Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/7/1/eabd2986/DC1

Supplementary Materials for

Zero–trade-off multiparameter quantum estimation via simultaneously saturating multiple Heisenberg uncertainty relations

Zhibo Hou, Jun-Feng Tang, Hongzhen Chen, Haidong Yuan*, Guo-Yong Xiang*, Chuan-Feng Li, Guang-Can Guo

*Corresponding author. Email: hdyuan@mae.cuhk.edu.hk (H.Y.); gyxiang@ustc.edu.cn (G.-Y.X.)

Published 1 January 2021, *Sci. Adv.* **7**, eabd2986 (2021)
DOI: 10.1126/sciadv.abd2986

This PDF file includes:

Supplementary Materials

GENERATORS FOR THE THREE PARAMETERS

The generator for a parameter x is given by $H_x \equiv \mathrm{i}(\partial_x U_s)U_s^\dagger$. For an SU(2) operator $U_s = e^{-\mathrm{i}\alpha \boldsymbol{n}\cdot\boldsymbol{\sigma}}$ with $\boldsymbol{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, the the generators corresponding to α, θ and ϕ can be calculated as

$$H_\alpha = \boldsymbol{n}_\alpha \cdot \boldsymbol{\sigma}, \quad H_\theta = \sin\alpha \boldsymbol{n}_\theta \cdot \boldsymbol{\sigma}, \quad H_\phi = \sin\alpha\sin\theta \boldsymbol{n}_\phi \cdot \boldsymbol{\sigma}$$

here

$$\boldsymbol{n}_\alpha = \boldsymbol{n}, \quad \boldsymbol{n}_\theta = \cos\alpha \boldsymbol{n}_1 + \sin\alpha \boldsymbol{n}_2, \quad \boldsymbol{n}_\phi = -\sin\alpha \boldsymbol{n}_1 + \cos\alpha \boldsymbol{n}_2, \tag{S1}$$

with $\boldsymbol{n}_1 = \partial_\theta \boldsymbol{n} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$, $\boldsymbol{n}_2 = \boldsymbol{n} \times \boldsymbol{n}_1 = (-\sin\phi, \cos\phi, 0)$. $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are Pauli matrices with

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The unit vectors $\boldsymbol{n}, \boldsymbol{n}_1, \boldsymbol{n}_2$ are orthogonal to each other and $\boldsymbol{n}_\theta$ and $\boldsymbol{n}_\phi$ are rotated from $\boldsymbol{n}_1, \boldsymbol{n}_2$ by an angle α along $\boldsymbol{n}$, thus vectors $\boldsymbol{n}, \boldsymbol{n}_\theta, \boldsymbol{n}_\phi$ for the three generators are also orthogonal with each other. The three generators are non-commuting and it is easy to obtain

$$\langle \Delta H_\alpha^2 \rangle \leq 1, \quad \langle \Delta H_\theta^2 \rangle \leq \sin^2\alpha, \quad \langle \Delta H_\phi^2 \rangle \leq \sin^2\alpha \sin^2\theta. \tag{S2}$$

DIFFERENT SCHEMES FOR MULTI-PARAMETER QUANTUM ESTIMATION

We present the common schemes that are used in multi-parameter quantum estimation, and compare them with the optimal scheme implemented in our experiment. In each scheme, the experiments are repeated for the same(n) number of times. This repetition is classical and has the same effect in all schemes, which will thus be neglected.

A simple way to estimate multiple parameters is to estimate each parameter separately. This can be achieved through individual measurements as in Fig. 2A and B, which only estimate one parameter at a time, essentially reducing the multi-parameter quantum estimation to several single-parameter quantum estimations. The scheme in Fig. 2A only uses separable states and separable measurements while the scheme in Fig. 2B allows to use entangled states and collective measurements in each group for the estimation of one parameter. The parallel scheme in Fig. 2C estimates multiple parameters simultaneously, which allows the use of the entangled states and collective measurements. This parallel scheme has been regarded as the scheme that can offer the most quantum advantages, since it can use large entangled probe states. It has been shown that the parallel scheme can outperform the scheme in Fig. 1B with a d-fold improvement for the estimation of $SU(d)$ operators[15-17]. However, due to the existence of the tradeoff, the parallel scheme does not achieve the highest precision simultaneously for all three parameters in the SU(2) operators.

The precision limit for the estimation of one parameter is lower bounded by the reciprocal of the largest uncertainty of its generator, which is

$$\delta\hat{\alpha} \geq \frac{1}{2}, \quad \delta\hat{\theta} \geq \frac{1}{2|\sin\alpha|}, \quad \delta\hat{\phi} \geq \frac{1}{2|\sin\alpha\sin\theta|}. \tag{S3}$$

If N copies of the operators are used, the individual measurement schemes in Fig. 2A,B divide the N operators equally, where each $\frac{N}{3}$ copies are used to estimate one parameter. For the scheme in Fig. 2A , the precision is limited by the shot-noise limit, i.e.,

$$\delta\hat{\alpha} \geq \frac{1}{2\sqrt{\frac{N}{3}}} = \frac{\sqrt{3}}{2\sqrt{N}}, \quad \delta\hat{\theta} \geq \frac{\sqrt{3}}{2\sqrt{N}|\sin\alpha|}, \quad \delta\hat{\phi} \geq \frac{\sqrt{3}}{2\sqrt{N}|\sin\alpha\sin\theta|}. \tag{S4}$$

