)^{n} p^{n}}{2 [1 - \epsilon'(t)]^{n} + \epsilon'(t)^{n} p^{n-1}} L \omega t. \quad (S45)$$
\( \delta^2 \omega \) in this protocol is calculated in the same way as in the main text and is shown in Fig. S1 we have used the same parameters and settings as in the main text. The behavior is the same as that for local amplitude damping noise in the main text. Indeed, \( |x - x_{\text{e}}| \sim 2^{-L(n-1)} \), since \( \|p(t)\|^n \sim 2^{-L(n-1)} \) as calculated with Eqs. (S20) and (S39). As shown in Sec. II, this exponential suppression of the systematic error is due to the \( L \) dependence of the Rényi entropy of the error states, which is calculated to be

\[
H_n = \frac{1}{1 - n} \ln \left( e^{(t^{\text{opt}}/n)^n} \sum_{k=1}^{2^L-1} p^n \right) \sim \ln(2^L - 1).
\] (S46)

Intuitively, the number of the error states under global depolarizing noise increases exponentially with \( L \), and the Rényi entropy of the error states increases. This would be a merit of using our error-mitigated quantum metrology in cases with initial GHZ states.

V. SEPARABLE STATES OF A THREE-LEVEL SYSTEM FOR DESCRIBING THE NITROGEN-VACANCY CENTER

So far, we have considered the GHZ state as an initial state because of its potential for entanglement metrology. Here, we turn our attention to a sensor that has been realized with current technology: the nitrogen-vacancy (NV) center \[5, 17\]. The NV center is composed of a nearest-neighbor nitrogen atom and a vacancy in diamond and provides a three-level system. Even at room temperature, we can polarize the NV center by illuminating green light, and we can also readout the state of the electron spins by measuring the photon emission from the NV center \[79\]. We can also control the NV center by using microwave pulses \[80\]. Moreover, the NV center in diamond has a long coherence time (a few milliseconds) even at room temperature \[81, 82\]. These properties make it a promising candidate for a highly-sensitive quantum sensor \[5, 17\].

One of the challenges to improving the sensitivity of the diamond-based quantum sensor is decoherence. NV centers are affected by dephasing and energy relaxation. It is known that we can suppress dephasing on the NV center by using dynamical decoupling \[83\]. This means that the coherence time of the NV center is ultimately limited by energy relaxation \[84\].

In this section, to show the performance of our protocol for sensing with the NV center, we consider \( L \) separable states of a three-level system under depolarizing noise, which corresponds to an energy relaxation in the high temperature limit.

We consider a system composed of three states, \( |\pm 1\rangle, |0\rangle \), where \( |0\rangle \) and \( |\pm 1\rangle \) are the eigenstates of the Pauli Z operator \( \sigma_z \), such as \( \sigma_z |0\rangle = 0 \) and \( \sigma_z |\pm 1\rangle = \pm |1\rangle \). We choose an initial probe state \( \rho(0) = (|+\rangle \langle +|)^{\otimes L} \) with \( |+\rangle = (|1\rangle + |\pm 1\rangle)/\sqrt{2} \). We prepare \( 2^n \) copies of \( \rho(0) \) for error-mitigated quantum metrology. We consider the Zeeman Hamiltonian \( H = \sum_{j=1}^{L} \omega \sigma_z^{(j)} / 2 \) with a parameter \( \omega \) determined by the target field. After the time evolution in the...
FIG. S2. Estimation uncertainty $\delta^2 \omega$ for $L$ separable states of a three-level system under (a) Markovian and (b) time-inhomogeneous depolarizing noise.

magnetic field with noise for the interaction time $t$, the state at the $i$th experimental run is described as

$$\rho_i(t) = \mathcal{E}_i \left[ e^{-iHt} \rho(0) e^{iHt} \right] = \left[ 1 - \epsilon_i(t) \right] e^{-iHt} \rho(0) e^{iHt} + \frac{\epsilon_i(t)}{3} I$$

