# Experimental generalized contextuality with single-photon qubits 

Xiang Zhan, ${ }^{1}$ Eric G. Cavalcanti, ${ }^{2}$ Jian Li, ${ }^{1}$ Zhihao Bian, ${ }^{1}$<br>Yongsheng Zhang, ${ }^{3,4, *}$ Howard M. Wiseman, ${ }^{2,5, 母}$, and Peng Xue ${ }^{1,6, ~}{ }^{\text {, }}$<br>${ }^{1}$ Department of Physics, Southeast University, Nanjing 211189, China<br>${ }^{2}$ Centre for Quantum Dynamics, Griffith University, Gold Coast, QLD 4222, Australia<br>${ }^{3}$ Key Laboratory of Quantum Information, University of Science and Technology of China, CAS, Hefei 230026, China<br>${ }^{4}$ Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China<br>${ }^{5}$ Centre for Quantum Computation and Communication Technology<br>(Australian Research Council), Griffith University, Brisbane 4111, Australia<br>${ }^{6}$ State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai 200062, China


#### Abstract

Contextuality is a phenomenon at the heart of the quantum mechanical departure from classical behaviour, and has been recently identified as a resource in quantum computation. Experimental demonstration of contextuality is thus an important goal. The traditional form of contextuality - as violation of a Kochen-Specker inequality - requires a quantum system with at least three levels, and the status of the assumption of determinism used in deriving those inequalities has been controversial. By considering 'unsharp' observables, Liang, Spekkens and Wiseman (LSW) derived an inequality for generalized noncontextual models that doesn't assume determinism, and applies already to a qubit. We experimentally implement the LSW test using the polarization states of a heralded single photon and three unsharp binary measurements. We violate the LSW inequality by more than 16 standard deviations, thus showing that our results cannot be reproduced by a noncontextual subset of quantum theory.


## I. INTRODUCTION

There are a number of proposals for tests which pit quantum mechanics against alternative views of reality, including the theorems of Bell [1] and of Kochen and Specker (KS) [2]. Corresponding experimental tests [3-8] have been performed and support the validity of quantum mechanics. Bell's theorem refers to a situation with two or more spatially separate particles and states that local hidden variable theories are incompatible with the statistical predictions of quantum mechanics. The KS theorem has the advantage of applying to a single system, and states that noncontextual hidden variable theories are incompatible with quantum predictions, under the assumption that the measurements can be described by projectors. A qutrit (three-level system) and five projectors are required for a proof of the traditional KS contextuality in a state-dependent manner [9, 10], while a qutrit and thirteen projectors for such a proof in a stateindependent manner [11-17].

To find simpler proofs of contextuality, applicable to a qubit (two-level system), generalizations of KS noncontextuality have been proposed [18 21]. These all utilise generalized measurements, described by positive operator-valued measures (POVMs). It has been argued, however [22] that these works make an unwarranted assumption of determinism for unsharp measurements.

[^0]More recently, Liang, Spekkens and Wiseman (LSW) [23] (Sec. 7.3) followed a different approach to derive noncontextuality inequalities for a particular class of non-projective measurements. The relevant class is the unsharp projective measurements, in which each of the set of orthogonal projectors is mixed in some ratio with other projectors from the same set, in order to make the POVM. (Thus each element of the POVM commutes with each other element, just as for a projective measurement.) The LSW assumption is that the response function is likewise a mixture of the deterministic response functions assumed by KS for projective measurements, in the same ratios. Using this principle, LSW derived a generalized noncontextuality inequality involving three different unsharp projective measurements on a qubit. Subsequently, Kunjwal and Ghosh [24] found a triple of unsharp observables that, according to the predictions of quantum mechanics, would give a significant violation of the LSW inequality, in a state-dependent manner.

Here, we experimentally violate the LSW inequality for the first time, via three unsharp binary qubit measurements that are pairwise jointly measurable. We use a photon polarization qubit, and the scheme of Ref. [24]. Our work verifies experimentally that even a single qubit is enough to demonstrate quantum contextuality, under the weak assumptions of Ref. [23]. As we assume the validity of operational quantum theory for the error analysis, our work demonstrates that our results cannot be reproduced by a noncontextual fragment of quantum theory - an important experimental benchmark. We exceed the LSW bound by many standard deviations, in an experimentally verified regime of validity for the inequality.