For the scheme in Fig. 2B, the precision for each parameter can achieve the Heisenberg scaling, as $\frac{N}{3}$ copies of the operator are used to estimate each parameter, we have

$$\delta\hat{\alpha} \geq \frac{1}{2}\frac{1}{\frac{N}{3}} = \frac{3}{2N}, \quad \delta\hat{\theta} \geq \frac{3}{2N|\sin\alpha|}, \quad \delta\hat{\phi} \geq \frac{3}{2N|\sin\alpha\sin\theta|}. \tag{S5}$$

For the scheme in Fig. 2C, the precisions for each parameter are[17]

$$\delta\hat{\alpha} \geq \frac{\sqrt{3}}{2\sqrt{N(N+2)}}, \quad \delta\hat{\theta} \geq \frac{\sqrt{3}}{2\sqrt{N(N+2)}|\sin\alpha|}, \quad \delta\hat{\phi} \geq \frac{\sqrt{3}}{2\sqrt{N(N+2)}|\sin\alpha\sin\theta|}, \tag{S6}$$

which has approximately a $\sqrt{3}$-fold improvement over the individual scheme in Fig. 2B. However, due to the tradeoffs among the precisions of different parameters, the parallel scheme does not achieve the highest precisions for all parameters simultaneously. If N operators are used for the estimation, the Heisenberg limit for each parameter should reach

$$\delta\hat{\alpha} \geq \frac{1}{2N}, \quad \delta\hat{\theta} \geq \frac{1}{2N|\sin\alpha|}, \quad \delta\hat{\phi} \geq \frac{1}{2N|\sin\alpha\sin\theta|}. \tag{S7}$$

This is approximately another $\sqrt{3}$-fold improvement. This can only be achieved if the tradeoffs among the precisions of different parameters are counteracted, as the highest precisions for three parameters with N copies of the operator are achieved simultaneously. Our control-enhanced sequential scheme achieves this. To highlight the importance of the controls in our method, we show that without controls the sequential measurements in our scheme can even perform worse than the individual measurement in Fig. 2A . Here we calculate the variance of three generators $H_x^{(N)}$ with $x \in \{\alpha, \theta, \phi\}$ when N copies of operators are used without controls. As in Eq. (6), we have $H_x^{(N)} = \sum_{k=0}^{N-1} U_s^k H_x \left[U_s^k\right]^\dagger$. For $x = \alpha$, H_α commutes with U_s, thus $H_\alpha^{(N)} = NH_\alpha$. For $x \in \{\theta, \phi\}$, H_x does not commute with U_s and the Bloch vector of $U_s^k H_x \left[U_s^k\right]^\dagger$ is rotated from $\boldsymbol{n}_x$ along $\boldsymbol{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ by an angle of $2k\alpha$. We then get $H_x^{(N)} = \boldsymbol{n}_x^{(N)} \cdot \boldsymbol{\sigma}$ with

$$\begin{aligned}
\boldsymbol{n}_\theta^{(N)} &= \sin\alpha \sum_{k=0}^{N-1} (\cos 2k\alpha \boldsymbol{n}_\theta + \sin 2k\alpha \boldsymbol{n}_\phi) = \sin N\alpha[\cos(N-1)\alpha \boldsymbol{n}_\theta + \sin(N-1)\alpha \boldsymbol{n}_\phi], \\
\boldsymbol{n}_\phi^{(N)} &= \sin\alpha\sin\theta \sum_{k=0}^{N-1} (\cos 2k\alpha \boldsymbol{n}_\phi - \sin 2k\alpha \boldsymbol{n}_\theta) = \sin N\alpha \sin\theta \left[\cos(N-1)\alpha \boldsymbol{n}_\phi - \sin(N-1)\alpha \boldsymbol{n}_\theta\right].
\end{aligned} \tag{S8}$$

We note that the variance of a generator is upper bounded by $\langle \Delta \left[H_x^{(N)}\right]^2 \rangle \leq \mathrm{tr}\left[H_x^{(N)}\right]^2 = \left|\boldsymbol{n}_x^{(N)}\right|^2$, which is $N^2, \sin^2 N\alpha$ and $\sin^2 N\alpha \sin^2\theta$ for α, θ and ϕ, respectively. Hence, the highest precisions achievable for the three parameters in the sequential scheme without control are

$$\delta\hat{\alpha} \geq \frac{1}{2N}, \quad \delta\hat{\theta} \geq \frac{1}{2|\sin N\alpha|}, \quad \delta\hat{\phi} \geq \frac{1}{2|\sin N\alpha\sin\theta|} \tag{S9}$$

respectively. Except for the estimation of α which achieves the Heisenberg limit, the direct sequential scheme without control can have worse precisions than the shot-noise limit, which scales as $\frac{1}{\sqrt{N}}$, for the other two parameters.

The precision of different schemes can also be characterized by the quantum Fisher information matrix[19]. With N uses of the operators, the quantum Fisher information matrices of the classical individual, entangled individual, entangled simultaneous scheme shown in Fig.2A,B,C, and the control-enhanced sequential scheme shown in Fig.1A are

$$J_1 = \frac{N}{3}J_0, \quad J_2 = \frac{N^2}{9}J_0, \quad J_3 = \frac{N(N+2)}{3}J_0, \quad J_4 = N^2 J_0, \tag{S10}$$

respectively, here

$$J_0 = 4\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin^2\alpha & 0 \\ 0 & 0 & \sin^2\alpha\sin^2\theta \end{pmatrix}. \tag{S11}$$

For any weighted matrix, W, we have $\mathrm{Tr}(W\mathrm{Cov}(\alpha,\theta,\phi)) \geq Tr(WJ^{-1})$. It can be seen that compared to the schemes of classical individual, entangled individual and entangled simultaneous, the control-enhanced sequential scheme has a constant improvement of $3N$, 9 and $3N/(N+2)$ respectively. Then at $N = 8$, for any weighted matrix the control-enhanced sequential scheme has a $10Log_{10}(3N) = 13.8dB$ improvement over the classical individual scheme.

