Using entanglement against noise in quantum metrology

Rafal Demkowicz-Dobrzański¹ and Lorenzo Maccone²

¹Faculty of Physics, University of Warsaw, ul. Hoża 69, PL-00-681 Warszawa, Poland
²Dip. Fisica and INFN Sez. Pavia, Univ. of Pavia, via Bassi 6, I-27100 Pavia, Italy

In this paper we analyze the performance of these strategies in the presence of specific noise models, and use the results to conjecture a general hierarchy for quantum metrology strategies in the presence of noise. Noise in quantum metrology has been extensively studied, e.g. see [12–25], but previous results [12–14] compared parallel-entangled with parallel-unentangled strategies which do not match in the noiseless case. In the noiseless case, classical single-probe sequential schemes (i) can attain the same 1/N scaling in precision in two ways: sampling the system in parallel with N probes in a joint entangled state, or sampling it N times sequentially with a single probe. The latter strategy does not employ entanglement, but both sample the system the same number of times and achieve the same precision. Their yield becomes inequivalent when noise is added and, surprisingly, the entangled strategy tolerates noise better (i.e. is more precise) than the entanglement-free one for many common noise models, an effect that was never observed before in quantum-enhanced protocols. We also analyze the asymptotic benefits of entangling ancillary systems in the estimation procedure. It is known that ancillary systems are useless in the noiseless scenario. We show that this is also the case for the erasure and dephasing models but, surprisingly, not for the amplitude-damping noise where ancilla-enhanced strategies permit a better precision. We use these results to conjecture a general hierarchy for quantum metrology strategies in the presence of noise.

PACS numbers: 03.65.Ta, 06.20.-f, 42.50.Lc

Quantum metrology [1, 2] describes parameter estimation techniques that, by sampling a system N times, achieve precision better than the 1/√N scaling of the central limit theorem of classical strategies. Different schemes can beat such limit (Fig. 1): (i) entanglement-free classical schemes where N/n independent probes sense the system sequentially thus rescaling the parameter, and hence the error, by n for each probe [3, 4]; (ii) entangled parallel schemes that employ a collective entangled state of the N probes that sample the system in parallel [5–8]; (iii) passive ancilla schemes, where the N probes may also be entangled with noiseless ancillas; (iv) active ancilla-assisted schemes (comprising all the previous cases) that also encompass all schemes employing feedback: adaptive procedures are described as general unitary operations that act on the probes and ancillas between the sensing stage and the final measurement [9, 10].

In this paper we analyze the performance of these strategies in the presence of specific noise models, and use the results to conjecture a general hierarchy of protocols. Noise in quantum metrology has been extensively studied, e.g. see [12–25], but previous results [12–14] compared parallel-entangled with parallel-unentangled strategies which do not match in the noiseless case. In the noiseless case, classical single-probe sequential schemes (i) can attain the same 1/N scaling as parallel entangled ones (ii) at the expense of an N-times longer sampling time, whereas passive and active ancilla schemes (iii) and (iv) offer no additional advantage [3, 11]. Single probe states are typically less sensitive to decoherence and much simpler to prepare than an N-probe entangled state, so it would seem [3] that the sequential strategy should be preferable in the presence of noise. Our first result is that this is not true: in the presence of noise (here we analyze dephasing, erasure and damping), the optimal sequential strategy achieves a precision strictly smaller than the optimal parallel-entangled strategy. Entanglement among probes increases the estimation precision over the unentangled sequential strategy that samples the system the same number N of times in the presence of noise, even though they match in the noiseless case. Our second main result is to show (ii) and (iii) are in general asymptotically inequivalent, by demonstrating that (iii) is strictly better than (ii) for amplitude-damping noise. Our third main result is to show that

---

**FIG. 1:** Quantum metrology strategies. The maps $\Lambda_\varphi$ encode on the probe the parameter $\varphi$ to be estimated. (i) sequential scheme: $\Lambda_\varphi$ acts $n$ times sequentially on $N/n$ input probes $\rho$ (this is an entanglement-free classical scheme); (ii) entangled parallel scheme: an entangled state of $N$ probes $\rho_N$ goes through $N$ maps $\Lambda_\varphi$ in parallel; (iii) passive ancilla scheme: the $N$ probes are also entangled with $M$ noiseless ancillas; (iv) active ancilla-assisted scheme: the action of $N$ channels $\Lambda_\varphi$ is interspersed with arbitrary unitaries $U_i$ representing interactions of the probe with ancillas. [Note that all the other schemes can be derived from (iv) choosing swap or identity unitaries $U_i$.]
the bounds to parallel-entangled strategies (ii) and (iii) derived for a large class of noise models \cite{13, 14} apply asymptotically in \( N \) also to the most general strategies (iv), suggesting that active ancilla-based schemes are not helpful in increasing the precision in the presence of noise \cite{43} (a result known only for the noiseless case \cite{4, 44}). Finally, even though we derive our results for specific classes of noise, we present a conjecture that these results are general for any Markovian noise, which provides a hierarchy of quantum metrology schemes

\[
\begin{align*}
(i) &= (ii) = (iii) = (iv) &\text{decoherence free}, \\
(i) &< (ii) = (iii) = (iv) &\text{dephasing, erasure}, \\
(i) &< (ii) < (iii) < (iv) &\text{amplitude-damping}, \\
(i) &\leq (ii) < (iii) < (iv) &\text{general conjecture}.
\end{align*}
\]

Namely in general, sequential strategies (i) are worse \cite{14} than parallel-entangled ones (ii), which might in some cases be improved by entangling the probes with noiseless ancillas (iii), but there is no additional asymptotic gain from using active ancilla-aided schemes (iv). Question marks represent our conjectures and the equality symbol “=” should be interpreted as asymptotically equivalent, though in the decoherence-free case as well as in the case of equality between (ii) and (iii) for erasure and dephasing noise this is actually a strict equality for any finite \( N \).