We note that an independent experimental demonstra-
tion of contextuality with qubit systems, following techniques complementary to the present work, is reported in 25]. There, the state preparations and measurements are realized with time-sharing methods, and the problem of noises in measurements is solved with a technique derived within the framework of generalised probabilistic theories.

## II. THEORETICAL IDEA

## A. Scheme for violating the LSW inequality

A generalized noncontextual model, referred to as a LSW model, can be realized using noisy spin- $\frac{1}{2}$ observables [23]. Specifically, three such observables, $\mathcal{M}_{k}(k=$ $1,2,3$ ), are required, each described by a two-outcome POVM, $\mathcal{M}_{k}=\left\{E_{+}^{k}, E_{-}^{k}\right\}$, of the form [23]

$$
\begin{equation*}
E_{ \pm}^{k} \equiv \frac{1}{2} \mathbb{1} \pm \frac{\eta}{2} \vec{\sigma} \cdot \hat{n}_{k}=\frac{1-\eta}{2} \mathbb{1}+\eta \Pi_{ \pm}^{k} \tag{1}
\end{equation*}
$$

Here $\mathbb{1}$ is the $2 \times 2$ identity matrix, $\vec{\sigma}$ is the vector of Pauli matrices $\left(\sigma_{x}, \sigma_{y}, \sigma_{z}\right), \hat{n}_{k}$ is the axis for measurement $k$, and $\eta \in[0,1]$ is the sharpness associated with each observable. For $\eta=1$, these reduce to projective measurements, $\mathcal{P}_{k}=\left\{\Pi_{+}^{k}, \Pi_{-}^{k}\right\}$. In our experiment, we choose a special case of trine spin axes

$$
\begin{equation*}
\hat{n}_{1}=(0,0,1), \hat{n}_{2}=\left(\frac{\sqrt{3}}{2}, 0,-\frac{1}{2}\right), \hat{n}_{3}=\left(-\frac{\sqrt{3}}{2}, 0,-\frac{1}{2}\right) \tag{2}
\end{equation*}
$$

equally spaced in the $z-x$ plane.
Testing the LSW inequality for a quantum mechanical violation requires a special kind of joint measurability, denoted by joint measurability contexts $\left\{\left\{\mathcal{M}_{1}, \mathcal{M}_{2}\right\},\left\{\mathcal{M}_{2}, \mathcal{M}_{3}\right\},\left\{\mathcal{M}_{1}, \mathcal{M}_{3}\right\}\right\}$. That is, the three observables $\mathcal{M}_{k}(k=1,2,3)$ are pairwise jointly measurable, for all three pairs, but not triply jointly measurable. Pairwise joint measurability is possible only if $\eta \leq(\sqrt{3}-1) \approx 0.732$ 23]. Triple-wise joint measurability - which would eliminate any possibility of contextuality since the entire experiment could be performed using a single context $\left\{\mathcal{M}_{1}, \mathcal{M}_{2}, \mathcal{M}_{3}\right\}$ - is possibly only if $\eta<2 / 3$ [23]. Here we restrict our consideration of $\eta$ to the narrow range $2 / 3<\eta \leq(\sqrt{3}-1)$.

The joint measurability context $\left\{\mathcal{M}_{i}, \mathcal{M}_{j}\right\}$ means that there exists a POVM $\mathcal{J}_{i j} \equiv\left\{G_{++}^{i j}, G_{+-}^{i j}, G_{-+}^{i j}, G_{--}^{i j}\right\}$ satisfying the marginal condition that $\sum_{\varepsilon} G_{\epsilon \varepsilon}^{i j}=E_{\epsilon}^{i}$, and $\sum_{\epsilon} G_{\epsilon \varepsilon}^{i j}=E_{\varepsilon}^{j}$, where $\epsilon, \varepsilon \in\{+1,-1\}$. We follow Ref. [24]
in using joint POVMs with the following general form:

$$
\begin{align*}
& G_{++}^{i j}=\frac{1}{2}\left\{\frac{\alpha_{i j}}{2} \mathbb{1}+\vec{\sigma} \cdot \frac{1}{2}\left[\eta\left(\hat{n}_{i}+\hat{n}_{j}\right)-\vec{a}_{i j}\right]\right\} \\
& G_{+-}^{i j}=\frac{1}{2}\left\{\left(1-\frac{\alpha_{i j}}{2}\right) \mathbb{1}+\vec{\sigma} \cdot \frac{1}{2}\left[\eta\left(\hat{n}_{i}-\hat{n}_{j}\right)+\vec{a}_{i j}\right]\right\} \\
& G_{-+}^{i j}=\frac{1}{2}\left\{\left(1-\frac{\alpha_{i j}}{2}\right) \mathbb{1}+\vec{\sigma} \cdot \frac{1}{2}\left[\eta\left(-\hat{n}_{i}+\hat{n}_{j}\right)+\vec{a}_{i j}\right]\right\} \\
& G_{--}^{i j}=\frac{1}{2}\left\{\frac{\alpha_{i j}}{2} \mathbb{1}+\vec{\sigma} \cdot \frac{1}{2}\left[\eta\left(-\hat{n}_{i}-\hat{n}_{j}\right)-\vec{a}_{i j}\right]\right\}, \tag{3}
\end{align*}
$$

where $\alpha_{i j} \in \mathbb{R}$ and $\vec{a}_{i j} \in \mathbb{R}^{3}$, and the relation $G_{\epsilon \varepsilon}^{i j}=G_{\epsilon \varepsilon}^{j i}$ with $\epsilon, \varepsilon \in\{+1,-1\}$ is satisfied.

The LSW inequality is the following [23]

$$
\begin{equation*}
R_{3} \equiv \frac{1}{3} \sum_{(i j) \in\{(12),(23),(13)\}} \operatorname{Pr}\left(X_{i} \neq X_{j} \mid \mathcal{J}_{i j}\right) \leq 1-\frac{\eta}{3} \tag{4}
\end{equation*}
$$

where $\operatorname{Pr}\left(X_{i} \neq X_{j} \mid \mathcal{J}_{i j}\right)$ denotes the probability of obtaining anticorrelated outcomes in a joint measurement denoted $\mathcal{J}_{i j}$. Note that by the (unreasonable) assumption of outcome determinism for POVMs in Refs. 18 21], the bound on the right-hand-side would be $2 / 3$ [23], whereas the LSW bound is at least 0.756 (since we require $\eta<0.732$ for pairwise joint measurability).

In quantum theory, where $\mathcal{J}_{i j}$ is described by a joint POVM as defined above, the average anticorrelation probability $R_{3}$ takes the form [24]

$$
\begin{equation*}
R_{3}^{Q}=\frac{1}{3} \sum_{(i j) \in\{(12),(23),(13)\}} \operatorname{Tr}\left[\left(G_{+-}^{i j}+G_{-+}^{i j}\right)\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right], \tag{5}
\end{equation*}
$$

where $\left|\phi_{0}\right\rangle$ is the qubit state being measured. It follows that a necessary condition for statedependent violation of the LSW inequality is $\operatorname{Tr}\left[\sum_{i j}\left(\alpha_{i j} \mathbb{1}-\vec{\sigma} \cdot \vec{a}_{i j}\right)\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right]<2 \eta$. It has been shown [24] that the largest violation of the LSW inequality for observables defined by Eq. (21) can be obtained by the state $\left|\phi_{0}\right\rangle=(|0\rangle+i|1\rangle) / \sqrt{2}$, and joint POVM $\mathcal{J}_{i j}$ in Eq. (3) defined by $\alpha_{i j}=1+\eta^{2} \hat{n}_{i} \cdot \hat{n}_{j}$ and a vector $\vec{a}_{i j}$ satisfying $\vec{a}_{i j}=\left(0, \sqrt{1+\eta^{4}\left(\hat{n}_{i} \cdot \hat{n}_{j}\right)^{2}-2 \eta^{2}}, 0\right)$. Moreover, the optimal violation for $\eta$ in the range $[2 / 3,0.732]$ is as $\eta \rightarrow 2 / 3$, so that $\alpha_{i j} \rightarrow 7 / 9$ and $\left|\vec{a}_{i j}\right| \rightarrow \sqrt{13} / 9$ for any $(i j) \in\{(12),(23),(13)\}$. Then the quantum average probability of anticorrelation is $R_{3}^{Q} \rightarrow 0.8114$ and exceeds the LSW noncontextual bound of $7 / 9 \approx 0.7778$.

