Minimization of the estimation error for entanglement distribution networks with arbitrary noise

Liangzhong Ruan
School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
and Qutech, Delft University of Technology, 2628 CJ Delft, The Netherlands

(Received 21 June 2023; accepted 12 July 2023; published 23 August 2023)

Fidelity estimation is essential for the quality control of entanglement distribution networks. Because measurements collapse quantum states, we consider a setup in which nodes randomly sample a subset of the entangled qubit pairs to measure and then estimate the average fidelity of the unsampled pairs conditioned on the measurement outcome. The proposed estimation protocol achieves the lowest mean-square estimation error in a difficult scenario with arbitrary noise and no prior information. Moreover, this protocol is implementation friendly because it only performs local Pauli operators according to a predefined sequence. Numerical studies show that compared to existing fidelity estimation protocols, the proposed protocol reduces the estimation error in both scenarios with independent and identically distributed noise and correlated noise.

DOI: 10.1103/PhysRevA.108.022418

I. INTRODUCTION

Entanglement distribution networks [1] are an important developmental stage on the way to a full-blown quantum Internet [2–4]. Such networks enable device-independent protocols [5–9], thereby achieving the highest level of quantum security [10]. Entanglement distribution networks do not require nodes to have local quantum memory. Since efficient communication-compatible quantum memories are still under development [11], entanglement distribution networks serve as a cornerstone for realizing trustworthy quantum applications with state-of-the-art quantum technology [12–15].

Entanglement quality assessment is a key building block for entanglement distribution networks. To this end, fidelity estimation for entangled states is a promising candidate. Fidelity is a metric that indicates the quality of quantum states [16–19] and can be estimated with separable quantum measurements and classical postprocessing. Several fidelity estimation protocols have been proposed [20–23], and fidelity estimation protocols of entangled states have been implemented in several recent experiments [12–14,24]. However, there are still two challenges that need to be addressed.

(i) Excessive estimation error due to arbitrary noise. Quantum networks often face heterogeneous and correlated noise. When distributing quantum keys and estimating channel capacities, this noise leads to excessive estimation errors [25,26]. Therefore, designing low-error fidelity estimation protocols in the presence of arbitrary noise is an interesting area of research.

(ii) Efficiency loss due to separable operations. In the absence of quantum memory, nodes in entanglement distribution networks perform operations that are separable between all qubits. Such operations result in significantly lower estimation efficiency [27] compared to joint operations. Minimizing the loss of efficiency due to separable operations is a major challenge for fidelity estimation.

This work focuses on fidelity estimation for entanglement distribution networks. Since measurements collapse quantum states, we consider a network in which nodes randomly sample a subset of qubit pairs to measure and estimate the average fidelity of unsampled pairs conditioned on the measurement outcome.

We prove that the protocol proposed in this paper achieves the lowest estimation error in the difficult scenario with arbitrary noise and no prior information, thereby overcoming the challenges listed above. Moreover, the proposed protocol is implementation friendly. This protocol uses only local Pauli measurements, standard operations that can be implemented on a variety of quantum platforms [12,24,28,29], and determines the basis of each Pauli measurement according to a predefined sequence so that no adaptive operation is required.

The remainder of the paper is organized as follows. Section II formulates the problem of minimizing the estimation error of fidelity in a scenario with arbitrary noise and no prior information. Sections III and IV solve the formulated problem and present a protocol that minimizes the estimation error of fidelity. Section V evaluates the proposed protocol by comparing it with existing ones. Section VI presents a brief summary.

Regarding the notation in this paper, random variables and their realizations are represented in uppercase and lowercase letters, respectively, e.g., F and f. Vectors and matrices are denoted by bold letters, e.g., $\boldsymbol{\rho}$ and $\boldsymbol{M}$. The symbols $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance of a random variable, respectively, $\mathbb{I}_n$ denotes an $n \times n$ identity matrix, $\Pr(\cdot)$ denotes the probability of an event, $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, and $\mathrm{Tr}_n(\cdot)$ denotes the partial trace of the nth subsystem.

II. SYSTEM SETUP AND PROBLEM FORMULATION

Consider a scenario in which two nodes share N noisy qubit pairs and have no prior information on the noise. The nodes tend to estimate the fidelity of the qubit pairs with respect to the target maximally entangled state. Since maximally entangled states are mutually convertible via local operations, we set the target state to $|\Psi^-\rangle$ without loss of generality, where $|\Psi^{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}$ and $|\Phi^{\pm}\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}$ are the four Bell states. We denote the set of all qubit pairs by $\mathcal{N}$. The nodes randomly sample a number M ($<N$) of qubit pairs to measure. The set of sampled pairs $\mathcal{M}$ is drawn from all M subsets of $\mathcal{N}$ with equal probability.

Entanglement distribution networks do not require nodes to have local quantum memory [1]. Consequently, the nodes execute operators that are separable between all qubits. We define the separable operators as

$$\boldsymbol{M}_r = \sum_k \otimes^{n\in\mathcal{M}} \left(\boldsymbol{M}^{(A)}_{r,n,k} \otimes \boldsymbol{M}^{(B)}_{r,n,k}\right), \tag{1}$$

where the node index $X \in \{A, B\}$, the qubit pair index $n \in \mathcal{M}$, and the operator index $r \in \mathcal{R}$, with $\mathcal{R}$ denoting the set of all possible measurement outcomes. As positive-operator-valued measure operators

$$\boldsymbol{M}^{(X)}_{r,n,k} \succcurlyeq \boldsymbol{0}, \quad \sum_r \boldsymbol{M}_r = \mathbb{I}_{4^M}, \tag{2}$$

where $\succcurlyeq$ denotes the matrix inequality. The measurement operation performed on all qubit pairs can be defined as

$$\mathcal{O} = \{\boldsymbol{M}_r, r \in \mathcal{R}\}. \tag{3}$$

We denote by $\boldsymbol{\rho}_{\text{all}}$ the joint state of all qubit pairs before the measurements and by $\boldsymbol{\rho}^{(r)}_{\text{all}}$ the joint state of the qubit pairs conditioned on the measurement outcome r. In this case, the state of the nth pair is

$$\boldsymbol{\rho}^{(r)}_n = \text{Tr}_{i\in\mathcal{N}\backslash\{n\}} \boldsymbol{\rho}^{(r)}_{\text{all}}. \tag{4}$$

Because measurements collapse quantum states, the nodes estimate the average fidelity of unsampled pairs conditioned on the measurement outcome r, i.e.,

$$\bar{f} = \frac{1}{N-M} \sum_{n\in\mathcal{N}\backslash\mathcal{M}} \langle\Psi^-|\boldsymbol{\rho}^{(r)}_n|\Psi^-\rangle. \tag{5}$$

The imperfection of the measurements is considered as part of the noise. Based on the measurement outcome, the estimator $\mathcal{D}$ is used to estimate the average fidelity of the unsampled qubit pairs $\bar{f}$, i.e.,

$$\check{f} = \mathcal{D}(r). \tag{6}$$

The nodes target at minimizing the mean-square error of the estimated fidelity. Because the set of sampled qubit pairs $\mathcal{M}$ is drawn from all M subsets of $\mathcal{N}$ with equal probability, each M subset is selected with probability $\binom{N}{M}^{-1}$. Therefore, the mean-square estimation error is given by

$$\binom{N}{M}^{-1} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{F})^2], \tag{7}$$

where $\mathbb{E}_R$ is the expectation averaged over all possible values of R. Since the factor $\binom{N}{M}^{-1}$ does not affect the optimization of the measurement protocol $\{\mathcal{O}, \mathcal{D}\}$, it is omitted in the following problem formulation for clarity.

To ensure the robustness of the fidelity estimation protocol, we consider the scenario with arbitrary noise and no prior information. Since there is arbitrary noise, the state of all qubit pairs $\boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{arb}}$, where $\mathcal{S}_{\text{arb}}$ is the set of all N qubit-pair states. Since there is no prior information, the state $\boldsymbol{\rho}_{\text{all}}$ that leads to the largest error must be considered. This optimization problem is formulated as follows.

Problem 1. This problem regards the error minimization for arbitrary states,

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \ \max_{\boldsymbol{\rho}_{\text{all}}\in\mathcal{S}_{\text{arb}}} \binom{N}{M}^{-1} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{F})^2] \tag{8a}$$

$$\text{subject to} \ \sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{F}] = 0 \, \forall \, \boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{arb}}, \tag{8b}$$

where $\check{F}$, $\bar{F}$, and R are the random variable forms of the estimated fidelity $\check{f}$, the average fidelity $\bar{f}$, and the measurement outcome r, respectively, and (8b) is the unbiased constraint of the estimate.

The following two sections elaborate the key procedures and ideas for solving Problem 1, which involves two main steps, namely, problem transformation and operation construction.

III. PROBLEM TRANSFORMATION

This step shows that Problem 1 can be transformed into an equivalent problem with independent noise. It consists of three substeps, i.e., steps A1, A2, and A3. The detailed derivations of these three substeps are in Appendixes A 1–A 3, respectively.

Step A1. This substep simplifies Problem 1 to an equivalent one with classical correlated noise.

Problem 2. This is a special case of Problem 1, with $\mathcal{S}_{\text{arb}}$ replaced by $\mathcal{S}_{\text{sp}}$, i.e., the set of $\boldsymbol{\rho}_{\text{all}}$ that are separable among all qubit pairs.

The key concept is to construct an operation $\mathcal{T}$ that removes the entanglement among different qubit pairs without changing the fidelity. Specifically, the probabilistic rotation $\mathcal{T}$ is defined below.

Definition 1 (probabilistic bilateral rotation). The operation $\mathcal{T}$ acts independently on each qubit pair. For each pair, $\mathcal{T}$ rotates both qubits first along the x axis for 180° with probability 0.5 and then along the y axis for 180° with probability 0.5. The Kraus operators of the first and second probabilistic rotations are

$$\boldsymbol{T}_{x,1} = \frac{\mathbb{I}_4}{\sqrt{2}}, \quad \boldsymbol{T}_{x,2} = -\frac{\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x}{\sqrt{2}}, \tag{9a}$$

$$\boldsymbol{T}_{y,1} = \frac{\mathbb{I}_4}{\sqrt{2}}, \quad \boldsymbol{T}_{y,2} = -\frac{\boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y}{\sqrt{2}}, \tag{9b}$$

respectively.

For a Bell state $|\phi\rangle \in \{|\Phi^{\pm}\rangle, |\Psi^{\pm}\rangle\}$,

$$\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x|\phi\rangle = \begin{cases} |\phi\rangle & \text{if } |\phi\rangle \in \{|\Phi^+\rangle, |\Psi^+\rangle\} \\ -|\phi\rangle & \text{if } |\phi\rangle \in \{|\Phi^-\rangle, |\Psi^-\rangle\}, \end{cases} \tag{10a}$$

$$\boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y|\phi\rangle = \begin{cases} |\phi\rangle & \text{if } |\phi\rangle \in \{|\Phi^-\rangle, |\Psi^+\rangle\} \\ -|\phi\rangle & \text{if } |\phi\rangle \in \{|\Phi^+\rangle, |\Psi^-\rangle\}. \end{cases} \tag{10b}$$

According to (9) and (10), it can be shown that

$$\mathcal{T}(|\phi\rangle\langle\psi|) = \begin{cases} |\phi\rangle\langle\psi| & \text{if } |\phi\rangle = |\psi\rangle \quad \text{(11a)} \\ \boldsymbol{0} & \text{if } |\phi\rangle \neq |\psi\rangle \quad \text{(11b)} \end{cases}$$

Equation (11a) ensures that $\mathcal{T}$ does not change the fidelity of the qubit pairs. Equation (11b) shows that $\mathcal{T}$ removes the off-diagonal terms of a density matrix expressed in the Bell basis. Such a diagonal density matrix corresponds to states separable among all qubit pairs. Therefore, operation $\mathcal{T}$ transforms the states in $\mathcal{S}_{\rm arb}$ to those in $\mathcal{S}_{\rm sp}$. Then the transformed state can be estimated by the optimal solution of Problem 2. This approach ensures that the minimum estimation error for states in $\mathcal{S}_{\rm arb}$ is no higher than that for states in $\mathcal{S}_{\rm sp}$. Moreover, because $\mathcal{S}_{\rm arb} \supset \mathcal{S}_{\rm sp}$, the minimum estimation errors must be the same for the two cases. This result shows the equivalence of Problems 1 and 2. Lemma 1 provides the formal statement of the above analysis.

Step A2. This substep shows that to solve Problem 2, it is sufficient to consider the special case with independent noise, i.e., the following problem.