We emphasize that even in the cases where our bounds prove equivalent, the related strategies may not be if the bounds are not achievable, which must be verified on a case-by-case basis.

Schemes that employ quantum-error correction \cite{27, 29} are of type (iv) [or of type (iii) if no adaptive probe transformation is performed], so our claim might be misinterpreted as saying that error correction schemes are useless. Actually, our results just show that their asymptotic precision can also be achieved through (possibly unknown) strategies of type (ii,iii): e.g. the noise models considered in \cite{27, 29} allow for decoupling the decoherence from the parameter sensing transformation at short evolution times: so, the bounds derived for (ii,iii) also allow for the possibility of better than \( 1/\sqrt{N} \) scaling \cite{14}.

Outline of the paper: we first introduce the quantum Cramer–Rao bound for the strategies (i-iv), and derive some general bounds for their quantum Fisher information. We then prove a gap in precision between (i) and (ii), the equivalence of (ii), (iii), and (iv) in case of dephasing and erasure noise and finally inequivalence of (ii) and (iii) for amplitude-damping.

The map \( \Lambda_\varphi \) that writes the parameter \( \varphi \) on the state \( \rho \) of the probe acts as

\[
\rho_\varphi = \Lambda_\varphi(\rho) = \sum_k K_k^\varphi \rho K_k^{\varphi\dagger},
\]

with \( K_k^\varphi \) the Kraus operators. The precision of an estimation strategy can be gauged through the root mean square error \( \Delta \varphi \) of the measurement of \( \varphi \). It is lower-bounded by the quantum Cramer-Rao bound \cite{2, 5–8, 2}:

\[
\Delta \varphi \geq \frac{1}{\sqrt{\nu F(\rho_\varphi)}},
\]

where \( \nu \) is the number of times the estimation is repeated, and \( F(\rho) \) is the quantum Fisher information (QFI) of a state \( \rho \). The bound (3) is guaranteed to be achievable in general only asymptotically for \( \nu \to \infty \), but in case of noise models with linear QFI it is also tight for a single shot setting, \( \nu = 1 \), provided one considers the asymptotics with respect to the number of probe particles, \( N \to \infty \).

The QFIs for the schemes (i-iv) are defined as

\[
F^{(i)} = \max_{\rho, n} F\{[\Lambda_\varphi^n(\rho)]^{\otimes N/n}\},
F^{(ii)} = \max_{\rho_N} F\{\Lambda_\varphi^{\otimes N}(\rho_N)\},
F^{(iii)} = \max_{\rho_M} F\{\Lambda_\varphi^{\otimes N} \otimes \mathbb{1}^{\otimes M}(\rho_M)\},
F^{(iv)} = \max_{\rho_M, \{U_i\}} F\{U_N \Lambda_\varphi \ldots U_1 \Lambda_\varphi(\rho_M)\},
\]

where \( \rho \) denotes an input state of a single probe and we look for the optimal sequential-parallel splitting of the \( N \) probes in \( n \) channels for strategies (i), \( \rho_N \) is the global state of \( N \) probes in (ii), while \( \rho_M \) denotes the global probes-ancilla input state in (iii) and (iv). In the formula for \( F^{(iv)} \), the \( U_i \)'s act on all the probes while \( \Lambda_\varphi \) without loss of generality may be assumed to act on the first probe only. Due to the convexity of the QFI, the optimal input probes are pure.

The hierarchy conjecture (i) should be understood in terms of corresponding inequalities on QFIs: \( F^{(ii)} \leq F^{(iii)} \) is obvious as (ii) is a special case of (iii), the inequality may be strict as is the case of the amplitude-damping discussed below; \( F^{(iii)} \leq F^{(iv)} \) is also easy to show since taking swap operators \( U_i \) in (iv) one can obtain the action of parallel channels on an entangled input state (iii). It is less trivial to determine the cases when inequalities turn to equalities and the corresponding schemes become asymptotically equivalent. Finally, the \( F^{(i)} \leq F^{(ii)} \) inequality is more challenging to prove in general, but we show below it applies to dephasing, erasure and amplitude damping noise, by showing that using \( N00N \) states \((|0\rangle^{\otimes N} + |1\rangle^{\otimes N})/\sqrt{2}\) in (ii) gives the same performance as (i), but that \( N00N \) states are not the optimal ones for the noisy (ii). Indeed we show that for dephasing, erasure or amplitude damping the inequality \( F^{(i)} < F^{(ii)} \) is strict, proving the advantage of parallel schemes (i). We also present general tools to derive bounds for (iv) and show that they are asymptotically equivalent to known bounds for (ii,iii). Moreover, since these bounds are saturable for dephasing and erasure using (ii) schemes, there is no asymptotic advantage of (iv) over the simpler (ii) and (iii) in these cases.
Calculating QFI explicitly for large N is in general not possible but bounds to it are known. The most versatile ones employ the non-uniqueness of the Kraus representation [13, 34]: \( \Lambda_\varphi \) is unchanged if one replaces \( K^\varphi_k \) with \( \hat{K}^\varphi_k = \sum u_k K^\varphi_k \), where \( u_k \) is an arbitrary \( \varphi \)-dependent unitary matrix. This produces bounds on the maximal QFI of a transformation \( \Lambda_\varphi \) in terms of minimizing over the possible Kraus representations [14, 34]:

\[
\max_{\rho} F[\Lambda_\varphi(\rho)] \leq 4 \min_{K^\varphi_k} \| \sum_k K^\varphi_k \|, \tag{8}
\]

where \( K^\varphi_k = \frac{\partial K^\varphi_k}{\partial \varphi} \) and \( \| \cdot \| \) is the operator norm. The above inequality becomes an equality provided one replaces \( \Lambda_\varphi \) with a trivially-extended channel \( \Lambda_\varphi \otimes \mathbb{1} \) which represents the possibility of entangling the probes with an ancilla [34]. This immediately implies that the bounds derived for (ii) will also be valid for (iii).

We now recall known bounds for \( F^{(ii/iii)} \) and derive a new bound for \( F^{(iv)} \) using the minimization of Eq. (8). Bounds for (ii) and (iii) are equivalent (as argued above) so we use a combined notation (ii/iii). For any Kraus representation \( K^\varphi_k \) of a single channel \( \Lambda_\varphi \) one can write a product Kraus representation for channels \( \Lambda_{N\varphi} \), \( U_1 \Lambda_\varphi \ldots U_N \Lambda_\varphi \) corresponding to schemes (ii/iii), (iv) respectively: \( K^{(ii/iii)}_k \equiv K^\varphi_k \otimes \ldots \otimes K^\varphi_k \); \( K^{(iv)}_k \equiv U_N K_{k_N} \ldots \ldots U_1 K_{k_1} \), where \( k = \{k_1, \ldots, k_N\} \).

For (ii/iii) the minimization [8] gives a simple bound expressed in terms of single-channel Kraus operators [34]:

\[
F^{(ii/iii)} \leq 4 \min N \| \alpha \| + N(N - 1) \| \beta \|^2 \leq 4 \min_{K^\varphi_k, \| \beta \| = 0} N \| \alpha \|, \tag{9}
\]

with \( \alpha \equiv \sum_k K^\varphi_k K^\varphi_k \) and \( \beta \equiv \sum_k K^\varphi_k K^\varphi_k \). The last inequality in [8] may be used without loss of efficiency for large N provided there is a Kraus representation for which \( \beta = 0 \) (it exists for many noisy maps), which immediately implies linear QFI scaling with \( N \) for such maps [13, 34]. The minimization in Eq. (8) can be performed with semi-definite programming [14, 36] or with the educated-guess method [35].

The derivation of the general bound for (iv) uses again [8] and a product Kraus representation. It gives (see supplemental material for the details)

\[
F^{(iv)} \leq 4 \min_{K^\varphi_k} N \| \alpha \| + N(N - 1) \| \beta \| \left( \| \alpha \| + \| \beta \| + 1 \right) \leq 4 \min_{K^\varphi_k, \| \beta \| = 0} N \| \alpha \|. \tag{10}
\]

Importantly, the asymptotic form of the bound is equivalent to [8], the one derived for (ii/iii) if \( \beta = 0 \) is feasible.

One can also resort to less powerful but more intuitive derivations of the bounds for (iv), based on the concept of classical and quantum simulations of the channel [14, 37, 38]. These techniques were originally proposed for (ii/iii), but they trivially generalize to (iv) as discussed below. The idea behind classical-quantum simulation [39] is to formally replace the action of the parameter-dependent channel \( \Lambda_\varphi \) with a parameter-independent extended map \( \Lambda_\varphi \otimes \mathbb{1} \) where the parameter dependence is put into the ancillary input system \( \sigma_\varphi \), so that for any input state \( \rho_\varphi \) : \( \Lambda_\varphi(\rho_\varphi) = \Lambda(\rho \otimes \sigma_\varphi) \). Since QFI is nonincreasing under parameter-independent quantum maps we have that for any \( p \) : \( F(\Lambda(\rho \otimes \sigma_\varphi)) \leq F(\sigma_\varphi) \). In case \( \sigma_\varphi \) is a diagonal density matrix we call the simulation classical (the parameter dependence is moved to classical probabilities) otherwise we call it quantum. This immediately implies for the schemes (ii) that \( F(\Lambda_{N\varphi}^{(ii/iii)}(\rho_\varphi)) \leq F(\Lambda_{N\varphi}^{(ii/iii)}(\rho_\varphi \otimes \sigma_\varphi^{(N)})) \leq NF(\sigma_\varphi) \) and proves the linear scaling of QFI for channels for which a simulation exists with \( F(\sigma_\varphi) < \infty \) [14, 36, 38]. A trivial fact that has not been noticed before is that the same method can be applied to schemes (iv) involving \( N \) uses of the channel \( \Lambda_\varphi \) as all parameter independent operations can be rewritten as one black-box quantum operation \( \Lambda \) that is fed with \( \sigma_\varphi^{(N)} \) being the only parameter dependent element in the construction, see Fig. 2. Moreover, since the QFI at a given point \( \varphi = \varphi_0 \) depends only on the form of the channel and its first derivative at \( \varphi_0 \) it is sufficient to construct a local simulation of the channel at a given point \( \Lambda_\varphi(\rho) = \Lambda(\rho \otimes \sigma_\varphi) + O(\delta \varphi^2) \), where \( \delta \varphi = \varphi - \varphi_0 \). Hence we get:

\[
F^{(ii/iii)} \leq N \min_{\Lambda, \sigma_\varphi} F(\sigma_\varphi). \tag{11}
\]

As discussed in [14, 36] this bound often coincides with the asymptotic bound in [8], e.g. in the case of erasure or dephasing (but not amplitude-damping, see [38]).

We now analyze dephasing, erasure and amplitude-damping noise. Let \( |0\rangle, |1\rangle \) be the eigenbasis of the phase encoding unitary \( U_\varphi = |0\rangle + e^{i\varphi}|1\rangle/\sqrt{2} \). We assume that the dephasing is defined with respect to the same basis so the corresponding Kraus operators read:

\[
K_0 = \mathbb{1} \left( \frac{1 + \sqrt{\eta}}{2} \right)^{1/2}, \quad K_1 = \sigma_z \left( \frac{1 - \sqrt{\eta}}{2} \right)^{1/2}, \tag{12}
\]

where \( \mathbb{1} = |0\rangle \langle 0| + |1\rangle \langle 1| \), \( \sigma_z = |0\rangle \langle 0| - |1\rangle \langle 1| \), and \( \sqrt{\eta} \) is the decoherence rate of the off-diagonal terms in the
density matrix. Since both Kraus operators commute with the unitary $U_\varphi$, we can separate the noise map from the sampling and consider a total evolution of the form

$$\rho_\varphi = \Lambda_\varphi(\rho) = \sum_k K_k U_\varphi \rho U_\varphi^\dagger K_k^\dagger.$$  \hspace{1cm} (13)

Instead, for erasure noise the probe is untouched with probability $\eta$ while with probability $1 - \eta$ its state is replaced with one in a subspace orthogonal to the subspace where the estimation takes place (again the noise map and $U_\varphi$ commute and the map can be written in a Kraus form, see supplemental material). The erasure map is isomorphic to optical loss applied to a state with fixed number of distinguishable photons with transmission coefficient $\eta$ in both arms of an interferometer \cite{14, 35}. Finally, in case of amplitude damping the two Kraus operators read

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{\eta} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{1 - \eta} \\ 0 & 0 \end{pmatrix},$$  \hspace{1cm} (14)

where $\eta$ represents the probability of a particle to switch from the excited to the ground state.

We start with the performance of entanglement-free strategies, calculating $F^{(i)}$. In case of the erasure, since in the noiseless case the optimal probe state is $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, while the probability of erasure event that removes all the phase information does not depend on the state itself, the optimal input state remains the same, and yields $F[\Lambda_\varphi(|+\rangle\langle+|)] = \eta$. For dephasing and amplitude damping the situation is less obvious but the optimal probe state is again $|+\rangle$ and the QFI at the output can be given again $F[\Lambda_\varphi(|+\rangle\langle+|)] = \eta$ \cite{36}, see supplemental material for a simple proof in case of dephasing.

To calculate $F^{(i)}$ it remains to optimize the number $n$ of sequential maps for each probe, see Fig. 3. Using $n$ maps in a sequence increases the overall phase rotation $n$ times at the cost of increasing the decoherence parameter $\eta$ to $\eta^n$, whereas considering parallel channels simply adds their QFIs. Therefore, $F[\Lambda_\varphi^n(\rho_{+\rangle\langle+|})^\otimes n] = N/n \cdot n^2 \eta^n$. This is the same formula which would be obtained for (ii) with input $\ket{0}_{00} \ket{N}_{\text{state}}$ \cite{17}. Treating $1 \leq n \leq N$ as a continuous parameter, we find the optimal value $n = \lceil \ln(1/\eta) \rceil^{-1}$ provided $e^{-1} \leq \eta \leq e^{-1/N}$, which corresponds to (i)

$$F^{(i)} = \frac{N}{e \ln(1/\eta)}.$$  \hspace{1cm} (15)

For erasure and dephasing, we use the inequality \cite{7} to calculate (see the supplemental material for the optimal Kraus representation that yields this)

$$F^{(i/\text{ii})} \lesssim \frac{N \eta}{1 - \eta}.$$  \hspace{1cm} (16)

Importantly, this bound is asymptotically saturable for both models with a scheme (ii) where the optimal input probes are prepared in a spin-squeezed states for atomic systems \cite{13}, \cite{15}, or in squeezed states of light for optical implementations \cite{14}, \cite{32}, \cite{33}. One can also derive the corresponding bound for the amplitude damping (see supplemental material) which reads $F^{(i/\text{iii})} \lesssim \frac{4N \eta}{1 - \eta}$. This bound, however, is not tight for (ii) strategies, which has been proven recently in \cite{46} using an alternative method based on the calculus of variations—the actual tight bound for (ii) coincides with Eq. (14). This makes the case of amplitude damping distinct from the other two and opens up a possibility of proving the asymptotic benefits of using the ancillas, see below.