In our experiment, we aim for $\eta=0.670$, strictly within the range $[2 / 3,0.732]$ but close to the optimum at $2 / 3$.

## B. Implementation of joint POVMs

For the pairwise joint measurements described above, each element of the POVM is rank one, and can be rewritten as $G_{\epsilon \varepsilon}^{i j}=\lambda_{\epsilon \varepsilon}\left|\xi_{\epsilon \varepsilon}^{i j}\right\rangle\left\langle\xi_{\epsilon \varepsilon}^{i j}\right|$ with $\epsilon, \varepsilon \in\{+1,-1\}$. Here,


FIG. 1: Experimental setup. Single photons are created via type-I spontaneous parametric down-conversion (SPDC) in a 0.5 mm -thick nonlinear- $\beta$-barium-borate (BBO) crystal which is pumped by a CW diode laser with 80 mW of power. One photon in the pair is detected to herald the other photon, which is injected into the optical network. To realize the joint POVM $\mathcal{J}_{i j}$, the single-qubit rotations $U_{\epsilon \varepsilon}$ are realized by sandwich-type QWP-HWP-QWP sets with different angles placed in different optical modes. The first two projectors are realized by PPBSs with different transmission probabilities for vertical polarization. The last projector is realized by a PBS. Detecting heralded single photons means in practice registering coincidences between single photon detectors: D0 and each of D1-D4.
$\lambda_{++}=\lambda_{--}=\left(2-\eta^{2}\right) / 4$ and $\lambda_{+-}=\lambda_{-+}=\left(2+\eta^{2}\right) / 4$. We propose a scheme for implementing the joint POVMs in three stages, each of which is a single-qubit rotation followed by a two-outcome measurement. In each, the positive result (i.e. a detector click) has a POVM element proportional to the appropriate projector, while the null result qubit is fed into the next stage. The null result qubit from the third stage is then also detected.

To be more specific, the three single-qubit rotations are designed as

$$
\begin{equation*}
U_{\epsilon \varepsilon}=|0\rangle\left\langle\phi_{\epsilon \varepsilon}^{\prime}\right|+|1\rangle\left\langle\phi_{\epsilon \varepsilon}^{\prime \perp}\right|, \epsilon \varepsilon \in\{+-,-+,++\} \tag{6}
\end{equation*}
$$

while the POVM elements $\left\{P_{\epsilon \varepsilon}^{0}, P_{\epsilon \varepsilon}^{1}\right\}$ take the form

$$
\begin{equation*}
P_{\epsilon \varepsilon}^{0}=|0\rangle\langle 0|+\left(1-\chi_{\epsilon \varepsilon}\right)|1\rangle\langle 1|, P_{\epsilon \varepsilon}^{1}=\chi_{\epsilon \varepsilon}|1\rangle\langle 1| \tag{7}
\end{equation*}
$$

