arXiv:2408.00079v4 [quant-ph] 26 Mar 2026

Metrologically optimal quantum states under noise

Chao Yin[1], Victor V. Albert[2], and Sisi Zhou[3,4]

[1]Department of Physics and Center for Theory of Quantum Matter, University of Colorado, Boulder CO 80309, USA

[2]Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland, USA

[3]Perimeter Institute for Theoretical Physics, Waterloo, ON, N2L 2Y5, Canada

[4]Department of Physics and Astronomy, Department of Applied Mathematics, and Institute for Quantum Computing, University of Waterloo, Ontario N2L 2Y5, Canada

We propose a class of metrological resource states whose quantum Fisher information scales optimally in both system size and noise rate. In these states, qubits are partitioned into sensing groups with relatively large correlations within a group but small correlations between groups. The states are obtainable from local Hamiltonian evolution, and we design a metrologically optimal and efficient measurement protocol utilizing time-reversed dynamics and single-qubit on-site measurements. Using quantum domino dynamics, we also present a protocol free of the time-reversal step that has an estimation error roughly twice the best possible value. Finally, we show that spin squeezed states are also optimal for noisy metrology under general conditions.

Contents

1 Introduction 2

2 Setting and a Bound on quantum Fisher information 3

3 An optimal example: quantum parity state 5

4 Sufficient condition for optimality 6

5 Locally generated optimal states 7

6 On-site measurement scheme with time-reversal 8

7 Quantum domino dynamics with on-site measurement 10

8 Spin-squeezing 12

9 Discussion 13

Chao Yin: chao.yin@colorado.edu
Victor V. Albert: vva@umd.edu
Sisi Zhou: sisi.zhou26@gmail.com

A Proof of Theorem 1: large QFI per group and small inter-group correlation guarantee optimal noisy metrology **14**
A.1 Proof of Lemma 2: Optimal observable with bounded operator norm 17
A.2 Proof of Lemma 3: Correlations do not increase by local channels 17

B Simplifying single-group QFI for a pure state via Wigner-Yanase correlation functions **18**
B.1 Application: Large correlation for locally generated states implies large QFI per group 19

C Proof of Theorem 2: on-site measurement scheme with time-reversal **20**
C.1 Lieb-Robinson bound: Rigorous treatment of light cones 20
C.2 Proof of Theorem 2 21
C.3 Proof of Lemma 6: The local version of the Loschmidt echo 22

D States generated from quantum domino dynamics **23**
D.1 Proof of Theorem 3: On-site measurements for quantum domino dynamics . 23
D.2 Partitioned domino state 25

E Generalizations **26**
E.1 General QFI upper bound 26
E.2 General optimality of states constructed in the main text 28

F Spin-squeezed states as metrological optimal states **30**

1 Introduction

Quantum metrology [1, 2, 3, 4] is the study of utilizing entanglement and other quantum effects to measure unknown data, such as gravitational waves [5, 6, 7, 8], images [9, 10, 11, 12], magnetic fields [13, 14, 15, 16, 17], and time [18, 19, 20, 21, 22]. A typical metrological protocol involves initializing N probes in a particular (possibly entangled) state, letting the value of an unknown quantity θ be imparted on the state, and estimating θ via a measurement. The Heisenberg limit (HL) [23] restricts the best-case estimation precision to scale inversely with the number N of entangled probes, $\delta\theta \sim 1/N$.

The prototypical example of a "metrologically optimal" state is the Greenberger-Horne-Zeilinger (GHZ) state [23], which is used to attain the HL in estimating a Z-axis magnetic field in a noiseless N-spin system. Unfortunately, the useful but fragile quantum correlations of this state quickly break down in the presence of noise. Instead of attaining the HL, the optimal scaling in a generically noisy setting is the standard quantum limit (SQL), $\delta\theta \sim 1/\sqrt{N}$ [23, 24, 25, 26, 27, 28, 29].

For quantum metrology to continue to bear fruit, it is important to quantify the performance of noise-robust metrological protocols not only in terms of the number of probes N [23, 24, 25, 26, 27, 28, 29], *but also* in terms of the noise rate p of each probe. A more precise calculation for several models with non-negligible noise yields a precision of $\delta\theta \gtrsim \sqrt{p/N}$ [30, 24, 25, 26, 31], meaning that reducing the noise rate p can lead to a $\sqrt{p}$ improvement on top of the naive scaling with $1/\sqrt{N}$. While the improvement can be significant for a small p, it was rarely explored previously what types of quantum states and measurement protocols can achieve it. Numerical evidence [32, 33] suggests that matrix product states with low bond dimension may be optimal for strong noise. However, these

results are based on specific examples, without a general theoretical framework, and the underlying mechanisms behind this behavior remain unclear.

In this work, we pioneer the study of quantum estimation limits as functions of both N and p. We successfully identify criteria on and examples of protocols whose variance scales optimally [1] with N and p and also construct efficient state preparation and measurement protocols. See Fig. 1 for a summary. The extensive entanglement of the GHZ state is optimal in the relatively noiseless setting of $pN \sim$ constant [34], while unentangled states do not yield any p-dependent precision improvement. Interpolating between the two edge cases, we show that a "Goldilocks" amount of quantum correlation—quantified by a correlation length $M \sim 1/p$—is optimal. We also establish the optimality criteria on a different class of states that may be long-range correlated: spin squeezed states. Throughout, we rigorously prove our results in Appendix, and provide intuitions and proof sketches in the main text. Although we are primarily interested in saturating the asymptotic scaling $\delta\theta \sim \sqrt{p/N}$ rather than the numerical prefactor, we will show examples with good quantitative performance.

2 Setting and a Bound on quantum Fisher information

In this work, we will state our optimality conditions for sensing a Z-axis magnetic field under arbitrary noise, with a straightforward extension to any single-qudit Hamiltonian rotations presented in Appendix.

We use dephasing noise as an illuminating warm-up example. Let N qubits, labeled by $i \in \Lambda$ with Pauli matrices X_i, Y_i, Z_i, be prepared in a state ρ, with each qubit i undergoing the channel

$$\mathcal{D}_{\theta,p}(\rho_i) = (1-p)\mathrm{e}^{-\mathrm{i}\theta Z_i}\rho_i \mathrm{e}^{\mathrm{i}\theta Z_i} + pZ_i\mathrm{e}^{-\mathrm{i}\theta Z_i}\rho_i \mathrm{e}^{\mathrm{i}\theta Z_i} Z_i \tag{1}$$

for $0 < p < 1/2$. For simplicity, we assume $p > 1/N$ (otherwise the noise is asymptotically negligible and the HL is achievable). One then measures the final state $\rho_\theta = \mathcal{D}_{\theta,p}^{\otimes N}(\rho)$ to estimate θ.

The estimation precision can be bounded by the state's quantum Fisher information (QFI), $\delta\theta \geq 1/\sqrt{N_\mathrm{e}\mathcal{F}(\rho_\theta)}$, via the quantum Cramér-Rao bound [35, 36, 37, 38, 39, 40], saturable as the number of repeated experiments $N_\mathrm{e} \to \infty$. The QFI for the above example is bounded by

Theorem 0 (Bound on QFI [30, 25, 24, 31])**.**

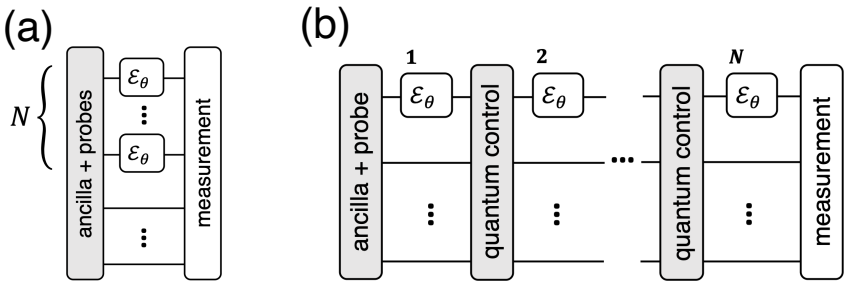
$$\mathcal{F}(\rho_\theta) \leq 4N/[1-(1-2p)^2] \sim N/p. \tag{2}$$

This yields the aforementioned scaling of $\delta\theta \sim \sqrt{p/N}$ [2], which also applies to general noise channels. This QFI bound cannot be surpassed with arbitrary state preparation and measurement protocols, including quantum error correction [27, 29, 28].

Note that Eq. (2) cannot be attained by separable states, for which $\mathcal{F}(\rho_\theta) \leq 4N$ scales constantly with p. The optimal product state is $\rho = \left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\frac{\langle 0|+\langle 1|}{\sqrt{2}}\right)^{\otimes N}$, which achieves $\mathcal{F}(\rho_\theta) = 4(1-2p)^2N$. This clearly has a gap between the $\sim N/p$ upper bound for small p.

[1] We say a family of quantum states ρ as functions of N and p is *optimal*, if and only if for any N and p, there exists a universal constant c such that the QFI $\mathcal{F}(\rho_\theta) \geq c\min\{N/p, N^2\}$.

[2] We use $f \sim g, f \gtrsim g$ and $f \lesssim g$ to denote $c_1 g \leq f \leq c_2 g, f \geq c_3 g$ and $f \leq c_4 g$, respectively, where $c_{1,2,3,4}$ are positive constants.



(c)

Optimal states for noisy metrology	State preparation	Measurement
Large QFI per group of size $\sim p^{-1}$, small correlation among groups (Theorem 1)	?	Measure each group separately
⋃ Suitable-range correlated states in d-dimension	Local Hamiltonian evolution with time $t \sim p^{-1/d}$	On-site measurements after time-reversal (Theorem 2)
⋃ Partitioned domino state	1d local time-independent Hamiltonian with $t \sim p^{-1}$	On-site measurements (Theorem 3)
Spin squeezed states with minimal stretching in orthogonal directions (Theorem 4)	In general not preparable by local dynamics	On-site measurements

Figure 1: Quantum metrology can be done in either parallel (a) or sequential (b) strategies. Although examples are known for both strategies that have the optimal performance under noisy metrology in terms of scaling [26, 31], it has remained unclear whether the resource states used in parallel strategies extend beyond isolated examples to form general and systematic constructions. In this work, we find (c) four *families* of metrological optimal states and their corresponding preparation and measurement protocols. The first three satisfy the small correlation condition Eq. (10) that generalizes the quantum parity state example, and each row is a special case of the previous row as indicated by ⋃. Each nested subset admits a more efficient scheme for state preparation and measurement than the previous one. Spin-squeezed states, described in the fourth row, violate Eq. (10) and do not belong to the previous families.

3 An optimal example: quantum parity state

The QFI upper bound Eq. (2) is the optimal scaling with respect to both N and p, because it can be saturated (up to a constant factor) by the following simple example: Consider the quantum parity state,

$$\rho = \bigotimes_I |\mathrm{GHZ}\rangle_I \langle \mathrm{GHZ}|_I \tag{3}$$

with $|\mathrm{GHZ}\rangle_I = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes M} + |1\rangle^{\otimes M})$ (an instance of the quantum parity code [41, 42, 43]). Here we have partitioned the N-qubit system into N/M groups labeled by $I \subset \Lambda$, each containing M qubits (with N/M assumed to be an integer for simplicity). We do not want to have the global GHZ state $M = N$ because its quantum entanglement is destroyed for any $p \gg 1/N$. To find a suitable choice of M, we first note that in the noiseless limit $p = 0$, each group I gains a phase $M\theta$ independently that corresponds to a QFI $\sim M^2$, so the QFI of the entire state, equal to the sum of individual QFIs, scales as $(N/M)M^2 = NM$. In addition, this scaling is robust against noise if pM is upper bounded by a constant, because then dephasing noise will only cause the individual QFI to reduce by a constant factor [30]. Therefore, Eq. (2) is saturated by tuning $M \sim 1/p$ (see also [44]). To see these claims concretely,

$$\begin{aligned}
\mathcal{F}(\rho_\theta) &= \frac{N}{M}\mathcal{F}(\rho_{I,\theta}) && (\rho_\theta = \textstyle\bigoplus_I \rho_{I,\theta}; \text{Additivity of QFI}) \\
&= \frac{N}{M} 4 \sum_{a<b} \frac{(\lambda_a - \lambda_b)^2}{\lambda_a + \lambda_b} |\langle a|Z_I|b\rangle|^2 && (Z_I := \textstyle\sum_{i\in I} Z_i) \\
&= 4NM(1-2p)^{2M}. && (4)
\end{aligned}$$

Here the second line uses the definition of QFI, where $\{(\lambda_a, |a\rangle)\}$ are eigenvalues and eigenstates of the density matrix in group I

$$\rho_{I,\theta} = \frac{1}{2}\begin{pmatrix} 1 & (1-2p)^M \\ (1-2p)^M & 1 \end{pmatrix}, \tag{5}$$

(setting $\theta = 0$) supported in the subspace containing $|0\cdots 0\rangle, |1\cdots 1\rangle$. The eigenstates are therefore $(|0\cdots 0\rangle \pm |1\cdots 1\rangle)/\sqrt{2}$ with $\lambda_a + \lambda_b = 1$ and $|\lambda_a - \lambda_b| = (1-2p)^M$, which leads to the last line of Eq. (4) using $\langle a|Z_I|b\rangle = M$. At $p \ll 1$ (p does not need to scale with N),

$$\mathcal{F}(\rho_\theta) = 4NM(1-2p)^{2M} \approx 4NM\mathrm{e}^{-4Mp} \approx \frac{N}{\mathrm{e}p}, \tag{6}$$

saturates the optimal scaling Eq. (2) up to a factor 1/e, where we have maximized the expression at $M \approx 1/(4p)$.

Replacing the GHZ state in each group by another state that achieves the HL within the group, such as a spin-squeezed state [45, 46, 47, 48, 49], does not change that arguments above and yields another optimal state.

The state above has a tensor-product structure w.r.t. the partitioning into groups. This may be too strict of a condition for practical purposes, e.g., when nearest neighbor interactions in the quantum sensor cannot be turned off. Therefore, it is crucial to understand the *robustness* of this partitioning strategy in the presence of correlations among the groups. This is a nontrivial task since the QFI of noisy mixed states is difficult to analyze and has no obvious relation with usual correlation measures [50]. We solve this problem below by showing that any state with sufficiently large intra-group and sufficiently small inter-group correlations achieves the optimal QFI scaling $\sim N/p$. Moreover, our results

throughout hold for general noise channels: The final state is $\rho_\theta = \mathcal{N}_{\theta,p}^{\otimes N}(\rho)$, where $\mathcal{N}_{\theta,p}$ is a general single-qubit noisy channel depending on θ, and the noise parameter p quantifies the diamond distance between noiseless ($\mathcal{N}_{\theta,0}$) and noisy ($\mathcal{N}_{\theta,p}$) evolutions (see Appendix E.1 for precise definitions).

As a remark of the scope of this work, we focus on parallel strategies for quantum metrology [26], where an entangled state is sent into the sensing channel once. The signal can also be sensed multiple times where coherence of the quantum register could provide quantum advantage (see Fig. 1). The quantum parity state example above can be mapped to an optimal sequential protocol using one qubit, which saturates the scaling Eq. (2) where now N means the number of sensing channels. The sequential protocol is simply letting the qubit coherently sense as many $\sim 1/p$ channels as possible set by the decoherence rate p. However, there are situations where one would prefer a parallel strategy rather than a sequential one, e.g. when saving the total sensing time, which motivates this work on *robust* families of parallel strategies.

4 Sufficient condition for optimality

Before stating our general sufficient condition (see Appendix for the proof), we first introduce some notation. For any state σ and operators A, B, define the expectation $\langle A\rangle_\sigma := \mathrm{tr}\,(\sigma A)$, the correlation function

$$C_\sigma(A,B) := \frac{1}{2}\langle AB + BA\rangle_\sigma - \langle A\rangle_\sigma \langle B\rangle_\sigma\,, \tag{7}$$

and the variance $\langle \Delta A^2\rangle_\sigma = C_\sigma(A,A)$. Let ρ_I be the reduced density matrix (RDM) of group I before sensing, and let $\rho_{I,\theta}$ ($\rho_{I,\theta}^{(0)}$) be the noisy (noiseless) RDM after sensing. We use A_I to indicate that operator A acts inside group I, and denote the operator norm by $\|\cdot\|$. The correlation between two groups is $C_\sigma(I;J) := \max_{\|A_I\|=\|A_J\|=1} |C_\sigma(A_I, A_J)|$.

Theorem 1. *For any given p, choose any $M < N$ with*

$$M \le c_{\mathrm{no}}/p, \tag{8}$$

where c_{no} is a positive constant, and partition the qubits into N/M groups as above. Given any state ρ, suppose for any group I, its QFI in the noiseless limit is large,

$$\mathcal{F}\left(\rho_{I,\theta}^{(0)}\right) \sim M^2, \tag{9}$$

and its correlations with other groups are small

$$\textstyle\sum_{J:J\notin\mathcal{S}_I} C_\rho(I;J) \lesssim M^{-2}, \tag{10}$$

where $\mathcal{S}_I$ is some set of groups containing a finite (independent of N, p, M) number of elements (including I) that may have large correlation with I. Then there exists an observable $\mathcal{O} = \sum_I \mathcal{O}_I$ such that

$$\mathcal{F}(\rho_\theta) \ge \frac{1}{\langle \Delta\mathcal{O}^2\rangle_{\rho_\theta}} \left(\partial_\theta \langle\mathcal{O}\rangle_{\rho_\theta}\right)^2 \sim NM. \tag{11}$$

Note that $\langle \Delta\mathcal{O}^2\rangle_{\rho_\theta}$ and $\langle\mathcal{O}\rangle_{\rho_\theta}$ represent its variance and expectation value w.r.t. ρ_θ, respectively. See, e.g., Ref. [51] for the lower bound on QFI in Eq. (11). $\mathcal{O}$ here give rise to an optimal observable to measure, which is a sum of individual observables $\mathcal{O}_I$ on each

group. It implies individual measurements on each group suffice to achieve the optimal QFI, which may still be challenging to implement in cases when the group size M is large.

Theorem 1 implies the optimal QFI scaling (Eq. (2)) is achieved for a family of states when $M \sim 1/p$ and two conditions in Eq. (9) and Eq. (10) are satisfied. They generalize the quantum parity state example:

(i) Large intra-group correlation: Eq. (9) requires that each group has the asymptotically optimal QFI $\sim M^2$ [3]. (For the quantum parity state, the RDM of each group is a M-qubit GHZ state, satisfying this condition.)

