# Supplemental Information for "Quantification of Entanglement and Coherence with Purity Detection"


Ting Zhang,[1] Graeme Smith,[2,3,4] John A. Smolin,[5] Lu Liu,[1] Xu-Jie Peng,[1] Qi Zhao,[6] Davide Girolami,[7] Xiongfeng Ma,[8] Xiao Yuan,[9,10,*] and He Lu[1,11,†]

[1]*School of Physics, State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China*
[2]*JILA, University of Colorado/NIST, 440 UCB, Boulder, CO 80309, USA*
[3]*Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA*
[4]*Department of Physics, University of Colorado, 390 UCB, Boulder, CO 80309, USA*
[5]*IBM T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598*
[6]*QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[7]*DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy*
[8]*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084 China*
[9]*Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China*
[10]*School of Computer Science, Peking University, Beijing 100871, China*
[11]*Shenzhen Research Institute of Shandong University, Shenzhen 518057, China.*


## SUPPLEMENTARY NOTE 1: DERIVATION OF THE BOUNDS

Given a quantum state $\rho$ in a $d$-dimensional Hilbert space, our task is to bound the Von Neumann entropy of $\rho$ with a function of the state purity $\mathcal{P}(\rho) := \mathrm{Tr}(\rho^2)$. The spectral decomposition of the quantum state is $\rho = \sum_{i=1}^{d} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$, where $\{\left|\psi_i\right\rangle\}$ forms an orthonormal basis of the $d$-dimensional Hilbert space. The variational problem is then formulated as

$$\begin{aligned} \max/\min S(\rho) &= -\sum_{i=1}^{d} \lambda_i \log(\lambda_i) \\ s.t.\ \ &\sum_{i=1}^{d} \lambda_i^2 = \gamma \\ &\sum_{i=1}^{d} \lambda_i = 1 \\ &0 \le \lambda_i \le 1, \forall i, \end{aligned} \tag{1}$$

where $\mathcal{P}(\rho) = \mathrm{Tr}\rho^2$ is the purity of $\rho$.

Intuitively, the vector $\lambda$ that maximizes $S$ is the one that spread as uniformly as possible; while the vector $\lambda$ that minimizes $S$ is the one that has the minimal number of nonzero large values. In the following, we will analytically solve this problem and confirm this intuition.

### A. Maximization

First, we focus on the maximization problem with $d = 3$. Note that when $d = 2$, the solution to the constraints of Eq. (1) is unique and the optimization problem will be trivial. Without loss of generality, we assume $\lambda_1 \ge \lambda_2 \ge \lambda_3$. Then the problem can be stated as

$$\begin{aligned} \max S(\rho) &= -\lambda_1 \log(\lambda_1) - \lambda_2 \log(\lambda_2) - \lambda_3 \log(\lambda_3) \\ s.t.\ \ &\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \gamma \\ &\lambda_1 + \lambda_2 + \lambda_3 = 1 \\ &1 \ge \lambda_1 \ge \lambda_2 \ge \lambda_3 \ge 0. \end{aligned} \tag{2}$$

We prove that the maximum is reached with the following Lemma.


* xiaoyuan@pku.edu.cn
† luhe@sdu.edu.cn

**Lemma 1.** *The solution to the maximization problem in Eq. (2) is given by*

$$\begin{aligned}\lambda_1 &= \frac{1}{3} + \sqrt{\frac{2}{3}\left(\mathcal{P}(\rho) - \frac{1}{3}\right)},\\ \lambda_2 &= \lambda_3 = \frac{1-\lambda_1}{2}.\end{aligned} \tag{3}$$

*Proof.* The differential of the entropy function $S(\rho)$ and the constraints are given by

$$\begin{aligned}dS = &-\left(\frac{1}{\ln(2)} + \log\lambda_1\right)d\lambda_1 - \left(\frac{1}{\ln(2)} + \log\lambda_2\right)d\lambda_2\\ &-\left(\frac{1}{\ln(2)} + \log\lambda_3\right)d\lambda_3\end{aligned} \tag{4}$$

and

$$\begin{aligned}&\lambda_1 d\lambda_1 + \lambda_2 d\lambda_2 + \lambda_3 d\lambda_3 = 0\\ &d\lambda_1 + d\lambda_2 + d\lambda_3 = 0,\end{aligned} \tag{5}$$

respectively. We rewrite Eq. (5) to

$$\begin{aligned}d\lambda_1 &= -\frac{(\lambda_3-\lambda_2)}{\lambda_1-\lambda_2}d\lambda_3,\\ d\lambda_2 &= -\frac{(\lambda_1-\lambda_3)}{\lambda_1-\lambda_2}d\,\lambda_3.\end{aligned}$$

Thus, the differential of the entropy function becomes

$$\begin{aligned}dS(\rho) &= \frac{d\lambda_3}{\lambda_1-\lambda_2}[(\lambda_3-\lambda_2)\log\lambda_1 + (\lambda_1-\lambda_3)\log\lambda_2\\ &\quad + (\lambda_2-\lambda_1)\log\lambda_3]\\ &= (\lambda_2-\lambda_3)\left[-\frac{\log\lambda_1-\log\lambda_2}{\lambda_1-\lambda_2} + \frac{\log\lambda_3-\log\lambda_2}{\lambda_3-\lambda_2}\right]d\lambda_3\end{aligned} \tag{6}$$

Since the function $\log\lambda$ is concave for $\lambda\in[0,1]$, for $\lambda_1\geq\lambda_2\geq\lambda_3$,

$$\frac{\log\lambda_1-\log\lambda_2}{\lambda_1-\lambda_2} \leq \frac{\log\lambda_3-\log\lambda_2}{\lambda_3-\lambda_2}.$$

Thus, $dS(\rho)/d\lambda_3\geq 0$. To reach the maximum of $S(\rho)$, we thus only need to set $\lambda_3$ to be its maximum, which happens when $\lambda_2=\lambda_3$. Together with the constraints, then we can solve the equations and show that the solution to the maximization problem is given in Eq. (3). □

Now, we can solve the maximization problem of Eq. (1) for a general case of $d$.