In the above, $\left|\phi_{\epsilon \varepsilon}^{\prime}\right\rangle$ and $\chi_{\epsilon \varepsilon}$ are chosen such that after the projector is applied, the probability of the click of $P_{\epsilon \varepsilon}^{1}$ is $\lambda_{\epsilon \varepsilon} \operatorname{Tr}\left(\left|\xi_{\epsilon \varepsilon}^{i j}\right\rangle\left\langle\xi_{\epsilon \varepsilon}^{i j} \mid \phi_{0}\right\rangle\left\langle\phi_{0}\right|\right)=\operatorname{Tr}\left(G_{\epsilon \varepsilon}^{i j}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right)$. If the other (null) result is obtained, the qubit enters the next stage of the apparatus, having had the operator $\sqrt{P_{\epsilon \varepsilon}^{0}}$ applied to its state. For example, we implement $G_{+-}^{i j}$ at the first stage by choosing $\left|\phi_{+-}^{\prime}\right\rangle=\left|\xi_{+-}^{i j \perp}\right\rangle$ and $\chi_{+-}=\lambda_{+-}$. The first detector clicks with probability $\operatorname{Tr}\left(G_{+-}^{i j}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right)$, and if it does not click then the qubit state entering the next stage of apparatus is $\sqrt{P_{+-}^{0}} U_{+-}\left|\phi_{0}\right\rangle=\left\langle\xi_{+-}^{i j \perp} \mid \phi_{0}\right\rangle|0\rangle+\sqrt{1-\chi_{+-}}\left\langle\xi_{+-}^{i j} \mid \phi_{0}\right\rangle|1\rangle$. We design the apparatus so as to next measure $G_{-+}^{i j}$, then $G_{++}^{i j}$, in the same way. Since $\sum_{\epsilon \varepsilon} G_{\epsilon \varepsilon}^{i j}=\mathbb{1}$, the fourth possible click, following the null outcome at the third state, corresponds to the implementation of $G_{-}^{i j}$. For further details see the Supplemental Material.

## III. EXPERIMENTAL REALIZATION

## A. Experimental violation of the LSW inequality

We perform the test of the LSW inequality with single photons. The basis states of the qubit, $|0\rangle$ and $|1\rangle$, are encoded by the polarizations of single photons, $|H\rangle$ and $|V\rangle$. We generate contextual quantum correlations by performing the four-outcome joint POVM on this qubit.

The experimental setup shown in Fig. 1 involves preparing the specific state (preparation stage) and then performing the joint POVM (measurement stage). In the preparation stage, polarization-degenerate trigger-herald photon pairs are produced and are registered by a coincidence count at two single-photon avalanche photodiodes (APDs) with 7 ns time window. Total coincidence counts are about $10^{5}$ over a collection time of 60 s , and the probability of randomly creating more than one simultaneous photon pair is thus of order $10^{-4}$, which is negligible. The second-order correlation $g^{(2)}$ is measured as $0.0089 \pm 0.0018$ which shows that the single-photon source is extremely non-classical [26]. The heralded single photons are prepared in state $\left|\phi_{0}\right\rangle=(|H\rangle+i|V\rangle) / \sqrt{2}$ after passing through a polarizing beam splitter (PBS), a half-wave plate (HWP, H0), and a quarter-wave plate (QWP, Q0).

In the measurement stage, to implement the twooutcome measurements, partially projecting polarizing elements are added to the setup and allow us to produce the required projectors with the appropriate weights. We employ partially polarizing beam splitters (PPBSs) with specific transmission probabilities for vertical polarization $T_{V}$ and same transmission probability for horizontal

TABLE I: Experimental results. The 'element' columns contain the elements of joint POVM corresponding to heralded singleclick events. The 'probability' columns contain the measured probabilities of joint POVM. The 'condition' column contains assigned measurement values. The 'value' column contains the measured probabilities of the anticorrelations. Rows 1-3 correspond to the results of joint POVMs on the state $\left|\phi_{0}\right\rangle$. Row 4 presents the measured $\tilde{R}_{3}^{Q}$. Error bar indicates the statistical uncertainty which is obtained based on assuming Poissonian statistics.

| D 1 |  | D 2 |  | D 3 | D 4 |  | Calculated contribution |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| condition |  |  |  |  |  |  |  |  |

polarization $T_{H}=1$. This allows us to project the state onto $|V\rangle$ on the reflected port of the PPBSs.

After passing through each QWP-HWP-QWP set and the following PPBS, the photons are detected by APDs on the reflected port, in coincidence with the trigger photons. The transmitted photons go into the next QWP-HWP-QWP set and PPBS (or, at the final stage, a PBS which can be regarded as a special PPBS with equal transmission and reflection probabilities). The relative detection efficiencies of the detectors D2-D4 with respect to D1, are measured as $0.9499 \pm 0.0070,0.9199 \pm 0.0069$ and $0.9801 \pm 0.0063$ respectively and these figures are used to correct the coincidence counts (see the Supplemental Material). The probability of measuring the photons is obtained by normalizing the corrected coincidence counts on each mode with respect to the total corrected coincidence counts. The overall detection efficiency of the heralded photons in our experiment is approximately $11 \%$. Thus we make the fair-sampling assumption: that the event selected out by the photonic coincidence is an unbiased representation of the whole sample.