Problem 3. This is a special case of Problem 2, with $\mathcal{S}_{\rm sp}$ replaced by $\mathcal{S}_{\rm id}$, i.e., the set of $\boldsymbol{\rho}_{\rm all}$ that are products of entangled bipartite states.

We denote by $\check{f}(r)$ the estimated fidelity given the measurement outcome r and by $\bar{f}(\boldsymbol{\rho}_{\rm all}, \mathcal{M}, r)$ the average fidelity given the state $\boldsymbol{\rho}_{\rm all}$, the sample set $\mathcal{M}$, and the measurement outcome r. Separable states $\boldsymbol{\rho}_{\rm all} \in \mathcal{S}_{\rm sp}$ are ensembles of product states, i.e., $\boldsymbol{\rho}_{\rm all} = \sum_k p_k \boldsymbol{\rho}_{\rm all}^{(k)}$, where $\boldsymbol{\rho}_{\rm all}^{(k)} \in \mathcal{S}_{\rm id}$, ensemble probabilities p_k satisfy $p_k \geqslant 0$, and $\sum_k p_k = 1$. Consequently, for each set $\mathcal{M}$, the estimation error with a separable state $\boldsymbol{\rho}_{\rm all}$ and that with product states $\{\boldsymbol{\rho}_{\rm all}^{(k)}\}$ are related by the inequality

$$\begin{align}
&\mathbb{E}_R[(\check{F}-\bar{F})^2|\boldsymbol{\rho}_{\rm all},\mathcal{M}] \nonumber\\
&\quad= \sum_r \Pr(r|\mathcal{M})[\check{f}(r)-\bar{f}(\boldsymbol{\rho}_{\rm all},\mathcal{M},r)]^2 \tag{12a}\\
&\quad= \sum_r \Pr(r|\mathcal{M})\left(\check{f}(r)-\sum_k \Pr\big(\boldsymbol{\rho}_{\rm all}^{(k)}\big|\mathcal{M},r\big)\bar{f}\big(\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M},r\big)\right)^2 \tag{12b}\\
&\quad\leqslant \sum_{r,k} \Pr(r|\mathcal{M})\Pr\big(\boldsymbol{\rho}_{\rm all}^{(k)}\big|\mathcal{M},r\big)\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M},r\big)\big]^2 \tag{12c}\\
&\quad= \sum_k p_k \sum_r \Pr\big(r\big|\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M}\big)\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M},r\big)\big]^2 \tag{12d}\\
&\quad= \sum_k p_k \mathbb{E}_R\big[(\check{F}-\bar{f})^2|\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M}\big], \tag{12e}
\end{align}$$

where $\Pr(\cdot)$ denotes the probability; Eq. (12c) is true because $(\sum_k w_k x_k)^2 \leqslant \sum_k w_k x_k^2 \,\forall x_k \in \mathbb{R}$, $w_k \geqslant 0$, and $\sum_k w_k = 1$; Eq. (12d) holds because according to Bayes' theorem

$$\begin{align}
\Pr(r|\mathcal{M})\Pr\big(\boldsymbol{\rho}_{\rm all}^{(k)}\big|\mathcal{M},r\big) &= \Pr\big(\boldsymbol{\rho}_{\rm all}^{(k)}\big|\mathcal{M}\big)\Pr\big(r\big|\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M}\big) \nonumber\\
&= \Pr\big(\boldsymbol{\rho}_{\rm all}^{(k)}\big)\Pr\big(r\big|\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M}\big) \nonumber\\
&= p_k \Pr\big(r\big|\boldsymbol{\rho}_{\rm all}^{(k)},\mathcal{M}\big); \tag{13}
\end{align}$$

and the lowercase letter $\bar{f}$ is used in (12e) to represent the average fidelity of the unsampled qubit pairs because $\bar{f}$ is deterministic given the product state $\boldsymbol{\rho}_{\rm all}^{(k)}$ and the sample set $\mathcal{M}$.

Because $\boldsymbol{\rho}_{\rm all} \in \mathcal{S}_{\rm sp}$ and $\boldsymbol{\rho}_{\rm all}^{(k)} \in \mathcal{S}_{\rm id} \,\forall\, k$, Eq. (12) shows that the minimum estimation error of Problem 2 is upper bounded by that of Problem 3. Moreover, because $\mathcal{S}_{\rm sp} \supset \mathcal{S}_{\rm id}$, the minimum estimation errors must be the same for the two cases. This result shows the equivalence of Problems 2 and 3. Lemma 2 provides the formal statement of the above analysis.

Remark 1 (limit the effect of correlation via postselection). In previous studies that made estimates with correlated noise, the estimation targets, e.g., the length of the quantum secret keys [25] and the capacity of the quantum channels [26], were evaluated conditioned on the error rates of the measurements being below certain thresholds. Such conditional evaluation is important because it postselects the quantum states according to the measurement outcome, which increases the effectiveness of the measurements and limits the negative effects of correlation.

The result of step A2 is consistent with the above studies. By evaluating the average fidelity conditioned on the exact value of the measurements outcome r, we bound the estimation errors in scenarios with correlated noise by those with independent noise. Specifically, the conditional distribution $\Pr(\boldsymbol{\rho}_{\rm all}^{(k)}|\mathcal{M}, r)$ in (12c) represents the postselection effect. This conditional probability enables the application of Bayes' theorem in (13), thereby bounding the estimation error with the states in $\mathcal{S}_{\rm sp}$ by that with the states in $\mathcal{S}_{\rm id}$.

Step A3. This substep shows that to solve Problem 3 it is sufficient to minimize the estimation error of the sampled pairs, i.e., the following problem.

Problem 4. This problem regards the error minimization for sampled qubit pairs,

$$\begin{aligned}
&\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \max_{\boldsymbol{\rho}_{\rm all}\in\mathcal{S}_{\rm id}(\boldsymbol{f}_{\rm all})} \binom{N}{M}^{-1} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F}-\bar{f}_{\mathcal{M}})^2]\\
&\text{subject to } \mathbb{E}_R[\check{F}-\bar{f}_{\mathcal{M}}] = 0 \,\forall\, \boldsymbol{\rho}_{\mathcal{M}},
\end{aligned}$$

where $\mathcal{S}_{\rm id}(\boldsymbol{f}_{\rm all})$ is the set of products of N entangled bipartite states with fidelity composition $\boldsymbol{f}_{\rm all} = \{f_n, n \in \mathcal{N}\}$, in which f_n is the fidelity of the nth qubit pair, $\bar{f}_{\mathcal{M}}$ is the average fidelity of the sampled qubit pairs, and $\boldsymbol{\rho}_{\mathcal{M}}$ is the state of all sampled qubit pairs.

In the case of independent noise, the measurement does not affect the fidelity of the unsampled qubit pairs. Consequently, the estimation error of Problem 3 can be decomposed into two parts, namely, the estimation error of the sampled qubit pairs

$$\binom{N}{M}^{-1} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F}-\bar{f}_{\mathcal{M}})^2] \tag{14}$$

and the sampling error, i.e., the deviation between the average fidelity of the sampled and the unsampled qubit pairs

$$\binom{N}{M}^{-1} \sum_{\mathcal{M}} (\bar{f}_{\mathcal{M}}-\bar{f})^2. \tag{15}$$

The sampling error (15) is not affected by the estimation protocol. Therefore, Problem 3 can be simplified to Problem 4, which minimizes (14). Lemma 3 provides the formal statement of the above analysis and Theorem 1 summarizes the results of this section.

IV. OPERATION CONSTRUCTION

The next step is to construct the optimal solution of Problem 1. This step consists of two substeps, i.e., steps B1 and

B2. The detailed derivations of these two substeps are in Appendixes B 1 and B 2, respectively.

Step B1. This substep characterizes the minimum estimation error of Problem 4.

The characterization of the minimal estimation error has been a subject of intense research. For Problem 4, the Cramér-Rao bound [30,31] characterizes the lower bound of the estimation error for a given measurement operation $\mathcal{O}$, and the quantum Fisher information [32,33] identifies this lower bound under the condition that all measurement operations are available.

However, Problem 4 aims to minimize the estimation error when all separable measurement operations are available. Therefore, the Cramér-Rao bound alone does not provide the lowest bound, while the quantum Fisher information provides a lower bound that is infeasible. We will close this gap by first further simplifying Problem 3 and then characterizing the limit of separable operators.

The derivation of a Cramér-Rao bound [30,31] requires knowledge of the distribution of a measurement outcome. In Problem 4 this distribution is determined by the state of the qubit pairs $\boldsymbol{\rho}_{\text{all}}$ and the measurement operation $\mathcal{O}$. However, even with independent noise, the state of each qubit pair $\boldsymbol{\rho}_n \in \mathcal{H}^{4\times 4}$ has several parameters other than the fidelity. This fact complicates the expression of the measurement outcome and makes the analysis of the Cramér-Rao bound not feasible. To overcome this challenge, we show that the minimum estimation error of general independent states is bounded below by that of independent Werner states, i.e., the following problem.

Problem 5. This is a special case of Problem 4, with $\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}})$ replaced by $\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}) = \{\otimes^{n\in\mathcal{N}}\boldsymbol{\sigma}_n\}$, where

$$\boldsymbol{\sigma}_n = f_n|\Psi^-\rangle\langle\Psi^-| + \frac{1-f_n}{3}(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|). \qquad (16)$$

Specifically, recall the bilateral rotation operation $\mathcal{B}$ in [34], which transforms generic states of a qubit pair into the Werner state with the same fidelity. Using logic similar to that in step A1, one can see that because the states in $\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}})$ can be transformed into those in $\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}})$ via the separable operation $\mathcal{B}$, the minimum estimation error of Problem 4 is no higher than that of Problem 5. Moreover, because $\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}) \supset \mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}})$, the minimum estimation errors must be the same for the two cases. This result shows the equivalence of Problems 4 and 5. Lemma 4 provides the formal statement of the above analysis. The states in $\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}})$ are fully parametrized by the fidelity composition $\boldsymbol{f}_{\text{all}}$, which allows the Cramér-Rao bound analysis of Problem 5.

The other factor affecting the distribution of measurement outcome is the separable measurement operation $\mathcal{O}$. Since the fidelity of each separable state with respect to a maximally entangled state lies in the interval $[0, \frac{1}{2}]$ [35], we have that for any separable operator $\boldsymbol{M}$,

$$0 \leqslant \frac{\text{Tr}(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M})}{\text{Tr}(\boldsymbol{M})} \leqslant \frac{1}{2}. \qquad (17)$$

For all separable operators, Eq. (17) limits the sensitivity of the measurement outcome to the changes in fidelity. Plugging this result into the Cramér-Rao bound, we lower bound the estimation error of Problem 5 by

$$\sum_{n\in\mathcal{N}} \frac{(2f_n+1)(1-f_n)}{2MN}. \qquad (18)$$

Lemma 5 provides the formal statement about the limit of separable operators and Lemma 6 characterizes the lower bound of the estimation error.

Protocol 1. Fidelity estimation.

1: *Preprocessing.* The nodes select the sample set $\mathcal{M}$ completely at random, i.e., select $\mathcal{M}$ from all M subsets of $\mathcal{N}$ with equal probability, and generate a number M of independent and identically distributed random variables $A_n \in \{x, y, z\}$, $n \in \mathcal{M}$, with distribution $\Pr(A_n = u) = \frac{1}{3},\ u \in \{x, y, z\}$.
2: *Perform measurements.* For the qubit pair $n \in \mathcal{M}$, both nodes measure the qubit in the A_n basis. If the measurement results of the two nodes match, record measurement outcome $r_n = 1$; otherwise record $r_n = 0$.
3: *Estimate fidelity.* The number of errors and the quantum bit error rate (QBER) are expressed as $e_{\mathcal{M}} = \sum_{n\in\mathcal{M}} r_n$ and $\varepsilon_{\mathcal{M}} = \frac{e_{\mathcal{M}}}{M}$, respectively. The estimated fidelity is

$$\check{f} = 1 - \tfrac{3}{2}\varepsilon_{\mathcal{M}}. \qquad (19)$$

Step B2. The second substep constructs an estimation protocol that is the optimal solution of Problem 1.