**Theorem 1.** *Suppose $\lambda_1\geq\lambda_2\geq\dots\lambda_d$. The solution to the maximization problem in Eq. (1) is*

$$\begin{aligned}\lambda_1 &= \frac{1}{d} + \sqrt{\frac{d-1}{d}\left(\mathcal{P}(\rho) - \frac{1}{d}\right)},\\ \lambda_2 &= \lambda_3 = \cdots = \lambda_d = \frac{1-\lambda_1}{d-1}.\end{aligned} \tag{7}$$

*Proof.* The solution in Eq. (7) is exactly determined when setting $\lambda_2=\lambda_3=\cdots=\lambda_d$. Suppose the maximization problem solution is not this one, then we must have that $\lambda_2>\lambda_d$. In the following, we prove the contradiction by showing that changing the values of $\lambda_1,\lambda_2,\lambda_d$ would make the entropy $S(\rho)$ larger while fixing all other values $(\lambda_3,\lambda_4,\dots,\lambda_{d-1})$ and the constraints. Now the constraints for $\lambda_1$, $\lambda_2$, and $\lambda_d$ becomes

$$\begin{aligned}\lambda_1^2+\lambda_2^2+\lambda_d^2 &= a\\ \lambda_1+\lambda_2+\lambda_d &= b.\end{aligned} \tag{8}$$

By defining $\lambda_1' = \lambda_1/b$, $\lambda_2' = \lambda_2/b$, $\lambda_d' = \lambda_d/b$, the relations become

$$\begin{aligned} \lambda_1'^2 + \lambda_2'^2 + \lambda_d'^2 = a/b^2 \\ \lambda_1' + \lambda_2' + \lambda_d' = 1. \end{aligned} \tag{9}$$

The entropy function is

$$\begin{aligned} S(\rho) = -\sum_{i=1}^{d} \lambda_i \log(\lambda_i), \\ = S_{1,2,d}(\rho) + S_r(\rho), \end{aligned} \tag{10}$$

where $S_{1,2,d}(\rho) = -\lambda_1 \log(\lambda_1) - \lambda_2 \log(\lambda_2) - \lambda_d \log(\lambda_d)$ and $S_r(\rho) = -\sum_{i=3}^{d-1} \lambda_i \log(\lambda_i)$. Since $S_r(\rho)$ is fixed, we need to maximize $S_{1,2,d}(\rho)$, which can also be represented as

$$\begin{aligned} S_{1,2,d}(\rho) &= -b\lambda_1' \log(b\lambda_1') - b\lambda_2' \log(b\lambda_2') - b\lambda_d' \log(b\lambda_d') \\ &= b[-\lambda_1' \log(\lambda_1') - \lambda_2' \log(\lambda_2') - \lambda_d' \log(\lambda_d')] - b \log b \end{aligned} \tag{11}$$

Denoting $S_{1,2,d}'(\rho) = -\lambda_1' \log(\lambda_1') - \lambda_2' \log(\lambda_2') - \lambda_d' \log(\lambda_d')$, this optimization problem has the same form of Eq. (2). Then Lemma 1 indicates that the maximum of $S_{1,2,d}'(\rho)$ given the constraints in Eq. (9) is reached when $\lambda_2' = \lambda_d'$. In other words, the maximum of $S_{1,2,d}(\rho)$ given the constrains of Eq. (8) is saturated with $\lambda_2 = \lambda_d$, which contradicts $\lambda_2 > \lambda_d$. Therefore, the solution to the maximization problem is given by Eq. (7). □

### B. Minimization

Now, we consider the solution to the minimization of Eq. (1). Similarly, we first consider the minimization with $d = 3$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3$,

$$\begin{aligned} \min S(\rho) = -\lambda_1 \log(\lambda_1) - \lambda_2 \log(\lambda_2) - \lambda_3 \log(\lambda_3) \\ s.t.\ \ \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \mathcal{P}(\rho) \\ \lambda_1 + \lambda_2 + \lambda_3 = 1 \\ 1 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \end{aligned} \tag{12}$$

**Lemma 2.** *The solution to the minimization problem in Eq. (12) is reached either when $\lambda_1 = \lambda_2$ or $\lambda_3 = 0$.*

*Proof.* From the proof of Lemma 1, we already showed that $dS(\rho)/d\lambda_3 \geq 0$. Therefore, the lower bound of $S(\rho)$ is reached when $\lambda_3$ takes its minimum. As $2(\lambda_1^2 + \lambda_2^2) \geq (\lambda_1 + \lambda_2)^2$, according to Eq. (12), we have

$$2(\mathcal{P}(\rho) - \lambda_3^2) \geq (1 - \lambda_3)^2. \tag{13}$$

The lower bound for $\lambda_3$ is

$$\lambda_3 \geq \max\left\{0, \frac{1 - \sqrt{6\mathcal{P}(\rho) - 2}}{3}\right\}$$

Thus, when $\mathcal{P}(\rho) \geq 1/2$, the minimal possible value for $\lambda_3$ is 0. When $1/3 \leq \mathcal{P}(\rho) < 1/2$, the minimal possible value for $\lambda_3$ is $\frac{1-\sqrt{6\mathcal{P}(\rho)-2}}{3}$ and $\lambda_1 = \lambda_2 = (1 - \lambda_3)/2$. Note that $\mathcal{P}(\rho) \geq 1/3$ for $d = 3$. □

Now, we can show the general solution to the minimization of Eq. (1).