OPTIMAL PROBE STATE

The probe state $|\Psi\rangle$, chosen as the maximally entangled state, can maximize the variances of all three generators simultaneously.

With the maximally entangled state, the variance of a generator is given by

$$\langle H_x^2\rangle = \langle\Psi|H_x^2|\Psi\rangle - \langle\Psi|H_x|\Psi\rangle^2 = \frac{1}{2}\,\mathrm{tr}(H_x^2) - \frac{1}{4}(\mathrm{tr}\,H_x)^2,$$

here H_x only acts on the probe qubit, but not the ancillary qubit, and we have used the fact that the reduced probe state of the maximally entangled state is the Identity. For the three generators

$$H_\alpha = \boldsymbol{n}_\alpha\cdot\boldsymbol{\sigma}, \quad H_\theta = \sin\alpha\boldsymbol{n}_\theta\cdot\boldsymbol{\sigma}, \quad H_\phi = \sin\alpha\sin\theta\boldsymbol{n}_\phi\cdot\boldsymbol{\sigma},$$

where $\boldsymbol{n}_\alpha, \boldsymbol{n}_\theta, \boldsymbol{n}_\phi$ are unit vectors given in Eq.(S1), the variance of the generators with the maximally entangled state are thus given by

$$\langle H_\alpha^2\rangle = \frac{1}{2}\,\mathrm{tr}(H_\alpha^2) = 1, \quad \langle H_\theta^2\rangle = \frac{1}{2}\,\mathrm{tr}(H_\theta^2) = \sin^2\alpha, \quad \langle H_\phi^2\rangle = \frac{1}{2}\,\mathrm{tr}(H_\phi^2) = \sin^2\alpha\sin^2\theta,$$

which saturate the upper bounds in Inequality (S2) simultaneously.

OPTIMAL MEASUREMENT

We show the projective measurements on a set of maximally entangled states are optimal for the estimation of all three parameters.

Without loss of generality, the optimal probe state is chosen as $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, with the first qubit representing system and the second as ancilla. As $\langle H_x\rangle = 0$ and γ in Eq. (3) is an arbitrary real scalar, the condition of observable O_x in Eq. (3) becomes

$$\boldsymbol{n}_x\cdot\boldsymbol{\sigma}(|00\rangle + |11\rangle) = \mathrm{i}\gamma(O_x - \langle O_x\rangle)(|00\rangle + |11\rangle). \tag{S12}$$

We first show that when $\boldsymbol{n}_x\cdot\boldsymbol{\sigma}$ are taken as the Pauli operators on the system qubit, the corresponding optimal observables commute with each other. With this we then show for arbitrary $\boldsymbol{n}_x\cdot\boldsymbol{\sigma} = n_{x,1}\sigma_1 + n_{x,2}\sigma_2 + n_{x,3}\sigma_3$ acts on the system qubit, the corresponding observables commute with each other.

When $\boldsymbol{n}_x\cdot\boldsymbol{\sigma}$ is taken as the three Pauli operators respectively, the left side of Eq.(S12) becomes

$$\begin{aligned} XI(|00\rangle + |11\rangle) &= (|01\rangle + |10\rangle),\\ YI(|00\rangle + |11\rangle) &= -\mathrm{i}(|01\rangle - |10\rangle),\\ ZI(|00\rangle + |11\rangle) &= (|00\rangle - |11\rangle), \end{aligned} \tag{S13}$$

where we use X, Y, Z to represent $\sigma_1, \sigma_2, \sigma_3$ for simplicity. The three observables can be chosen as ZY, XZ and YX, as their operations on the state are

$$\begin{aligned} ZY(|00\rangle + |11\rangle) &= \mathrm{i}(|01\rangle + |10\rangle) = \mathrm{i}XI(|00\rangle + |11\rangle),\\ XZ(|00\rangle + |11\rangle) &= (|01\rangle - |10\rangle) = \mathrm{i}YI(|00\rangle + |11\rangle),\\ YX(|00\rangle + |11\rangle) &= -\mathrm{i}(|00\rangle - |11\rangle) = -\mathrm{i}ZI(|00\rangle + |11\rangle). \end{aligned} \tag{S14}$$

Since the average of these three observables under the maximally entangled state is 0, i.e., $\langle ZY\rangle = \langle XZ\rangle = \langle YX\rangle = 0$, thus, ZY, XZ and YX are observables satisfying Eq.(S12) for generators XI, YI

and ZI, respectively. It is easy to check these three observables commute with each other, and thus they can be implemented simultaneously. Their common eigenvectors are four maximally entangled states rotated from four Bell states by a local operation $U_0 = e^{\mathrm{i}\frac{\pi}{3}\frac{1}{\sqrt{3}}(\sigma_x+\sigma_y+\sigma_z)}$ on the system qubit. This can also be understood intuitively as ZY, XZ and YX are rotated from YY, ZZ and XX along a direction of $\frac{1}{\sqrt{3}}(1,1,1)$ by an angle of $\frac{2}{3}\pi$ and note that the common eigenvectors of YY, ZZ and XX are the four Bell states.