The probabilities of photons being measured on the reflected ports (clicks on the detectors D1-D3) correspond to those of the joint POVM elements $G_{+-}^{i j}, G_{-+}^{i j}$, and $G_{++}^{i j}$, whereas the probability of photons being measured on the transmitted port of the PBS (click on the detector D4) corresponds to that of the element $G_{-\ldots}^{i j}$. We can estimate the matrix forms of the joint POVM elements from the measured probabilities (see Subsec. IIIB). The negligible difference from the theoretical prediction guarantees successful experimental realization of joint POVMs by taking into account of all the imperfections of the experimental setup.

In Table I, we present the measured probabilities and the outcomes of the joint POVM with noise parameter $\eta=0.67$ on the specific state $\left|\phi_{0}\right\rangle$. The result of measured average probability of anticorrelations is $\tilde{R}_{3}^{Q}=0.8125 \pm 0.0010$. Here, and below, the tilde relates to the experimentally implemented POVMs, as opposed to the theoretical ones aimed for; see Subsec. IIIB. This $\tilde{R}_{3}^{Q}$ violates the bound set by the noncontextual hidden variable theory $1-\eta / 3=0.7767$ by 35 standard deviations. Furthermore, in our experiment the noise parameter can be estimated by the experimental data (see

Subsec. IIIB). The average value of the estimated noise parameters in the experiment is $\bar{\eta}=0.6690 \pm 0.0019$. Using this value, rather than the aimed-for 0.670 , makes almost no difference in the LSW bound: the bound set by the noncontextual hidden variable theory can be calculated as $1-\overline{\tilde{\eta}} / 3=0.7770 \pm 0.0006$ compared to $1-0.670 / 3=0.7767$. Even including the uncertainty in the former bound, the experimentally measured average probability of anticorrelation, $\tilde{R}_{3}^{Q}=0.8125 \pm 0.0010$, still implies a violation of this experimental bound of the LSW inequality $1-\bar{\eta} / 3$ by 22 standard deviations. In Subsec. IIIB we give an alternate way of comparing the correlations and the bound, which also gives a violation by many standard deviations. Here, we finish by noting that the experimental value $\tilde{R}_{3}^{Q}$ is in agreement (1.6 standard deviations) with its theoretical prediction $0.8087 \pm 0.0022$, predicted via the estimated noise parameter $\overline{\tilde{\eta}}$.

## B. Evaluating the quality of experimental realization of POVM

We consider the effect on the implementation of the joint POVM due to all the important imperfections, namely in the PPBSs $\left(T_{V}^{1}=0.3904 \pm 0.0045, T_{V}^{2}=\right.$ $0.2897 \pm 0.0050$ ), WPs (typical retardance accuracy< 2.67 nm ), PBSs (typical extinction ratio $\sim 10^{5}: 1$ ), and detectors. We define a modified 2-norm distance $D(A, B)$ between the matrix form of the theoretical prediction of POVM element $A$ and that of experimental implementation of the corresponding POVM element $B$ as

$$
\begin{equation*}
D(A, B)=\sqrt{\frac{\operatorname{Tr}\left[(A-B)^{2}\right]}{\operatorname{Tr}\left(A^{2}\right)}} \tag{8}
\end{equation*}
$$

For the particular forms of the POVM described in our paper, the distance ranges between 0 for a perfect match and $\sqrt{2}$ for a complete mismatch. For example, we use the distance $D\left(G_{\epsilon \varepsilon}^{i j}, \tilde{G}_{\epsilon \varepsilon}^{i j}\right)$ to measure the mismatch between the theoretical prediction of $G_{\epsilon \varepsilon}^{i j}$ with $i \neq j \in$ $\{1,2,3\}, \epsilon, \varepsilon \in\{+1,-1\}$, and the corresponding experimental implementation $\tilde{G}_{\epsilon \varepsilon}^{i j}$.