**Theorem 2.** *Suppose $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k$, the solution to the minimization problem in Eq. (1) is*

$$\begin{aligned} \lambda_1 = \lambda_2 = \cdots = \lambda_{k-1} = \frac{1-\alpha}{k-1}, \\ \lambda_k = \alpha, \\ \lambda_{k+1} = \cdots = \lambda_d = 0. \end{aligned} \tag{14}$$

*Here,*

$$\alpha = \frac{1}{k} - \sqrt{(1 - 1/k)(\mathcal{P}(\rho) - 1/k)} \tag{15}$$

*and $k$ is the integer such that $\frac{1}{k} \leq \mathcal{P}(\rho) \leq \frac{1}{k-1}$.*

*Proof.* Suppose we always have the solution in the form as

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{k-1}, \lambda_k, \lambda_{k+1} = \ldots \lambda_d = 0. \tag{16}$$

Otherwise, there must exist three $\lambda_i, \lambda_j, \lambda_k$ such that $\lambda_i > \lambda_j \geq \lambda_k$ and $\lambda_k \neq 0$. Following a similar argument in the proof of Theorem 1, we can show that this contradicts Lemma 2.

According to Eq. (16), we have

$$\begin{aligned}(k-1)\lambda_1^2 + \lambda_k^2 &= \mathcal{P}(\rho),\\ (k-1)\lambda_1 + \lambda_k &= 1,\\ k &\leq d\end{aligned}$$

We can show that the possible integer value for $k$ is unique. That is,

$$\begin{aligned}k[(k-1)\lambda_1^2 + \lambda_k^2] &\geq [(k-1)\lambda_1 + \lambda_k]^2\\ &\geq (k-1)[(k-1)\lambda_1^2 + \lambda_k^2]\end{aligned}$$

Equivalently, we have

$$k\mathcal{P}(\rho) \geq 1 \geq (k-1)\mathcal{P}(\rho),$$

hence

$$\begin{aligned}\frac{1}{\mathcal{P}(\rho)} \leq k \ &\leq \frac{1}{\mathcal{P}(\rho)} + 1,\\ \frac{1}{k} \leq \mathcal{P}(\rho) \ &\leq \frac{1}{k-1}.\end{aligned}$$

□

### C. Upper and lower bounds to coherence and entanglement

We now call $\{\lambda^{\mathrm{M}}_{i,\rho}\}, \{\lambda^{\mathrm{m}}_{i,\rho}\}$ the vectors solving the maximization and the minimization, respectively. Given a bipartite state $\rho_{AB}$, by minimizing (maximizing) the marginal purity on $B$ subsystem and maximizing (minimizing) the global purity, one has
*Result 1 (Eq. (2) of the main text)*— Given a quantum state $\rho_{AB} \in \mathcal{H}_{d_A} \otimes \mathcal{H}_{d_B}$, and defining $\rho_B = \mathrm{Tr}_A \rho_{AB}$, its coherent information $I(A\rangle B)$ is bounded as follows:

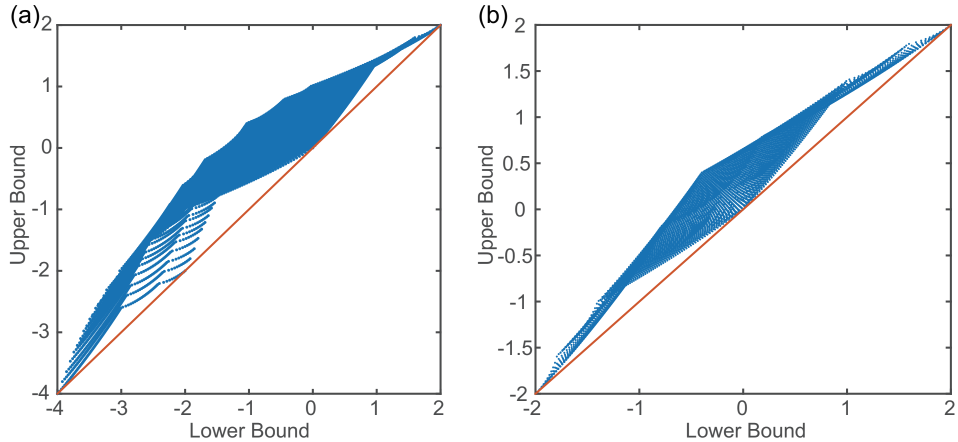
$$l_e(\rho_{AB}) \leq I(A\rangle B) \leq u_e(\rho_{AB}), \tag{17}$$