Finally, regarding scheme (iv), we note that since the asymptotic bound on $F_Q^{(iv)}$ coincides with the bound on $F_Q^{(i/\text{iii})}$ (since $\beta = 0$) and the latter is asymptotically saturable using (ii) for erasure and dephasing, this immediately implies that there is no asymptotic benefit in using (iv) in these case.

In order to inspect the benefits of entangled-based strategies over sequential ones we plot in Fig. 3 the ratio of formulas in Eqs. (14) and (15) as a function of $\eta$. In the limit of weak decoherence $\eta \approx 1$ we recover the known result \cite{12, 13} that the quantum enhancement in frequency estimation is bounded by $\exp(1)$. The correspondence should be clear when one notices that the weak decoherence limit $\eta \rightarrow 1$ corresponds to the limit of vanishing interrogation time for frequency estimation schemes. We stress again however, that in the noiseless case $\eta = 1$ all four metrology schemes perform equally well, achieving the Heisenberg scaling.
(iii) are equivalent, in case of dephasing and erasure. Surprisingly, they are inequivalent in general, and in particular in case of the amplitude-damping noise. In this case, as mentioned above and proven in [11], bound (16) is tight for (ii). A numerical search for optimal ancilla assisted strategies (iii) for small number of probes $N \leq 4$ gives a QFI that exceeds the bound (16) for $\eta \lesssim 0.5$, see supplemental material. Most importantly, this advantage of (iii) over (ii) strategies will be preserved also in the asymptotic limit, since the bound (16) is linear in $N$ and the same linear gain can be achieved by simply repeating experiment, e.g. using the optimal 4 particle strategy $N/4$ times. This gives a (numerical) proof that (ii) is strictly less powerful than (iii) for amplitude-damping.

In conclusion, we have presented a hierarchy for the performance of quantum metrology in the presence of dephasing, erasure and amplitude-damping noise, and illustrated a conjecture on how this hierarchy can be extended to arbitrary noise models, based on new general bounds. Surprisingly, in this hierarchy entanglement-free schemes perform worse than entangled ones, and in some cases schemes with passive ancillas perform better than unaided ones, even though they are all equivalent in the noiseless case.

RDD thanks Marcin Jarzyna for useful discussions. This research work supported by the FP7 IP project SIQS co-financed by the Polish Ministry of Science and Higher Education.

Supplemental Material

Optimal single particle probe states for the dephasing noise model

Here we prove that the $|+\rangle$ input probe state is indeed optimal for the dephasing channel defined by Kraus representations [12]. This input probe yields $F[\Lambda_{\varphi}(|+\rangle\langle+|)] = \eta$. To see that this is indeed the optimal probe, it is enough to show that it achieves the bound (8) for a particular Kraus decomposition of the channel. The canonical Kraus representations $K_0^\varphi = K_1 U_{\varphi}$, with $K_1$ given in (12) is not appropriate as the corresponding $4\|\sum K_0^\varphi K_0^\varphi\|$ equals 1. If we, however, replace $K_1$ with $\tilde{K}_1$ given by

\begin{align}
\tilde{K}_0 &= \cos(\varphi) K_0 - i \sin(\varphi) K_1, \\
\tilde{K}_1 &= \cos(\varphi) K_1 - i \sin(\varphi) K_0,
\end{align}

with $\xi = \sqrt{1-\eta}$, we get $4\|\sum \tilde{K}_0^\varphi \tilde{K}_0^\varphi\| = \eta$ proving that $|+\rangle$ is indeed optimal.

Optimal Kraus representations yielding asymptotically tight bounds for dephasing and erasure noise models

Here we present explicit Kraus representations for the dephasing, and erasure noise that yield asymptotically tight bounds on QFI given in Eq. (14) of the main text. Detailed discussion of amplitude damping model is given in . All Kraus operators are assumed to be additionally multiplied on the r.h.s. by the unitary evolution $U_{\varphi} = |0\rangle\langle 0| + e^{i\varphi}|1\rangle\langle 1|$, i.e.: $K_0^\varphi = K_0 U_{\varphi}$ before being used to calculate the bound:

$$F^{(ii-iv)}_{\beta=0} \min_{K_0^\varphi, K_1^\varphi} N \|\alpha\|,$$

where

$$\alpha = \sum_k \hat{K}_0^\varphi \hat{K}_k^\varphi, \quad \beta = \sum_k \hat{K}_0^\varphi \hat{K}_k^\varphi.$$

Dephasing

While the canonical Kraus representations is given by [12], the optimal Kraus representation that gives minimal $\|\alpha\|$ under the constraint $\beta = 0$ reads

\begin{align}
\tilde{K}_0 &= \cos(\chi \varphi) K_0 - i \sin(\chi \varphi) K_1, \\
\tilde{K}_1 &= \cos(\chi \varphi) K_1 - i \sin(\chi \varphi) K_0
\end{align}

where $\chi = 1/[2\sqrt{1-\eta}]$ and yields $\|\alpha\| = \eta/[4(1-\eta)]$ which reproduces the asymptotically tight bound given in Eq. (14). Note the difference between this representation and the one used to derive the optimal single probe QFI given in Eqs. (17) and (18) amount to a different “Kraus rotation speed”: $\chi$ instead of $\xi$, which guarantees that $\beta = 0$.