To obtain the distance, we perform measurement tomography [27, 28]. Single photons, prepared in the
states $|H\rangle,|V\rangle,|R\rangle=(|H\rangle+i|V\rangle) / \sqrt{2}$ and $|D\rangle=$ $(|H\rangle+|V\rangle) / \sqrt{2}$, are passed through the optical circuit and are detected by APDs in coincidence with the trigger photons. After correcting for the relative efficiencies of the different detectors, the photon counts give the measured probabilities. From these we can obtain the matrix forms of all twelve elements of the joint POVMs $\tilde{G}$ via maximum-likelihood estimation.

In our experiment the accuracy of the experimental implementation of the measurements described by the POVM $\tilde{E}_{\epsilon}^{i(j)}$-the noisy version of the projective measurements-are more important. Here $\tilde{E}_{\epsilon}^{i(j)} \approx$ $\sum_{\varepsilon} G_{\epsilon \varepsilon}^{i j}$, so the $(j)$ superscript indicates any dependence on context in which it is performed (see below). The element $\tilde{E}_{\epsilon}^{i(j)}$ is estimated by minimizing the 2-norm distance $D\left(\tilde{E}_{\epsilon}^{i(j)}, \sum_{\varepsilon} \tilde{G}_{\epsilon \varepsilon}^{i j}\right)$ which is defined in Eq. (8). This minimization is subject to the constraints that the sum $\tilde{E}_{+}^{i(j)}+\tilde{E}_{-}^{i(j)}$ equals the identity operator, whilst the difference is traceless, as per Eq. (1). The minimum distances found by this procedure are very small (all less than $6.2 \times 10^{-5}$ ), which justifies our approach. The theoretical prediction of the element satisfies the marginal condition $E_{\epsilon}^{i(j)}=\sum_{\varepsilon} G_{\epsilon \varepsilon}^{i j}=E_{\epsilon}^{i(k)}=\sum_{\varepsilon} G_{\epsilon \varepsilon}^{i k}(i \neq j \neq$ $k \in\{1,2,3\})$. However due to the imperfections in the experiment, there might be a slight difference between $\sum_{\varepsilon} \tilde{G}_{\epsilon \varepsilon}^{i j}$ and $\sum_{\varepsilon} \tilde{G}_{\epsilon \varepsilon}^{i k}$. Hence we use two superscripts $i$ and $j$ to represent the estimated element of POVM $\tilde{E}_{\epsilon}^{i(j)}$ which is estimated by $\tilde{G}_{\epsilon \varepsilon}^{i j}$ and corresponds to the joint measurable context $\left\{\mathcal{M}_{i}, \mathcal{M}_{j}\right\}$. The difference between the elements $\tilde{E}_{\epsilon}^{i(j)}$ and $\tilde{E}_{\epsilon}^{i(k)}$ can also be measured by the 2-norm distance $D\left(\tilde{E}_{\epsilon}^{i(j)}, \tilde{E}_{\epsilon}^{i(k)}\right)$ defined in Eq. (8). Thus the distances satisfy $D\left(\tilde{E}_{+}^{i(j)}, \tilde{E}_{+}^{i(k)}\right)=D\left(\tilde{E}_{-}^{i(j)}, \tilde{E}_{-}^{i(k)}\right)$. As shown in Fig. 2, all the distances $D\left(\tilde{E}^{i(j)}, \tilde{E}^{i(k)}\right)$ are smaller than 0.0006 , which validates the experimental realizations of pairwise jointly measurable POVMs.

We also compare the estimated element $\tilde{E}_{\epsilon}^{i(j)}$ with the theoretical ideal $E_{\epsilon}^{i}$ by calculating the 2-norm distance $D\left(E_{\epsilon}^{i}, \tilde{E}_{\epsilon}^{i(j)}\right)$. Since the distances satisfy $D\left(E_{+}^{i}, \tilde{E}_{+}^{i(j)}\right)=$ $D\left(E_{-}^{i}, \tilde{E}_{-}^{i(j)}\right)$ we only show six values of the distances $D\left(E^{i}, \tilde{E}^{i(j)}\right)$ in Fig. 3 All the distances are smaller than 0.0007 , which shows the successful experimental realizations of the POVMs with the chosen noise parameter $\eta=0.67$.