The analysis in step B1, particularly the proof of Lemma 6, shows that to minimize the estimation error, each measurement operator must balance either of the two inequalities in (17). Therefore, a measurement operator $\boldsymbol{M}$ will either be best aligned with the target state among all separable operators, i.e.,

$$\frac{\text{Tr}(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M})}{\text{Tr}(\boldsymbol{M})} = \frac{1}{2}, \qquad (20)$$

or be orthogonal to the target state, i.e.,

$$\text{Tr}(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M}) = 0. \qquad (21)$$

The measurement operation $\mathcal{O}^*$ of Protocol 1 is constructed according to the above principle. For example, when nodes measure in the z basis, the operators corresponding to $r = 0$ and $r = 1$ are, respectively,

$$\begin{aligned}\boldsymbol{M}_{z,0} &= |01\rangle\langle 01| + |10\rangle\langle 10| = |\Psi^-\rangle\langle\Psi^-| + |\Psi^+\rangle\langle\Psi^+|,\\ \boldsymbol{M}_{z,1} &= |00\rangle\langle 00| + |11\rangle\langle 11| = |\Phi^-\rangle\langle\Phi^-| + |\Phi^+\rangle\langle\Phi^+|. \qquad (22)\end{aligned}$$

According to (22), $\boldsymbol{M}_{z,0}$ and $\boldsymbol{M}_{z,1}$ satisfy (20) and (21), respectively. Moreover, the basis of measurement on each qubit pair is chosen in an independent and identically distributed manner. With these properties, we show that in scenarios with independent noise, the measurement outcome $R_n \in \mathcal{M}$ is independent random variables with variance

$$\frac{2(2f_n+1)(1-f_n)}{9}. \qquad (23)$$

Further noticing that the sample set $\mathcal{M}$ is selected from all M subsets of $\mathcal{N}$ with equal probability, we substitute (23) into

(19) and obtain that the operation $\mathcal{O}^*$ achieves the minimum estimation error given by (18) for states in $\mathcal{S}_{\text{id}}$. Therefore, $\mathcal{O}^*$ is optimal for Problem 4. Lemmas 7 and 8 provide the formal statement of the above analysis. Specifically, Lemma 7 shows that operation $\mathcal{O}^*$ of Protocol 1 is a legitimate solution of Problem 4 and Lemma 8 shows that $\mathcal{O}^*$ is optimal for Problem 4.

Since $\mathcal{O}^*$ is optimal for Problem 4, the composite operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T}$ is optimal for Problem 1 according to the results given in Sec. III. To further simplify the measurement operation, we note that according to (11), the operation $\mathcal{T}$ on each qubit pair can be expressed by the four Kraus operators

$$\boldsymbol{T}_\phi = |\phi\rangle\langle\phi|, \quad \phi \in \{\Phi^\pm, \Psi^\pm\}. \tag{24}$$

Substituting (24) into (22) shows that the operation $\mathcal{T}$ does not change the operators of $\mathcal{O}^*$, e.g.,

$$\boldsymbol{M}_{z,r} = \sum_{\phi\in\{\Phi^\pm,\Psi^\pm\}} \boldsymbol{T}_\phi^\dagger \boldsymbol{M}_{z,r} \boldsymbol{T}_\phi, \quad r \in \{0, 1\}, \tag{25}$$

which shows that $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T} = \mathcal{O}^*$, i.e., $\hat{\mathcal{O}}^*$ and $\mathcal{O}^*$ are equivalent. Therefore, $\mathcal{O}^*$ is optimal for Problem 1. Theorem A2 provides the formal statement of the above analysis.

Remark 2 (the implementability of Protocol 1). It may seem that using bilateral Pauli measurements as in Protocol 1 is a quite natural choice for estimating the fidelity of entangled qubit pairs. In fact, using these standard measurements to obtain optimal performance in scenarios with arbitrary noise is a key advantage of Protocol 1.

To ensure implementability of Protocol 1, simpler operations are chosen with priority. In fact, by using more complicated operations, one can obtain optimal solutions of Problem 1 other than Protocol 1. For example, consider a protocol in which the nodes perform the operation $\mathcal{T}$ defined in (11) before making the Pauli measurements in Protocol 1. Then according to the paragraph after (22), this protocol is also optimal for Problem 1. We have made additional efforts to simplify the measurement operators while preserving the optimality of the protocol.

To achieve optimal performance in scenarios with arbitrary noise, the preprocessing step, i.e., step 1, is developed. The complete randomness of the sampling set $\mathcal{M}$ and the independent and identically distributed distribution of the bases of the measurements are necessary for handling arbitrary noise. For example, when the measurements in the x, y, or z basis are made in a clustered manner, the estimated fidelity is biased or has a higher estimation error for some non–independent and identically distributed noises. The preprocessing in step 1 neutralizes the effect of the arbitrary nose without compromising the implementability of Protocol 1.

Remark 3 (the advantage of single intermediate observable). Given the target state $|\Psi^-\rangle$, the expectation of the observable $|\Psi^-\rangle\langle\Psi^-|$ is the fidelity. However, since the target state is entangled, the observable $|\Psi^-\rangle\langle\Psi^-|$ cannot be realized with separable operators. To address this problem, previous studies on fidelity estimation introduced intermediate observables. For example, to estimate the fidelity of noisy Bell states, Ref. [21] used two observables $|0\rangle\langle 0|^{\otimes 2} + |1\rangle\langle 1|^{\otimes 2}$ and $|0\rangle\langle 1|^{\otimes 2} + |1\rangle\langle 0|^{\otimes 2}$ and [22] used four Pauli observables.

To estimate the value of multiple intermediate observables, the measurements must be split into multiple clusters. In this case, minimizing the estimation error involves both inter- and intracluster designs, which makes it difficult to find an optimal solution.

To overcome the above challenge, Protocol 1 uses only one intermediate variable $\varepsilon_{\mathcal{M}}$, which corresponds to the observable

$$\tfrac{2}{3}(|\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+|). \tag{26}$$

With one intermediate observable, the design of measurement operators on multiple qubit pairs is simplified to that on one pair. Moreover, it can be further simplified by the fact that the measurement operators must satisfy either (20) or (21). In this way, we have found the measurement operators that minimize the estimation error.

V. PROTOCOL EVALUATION

In this section we evaluate the proposed protocol via a demonstrative example.

A. Noise model

The noise is modeled as correlated and heterogeneous depolarizing channels. Specifically, the corresponding density matrix of all qubit pairs is

$$\begin{aligned}\boldsymbol{\rho}_{\text{all}} = \tfrac{1}{2}[(\otimes^{N/4}\boldsymbol{\rho}^{(g)}) \otimes (\otimes^{3N/4}\boldsymbol{\rho}^{(b)}) \\ + (\otimes^{3N/4}\boldsymbol{\rho}^{(g)}) \otimes (\otimes^{N/4}\boldsymbol{\rho}^{(b)})],\end{aligned} \tag{27}$$

where the state of each qubit pair

$$\boldsymbol{\rho}^{(s)} = p^{(s)}\frac{\mathbb{I}_4}{4} + (1 - p^{(s)})|\Psi^-\rangle\langle\Psi^-|, \quad s \in \{g, b\} \tag{28}$$

in which the error probabilities $p^{(s)}$, $s \in \{g, b\}$, of good and bad channels satisfy $0 \leqslant p^{(g)} \leqslant p^{(b)} \leqslant 1$ and N is a multiple of 4.

The mean of the error probabilities of the bad and good channels, i.e.,

$$p = \frac{p^{(b)} + p^{(g)}}{2} \in [0, 1], \tag{29}$$

represents the noisy intensity, and the difference between the two probabilities, i.e.,

$$d = p^{(b)} - p^{(g)} \in [0, 1], \tag{30}$$

represents the degree of correlation and heterogeneity of the noise. In particular, the noise is independent and identically distributed when $d = 0$. ■

B. Computation method

Given the noise model, the fidelity of a qubit pair with good or bad channels is given by

$$f^{(s)} = 1 - \tfrac{3}{4}p^{(s)}, \quad s \in \{g, b\}. \tag{31}$$

Defining the number of unsampled qubit pairs with good or bad channels as

$$N^{(s)} = \sum_{n\in\mathcal{N}\setminus\mathcal{M}} \mathbb{1}(\boldsymbol{\rho}_n = \boldsymbol{\rho}^{(s)}), \quad s \in \{g, b\}, \tag{32}$$

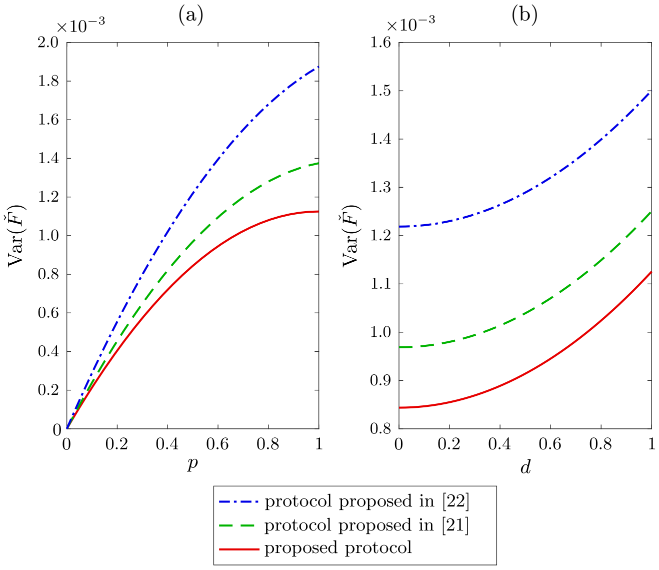

FIG. 1. Variance of the estimated fidelity given by different protocols as a function of the noisy intensity p and the degree of correlation and heterogeneity d for $N = 1000$, $M = 500$, and (a) $d = 0$, i.e., the noise is independent and identically distributed, and (b) $p = 0.5$.

then the average fidelity of the unsampled qubit pairs $\bar{F}$ is determined by $F^{(s)}$ and $N^{(s)}$, $s \in \{g, b\}$, as

$$\bar{F} = \frac{f^{(g)}N^{(g)} + f^{(b)}N^{(b)}}{N - M}. \tag{33}$$

Furthermore, according to (19), the estimated fidelity $\check{F}$ is determined the QBER $\mathcal{E}_{\mathcal{M}}$. Hence, given $N^{(g)}$, $N^{(b)}$, and $\mathcal{E}_{\mathcal{M}}$, the variables of interest, i.e., $\bar{F}$ and $\check{F}$, are determined.

Here $N^{(g)}$, $N^{(b)}$, and $\mathcal{E}_{\mathcal{M}}$ are discrete random variables with finite support. Therefore, their exact joint distribution can be computed in principle. To avoid excessive memory requirements, we ignore realizations with probability less than 10^{-10} in the computation. The probabilities of the ignored realizations have a sum no higher than 10^{-8}, which upper bounds the computational error.

Figure 1 shows the mean-square estimation error, i.e., $\mathrm{Var}(\check{F})$, as a function of the noisy intensity p and the degree of correlation and heterogeneity d. The proposed protocol is compared with those proposed in [21,22]. This figure shows that the estimation error is an increasing function of both p and d. In all cases, the proposed protocol has the lowest estimation error. This result is consistent with Remark 3, which explains how Protocol 1 minimizes the estimation error. Because the minimum estimation error is achieved, Protocol 1 performs better than protocols that are not fully optimized.

VI. CONCLUSION

In this paper we proposed a protocol to estimate the fidelity of entangled qubit pairs shared by remote nodes. The proposed protocol used only Pauli measurements and achieved the minimum mean-square error in a challenging scenario with arbitrary noise and no prior information. Numerical tests confirmed the efficiency and reliability of the proposed protocol.

ACKNOWLEDGMENTS

The author would like to thank Stephanie Wehner and Bas Dirkse at TUDelft, The Netherlands, for refining the structure of the paper and stimulating discussions on the treatment of non–independent and identically distributed noise. The author is also indebted to Wenhan Dai at UMass, USA, who distilled the logic of the derivation and provided feedback on the manuscript. The author is grateful to the EU Flagship on Quantum Technologies and the Quantum Internet Alliance (funded by the EU Horizon 2020 Research and Innovation Programme, Grant No. 820445).

APPENDIX A: PROBLEM TRANSFORMATION

Recall the problem of minimizing the mean-square estimation error of fidelity in scenarios with arbitrary noise, i.e., Problem 1, formulated in (8a) and (8b). In this Appendix we prove that in order to solve Problem 1, it suffices to consider the scenario with independent noise.

1. Sufficiency of considering classically correlated noise

We first show that to solve Problem 1, it is sufficient to consider a problem with no entanglement among different qubit pairs. This problem is given by Problem 2.

Problem 2 (error minimization for separable states). This problem is to

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \ \max_{\boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{sp}}} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{F})^2] \tag{A1a}$$

$$\text{subject to} \ \sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{F}] = 0 \, \forall \, \boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{sp}}, \tag{A1b}$$

where $\mathcal{S}_{\text{sp}}$ denotes the set of states $\boldsymbol{\rho}_{\text{all}}$ that are separable among all qubit pairs.

We construct a probabilistic rotation operation as given by Definition 1. By applying operation $\mathcal{T}$, the following lemma shows the equivalence of Problems 1 and 2.