For three general generators, the same maximally entangled projective measurements for ZY, XZ and YX also satisfies Eq.(S12) because each observable can be constructed as a linear combination of ZY, XZ and YX. For example, if a generator is given by $\boldsymbol{n}_x \cdot \boldsymbol{\sigma} = n_{x,1}\sigma_1 + n_{x,2}\sigma_2 + n_{x,3}\sigma_3$, then the observable $O_x = \boldsymbol{r}_x \cdot (ZY, XZ, YX)$ with $\boldsymbol{r}_x = (n_{x,1}, n_{x,2}, -n_{x,3})$ satisfies Eq.(S12). The observables for the three parameters of the SU(2) operator can all be taken as the linear combinations of ZY, XZ and YX. Thus, the optimal measurement can be taken as the projective measurement on the common eigenvectors of ZY, XZ and YX, which is implemented in our experiment.

We note this is not the only choice. Actually the projective measurements on any set of maximally entangled states can satisfy Eq.(S12), which we now show. First, any set of maximally entangled states can be represented as the common eigenvectors of Z_1Y_2, X_2Z_2 and Y_2X_2, where $\{X_1, Y_1, Z_1\}$ acts on the system qubit and $\{X_2, Y_2, Z_2\}$ acts on the ancilla qubit, which are obtained from the Pauli operators, $\{X, Y, Z\}$, with local operations U_1 on the system qubit and U_2 on the ancilla qubit respectively. These local operations can be represented as rotations, R^1 and R^2, in a three dimensional space as $X_k = (R^k_{11}, R^k_{21}, R^k_{31}) \cdot \boldsymbol{\sigma}, Y_k = (R^k_{12}, R^k_{22}, R^k_{32}) \cdot \boldsymbol{\sigma}, Z_k = (R^k_{13}, R^k_{23}, R^k_{33}) \cdot \boldsymbol{\sigma}$ for $k = 1, 2$. The operations of Z_1Y_2, X_2Z_2 and Y_2X_2 on the probe state are

$$\begin{aligned}
Z_1Y_2(|00\rangle + |11\rangle) &= [(R^1_{33}R^2_{22} + R^1_{32}R^2_{23})ZY + (R^1_{31}R^2_{23} - R^1_{33}R^2_{21})XZ + (R^1_{32}R^2_{21} + R^1_{31}R^2_{22})YX \\
&\quad + (R^1_{31}R^2_{21} - R^1_{32}R^2_{22} + R^1_{33}R^2_{23})](|00\rangle + |11\rangle), \\
X_1Z_2(|00\rangle + |11\rangle) &= [(R^1_{13}R^2_{32} + R^1_{12}R^2_{33})ZY + (R^1_{11}R^2_{33} - R^1_{13}R^2_{31})XZ + (R^1_{12}R^2_{31} + R^1_{11}R^2_{32})YX \\
&\quad + (R^1_{11}R^2_{31} - R^1_{12}R^2_{32} + R^1_{13}R^2_{33})](|00\rangle + |11\rangle), \\
Y_1X_2(|00\rangle + |11\rangle) &= [(R^1_{23}R^2_{12} + R^1_{22}R^2_{13})ZY + (R^1_{21}R^2_{13} - R^1_{23}R^2_{11})XZ + (R^1_{22}R^2_{11} + R^1_{21}R^2_{12})YX \\
&\quad + (R^1_{21}R^2_{11} - R^1_{22}R^2_{12} + R^1_{23}R^2_{13})](|00\rangle + |11\rangle).
\end{aligned} \tag{S15}$$

This can be rewritten in a compact form as

$$\begin{pmatrix} Z_1Y_2 - \langle Z_1Y_2\rangle \\ X_1Z_2 - \langle X_1Z_2\rangle \\ Y_1X_2 - \langle Y_1X_2\rangle \end{pmatrix} (|00\rangle + |11\rangle) = T \begin{pmatrix} ZY \\ XZ \\ YX \end{pmatrix} (|00\rangle + |11\rangle) \tag{S16}$$

with

$$T = \begin{pmatrix} R^1_{33}R^2_{22} + R^1_{32}R^2_{23} & R^1_{31}R^2_{23} - R^1_{33}R^2_{21} & R^1_{32}R^2_{21} + R^1_{31}R^2_{22} \\ R^1_{13}R^2_{32} + R^1_{12}R^2_{33} & R^1_{11}R^2_{33} - R^1_{13}R^2_{31} & R^1_{12}R^2_{31} + R^1_{11}R^2_{32} \\ R^1_{23}R^2_{12} + R^1_{22}R^2_{13} & R^1_{21}R^2_{13} - R^1_{23}R^2_{11} & R^1_{22}R^2_{11} + R^1_{21}R^2_{12} \end{pmatrix}. \tag{S17}$$

Note that T is generated from U_1 and U_2 and is invertible. Thus

$$T^{-1} \begin{pmatrix} Z_1Y_2 - \langle Z_1Y_2\rangle \\ X_1Z_2 - \langle X_1Z_2\rangle \\ Y_1X_2 - \langle Y_1X_2\rangle \end{pmatrix} (|00\rangle + |11\rangle) = \begin{pmatrix} ZY \\ XZ \\ YX \end{pmatrix} (|00\rangle + |11\rangle). \tag{S18}$$

Any linear combination of ZY, XZ, YX on the maximal entangled state can thus also be written as a linear combination of Z_1Y_2, X_1Z_2, Y_1X_2 on the maximal entangled state. The observable can just also be written as a linear combination of Z_1Y_2, X_1Z_2, Y_1X_2 with the corresponding measurement as the projective measurement on the common eigenvectors of Z_1Y_2, X_1Z_2, Y_1X_2.