(ii) Small inter-group correlation: Eq. (10) requires each group to have a large correlation with *at most* a finite number of other groups. (For the quantum parity state, each group has zero correlations with others.)

To prove Theorem 1, we first use the large intra-group correlation condition to prove that the QFI is $\sim NM$ when $p = 0$, similar to the quantum parity state. Next, we use the small inter-group correlation condition to prove, when M is sufficiently small, i.e., $M \sim 1/p$, that the QFI is only negligibly perturbed by noise. We also adopt similar proof ideas for all other theorems below, where we approach the problem by first calculating the noiseless case and then show negligible perturbation under noise. As a result, our theorems apply for general Hamiltonian and noise models. As we will see below, our general criteria for optimal noisy metrology can be satisfied by a large family of many-body states, where efficient preparation and measurement protocols exist.

5 Locally generated optimal states

We show Theorem 1 can be naturally satisfied by states generated by local Hamiltonian evolution. Here, local Hamiltonians are sums of geometrically local terms supported on a constant-size sphere on a lattice.

Suppose the qubits lie on vertices of a d-dimensional square lattice, start from $|\mathbf{0}\rangle := |0\rangle^{\otimes N}$. We consider states $|\psi\rangle = U|\mathbf{0}\rangle$, where unitaries U are generated by a (possibly time-dependent) local Hamiltonian for time T. Such evolution is *quasi-causal* because it satisfies the following Lieb–Robinson bound [52, 53]: if an operator A acts within set $S \subseteq \Lambda$, then, in the Heisenberg picture,

$$A(T) := U^\dagger A U \approx \widetilde{A}_{\widetilde{S}}, \tag{12}$$

i.e., $A(T)$ can be approximated (up to small error) by an operator $\widetilde{A}_{\widetilde{S}}$. It is strictly supported in the "light cone" region

$$\widetilde{S} := \{i \in \Lambda : \mathsf{d}(i, S) \le T\}, \tag{13}$$

This region includes all vertices within a distance T from S. Note that here T is rescaled such that the Lieb–Robinson velocity [53] is set to 1, $\mathsf{d}(\cdot,\cdot)$ is the distance function with $\mathsf{d}(i, S) = \min_{j\in S} \mathsf{d}(i, j)$, and we will assume Eq. (12) is exact from now on, with the rigorous treatment given in Appendix.

The state $\rho = |\psi\rangle\langle\psi| = U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger$ has correlation length at most $2T$ [54] because faraway light cones do not intersect. The condition Eq. (10) is then satisfied if each group

[3] In Appendix, we show that when ρ is pure, the QFI of each RDM can be interpreted as certain correlation functions of ρ.

I is chosen to be an $L \times \cdots \times L$ hypercube with $L = M^{1/d} \gtrsim T$. Each group can have a large correlation with only its neighboring groups.

To satisfy the condition Eq. (9), we first choose $T \sim L$ so that qubits within each group are sufficiently correlated so as to satisfy Eq. (9) [55]. Then Eq. (9) is equivalent to the requirement of a large intra-group ZZ correlation [3], i.e.,

$$C_\rho(Z_I, Z_I) \geq c_Z L^d T^d, \tag{14}$$

where $Z_I = \sum_{i\in I} Z_i$ and $c_Z \sim 1$.

We will give an example $|\psi\rangle = U|0\rangle$ that satisfies Eq. (14) later. For now, we consider *any* optimal state that is efficiently preparable by local Hamiltonian evolution, and for which we propose an efficient measurement protocol below.

6 On-site measurement scheme with time-reversal

For locally generated states, we propose an on-site measurement scheme by assuming the ability to time-reverse the evolution U (e.g., by changing the sign of the generating Hamiltonian).

After sensing, one applies a unitary $U_{\text{rev}} := U^\dagger e^{i\theta_{\text{pr}} Z}$ to the state ρ_θ, and then measures in the computational basis before post-processing the data. Here, θ_{pr} represents prior knowledge of θ, such that $\tilde{\theta} := \theta - \theta_{\text{pr}}$ satisfies $|\tilde{\theta}| \sim L^{-d}$ [4]. The idea comes from the Loschmidt echo protocol working in the noiseless case [59], where one processes the measurement data by simply projecting the state to $|\mathbf{0}\rangle$. The probability is

$$\mathsf{P}_{\text{echo}} := \left|\langle \mathbf{0}| U^\dagger e^{-i\tilde{\theta} Z} U |\mathbf{0}\rangle\right|^2 = 1 - \frac{\tilde{\theta}^2}{4}\mathcal{F}(\rho_\theta) + \mathrm{O}(\tilde{\theta}^4). \tag{15}$$

The projection is optimal because the corresponding classical Fisher information $\frac{1}{\Delta \mathsf{P}^2_{\text{echo}}} (\partial_\theta \mathsf{P}_{\text{echo}})^2 \approx \mathcal{F}(\rho_\theta)$ at small $\tilde{\theta}$, where $\Delta \mathsf{P}^2_{\text{echo}} = \mathsf{P}_{\text{echo}}(1 - \mathsf{P}_{\text{echo}})$ is the variance of a Bernoulli random variable with probability P_{echo}.

However, the Loschmidt echo protocol above through measuring the observable $|\mathbf{0}\rangle\langle\mathbf{0}|$ no longer works under noise. To guarantee robustness against noise, we need to post-process the data from computational-basis measurement in a different way. Let $P_{\widetilde{I}} := \bigotimes_{i\in\widetilde{I}} |0\rangle_i \langle 0|$ be projections onto small regions $\widetilde{I}$ (defined via Eq. (13)), we measure instead the overall observable $\langle \mathcal{O}_{\text{rev}} \rangle_{\rho_\theta}$ with

$$\mathcal{O}_{\text{rev}} := \sum_I \widetilde{P}_I := \sum_I U_{\text{rev}}^\dagger P_{\widetilde{I}} U_{\text{rev}}. \tag{16}$$

Note that we often omit the identity operator on the rest of the system for simplicity, e.g., here $P_{\widetilde{I}}$ means $P_{\widetilde{I}} \otimes \mathrm{id}$. This time-reversal on-site measurement scheme is asymptotically optimal.

Theorem 2. *For any p, choose $L = M^{1/d}$ such that Eq. (8) holds for some constant c_{no} (which is not necessarily the same constant in Theorem 1). For any locally generated state ρ satisfying Eq. (14), there exist constants $C > c > 1$ determined by c_Z and d, such that for any $C \geq L/T \geq c$,*

$$\frac{1}{\langle \Delta \mathcal{O}^2_{\text{rev}} \rangle_{\rho_\theta}} \left(\partial_\theta \langle \mathcal{O}_{\text{rev}} \rangle_{\rho_\theta}\right)^2 \sim N L^d. \tag{17}$$

[4] The requirement $|\tilde{\theta}| \lesssim L^{-d}$ is a common assumption in metrology because one can always gain the prior knowledge on the "digits" of θ one by one with negligible overhead until the requirement is satisfied [56, 57, 58]. We also assume $|\tilde{\theta}| \gtrsim L^{-d}$, achievable by shifting θ_{pr} if necessary, so that the slope Eq. (21) is large.

Theorem 2 implies locally generated optimal states, combined with the time-reversal measurement protocol, achieve the optimal QFI when choosing $M \sim 1/p$. An additional advantage of the time-reversal measurement protocol is the observable $\mathcal{O}_{\text{rev}}$ is independent of p, in contrast to $\mathcal{O}$ in Theorem 1 that may be a function of p.

Below we sketch the proof idea of Theorem 2, leaving the full proof to Appendix. Note that all approximations "$\approx$" below stem from the light-cone approximation in Eq. (12).

Focusing on $d = 2$ (generalization is straightforward) as shown in Fig. 2 and recalling that $|\tilde{\theta}| \sim L^{-2}$ and $T \sim L$, we first expand in the noiseless case,

$$\begin{aligned}\langle \widetilde{P}_I \rangle_{\rho_\theta^{(0)}} &= \langle \mathbf{0} | \, \mathrm{e}^{\mathrm{i}\tilde{\theta} Z(T)} P_{\widetilde{I}} \mathrm{e}^{-\mathrm{i}\tilde{\theta} Z(T)} \, | \mathbf{0} \rangle \\ &= 1 - \frac{\tilde{\theta}^2}{2} \langle \mathbf{0} | \, [Z(T), [Z(T), P_{\widetilde{I}}]] \, | \mathbf{0} \rangle + \cdots . \end{aligned} \tag{18}$$

The first-order term vanishes similarly as Eq. (15), and we show that the second-order term is $\sim \tilde{\theta}^2 L^4$. (Note that we have required that $|\tilde{\theta}|$ is sufficiently large to guarantee that the second-order derivative is dominant.) We expand $Z = \sum_J Z_J$ and observe that only $J = I$ or neighbors of I contribute, because otherwise $[Z_J(T), P_{\widetilde{I}}] = 0$ since they do not overlap in space from Eq. (12). Furthermore, when L/T is a sufficiently large constant, the coefficient is dominated by the term

$$\langle \mathbf{0} | \, [Z_I(T), [Z_I(T), P_{\widetilde{I}}]] \, | \mathbf{0} \rangle \approx 2 \, C_\rho(Z_I, Z_I) \geq 2 c_Z L^2 T^2, \tag{19}$$

which can be verified from direct computation and Eq. (14). The reason is that the sum of all other terms is $\lesssim LT^3$, which becomes subdominant when L/T is sufficiently large. To show this, we note that for $i \in I, j \in J \neq I$,

$$\langle \mathbf{0} | \, [Z_i(T), [Z_j(T), P_{\widetilde{I}}]] \, | \mathbf{0} \rangle \approx 0, \tag{20}$$

unless the two sites are close ($\mathrm{d}(i,j) \leq 2T$), as shown in Fig. 2. Then the sum of Eq. (20) over i, j is bounded by $\sum_{j \text{ close to } I} \sum_{i \text{ close to } j} \lesssim LT \cdot T^2$.

Above we showed, at sufficiently large L/T, the second-order coefficient in Eq. (18) is $\sim L^4$. This guarantees $-\partial_\theta \langle \widetilde{P}_I \rangle_{\rho_\theta^{(0)}} \sim \tilde{\theta} \times$ [the second-order coefficient] $\sim L^2$, and thus

$$\partial_\theta \langle \mathcal{O}_{\text{rev}} \rangle_{\rho_\theta^{(0)}} \sim N. \tag{21}$$

(Note that the higher orders of $\tilde{\theta}$ in Eq. (18) are safely ignored by similar locality analysis.) Since faraway groups are hardly correlated,

$$\left\langle \Delta \mathcal{O}_{\text{rev}}^2 \right\rangle_{\rho_\theta^{(0)}} \sim \sum_I \langle \Delta \widetilde{P}_I^2 \rangle_{\rho_\theta^{(0)}} \lesssim N/L^2, \tag{22}$$

leading to Eq. (17) for the noiseless case.

Finally, we consider the noisy case $p > 0$. Eq. (22) still holds (where $\rho_\theta^{(0)}$ is replaced with ρ_θ) because adding noise does not create long-range correlation [60]. It then remains to show Eq. (21) is also robust. For illustrative purposes, we focus on dephasing noise here, i.e., $\mathcal{N}_{\theta,p} = \mathcal{D}_{\theta,p}$, yielding

$$\begin{aligned}&\left| \partial_\theta (\langle \widetilde{P}_I \rangle_{\rho_\theta} - \langle \widetilde{P}_I \rangle_{\rho_\theta^{(0)}}) \right| = \left| \langle [Z, \mathcal{D}_{0,p}^{\otimes N}(\widetilde{P}_I) - \widetilde{P}_I] \rangle_{\rho_\theta^{(0)}} \right| \\ &\lesssim L^2 \left\| \mathcal{D}_{0,p}^{\otimes c_0 L^2}(\widetilde{P}_I) - \widetilde{P}_I \right\| \lesssim L^4 p, \end{aligned} \tag{23}$$

where we recall that the dual map of the dephasing noise $\mathcal{D}_{0,p}$ is itself, and use the locality of $\widetilde{P}_I$ to restrict the channel to a region of size $c_0 L^2$ with $c_0 \sim 1$. Taking $pL^2 \leq c_{\text{no}}$ and c_{no} sufficiently small guarantees the contribution from noise to Eq. (21) is subdominant relative to the noiseless result. This completes the proof sketch of Theorem 2.

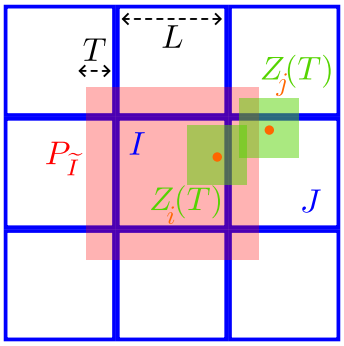


Figure 2: A coarse-grained sketch for the time-reversal on-site measurement protocol in 2d, where the underlying lattice is not shown. There is an observable $P_{\tilde{I}}$ for each group I after the time-reversal $U_{\rm rev}$, whose expectation is only affected by dynamics in the neighboring groups since $L > T$. As a crucial step to establish large sensitivity Eq. (21), we show terms like $\langle \mathbf{0}| [Z_I(T), [Z_J(T), P_{\tilde{I}}]] |\mathbf{0}\rangle$ contribute subdominantly in Eq. (18), using the fact that only sites $i \in I, j \in J$ (orange dots) whose operators $Z_i(T)$ and $Z_j(T)$ (green regions) have overlapping support contribute to these terms.

7 Quantum domino dynamics with on-site measurement

We have shown that any state generated by local dynamics with correlations of the form Eq. (14) is metrologically optimal when $L^d \sim 1/p$. Here we present an explicit example of this framework using quantum domino dynamics [61, 62], where an on-site measurement protocol without time-reversal is sufficient to achieve the optimal QFI.

Consider the 3-local spin-chain Hamiltonian that flips a spin only if its two neighbors are anti-aligned,

$$H_{\rm do} = \frac{1}{2} \sum_{i=2}^{N-1} X_i (1 - Z_{i-1} Z_{i+1}), \tag{24}$$

which can be engineered as the effective Hamiltonian for the transerve-field Ising model at weak X field [61]. The vacuum $|\mathbf{0}\rangle$ is an eigenstate of $H_{\rm do}$, and, more generally, the number of domain walls (DWs)—transitions from $|0\rangle$ to $|1\rangle$ or vice versa—is conserved. If the system starts from the initial state $|1_1\rangle := |10\cdots 0\rangle$, it only traverses the state space with one DW $\mathrm{e}^{-\mathrm{i}TH_{\rm do}} |1_1\rangle = \sum_{i=1}^{N} a_i |1_i\rangle$, where $|1_i\rangle$ is the state with 1 on the first i qubits and 0 on the rest and the coefficients a_i are determined by T. This single-particle hopping process of the DW justifies the name "domino". Before $vT \gtrsim N$ (i.e., when the DW propagates to the right), a_i is centered around $i = vT$ with a velocity v. Then, an initial state $|+\rangle_1 \otimes |\mathbf{0}\rangle$ would evolve to

$$|\psi_{\rm do}\rangle := \mathrm{e}^{-\mathrm{i}TH_{\rm do}} |+\rangle_1 \otimes |\mathbf{0}\rangle = \frac{1}{\sqrt{2}} \left(|\mathbf{0}\rangle + \sum_{i=1}^{N} a_i |1_i\rangle \right), \tag{25}$$

which is roughly a GHZ-like state on the left $\sim vT$ qubits (with other qubits in state $|0\rangle$), useful for metrology [62].

Adapting the domino dynamics to our partition framework, we propose to evolve the initial state containing one $|+\rangle$ in every L qubits:

$$|\Psi_{\rm do}\rangle := \mathrm{e}^{-\mathrm{i}TH_{\rm do}} \left(|+\rangle_1 \otimes |+\rangle_{L+1} \otimes \cdots \right) \otimes |\mathbf{0}\rangle_{\rm rest}. \tag{26}$$

This final state, which we call the *partitioned domino state*, satisfies Eq. (14) for $T \lesssim L$. We numerically study noisy metrology using $|\Psi_{\rm do}\rangle$ on a chain of 200 qubits, and Fig. 3 shows the results. Remarkably, this protocol is not only optimal asymptotically, but also

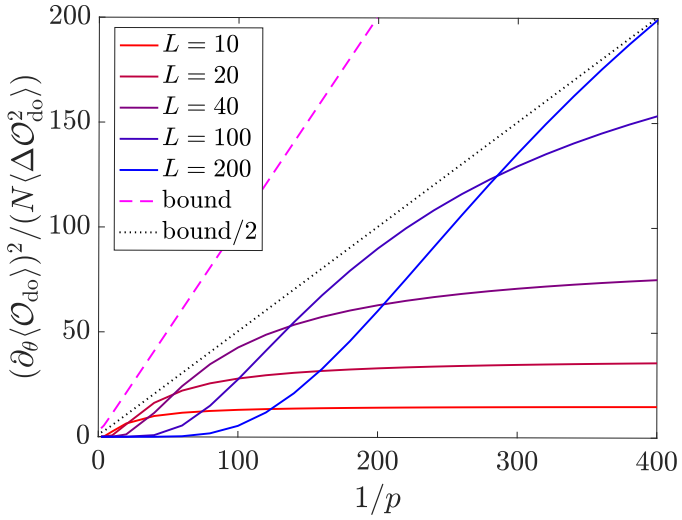


Figure 3: Optimal noisy metrology for $N = 200$ qubits under dephasing, using partitioned domino state Eq. (26) generated from quantum domino dynamics, and on-site measurements $\mathcal{O}_{\rm do}$ of the form Eq. (30). The y-axis quantifies $(\delta\theta)^{-2}/N$ for local estimation at $\theta = 0$. For each solid line, the spin chain is divided to N/L groups each of size L, and the generating time $T \sim L$ is suitably chosen (see Appendix for details). For any noise probability $p \ll 1$ in Eq. (1), by choosing suitable T and L, the performance approaches the bound in Eq. (2) (pink dashed line) up to a factor ≈ 2.

quantitatively. By adjusting L for different p, the precision $(\delta\theta)^2$ is only roughly twice larger than the known lower bound Eq. (2).