Lemma 1 (equivalence of Problems 1 and 2). If a measurement operation $\mathcal{O}^*$ is optimal in Problem 2, the composite operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T}$ is optimal in Problem 1.

Proof. The four Bell states $\{|\Phi^{\pm}\rangle, |\Psi^{\pm}\rangle\}$ form a basis for a qubit pair. Therefore, the state of each qubit pair $n \in \mathcal{N}$ can be written as

$$\boldsymbol{\rho}_n = \sum_{k \in \mathcal{K}_n} c_{k,n} |\phi_k\rangle\langle\psi_k|, \tag{A2}$$

where $|\phi_k\rangle$ and $|\psi_k\rangle$ are Bell states, i.e., $|\Phi^{\pm}\rangle$ and $|\Psi^{\pm}\rangle$, and $\mathcal{K}_n$ is the set of all terms with nonzero coefficient $c_{k,n} \in \mathbb{C}$. Given (A2), the fidelity of qubit pair n is given by

$$f_n = \sum_{k \in \mathcal{K}_n} [c_{k,n} \mathbb{1}(|\phi_k\rangle = |\psi_k\rangle = |\Psi^-\rangle)], \tag{A3}$$

where $\mathbb{1}(\cdot)$ is the indicator function.

According to (11), for Bell states $|\phi\rangle, |\psi\rangle \in \{|\Phi^{\pm}\rangle, |\Psi^{\pm}\rangle\}$,

$$\mathcal{T}(|\phi\rangle\langle\psi|) = \begin{cases} |\phi\rangle\langle\psi| & \text{if } |\phi\rangle = |\psi\rangle \\ 0 & \text{if } |\phi\rangle \neq |\psi\rangle. \end{cases} \tag{A4}$$

By substituting (A4) into (A3), it is clear that the rotation $\mathcal{T}$ does not change the fidelity of any qubit pair.

Consider a composite operation in which the rotation $\mathcal{T}$ is performed on the sampled qubit pairs $n \in \mathcal{M}$ before the measurement $\mathcal{O}$. Subsequently, this rotation is performed on the unsampled qubit pairs $n \in \mathcal{N}\backslash\mathcal{M}$. According to the analysis described in the preceding paragraph, the rotation $\mathcal{T}$ on the unsampled qubit pairs does not affect the measurement outcome $\boldsymbol{r}$ or the average fidelity of the unsampled qubit pairs $\bar{f}$. In this sense, the considered composite operation is equivalent to the operation $\hat{\mathcal{O}} = \mathcal{O} \circ \mathcal{T}$, i.e., both the rotation and the measurement are performed only on the sampled qubit pairs. Moreover, since the rotation $\mathcal{T}$ on the unsampled qubit pairs is commutable with the measurement $\mathcal{O}$, the operation $\hat{\mathcal{O}}$ is also equivalent to an operation in which the rotation $\mathcal{T}$ is performed on all qubit pairs before the measurement $\mathcal{O}$. Given the above two equivalences, when analyzing the performance of the operation $\hat{\mathcal{O}}$, we consider the operation in which the rotation is performed on all qubit pairs before the measurement.

An arbitrary state of all N qubit pairs can be expressed in the Bell basis as

$$\boldsymbol{\rho}_{\text{all}} = \sum_{k\in\mathcal{K}} c_k \otimes^{n\in\mathcal{N}} |\phi_{k,n}\rangle\langle\psi_{k,n}|, \tag{A5}$$

where $|\phi_{k,n}\rangle$ and $|\psi_{k,n}\rangle$ are Bell states, i.e., $|\Phi^{\pm}\rangle$ and $|\Psi^{\pm}\rangle$, and $\mathcal{K}$ is the set of all terms with nonzero coefficient $c_k \in \mathbb{C}$. Substituting (A4) into (A5), we find that after performing the rotation $\mathcal{T}$ on all qubit pairs, their joint state is

$$\mathcal{T}(\boldsymbol{\rho}_{\text{all}}) = \sum_{k\in\mathcal{K}} c_k \otimes^{n\in\mathcal{N}} \mathbb{1}(|\phi_{k,n}\rangle = |\psi_{k,n}\rangle)|\phi_{k,n}\rangle\langle\psi_{k,n}|. \tag{A6}$$

Equation (A6) shows that after performing $\mathcal{T}$ on all qubit pairs, the state of all qubit pairs is a composition of product states and therefore separable.

We define the function $e(\mathcal{S}, \mathcal{O}, \mathcal{D})$ as the worst-case-square estimation error for states in the set $\mathcal{S}$ by using the measurement operator $\mathcal{O}$ and the estimator $\mathcal{D}$, i.e.,

$$e(\mathcal{S}, \mathcal{O}, \mathcal{D}) = \max_{\boldsymbol{\rho}_{\text{all}}\in\mathcal{S}} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{F})^2|\mathcal{M}, \mathcal{O}, \mathcal{D}]. \tag{A7}$$

The objective functions of Problems 1 and 2 can be rewritten as

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{arb}}, \mathcal{O}, \mathcal{D}), \quad \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{sp}}, \mathcal{O}, \mathcal{D}), \tag{A8}$$

respectively.

We denote by $\mathcal{D}^*$ the estimator used with the measurement operation $\mathcal{O}^*$ to achieve the minimum mean-square error. The following inequalities involving the objective functions of Problems 1 and 2 can be derived:

$$\begin{aligned} &\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{arb}}, \mathcal{O}, \mathcal{D}) \\ &\leqslant e(\mathcal{S}_{\text{arb}}, \hat{\mathcal{O}}^*, \mathcal{D}^*) && \text{(A9a)} \\ &\leqslant e(\mathcal{S}_{\text{sp}}, \mathcal{O}^*, \mathcal{D}^*) && \text{(A9b)} \\ &= \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{sp}}, \mathcal{O}, \mathcal{D}). && \text{(A9c)} \end{aligned}$$

Equation (A9b) holds because rotation $\mathcal{T}$ converts an arbitrary state to a separable state and (A9c) holds because $(\mathcal{O}^*, \mathcal{D}^*)$ is the optimal solution to Problem 2.

Moreover, because $\mathcal{S}_{\text{arb}} \supset \mathcal{S}_{\text{sp}}$,

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{arb}}, \mathcal{O}, \mathcal{D}) \geqslant \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{sp}}, \mathcal{O}, \mathcal{D}). \tag{A10}$$

According to (A9) and (A10), we can obtain

$$\begin{aligned} \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{arb}}, \mathcal{O}, \mathcal{D}) &= e(\mathcal{S}_{\text{arb}}, \hat{\mathcal{O}}^*, \mathcal{D}^*) \\ &= \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{sp}}, \mathcal{O}, \mathcal{D}). \qquad \text{(A11)} \end{aligned}$$

The first equality in (A11) shows that $\hat{\mathcal{O}}^*$ is optimal for Problem 1. This completes the proof of Lemma 1. ■

2. Sufficiency of considering independent noise

We next show that to solve Problem 2, it is sufficient to consider a problem in which there is no correlation between different qubit pairs. This problem is given by Problem 3.

Problem 3 (error minimization for independent states). This problem is to

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \max_{\boldsymbol{\rho}_{\text{all}}\in\mathcal{S}_{\text{id}}} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f})^2] \tag{A12a}$$

$$\text{subject to} \sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{f}] = 0\, \forall\, \boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{id}}, \tag{A12b}$$

where $\mathcal{S}_{\text{id}}$ denotes the set of states $\boldsymbol{\rho}_{\text{all}}$ in which the state $\boldsymbol{\rho}_n$ of every qubit pair n, $n \in \mathcal{N}$, is independent, i.e., $\boldsymbol{\rho}_{\text{all}}$ can be expressed as

$$\boldsymbol{\rho}_{\text{all}} = \otimes^{n\in\mathcal{N}} \boldsymbol{\rho}_n. \tag{A13}$$

In the case of (A13), the average fidelity of the unsampled pairs $\bar{f}$ is deterministic for each sample set $\mathcal{M}$. Thus, lowercase letter $\bar{f}$ is used in (A12).

The following lemma shows the sufficiency of considering independent noise.

Lemma 2 (equivalence of Problems 2 and 3). If a measurement operation $\mathcal{O}^*$ is optimal in Problem 3, it is also optimal in Problem 2.

Proof. A separable state $\boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{sp}}$ can be expressed as

$$\boldsymbol{\rho}_{\text{all}} = \sum_{k\in\mathcal{K}} p_k \otimes^{n\in\mathcal{N}} \boldsymbol{\rho}_{k,n}, \tag{A14}$$

where p_k is the probability of the ensemble in case k, with $p_k \geqslant 0$, $\sum_{k\in\mathcal{K}} p_k = 1$, and $\boldsymbol{\rho}_{k,n}$ is the density matrix of qubit pair n in case k. We define $\boldsymbol{\rho}_{\text{all}}^{(k)} = \otimes^{n\in\mathcal{N}} \boldsymbol{\rho}_{k,n}$ and denote the fidelity of $\boldsymbol{\rho}_{k,n}$ by $f_{k,n}$. Then $\boldsymbol{\rho}_{\text{all}}^{(k)} \in \mathcal{S}_{\text{id}}\, \forall\, k \in \mathcal{K}$.

We denote by $\check{f}(r)$ the estimated fidelity given the measurement outcome r and by $\bar{f}(\boldsymbol{\rho}_{\text{all}}, \mathcal{M}, r)$ the average fidelity of the unsampled qubit pairs given the state $\boldsymbol{\rho}_{\text{all}}$, sample set $\mathcal{M}$, and measurement outcome r. With (A14), for every estimation protocol $\{\mathcal{O}, \mathcal{D}\}$ and every state $\boldsymbol{\rho}_{\text{all}} \in \mathcal{S}_{\text{sp}}$, we have that

$$\begin{aligned} &\sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{F}] \\ &\quad = \sum_{\mathcal{M}} \sum_{r} \Pr(r|\mathcal{M})[\check{f}(r) - \bar{f}(\boldsymbol{\rho}_{\text{all}}, \mathcal{M}, r)] \qquad \text{(A15a)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathcal{M}}\sum_{r}\sum_{k}\Pr(r|\mathcal{M})\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)\\
&\quad\times\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M},r\big)\big] && \text{(A15b)}\\
&= \sum_{k}p_k\sum_{\mathcal{M}}\sum_{r}\Pr\big(r\big|\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M}\big)\\
&\quad\times\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M},r\big)\big] && \text{(A15c)}\\
&= \sum_{k}p_k\sum_{\mathcal{M}}\mathbb{E}_R\big[\check{F}-\bar{f}\big|\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big] && \text{(A15d)}
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\sum_{\mathcal{M}}\mathbb{E}_R[(\check{F}-\bar{F})^2]\\
&= \sum_{\mathcal{M}}\sum_{r}\Pr(r|\mathcal{M})\bigg(\check{f}(r)-\sum_{k}\Pr\\
&\quad\times\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)\bar{f}\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M},r\big)\bigg)^2 && \text{(A16a)}\\
&\leqslant \sum_{\mathcal{M}}\sum_{r}\sum_{k}\Pr(r|\mathcal{M})\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)\\
&\quad\times\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M},r\big)\big]^2 && \text{(A16b)}\\
&= \sum_{k}\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M}\big)\sum_{\mathcal{M}}\sum_{r}\Pr\big(r\big|\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M}\big)\\
&\quad\times\big[\check{f}(r)-\bar{f}\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M},r\big)\big]^2 && \text{(A16c)}\\
&= \sum_{k}p_k\sum_{\mathcal{M}}\mathbb{E}_R\big[(\check{F}-\bar{f})^2\big|\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big] && \text{(A16d)}\\
&\leqslant \max_{\boldsymbol{\rho}_{\mathrm{all}}\in\mathcal{S}_{\mathrm{id}}}\sum_{\mathcal{M}}\mathbb{E}_R[(\check{F}-\bar{f})^2], && \text{(A16e)}
\end{aligned}
$$

where (A15b) and (A16b) hold because

$$
\sum_{k\in\mathcal{K}}\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)=1, \tag{A17}
$$

$$
\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)\geqslant 0\,\forall\,k\in\mathcal{K}, \tag{A18}
$$

and the function x^2 is convex; Eqs. (A15c) and (A16c) hold because according to Bayes' theorem

$$
\Pr(r|\mathcal{M})\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M},r\big)=\Pr\big(\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\big|\mathcal{M}\big)\Pr\big(r\big|\boldsymbol{\rho}_{\mathrm{all}}^{(k)},\mathcal{M}\big); \tag{A19}
$$

and (A16e) holds because $\boldsymbol{\rho}_{\mathrm{all}}^{(k)}\in\mathcal{S}_{\mathrm{id}}$, $p_k\geqslant 0\,\forall\,k\in\mathcal{K}$, and $\sum_k p_k=1$.