$$\begin{aligned}&l_e(\rho_{AB}) =\\ &-(1-\lambda^{\mathrm{m}}_{k_{\rho_B},\rho_B})\log\lambda^{\mathrm{m}}_{1,\rho_B} - \lambda^{\mathrm{m}}_{k_{\rho_B},\rho_B}\log\lambda^{\mathrm{m}}_{k_{\rho_B},\rho_B}\\ &+(1-\lambda^{\mathrm{M}}_{1,\rho_{AB}})\log\frac{(1-\lambda^{\mathrm{M}}_{1,\rho_{AB}})}{(d-1)} + \lambda^{\mathrm{M}}_{1,\rho_{AB}}\log\lambda^{\mathrm{M}}_{1,\rho_{AB}},\\ &u_c(\rho_{AB}) =\\ &(1-\lambda^{\mathrm{m}}_{k_{\rho_{AB}},\rho_{AB}})\log\lambda^{\mathrm{m}}_{1,\rho_{AB}} + \lambda^{\mathrm{m}}_{k_{\rho_{AB}},\rho_{AB}}\log\lambda^{\mathrm{m}}_{k_{\rho_{AB}},\rho_{AB}}\\ &-(1-\lambda^{\mathrm{M}}_{1,\rho_B})\log\frac{(1-\lambda^{\mathrm{M}}_{1,\rho_B})}{(d_B-1)} - \lambda^{\mathrm{M}}_{1,\rho_B}\log\lambda^{\mathrm{M}}_{1,\rho_B}.\end{aligned}$$

By minimizing (maximizing) the coherence of the dephased state $\rho_d = \sum_i |i\rangle\langle i|\,\rho\,|i\rangle\langle i|$, and maximizing (minimizing) the coherence of the state under study, we obtain lower (upper) bounds to the relative entropy of coherence:
*Result 2 (Eq. (5) of the main text)* — The relative entropy of coherence $C_{\mathrm{RE}}(\rho)$ is bounded as follows:

$$l_c(\rho) \leq C_{\mathrm{RE}}(\rho) \leq u_c(\rho), \tag{18}$$

$$\begin{aligned}&l_c(\rho) = -(1-\lambda^{\mathrm{m}}_{k_{\rho_d},\rho_d})\log\lambda^{\mathrm{m}}_{1,\rho_d} - \lambda^{\mathrm{m}}_{k_{\rho_d},\rho_d}\log\lambda^{\mathrm{m}}_{k_{\rho_d},\rho_d}\\ &+(1-\lambda^{\mathrm{M}}_{1,\rho})\log\frac{(1-\lambda^{\mathrm{M}}_{1,\rho})}{(d-1)} + \lambda^{\mathrm{M}}_{1,\rho}\log\lambda^{\mathrm{M}}_{1,\rho},\\ &u_c(\rho) = (1-\lambda^{\mathrm{m}}_{k_\rho,\rho})\log\lambda^{\mathrm{m}}_{1,\rho} + \lambda^{\mathrm{m}}_{k_\rho,\rho}\log\lambda^{\mathrm{m}}_{k_\rho,\rho}\\ &-(1-\lambda^{\mathrm{M}}_{1,\rho_d})\log\frac{(1-\lambda^{\mathrm{M}}_{1,\rho_d})}{(d-1)} - \lambda^{\mathrm{M}}_{1,\rho_d}\log\lambda^{\mathrm{M}}_{1,\rho_d}.\end{aligned}$$




**Supplementary Fig** 1. Numerical simulations of lower and upper bounds of (a) coherent information and (b) relative entropy of coherence.

### D. Tightness of the bounds

To investigate the tightness of our bounds, we simulate the lower and upper bounds of coherent information on four-qubit state $\rho_{AB}$ with $|A| = |B| = 2$, and the results are shown in Supplementary Fig. 1(a). Each point displayed corresponds to the bounds at some pair of purity values $\mathcal{P}(\rho_{AB})$ and $\mathcal{P}(\rho_B)$. The results are shown in Supplementary Fig. 1(a). Also, we simulate the lower and upper bounds of relative entropy of coherence on four-dimensional state ($d = 4$), and the results are shown in Supplementary Fig. 1(b). Each point displayed corresponds to the bounds at some pair of purity values $\mathcal{P}(\rho)$ and $\mathcal{P}(\rho_d)$. We can see that overall the upper and lower bounds give a good estimation of the the coherent information and the relative entropy of coherence. These bounds are tight when their values are close to the minimal (maximally mixed states) and maximal (maximally entangled states and maximally coherent states).

We use $\epsilon$ to characterize the "tightness" of lower and upper bounds for *pure states* $\rho$. For coherent information, the bounds are tight ($u_e(\rho) = l_e(\rho)$) for pure states ($\mathcal{P}(\rho_{AB} = 1)$) with $\mathcal{P}(\rho_B) = \frac{1}{d_B}$, where $B$ is the subsystem of $\rho = \rho_{AB}$ and $d_B$ is the dimension of system $B$. The difference $\epsilon_e = \mathcal{P}(\rho_B) - 1/d_B$ is related to the tightness of $u_e(\rho)$ and $l_e(\rho)$. For example, we consider a 4-qubit GHZ state $|\mathrm{GHZ}_4\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)_{1234}$ with bi-partition of $A = \{1, 2, 3\}$ and $B = \{4\}$. The purity of subsystem $B$ is $\mathcal{P}(\rho_B) = \frac{1}{2}$ indicating the maximal entanglement of $|\mathrm{GHZ}_4\rangle$.

For relative entropy of coherence, the bounds are tight ($u_c(\rho) = l_c(\rho)$) for pure state ($\mathcal{P}(\rho) = 1$) with diagonal matrix of $\rho_d = \frac{1}{d}\mathbb{I}_d$ ($\mathcal{P}(\rho_d) = \frac{1}{d}$). The difference $\epsilon_c = \mathcal{P}(\rho_d) - 1/d$ is related to the tightness of $u_c(\rho)$ and $l_c(\rho)$. For example, we consider a single-qubit state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, the purity of its diagonal matrix is $\mathcal{P}(\rho_d) = \frac{1}{2}$ indicating $|+\rangle$ is a maximally coherent state.