Interestingly, the QFI here even outperforms the quantum parity state Eq. (6) with a prefactor $1/2 > 1/\mathrm{e}$. The quantum parity state, as a parallel metrology strategy, can be mapped to a sequential strategy, and this e-factor suppression of the QFI is fundamental for any sequential metrology strategies [26]. Although the quantum parity state inspires our construction of the partitioned domino state, it turns out the latter contains entanglement structure that cannot be captured by mappings from sequential strategies. Intuitively, the partitioned domino state can be mapped to a "coherent superposition" of sequential strategies with varying duration for a quantum metrological register. The fact that such a superposition enhances metrological power could be interesting to investigate in future work.

Besides having the optimal metrological power, the domino-state protocol requires only on-site measurements without any time-reversed evolution. To show this, we focus first on the simplest one-DW case with state Eq. (25) and prove the following in Appendix:

Theorem 3. *In a rotated on-site basis (i.e., redefining states $|1\rangle_i$ with extra phases) such that $\mathrm{Arg}(a_i) = -\frac{\pi}{2}(1+i) - 2\theta_{\rm pr} i$, define the measurement observable*

$$\mathcal{O}_{\rm do} := \sum_{vT/2 \le i \le 2vT} |a_i| \mathbf{Y}_{1,i}, \tag{27}$$

where $\mathbf{Y}_{1,i} := \bigotimes_{j=1}^{i} Y_j$. If

$$\sum_{vT/2 \le i \le 2vT} |a_i|^2 \ge c_{\rm do} \sim 1, \tag{28}$$

then

$$\frac{1}{\langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta}} \left(\partial_\theta \langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta} \right)^2 \sim T^2, \tag{29}$$

if noise $p \le c_{\rm do}/(vT)$ for a constant $c_{\rm do}$.

Eq. (27) can be realized by Y-basis on-site measurements with efficient classical post-processing. The condition Eq. (28) means that $|a_i|^2$, as a probability distribution, is concentrated around vT, which holds before $vT \gtrsim N$.

The idea to choose Eq. (27) as the observable is as follows. $|\psi_{\rm do}\rangle$ is roughly a superposition of GHZ states $|\mathrm{GHZ}\rangle_{1\cdots i} \otimes |\mathbf{0}\rangle$ of different size i. For each GHZ state, $\mathbf{Y}_{1,i}$ is an optimal observable to sense the Z-axis magnetic field. However, each such observable does not have large sensitivity, i.e., $\partial_\theta \langle \mathbf{Y}_{1,i}\rangle \sim |a_i| i \ll i$, because the DW wave packet spreads. Instead, we use $\mathcal{O}_{\rm do}$, a linear combination of them, to accumulate the small derivatives into $\partial_\theta \mathcal{O}_{\rm do} \sim \sum_i |a_i|^2 \sim 1$. $\langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta}$ is also bounded by a constant due to the lack of interference among different GHZ states in $|\psi_{\rm do}\rangle$, making $\mathcal{O}_{\rm do}$ a suitable observable.

Generalizing the idea, the observable we use in Fig. 3 for state Eq. (26) is of the form (see Appendix)

$$\sum_{\text{integer } I} \sum_{1-L/2 \le i \le j \le L/2} b_{i,j} \mathbf{Y}_{IL+i, IL+j} \ , \tag{30}$$

where the real coefficients $b_{i,j}$ come from simulating the two-DW subspace for the group $I = 0$, which only requires *polynomial* (in L) resource of classical computation.

8 Spin-squeezing

We have established in Theorem 1 a sufficient condition for metrological optimality. Here we show this small-correlation condition (Eq. (10)) is *not necessary*: spin-squeezed states (SSS) [45, 46] do not satisfy it in general, but can be optimal (see Appendix).

Theorem 4. *Suppose state ρ satisfies*

$$\langle X \rangle_\rho = \Theta(N), \quad \left\langle \Delta Y^2 \right\rangle_\rho = N\xi^2, \tag{31}$$

$$\left\langle \Delta S^2 \right\rangle_\rho = \mathrm{O}(N/\xi^2), \quad \forall S := \textstyle\sum_i S_i \ \textit{with} \ \|S_i\| = 1 \tag{32}$$

where $X = \sum_i X_i$ etc and S_i denotes an operator on i. Its metrological power is robust in noisy metrology with $p = \mathrm{O}(\xi^2)$:

$$\mathcal{F}(\rho_\theta) \ge \frac{1}{\langle \Delta Y^2 \rangle_{\rho_\theta}} \left(\partial_\theta \langle Y \rangle_{\rho_\theta} \right)^2 \Big|_{\theta=0} = \Theta(N/\xi^2). \tag{33}$$

Theorem 32 generalizes special SSSs and noise analyzed previously [47, 31, 48], where it was shown that SSS could saturate the numerical prefactor in Eq. (2) as well. Eq. (31) is a standard definition of SSS: ρ is polarized in the X direction and squeezed by a factor $\xi \ll 1$ in the Y direction. For example, extremely squeezed states with $\xi = \Theta(1/\sqrt{N})$ exist in the permutation-symmetric subspace of the N qubits, for which any two groups of qubits have a constant correlation $C_\rho(I;J) = \Theta(1)$, violating Eq. (10). Explanations on the violation of Eq. (10) for any sufficiently small ξ can be found in Appendix. Here we require the "polarization" S along any direction in Eq. (32) is at most stretched by a factor $\mathrm{O}(1/\xi)$, the minimum amount allowed by the uncertainty principle. The intuition is, under general noise, the observable Y may mix with "orthogonal" directions S and thus increase its variance. Eq. (32) ensures its impact is sufficiently weak. The condition Eq. (32) suggests the general usefulness of correlation bounds for noisy metrology.

We remark that SSSs were known to achieve the optimal QFI (with the optimal prefactor) for parameter estimation in general noisy quantum channels, when concatenated with an inner layer of quantum error correcting codes [31, 63]. This endows SSSs with a

particular advantage in noisy quantum metrology relative to other state families discussed previously, although achieving the optimal prefactor is not a primary focus of this work.

In general, it is not straightforward to prepare SSSs from natural (i.e., time-independent) local dynamics: The squeezing parameter ξ has a limited scaling with N in recent proposals [64, 65, 66], and it is argued that the 1d case can only generate $\xi \sim 1$ independent of N. This is in sharp contrast with the preparable domino state above, which has optimal robustness against noise and on-site measurement schemes. Nevertheless, it is unclear whether the domino dynamics can be generalized to higher dimensions. The small-correlation states and SSSs are thus complementary classes of metrological resources, with each offering advantages depending on the context.

9 Discussion

In this work, we prove sufficient conditions for optimal noisy metrology and illustrate them with several examples, from simple quantum parity states to quantum domino dynamics and spin-squeezed states. Beyond state optimality, we also address the problem of efficient measurement schemes for metrology. In particular, the domino example suggests that on-site measurements apply to wider sensing scenarios beyond GHZ.

Acknowledgements

C.Y. was supported by the Department of Energy under Quantum Pathfinder Grant DE-SC0024324. S.Z. acknowledges funding provided by Perimeter Institute for Theoretical Physics, a research institute supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. V.V.A. thanks Ryhor Kandratsenia and Olga Albert for providing daycare support throughout this work.

Appendix

For illustrative purpose, we mainly focus on sensing Z magnetic field under dephasing noise $\mathcal{D}_{\theta,p}$ (Eq. (1) in the main text), and prove the three Theorems in this restricted setting first. In the last Section, we prove the full versions in Theorem 8, after formal definitions of general noise. The reason for such arrangement is two-fold: First, the general case requires heavier notations and sometimes more complicated proof techniques. Second, dephasing is intuitively the most dangerous type of noise that would ruin sensing Z field, so if a state is optimal against dephasing, it is expected to be optimal more generally.

We use a stronger non-asymptotic definition of big-O notation to avoid issues like which of the two limits $N \to \infty$ and $p \to 0$ are taken first. Namely, we define

$$\begin{aligned} f = \mathrm{O}(g) \quad &\leftrightarrow \quad f \le c_1 g \\ f = \Theta(g) \quad &\leftrightarrow \quad c_2 g \le f \le c_3 g \\ f = \Omega(g) \quad &\leftrightarrow \quad c_4 g \le f \end{aligned}$$

for some positive constants $c_{1,2,3,4}$. Here, f and g are any positive functions of N (any integer ≥ 3), p (any real number $\in (1/N, 1/2)$), M (any integer satisfying Eq. (36)), and all other parameters whose domains are specified in the contexts. In other words, we say $f = \mathrm{O}(g)$ if there exists a constant $c_{1,2,3,4}$ independent of N, p, M, such that $f \le c_1 g$ holds for any parameters $N, p, \cdots$ in the aforementioned regime.

A Proof of Theorem 1: large QFI per group and small inter-group correlation guarantee optimal noisy metrology

We rephrase Theorem 1 in the main text here. We consider here Pauli-Z signal and dephasing noise for simplicity, and the extension to general signal and noise models will be explained in Section E.

Theorem 1. *For each group I, suppose it has large QFI in the noiseless limit*

$$\mathcal{F}\left(\rho_{I,\theta}^{(0)}\right) \ge c^2 M^2, \tag{34}$$

and its correlation with other groups is bounded by

$$\sum_{J: J \notin \mathcal{S}_I} C_\rho(I; J) \le c' M^{-2}, \quad \text{where} \quad |\mathcal{S}_I| \le c'', \tag{35}$$

for some set $\mathcal{S}_I$. Here, $\mathcal{S}_I \ni I$ contains a finite number of groups that may have large correlation with I. (c, c', c'' and c_{no} below can be any positive constants independent of N, p.) Then for noise

$$p \le c_{\mathrm{no}}/M, \tag{36}$$

there exists an observable $\mathcal{O} = \sum_I \mathcal{O}_I$ such that

$$\mathcal{F}(\rho_\theta) \ge \frac{1}{\langle \Delta \mathcal{O}^2 \rangle_{\rho_\theta}} \left(\partial_\theta \langle \mathcal{O} \rangle_{\rho_\theta}\right)^2 = \Theta(NM) \tag{37}$$

Proof. Our proof strategy is as follows: Following the intuition of the partitioned GHZ example, we first show that QFI of each group Eq. (34) is robust to noise Eq. (36) such that

$$\mathcal{F}\left(\rho_{I,\theta}; Z_I\right) = \Theta(M^2), \tag{38}$$

where $Z_I := \sum_{i\in I} Z_i$, and

$$\mathcal{F}(\sigma; A) := 2 \sum_{\alpha,\beta: p_\alpha + p_\beta > 0} \frac{(p_\alpha - p_\beta)^2}{p_\alpha + p_\beta} |\langle\alpha|A|\beta\rangle|^2 , \tag{39}$$

given the eigendecomposition of σ:

$$\sigma = \sum_\alpha p_\alpha |\alpha\rangle \langle\alpha| . \tag{40}$$

Note that beyond the case when $\partial_\theta \rho_{I,\theta} = -\mathrm{i}[Z_I, \rho_{I,\theta}]$, $\mathcal{F}(\rho_{I,\theta}; Z_I)$ is not equal to $\mathcal{F}(\rho_{I,\theta})$ in general. We then show that Eq. (38) guarantees the existence of an observable $\mathcal{O}_I$ acting in the group that saturates Eq. (38) parametrically. Furthermore, the $\mathcal{O}_I$'s for different groups are roughly independent using the bound Eq. (35) on inter-group correlation, so that Eq. (37) follows roughly by multiplying Eq. (38) with the number of groups N/M.

More precisely, let $\mathcal{D}_{0,p}$ be the pure dephasing channel on a single qubit. Since $\rho_{I,\theta} = \mathcal{D}_{0,p}^{\otimes M}(\rho_{I,\theta}^{(0)})$ and $\mathcal{F}\left(\rho_{I,\theta}^{(0)}; Z_I\right) = \mathcal{F}\left(\rho_{I,\theta}^{(0)}\right)$, the continuity of QFI (see Eq.(A7) in [67]; see also [68]) implies

$$\begin{aligned}
\left|\mathcal{F}(\rho_{I,\theta}; Z_I) - \mathcal{F}(\rho_{I,\theta}^{(0)})\right| &\le 32M^2 \sqrt{\left\|\mathcal{D}_{0,p}^{\otimes M}(\rho_{I,\theta}^{(0)}) - \rho_{I,\theta}^{(0)}\right\|_1} \\
&\le 32M^2 \sqrt{\left\|(\mathcal{D}_{0,p}\otimes \mathrm{id})(\rho_{I,\theta}^{(0)}) - \rho_{I,\theta}^{(0)}\right\|_1 + \left\|(\mathcal{D}_{0,p}\otimes\mathcal{D}_{0,p}\otimes \mathrm{id})(\rho_{I,\theta}^{(0)})) - (\mathcal{D}_{0,p}\otimes \mathrm{id})(\rho_{I,\theta}^{(0)})\right\|_1 + \cdots} \\
&\le 32M^2\sqrt{2pM} \le 32\sqrt{2c_{\mathrm{no}}}M^2,
\end{aligned} \tag{41}$$

where $\|\cdot\|_1$ denotes the trace norm, id denotes the identity operation (or the identity operator as used later), the second line uses triangle inequality to give an upper bound containing M terms, and the third lines uses $\|\mathcal{D}_{0,p}(\sigma) - \sigma\|_1 = \|-p\sigma + pZ_i\sigma Z_i\|_1 \le 2p$ for any density matrix σ. Then according to Eq. (34), Eq. (38) holds as long as c_{no} is a sufficiently small constant $32\sqrt{2c_{\mathrm{no}}} \le c^2/2$.

Then we invoke the following Lemma that we prove later:

Lemma 2. *There exists a positive constant $c_{\mathrm{tri}} = \Theta(1)$ such that the following holds: For any 2^M-dimensional state σ_θ whose dependence on θ is given by*

$$\partial_\theta \sigma_\theta = -\mathrm{i}[A, \sigma_\theta] + B, \tag{42}$$

for some operators A, B with A normalized by $\|A\| = M$, there exists an observable $\mathcal{L}$ such that

$$\partial_\theta \langle\mathcal{L}\rangle_{\sigma_\theta} \ge \frac{1}{2M}\mathcal{F}(\sigma_\theta; A) - c_{\mathrm{tri}} M \|B\|_1 , \tag{43a}$$

$$\left\langle \Delta\mathcal{L}^2 \right\rangle_{\sigma_\theta} \le 1. \tag{43b}$$

Furthermore, it has bounded operator norm

$$\|\mathcal{L}\| \le c_{\mathrm{tri}} M. \tag{44}$$

Then in the case when $\mathcal{F}(\sigma_\theta; A) = \Theta(M^2)$ and B is sufficiently small, measuring $\mathcal{L}$ achieves the HL while at the same time has not-too-large operator norm (e.g., not $\|\mathcal{L}\| = \Theta(\|A\| 2^M)$). Note that the observable $\mathcal{L}$ is similar to but not identical with the symmetric logarithmic derivative [40]: We define $\mathcal{L}$ in Eq. (54) below, so that we can bound its operator norm. It is an interesting open question whether Eq. (44) can be

further improved (e.g., getting rid of the M dependence that allows M^{-2} to be removed from Eq. (35)), and whether the operator norm of the symmetric logarithmic derivative can also be bounded.

In our case, let

$$\sigma_\theta = \rho_{I,\theta}, \tag{45}$$

and the generator

$$A = Z_I. \tag{46}$$

For the dephasing noise here, $B = 0$ because the unitary evolution commutes with the noise channel. For more general noise in Section E, we will show that

$$\|B\|_1 = \mathrm{O}(pM). \tag{47}$$

Eq. (47) implies that with a sufficiently small constant c_{no} in Eq. (36), the second term in Eq. (43a) $= \mathrm{O}(pM^2)$ is subdominant comparing to the first $= \Theta(M)$, so that Lemma 2 implies the existence of an observable $\mathcal{O}_I$ that parametrically saturates the single-group QFI (Eq. (38)):

$$\partial_\theta \langle \mathcal{O}_I \rangle_{\rho_\theta} = \Theta(M), \quad \left\langle \Delta \mathcal{O}_I^2 \right\rangle_{\rho_\theta} \leq 1, \tag{48}$$

$$\|\mathcal{O}_I\| = \mathrm{O}(M), \tag{49}$$

leading to

$$\partial_\theta \langle \mathcal{O} \rangle_{\rho_\theta} = N/M \cdot \Theta(M) = \Theta(N), \tag{50}$$

by summing over I.

Finally, to bound the variance, we derive a correlation bound for the final state ρ_θ from the initial one Eq. (35):

Lemma 3. *Suppose Eq. (35) holds. Then for any set of operators $\{A_J\}$ normalized by*

$$\left\langle \Delta A_J^2 \right\rangle_{\rho_\theta} \leq 1, \tag{51}$$

we have

$$\sum_{J \neq I} |C_{\rho_\theta}(A_I, A_J)| \leq c'' - 1 + c' M^2 \mathsf{A}^2, \ \ \forall I, \ \ \textit{where} \ \ \mathsf{A} = \max_J \|A_J\| . \tag{52}$$

We will prove this Lemma in Section A.2. Note that this is the only place that uses Eq. (35). In other words, the condition Eq. (35) can be replaced by a weaker condition Eq. (52), and Theorem 1 still holds. Nevertheless, we use Eq. (35) because it does not depend on the noise channel.

Let A_I's in Lemma 3 be the observables $\mathcal{O}_I$'s, we bound the total variance:

$$\begin{aligned} \left\langle \Delta \mathcal{O}^2 \right\rangle_{\rho_\theta} &= \sum_I \left\langle \Delta \mathcal{O}_I^2 \right\rangle_{\rho_\theta} + \sum_I \sum_{J \neq I} C_{\rho_\theta}(\mathcal{O}_I, \mathcal{O}_J) \\ &\leq N/M \left(1 + c'' - 1 + c' M^{-2} \mathrm{O}(M^2)\right) \\ &= \Theta(N/M), \end{aligned} \tag{53}$$

which combines with Eq. (50) to yield Eq. (37).