According to (A15), if the estimated fidelity $\check{F}$ satisfies (A12b), it also satisfies (A1b). Therefore, an estimate that is unbiased for independent states is also unbiased for separable states.

Recall the worst-case-square estimation error function $e(\mathcal{S},\mathcal{O},\mathcal{D})$ defined in (A7) and denote by $\mathcal{D}^*$ the estimator used with the measurement operation $\mathcal{O}^*$ to achieve the minimum mean-square error. In this case, the following inequalities involving the objective functions of Problems 2 and 3 can be derived:

$$
\begin{aligned}
&\operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{sp}},\mathcal{O},\mathcal{D})\\
&\quad\leqslant e(\mathcal{S}_{\mathrm{sp}},\mathcal{O}^*,\mathcal{D}^*) && \text{(A20a)}\\
&\quad\leqslant e(\mathcal{S}_{\mathrm{id}},\mathcal{O}^*,\mathcal{D}^*) && \text{(A20b)}\\
&\quad= \operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{id}},\mathcal{O},\mathcal{D}). && \text{(A20c)}
\end{aligned}
$$

Equation (A20b) is true because (A16) holds for every estimation protocol $\{\mathcal{O},\mathcal{D}\}$ and every state $\boldsymbol{\rho}_{\mathrm{all}}\in\mathcal{S}_{\mathrm{sp}}$ and (A20c) holds because $\{\mathcal{O}^*,\mathcal{D}^*\}$ is the optimal solution to Problem 3. Moreover, because $\mathcal{S}_{\mathrm{sp}}\supset\mathcal{S}_{\mathrm{id}}$,

$$
\operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{sp}},\mathcal{O},\mathcal{D})\geqslant\operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{id}},\mathcal{O},\mathcal{D}). \tag{A21}
$$

According to (A20) and (A21),

$$
\begin{aligned}
\operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{sp}},\mathcal{O},\mathcal{D}) &= e(\mathcal{S}_{\mathrm{sp}},\mathcal{O}^*,\mathcal{D}^*)\\
&= \operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\, e(\mathcal{S}_{\mathrm{id}},\mathcal{O},\mathcal{D}). && \text{(A22)}
\end{aligned}
$$

The first equality in (A22) shows that $\mathcal{O}^*$ is optimal for Problem 2. This completes the proof of Lemma 2. ■

3. Sufficiency of considering the sampled qubit pairs

In a third step, we show that in the case of independent noise, to minimize the estimation error with respect to the average fidelity of the unsampled qubit pairs, it is sufficient to minimize the corresponding value of the sampled qubit pairs. To this end, transform Problem 3 into Problem 4.

Problem 4 (error minimization for sampled qubit pairs). This problem is to

$$
\operatorname*{minimize}_{\mathcal{O},\mathcal{D}}\ \max_{\boldsymbol{\rho}_{\mathrm{all}}\in\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}})}\sum_{\mathcal{M}}\mathbb{E}_R[(\check{F}-\bar{f}_{\mathcal{M}})^2] \tag{A23a}
$$

$$
\text{subject to } \mathbb{E}_R[\check{F}-\bar{f}_{\mathcal{M}}]=0\,\forall\,\boldsymbol{\rho}_{\mathcal{M}}, \tag{A23b}
$$

where $\boldsymbol{\rho}_{\mathcal{M}}$ is the state of all sampled qubit pairs and $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}})$ denotes the set of N qubit pair states with independent noise and fidelity composition $\boldsymbol{f}_{\mathrm{all}}=\{f_n, n\in\mathcal{N}\}$, i.e.,

$$
\boldsymbol{\rho}_{\mathrm{all}}=\otimes^{n\in\mathcal{N}}\boldsymbol{\rho}_n, \tag{A24}
$$

in which $\boldsymbol{\rho}_n$ is the state of the nth qubit pair,

$$
f_n=\langle\Psi^-|\boldsymbol{\rho}_n|\Psi^-\rangle,\quad n\in\mathcal{N}, \tag{A25}
$$

is the fidelity of the nth qubit pair, and

$$
\bar{f}_{\mathcal{M}}=\frac{1}{M}\sum_{n\in\mathcal{M}}f_n \tag{A26}
$$

is the average fidelity of the sampled qubit pairs.

The following lemma shows that a measurement operation $\mathcal{O}^*$ which is optimal for Problem 4 is also optimal for Problem 3.

Lemma 3 (equivalence of Problems 3 and 4). If a measurement operation $\mathcal{O}^*$ is optimal in Problem 4 for all compositions of fidelity, i.e., for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$, it is also optimal in Problem 3.

Proof. Because the sample set $\mathcal{M}$ is drawn completely at random,

$$\sum_{\mathcal{M}} \bar{f} - \bar{f}_{\mathcal{M}} = 0\,\forall\, \boldsymbol{\rho}_{\mathrm{all}} \in \mathcal{S}_{\mathrm{id}}. \tag{A27}$$

According to (A27), Eq. (A12b) is equivalent to

$$\sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{f}_{\mathcal{M}}] = 0\,\forall\, \boldsymbol{\rho}_{\mathrm{all}} \in \mathcal{S}_{\mathrm{id}}. \tag{A28}$$

Equation (A28) indicates that any estimation protocol $\{\mathcal{O}, \mathcal{D}\}$ satisfies (A23b) for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$ and $\mathcal{M} \subset \mathcal{N}$ also satisfies (A12b). In the following, we prove the converse statement by contradiction.

Suppose there is an estimation protocol $\{\mathcal{O}, \mathcal{D}\}$ that satisfies (A28) but does not satisfy (A23b). In this case, there must exist some $\boldsymbol{\rho}_{\mathcal{M}}$ such that

$$\mathbb{E}_R[\check{F} - \bar{f}_{\mathcal{M}}] \neq 0. \tag{A29}$$

We denote by $m \in \mathcal{M}$ one of the sampled qubit pairs, by l the number of sampled qubit pairs whose state is the same as that of the mth pair, i.e.,

$$\sum_{n\in\mathcal{M}} \mathbb{1}(\boldsymbol{\rho}_n = \boldsymbol{\rho}_m) = l, \tag{A30}$$

and by $\boldsymbol{\rho}^{(l)}$ the state $\boldsymbol{\rho}_{\mathcal{M}}$ that satisfies (A29). Because $m \in \mathcal{M}$, it is clear that $l \geqslant 1$.

Subsequently, we consider the state

$$\boldsymbol{\rho}_{\mathrm{all}} = \boldsymbol{\rho}^{(l)} \otimes (\otimes^{N-M} \boldsymbol{\rho}_m). \tag{A31}$$

With the state defined in (A31), the sampled set $\mathcal{M}$ has at least a number l of qubit pairs in the state $\boldsymbol{\rho}_m$. When the number of sampled qubit pairs in state $\boldsymbol{\rho}_m$ is equal to l, the joint state of the sampled qubit pairs is equal to $\boldsymbol{\rho}^{(l)}$. Thus, according to (A28) and (A29), there must exist a sample set $\tilde{\mathcal{M}}$ such that

$$\mathbb{E}_R[\check{F} - \bar{f}_{\tilde{\mathcal{M}}}] \neq 0, \quad \sum_{n\in\tilde{\mathcal{M}}} \mathbb{1}(\boldsymbol{\rho}_n = \boldsymbol{\rho}_m) = \tilde{l} \geqslant l + 1. \tag{A32}$$

We denote the state $\boldsymbol{\rho}_{\tilde{\mathcal{M}}}$ by $\boldsymbol{\rho}^{(\tilde{l})}$. Next we consider the state

$$\boldsymbol{\rho}_{\mathrm{all}} = \boldsymbol{\rho}^{(\tilde{l})} \otimes (\otimes^{N-M} \boldsymbol{\rho}_m) \tag{A33}$$

and repeat the analysis above. Because $\tilde{l} \geqslant l + 1$, by repeating this analysis for at most $M - l$ times, we can obtain that when $\boldsymbol{\rho}_{\mathcal{M}} = \boldsymbol{\rho}^{(M)} = \otimes^M \boldsymbol{\rho}_m$, the unbiased constraint is not met, i.e.,

$$\mathbb{E}_R[\check{F} - \bar{f}_{\mathcal{M}}] \neq 0. \tag{A34}$$

However, according to (A34), Eq. (A28) does not hold for the state $\boldsymbol{\rho}_{\mathrm{all}} = \otimes^N \boldsymbol{\rho}_m$. This contradiction shows that any estimation protocol $\{\mathcal{O}, \mathcal{D}\}$ that satisfies (A12b) also satisfies (A23b).

Using the strengthened unbiased constraint (A23b), the estimation error in Problem 3 can be decomposed as

$$\begin{aligned}
&\sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f})^2] \\
&= \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}} + \bar{f}_{\mathcal{M}} - \bar{f})^2] && \text{(A35a)} \\
&= \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] + 2\sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})(\bar{f}_{\mathcal{M}} - \bar{f})] \\
&\quad + \sum_{\mathcal{M}} \mathbb{E}_R[(\bar{f}_{\mathcal{M}} - \bar{f})^2] && \text{(A35b)} \\
&= \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] + 2\sum_{\mathcal{M}} \mathbb{E}_R[\check{F} - \bar{f}_{\mathcal{M}} | \boldsymbol{\rho}_{\mathcal{M}}] \\
&\quad \times (\bar{f}_{\mathcal{M}} - \bar{f}) + \sum_{\mathcal{M}} (\bar{f}_{\mathcal{M}} - \bar{f})^2 && \text{(A35c)} \\
&= \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] + \sum_{\mathcal{M}} (\bar{f}_{\mathcal{M}} - \bar{f})^2, && \text{(A35d)}
\end{aligned}$$

where (A35c) holds because in cases with independent noise, the average fidelity of sampled qubit pairs $\bar{f}_{\mathcal{M}}$ and that of unsampled qubit pairs $\bar{f}$ are deterministic given each sample set $\mathcal{M}$; Eq. (A35d) is true due to (A23b).

Equation (A35) decomposes the estimation error into two parts, i.e., the estimation error with respect to the average fidelity of the sampled qubit pairs

$$\sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2], \tag{A36}$$

and the deviation between the average fidelity of sampled and unsampled qubit pairs

$$\sum_{\mathcal{M}} (\bar{f}_{\mathcal{M}} - \bar{f})^2. \tag{A37}$$

The value of (A37) is determined by the fidelity composition of all qubit pairs, i.e.,

$$\boldsymbol{f}_{\mathrm{all}} = \{f_n, n \in \mathcal{N}\}, \tag{A38}$$

and not affected by the estimation protocol $\{\mathcal{O}, \mathcal{D}\}$. Therefore, for every fidelity composition $\boldsymbol{f}_{\mathrm{all}}$, minimizing the estimation error

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \max_{\boldsymbol{\rho}_{\mathrm{all}} \in \mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}})} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f})^2] \tag{A39}$$

is equivalent to

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \max_{\boldsymbol{\rho}_{\mathrm{all}} \in \mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}})} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2]. \tag{A40}$$

Hence, if a protocol $\{\mathcal{O}^*, \mathcal{D}^*\}$ is optimal in Problem 4 for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$, it is also the optimal solution to (A39) for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$. Furthermore, because

$$\mathcal{S}_{\mathrm{id}} = \bigcup_{\boldsymbol{f}_{\mathrm{all}}} \mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}), \tag{A41}$$

protocol $\{\mathcal{O}^*, \mathcal{D}^*\}$ minimizes

$$\max_{\boldsymbol{\rho}_{\mathrm{all}} \in \mathcal{S}_{\mathrm{id}}} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}\,)^2]. \tag{A42}$$

Therefore, $\{\mathcal{O}^*, \mathcal{D}^*\}$ is also an optimal solution to Problem 3. This completes the proof of Lemma 3. ■

The following theorem summarizes the results of this section.

Theorem 1 (generality of optimality with independent noise). If a measurement operation $\mathcal{O}^*$ is optimal in Problem 4 for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$, then a composite measurement operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T}$ is optimal in Problem 1, where the operation $\mathcal{T}$ is defined as in Definition 1.

Proof. The theorem is a direct consequence of Lemmas 1–3. ■

APPENDIX B: OPERATION CONSTRUCTION

In this Appendix we first derive a lower bound of the estimation error and then, based on the conditions for achieving this lower bound, construct an optimal fidelity estimation protocol.

1. Lower bound of the estimation error

We will first further simplify Problem 4 (Lemma 4), then characterize the limit of separable operators (Lemma 5), and finally determine the lowest feasible bound of the estimation error (Lemma 6).