Erasure

The canonical Kraus operators for the erasure map is

\begin{align}
K_0 &= \begin{pmatrix} \sqrt{\eta} & 0 & 0 \\ 0 & \sqrt{\eta} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
K_2 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \sqrt{1-\eta} & 0 & 0 \end{pmatrix}, \quad K_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1-\eta \end{pmatrix}
\end{align}

where the third dimension corresponds to the state from the phase insensitive subspace. $K_0, K_1$ correspond to the probe remaining untouched within the phase-sensitive or phase-insensitive subspace respectively while $K_2, K_3$ represent the events of probe being erased to the phase insensitive state from either $|0\rangle$ or $|1\rangle$. The optimal Kraus representation reads:

$$\tilde{K}_0 = K_0, \quad \tilde{K}_1 = K_1, \quad \tilde{K}_2 = e^{-i\xi \varphi} K_2, \quad \tilde{K}_3 = e^{i\xi \varphi} K_3$$
where $\zeta = 1/[2(1-\eta)]$ and yields $||\alpha|| = \eta/[4(1-\eta)]$ in agreement with Eq. [16].

**Asymptotic bounds and inequivalence of (ii) and (iii) for amplitude-damping noise model**

Here we discuss the asymptotic bounds for amplitude-damping noise model and show numerically that (ii) and (iii) are inequivalent as there is a finite gap in precision between (ii) and (iii). Making use of the formula [16] we obtain the asymptotic bound in the form [14, 36]

$$F^{(ii-iv)} \leq \frac{4N\eta}{1-\eta},$$

(25)

which corresponds to the following optimal choice of Kraus operators $\tilde{K}_i$ expressed in terms of canonical Kraus operators given in Eq. [14]:

$$\tilde{K}_1 = e^{-i\varphi/2}K_1, \quad \tilde{K}_2 = e^{i\xi\varphi/2}K_2,$$

(26)

where $\xi = (1+\eta)/(1-\eta)$, which yields $||\alpha|| = \eta/(1-\eta)$. Using alternative method methods of variational calculus a tighter and asymptotically saturable bound for (ii) strategies has been derived in [40]:

$$F^{(ii)} \lesssim \frac{N\eta}{1-\eta},$$

(27)

which coincides with the asymptotic bounds for the dephasing and erasure noise, see Eq. [16]. Clearly, there is a significant gap between asymptotically tight bound (27) valid for (ii) strategies and the more general bound (25) covering all ancilla-assisted strategies. However, since the bound (27) is not guaranteed to be saturable even with (iii) or (iv) strategies, it is not yet a proof that there is in fact asymptotic ancilla-assisted precision enhancement. Therefore we have performed a numerical search for the optimal (iii) strategies for low $N$ for which the numerical search is feasible. We have achieved it by implementing a semi-definite program minimizing the right-hand side of (3) over Kraus representations in a analogous way as in [14, 36], but this time without assuming the product Kraus representation structure $K^{\varphi(ii/iii)}_k = K^{\varphi}_{k_1} \otimes \cdots \otimes K^{\varphi}_{k_N}$ for $N$ parallel channels $\Lambda^{\otimes N}$, but rather allowing for an arbitrary Kraus representation. This complicates numerics significantly and is therefore feasible only for small $N$, but guarantees that the resulting value of QFI is achievable using passive-ancilla assisted strategies (iii).

We depict the results in Fig. 4 where it is evident that numerically obtained QFI for the (iii) schemes (solid, gray) surpasses the bound for (ii) (dashed, black) strategies already for $N = 4$. As argued in the main text, although this advantage is demonstrated here only for finite $N = 4$, it can be pushed to the interesting asymptotic regime of infinite $N$ by simply repeating the experiment many times independently and averaging the outcomes. This proves that indeed (ii) is inequivalent to (iii): surprisingly the use of passive ancillas allows one to achieve a strictly higher accuracy in estimation with amplitude-damping noise. For comparison we also present numerically obtained results of achievable QFI in case of (ii) strategies (solid, black) which were obtained using a numerical iterative algorithm proposed in [11, 12], as well as the universally valid bound (25) (dashed, gray). It still remains an interesting open question whether the bound (25) is tight, i.e. whether the solid gray line will asymptotically approach it, and if not whether allowing for active adaptive schemes (iv) would be sufficient to actually reach the bound. We also cannot exclude the possibility that the bound is simply not tight and no strategies, even the most general (iv), can approach it asymptotically.