□

A.1 Proof of Lemma 2: Optimal observable with bounded operator norm

Proof. Using the eigendecomposition Eq. (40) for σ_θ with $p_\alpha \geq 0$, define a Hermitian observable as

$$\mathcal{L} = \frac{\mathrm{i}}{M} \sum_{\alpha\beta} \mathrm{sign}(p_\alpha - p_\beta) \langle\alpha|A|\beta\rangle |\alpha\rangle \langle\beta| , \tag{54}$$

where the sign function $\mathrm{sign}(p_\alpha - p_\beta) = 1, -1, 0$ if $p_\alpha - p_\beta > 0, < 0, = 0$ respectively. Assume p_α decreases, after arranging the basis $\{|\alpha\rangle\}$ in order. Up to a factor of i/M, $\mathcal{L}$ is equal to A after the sign of the lower triangular part is flipped. Such operation does not change the norm of the operator too much. According to Eq. (15) in [69],

$$\|\mathcal{L}\| \leq \frac{1}{M} \|\text{upper triangular part of } A\| \leq \frac{1}{M} \left(\frac{8}{\pi^2} \log 2^M + \mathrm{O}(1) \right) \|A\| , \tag{55}$$

which leads to Eq. (44) for some constant c_{tri}.

We then verify Eq. (43) by explicit computation,

$$\begin{aligned} \partial_\theta \langle\mathcal{L}\rangle_{\sigma_\theta} &= \frac{1}{M} \sum_{\alpha\beta} |p_\alpha - p_\beta| \, |\langle\alpha|A|\beta\rangle|^2 + \mathrm{tr}\left(\mathcal{L}B\right) \\ &\geq \frac{1}{M} \sum_{\alpha\beta : p_\alpha + p_\beta > 0} \frac{(p_\alpha - p_\beta)^2}{p_\alpha + p_\beta} |\langle\alpha|A|\beta\rangle|^2 - \|B\|_1 \|\mathcal{L}\| \\ &\geq \frac{1}{2M} \mathcal{F}(\sigma_\theta; A) - c_{\mathrm{tri}} M \|B\|_1 , \end{aligned} \tag{56}$$

where we have used $|p_\alpha - p_\beta| < p_\alpha + p_\beta$ in the second line, and Eq. (44) in the last line. Furthermore,

$$\left\langle \Delta\mathcal{L}^2 \right\rangle_{\sigma_\theta} = \left\langle \mathcal{L}^2 \right\rangle_{\sigma_\theta} \leq \frac{1}{M^2} \sum_{\alpha\beta} p_\alpha |\langle\alpha|A|\beta\rangle|^2 \leq \frac{1}{M^2} \sum_{\alpha} p_\alpha \|A|\alpha\rangle\|^2 \leq \frac{1}{M^2} \|A\|^2 \sum_\alpha p_\alpha = 1, \tag{57}$$

because σ is normalized.

□

A.2 Proof of Lemma 3: Correlations do not increase by local channels

Proof. For the groups in $\mathcal{S}_I$ that may have large correlation with I, we have

$$\sum_{J \in \mathcal{S}_I : J \neq I} |C_{\rho_\theta}(A_I, A_J)| \leq \sum_{J \in \mathcal{S}_I : J \neq I} \sqrt{C_{\rho_\theta}(A_I, A_I) C_{\rho_\theta}(A_J, A_J)} \leq \sum_{J \in \mathcal{S}_I : J \neq I} 1 = |\mathcal{S}_I| - 1 \leq c'' - 1, \tag{58}$$

from Cauchy-Schwarz inequality and Eq. (51).

For the other groups, we have

$$\begin{aligned} \sum_{J \notin \mathcal{S}_I} |C_{\rho_\theta}(\mathcal{O}_I, \mathcal{O}_J)| &= \sum_{J \notin \mathcal{S}_I} \left| C_\rho(\widetilde{\mathcal{O}}_I, \widetilde{\mathcal{O}}_J) \right| \leq \sum_{J \notin \mathcal{S}_I} C_\rho(I; J) \\ &\leq c' M^{-2} \left(\max_J \left\| \widetilde{\mathcal{O}}_J \right\| \right)^2 \leq c' M^{-2} \left(\max_J \|\mathcal{O}_J\| \right)^2 , \end{aligned} \tag{59}$$

where $\widetilde{\mathcal{O}}_J = \mathcal{D}_{\theta,p}(\mathcal{O}_J)$ (the dual map of $\mathcal{D}_{\theta,p}$ is itself), and the second line comes from Eq. (35) and $\left\| \widetilde{\mathcal{O}}_J \right\| \leq \|\mathcal{O}_J\|$ (operator norm is non-increasing under any dual map).

Adding Eq. (58) with Eq. (59) leads to Eq. (52). □

B Simplifying single-group QFI for a pure state via Wigner-Yanase correlation functions

The single-group QFI, i.e., the QFI of the RDM in a single group (Eq. (34)) does not have an explicit expression in general. However, if the entire state ρ_θ is pure, we show that it is almost equivalent to certain correlation functions, which is more intuitive and easier to work with. The connection turns out to come from the quantity called Wigner-Yanase skew information [70]:

$$\mathcal{I}(\rho, A) := \operatorname{tr}\left(\rho A^2\right) - \operatorname{tr}\left(\sqrt{\rho} A \sqrt{\rho} A\right), \tag{60}$$

defined for any state ρ and operator A.

Proposition 4. *For a pure global state $\rho = |\psi\rangle\langle\psi|$, consider bipartitioning the qubits into I and I^{c} (complement of I). Then the QFI for group I in the noiseless case is equivalent to its Wigner-Yanase skew information up to a factor between 1 and 2:*

$$\mathcal{I}(\rho_{I,\theta}^{(0)}, Z_I) \le \mathcal{F}(\rho_{I,\theta}^{(0)}) \le 2\,\mathcal{I}(\rho_{I,\theta}^{(0)}, Z_I). \tag{61}$$

Moreover, the Wigner-Yanase skew information can be expressed as

$$\mathcal{I}(\rho_{I,\theta}^{(0)}, Z_I) = C_\psi(Z_I, Z_I) - C_\psi(Z_I, A^{\mathrm{c}}), \tag{62}$$

for some operator A^{c} acting on I^{c} with norm $\|A^{\mathrm{c}}\| \le 2M$.

Proof. Eq. (61) is known from [71], given that the generator evolving $\rho_{I,\theta}^{(0)}$ is Eq. (46). As a result, we only need to show Eq. (62).

Writing

$$|\psi\rangle = \sum_\alpha \sqrt{p_\alpha}\, |\alpha\rangle_I \otimes |\alpha\rangle_{I^{\mathrm{c}}}, \tag{63}$$

by Schmidt decomposition, we have

$$\rho_{I,\theta}^{(0)} = \mathrm{e}^{-\mathrm{i}\theta Z_I} \left(\sum_\alpha p_\alpha\, |\alpha\rangle_I \langle\alpha|_I \right) \mathrm{e}^{\mathrm{i}\theta Z_I}. \tag{64}$$

Defining

$$\widetilde{Z}_I := Z_I - \langle\psi| Z_I |\psi\rangle, \tag{65}$$

we have

$$\begin{aligned}
\mathcal{I}(\rho_{I,\theta}^{(0)}, Z_I) &= \mathcal{I}\left(\rho_{I,\theta}^{(0)}, \widetilde{Z}_I\right) = \mathcal{I}\left(\sum_\alpha p_\alpha\, |\alpha\rangle_I \langle\alpha|_I, \widetilde{Z}_I\right) \\
&= \sum_\alpha p_\alpha \langle\alpha| \widetilde{Z}_I^2 |\alpha\rangle_I - \sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| \widetilde{Z}_I |\beta\rangle_I \langle\alpha| \widetilde{Z}_I |\beta\rangle_I^* \\
&= \sum_\alpha p_\alpha \langle\alpha| \widetilde{Z}_I^2 |\alpha\rangle_I - \sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| \widetilde{Z}_I |\beta\rangle_I \langle\alpha| A^{\mathrm{c}} |\beta\rangle_{I^{\mathrm{c}}}, \\
&= \langle\psi| \widetilde{Z}_I^2 |\psi\rangle - \langle\psi| \widetilde{Z}_I A^{\mathrm{c}} |\psi\rangle = C_\psi(Z_I, Z_I) - C_\psi(Z_I, A^{\mathrm{c}}),
\end{aligned} \tag{66}$$

where * denotes complex conjugate. Here the first line comes from the fact that the skew information Eq. (60) is unchanged when adding identity to the operator Z_I, and that the

evolution $e^{-i\theta Z_I}$ commutes through $\widetilde{Z}_I$. The second line comes from directly expanding Eq. (60). In the third line, we defined an operator A^c acting on I^c by

$$\langle\alpha| A^c |\beta\rangle_{I^c} = \langle\alpha| \widetilde{Z}_I |\beta\rangle_I^*, \tag{67}$$

and setting other matrix elements to 0. The last line of Eq. (66) simplifies the terms by invoking the Schmidt decomposition Eq. (63), and the definition of the connected correlation function noticing $\langle\psi| \widetilde{Z}_I |\psi\rangle = 0$.

Finally,

$$\|A^c\| = \left\|(P\widetilde{Z}_I P)^*\right\| = \left\|P\widetilde{Z}_I P\right\| \le \left\|\widetilde{Z}_I\right\| \le 2\|Z_I\| = 2M. \tag{68}$$

from Eq. (67), where P is the projector onto $\mathrm{span}\{|\alpha\rangle_I, \forall\alpha\}$. □

B.1 Application: Large correlation for locally generated states implies large QFI per group

Here we show an application of Proposition 4: In the main text, the condition Eq. (14) for Theorem 2 implies condition Eq. (9) in Theorem 1, so the locally generated states for Theorem 2 are special cases of states in Theorem 1. Note that Theorem 2 holds by itself (as we will prove in the next Section) independent of discussion here.

To show Eq. (14) implies Eq. (9) in the main text, we only need to show

$$\mathcal{I}(\rho^{(0)}_{I,\theta}, Z_I) = C_\psi(Z_I, Z_I) - C_\psi(Z_I, A^c) = \Theta(L^{2d}), \tag{69}$$

according to Proposition 4, with A^c defined in Eq. (67). Since the first term $C_\psi(Z_I, Z_I) = \Theta(L^d T^d)$, Eq. (69) will hold as long as the second term scales as $O(L^{d-1}T^{d+1})$, and L/T is chosen as a sufficiently large constant. To bound the second term $C_\psi(Z_I, A^c)$, we separate

$$Z_I = Z_{\mathrm{bulk}} + Z_{\mathrm{boundary}}, \quad \text{where} \quad Z_{\mathrm{boundary}} = \sum_{i\in I:\mathsf{d}(i,I^c)\le 3T} Z_i, \tag{70}$$

(and $\widetilde{Z}_{\mathrm{bulk}}$, $\widetilde{Z}_{\mathrm{boundary}}$ similarly by Eq. (65); here the index I is omitted for simplicity). Here Z_{boundary} (Z_{bulk}) are summed over spins $i \in I$ close to (far from) the boundary. Since the correlation length is roughly $2T$,

$$C_\psi(Z_{\mathrm{bulk}}, A^c) \approx 0, \tag{71}$$

(see the later rigorous Eq. (80)). On the other hand, since Eq. (67) is linear, we can also separate

$$A^c = A^c_{\mathrm{bulk}} + A^c_{\mathrm{boundary}}, \quad \text{where} \quad \langle\alpha| A^c_{\mathrm{boundary}} |\beta\rangle_{I^c} = \langle\alpha| \widetilde{Z}_{\mathrm{boundary}} |\beta\rangle_I^* \tag{72}$$

accordingly. The contribution from A^c_{bulk} is then also negligible by $Z \leftrightarrow A$ symmetry:

$$\begin{aligned}
C_\psi(Z_I, A^c_{\mathrm{bulk}}) &= \sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| \widetilde{Z}_I |\beta\rangle_I \langle\alpha| A^c_{\mathrm{bulk}} |\beta\rangle_{I^c} \\
&= \sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| \widetilde{Z}_I |\beta\rangle_I \langle\alpha| Z_{\mathrm{bulk}} |\beta\rangle_I^* \\
&= \left(\sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| \widetilde{Z}_I |\beta\rangle_I^* \langle\alpha| Z_{\mathrm{bulk}} |\beta\rangle_I\right)^* \\
&= \left(\sum_{\alpha\beta} \sqrt{p_\alpha p_\beta}\, \langle\alpha| A^c |\beta\rangle_I \langle\alpha| Z_{\mathrm{bulk}} |\beta\rangle_I\right)^* = C_\psi(Z_{\mathrm{bulk}}, A^c)^* \approx 0.
\end{aligned} \tag{73}$$

As a result,

$$\begin{aligned} C_\psi(Z_I, A^{\mathrm{c}}) &= C_\psi(Z_I, A^{\mathrm{c}}_{\mathrm{bulk}}) + C_\psi(Z_{\mathrm{bulk}}, A^{\mathrm{c}}_{\mathrm{boundary}}) + C_\psi(Z_{\mathrm{boundary}}, A^{\mathrm{c}}_{\mathrm{boundary}}) \\ &\approx 0 + 0 + C_\psi(Z_{\mathrm{boundary}}, A^{\mathrm{c}}_{\mathrm{boundary}}) \\ &= \mathrm{tr}\left(\sqrt{\rho}\widetilde{Z}_{\mathrm{boundary}}\sqrt{\rho}\widetilde{Z}_{\mathrm{boundary}}\right) \le \mathrm{tr}\left(\rho\widetilde{Z}^2_{\mathrm{boundary}}\right) = C_\psi(Z_{\mathrm{boundary}}, Z_{\mathrm{boundary}}) \\ &= \mathrm{O}(L^{d-1}T \times T^d) = \mathrm{O}(L^{d-1}T^{d+1}). \end{aligned} \tag{74}$$

Here in the second line, we have used Eq. (73) and Eq. (71) where A^{c} is replaced by $A^{\mathrm{c}}_{\mathrm{boundary}}$ (because the only crucial thing is that Z_{bulk} is far from I^{c}). In the third line, we have used Cauchy-Schwarz inequality

$$\mathrm{tr}\left(\sqrt{\rho}\mathcal{O}\sqrt{\rho}\mathcal{O}\right) = \mathrm{tr}\left((\mathcal{O}\sqrt{\rho})^\dagger\sqrt{\rho}\mathcal{O}\right) \le \sqrt{\mathrm{tr}\left((\mathcal{O}\sqrt{\rho})^\dagger\mathcal{O}\sqrt{\rho}\right)\mathrm{tr}\left((\sqrt{\rho}\mathcal{O})^\dagger\sqrt{\rho}\mathcal{O}\right)} = \mathrm{tr}\left(\rho\mathcal{O}^2\right), \tag{75}$$

for Hermitian operator $\mathcal{O} = \widetilde{Z}_{\mathrm{boundary}}$. The last line of Eq. (74) uses the fact that each i is correlated to $\mathrm{O}(T^d)$ other qubits, while there are $\mathrm{O}(L^{d-1}T)$ qubits included in Z_{boundary}.

To conclude, we have shown that the second term of Eq. (69) is subdominant comparing to the first term. Thus, Eq. (14) implies Eq. (9) in the main text.

C Proof of Theorem 2: on-site measurement scheme with time-reversal

C.1 Lieb-Robinson bound: Rigorous treatment of light cones

We first introduce the rigorous version of Eq. (12) in the main text. Suppose the Hamiltonian evolution time during U is actually T' instead of T (although we still use notation $A(T) = U^\dagger A U$), the following Lieb-Robinson bound [52, 53] holds: There exist constants $\mu, \mu', v_{\mathrm{LR}} = \Theta(1)$, such that for any operator A supported on set $S \subset \Lambda$, there exists an operator $\widetilde{A}_{\widetilde{S}}$ acting on $\widetilde{S} = \{i \in \Lambda : \mathsf{d}(i, S) \le T\}$ such that

$$\left\|A(T) - \widetilde{A}_{\widetilde{S}}\right\| \le \mu' \|A\| \, \mathrm{e}^{-\mu(T - v_{\mathrm{LR}}T')}. \tag{76}$$

Therefore, we can choose

$$T = v_{\mathrm{LR}}T' + \Theta(\log L), \tag{77}$$

such that

$$\left\|A(T) - \widetilde{A}_{\widetilde{S}}\right\| \le \|A\| \, \epsilon, \quad \text{where} \quad \epsilon := L^{-\alpha}, \tag{78}$$

for some constant $\alpha > 0$. Later, we will see that ϵ always contributes in a subdominant way. The overhead $\log L \ll L = \Theta(T)$ in Eq. (77) is negligible so that we can still assume (Eq. (14) in the main text)

$$C_\psi(Z_I, Z_I) \ge c_Z L^d T^d, \tag{79}$$

which is equivalent to $C_\psi(Z_I, Z_I) = \Theta[L^d (T')^d]$.

As a consequence of the Lieb-Robinson bound Eq. (76), faraway regions have exponentially decaying correlation in ψ:

$$|C_\psi(A_I, B_J)| \le c'_{\mathrm{LRB}} \|A_I\| \, \|B_J\| \, \mathrm{e}^{-c_{\mathrm{LRB}} \mathsf{d}(I,J)}, \quad \mathsf{d}(I, J) \ge 3v_{\mathrm{LR}}T', \tag{80}$$

for some constants $c_{\mathrm{LRB}}, c'_{\mathrm{LRB}}$, where $\mathsf{d}(I, J) = \min_{i \in I, j \in J} \mathsf{d}(i, j)$ is the distance between the two sets. See Theorem 5.9 in [53] for a proof, where a prefactor $|\partial I| = \Theta(L^{d-1})$ is

absorbed into the exponential using $L = \Theta(T')$. As a result, Eq. (35) is naturally satisfied for ψ, because apart from a finite number of neighboring J's,

$$\sum_{J:\mathsf{d}(I,J)\geq 3T} |C_\rho(A_I, A_J)| \leq c'_{\mathrm{LRB}}\mathsf{A}^2 \sum_{J:\mathsf{d}(I,J)\geq 3T} \mathrm{e}^{-c_{\mathrm{LRB}}\,\mathsf{d}(I,J)} \leq \tilde{c}_{\mathrm{LRB}}\mathsf{A}^2\mathrm{e}^{-3c_{\mathrm{LRB}}T} \ll T^{-2d}\mathsf{A}^2. \tag{81}$$

Here we have used that Eq. (80) also holds for $\mathsf{d}(I, J) \geq 3T$, and $\tilde{c}_{\mathrm{LRB}}$ is a constant from the convergent sum of the exponentially decaying terms.