The state of a qubit pair $\boldsymbol{\rho}_n \in \mathcal{H}^{4\times 4}$ has several parameters other than fidelity. This property complicates the Fisher information analysis of Problem 4. To overcome this challenge, we further simplify Problem 4 to an equivalent one, Problem 5, in which the fidelity composition $\boldsymbol{f}_{\text{all}}$ is sufficient to parametrize the state of the qubit pairs.

Problem 5 (error minimization for Werner states). Given that $\boldsymbol{\rho}_{\text{all}} = \otimes^{n\in\mathcal{N}}\boldsymbol{\sigma}_n$,

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}} \sum_{\mathcal{M}} \mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] \tag{B1}$$

$$\text{subject to } \mathbb{E}_R[\check{F} - \bar{f}_{\mathcal{M}}|\mathcal{M}] = 0\,\forall\,\mathcal{M} \subset \mathcal{N}, \tag{B2}$$

where $\boldsymbol{\sigma}_n$ is the Werner state with fidelity f_n, i.e.,

$$\boldsymbol{\sigma}_n = f_n|\Psi^-\rangle\langle\Psi^-| + \frac{1-f_n}{3}(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|). \tag{B3}$$

Lemma 4 (equivalence of Problems 4 and 5). With the optimal estimation protocol $\{\mathcal{O},\mathcal{D}\}$, the objective functions of Problems 4 and 5 have the same value.

Proof. Recall the function of the worst-case-square estimation error $e(\mathcal{S},\mathcal{O},\mathcal{D})$ defined in (A7) and define $\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}) = \{\otimes^{n\in\mathcal{N}}\boldsymbol{\sigma}_n\}$. The objective functions of Problems 4 and 5 can be rewritten as

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}),$$

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}), \tag{B4}$$

respectively.

We denote by $\{\mathcal{O}^*,\mathcal{D}^*\}$ the optimal solution to Problem 5 and we denote the random bilateral rotation operation proposed in [34] by $\mathcal{B}$. We define the composition operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ (\otimes^M\mathcal{B})$. In this case, we have that

$$\begin{aligned}
&\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D})\\
&\leqslant e(\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}),\hat{\mathcal{O}}^*,\mathcal{D}^*) &&\text{(B5a)}\\
&\leqslant e(\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}),\mathcal{O}^*,\mathcal{D}^*) &&\text{(B5b)}\\
&= \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}), &&\text{(B5c)}
\end{aligned}$$

where (B5b) holds because the operation $\mathcal{B}$ transforms a general qubit pair state $\boldsymbol{\rho}_n$ to a Werner state $\boldsymbol{\sigma}_n$ with the same fidelity and (B5c) holds because $\{\mathcal{O}^*,\mathcal{D}^*\}$ is the optimal solution to Problem 2. Moreover, because $\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}) \subset \mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}})$,

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}) \leqslant \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}). \tag{B6}$$

From (B5) and (B6) we have

$$\underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{w}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}) = \underset{\mathcal{O},\mathcal{D}}{\text{minimize}}\, e(\mathcal{S}_{\text{id}}(\boldsymbol{f}_{\text{all}}),\mathcal{O},\mathcal{D}), \tag{B7}$$

which proves Lemma 4. ■

The next lemma characterizes the limit of separable operators when measuring maximally entangled qubit pairs.

Lemma 5 (limit of separable operators). Here $|\Phi\rangle$ is a maximally entangled state of a qubit pair and $\boldsymbol{M} = \boldsymbol{M}^{(A)} \otimes \boldsymbol{M}^{(B)}$ is a separable operator, where $\boldsymbol{M}^{(X)} \succcurlyeq \boldsymbol{0}$, $X \in \{A,B\}$. In this case,

$$0 \leqslant \frac{\text{Tr}(|\Phi\rangle\langle\Phi|\boldsymbol{M}^{(AB)})}{\text{Tr}[(\mathbb{I}_4 - |\Phi\rangle\langle\Phi|)\boldsymbol{M}^{(AB)}]} \leqslant 1. \tag{B8}$$

Proof. We define

$$\boldsymbol{\rho} = \frac{\boldsymbol{M}^{(AB)}}{\text{Tr}(\boldsymbol{M}^{(AB)})}. \tag{B9}$$

Since $\boldsymbol{M}^{(X)} \succcurlyeq \boldsymbol{0}$, $X \in \{A,B\}$, $\boldsymbol{\rho}$ is the density matrix of a separable state of a qubit pair. For this state, its fidelity with respect to every maximally entangled state lies in the interval $[0,\frac{1}{2}]$ [36,37], i.e.,

$$\text{Tr}(|\Phi\rangle\langle\Phi|\boldsymbol{\rho}) \in \left[0,\tfrac{1}{2}\right]. \tag{B10}$$

Consequently,

$$\frac{\text{Tr}(|\Phi\rangle\langle\Phi|\boldsymbol{\rho})}{\text{Tr}[(\mathbb{I}_4 - |\Phi\rangle\langle\Phi|)\boldsymbol{\rho}]} = \frac{\text{Tr}(|\Phi\rangle\langle\Phi|\boldsymbol{\rho})}{1 - \text{Tr}(|\Phi\rangle\langle\Phi|\boldsymbol{\rho})} \in [0,1]. \tag{B11}$$

According to (B9) and (B11), Eq. (B8) is obtained. This completes the proof of Lemma 5. ■

Remark 4 (the role of inequality (B8)). To measure the fidelity with respect to a state $|\Phi\rangle\langle\Phi|$, it is most efficient to use $|\Phi\rangle\langle\Phi|$ and $\mathbb{I}_4 - |\Phi\rangle\langle\Phi|$ as measurement operators to ensure that the distribution of the measurement outcome is maximally sensitive to the changes in fidelity.

Since the target state $|\Psi^-\rangle$ is entangled, separable operations cannot realize the measurement operators mentioned above. To increase their efficiency, separable measurement operators should be designed to best mimic the ideal operator $|\Psi^-\rangle\langle\Psi^-|$. The inequality (B8) characterizes the best alignment between separable operators and the target state. Consequently, Eq. (B8) is the critical constraint that upper bounds the efficiency of separable operators in fidelity estimation.

Based on the above results, the following lemma characterizes a lower bound for the estimation error in Problem 4.

Lemma 6 (lower bound for the estimation error). The mean-square estimation error, i.e., the objective function of Problem 4 divided by $\binom{N}{M}$, is no less than

$$\sum_{n\in\mathcal{N}} \frac{(2f_n+1)(1-f_n)}{2MN}. \tag{B12}$$

Proof. According to Lemma 4, the minimum value of the objective function of Problem 4 is equal to that of Problem 5. Therefore, the following analysis lower bounds the objective function of Problem 5.

In Problem 5, $\boldsymbol{\rho}_{\text{all}} = \otimes^{n\in\mathcal{N}}\boldsymbol{\sigma}_n$, i.e., the qubit pairs are in independent Werner states. In this case, for each sampled set $\mathcal{M}$, the fidelity composition of the sampled qubit pairs,

$$\boldsymbol{f}_{\mathcal{M}} = \{f_n, n\in\mathcal{M}\}, \tag{B13}$$

is the unknown fixed parameter that determines the distribution of the measurement outcome R. According to the unbiased constraint (A23b) and the fact that $F_{\mathcal{M}} = \frac{1}{M}\sum_{n\in\mathcal{M}} f_n$,

$$\frac{\partial \mathbb{E}_R[\check{F}]}{\partial \boldsymbol{f}_{\mathcal{M}}} = \frac{\boldsymbol{1}_M}{M}, \tag{B14}$$

where $\boldsymbol{1}_M$ denotes the $1\times M$ all-1 vector. In this case, according to the Cramér-Rao bound [30,31],

$$\mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] \geqslant \frac{\partial \mathbb{E}_R[\check{F}]}{\partial \boldsymbol{f}_{\mathcal{M}}}[\boldsymbol{I}_R(\boldsymbol{f}_{\mathcal{M}})]^{-1}\left(\frac{\partial \mathbb{E}_R[\check{F}]}{\partial \boldsymbol{f}_{\mathcal{M}}}\right)^T, \tag{B15}$$

where $\boldsymbol{I}_R(\boldsymbol{f}_{\mathcal{M}})$ is the $M\times M$ Fisher information matrix of the measurement outcome R evaluated at point $\boldsymbol{f}_{\mathcal{M}}$. The elements of $\boldsymbol{I}_R(\boldsymbol{f}_{\mathcal{M}})$ are specified as

$$I_{n,m}(\boldsymbol{f}_{\mathcal{M}}) = \mathbb{E}_R\left[\frac{\partial}{\partial f_n}\ln\mathcal{P}(R;\boldsymbol{f}_{\mathcal{M}})\left(\frac{\partial}{\partial f_m}\ln\mathcal{P}(R;\boldsymbol{f}_{\mathcal{M}})\right)^T\right], \tag{B16}$$

where $n, m\in\mathcal{M}$ and $\mathcal{P}(r;\boldsymbol{f}_{\mathcal{M}})$ is the distribution of measurement outcome r given the fidelity composition $\boldsymbol{f}_{\mathcal{M}}$.

According to Proposition 1 in [38], the reciprocal of the diagonal elements of a Fisher information matrix lower bounds the inverse of that matrix, i.e.,

$$[\boldsymbol{I}_R(\boldsymbol{f}_{\mathcal{M}})]^{-1} \succcurlyeq \mathrm{diag}\big(I_{n,n}^{-1}(\boldsymbol{f}_{\mathcal{M}}), n\in\mathcal{M}\big), \tag{B17}$$

where $\succcurlyeq$ denotes the matrix inequality and $\mathrm{diag}(\cdot)$ denotes the diagonal matrix. Substituting (B14) and (B17) into (B15), we get

$$\mathbb{E}_R[(\check{F} - \bar{f}_{\mathcal{M}})^2] \geqslant \frac{1}{M^2}\sum_{n\in\mathcal{M}} I_{n,n}^{-1}(\boldsymbol{f}_{\mathcal{M}}). \tag{B18}$$

Given that the state of qubit pairs $\boldsymbol{\rho}_{\text{all}} = \otimes^{n\in\mathcal{N}}\boldsymbol{\sigma}_n$ and the measurement operator

$$\boldsymbol{M}_r = \sum_k \otimes^{n\in\mathcal{M}}\big(\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big), \tag{B19}$$

the distribution of the measurement outcome can be expressed as

$$\begin{aligned}\mathcal{P}(r;\boldsymbol{f}_{\mathcal{M}}) &= \mathrm{Tr}(\otimes^{n\in\mathcal{M}}\boldsymbol{\sigma}_n\boldsymbol{M}_r)\\ &= \sum_k \mathrm{Tr}\big[\otimes^{n\in\mathcal{M}}\big(\boldsymbol{\sigma}_n\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big)\big]\\ &= \sum_k\prod_{n\in\mathcal{M}}\mathrm{Tr}\big(\boldsymbol{\sigma}_n\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big).\end{aligned} \tag{B20}$$

To characterize the effect of fidelity f_n on the distribution of the measurement outcome, we define

$$\begin{aligned}a_{r,n} &= \sum_k \mathrm{Tr}\big(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big)\\ &\quad\times\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}\big(\boldsymbol{\sigma}_m\boldsymbol{M}^{(A)}_{r,m,k}\otimes\boldsymbol{M}^{(B)}_{r,m,k}\big),\\ b_{r,n} &= \sum_k \mathrm{Tr}\big[(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|)\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big]\\ &\quad\times\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}\big(\boldsymbol{\sigma}_m\boldsymbol{M}^{(A)}_{r,m,k}\otimes\boldsymbol{M}^{(B)}_{r,m,k}\big).\end{aligned} \tag{B21}$$

It is evident that the value of f_n has no effect on $a_{r,n}$ and $b_{r,n}$, i.e.,

$$\frac{\partial a_{r,n}}{\partial f_n} = \frac{\partial b_{r,n}}{\partial f_n} = 0\,\forall\, r, n. \tag{B22}$$

According to (B3) and (B20), the distribution of the measurement outcome can be rewritten as

$$\begin{aligned}\mathcal{P}(r|\boldsymbol{f}_{\mathcal{M}}) &= f_n\sum_k\mathrm{Tr}\big(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big)\\ &\quad\times\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}\big(\boldsymbol{\sigma}_m\boldsymbol{M}^{(A)}_{r,m,k}\otimes\boldsymbol{M}^{(B)}_{r,m,k}\big)\\ &\quad+\frac{1-f_n}{3}\sum_k\mathrm{Tr}\big[(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|)\boldsymbol{M}^{(A)}_{r,n,k}\\ &\quad\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big]\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}\big(\boldsymbol{\sigma}_m\boldsymbol{M}^{(A)}_{r,m,k}\otimes\boldsymbol{M}^{(B)}_{r,m,k}\big)\\ &= f_n a_{r,n} + \frac{1-f_n}{3}b_{r,n}.\end{aligned} \tag{B23}$$