In theory, $\epsilon_e$ and $\epsilon_c$ provide mathematical criteria for the tightness of our method. $\epsilon_e(\epsilon_c) \to 0$ indicates the maximally entangled state (maximally coherent state), which is of particular interest in quantum information science.

## SUPPLEMENTARY NOTE 2: DETAILS OF EXPERIMENTAL REALIZATIONS AND RESULTS

Before discussing the experimental details, we provide an overview of the crucial optical components and their respective functions employed in our experiment:

- Waveplates: We utilize both half-wave plates (HWPs) and quarter-wave plates (QWPs) to perform unitary operations ($U$) on photons. The angle between the fast axis of the waveplate and the vertical polarization direction is denoted as $\omega$ for HWPs and $\vartheta$ for QWPs. These waveplates induce specific unitary transformations on the quantum state of photons, enabling the desired experimental operations. The unitary transformations of HWPs and QWPs can be expressed as:

$$U_{\mathrm{HWP}}(\omega) = -\begin{pmatrix} \cos 2\omega & \sin 2\omega \\ \sin 2\omega & -\cos 2\omega \end{pmatrix}, \tag{19}$$

$$U_{\mathrm{QWP}}(\vartheta) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 + i\cos 2\vartheta & i\sin 2\vartheta \\ i\sin 2\vartheta & 1 - i\cos 2\vartheta \end{pmatrix} \tag{20}$$

- Polarization beam splitter (PBS): This component separates photons with perpendicular polarization directions. It transmits horizontally polarized photons while reflecting vertically polarized photons.

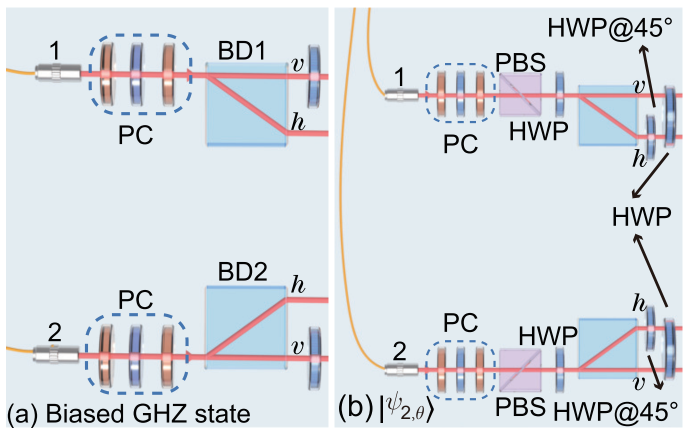
- Beam displacer (BD): The BD allows vertically polarized photons to pass through in the original direction, while horizontally polarized photons are displaced. In our experiment, we use a BD with a length of 28.3 mm, resulting in a displacement offset of 3 mm.

### A. Preparation of $|\mathrm{GHZ}_\theta\rangle$ and $|\psi_{2,\theta}\rangle$

To generate $|\mathrm{GHZ}_\theta\rangle$, photon 1 and 2 pass through two BDs after calibration of their polarization as shown in Fig. 2(a). Calling $h(v)$ the deviated (transmitted) spatial modes, the step-by-step description of optical elements to generate $|\mathrm{GHZ}_\theta\rangle$ is

$$\begin{aligned} &\cos\theta\,|HH\rangle_{12} + \sin\theta\,|VV\rangle_{12} \\ \xrightarrow{\mathrm{BD1}} &\cos\theta\,|HhH\rangle_{11'2} + \sin\theta\,|VvV\rangle_{11'2} \\ \xrightarrow{\mathrm{BD2}} &\cos\theta\,|HhHh\rangle_{11'22'} + \sin\theta\,|VvVv\rangle_{11'22'}\,. \end{aligned} \tag{21}$$

In the experiment, by changing the value of $\theta$, we totally generate eleven $\rho_{\mathrm{GHZ}_\theta}$. For each $\rho_{\mathrm{GHZ}_\theta}$, we reconstruct the correspond-




**Supplementary Fig** 2. (a) The setup to generate biased four-qubit GHZ states $|\mathrm{GHZ}_\theta\rangle$. (b) The setup to generate two-copy states $|\psi_{2,\theta}\rangle$. PC: polarization control is a combination of QWP, HWP and QWP which can be set to calibrate the variation of polarization caused by the twist of fibre.

ing density matrix $\rho_{\mathrm{GHZ}_\theta}$ using standard quantum state tomography (SQT). We calculate the fidelity $F = \mathrm{Tr}(\rho_{\mathrm{GHZ}_\theta}\,|\mathrm{GHZ}_\theta\rangle\langle\mathrm{GHZ}_\theta|)$. The results are summarized in Supplementary Table I.