C.2 Proof of Theorem 2

We repeat Theorem 2 in the main text for convenience:

Theorem 5. *If Eq. (79) holds, then there exist constants c_{no} and $C > c > 1$ determined by c_Z, d, such that for any $C \geq L/T \geq c$ and mild noise Eq. (36),*

$$\frac{1}{\langle \Delta\mathcal{O}^2_{\mathrm{rev}}\rangle_{\rho_\theta}} \left(\partial_\theta \langle \mathcal{O}_{\mathrm{rev}}\rangle_{\rho_\theta}\right)^2 = \Theta(NL^d). \tag{82}$$

Proof. We focus on $d = 2$ for concreteness as shown in Fig. 2 of the main text, and other dimensions follow analogously. We first consider the noiseless case and invoke the following Lemma that we prove later in the next subsection.

Lemma 6. *For sufficiently large constant $c \leq L/T$,*

$$-\partial_\theta \left\langle \widetilde{P}_I \right\rangle_{\rho_\theta^{(0)}} = c_1\tilde{\theta}L^4 + \mathrm{O}(\tilde{\theta}^2L^6) \tag{83}$$

$$= \Theta(L^2), \quad \forall\tilde{\theta} \in [\tilde{c}/(2L^2), \tilde{c}/L^2], \tag{84}$$

for some constants $c_1, \tilde{c} = \Theta(1)$.

According to Eq. (84), $\left|\partial_\theta \langle \mathcal{O}_{\mathrm{rev}}\rangle_{\rho_\theta^{(0)}}\right| = \sum_I \Theta(L^2) = N$. On the other hand,

$$\left\langle \Delta\mathcal{O}^2_{\mathrm{rev}} \right\rangle_{\rho_\theta^{(0)}} = \sum_I \sum_J C_{U_{\mathrm{rev}}\rho_\theta^{(0)}U^\dagger_{\mathrm{rev}}}(P_I, P_J)$$

$$\leq \sum_I \mathrm{O}(1) = \mathrm{O}(N/L^2), \tag{85}$$

which leads to Eq. (82) at $p = 0$. Eq. (85) holds because $U_{\mathrm{rev}}\rho_\theta U^\dagger_{\mathrm{rev}}$ is short-range entangled, so only nearby groups (I, J) have $C_{U_{\mathrm{rev}}\rho_\theta U^\dagger_{\mathrm{rev}}}(P_I, P_J) = \mathrm{O}(\|P_I\|\,\|P_J\|) = \mathrm{O}(1)$; the correlation between faraway groups are exponentially small and safely ignored.

Finally, we consider the noisy case $p > 0$. Eq. (85) still holds when we replace the noiseless $\rho_\theta^{(0)}$ with the noisy ρ_θ because adding noise does not create long-range correlation [60]. Therefore, we only need to show that Eq. (84) is robust. Viewing the dephasing channel as (its dual map) acting on the observable,

$$\left|\partial_\theta \left(\left\langle \widetilde{P}_I \right\rangle_{\rho_\theta} - \left\langle \widetilde{P}_I \right\rangle_{\rho_\theta^{(0)}}\right)\right| = \left|\left\langle [Z, \mathcal{D}_{0,p}^{\otimes N}(\widetilde{P}_I \otimes \mathrm{id}) - \widetilde{P}_I \otimes \mathrm{id}]\right\rangle_{\rho_\theta^{(0)}}\right|$$

$$\leq 2\left\|Z_{\widetilde{I}_T}\right\| \left\|(\mathcal{D}_{0,p}^{\otimes M'} \otimes \mathrm{id})(\widetilde{P}_I \otimes \mathrm{id}) - \widetilde{P}_I \otimes \mathrm{id}\right\| + \mathrm{O}(L^2\epsilon)$$

$$\leq 2M' \sum_{g=1}^{M'} \left\|(\mathcal{D}_{0,p}^{\otimes g} \otimes \mathrm{id})(\widetilde{P}_I \otimes \mathrm{id}) - (\mathcal{D}_{0,p}^{\otimes g-1} \otimes \mathrm{id})(\widetilde{P}_I \otimes \mathrm{id})\right\| + \mathrm{O}(L^2\epsilon) \leq \Theta(L^4)p + \mathrm{O}(L^2\epsilon). \tag{86}$$

Here in the second line, we have again used the locality of $\widetilde{P}_I$ to restrict Z and the dephasing channel to region $\widetilde{I}_T$ defined by $\widetilde{I}_T = \{i \in \Lambda : \mathsf{d}(i, \widetilde{I}) \le T\}$ with size $M' = \Theta(L^2)$. The $\mathrm{O}(L^2\epsilon)$ error comes from the exponentially decaying Lieb-Robinson bound Eq. (76), so that only $\Theta(L^2)$ many Z_i's near $\widetilde{I}_T$ dominates. In the last line, we have used the triangle inequality, $\|\mathcal{D}_{0,p}(A) - A\| = \|p(Z_i A Z_i - A)\| \le 2p\,\|A\|$, and $\left\|(\mathcal{D}_{0,p}^{\otimes g-1} \otimes \mathrm{id})(\widetilde{P}_I \otimes \mathrm{id})\right\| \le \|P_I\| \le 1$. Therefore, there exists constant c_{no} such that Eq. (86) is smaller than Eq. (84) at weak noise Eq. (36), which further leads to Eq. (82) by summing over I and combining with Eq. (85). □

C.3 Proof of Lemma 6: The local version of the Loschmidt echo

Proof of Lemma 6. Since the second line Eq. (84) follows immediately from the first line, we focus on showing the first line Eq. (83) for sufficiently large $c \le L/T$. From direct calculation and Taylor's theorem,

$$\begin{aligned}
-\partial_\theta \left\langle \widetilde{P}_I \right\rangle_{\rho_\theta^{(0)}} &= -\partial_{\tilde\theta} \langle \mathbf{0}| \, \mathrm{e}^{\mathrm{i}\tilde\theta Z(T)} P_{\widetilde{I}} \mathrm{e}^{-\mathrm{i}\tilde\theta Z(T)} \, |\mathbf{0}\rangle = -\mathrm{i} \langle \mathbf{0}| \, \mathrm{e}^{\mathrm{i}\tilde\theta Z(T)} [Z(T), P_{\widetilde{I}}] \mathrm{e}^{-\mathrm{i}\tilde\theta Z(T)} \, |\mathbf{0}\rangle \\
&= \tilde\theta \, \langle \mathbf{0}| \, [Z(T), [Z(T), P_{\widetilde{I}}]] \, |\mathbf{0}\rangle + \frac{\mathrm{i}}{2}\tilde\theta^2 \, \langle \mathbf{0}| \, \mathrm{e}^{\mathrm{i}\tilde\theta' Z(T)} [Z(T), [Z(T), [Z(T), P_{\widetilde{I}}]]] \mathrm{e}^{-\mathrm{i}\tilde\theta' Z(T)} \, |\mathbf{0}\rangle \,,
\end{aligned} \tag{87}$$

where $\tilde\theta' \in [0, \tilde\theta]$, and we have used that the zeroth order vanishes:

$$\langle \mathbf{0}| \, [Z(T), P_{\widetilde{I}}] \, |\mathbf{0}\rangle = \langle \mathbf{0}| \, Z(T) \, |\mathbf{0}\rangle - \langle \mathbf{0}| \, Z(T) \, |\mathbf{0}\rangle = 0. \tag{88}$$

To bound the first term in Eq. (87), we expand $Z = \sum_J Z_J$ and separate it to two kinds of terms (ignoring the prefactor $\tilde\theta$):

(i) The single term

$$\begin{aligned}
\langle \mathbf{0}| \, [Z_I(T), [Z_I(T), P_{\widetilde{I}}]] \, |\mathbf{0}\rangle &= 2 \, \langle \mathbf{0}| \, Z_I^2(T) P_{\widetilde{I}} |\mathbf{0}\rangle - 2 \, \langle \mathbf{0}| \, Z_I(T) P_{\widetilde{I}} Z_I(T) \, |\mathbf{0}\rangle \\
&= 2 \, \langle \mathbf{0}| \, Z_I^2(T) P_{\widetilde{I}} |\mathbf{0}\rangle - 2 \, \langle \mathbf{0}| \, Z_I(T) P_{\widetilde{I}} Z_I(T) P_{\widetilde{I}^{\mathrm{c}}} \, |\mathbf{0}\rangle \\
&= 2 \, \langle \mathbf{0}| \, Z_I^2(T) \, |\mathbf{0}\rangle - 2 \, \langle \mathbf{0}| \, Z_I(T) \left(P_{\widetilde{I}} \otimes P_{\widetilde{I}^{\mathrm{c}}} \right) Z_I(T) \, |\mathbf{0}\rangle + \mathrm{O}(L^4\epsilon) \\
&= 2 \, C_\psi(Z_I, Z_I) + \mathrm{O}(L^4\epsilon) \ge 2 c_Z L^2 T^2 + \mathrm{O}(L^4\epsilon).
\end{aligned} \tag{89}$$

Here we have used $|\mathbf{0}\rangle = P_{\widetilde{I}^{\mathrm{c}}} \, |\mathbf{0}\rangle$ (c means complement) in the second line, and $\left\|[Z_I(T), P_{\widetilde{I}^{\mathrm{c}}}]\right\| \le \|Z_I\| \, \epsilon = L^2\epsilon$ from Eq. (78) in the third line with

$$\langle \mathbf{0}| \, Z_I(T) P_{\widetilde{I}} Z_I(T) P_{\widetilde{I}^{\mathrm{c}}} \, |\mathbf{0}\rangle \le \|Z_I(T)\| \left\|P_{\widetilde{I}}\right\| \left\|[Z_I(T), P_{\widetilde{I}^{\mathrm{c}}}]\right\| \le L^2 \cdot 1 \cdot \epsilon. \tag{90}$$

We will use such manipulations many times later. In the last line of Eq. (89), we have used Eq. (79). One can see the ϵ term is subdominant because $\epsilon \ll 1$.

(ii) All other terms can be arranged by Jacobi identity and $[Z_J(T), Z_K(T)] = 0$ to the form

$$\langle \mathbf{0}| \, [Z_J(T), [Z_K(T), P_{\widetilde{I}}]] \, |\mathbf{0}\rangle \,, \tag{91}$$

where $K \ne I$ (J can be I, K or neither). In other words, if the original $K = I$, then $J \ne I$ (otherwise we return to the single term Eq. (89)), and we use

$$[Z_J(T), [Z_K(T), P_{\widetilde{I}}] = -[Z_K(T), [P_{\widetilde{I}}, Z_J(T)]] - [P_{\widetilde{I}}, [Z_J(T), Z_K(T)]] = [Z_K(T), [Z_J(T), P_{\widetilde{I}}]], \tag{92}$$

before relabeling $J \leftrightarrow K$. Moreover, we can assume J, K in Eq. (91) are neighbors of I (J can also be I itself), i.e. one of the 9 blocks in Fig. 2 of the main text; otherwise for example $\left\|[Z_K(T), P_{\widetilde{I}}]\right\| = \mathrm{O}(\mathrm{e}^{-\mu \mathsf{d}(K,I)}) = \mathrm{O}(\mathrm{e}^{-\mu L})$ is exponentially small from Lieb-Robinson bound Eq. (76), whose contribution to Eq. (87) is negligible.

For Eq. (91), since $K \neq I$, $Z_k(T)$ ($k \in K$) contributes only when it has support on $\widetilde{I}$, the red region in Fig. 2 of the main text. This requires $k \in \widetilde{I}_T = \{i \in \Lambda : \mathsf{d}(i, \widetilde{I}) \leq T\}$, otherwise $\left\|[Z_k(T), P_{\widetilde{I}}]\right\| \leq \epsilon$. This leads to the first line of

$$\begin{aligned}
\langle \mathbf{0}| [Z_J(T), [Z_K(T), P_{\widetilde{I}}]] |\mathbf{0}\rangle &= \sum_{k \in K \cap \widetilde{I}_T} \langle \mathbf{0}| [Z_J(T), [Z_k(T), P_{\widetilde{I}}]] |\mathbf{0}\rangle + \mathrm{O}(L^4 \epsilon) \\
&= \sum_{k \in K \cap \widetilde{I}_T} \sum_{j \in J} \langle \mathbf{0}| [Z_j(T), [Z_k(T), P_{\widetilde{I}}]] |\mathbf{0}\rangle + \mathrm{O}(L^4 \epsilon) \\
&= \sum_{k \in K \cap \widetilde{I}_T} \sum_{j \in J : \mathsf{d}(j,k) \leq 2T} \langle \mathbf{0}| [Z_j(T), [Z_k(T), P_{\widetilde{I}}]] |\mathbf{0}\rangle + \mathrm{O}(L^4 \epsilon) \\
&= \mathrm{O}(LT \cdot T^2) + \mathrm{O}(L^4 \epsilon) = \mathrm{O}(LT^3).
\end{aligned} \tag{93}$$

Here the second line comes from expanding $Z_J = \sum_{j \in J} Z_J$. To get the third line of Eq. (93), observe that for any pairs of j, k with $\mathsf{d}(j, k) > 2T$, there exists bipartition $\widetilde{I} = \widetilde{I}_j \otimes \widetilde{I}_k$ such that $Z_j(T)$ roughly has no support on $\widetilde{I}_k$: $\left\|[Z_j(T), P_{\widetilde{I}_k}]\right\| \leq \epsilon$, and vice versa. Therefore,

$$\begin{aligned}
\langle \mathbf{0}| [Z_j(T), [Z_k(T), P_{\widetilde{I}}]] |\mathbf{0}\rangle &= \langle \mathbf{0}| [Z_j(T), [Z_k(T), P_{\widetilde{I}_k}] \otimes P_{\widetilde{I}_j} + \mathrm{O}(\epsilon)] |\mathbf{0}\rangle \\
&= \langle \mathbf{0}| [Z_j(T), P_{\widetilde{I}_j}] \otimes [Z_k(T), P_{\widetilde{I}_k}] |\mathbf{0}\rangle + \mathrm{O}(\epsilon) \\
&= \langle \mathbf{0}| [Z_j(T), P_{\widetilde{I}_j}] |\mathbf{0}\rangle \langle \mathbf{0}| [Z_k(T), P_{\widetilde{I}_k}] |\mathbf{0}\rangle + \mathrm{O}(\epsilon) = \mathrm{O}(\epsilon),
\end{aligned} \tag{94}$$

where we have used Eq. (88) at last. This justifies the third line of Eq. (93), which leads to the last line because the number of j for each k is bounded by $\mathrm{O}(T^2)$, while the number of k is bounded by $\left|K \cap \widetilde{I}_T\right| = \mathrm{O}(LT)$.

Since there are finitely many pairs of (J, K) to choose, type (ii) terms contribute $\mathrm{O}(LT^3)$. Combining with Eq. (89), the first term in Eq. (87) is $\tilde{\theta}\left[2c_Z L^2 T^2 - \mathrm{O}\left(L^3 T\right)\right] = c_1 \tilde{\theta} L^4$ with $c_1 = \Theta(1)$ for sufficiently large L/T. Namely, the single term of type (i) dominates.

It remains to bound the second-order remainder term in Eq. (87). Using Jacobi identities and Eq. (78) as above, the three $Z(T)$'s in the commutator can all be replaced by $Z_{\widetilde{I}_T}(T) + \mathrm{O}(L^2 \epsilon)$. So the second term in Eq. (78) is bounded by

$$\mathrm{O}\left(\tilde{\theta}^3 \|Z_{\widetilde{I}_T}\|^3\right) + \mathrm{O}(L^6 \epsilon) = \mathrm{O}\left[\tilde{\theta}^3 (L + 2T)^6\right] = \mathrm{O}\left(\tilde{\theta}^3 L^6\right). \tag{95}$$

□

D States generated from quantum domino dynamics

D.1 Proof of Theorem 3: On-site measurements for quantum domino dynamics

We restate Theorem 3 in the main text here:

Theorem 7. *Consider the single-DW state $|\psi_{\rm do}\rangle$ in Eq. (19) in the main text. In a rotated on-site basis (i.e. redefining the local states $|1\rangle_i$ with an extra phase) such that* $\mathrm{Arg}(a_i) = -\frac{\pi}{2}(1+i) - 2\theta_{\rm pr} i$*, define the measurement observable*

$$\mathcal{O}_{\rm do} = \sum_{vT/2 \le i \le 2vT} |a_i| \mathbf{Y}_{1,i}, \tag{96}$$

where $\mathbf{Y}_{1,i} := \bigotimes_{j=1}^{i} Y_j$*. For any $v, T > 0$, if*

$$\sum_{vT/2 \le i \le 2vT} |a_i|^2 \ge c_{\rm do} = \Theta(1), \tag{97}$$

then

$$\frac{1}{\langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta}} \left(\partial_\theta \langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta} \right)^2 = \Theta(T^2), \tag{98}$$

for $\rho_\theta = \mathcal{D}_{\theta,p}^{\otimes N}(|\psi_{\rm do}\rangle\langle\psi_{\rm do}|)$ under noise $p \le 1/(vT)$.