![FIG. 4: Comparison between the yield of the amplitude-damping channel with and without passive ancillas for exemplary decoherence parameter $\eta = 0.5$. We plot the Quantum Fisher Information (QFI) $F$ and its bounds in the presence of amplitude-damping as a function of the number $N$ of maps employed in the estimation. Solid black curve: attainable QFI without ancillas, strategy (ii); solid gray curve: attainable QFI with passive ancillas, strategy (iii); dashed black curve: Knysh et al. asymptotically tight upper bound for the QFI for (ii) strategies from [40]; dashed gray curve: our universal bound for QFI for both passive (iii) and active (iv) ancillas, no strategy can achieve better precision. The gray box emphasizes that for $N = 4$, and hence also asymptotically, the strategy (iii) can beat the Knysh et al. bound for all strategies of type (ii).](image)

**Derivation of the bound for general feedback assisted schemes (iv)**

Let us first try to derive a bound for a simple sequential strategy (i) $(k = 1, n = N)$ making use of Eq. [8] and a
product Kraus representation $K^{\varphi_1}_{k} = K_{kN}^{\varphi} \cdots K_{k1}^{\varphi}$:

$$F^{(i)} \leq 4 \min_{K_k^{\varphi}} \| \sum_k K_k^{\varphi(i)\dagger} K_k^{\varphi(i)} \| =$$

$$4 \min_{K_k^{\varphi}} \| \sum_{k, i, j} K_{k1}^{\varphi} \cdots K_{kN}^{\varphi} K_{kN}^{\varphi'} K_{k1}^{\varphi'} \cdots K_{k1}^{\varphi} \| = \sum_{k} \sum_{i,j=1}^{N} \| K_{k1}^{\varphi} \cdots K_{kN}^{\varphi} K_{kN}^{\varphi'} K_{k1}^{\varphi'} \cdots K_{k1}^{\varphi} \| .$$

(28)

First note the following property of the operator norm

$$\| \sum_k L_k A L_k \| \leq \| A \| \sum_k L_k^{\dagger} L_k$$

(29)

valid for any operator $A$ and any set of operators $L_k$. Making use of the above monotonicity property and the trace preservation condition $\sum_k K_k^{\varphi(i)\dagger} K_k^{\varphi(i)} = 1$ we get:

$$F^{(i)} \leq 4 \min_{K_k^{\varphi}} \sum_k \| \sum_i K_{ki}^{\varphi} \| K_{ki}^{\varphi\dagger} +$$

$$+ \sum_{i<j} \| \sum_{k_i, \ldots, k_j} K_{k_i}^{\varphi} \cdots K_{k_j}^{\varphi} K_{k_j}^{\varphi\dagger} \cdots K_{k_i}^{\varphi\dagger} + h.c. \| .$$

(30)

Focusing on the second summation term, observe that the trace preservation condition implies that $\sum_k K_k^{\varphi(i)\dagger} K_k^{\varphi(i)}$ is anti-hermitian. Let $iA$ be an anti-hermitian operator and consider the following chain of inequalities

$$\| \sum_k K_k^{\varphi} i A K_k + h.c. \| = \| \sum_k K_k^{\varphi} i A K_k - K_k^{\varphi} i A K_k \| =$$

$$= \| (K_k + i K_k)^\dagger i A (K_k + i K_k) - K_k^{\varphi\dagger} i A K_k - K_k^{\varphi\dagger} i A K_k \| .$$

(31)

Making use of the triangle inequality together with Eq. (29) we upper bound the above expression by

$$\leq 2 \| A \| \left( \sum_k \| K_k^{\varphi\dagger} K_k \| + \sum_k \| K_k^{\varphi\dagger} K_k \| + 1 \right) .$$

(32)

The above inequality allows us to rewrite Eq. (30) in a final form:

$$F^{(i)} \leq 4 \min_{K_k^{\varphi}} N \| \alpha \| + N(N-1) \| \beta \| (\| \alpha \| + \| \beta \| + 1) ,$$

(33)

where

$$\alpha = \sum_k K_k^{\varphi\dagger} K_k^{\varphi}, \quad \beta = \sum_k K_k^{\varphi\dagger} K_k^{\varphi} .$$

(34)

All the steps in the above derivation are also valid for (iv) in addition to (i), as the intermediate unitary operations $U_i$ do not affect the values of the operator norms that appear in the derivation. Hence, inequality (33) applies also to this case:

$$F^{(iv)} \leq 4 \min_{K_k^{\varphi}} N \| \alpha \| + N(N-1) \| \beta \| (\| \alpha \| + \| \beta \| + 1) .$$

(35)