EXPERIMENT REALIZATION OF THE DETERMINISTIC COMPLETE BELL MEASUREMENTS

The complete Bell measurements on the polarization and the path of the photon are realized by a two-step quantum walk followed by a measurement on the polarization.

The system state in a one-dimensional discrete quantum walk can be described by two degrees of freedom $|x, c\rangle$ with $x = \cdots, -1, 0, 1, \cdots$ representing the walker position and $c = 0, 1$ denoting the coin state. In each step of the quantum walk, there are site-dependent coin operations $C(x, t)$ and the conditional translation operation

$$T = \sum_x |x+1, 0\rangle\langle x, 0| + |x-1, 1\rangle\langle x, 1|, \tag{S19}$$

adding up to a total unitary transformation $T \sum_x |x\rangle\langle x| \otimes C(x, t)$. In the following, we will calculate the realized measurements when coin operators are engineered and realized as in Fig. S1.

The coin operators are

$$\begin{aligned} C(-1,1) = C(-2,2) = C(2,2) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ C(-1,3) = C(1,3) &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \\ C(1,1) = C(0,2) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{S20}$$

Thus, the total dynamics is

$$\begin{aligned} T_a &= \sum_{x=-1,1} |x\rangle\langle x| \otimes C(x,3) T \sum_{x=-2,0,2} |x\rangle\langle x| \otimes C(x,2) T \sum_{x=-1,1} |x\rangle\langle x| \otimes C(x,1) \\ &= |1,0\rangle \frac{\langle 1,0| + \langle -1,1|}{\sqrt{2}} + |-1,0\rangle \frac{\langle 1,1| + \langle -1,0|}{\sqrt{2}} - |-1,1\rangle \frac{\langle 1,1| - \langle -1,0|}{\sqrt{2}} - |1,1\rangle \frac{\langle 1,0| - \langle -1,1|}{\sqrt{2}}. \end{aligned} \tag{S21}$$

The measurement probabilities of the four outcomes for any input state ρ are thus

$$P_1 = \langle 1,0|T_a \rho T_a^\dagger|1,0\rangle = \mathrm{tr}(\rho T_a^\dagger|1,0\rangle\langle 1,0|T_a) = \mathrm{tr}(\rho E_0), \tag{S22}$$

$$P_2 = \langle -1,0|T_a \rho T_a^\dagger|-1,0\rangle = \mathrm{tr}(\rho T_a^\dagger|-1,0\rangle\langle -1,0|T_a) = \mathrm{tr}(\rho E_1), \tag{S23}$$

$$P_3 = \langle -1,1|T_a \rho T_a^\dagger|-1,1\rangle = \mathrm{tr}(\rho T_a^\dagger|-1,1\rangle\langle -1,1|T_a) = \mathrm{tr}(\rho E_2), \tag{S24}$$

$$P_4 = \langle 1,1|T_a \rho T_a^\dagger|1,1\rangle = \mathrm{tr}(\rho T_a^\dagger|1,1\rangle\langle 1,1|T_a) = \mathrm{tr}(\rho E_3), \tag{S25}$$

where

$$\begin{aligned} E_1 &= T_a^\dagger|1,0\rangle\langle 1,0|T_a = \frac{1}{2}(|1,0\rangle + |-1,1\rangle)(\langle 1,0| + \langle -1,1|) = \left|\Phi^+\right\rangle\left\langle\Phi^+\right|, \\ E_2 &= T_a^\dagger|-1,0\rangle\langle -1,0|T_a = \frac{1}{2}(|1,1\rangle + |-1,0\rangle)(\langle 1,1| + \langle -1,0|) = \left|\Psi^+\right\rangle\left\langle\Psi^+\right|, \\ E_3 &= T_a^\dagger|-1,1\rangle\langle -1,1|T_a = \frac{1}{2}(|1,1\rangle - |-1,0\rangle)(\langle 1,1| - \langle -1,0|) = \left|\Psi^-\right\rangle\left\langle\Psi^-\right|, \\ E_4 &= T_a^\dagger|1,1\rangle\langle 1,1|T_a = \frac{1}{2}(|1,0\rangle - |-1,1\rangle)(\langle 1,0| - \langle -1,1|) = \left|\Phi^-\right\rangle\left\langle\Phi^-\right| \end{aligned} \tag{S26}$$

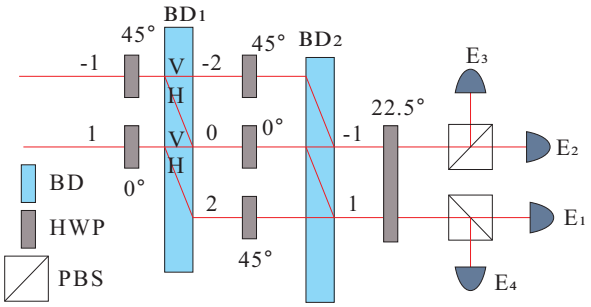

FIG. S1. **Bell measurement scheme using photonic quantum walks.**

with four Bell states

$$
\begin{aligned}
|\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|1,0\rangle + |-1,1\rangle), \\
|\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|1,1\rangle + |-1,0\rangle), \\
|\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|1,1\rangle - |-1,0\rangle), \\
|\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|1,0\rangle - |-1,1\rangle).
\end{aligned} \tag{S27}
$$

It can be directly seen from Eqs. (S27) that the measurement scheme in Fig. S1 can realize the complete set of Bell measurements on the path qubit with bases 1 (down mode) and -1 (up mode) and polarization qubit with bases 0 (Horizontal polarization) and 1 (Vertical polarization).