Proof. We first focus on the noiseless case. Direct computation yields expectation values

$$\langle \mathbf{Y}_{1,i} \rangle_{\rho_\theta^{(0)}} = 0, \quad \partial_\theta \langle \mathbf{Y}_{1,i} \rangle_{\rho_\theta^{(0)}} = 2|a_i| i, \tag{99}$$

where we have used the redefined local states $|1\rangle_i$ to make sure all $\partial_\theta \langle \mathbf{Y}_{1,i} \rangle_{\rho_\theta^{(0)}}$ are at its maximum. Note that i is the real number for qubit index, not the imaginary unit i. Although each $\mathbf{Y}_{1,i}$ does not have HL sensitivity because $|a_i| \ll 1$ and $\left\langle \Delta \mathbf{Y}_{1,i}^2 \right\rangle_{\rho_\theta^{(0)}} = \left\langle \mathbf{Y}_{1,i}^2 \right\rangle_{\rho_\theta^{(0)}} - 0 = 1$, a linear combination of them like Eq. (96) does. Here we assume a more general form $\mathcal{O}_{\rm do} = \sum_{i=1}^{N} b_i \mathbf{Y}_{1,i}$ for latter convenience. Eq. (99) then yields $\langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta^{(0)}} = 0$ and

$$\partial_\theta \langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta^{(0)}} = 2 \sum_i b_i |a_i| i \ge vT \sum_{vT/2 \le i \le 2vT} b_i |a_i|. \tag{100}$$

We then want to show the variance of $\mathcal{O}_{\rm do}$ is small in the state:

$$\begin{aligned} \left\langle \Delta \mathcal{O}_{\rm do}^2 \right\rangle_{\rho_\theta^{(0)}} = \left\langle \mathcal{O}_{\rm do}^2 \right\rangle_{\rho_\theta^{(0)}} &= \sum_i b_i^2 + 2 \sum_{1 \le i < j \le N} b_i b_j \langle \mathbf{Y}_{i+1,j} \rangle_{\rho_\theta^{(0)}} \\ &\le \sum_i b_i^2 + 2 \sum_{1 \le i < j \le N} b_i b_j |a_i||a_j| \\ &\le \sum_i b_i^2 + \left(\sum_i b_i |a_i| \right)^2. \end{aligned} \tag{101}$$

Here $\mathbf{Y}_{i,j} := \bigotimes_{k=i}^{j} Y_k$, and its expectation value is only contributed by $|1_i\rangle$ and $|1_j\rangle$, which leads to the second line. Finally, choosing $b_i = |a_i|$, $\langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta^{(0)}} \le 1 + 1 = 2$ due to normalization, and

$$\partial_\theta \langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta^{(0)}} \ge c_{\rm do} vT, \tag{102}$$

according to Eq. (100) and our assumption Eq. (97). This leads to Eq. (98) in the noiseless case.

With noise, the (dual map of) dephasing channel acting on observable Eq. (96) retains its form but with coefficients $b_i = |a_i|(1-2p)^i$. Then we can still use Eq. (100) and Eq. (101) with these coefficients, yielding

$$\partial_\theta \langle \mathcal{O}_{\rm do} \rangle_{\rho_\theta} \ge c_{\rm do} vT (1-2p)^{2vT}, \tag{103}$$

and $\langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta} \le \langle \Delta \mathcal{O}_{\rm do}^2 \rangle_{\rho_\theta^{(0)}} \le 2$. This establishes Eq. (98) for noise $p < 1/(vT)$ because $(1-2p)^{2vT} \ge (1-2p)^{2/p} \ge \mathrm{e}^{-4}$ for $p < 1/2$. □

If $T \ll N$, Eq. (97) can be verified by the exact solution $a_i \approx (-\mathrm{i})^{i-1}(i/t)J_i(vT)$ with $v = 2$ [72] where J_i is the Bessel function of the first kind. Here the exact solution comes from the fact that the restricted Hamiltonian in the 1-DW subspace is the solvable matrix

$$H_{ij}^{(1)} = \begin{cases} 1, & |i-j| = 1 \\ 0, & \text{otherwise} \end{cases} \tag{104}$$

In practice, Eq. (97) holds all the way up to $vT \approx N$ from numerics, when the wavefront bounces off the right boundary.

Note that [62] proposes another measurement scheme: after sensing, one further evolve the system under $H_{\rm do}$ for some time before the final on-site measurement, so that the DW bounces off the right boundary and approximately returns to the initial site 1. However, this is not parametrically optimal, because the wave packet actually *spreads* as Bessel function $a_i \propto (i/t)J_i(vT) = \mathrm{O}(T^{-1/3}) \ll 1, \forall i$, yielding a small final signal $\langle Y_1 \rangle_{\rm bounce} = \Theta(N^{-1/3})$ and $\partial_\theta \langle Y_1 \rangle_{\rm bounce} = \Theta(N^{2/3})$ that leads to $\delta\theta = \Theta(N^{-2/3})$ Here Y_1 is the observable on the first site in the final state after bouncing off.

D.2 Partitioned domino state

For partitioned domino state $|\Psi_{\rm do}\rangle$ (Eq. (20) in the main text), one wants to use

$$T \approx L/(2v), \tag{105}$$

with the DW velocity $v = 2$ so that DWs of each group propagate to the edges. To simplify the situation, we choose T to be slightly smaller than Eq. (105), to make sure the leakage of each DW to other groups is small, so that the groups evolve almost independently. For example, for Fig. 3 in the main text, we choose

$$T = 1.3, \quad 3.7, \quad 8.7, \quad 23.9, \quad 49, \tag{106}$$

for the five values of $L = 10, 20, 40, 100, 200$. In the numerical simulation, we simulate individual states $|\Psi_{\rm simu}\rangle$ generated from single $|+\rangle$'s and approximate $|\Psi_{\rm do}\rangle$ using tensor product of these states. We estimate that such approximation of independent dynamics causes small error $\left\| |\Psi_{\rm do}\rangle - |\Psi_{\rm simu}\rangle^{\otimes N/L} \right\|^2 \lessapprox 2\%$ for the choices Eq. (106).

Based on the T chosen above, $|\Psi_{\rm do}\rangle$ is approximately N/L copies of the L-qubit state

$$\frac{1}{\sqrt{2}} |\mathbf{0}\rangle + \frac{1}{\sqrt{2}} \sum_{-L/2+1 \le i,j \le L/2} a_{i,j} |1_{i,j}\rangle , \tag{107}$$

which is evolved from a single $|+\rangle$ at site 1. Here $|1_{i,j}\rangle$ is all-1 from i to j (including these two sites), and all-0 elsewhere. Since the Hamiltonian $H_{\rm do}$ is real, $a_{i,j}$ is either real for $|i-j|$ being even, or purely imaginary for odd $|i-j|$. This structure simplifies the observable for sensing $\theta \approx 0$: one just measures in the y on-site basis, because for GHZ metrology $|\widetilde{\rm GHZ}\rangle = \mathrm{e}^{-\mathrm{i}\theta \sum_{i=1}^L Z_i} \frac{1}{\sqrt{2}} \left(|0^{\otimes L}\rangle \pm \mathrm{i}^{L+1} |1^{\otimes L}\rangle \right)$ at $\theta = 0$,

$$\partial_\theta \langle \mathbf{Y}_{1,L} \rangle_{\widetilde{\rm GHZ}} = \pm L \left(\mathrm{i} \langle \mathbf{0}| \mathbf{Y}_{1,L} \mathrm{i}^{L+1} |\mathbf{1}\rangle - \mathrm{i} \langle \mathbf{1}| (-\mathrm{i})^{L+1} \mathbf{Y}_{1,L} |\mathbf{0}\rangle \right) = \mp 2L, \tag{108}$$

is an optimal observable for any L. Note that the following still applies to cases beyond $\theta \approx 0$, where one may want to optimize numerically over the local phases ϕ_i of local observables $Y_i \to \cos(\phi_i)X_i + \sin(\phi_i)Y_i$.

For sensing $\theta \approx 0$ with N/L copies of Eq. (107), we choose an observable

$$\mathcal{O}'_{\rm do} = \sum_{-L/2+1\le i\le 1, 1\le j\le L/2} b_{i,j}\mathbf{Y}_{i,j}, \tag{109}$$

for each L-qubit group with $b_{i,j} = |a_{i,j}|$, and the total observable would be a sum of $\mathcal{O}'_{\rm do}$ over groups, just like Theorem 1. This is equivalent to Eq. (24) in the main text, where $b_{i,j} = 0$ for (i,j) that does not appear in the sum Eq. (109). The idea of Eq. (109) is similar to Theorem 3 in the main text, where we want the signal to be contributed from terms like $\langle \mathbf{0}| \mathbf{Y}_{i,j} |1_{i,j}\rangle$. Note that one can also allow for terms with $i > 1$ in Eq. (109); our choice simplifies the calculation a bit. Still, it is tedious to perform a fully analytical treatment like Theorem 3 in the main text, so we numerically calculate the signal and variance of Eq. (109), as shown in Fig. 3 in the main text. The results are as expected.

E Generalizations

E.1 General QFI upper bound

Here we derive a general QFI upper bound of $\Theta(N/p)$ for any multi-probe systems where each probe is independently and identically subject to a quantum signal and a quantum noise with noise probability p. Specifically, we consider the following two situations:

(1) The noise quantum channel $\mathcal{N}_\theta$[5] acting on each probe describes the evolution of an open quantum system with a Hamiltonian $H_{\rm p}$

$$\mathcal{N}_\theta = \exp(\theta\mathscr{H}t + \gamma\mathscr{L}t), \tag{110}$$

where t is the evolution time [6], $\mathscr{H}(\rho) = -\mathrm{i}[H_{\rm p}, \rho]$ describes the Hamiltonian evolution, and the noise is described by

$$\mathscr{L}(\rho) = \sum_{i=1}^{k} L_i\rho L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho\}. \tag{111}$$

θ and γ represent the strengths of the signal and the noise, respectively, where θ is the unknown parameter to be detected. We consider the regime where $\gamma \ll 1$ and $t = 1$. The noise probability p is defined to be

$$p = 1 - \mathrm{e}^{-\gamma t}, \tag{112}$$

[5]Note that we use notations $\mathcal{N}_\theta, \mathcal{N}_\theta^{(0)}$ instead of $\mathcal{N}_{\theta,p}, \mathcal{N}_{\theta,0}$ in the main text, because for a general channel there is no unique way to define the error probability p. We will define p by a convenient choice shortly.

[6]Note that t is different than the time T (T') to prepare locally generated resource states in Section C. Nevertheless, in order to achieve optimal noisy metrology, one needs to choose T such that $\gamma t = \Theta(T^{-d})$ according to Theorem 5. Below we set $t = 1$ without loss of generality by rescaling the noise rate γ.

and we have $p = \Theta(\gamma)$. Viewing θ/γ as the parameter to be estimated, we can use results from previous works [29, 28, 73] to obtain

$$\begin{aligned}\mathcal{F}(\mathcal{N}_\theta^{\otimes N}) &\leq \left(\frac{1}{\gamma^2}\right) \cdot 4 \min_{\substack{\forall h,\mathbf{h},\mathfrak{h},\ s.t.,\\ H+h_{00}\mathrm{id}+\mathbf{h}^\dagger\mathbf{L}+\mathbf{h}\mathbf{L}^\dagger+\mathbf{L}^\dagger\mathfrak{h}\mathbf{L}=0}} \left\|(\mathbf{h}\mathrm{id}+\mathfrak{h}\mathbf{L})^\dagger(\mathbf{h}\mathrm{id}+\mathfrak{h}\mathbf{L})\right\| \cdot (N\gamma t) \\ &= \mathrm{O}\left(\frac{N}{\gamma}\right) = \mathrm{O}\left(\frac{N}{p}\right), \end{aligned} \tag{113}$$

where $\mathbf{L} = \begin{pmatrix} L_1 \\ \vdots \\ L_r \end{pmatrix}$, $h_{00} \in \mathbb{R}$, $\mathbf{h} \in \mathbb{C}^r$, $\mathfrak{h} \in \mathbb{C}^{r\times r}$ and is Hermitian. Here $\mathcal{F}(\mathcal{N}_\theta^{\otimes N})$ is the channel QFI defined by $\mathcal{F}(\mathcal{N}_\theta^{\otimes N})(\rho)$ maximized over all input states ρ. The only assumption is the channel violates the so-called "Hamiltonian-not-in-Lindblad-span" condition, i.e., $H \in \mathrm{span}\{\mathrm{id}, L_i, L_i^\dagger, L_i^\dagger L_j, \forall i,j\}$. The condition is violated for, e.g., Pauli-Z signal and dephasing noise, as discussed in the main text. This situation also encompasses many commonly encountered noise types, e.g., depolarizing noise, amplitude damping noise, erasure noise, etc. When the condition holds, the HL is in principle achievable using techniques like quantum error correction [29], which is beyond the scope of our work.

(2) Another broadly discussed scenario is the following, where the evolution of each probe is described by

$$\mathcal{N}_\theta = \exp(\gamma\mathscr{L}t)\exp(\theta\mathscr{H}t), \tag{114}$$

where $\mathscr{L}$ and $\mathscr{H}$ are defined as above, which is the composition of a Hamiltonian evolution and a noise channel afterwards. Note that situation (2) is the same as situation (1) when $\mathscr{H}$ and $\mathscr{L}$ commutes, e.g., Pauli-Z signal with depolarizing noise. The assumption that the noise acts after the Hamiltonian evolution is less commonly encountered in practice, but can be much simpler to deal with in theory when the noise channel $\exp(\gamma\mathscr{L}t)$ is analytically solvable, see, e.g. [24]. Again, we consider the limiting case where $\gamma \ll 1$ and t is a constant and p is the noise probability that satisfies $p = \Theta(\gamma)$. Note that here the noise channel $\mathcal{N} = \exp(\gamma\mathscr{L}t)$ can represent any CPTP map, i.e., any noisy quantum channel, and we further require the "Hamiltonian-not-in-Kraus-span" condition to be violated, i.e, $H \in \mathrm{span}\{\mathrm{id}, K_i^\dagger K_j, \forall i,j\}$. The assumption again holds in most noisy quantum systems with e.g., depolarizing noise, amplitude damping noise, etc. We assume $\mathcal{N}(\cdot) = \sum_{i=0}^r K_i(\cdot)K_i^\dagger$, where K_i are Kraus operators that we specify later. From previous works, the QFI is upper bounded by [24, 31]

$$\mathcal{F}(\mathcal{N}_\theta^{\otimes N}) \leq 4 \min_{\forall h,\ s.t., Ht+\mathbf{K}^\dagger h\mathbf{K}=0} \left\|\mathbf{K}^\dagger h^2\mathbf{K} - H^2t^2\right\| N, \tag{115}$$

where $\mathbf{K} = \begin{pmatrix} K_0 \\ \vdots \\ K_r \end{pmatrix}$, $h \in \mathbb{C}^{(r+1)\times(r+1)}$ is Hermitian. Expanding $\mathbf{K}$ and γh in Taylor expansions of γt, we have

$$K_0 = \mathrm{id} - \frac{\gamma t}{2}\sum_{i=1}^r L_i^\dagger L_i + \mathrm{O}((\gamma t)^{3/2}), \tag{116}$$

$$K_i = \sqrt{\gamma t}L_i + \mathrm{O}(\gamma t), \quad i = 1,\ldots,r, \tag{117}$$

$$\gamma h = \begin{pmatrix} h_{00}\gamma t + \mathrm{O}((\gamma t)^{3/2}) & \mathbf{h}^\dagger(\gamma t)^{1/2} + \mathrm{O}(\gamma t) \\ \mathbf{h}(\gamma t)^{1/2} + \mathrm{O}(\gamma t) & \mathfrak{h} + \mathrm{O}((\gamma t)^{1/2}) \end{pmatrix}, \tag{118}$$

such that $Ht + \mathbf{K}^\dagger h \mathbf{K} = 0$ translates to

$$H + h_{00}\mathrm{id} + \mathbf{h}^\dagger \mathbf{L} + \mathbf{L}^\dagger \mathbf{h} + \mathfrak{h}\mathbf{L}^\dagger \mathbf{L} = 0, \tag{119}$$

where $h_{00} \in \mathbb{R}$, $\mathbf{h} \in \mathbb{C}^r$, $\mathfrak{h} \in \mathbb{C}^{r\times r}$ and is Hermitian. Then have

$$\begin{aligned}&\min_{\forall h,\ s.t., Ht+\mathbf{K}^\dagger h\mathbf{K}=0} \left\|\mathbf{K}^\dagger h^2 \mathbf{K} - H^2 t^2\right\| \leq \\ &\frac{1}{\gamma^2}\left((\gamma t) \min_{\substack{\forall h,\mathbf{h},\mathfrak{h},\ s.t.,\\ H+h_{00}\mathrm{id}+\mathbf{h}^\dagger\mathbf{L}+\mathbf{h}\mathbf{L}^\dagger+\mathbf{L}^\dagger\mathfrak{h}\mathbf{L}=0}} \left\|(\mathbf{h}\mathrm{id} + \mathfrak{h}\mathbf{L})^\dagger(\mathbf{h}\mathrm{id} + \mathfrak{h}\mathbf{L})\right\| + \mathrm{O}((\gamma t)^{3/2})\right)\end{aligned} \tag{120}$$

where we ignored the H^2t^2 term because $\|A - B\| \leq \|A\|$ when $A - B$ and B are arbitrary positive semidefinite matrices. It implies

$$\mathcal{F}(\mathcal{N}_\theta^{\otimes N}) \leq \frac{Nt}{\gamma} \cdot \text{constant} + \mathrm{O}\left(\frac{N}{\gamma^{1/2}}\right) = \mathrm{O}\left(\frac{N}{p}\right), \tag{121}$$

where the constant above is the same as the one appearing in Eq. (113).

As discussed above, for general noise, including Eq. (110) and Eq. (114), we always have a QFI upper bound of $\mathrm{O}(N/p)$. We show next how to generalize our theorems to find metrologically optimal states achieving this scaling. We will use the following two properties

$$\left\|\partial_\theta \mathcal{N}_\theta - \partial_\theta \mathcal{N}_\theta^{(0)}\right\|_\diamond = \mathrm{O}(p), \quad \left\|\mathcal{N}_\theta - \mathcal{N}_\theta^{(0)}\right\|_\diamond = \mathrm{O}(p), \tag{122}$$

where $^{(0)}$ represents the noiseless channel when taking $\gamma = 0$ (i.e., $\mathcal{N}_\theta^{(0)} = \exp(\theta\mathscr{H}t)$), $\|\cdot\|_\diamond$ is the diamond norm, and Eq. (122) is satisfied for both Eq. (110) and Eq. (114). Here the correctness of Eq. (122) can be seen from the two expressions of the noisy channels (Eq. (110) and Eq. (114)), $\mathcal{N}_\theta^{(0)} = \mathcal{N}_\theta|_{\gamma=0}$ and the fact that $\gamma = \Theta(p)$.