According to (B16), (B22), and (B23),

$$\begin{aligned}I_{n,n}(\boldsymbol{f}_{\mathcal{M}}) &= \sum_r\mathcal{P}(r|f_n)\left(\frac{\partial}{\partial f_n}\ln\mathcal{P}(r|f_n)\right)^2\\ &= \sum_r\frac{\big(a_{r,n}-\frac{1}{3}b_{r,n}\big)^2}{f_n a_{r,n}+\frac{1-f_n}{3}b_{r,n}}.\end{aligned} \tag{B24}$$

According to the property of the positive-operator-valued measure, i.e., $\sum_r\boldsymbol{M}_r = \mathbb{I}_{4^M}$, it can be obtained that

$$\begin{aligned}\sum_r a_{r,n} &= \mathrm{Tr}\left(|\Psi^-\rangle\langle\Psi^-|\otimes(\otimes^{m\in\mathcal{M}\backslash\{n\}}\boldsymbol{\sigma}_m)\sum_r\boldsymbol{M}_r\right)\\ &= \mathrm{Tr}[|\Psi^-\rangle\langle\Psi^-|\otimes(\otimes^{m\in\mathcal{M}\backslash\{n\}}\boldsymbol{\sigma}_m)]\\ &= \mathrm{Tr}(|\Psi^-\rangle\langle\Psi^-|)\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}(\boldsymbol{\sigma}_m)\\ &= 1.\end{aligned} \tag{B25}$$

Similarly,

$$\begin{aligned}\sum_r b_{r,n} &= \mathrm{Tr}(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|)\prod_{m\in\mathcal{M}\backslash\{n\}}\mathrm{Tr}(\boldsymbol{\sigma}_m)\\ &= 3.\end{aligned} \tag{B26}$$

Applying Lemma 5, we get

$$\begin{aligned}0 &\leqslant \mathrm{Tr}\big(|\Psi^-\rangle\langle\Psi^-|\boldsymbol{M}^{(A)}_{r,n,k}\otimes \boldsymbol{M}^{(B)}_{r,n,k}\big)\\ &\leqslant \mathrm{Tr}\big[(\mathbb{I}_4 - |\Psi^-\rangle\langle\Psi^-|)\boldsymbol{M}^{(A)}_{r,n,k}\otimes\boldsymbol{M}^{(B)}_{r,n,k}\big]\,\forall\, r, n, k.\end{aligned} \tag{B27}$$

Substituting (B27) into (B21), we obtain that

$$0 \leqslant a_{r,n} \leqslant b_{r,n}\,\forall\, r, n. \tag{B28}$$

Defining

$$I_{r,n}(a_{r,n}, b_{r,n}) = \frac{\left(a_{r,n} - \frac{1}{3}b_{r,n}\right)^2}{f_n a_{r,n} + \frac{1-f_n}{3}b_{r,n}}, \tag{B29}$$

then because

$$\frac{\partial^2 I_{r,n}}{\partial^2 a_{r,n}} = \frac{6b^2_{r,n}}{[(1-f_n)b_{r,n} + 3f_n a_{r,n}]^3} \geqslant 0, \tag{B30}$$

the function $I_{r,n}$ is convex with respect to $a_{r,n}$. Therefore, according to (B28),

$$\begin{aligned}I_{r,n}(a_{r,n}, b_{r,n}) &\leqslant \frac{b_{r,n} - a_{r,n}}{b_{r,n}} I_{r,n}(0, b_{r,n}) + \frac{a_{r,n}}{b_{r,n}} I_{r,n}(b_{r,n}, b_{r,n})\\ &= \frac{1}{3(1-f_n)}(b_{r,n} - a_{r,n}) + \frac{4}{3(2f_n+1)}a_{r,n}.\end{aligned} \tag{B31}$$

Substituting (B25), (B26), and (B31) into (B24), we obtain that

$$\begin{aligned}I_{n,n}(\boldsymbol{f}_{\mathcal{M}}) &= \sum_r I_{r,n}(a_{r,n}, b_{r,n})\\ &\leqslant \frac{1}{3(1-f_n)}\sum_r (b_{r,n} - b_{r,n}) + \frac{4}{3(2f_n+1)}\sum_r b_{r,n}\\ &= \frac{2}{3(1-f_n)} + \frac{4}{3(2f_n+1)}\\ &= \frac{2}{(2f_n+1)(1-f_n)}.\end{aligned} \tag{B32}$$

Substituting (B32) into (B18), the mean-square estimation error is lower bounded as

$$\begin{aligned}&\binom{N}{M}^{-1}\sum_{\mathcal{M}}\mathbb{E}_{R}[(\check{F} - \bar{f}_{\mathcal{M}})^2]\\ &\geqslant \binom{N}{M}^{-1}\sum_{\mathcal{M}}\sum_{n\in\mathcal{M}}\frac{I^{-1}_{n,n}(\boldsymbol{f}_{\mathcal{M}})}{M^2}\\ &\geqslant \binom{N}{M}^{-1}\sum_{n\in\mathcal{N}}\binom{N-1}{M-1}\frac{(2f_n+1)(1-f_n)}{2M^2}\\ &= \sum_{n\in\mathcal{N}}\frac{(2f_n+1)(1-f_n)}{2MN}.\end{aligned} \tag{B33}$$

With (B33), Eq. (B12) is obtained, which completes the proof of Lemma 6. ■

2. Achieving the minimum estimation error

This Appendix describes the measurement operation used to achieve the minimum estimation error.

According to the proof of Lemma 6, a necessary condition for achieving the minimum estimation error is that the inequality in (B31) is balanced. To this end, the projection of a measurement operator onto the target state, i.e., $a_{r,n}$ defined in (B21), must balance either of the two inequalities in (B28). The estimation protocol is built according to this principle. Specifically, we consider bilateral local Pauli measurements in the same basis. When both nodes make measurements in the u basis, $u \in \{x, y, z\}$, the operators corresponding to the asymmetric and symmetric measurement results $\boldsymbol{M}_{u,0}$ and $\boldsymbol{M}_{u,1}$ are given, respectively, by

$$\begin{aligned}\boldsymbol{M}_{x,0} &= |+-\rangle\langle+-| + |-+\rangle\langle-+|\\ &= |\Psi^-\rangle\langle\Psi^-| + |\Phi^-\rangle\langle\Phi^-|,\end{aligned} \tag{B34a}$$

$$\begin{aligned}\boldsymbol{M}_{x,1} &= |++\rangle\langle++| + |--\rangle\langle--|\\ &= |\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+|,\end{aligned} \tag{B34b}$$

$$\begin{aligned}\boldsymbol{M}_{y,0} &= |\times\odot\rangle\langle\times\odot| + |\odot\times\rangle\langle\odot\times|\\ &= |\Psi^-\rangle\langle\Psi^-| + |\Phi^+\rangle\langle\Phi^+|,\end{aligned} \tag{B34c}$$

$$\begin{aligned}\boldsymbol{M}_{y,1} &= |\times\times\rangle\langle\times\times| + |\odot\odot\rangle\langle\odot\odot|\\ &= |\Psi^+\rangle\langle\Psi^+| + |\Phi^-\rangle\langle\Phi^-|,\end{aligned} \tag{B34d}$$

$$\begin{aligned}\boldsymbol{M}_{z,0} &= |01\rangle\langle01| + |10\rangle\langle10|\\ &= |\Psi^-\rangle\langle\Psi^-| + |\Psi^+\rangle\langle\Psi^+|,\end{aligned} \tag{B34e}$$

$$\begin{aligned}\boldsymbol{M}_{z,1} &= |00\rangle\langle00| + |11\rangle\langle11|\\ &= |\Phi^-\rangle\langle\Phi^-| + |\Phi^+\rangle\langle\Phi^+|,\end{aligned} \tag{B34f}$$

where $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$, $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$, $|\times\rangle = \frac{|0\rangle+i|1\rangle}{\sqrt{2}}$, and $|\odot\rangle = \frac{|0\rangle-i|1\rangle}{\sqrt{2}}$.

In (B34), the operators $\boldsymbol{M}_{u,0}$, $u \in \{x, y, z\}$, balance the second inequality in (B28) and the operators $\boldsymbol{M}_{u,1}$, $u \in \{x, y, z\}$, balance the first inequality in (B28). The following two lemmas confirm the optimality of the proposed measurement operator.

Lemma 7 (unbiased estimator). Given the measurements made in Protocol 1, the estimated fidelity

$$\check{f} = 1 - \frac{3\varepsilon_{\mathcal{M}}}{2} \tag{B35}$$

satisfies (A23b), where the QBER $\varepsilon_{\mathcal{M}}$ is expressed as

$$\varepsilon_{\mathcal{M}} = \frac{\sum_{n\in\mathcal{M}} r_n}{M}, \tag{B36}$$

in which r_n is the measurement result of qubit n.

Proof. We define the following four terms:

$$\begin{aligned}f^{(0)} &= \frac{1}{M}\sum_{n\in\mathcal{M}}\mathrm{Tr}(\boldsymbol{\rho}_n|\Psi^-\rangle\langle\Psi^-|),\\ f^{(1)} &= \frac{1}{M}\sum_{n\in\mathcal{M}}\mathrm{Tr}(\boldsymbol{\rho}_n|\Psi^+\rangle\langle\Psi^+|),\\ f^{(2)} &= \frac{1}{M}\sum_{n\in\mathcal{M}}\mathrm{Tr}(\boldsymbol{\rho}_n|\Phi^-\rangle\langle\Phi^-|),\\ f^{(3)} &= \frac{1}{M}\sum_{n\in\mathcal{M}}\mathrm{Tr}(\boldsymbol{\rho}_n|\Phi^+\rangle\langle\Phi^+|).\end{aligned} \tag{B37}$$

In this case, $\bar{f}_{\mathcal{M}} = f^{(0)}$. Moreover, as the Bell states form a basis of $\mathbb{C}^4$,

$$\sum_{i=0}^{3} f^{(i)} = \frac{1}{M}\sum_{n\in\mathcal{M}} \mathrm{Tr}(\boldsymbol{\rho}_n) = 1. \tag{B38}$$

Given the distribution of A_n, i.e., $\Pr(A_n = u) = \frac{1}{3}$, $u \in \{x, y, z\}$, and the expression of the measurement operators (B34), the following expression can be obtained:

$$\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{n\in\mathcal{M}} R_n \mathbb{1}(A_n = x)}{M}\middle|\bar{f}_{\mathcal{M}}\right]\\
&= \frac{1}{M}\sum_{n\in\mathcal{M}} \Pr(A_n = x)\mathbb{E}[R_n|A_n = x, \bar{f}_{\mathcal{M}}]\\
&= \frac{1}{3M}\sum_{n\in\mathcal{M}} \mathrm{Tr}[\boldsymbol{\rho}_n(\mathbb{I}_4 - \boldsymbol{M}_{x,0})]\\
&= \frac{1}{3M}\left(\sum_{n\in\mathcal{M}} \mathrm{Tr}(\boldsymbol{\rho}_n|\Psi^+\rangle\langle\Psi^+|) + \sum_{n\in\mathcal{M}} \mathrm{Tr}(\boldsymbol{\rho}_n|\Phi^+\rangle\langle\Phi^+|)\right)\\
&= \frac{f^{(1)} + f^{(3)}}{3}.
\end{aligned} \tag{B39}$$

Similar to (B39), we can obtain that

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{n\in\mathcal{M}} R_n \mathbb{1}(A_n = y)}{M}\middle|\bar{f}_{\mathcal{M}}\right]
&= \frac{1}{3M}\sum_{n\in\mathcal{M}} \mathrm{Tr}[\boldsymbol{\rho}_n(\mathbb{I}_4 - \boldsymbol{M}_{y,0})]\\
&= \frac{f^{(1)} + f^{(2)}}{3},
\end{aligned} \tag{B40}$$

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{n\in\mathcal{M}} R_n \mathbb{1}(A_n = z)}{M}\middle|\bar{f}_{\mathcal{M}}\right]
&= \frac{1}{3M}\sum_{n\in\mathcal{M}} \mathrm{Tr}[\boldsymbol{\rho}_n(\mathbb{I}_4 - \boldsymbol{M}_{z,0})]\\
&= \frac{f^{(2)} + f^{(3)}}{3}.
\end{aligned} \tag{B41}$$

Substituting (B38)–(B41) into (B35) yields

$$\begin{aligned}
&\mathbb{E}[\check{F}|\bar{f}_{\mathcal{M}}]\\
&= 1 - \frac{3}{2}\mathbb{E}[\mathcal{E}_{\mathcal{M}}|\bar{f}_{\mathcal{M}}]\\
&= 1 - \frac{3}{2}\mathbb{E}\left[\frac{\sum_{a\in\{x,y,z\}}\sum_{n\in\mathcal{M}} R_n \mathbb{1}\{A_n = a\}}{M}\middle|\bar{f}_{\mathcal{M}}\right]\\
&= 1 - \frac{3}{2}\left(\sum_{i=1}^{3}\frac{2f^{(i)}}{3}\right)\\
&= f^{(0)} = \bar{f}_{\mathcal{M}},
\end{aligned} \tag{B42}$$

where $\check{F}$ and $\mathcal{E}_{\mathcal{M}}$ are the random variable form of $\check{f}$ and $\varepsilon_{\mathcal{M}}$, respectively. Equation (B42) shows that the estimator given in (B35) satisfies (A23b). This completes the proof of Lemma 7. ■

Next Lemma 8 proves the optimality of the proposed measurement operation in scenarios with independent noise.