**Supplementary Table** I. The fidelities of the prepared $\rho_{\mathrm{GHZ}_\theta}$.

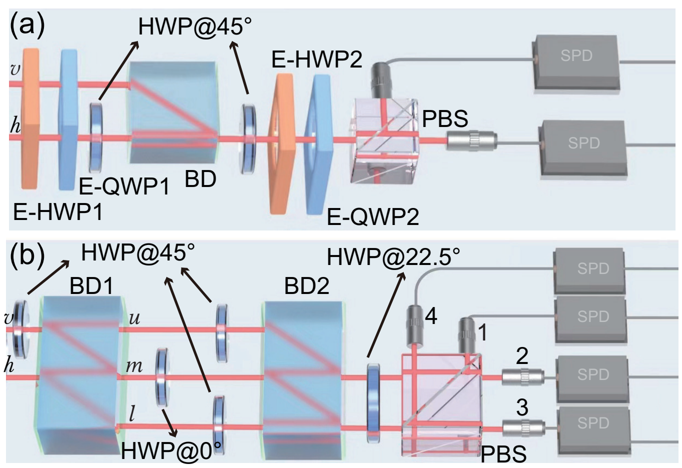
| $\theta$ | $F$ |
|---|---|
| 0 | $0.9293 \pm 0.0004$ |
| $\pi/20$ | $0.9050 \pm 0.0006$ |
| $2\pi/20$ | $0.9057 \pm 0.0007$ |
| $3\pi/20$ | $0.9331 \pm 0.0005$ |
| $4\pi/20$ | $0.9136 \pm 0.0006$ |
| $5\pi/20$ | $0.9124 \pm 0.0007$ |
| $6\pi/20$ | $0.9155 \pm 0.0003$ |
| $7\pi/20$ | $0.8850 \pm 0.0007$ |
| $8\pi/20$ | $0.8833 \pm 0.0009$ |
| $9\pi/20$ | $0.9155 \pm 0.0005$ |
| $10\pi/20$ | $0.9290 \pm 0.0004$ |

The setup to prepare $|\psi_{2,\theta}\rangle \otimes |\psi_{2,\theta}\rangle$ is illustrated in Supplementary Fig. 2(b). Photon 1 and 2 pass through two PBSs creating the resulting state of $|HH\rangle_{12}$. The step-by-step description of optical elements to generate $|\psi_{2,\theta}\rangle \otimes |\psi_{2,\theta}\rangle$ is

$$
\begin{aligned}
|HH\rangle_{12} &\xrightarrow[\text{are applied on photon 1 and 2}]{\text{Two HWPs@}\theta/2} (\cos\theta\,|H\rangle_1 + \sin\theta\,|V\rangle_1) \otimes (\cos\theta\,|H\rangle_2 + \sin\theta\,|V\rangle_2) \\
&\xrightarrow[\text{are applied on photon 1 and 2}]{\text{BD1\&BD2}} (\cos\theta\,|Hh\rangle_{11'} + \sin\theta\,|Vv\rangle_{11'}) \otimes (\cos\theta\,|Hh\rangle_{22'} + \sin\theta\,|Vv\rangle_{22'}) \\
&\xrightarrow[\text{are applied on path } h \text{ of photon 1 and 2}]{\text{Two HWPs@45}^\circ} |VV\rangle_{12} \otimes [(\cos\theta\,|h\rangle_{1'} + \sin\theta\,|v\rangle_{1'}) \otimes (\cos\theta\,|h\rangle_{2'} + \sin\theta\,|v\rangle_{2'})] = |VV\rangle_{12} \otimes |\psi_{2,\theta}\rangle_{1'2'} \\
&\xrightarrow[\text{are applied on both path of photon 1 and 2}]{\text{Two HWPs@}\theta/2} |\psi_{2,\theta}\rangle_{12} \otimes |\psi_{2,\theta}\rangle_{1'2'}
\end{aligned}
\tag{22}
$$

As shown in Eq. (22), the first $|\psi_{2,\theta}\rangle$ is encoded into the polarization degree of freedom (DOF) of both photons, while the second $|\psi_{2,\theta}\rangle$ is encoded on the path DOF of both photons.

### B. Experimental setups to perform shadow estimation and collective measurements

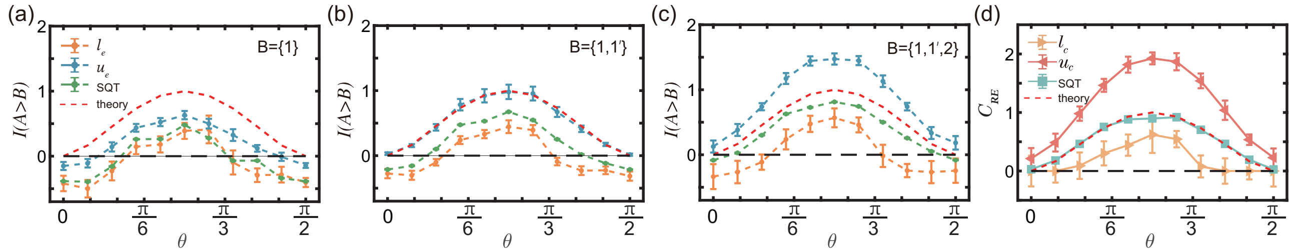



**Supplementary Fig** 3. Illustration of setups to perform (a) shadow estimation and (b) collective measurements.

In shadow estimation, the single-qubit Clifford unitary operation $U_n \in \mathrm{Cl}_2$ followed by a projective measurement on Pauli-Z basis is equivalent to measuring qubit on basis $J_n = U_n^\dagger Z U_n \in \{\pm X, \pm Y, \pm Z\}$. The experimental setup to perform measurement on arbitrary basis $(\alpha\,|H\rangle + \delta\,|V\rangle) \otimes (\gamma\,|h\rangle + \mu\,|v\rangle)$ is shown in Supplementary Fig. 3(a), where $\alpha\,|H\rangle + \delta\,|V\rangle$ is encoded on polarization DOF and $\gamma\,|h\rangle + \mu\,|v\rangle$ is encoded on path DOF. Projection on $\alpha\,|H\rangle + \delta\,|V\rangle$ is realized by combination of E-HWP1 (set at angle of $\theta_1$) and E-QWP1 (set at angle of $\upsilon_1$), while the projection of $\gamma\,|h\rangle + \mu\,|v\rangle$ and its orthogonal basis is realized by combination of E-HWP2 (set at angle of $\theta_2$), E-QWP2 (set at angle of $\upsilon_2$) and a PBS. A step-by-step description of implementation of such projective measurement is described by