One example of the above formalism of noise channel is the amplitude damping noise (by default we set the evolution time $t = 1$). We assume

$$\mathcal{N}_\theta = \exp(\gamma\mathscr{L})\exp(\theta\mathscr{H}), \tag{123}$$

$$\mathscr{L}(\cdot) = \gamma(|0\rangle\langle 1|(\cdot)|1\rangle\langle 0| - \frac{1}{2}\{|1\rangle\langle 1|, (\cdot)\}). \tag{124}$$

Or equivalently,

$$\mathcal{N}_\theta(\cdot) = K_1 e^{-iH_\mathrm{p}\theta}(\cdot)e^{iH_\mathrm{p}\theta}K_1^\dagger + K_2 e^{-iH_\mathrm{p}\theta}(\cdot)e^{iH_\mathrm{p}\theta}K_2^\dagger, \tag{125}$$

$$K_1 = |0\rangle\langle 0| + \sqrt{1-p}\,|1\rangle\langle 1|, \quad K_2 = \sqrt{p}\,|0\rangle\langle 1|. \tag{126}$$

where $p = 1 - e^{-\gamma}$. In the noise regime we consider, i.e., $1/N < p < 1/2$, we have $\gamma/2 \leq p \leq \gamma$ and clearly $p = \Theta(\gamma)$ as required.

E.2 General optimality of states constructed in the main text

The mapping from any qudit system to a qubit system is straightforward. To estimate θ in Eq. (110) or Eq. (114), we can consider the two-dimensional subspace spanned by two eigenstates corresponding to the highest and the lowest eigenvalues of H and denote them by $|0\rangle$ and $|1\rangle$. We further assume, without loss of generality, the highest and lowest eigenvalues are 1 and -1 and $t = 1$. Then any multi-qubit state described in the main text for sensing under dephasing noise naturally represents a multi-probe state in the general context here (Eq. (110) or Eq. (114)). We show the state optimality below.

Theorem 8. *Theorem 1,5,7 continue to hold for the generalized situations above.*

Proof. We first deal with Theorem 5 that is conceptually simpler. In its proof, the only step related to the noise channel is the final paragraph. Since the variance bound still holds for general single-qubit noise that cannot create long-range correlation, we only need to modify the derivative bound Eq. (86). In a nutshell, Eq. (86) generalizes because it only uses the fact that $\mathcal{N}_\theta$ is p-close to $\mathcal{N}_\theta^{(0)}$ in the sense of Eq. (122), not the specific form of $\mathcal{N}$.

More precisely, below we use $\mathcal{N}_\theta^\dagger, \mathcal{N}_\theta^{(0)\dagger}$ to represent the dual maps of $\mathcal{N}_\theta, \mathcal{N}_\theta^{(0)}$. Then the p-closeness Eq. (122) for the original channels implies

$$\left\|(\partial_\theta \mathcal{N}_\theta^\dagger \otimes \mathrm{id})(\mathcal{O}) - (\partial_\theta \mathcal{N}_\theta^{(0)\dagger} \otimes \mathrm{id})(\mathcal{O})\right\| \le \mathrm{O}(p)\,\|\mathcal{O}\|\,, \quad \left\|(\mathcal{N}_\theta^\dagger \otimes \mathrm{id})(\mathcal{O}) - (\mathcal{N}_\theta^{(0)\dagger} \otimes \mathrm{id})(\mathcal{O})\right\| \le \mathrm{O}(p)\,\|\mathcal{O}\|\,, \tag{127}$$

where id is the identity channel acting on ancillary degrees of freedom, and $\mathcal{O}$ is any operator acting on the whole system.

We can modify Eq. (86) as follows (we ignore the tensor product with id operators below for simplicity),

$$\begin{aligned}
&\left|\partial_\theta\left(\left\langle \widetilde{P}_I\right\rangle_{\rho_\theta} - \left\langle \widetilde{P}_I\right\rangle_{\rho_\theta^{(0)}}\right)\right| = \left|\partial_\theta\left\langle \mathcal{N}_\theta^{\dagger\otimes N}(\widetilde{P}_I) - \mathcal{N}_\theta^{\dagger(0)\otimes N}(\widetilde{P}_I)\right\rangle_\rho\right| \\
&= \left|\sum_i \left\langle (\partial_\theta \mathcal{N}_\theta^\dagger)_i \otimes \mathcal{N}_\theta^{\dagger\otimes N-1}(\widetilde{P}_I) - (\partial_\theta \mathcal{N}_\theta^{(0)\dagger})_i \otimes \mathcal{N}_\theta^{(0)\dagger\otimes N-1}(\widetilde{P}_I)\right\rangle_\rho\right| \\
&\le \sum_{i\in\widetilde{I}_T}\left|\left\langle (\partial_\theta \mathcal{N}_\theta^\dagger)_i \otimes \mathcal{N}_\theta^{\dagger\otimes M'-1}(\widetilde{P}_I) - (\partial_\theta \mathcal{N}_\theta^{(0)\dagger})_i \otimes \mathcal{N}_\theta^{(0)\dagger\otimes M'-1}(\widetilde{P}_I)\right\rangle_\rho\right| + \mathrm{O}(L^2\epsilon) \\
&\le M' \max_i \left\|(\partial_\theta \mathcal{N}_\theta^\dagger)_i \otimes \mathcal{N}_\theta^{\dagger\otimes M'-1}(\widetilde{P}_I) - (\partial_\theta \mathcal{N}_\theta^{(0)\dagger})_i \otimes \mathcal{N}_\theta^{\dagger\otimes M'-1}(\widetilde{P}_I)\right\| \\
&\quad + M' \max_i \sum_{g=1}^{M'-1} \left\|(\partial_\theta \mathcal{N}_\theta^{(0)\dagger})_i \otimes \left[\mathcal{N}_\theta^{\dagger\otimes g} \otimes \mathcal{N}_\theta^{(0)\dagger\otimes M'-1-g}(\widetilde{P}_I) - \mathcal{N}_\theta^{\dagger\otimes g-1} \otimes \mathcal{N}_\theta^{(0)\dagger\otimes M'-g}(\widetilde{P}_I)\right]\right\| + \mathrm{O}(L^2\epsilon) \\
&\le M' \cdot M'\mathrm{O}(p) + \mathrm{O}(L^2\epsilon) \le \Theta(L^4)p + \mathrm{O}(L^2\epsilon).
\end{aligned} \tag{128}$$

Here in the third line, we have again used that $\widetilde{P}_I$ is almost supported in region $\widetilde{I}_T$ with size $M' = \Theta(L^2)$: For a faraway site i, $(\partial_\theta \mathcal{N}_\theta)_i$ acts on identity operator and becomes zero. To get the last line, we have used Eq. (127) with the triangle inequality. The generalized Theorem 5 is then proven.

Theorem 7 generalizes in a similar way.

It remains to generalize Theorem 1. In its proof, Eq. (38) still holds because Eq. (41) only uses the continuity of QFI and the p-closeness Eq. (122) for $\mathcal{N}_\theta = \mathcal{D}_{\theta,p}$ and $\mathcal{N}_\theta^{(0)} = \mathcal{D}_{\theta,0}$. Then we still use Lemma 2 to choose an observable $\mathcal{O}_I$ for each group. Choose $\sigma_\theta = \rho_{I,\theta}$ and $A = H_I := \sum_{i\in I} H_i$ in Lemma 2, where H_i represents the signal Hamiltonian on each probe, it was shown in the proof of Theorem 1 that when Eq. (47) holds (which we will prove below), there exists $\mathcal{O}_I$ that satisfies Eq. (48) and Eq. (49), and the rest of the proof also applies to general signal and noise.

To prove Eq. (47) for general signal and noise, we note by direct calculation,

$$\begin{aligned}
B &= \partial_\theta \rho_{I,\theta} + \mathrm{i}[H_I, \rho_{I,\theta}] = \partial_\theta\left(\mathcal{N}_\theta^{\otimes M}(\rho_I)\right) + \mathrm{i}[H_I, \mathcal{N}_\theta^{\otimes M}(\rho_I)] \\
&= \sum_{i\in I} \mathcal{N}_\theta^{\otimes M-1} \otimes (\partial_\theta \mathcal{N}_\theta)_i(\rho_I) + \mathrm{i}[H_i, \mathcal{N}_\theta^{\otimes M}(\rho_I)] = \sum_{i\in I} \mathcal{N}_\theta^{\otimes M-1}\left((\partial_\theta \mathcal{N}_\theta)_i(\rho_I) + \mathrm{i}[H_i, (\mathcal{N}_\theta)_i(\rho_I)]\right),
\end{aligned} \tag{129}$$

where we have used the Leibniz product rule to get the second line, and that H_i commutes with channels acting on sites other than i in the end. Taking trace norm, we have

$$\begin{aligned}
&\|B\|_1 \le M \max_i \|(\partial_\theta \mathcal{N}_\theta)_i(\rho_I) + \mathrm{i}[H_i, (\mathcal{N}_\theta)_i(\rho_I)]\|_1 \\
&= M \max_i \left\| \left(\partial_\theta \mathcal{N}_\theta - \partial_\theta \mathcal{N}_\theta^{(0)}\right)_i (\rho_I) + \left(\partial_\theta \mathcal{N}_\theta^{(0)}\right)_i (\rho_I) + \mathrm{i}\left[H_i, \left(\mathcal{N}_\theta^{(0)}\right)_i (\rho_I)\right] + \mathrm{i}\left[H_i, \left(\mathcal{N}_\theta - \mathcal{N}_\theta^{(0)}\right)_i (\rho_I)\right] \right\|_1 \\
&= M \max_i \left\| \left(\partial_\theta \mathcal{N}_\theta - \partial_\theta \mathcal{N}_\theta^{(0)}\right)_i (\rho_I) + \mathrm{i}\left[H_i, \left(\mathcal{N}_\theta - \mathcal{N}_\theta^{(0)}\right)_i (\rho_I)\right] \right\|_1 = M \cdot \mathrm{O}(p). \qquad (130)
\end{aligned}$$

Here the middle two terms in the second line cancels because $\left(\partial_\theta \mathcal{N}_\theta^{(0)}\right)_i \rho_I = -\mathrm{i}\left[H_i, \left(\mathcal{N}_\theta^{(0)}\right)_i \rho_I\right]$, and we have used Eq. (122) together with $\|H_i\| = \mathrm{O}(1)$ in the end. This proves Eq. (47) and generalizes Theorem 1.

□

We expect our results generalize even beyond independent signal and noise on each probe to include signal and noise that might be correlated within each group of probes, because, for example, the generalized Theorem 1 only relies on the continuity of QFI and the existence of an observable in each group saturating its QFI.

In addition, our protocol may also tolerate noise in state preparation and measurement when the noise probability per group $p_{\text{per group}}$ is smaller than some constant $c_{\text{per group}}$: Take Theorem 5 with measurement noise for example. Suppose the measurement noise is characterized by a channel $\mathcal{E}$ prior to the ideal measurement. The ideal measurement observable $\widetilde{P}_I$ for each group becomes $\mathcal{E}^\dagger\left(\widetilde{P}_I\right)$. Let η represents the error bound Eq. (128) for replacing $\rho_\theta \to \rho_\theta^{(0)}$, the additional error caused by $\mathcal{E}$ is

$$\begin{aligned}
\left|\partial_\theta \left\langle \mathcal{E}^\dagger\left(\widetilde{P}_I\right) - \widetilde{P}_I \right\rangle_{\rho_\theta}\right| &\le \left|\partial_\theta \left\langle \mathcal{E}^\dagger\left(\widetilde{P}_I\right) - \widetilde{P}_I \right\rangle_{\rho_\theta^{(0)}}\right| + 2\eta = \left|\left\langle \left[H, \mathcal{E}^\dagger\left(\widetilde{P}_I\right) - \widetilde{P}_I\right] \right\rangle_{\rho_\theta^{(0)}}\right| + 2\eta \\
&\le 2M' \left\|\mathcal{E}^\dagger\left(\widetilde{P}_I\right) - \widetilde{P}_I\right\| + 3\eta \le 2M' \times p_{\text{per group}} + 3\eta, \qquad (131)
\end{aligned}$$

where $p_{\text{per group}}$ is the diamond-norm distance between $\mathcal{E}^\dagger$ and the identity channel. Here we changed $2\eta \to 3\eta$ to account for the error restricting $H := \sum_i H_i$ to $H_{\widetilde{I}_T}$. Eq. (131) is then bounded by a constant times M', which is exactly the same as Eq. (128), so the presence of $\mathcal{E}$ does not change the scaling of the final precision (because obviously the variance bound Eq. (85) is also robust). Note that for this generalization to hold, we need the observable to have a small operator norm $\left\|\widetilde{P}_I\right\| = 1$ comparing to $\partial_\theta \left\langle \widetilde{P}_I \right\rangle = \Theta(M)$; that does not hold in the case of Theorem 1 (see Eq. (42) and Eq. (44)).

F Spin-squeezed states as metrological optimal states

In this work we have mainly focused on optimal states given by Theorem 1 that have large intra-group and small inter-group correlations. However, there can be other classes of optimal states with different correlation structure. [31, 47, 48] studied a particular spin-squeezed state (SSS) sensing the channel Eq. (1) in the main text, and found optimal performance even if every pair of qubits shares $\Theta(1)$ correlation. Below we prove that very general SSS are optimal for sensing magnetic field under any general single-qubit noise.

Theorem 9 (Restatement of Theorem 4 in the main text)**.** *Suppose ρ is spin-squeezed in the sense that*

$$\langle X\rangle_\rho = \Theta(N), \tag{132}$$

$$\left\langle \Delta Y^2 \right\rangle_\rho = N\xi^2, \tag{133}$$

$$\left\langle \Delta S^2 \right\rangle_\rho = \mathrm{O}(N/\xi^2), \quad \forall S := \sum_i S_i \text{ with } \|S_i\| = 1 \tag{134}$$

where $X = \sum_i X_i$ etc. Each qubit undergoes a channel $\mathcal{N}_\theta$ satisfying Eq. (122) with $\mathcal{N}_\theta^{(0)}(\cdot) = \mathrm{e}^{-\mathrm{i}\theta Z}(\cdot)\mathrm{e}^{\mathrm{i}\theta Z}$ for some $\xi < 1$. Then if $p = \mathrm{O}(\xi^2)$, the final state ρ_θ has the parametrically the same metrological power as its noiseless limit at $\theta = 0$, because

$$\mathcal{F}(\rho_\theta) \geq \frac{1}{\langle \Delta Y^2\rangle_{\rho_\theta}} \left(\partial_\theta \langle Y\rangle_{\rho_\theta}\right)^2 \Big|_{\theta=0} = \Theta(N/\xi^2). \tag{135}$$

Eq. (135) then saturates the optimal bound $\Theta(N/p)$ at $p = \Theta(\xi^2)$. Furthermore, Y is an optimal observable here, indicating the existence of an optimal on-site measurement strategy. Note that here we must have $\xi = \Omega(1/\sqrt{N})$, because otherwise Eq. (132) and Eq. (133) cannot be simultaneously satisfied due to the uncertainty relation. Here we focus on sensing at $\theta = 0$ without loss of generality, because other values θ can be dealt with similarly by rotating the final observable $Y \to Y\cos(2\theta) + X\sin(2\theta)$. The conditions Eq. (132) and Eq. (133) are standard for SSS, which simply mean large polarization along x and small fluctuation in polarization along y. Since $\langle \Delta Y^2\rangle = N$ for a product state along x-direction, the $\xi^2 \ll 1$ is roughly the squeezing parameter. Taking $S = Z$ for example, the third condition Eq. (134) requires that fluctuation in other directions cannot be much larger than what is predicted from Eq. (133) and the saturation of the uncertainty relation $\langle \Delta Y^2\rangle \langle \Delta Z^2\rangle \geq |\langle X\rangle|^2 = \Omega(N^2)$.

Proof. Acting the dual maps $\mathcal{N}_\theta^\dagger$ on the observable, we have

$$\langle Y\rangle_{\rho_\theta} = \sum_i \left\langle \mathcal{N}_\theta^\dagger(Y_i)\right\rangle_\rho = \sum_i \left\langle \mathcal{N}_\theta^{(0)\dagger}(Y_i)\right\rangle_\rho + \left\langle (\mathcal{N}_\theta^\dagger - \mathcal{N}_\theta^{(0)\dagger})(Y_i)\right\rangle_\rho = \sum_i \left\langle \mathrm{e}^{\mathrm{i}\theta Z_i} Y_i \mathrm{e}^{-\mathrm{i}\theta Z_i}\right\rangle_\rho + \mathrm{O}(pN), \tag{136}$$

$$\partial_\theta \langle Y\rangle_{\rho_\theta}|_{\theta=0} = 2\sum_i \langle X_i\rangle_\rho + \mathrm{O}(pN) = \Theta(N), \tag{137}$$

where we have used Eq. (127), and plugged in Eq. (132) in the second line. Similarly,

$$\begin{aligned}\left\langle Y^2\right\rangle_{\rho_\theta}|_{\theta=0} &= N + \sum_{i\neq j}\left\langle \mathcal{N}_0^\dagger(Y_i)\otimes\mathcal{N}_0^\dagger(Y_j)\right\rangle_\rho = N + \left\langle \left(\sum_i \mathcal{N}_0^\dagger(Y_i)\right)^2\right\rangle_\rho - \sum_i \left\langle \left(\mathcal{N}_0^\dagger(Y_i)\right)^2\right\rangle_\rho \\ &= \left\langle \left[\sum_i (Y_i + \mathrm{O}(p)S_i)\right]^2\right\rangle_\rho + \sum_i \left\langle \mathrm{id} - (Y_i + \mathrm{O}(p)S_i)^2\right\rangle_\rho = \left\langle \left[\sum_i (Y_i + \mathrm{O}(p)S_i)\right]^2\right\rangle_\rho + \mathrm{O}(pN),\end{aligned} \tag{138}$$

where we have introduced an operator $S_i \propto \mathcal{N}_0^\dagger(Y_i) - Y_i$ on spin i with $\|S_i\| = 1$, and used

$\mathrm{id} - Y_i^2 = 0$. Combining Eq. (136) (the second term below) and Eq. (138),

$$\begin{aligned}\left\langle \Delta Y^2 \right\rangle_{\rho_\theta}|_{\theta=0} &= \left\langle \left[\sum_i (Y_i + \mathrm{O}(p) S_i)\right]^2 \right\rangle_\rho - \left[\left\langle \sum_i (Y_i + \mathrm{O}(p) S_i)\right\rangle_\rho\right]^2 + \mathrm{O}(pN) \\ &= \left\langle \Delta Y^2 \right\rangle_\rho + \mathrm{O}(p)\left[C_\rho(Y,S) + C_\rho(S,Y)\right] + \mathrm{O}(p^2)\left\langle \Delta S^2 \right\rangle_\rho + \mathrm{O}(pN) \\ &= N\xi^2 + \mathrm{O}(p\sqrt{\langle \Delta Y^2\rangle \langle \Delta S^2\rangle}) + \mathrm{O}(p^2 N/\xi^2) + \mathrm{O}(pN) = N\xi^2 + \mathrm{O}(N(p + p^2/\xi^2)) \\ &= \Theta(N\xi^2), \quad \text{if} \quad p = \mathrm{O}(\xi^2), \end{aligned} \tag{139}$$

where we have used Cauchy-Schwarz inequality and Eq. (134) in the third line. Combining Eq. (137) with Eq. (139) yields Eq. (135).