(S47)

with $\epsilon_i(t)$ being the error rate at the $i$th experimental run. The effect of fluctuating noise is included in $\epsilon_i(t) = 1 - \exp(-t/T_i)$ for Markovian noise and $\epsilon_i(t) = 1 - \exp(-t^2/T_i^2)$ for time-inhomogeneous noise, where $T_i$ is the coherence time at the $i$th experimental run and fluctuates from one experimental run to another. From the input $2n$ copies of $\rho_i(t)$, we can obtain the error-mitigated expectation value in Eq. (4) by computing the average of the denominators $\text{Tr} [\rho(t)^n]$ and numerators $\text{Tr} [\rho(t)^n Y]$ as

$$\text{Tr} [\rho(t)^n] = \left[ 1 - \frac{2}{3} \epsilon(t) \right]^n + 2 \left( \frac{\epsilon(t)}{3} \right)^n,$$

(S48)

$$\text{Tr} [\rho(t)^n P_y] \approx \frac{1}{2} \left[ \left( 1 - \frac{2}{3} \epsilon(t) \right)^n + \left( \frac{\epsilon(t)}{3} \right)^n \right] + \frac{1}{2} \left[ \left( 1 - \frac{2}{3} \epsilon(t) \right)^n - \left( \frac{\epsilon(t)}{3} \right)^n \right] \omega t,$$

(S49)

$$\text{Tr} [\rho(t)^n Y] = 2 \text{Tr} [\rho(t)^n P_y] - \text{Tr} [\rho(t)^n].$$

(S50)

These expectation values lead to an error-mitigated probability,

$$\langle P_y \rangle_{\text{mit}} = \frac{\text{Tr}[\rho(t)^n P_y]}{\text{Tr}[\rho(t)^n]} = x + y \omega,$$

(S51)

where

$$x = \frac{1}{2} \frac{\left[ 1 - \frac{2}{3} \epsilon(t) \right]^n + \left( \frac{\epsilon(t)}{3} \right)^n}{\left( 1 - \frac{2}{3} \epsilon(t) \right)^n + 2 \left( \frac{\epsilon(t)}{3} \right)^n},$$

(S52)

$$y = \frac{1}{2} \frac{\left[ 1 - \frac{2}{3} \epsilon(t) \right]^n - \left( \frac{\epsilon(t)}{3} \right)^n}{\left( 1 - \frac{2}{3} \epsilon(t) \right)^n + 2 \left( \frac{\epsilon(t)}{3} \right)^n} L t.$$  

(S53)

The estimation uncertainty in our protocol is calculated in the same way as in the main text and is shown in Fig. S2. We used the same parameters and settings as in the main text; the only difference is that we consider $N_{\text{samp}} \rightarrow LN_{\text{samp}}$ because we treat $L$-qubit separable states. Although our scheme works well and recovers the scaling of $\delta^2 \omega$ in the present case, the behavior of $\delta^2 \omega$ is different from the previous case with initial GHZ states. The qualitative behavior
of $\delta^2 \omega$ is independent of both Markovian and time-inhomogeneous noise as shown in Fig. S2, since the initial states are separable and not entangled; it is known that the scaling of $\delta^2 \omega$ for separable states is limited to SQL [5]. Moreover, when the systematic error is dominant, the optimal $n$ increases as $L$ increases. This is because the systematic error in the present case is independent of $L$ as $x - x_e$ is independent of $L$: the Rényi entropy is also independent of $L$. The systematic error is, therefore, suppressed only by increasing $n$, involving an exponentially increasing sampling cost [50]. When the statistical error is dominant, i.e. for small $L$, the increased sampling cost degrades $\delta^2 \omega$, and therefore, a smaller $n$ is optimal. When the systematic error is dominant, i.e. for large $L$, since increasing $n$ exponentially suppresses the systematic error, a large $n$ is optimal. This is the reason for the crossover behavior between the cases of $n = 2$ and $n = 3$ observed in Fig. S2.