$$
\begin{aligned}
(\alpha\,|H\rangle + \beta\,|V\rangle) \otimes (\gamma\,|h\rangle + \delta\,|v\rangle) &\xrightarrow[\text{on both path}]{\text{E-HWP1@}\omega_1\text{\&E-QWP1@}\vartheta_1} |H\rangle \otimes (\gamma\,|h\rangle + \delta\,|v\rangle) \\
&\xrightarrow[\text{on path h}]{\text{HWP@45}^\circ} \gamma\,|Vh\rangle + \delta\,|Hv\rangle \\
&\xrightarrow[\text{combine the path } h \text{ and } v]{\text{BD3}} \gamma\,|V\rangle + \delta\,|H\rangle \\
&\xrightarrow{\text{HWP@45}^\circ} \gamma\,|H\rangle + \delta\,|V\rangle \\
&\xrightarrow{\text{E-HWP2@}\omega_2\text{\&E-QWP2@}\vartheta_2} |H\rangle \\
&\xrightarrow{\text{PBS}} \begin{cases} \text{transmitted}, & \text{results of } (\alpha\,|H\rangle + \beta\,|V\rangle) \otimes (\gamma\,|h\rangle + \delta\,|v\rangle) \\ \text{reflected}, & \text{results of } (\alpha\,|H\rangle + \beta\,|V\rangle) \otimes (\gamma\,|h\rangle + \delta\,|v\rangle)^\perp \end{cases},
\end{aligned}
\tag{23}
$$

where $(\gamma\,|h\rangle + \delta\,|v\rangle)^\perp$ is the orthogonal state of $\gamma\,|h\rangle + \delta\,|v\rangle$.




**Supplementary Fig** 4. Experimental results of lower and upper bounds of $I(A\rangle B)$ and $C_{\mathrm{RE}}$ for the prepared $\rho_{\mathrm{GHZ}_\theta}$. (a)-(c), the lower bound $l_e$ (orange dots) and upper bound $u_e$ (blue dots) with $B=\{1\},\{1,1'\}$ and $\{1,1',2\}$ respectively. The green dots represent $I(A\rangle B)$ calculated with reconstructed $\rho_{\mathrm{GHZ}_\theta}$, and the dashed red lines represents the theoretical predictions of ideal $|\mathrm{GHZ}_\theta\rangle$. (d), the the lower bound $l_c$ (orange dots) and upper bound $u_c$ (red dots) of $C_{\mathrm{RE}}(\rho_{\psi_{2,\theta}})$. The cyan dots represent $C_{\mathrm{RE}}(\rho_{\psi_{2,\theta}})$ calculated with reconstructed $\rho_{\mathrm{GHZ}_\theta}$ and the dashed red line represents theoretical prediction of ideal $|\mathrm{GHZ}_\theta\rangle$.

The experimental setup to implement BSM between polarization-encoded qubit and path-encoded qubit is shown in Fig. 3(b). Note that the projection on four Bell states is demonstrated with single experimental setting. A detailed description of Fig. 3(b) is

$$
\begin{cases}
|\Psi^+\rangle=\frac{1}{\sqrt{2}}(|Hv\rangle+|Vh\rangle)\\
|\Psi^-\rangle=\frac{1}{\sqrt{2}}(|Hv\rangle-|Vh\rangle)\\
|\Phi^+\rangle=\frac{1}{\sqrt{2}}(|Hh\rangle+|Vv\rangle)\\
|\Phi^-\rangle=\frac{1}{\sqrt{2}}(|Hh\rangle-|Vv\rangle)
\end{cases}
\xrightarrow[\text{on path } v]{\text{HWP@45}^\circ}
\begin{cases}
\frac{1}{\sqrt{2}}(|Vv\rangle+|Vh\rangle)\\
\frac{1}{\sqrt{2}}(|Vv\rangle-|Vh\rangle)\\
\frac{1}{\sqrt{2}}(|Hh\rangle+|Hv\rangle)\\
\frac{1}{\sqrt{2}}(|Hh\rangle-|Hv\rangle)
\end{cases}
\xrightarrow[\text{Re-encoding}]{\text{BD1}}
\begin{cases}
\frac{1}{\sqrt{2}}(|Vu\rangle+|Vm\rangle)\\
\frac{1}{\sqrt{2}}(|Vu\rangle-|Vm\rangle)\\
\frac{1}{\sqrt{2}}(|Hl\rangle+|Hm\rangle)\\
\frac{1}{\sqrt{2}}(|Hl\rangle-|Hm\rangle)
\end{cases}
$$
$$
\xrightarrow[\text{HWP@0}^\circ\text{on path } m]{\text{HWPs@45}^\circ\text{on path } u \text{ and } l}
\begin{cases}
\frac{1}{\sqrt{2}}(|Hu\rangle-|Vm\rangle)\\
\frac{1}{\sqrt{2}}(|Hu\rangle+|Vm\rangle)\\
\frac{1}{\sqrt{2}}(|Vl\rangle+|Hm\rangle)\\
\frac{1}{\sqrt{2}}(|Vl\rangle-|Hm\rangle)
\end{cases}
\xrightarrow[\text{combine path}]{\text{BD2}}
\begin{cases}
|-\rangle\,|m\rangle\\
|+\rangle\,|m\rangle\\
|+\rangle\,|l\rangle\\
|-\rangle\,|l\rangle
\end{cases}
\xrightarrow[\text{on path } m \text{ and } l]{}
\begin{cases}
|V\rangle\,|m\rangle\\
|H\rangle\,|m\rangle\\
|H\rangle\,|l\rangle\\
|V\rangle\,|l\rangle
\end{cases}
\xrightarrow{\text{PBS}}
\begin{cases}
\text{clicks on detector 1}\\
\text{clicks on detector 2}\\
\text{clicks on detector 3}\\
\text{clicks on detector 4}
\end{cases}. \tag{24}
$$