This should sound an alarm that most probably the bound is far from tight for (i), as it is invariant under such a significant generalization of the scheme. This is indeed the case. As discussed in the main text, for the case of erasure and dephasing decoherence models the above bound is asymptotically tight when (iv) schemes are considered, but is hardly useful in analyzing the performance of (i). The intuitive reason behind this, is that when minimizing over Kraus representations of the channels we have restricted ourselves to product Kraus representations derived from single channel Kraus representations. When thinking of $\Lambda_{\varphi}$ this makes an artificial formal separation of the channels acting in a sequence, and significantly better bounds may be derived by performing a minimization over general Kraus representation of $\Lambda_{\varphi}$ instead of just product ones.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Science 306, 1330 (2004).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Nature Phot. 5, 222 (2011).
[3] A. Luis, Phys. Rev. A 65, 025802 (2002).
[4] B. L. Higgins, D.W. Berry, S.D. Bartlett, H.M. Wiseman, and G.J. Pryde, Nature 450, 393 (2007).
[5] S. L. Braunstein, C. M. Caves and G. J. Milburn, Ann. Phys. 247, 135 (1996).
[6] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. 72, 3439 (1994).
[7] C. W. Helstrom, “Quantum Detection and Estimation Theory,” (Academic Press, New York, 1976).
[8] A.S. Holevo, “Probabilistic and Statistical Aspects of Quantum Theory,” (North-Holland, Amsterdam, 1982).
[9] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. 96, 010401 (2006).
[10] W. van Dam, G.M. D’Ariano, A. Ekert, C. Macchiavello, and M. Mosca, Phys. Rev. Lett. 98, 090501 (2007).
[11] A. M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. 47, 155 (2000).
[12] S. F. Huelga, et al., Phys. Rev. Lett. 79, 3865 (1997).
[13] B.M. Escher, R.L. de Matos Filho, and L. Davidovich, Nature Phys. 7, 406 (2011).
[14] R. Demkowicz-Dobrzański, J. Kolodyński, and M. Guta, Nature Comm. 3, 1063 (2012).
[15] D. Ulam-Orgikh M. Kitagawa, Phys. Rev. A, 64, 052106 (2001).
[16] A. Shaji and C. M. Caves, Phys. Rev. A 76, 032111 (2007).
[17] U. Dorner, et al., Phys. Rev. Lett. 102, 040403 (2009).
[18] J. Kolodyński and R. Demkowicz-Dobrzański, Phys. Rev. A 82, 053804 (2010).
[19] S. V. Knysh, N. Smelyanskiy, and G. A. Durkin, Phys. Rev. A 83, 021804 (2011).
[20] C. M. Caves, Phys. Rev. D 23, 1693 (1981).
[21] M. Xiao, L.-A. Wu, and H. J. Kimble, Phys. Rev. Lett. 59, 278 (1987).
[22] M. A. Rubin and S. Kaushik, Phys. Rev. A 75, 053805 (2007).
[23] R. Demkowicz-Dobrzański, Laser Phys. 20, 1197 (2010).
[24] S. Alipour, M. Mehboudi, A.T. Rezakhani, Phys. Rev. Lett. 112, 120405 (2014).
[25] S. Alipour, arXiv:1403.8033 (2014).
[26] M. Jarzyna and R. Demkowicz-Dobrzański, arxiv:1407.4805 (2014)
[27] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. 112, 150802 (2014)
[28] W. Dur, M. Skotniotis, F. Frwis, and B. Kraus, Phys. Rev. Lett. 112, 080801 (2014)
[29] G. Arrad, Y. Vinkler, D. Aharanov and A. Retzker, Phys. Rev. Lett. 112, 150801 (2014)
[30] R. Chaves, J.B. Brask, M. Markiewicz, J. Kolodynski, and A. Acin, Phys. Rev. Lett. 111, 120401 (2013).
[31] A. Shaji, C.M. Caves, Phys. Rev. A 76, 032111 (2007).
[32] S. Boixo, C.M. Caves, A. Datta, A. Shaji, Laser Phys. 16, 1525 (2006).
[33] L. Maccone, Phys. Rev. A 88, 042109 (2013).
[34] A. Fujiwara and H. Imai, J. Phys. A: Math. Theor. 41, 255304 (2008)
[35] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kolodynski, ArXiv e-prints, 1405.7703 (2014).
[36] J. Kolodynski and R. Demkowicz-Dobrzański, New J. Phys. 15, 073043 (2013).
[37] R. Demkowicz-Dobrzański, K. Banaszek, and R. Schnabel, Phys. Rev. A 88, 041802(R) (2013)
[38] S. Boixo and C. Heunen, Phys. Rev. Lett. 108, 120402 (2012).
[39] K. Matsumoto, ArXiv e-prints, 1006.0300 (2010).
[40] S. V. Knysh, E. H. Chen, and G. A. Durkin, ArXiv e-prints, 1402.0495(2014)
[41] K. Macieszczak, M. Fraas, and R. Demkowicz-Dobrzański, Arxiv e-prints, arXiv:1311.5576 (2013)
[42] K. Macieszczak, Arxiv e-prints, Arxiv e-prints, arXiv:1312.1356 (2013)
[43] Our claim is stronger than the one in [33] that if (ii) and (iii) are limited by a $1/\sqrt{N}$ precision scaling, so is (iv): we derive an explicit bound for (iv) and show that it asymptotically coincides with the bound for (ii) and (iii).
[44] But they are equivalent in the noiseless case.
[45] This claim may seem in contradiction with the one of [33] where equivalence of the sequential and parallel-entangled strategy is proven in for dephasing. The contradiction is only apparent, as no optimization over the input state is performed there: only the response of the channel is analyzed.
[46] For completeness we should add that if $\eta < e^{-1}$ we take $n = 1$ which gives $F = N\eta$, while for $\eta > e^{-1/N}$ we take $n = N$ which gives $F = N^2\eta^N$. Note that in the asymptotic limit $N \to \infty$ and $\eta < 1$, we can ignore the case $\eta > e^{-1/N}$.  

[47]