The quality of the actually realized measurement operators is experimentally verified as the interference visibility when input states are the four Bell states. The experimentally measured visibilities are as high as 99.8% with N up to 8.

THE OPTIMAL OBSERVABLE IN THE UNCERTAINTY RELATION AND THE SLD

In the main text we have shown that the condition for an observable O_x to achieve the minimal uncertainty for a particular parameter x is

$$
(H_x - \langle H_x\rangle)|\Psi_x\rangle = i\gamma O_x|\Psi_x\rangle, \tag{S28}
$$

where γ is arbitrary real scalar, and without loss of generality we have shifted the mean value of the observable O_x to zero, i.e. $\langle O_x\rangle = 0$. The optimal measurement that achieves the highest precision corresponds to the set of projection on the eigenvectors of O_x.

With the statistical approach, Helstrom has shown that the necessary and sufficient condition for a set of POVM, $\{E_\xi\}$, to saturate the quantum Cramér-Rao bound is given by[1]

$$
E_\xi^{1/2}\rho_x^{1/2} = \lambda_\xi E_\xi^{1/2} L_x \rho_x^{1/2}, \lambda_\xi \in \mathbb{R} \quad \forall \xi, \tag{S29}
$$

where L_x is the symmetric logarithmic derivative (SLD) which can be obtained by solving the equation $\partial\rho_x/\partial x = (\rho_x L_x + L_x\rho_x)/2$. For a pure state $\rho_x = |\Psi_x\rangle\langle\Psi_x|$, it is straightforward to verify that

$$
L_x = 2(|\partial_x\Psi_x\rangle\langle\Psi_x| + |\Psi_x\rangle\langle\partial_x\Psi_x|) = 2\mathrm{i}\,[|\Psi_x\rangle\langle\Psi_x|, H_x] \tag{S30}
$$

is a solution, where for the second equality we used the fact that $|\partial_x\Psi_x\rangle = -iH_x|\Psi_x\rangle$ which can be obtained directly from the definition of the generator H_x. Note that the condition in Eq.(S29) does not imply that E_ξ has to be the projective measurement on the eigen-vector of $L_x = 2\mathrm{i}\,[|\Psi_x\rangle\langle\Psi_x|, H_x]$. Actually only if ρ_x is full rank then we can get rid of ρ_x in Eq.(S29) to obtain $E_\xi^{1/2} = \lambda_\xi E_\xi^{1/2} L_x$ which implies that E_ξ is the projection on the eigen-vector of L_x, if it is a projective measurement. But if ρ_x is not full rank, E_ξ does not have to be the projection on the eigenvector of the given L_x. In particular, for pure states the optimal measurement does not have to be the projective measurement on the eigen-vector of $L_x = 2\mathrm{i}\,[|\Psi_x\rangle\langle\Psi_x|, H_x]$.

The condition provided by the uncertainty relation is equivalent to the necessary and sufficient

condition in Eq.(S29). By writing the projective measurement as $E_\xi = |o_\xi\rangle\langle o_\xi|$, we have

$$\begin{aligned}
&E_\xi^{1/2}\rho_x^{1/2} = \lambda_\xi E_\xi^{1/2} L_x \rho_x^{1/2}, \lambda_\xi \in \mathbb{R} \quad \forall\xi, \\
\Leftrightarrow &|o_\xi\rangle\langle o_\xi|\Psi_x\rangle\langle\Psi_x| = \lambda_\xi |o_\xi\rangle\langle o_\xi| L_x |\Psi_x\rangle\langle\Psi_x|, \lambda_\xi \in \mathbb{R} \quad \forall\xi, \\
\Leftrightarrow &|o_\xi\rangle\langle o_\xi|\Psi_x\rangle\langle\Psi_x| = -2\mathrm{i}\lambda_\xi |o_\xi\rangle\langle o_\xi| \left(H_x - \langle H_x\rangle\right) |\Psi_x\rangle\langle\Psi_x|, \lambda_\xi \in \mathbb{R} \quad \forall\xi, \\
\Leftrightarrow &|o_\xi\rangle\langle o_\xi| \left(H_x - \langle H_x\rangle\right) |\Psi_x\rangle\langle\Psi_x| = \mathrm{i}\frac{1}{2\lambda_\xi} |o_\xi\rangle\langle o_\xi|\Psi_x\rangle\langle\Psi_x|, \lambda_\xi \in \mathbb{R} \quad \forall\xi, \\
\Leftrightarrow &\sum_\xi \left(|o_\xi\rangle\langle o_\xi|\right) \left(H_x - \langle H_x\rangle\right) |\Psi_x\rangle\langle\Psi_x| = \mathrm{i}\sum_\xi \left(\frac{1}{2\lambda_\xi}|o_\xi\rangle\langle o_\xi|\right) |\Psi_x\rangle\langle\Psi_x|, \lambda_\xi \in \mathbb{R}, \\
\Leftrightarrow &\left(H_x - \langle H_x\rangle\right) |\Psi_x\rangle = \mathrm{i}O_x|\Psi_x\rangle,
\end{aligned} \tag{S31}$$