### C. Robustness of coherent information and relative entropy of coherence

In our experiment, the fidelity of prepared states are 0.9% with respect to the target state, so that the results with standard state tomography (QST) are treated as "true values". This is because the coherent information and relative entropy of coherence are sensitive to noise. To clarify this, we present the theoretical predictions of $I(A\rangle B)$ with $B=\{1\},\{1,1'\}$ and $\{1,1',2\}$ (red dashed lines) in Supplementary Fig. 4(a)-(c) and that of $C_{\mathrm{RE}}$ in Supplementary Fig. 4(d). From Supplementary Fig. 4, we observe that

1. There are always gaps between the theoretical predictions and the results from SQT for $I(A\rangle B)$, and the smaller the $|B|$ is the larger the gap is.

2. The theoretical predictions of $C_{\mathrm{RE}}$ agrees well with that from SQT, except the state $\rho_{\mathrm{GHZ}_\theta}$ with $\theta=\pi/4$.

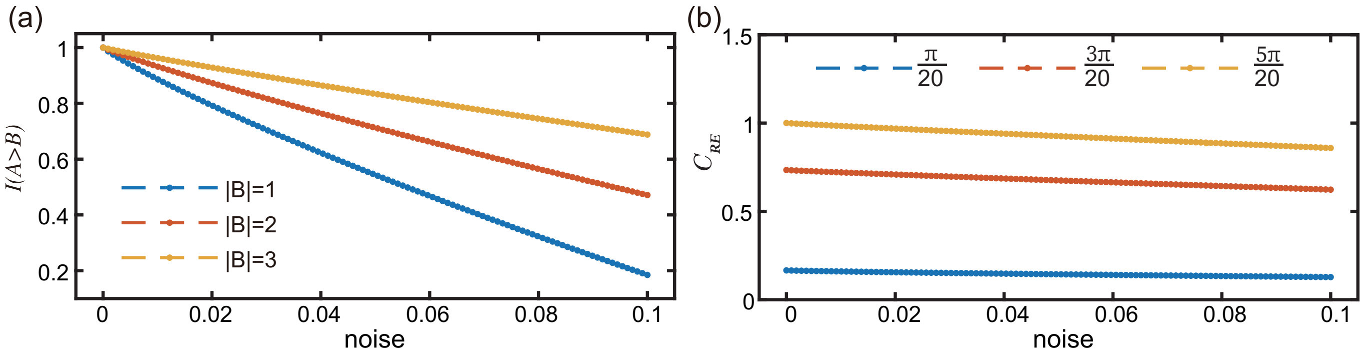
The discrepancy between theoretical predictions and experimental results are mainly caused by experimental noise, which we will discuss separately in the following.

For the coherent information $I(A\rangle B)$ and $C_{\mathrm{RE}}$, we calculate the values of $I(A\rangle B)$ on four-qubit state

$$\rho_{\text{noise}}=(1-p)\,|\mathrm{GHZ}_{\pi/4}\rangle\,\langle\mathrm{GHZ}_{\pi/4}|+p\frac{\mathbb{I}}{16}. \tag{25}$$

The results of $I(A\rangle B)$ with $p\in[0,0.1]$ are shown in Supplementary Fig. 5(a), where $I(A\rangle B)$ decreases faster with small $|B|$. In our experiment, the average fidelity of prepared $|\mathrm{GHZ}_\theta\rangle$ is about 0.9 as listed in Supplementary Table I, which introduces different discrepancies of $I(A\rangle B)$ with $|B|=1,2$ and 3 as shown in Supplementary Fig. 4(a)-(c). For the relative entropy of coherence $C_{\mathrm{RE}}$, we calculate the value of $C_{\mathrm{RE}}$ on four-qubit state

$$\rho_{\text{noise}}=(1-p)\,|\mathrm{GHZ}_\theta\rangle\,\langle\mathrm{GHZ}_\theta|+p\frac{\mathbb{I}}{16}. \tag{26}$$




**Supplementary Fig** 5. Simulated results of $I(A\rangle B)$ and $C_{\mathrm{RE}}$ of noisy states. (a), $I(A\rangle B)$ of noisy state in Eq. 25 with $|B| = 1, 2$ and 3. (b), $C_{\mathrm{RE}}$ of noisy state in Eq. 26 with $\theta = \pi/20, 3\pi/20$ and $5\pi/20$.

with $\theta = \pi/20, 3\pi/20$ and $5\pi/20$ respectively. As shown in Supplementary Fig. 5(b), $C_{\mathrm{RE}}$ decreases faster with larger $\theta$ as $p$ increases, which explains the biggest discrepancy at $\theta = \pi/4$ in Supplementary Fig. 4(d).