□

One example of optimal SSS's that saturate the conditions in Theorem 9 are SSS's generated by one-axis twisting [31, 47, 45, 48],

$$|\psi_{\mu,\nu}\rangle = \exp(-i\nu X/2)\exp(-i\mu Z^2/8)\exp(-i\pi Y/4)\,|\mathbf{0}\rangle\,, \tag{140}$$

and taking $\xi = \sqrt{p}$ and

$$\mu = \frac{1}{N\xi}, \quad \nu = \frac{\pi}{2} - \frac{1}{2}\arctan\frac{4\sin(\mu/2)\cos^{N-2}(\mu/2)}{1-\cos^{N-2}\mu}. \tag{141}$$

In this case, let $a = 1-\cos^{N-2}\mu$, $b = 4\sin(\mu/2)\cos^{N-2}(\mu/2)$, $\rho = |\psi_{\mu,\nu}\rangle\langle\psi_{\mu,\nu}|$, we have

$$\langle X\rangle_\rho = N\cos\left(\frac{\mu}{2}\right)^{N-1} = \Theta(N), \quad \langle Y\rangle_\rho = 0, \quad \langle Z\rangle_\rho = 0, \tag{142}$$

$$\langle \Delta X^2\rangle_\rho = N\left(N(1-\cos^{2(N-1)}\left(\frac{\mu}{2}\right) - (N-1)\frac{a}{2}\right) = \mathrm{O}(1/\xi^4), \tag{143}$$

$$\langle \Delta Y^2\rangle_\rho = N\left(1 + \frac{1}{4}(N-1)(a - \sqrt{a^2+b^2})\right) = \Theta(N\xi^2), \tag{144}$$

$$\langle \Delta Z^2\rangle_\rho = N\left(1 + \frac{1}{4}(N-1)(a + \sqrt{a^2+b^2})\right) = \Theta(N/\xi^2). \tag{145}$$

Eq. (132) and Eq. (133) hold true from above, and Eq. (134) further holds because let $S = s_x X + s_y Y + s_z Z$, we have $\langle \Delta S^2\rangle_\rho \le s_x^2\langle \Delta X^2\rangle_\rho + s_y^2\langle \Delta Y^2\rangle_\rho + s_z^2\langle \Delta Z^2\rangle_\rho + 2|s_x s_y|\sqrt{\langle X^2\rangle_\rho\langle Y^2\rangle_\rho} + 2|s_y s_z|\sqrt{\langle Y^2\rangle_\rho\langle Z^2\rangle_\rho} + 2|s_z s_x|\sqrt{\langle Z^2\rangle_\rho\langle X^2\rangle_\rho} \le \mathrm{O}(N/\xi^2)$.

Note that the spin-squeezed states have a different correlation structure from the short-ranged correlated states we introduced previously. Take Eq. (140) as an example. It is a permutation invariant state. We consider a sufficiently small $\xi = \sqrt{p}$ such that $\langle \Delta Z^2\rangle_\rho / N$ is bounded below by a constant larger than 1. Then Eq. (145) implies $\langle Z_i Z_j\rangle_\rho - \langle Z_i\rangle_\rho\langle Z_j\rangle_\rho = 1/\Theta(N\xi^2)$ when $i \neq j$. Thus, $C_\rho(I;J)$ in Eq. (35) at least $1/\Theta(N\xi^2)$ when $I \neq J$, and $\sum_{J:J\notin\mathcal{S}_I} C_\rho(I;J) = \Omega(1/M\xi^2) = \Omega(M^{-1})$ because only a finite number of Js in $\mathcal{S}_I$ are excluded in the sum. Therefore, the spin-squeezed state does not satisfy Eq. (140).

The above proofs can generalize to cover signal/noise that acts independently but not identically on each qubit. Finally, we note that for purely dephasing channel (Eq. (1) in the main text), we can take $S_i = Y_i$ in the proof, and one does not require condition Eq. (134) any more.

References

[1] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". Nature photonics **5**, 222–229 (2011).

[2] C. L. Degen, F. Reinhard, and P. Cappellaro. "Quantum sensing". Rev. Mod. Phys. **89**, 035002 (2017).

[3] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. "Quantum metrology with nonclassical states of atomic ensembles". Rev. Mod. Phys. **90**, 035005 (2018).

[4] Stefano Pirandola, Bhaskar Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd. "Advances in photonic quantum sensing". Nat. Photonics. **12**, 724 (2018).

[5] Carlton M. Caves. "Quantum-mechanical noise in an interferometer". Phys. Rev. D **23**, 1693–1708 (1981).

[6] Bernard Yurke, Samuel L. McCall, and John R. Klauder. "Su(2) and su(1,1) interferometers". Phys. Rev. A **33**, 4033–4054 (1986).

[7] LIGO Collaboration. "A gravitational wave observatory operating beyond the quantum shot-noise limit". Nature Physics **7**, 962–965 (2011).

[8] LIGO Collaboration. "Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light". Nature Photonics **7**, 613–619 (2013).

[9] David Le Sage, Koji Arai, David R Glenn, Stephen J DeVience, Linh M Pham, Lilah Rahn-Lee, Mikhail D Lukin, Amir Yacoby, Arash Komeili, and Ronald L Walsworth. "Optical magnetic imaging of living cells". Nature **496**, 486–489 (2013).

[10] Gabriela Barreto Lemos, Victoria Borish, Garrett D Cole, Sven Ramelow, Radek Lapkiewicz, and Anton Zeilinger. "Quantum imaging with undetected photons". Nature **512**, 409–412 (2014).

[11] Mankei Tsang, Ranjith Nair, and Xiao-Ming Lu. "Quantum theory of superresolution for two incoherent optical point sources". Physical Review X **6**, 031033 (2016).

[12] MH Abobeih, J Randall, CE Bradley, HP Bartling, MA Bakker, MJ Degen, M Markham, DJ Twitchen, and TH Taminiau. "Atomic-scale imaging of a 27-nuclear-spin cluster using a quantum sensor". Nature **576**, 411–415 (2019).

[13] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen. "Spin squeezing and reduced quantum noise in spectroscopy". Phys. Rev. A **46**, R6797–R6800 (1992).

[14] J. J. Bollinger, Wayne M. Itano, D. J. Wineland, and D. J. Heinzen. "Optimal frequency measurements with maximally correlated states". Phys. Rev. A **54**, R4649–R4652 (1996).

[15] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland. "Toward heisenberg-limited spectroscopy with multiparticle entangled states". Science **304**, 1476–1478 (2004).

[16] JM Taylor, Paola Cappellaro, L Childress, Liang Jiang, Dmitry Budker, PR Hemmer, Amir Yacoby, R Walsworth, and MD Lukin. "High-sensitivity diamond magnetometer with nanoscale resolution". Nature Physics **4**, 810–816 (2008).

[17] Hengyun Zhou, Joonhee Choi, Soonwon Choi, Renate Landig, Alexander M Douglas, Junichi Isoya, Fedor Jelezko, Shinobu Onoda, Hitoshi Sumiya, Paola Cappellaro, et al. "Quantum metrology with strongly interacting spin systems". Physical review X **10**, 031003 (2020).

[18] Till Rosenband, DB Hume, PO Schmidt, Chin-Wen Chou, Anders Brusch, Luca Lorini, WH Oskay, Robert E Drullinger, Tara M Fortier, Jason E Stalnaker, et al. "Frequency ratio of al+ and hg+ single-ion optical clocks; metrology at the 17th decimal place". Science **319**, 1808–1812 (2008).

[19] Jürgen Appel, Patrick Joachim Windpassinger, Daniel Oblak, U Busk Hoff, Niels Kjærgaard, and Eugene Simon Polzik. "Mesoscopic atomic entanglement for precision measurements beyond the standard quantum limit". Proceedings of the National Academy of Sciences **106**, 10960–10965 (2009).

[20] Andrew D Ludlow, Martin M Boyd, Jun Ye, Ekkehard Peik, and Piet O Schmidt. "Optical atomic clocks". Reviews of Modern Physics **87**, 637 (2015).

[21] Raphael Kaubruegger, Denis V Vasilyev, Marius Schulte, Klemens Hammerer, and Peter Zoller. "Quantum variational optimization of ramsey interferometry and atomic clocks". Physical review X **11**, 041045 (2021).

[22] Christian D Marciniak, Thomas Feldker, Ivan Pogorelov, Raphael Kaubruegger, Denis V Vasilyev, Rick van Bijnen, Philipp Schindler, Peter Zoller, Rainer Blatt, and Thomas Monz. "Optimal metrology with programmable quantum sensors". Nature **603**, 604–609 (2022).

[23] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum metrology". Phys. Rev. Lett. **96**, 010401 (2006).

[24] Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Mădălin Guţă. "The elusive heisenberg limit in quantum-enhanced metrology". Nature communications **3**, 1063 (2012).

[25] BM Escher, Ruynet Lima de Matos Filho, and Luiz Davidovich. "General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology". Nature Physics **7**, 406–411 (2011).

[26] Rafal Demkowicz-Dobrzański and Lorenzo Maccone. "Using entanglement against noise in quantum metrology". Physical review letters **113**, 250801 (2014).

[27] Pavel Sekatski, Michalis Skotiniotis, Janek Kołodyński, and Wolfgang Dür. "Quantum metrology with full and fast quantum control". Quantum **1**, 27 (2017).

[28] Rafał Demkowicz-Dobrzański, Jan Czajkowski, and Pavel Sekatski. "Adaptive quantum metrology under general markovian noise". Phys. Rev. X **7**, 041009 (2017).

[29] Sisi Zhou, Mengzhen Zhang, John Preskill, and Liang Jiang. "Achieving the heisenberg limit in quantum metrology using quantum error correction". Nature communications **9**, 78 (2018).

[30] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac. "Improvement of frequency standards with quantum entanglement". Phys. Rev. Lett. **79**, 3865–3868 (1997).

[31] Sisi Zhou and Liang Jiang. "Asymptotic theory of quantum channel estimation". PRX Quantum **2**, 010343 (2021).

[32] Marcin Jarzyna and Rafał Demkowicz-Dobrzański. "Matrix product states for quantum metrology". Phys. Rev. Lett. **110**, 240405 (2013).

[33] Krzysztof Chabuda, Jacek Dziarmaga, Tobias J Osborne, and Rafał Demkowicz-Dobrzański. "Tensor-network approach for quantum metrology in many-body quantum systems". Nature communications **11**, 250 (2020).

[34] Masahito Hayashi, Zi-Wen Liu, and Haidong Yuan. "Global heisenberg scaling in noisy and practical phase estimation". Quantum Science and Technology **7**, 025030 (2022).

[35] Carl W Helstrom. "Quantum detection and estimation theory". Journal of statistical physics **1**, 231–252 (1969).

[36] Alexander S Holevo. "Probabilistic and statistical aspects of quantum theory". Volume 1. Springer Science & Business Media. (2011).

[37] Samuel L. Braunstein and Carlton M. Caves. "Statistical distance and the geometry of quantum states". Phys. Rev. Lett. **72**, 3439–3443 (1994).

[38] Ole E Barndorff-Nielsen and Richard D Gill. "Fisher information in quantum statistics". Journal of Physics A: Mathematical and General **33**, 4481 (2000).

[39] Richard D Gill and Serge Massar. "State estimation for large ensembles". Physical Review A **61**, 042312 (2000).

[40] Matteo G. A. Paris. "Quantum estimation for quantum technology". International Journal of Quantum Information **07**, 125–137 (2009).

[41] Peter W. Shor. "Scheme for reducing decoherence in quantum computer memory". Phys. Rev. A **52**, R2493–R2496 (1995).

[42] E. Knill, R. Laflamme, and G. Milburn. "Efficient linear optics quantum computation" (2000). quant-ph:quant-ph/0006088.

[43] T. C. Ralph, A. J. F. Hayes, and Alexei Gilchrist. "Loss-tolerant optical qubits". Phys. Rev. Lett. **95**, 100501 (2005).

[44] Allen Zang, Tian-Xing Zheng, Peter C. Maurer, Frederic T. Chong, Martin Suchara, and Tian Zhong. "Enhancing Noisy Quantum Sensing by GHZ State Partitioning" (2025). arXiv:2507.02829.

[45] Masahiro Kitagawa and Masahito Ueda. "Squeezed spin states". Phys. Rev. A **47**, 5138–5143 (1993).

[46] Jian Ma, Xiaoguang Wang, C.P. Sun, and Franco Nori. "Quantum spin squeezing". Physics Reports **509**, 89–165 (2011).

[47] Duger Ulam-Orgikh and Masahiro Kitagawa. "Spin squeezing and decoherence limit in ramsey spectroscopy". Physical Review A **64**, 052106 (2001).

[48] Jonatan Bohr Brask, Rafael Chaves, and Janek Kołodyński. "Improved quantum magnetometry beyond the standard quantum limit". Physical Review X **5**, 031010 (2015).

[49] Anders S Sørensen and Klaus Mølmer. "Entanglement and extreme spin squeezing". Physical review letters **86**, 4431 (2001).

[50] Géza Tóth and Iagoba Apellaniz. "Quantum metrology from a quantum information science perspective". Journal of Physics A: Mathematical and Theoretical **47**, 424006 (2014).

[51] Luca Pezzé and Augusto Smerzi. "Entanglement, nonlinear dynamics, and the heisenberg limit". Physical review letters **102**, 100401 (2009).

[52] Elliott H. Lieb and Derek W. Robinson. "The finite group velocity of quantum spin systems". Commun. Math. Phys. **28**, 251–257 (1972).

[53] Chi-Fang (Anthony) Chen, Andrew Lucas, and Chao Yin. "Speed limits and locality in many-body quantum dynamics". Reports on Progress in Physics **86**, 116001 (2023).

[54] S. Bravyi, M. B. Hastings, and F. Verstraete. "Lieb-robinson bounds and the generation of correlations and topological quantum order". Phys. Rev. Lett. **97**, 050401 (2006).

[55] Yaoming Chu, Xiangbei Li, and Jianming Cai. "Strong quantum metrological limit from many-body physics". Phys. Rev. Lett. **130**, 170801 (2023).

[56] B L Higgins, D W Berry, S D Bartlett, M W Mitchell, H M Wiseman, and G J Pryde. "Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements". New Journal of Physics **11**, 073023 (2009).

[57] Shelby Kimmel, Guang Hao Low, and Theodore J. Yoder. "Robust calibration of a universal single-qubit gate set via robust phase estimation". Phys. Rev. A **92**, 062315 (2015).

[58] Federico Belliardo and Vittorio Giovannetti. "Achieving heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error". Phys. Rev. A **102**, 042613 (2020).

[59] Tommaso Macrì, Augusto Smerzi, and Luca Pezzè. "Loschmidt echo for quantum metrology". Phys. Rev. A **94**, 010102 (2016).

[60] David Poulin. "Lieb-robinson bound and locality for general markovian quantum dynamics". Phys. Rev. Lett. **104**, 190401 (2010).

[61] Jae-Seung Lee and A. K. Khitrin. "Stimulated wave of polarization in a one-dimensional ising chain". Phys. Rev. A **71**, 062338 (2005).

[62] Atsuki Yoshinaga, Mamiko Tatsuta, and Yuichiro Matsuzaki. "Entanglement-enhanced sensing using a chain of qubits with always-on nearest-neighbor interactions". Phys. Rev. A **103**, 062602 (2021).

[63] Arkadiusz Kobus and Rafał Demkowicz-Dobrzański. "Asymptotically optimal joint phase and dephasing strength estimation using spin-squeezed states" (2025).

[64] Maxwell Block, Bingtian Ye, Brenden Roberts, Sabrina Chern, Weijie Wu, Zilin Wang, Lode Pollet, Emily J Davis, Bertrand I Halperin, and Norman Y Yao. "Scalable spin squeezing from finite-temperature easy-plane magnetism". Nature PhysicsPages 1–7 (2024).

[65] Tommaso Comparin, Fabio Mezzacapo, and Tommaso Roscilde. "Robust spin squeezing from the tower of states of u(1)-symmetric spin hamiltonians". Phys. Rev. A **105**, 022625 (2022).

[66] Tommaso Roscilde, Filippo Caleca, Adriano Angelone, and Fabio Mezzacapo. "Scalable spin squeezing from critical slowing down in short-range interacting systems". Phys. Rev. Lett. **133**, 210401 (2024).

[67] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acín, and M. Lewenstein. "Asymptotic role of entanglement in quantum metrology". Phys. Rev. A **94**, 012339 (2016).

[68] A. T. Rezakhani, M. Hassani, and S. Alipour. "Continuity of the quantum fisher information". Phys. Rev. A **100**, 032317 (2019).

[69] Rajendra Bhatia. "Pinching, trimming, truncating, and averaging of matrices". The American Mathematical Monthly **107**, 602–608 (2000).

[70] E. P. Wigner and Mutsuo M. Yanase. "Information contents of distributions". Proceedings of the National Academy of Sciences **49**, 910–918 (1963).

[71] Shunlong Luo. "Wigner-yanase skew information vs. quantum fisher information". Proceedings of the American Mathematical Society **132**, 885–890 (2004).

[72] Benoit Roubert, Petr Braun, and Daniel Braun. "Large effects of boundaries on spin amplification in spin chains". Phys. Rev. A **82**, 022302 (2010).

[73] Kianna Wan and Robert Lasenby. "Bounds on adaptive quantum metrology under markovian noise". Physical Review Research **4**, 033092 (2022).