where the hermitian operator $O_x \equiv \sum_\xi \left(\frac{1}{2\lambda_\xi}|o_\xi\rangle\langle o_\xi|\right)$ can be viewed as an observable satisfying Eq.(S28) with $\gamma = 1$. Here without loss of generality we have assumed that $\lambda_\xi \neq 0$ since if $\lambda_\xi = 0$, the corresponding probability $p(\xi) = \mathrm{tr}(E_\xi \rho_x) = 0$ and will not contribute to the quantum Fisher information. Therefore, each measurement in Eq.(S29) corresponds to an observable satisfying Eq.(S28). But again they do not have to be the projections on the eigenvectors of $L_x = 2\mathrm{i}\left[|\Psi_x\rangle\langle\Psi_x|, H_x\right]$. In fact in our case the three SLDs, $L_\alpha = 2i[|\Psi(\alpha,\theta,\phi)\rangle\langle\Psi(\alpha,\theta,\phi)|, H_\alpha]$, $L_\theta = 2i[|\Psi(\alpha,\theta,\phi)\rangle\langle\Psi(\alpha,\theta,\phi)|, H_\theta]$ and $L_\phi = 2i[|\Psi(\alpha,\theta,\phi)\rangle\langle\Psi(\alpha,\theta,\phi)|, H_\phi]$ actually do not commute with each other, thus one can not obtain the optimal measurement from these three SLDs.

To see these three SLDs do not commute, we write

$$L_x = 2\mathrm{i}\left[|\Psi(\alpha,\theta,\phi)\rangle\langle\Psi(\alpha,\theta,\phi)|, H_x\right] = 2\mathrm{i}U_s[|\Psi\rangle\langle\Psi|, H'_x]U_s^\dagger, \tag{S32}$$

where $|\Psi(\alpha,\theta,\phi)\rangle = U_s|\Psi\rangle$ with $|\Psi\rangle$ as the initial probe state, $H'_x = U_s^\dagger H_x U_s = \mathrm{i}U_s^\dagger(\partial_x U_s)$, here $x \in \{\alpha,\theta,\phi\}$. Then

$$\begin{aligned}
[L_\alpha, L_\theta] =& -4U_s\left(\langle\Psi|H'_\alpha|\Psi\rangle\left[|\Psi\rangle\langle\Psi|, H'_\theta\right] - \langle\Psi|H'_\theta|\Psi\rangle\left[|\Psi\rangle\langle\Psi|, H'_\alpha\right]\right)U_s^\dagger \\
&+ 4\langle\Psi|\left[H'_\alpha, H'_\theta\right]|\Psi\rangle U_s|\Psi\rangle\langle\Psi|U_s^\dagger + 4U_s(H'_\alpha|\Psi\rangle\langle\Psi|H'_\theta - H'_\theta|\Psi\rangle\langle\Psi|H'_\alpha)U_s^\dagger \\
=&4U_s(H'_\alpha|\Psi\rangle\langle\Psi|H'_\theta - H'_\theta|\Psi\rangle\langle\Psi|H'_\alpha)U_s^\dagger \neq \mathbb{O},
\end{aligned} \tag{S33}$$

for the second equality we used $\langle\Psi|H'_\alpha|\Psi\rangle = \mathrm{tr}(H'_\alpha) = 0$, $\langle\Psi|H'_\theta|\Psi\rangle = \mathrm{tr}(H'_\theta) = 0$ and $\langle\Psi|\left[H'_\alpha, H'_\theta\right]|\Psi\rangle = \langle\Psi|H'_\phi|\Psi\rangle = \mathrm{tr}(H'_\phi) = 0$. It is also easy to verify that the forth term is not a zero matrix $\mathbb{O}$ by specifying one matrix element of it, i.e.

$$\begin{aligned}
\langle 01|U_s^\dagger\left[L_\alpha, L_\theta\right]U_s|00\rangle &= 4\langle 01|(H'_\alpha|\Psi\rangle\langle\Psi|H'_\theta - H'_\theta|\Psi\rangle\langle\Psi|H'_\alpha)|00\rangle \\
&= -2e^{-\mathrm{i}\phi}\sin\alpha(\cos\alpha + \mathrm{i}\cos\theta\sin\alpha) \neq 0.
\end{aligned} \tag{S34}$$

Similarly we have $[L_\theta, L_\phi] \neq \mathbb{O}$, and $[L_\phi, L_\alpha] \neq \mathbb{O}$. The optimal measurements thus can not be the projective measurements on the common eigen-vectors of these three SLDs.

THE EFFECTS OF THE LOSS

Here we consider two noises that lead to the photon loss in the control-enhanced sequential scheme and their effects on the precision. The first is the channel loss. For one pass through the channel (i.e., the photon goes one loop), there are four mirrors (MT-optics, coating: HR@808 nm, reflectivity 0.999) and six wave plates (MT-optics, coating: AR@808 nm, transmissivity 0.998). The transmission efficiency is then $\eta_1 = 0.999^4 \times 0.998^6 = 0.984$ when the photon goes through one loop. For N passes the transmission efficiency is η_1^N. The second is the loss in the measurement. In our experiment the collection efficiency of one photon is $\eta_2 = 0.48$, which includes the coupling efficiency(0.80) and the detection efficiency(0.60). After including these noises, with N loops the covariance error matrix of the estimation becomes $\eta_1^{-N}\eta_2^{-1}N^{-2}J_0^{-1}$. We can see that η_1 and η_2 have different effects on the scaling, where η_2 only affects the precision with a constant factor, η_1 leads to an exponential factor. When $N \approx -1/\log\eta_1 = 62$ the precision becomes close to the shot noise limit $N^{-1/2}$. For