Lemma 8 (optimality with independent noise). Protocol 1 achieves the minimum estimation error in Problem 4.

Proof. By repeating the derivation from (B39)–(B42) for qubit pair n, the following expression can be obtained:

$$\Pr(R_n = 1) = \tfrac{2}{3}(1 - f_n). \tag{B43}$$

Hence, the variance of R_n is given by

$$\begin{aligned}
\mathbb{V}[R_n] &= \Pr(R_n = 1)[1 - \Pr(R_n = 1)]\\
&= \frac{(2f_n + 1)(2 - 2f_n)}{9}.
\end{aligned} \tag{B44}$$

In the case of independent noise, the measurement outcomes R_n, $n \in \mathcal{M}$, on different qubit pairs are independent. Therefore, according to (B35) and (B36),

$$\begin{aligned}
\mathbb{V}[\check{F}|\mathcal{M}] &= \frac{9}{4M^2}\sum_{n\in\mathcal{M}}\mathbb{V}[R_n]\\
&= \sum_{n\in\mathcal{M}}\frac{(2f_n + 1)(1 - f_n)}{2M^2}.
\end{aligned} \tag{B45}$$

According to Lemma 7, $\check{F}$ satisfies (A23b); hence

$$\mathbb{E}_R[(\check{F} - \bar{F}_{\mathcal{M}})^2|\mathcal{M}] = \mathbb{V}[\check{F}|\mathcal{M}]. \tag{B46}$$

As the sample set $\mathcal{M}$ is selected completely at random, Eqs. (B45) and (B46) indicate that for all $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$,

$$\begin{aligned}
&\binom{N}{M}^{-1}\sum_{\mathcal{M}\subset\mathcal{N}}\mathbb{E}_R[(\check{F} - \bar{F}_{\mathcal{M}})^2]\\
&= \binom{N}{M}^{-1}\sum_{\mathcal{M}\subset\mathcal{N}}\sum_{n\in\mathcal{M}}\frac{(2f_n + 1)(1 - f_n)}{2M^2}\\
&= \binom{N}{M}^{-1}\sum_{n\in\mathcal{N}}\binom{N-1}{M-1}\frac{(2f_n + 1)(1 - f_n)}{2M^2}\\
&= \sum_{n\in\mathcal{N}}\frac{(2f_n + 1)(1 - f_n)}{2MN}.
\end{aligned} \tag{B47}$$

According to Lemma 6, the estimation error of Problem 4 is no less than (B47). This aspect shows that Protocol 1 is optimal in Problem 4 for all fidelity compositions $\mathcal{S}_{\mathrm{id}}(\boldsymbol{f}_{\mathrm{all}}) \subset \mathcal{S}_{\mathrm{id}}$. This completes the proof of Lemma 8. ■

Finally, Theorem 2 summarizes the results of this section.

Theorem 2 (optimal estimation protocol). Protocol 1 is optimal for Problem 1.

Proof. Denote the measurement operation and estimator of Protocol 1 by $\mathcal{O}^*$ and $\mathcal{D}^*$, respectively. According to Lemma 8 and Theorem 1, the estimation protocol $\{\hat{\mathcal{O}}^*, \mathcal{D}^*\}$ is optimal for Problem 1, where the composite measurement operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T}$.

Recall the property of the operation $\mathcal{T}$ given in (A4), where $|\phi\rangle$ and $|\psi\rangle$ are Bell states $\{|\Phi^\pm\rangle, |\Psi^\pm\rangle\}$. Given (A4) and the fact that Bell states form a basis of $\mathcal{H}^4$, the Kraus operators of $\mathcal{T}$ processing one qubit pair can be expressed as

$$\boldsymbol{T}_\phi = |\phi\rangle\langle\phi|, \quad \phi \in \{\Phi^\pm, \Psi^\pm\}. \tag{B48}$$

In this case of (B34) and (B48), the operation $\mathcal{T}$ does not change the operators of $\mathcal{O}^*$, i.e.,

$$\boldsymbol{M}_{u,r} = \sum_{\phi\in\{\Phi^\pm,\Psi^\pm\}} \boldsymbol{T}_\phi^\dagger \boldsymbol{M}_{u,r}\boldsymbol{T}_\phi \; \forall\, u \in \{x, y, z\},\;\; r \in \{0, 1\}, \tag{B49}$$

where the dagger is the Hermitian transpose. Equation (B49) shows that the composite measurement operation $\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T}$ is equivalent to $\mathcal{O}^*$, i.e.,

$$\hat{\mathcal{O}}^* = \mathcal{O}^* \circ \mathcal{T} = \mathcal{O}^*. \tag{B50}$$

According to (B50), the estimation protocol $\{\mathcal{O}^*, \mathcal{D}^*\}$ of Protocol 1 is optimal for Problem 1. This proves Theorem 2. ■

[1] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).

[2] H. J. Kimble, The quantum internet, Nature (London) **453**, 1023 (2008).

[3] D. Castelvecchi, The quantum internet has arrived (and it hasn't), Nature (London) **554**, 289 (2018).

[4] M. Pant, H. Krovi, D. Towsley, L. Tassiulas, L. Jiang, P. Basu, D. Englund, and S. Guha, Routing entanglement in the quantum internet, npj Quantum Inf. **5**, 25 (2019).

[5] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, Practical device-independent quantum cryptography via entropy accumulation, Nat. Commun. **9**, 459 (2018).

[6] M. Avesani, D. G. Marangon, G. Vallone, and P. Villoresi, Source-device-independent heterodyne-based quantum random number generator at 17 Gbps, Nat. Commun. **9**, 5365 (2018).

[7] C. M. Lee and M. J. Hoban, Towards Device-Independent Information Processing on General Quantum Networks, Phys. Rev. Lett. **120**, 020504 (2018).

[8] R. Schwonnek, K. T. Goh, I. W. Primaatmaja, E. Y. Z. Tan, R. Wolf, V. Scarani, and C. C. W. Lim, Device-independent quantum key distribution with random key basis, Nat. Commun. **12**, 2880 (2021).

[9] M.-H. Li, X. Zhang, W.-Z. Liu, S.-R. Zhao, B. Bai, Y. Liu, Q. Zhao, Y. Peng, J. Zhang, Y. Zhang, W. J. Munro, X. Ma, Q. Zhang, J. Fan, and J.-W. Pan, Experimental Realization of Device-Independent Quantum Randomness Expansion, Phys. Rev. Lett. **126**, 050503 (2021).

[10] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photon. **12**, 1012 (2020).

[11] M. Körber, O. Morin, S. Langenfeld, A. Neuzner, S. Ritter, and G. Rempe, Decoherence-protected memory for a single-photon qubit, Nat. Photon. **12**, 18 (2018).

[12] J. Yin, Y. Cao, Y.-H. Li, S.-K. Liao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia *et al.*, Satellite-based entanglement distribution over 1200 kilometers, Science **356**, 1140 (2017).

[13] P. C. Humphreys, N. Kalb, J. P. J. Morits, R. N. Schouten, R. F. L. Vermeulen, D. J. Twitchen, M. Markham, and R. Hanson, Deterministic delivery of remote entanglement on a quantum network, Nature (London) **558**, 268 (2018).

[14] M. Pompili, S. L. N. Hermans, S. Baier, H. K. C. Beukers, P. C. Humphreys, R. N. Schouten, R. F. L. Vermeulen, M. J. Tiggelman, L. dos Santos Martins, B. Dirkse, S. Wehner, and R. Hanson, Realization of a multinode quantum network of remote solid-state qubits, Science **372**, 259 (2021).

[15] H.-Y. Liu, X.-H. Tian, C. Gu, P. Fan, X. Ni, R. Yang, J.-N. Zhang, M. Hu, J. Guo, X. Cao, X. Hu, G. Zhao, Y.-Q. Lu, Y.-X. Gong, Z. Xie, and S.-N. Zhu, Drone-based entanglement distribution towards mobile quantum networks, Natl. Sci. Rev. **7**, 921 (2020).

[16] J. R. West, D. A. Lidar, B. H. Fong, and M. F. Gyure, High Fidelity Quantum Gates via Dynamical Decoupling, Phys. Rev. Lett. **105**, 230503 (2010).

[17] M. G. Bason, M. Viteau, N. Malossi, P. Huillery, E. Arimondo, D. Ciampini, R. Fazio, V. Giovannetti, R. Mannella, and O. Morsch, High-fidelity quantum driving, Nat. Phys. **8**, 147 (2012).

[18] S. Arroyo-Camejo, A. Lazariev, S. W. Hell, and G. Balasubramanian, Room temperature high-fidelity holonomic single-qubit gate on a solid-state spin, Nat. Commun. **5**, 4870 (2014).

[19] X. Zhang, M. Luo, Z. Wen, Q. Feng, S. Pang, W. Luo, and X. Zhou, Direct Fidelity Estimation of Quantum States Using Machine Learning, Phys. Rev. Lett. **127**, 130503 (2021).

[20] R. D. Somma, J. Chiaverini, and D. J. Berkeland, Lower bounds for the fidelity of entangled-state preparation, Phys. Rev. A **74**, 052302 (2006).

[21] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, Toolbox for entanglement detection and fidelity estimation, Phys. Rev. A **76**, 030305(R) (2007).

[22] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[23] H. Zhu and M. Hayashi, Optimal verification and fidelity estimation of maximally entangled states, Phys. Rev. A **99**, 052346 (2019).

[24] N. Kalb, A. A. Reiserer, P. C. Humphreys, J. J. W. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson, Entanglement distillation between solid-state quantum network nodes, Science **356**, 928 (2017).

[25] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, Tight finite-key analysis for quantum cryptography, Nat. Commun. **3**, 634 (2012).

[26] C. Pfister, M. A. Rol, A. Mantri, M. Tomamichel, and S. Wehner, Capacity estimation and verification of quantum channels with arbitrarily correlated errors, Nat. Commun. **9**, 27 (2018).

[27] E. Bagan, M. A. Ballester, R. D. Gill, R. Muñoz-Tapia, and O. Romero-Isart, Separable Measurement Estimation of Density Matrices and its Fidelity Gap with Collective Protocols, Phys. Rev. Lett. **97**, 130501 (2006).

[28] K. De Greve, P. L. McMahon, L. Yu, J. S. Pelc, C. Jones, C. M. Natarajan, N. Y. Kim, E. Abe, S. Maier, C. Schneider, M. Kamp, S. Höfling, R. H. Hadfield, A. Forchel, M. M. Fejer, and Y. Yamamoto, Complete tomography of a high-fidelity solid-state entangled spin–photon qubit pair, Nat. Commun. **4**, 2228 (2013).

[29] M. Bock, P. Eich, S. Kucera, M. Kreis, A. Lenhard, C. Becher, and J. Eschner, High-fidelity entanglement between a trapped ion and a telecom photon via quantum frequency conversion, Nat. Commun. **9**, 1998 (2018).

[30] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1999).

[31] C. R. Rao, *Selected Papers of C.R. Rao* (Wiley, New York, 1994).

[32] M. G. A. Paris, Quantum estimation for quantum technology, Int. J. Quantum Inf. **07**, 125 (2009).

[33] D. Šafránek, Simple expression for the quantum Fisher information matrix, Phys. Rev. A **97**, 042322 (2018).

[34] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels, Phys. Rev. Lett. **76**, 722 (1996).

[35] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[36] R. Horodecki and M. Horodecki, Information-theoretic aspects of inseparability of mixed states, Phys. Rev. A **54**, 1838 (1996).

[37] M. Horodecki, P. Horodecki, and R. Horodecki, Inseparable Two Spin-$\frac{1}{2}$ Density Matrices Can Be Distilled to a Singlet Form, Phys. Rev. Lett. **78**, 574 (1997).

[38] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, Some Classes of Global Cramer-Rao Bounds, Ann. Statis. **15**, 1421 (1987).