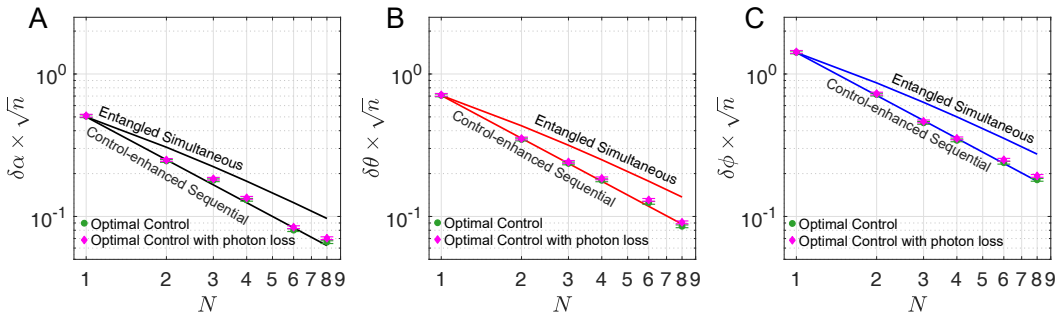

FIG. S2. **Control-enhanced sequential scheme with channel loss.** The channel loss for N passes degrades the standard deviation by a factor of $\eta_1^{N/2}$, which is only slightly larger than the lossless case since $\eta_1 = 0.984$ in the experiment is quite high.

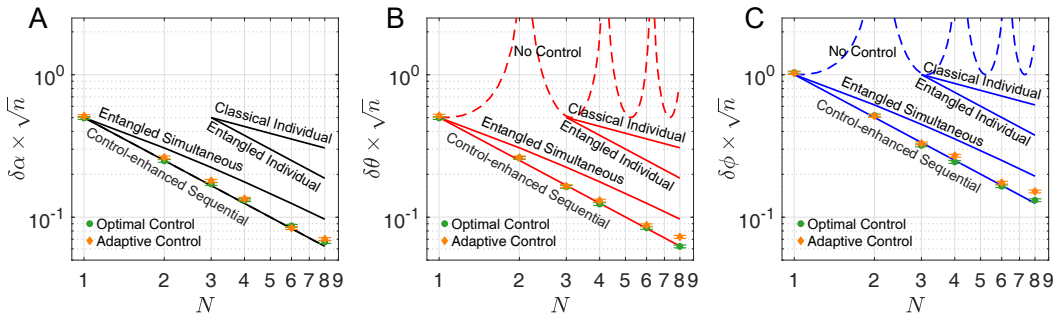

FIG. S3. **Experimental results for U_s with $\alpha = 1.5$ and $\theta = \phi = \frac{\pi}{6}$.**

$1 \leq N \leq 62$, the scaling of the precision is between the Heisenberg limit and the shot noise limit. In our experiment, the scalings range from $N^{-0.992}$ to $N^{-0.936}$ for $1 \leq N \leq 8$, which are very close to the Heisenberg scaling, as illustrated in Fig. S2.

For the classical individual measurement, at the presence of the loss the covariance error matrix is given by $3/(\eta_1\eta_2 N)J_0^{-1}$, here the transmission efficiency $\eta_1 = 0.998^3 = 0.994$ since only three wave plates are used in the classical scheme, the detection efficiency is the same, which is given by $\eta_2 = 0.48$. At the presence of these losses, the control-enhanced sequential scheme then has a 13.27dB(in terms of variance) improvement over the classical scheme at $N = 8$. If we compare with the lossless classical scheme, the improvement of the control-enhanced sequential scheme(at the presence of the stated loss) at $N = 8$ is 10.05dB (in terms of variance).

TABLE S1. **Quantitative distances $(\delta\hat{x} - \delta x) \times \sqrt{n}$ between the experimentally simultaneously measured $\delta\hat{x}$ and the theoretical optimal precision for each parameter δx $(x = \alpha, \theta, \phi)$ in Fig. 4.**

$(\delta\hat{x} - \delta x) \times \sqrt{n}$		$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 6$	$N = 8$
Optimal Control	α	0.006 ± 0.011	-0.005 ± 0.005	0.013 ± 0.004	0.006 ± 0.003	-0.003 ± 0.002	0.004 ± 0.002
	θ	0.001 ± 0.016	-0.007 ± 0.008	0.000 ± 0.005	0.002 ± 0.004	0.007 ± 0.003	-0.003 ± 0.002
	ϕ	0.004 ± 0.032	0.010 ± 0.016	-0.017 ± 0.010	-0.011 ± 0.008	0.003 ± 0.005	0.004 ± 0.004
Adaptive Control	α	0.006 ± 0.011	0.001 ± 0.006	0.008 ± 0.004	0.003 ± 0.003	0.002 ± 0.002	0.010 ± 0.002
	θ	0.009 ± 0.016	0.002 ± 0.008	0.001 ± 0.005	0.028 ± 0.005	0.010 ± 0.003	0.012 ± 0.002
	ϕ	0.09 ± 0.03	0.06 ± 0.02	0.02 ± 0.01	0.006 ± 0.008	0.040 ± 0.006	0.029 ± 0.005