For determining the LSW bound used in Subsec. III A it is important to know the noise parameter of $\tilde{\eta}$ associated with the POVM. This can be estimated as

$$
\begin{equation*}
\tilde{\eta}_{\epsilon}^{i(j)}=\sqrt{2 \operatorname{Tr}\left[\tilde{E}_{\epsilon}^{i(j)}\left(\tilde{E}_{\epsilon}^{i(j)}\right)^{\dagger}\right]-1} \tag{9}
\end{equation*}
$$

The condition $\tilde{E}_{+}^{i(j)}+\tilde{E}_{-}^{i(j)}=\mathbb{1}$ guarantees that we have the same values of $\tilde{\eta}_{+}^{i(j)}$ and $\tilde{\eta}_{-}^{i(j)}$. Compared to the value of the noise parameter we aimed for in the experiment $\eta=0.67$, all the differences $\left|\tilde{\eta}^{i(j)}-\eta\right|$ are smaller than 0.0015 . The average value of the estimated


FIG. 2: The distance $D\left(\tilde{E}^{i(j)}, \tilde{E}^{i(k)}\right)$ between the estimated POVM elements for different contexts $\left\{\mathcal{M}_{i}, \mathcal{M}_{j}\right\}$ and $\left\{\mathcal{M}_{i}, \mathcal{M}_{k}\right\}$. Error bars indicate the statistical uncertainty, obtained from Monte-Carlo simulations assuming Poissonian photon-counting statistics.
noise parameters in the experiment is $\overline{\tilde{\eta}}=\frac{1}{6} \sum_{i(j)} \tilde{\eta}^{i(j)}=$ $0.6690 \pm 0.0019$.

Finally, the value $R_{3}^{Q}$ corresponding to the ideal POVMs can also be bounded, as follows. An arbitrary qubit POVM element $G$ can be written as $G=a \mathbb{1}+b \hat{n} \cdot \vec{\sigma}$, where $\hat{n}$ is a unit vector and $a$ and $b$ are nonnegative numbers satisfying $b \leq a$ and $b \leq 1-a$. An arbitrary qubit density operator can be written as $\rho=\frac{1}{2}(\mathbb{1}+\vec{r} \cdot \vec{\sigma})$, where $|\vec{r}| \leq 1$. The probability of obtaining the outcome corresponding to $G$ for a POVM containing $G$ on state $\rho$ is given by $\operatorname{Pr}(G)=\operatorname{Tr}[G \rho]=\frac{1}{2}(a+b \vec{r} \cdot \hat{n})$. Let $\vec{g}=\left(a, b n_{x}, b n_{y}, b n_{z}\right)$ and $\vec{s}=\frac{1}{2}\left(1, r_{x}, r_{y}, r_{z}\right)$. Then $\operatorname{Pr}(G)=\vec{g} \cdot \vec{s}$. Let $\tilde{G}$ denote the experimentally realized POVM element corresponding to $G$, and likewise $\overrightarrow{\tilde{g}}$ to $\vec{g}$. Then $|\operatorname{Pr}(G)-\operatorname{Pr}(\tilde{G})|=|(\vec{g}-\overrightarrow{\tilde{g}}) \cdot \vec{s}| \leq|\vec{g}-\overrightarrow{\tilde{g}}||\vec{s}| \leq$ $\frac{1}{\sqrt{2}}|\vec{g}-\overrightarrow{\tilde{g}}|$. Thus we obtain the bound

$$
\begin{equation*}
\operatorname{Pr}(G) \geq \operatorname{Pr}(\tilde{G})-\frac{1}{\sqrt{2}}|\vec{g}-\overrightarrow{\tilde{g}}| \tag{10}
\end{equation*}
$$

Let $\tilde{g}_{+-}^{i j}\left(\tilde{g}_{-+}^{i j}\right)$ be the vector representation of $\tilde{G}_{+-}^{i j}\left(\tilde{G}_{-+}^{i j}\right)$ in our experiment, obtained above by tomography. Then from (10) we obtain a lower bound for the ideal value

$$
\begin{align*}
& R_{3}^{Q} \geq \tilde{R}_{3}^{Q} \\
&-\frac{1}{3 \sqrt{2}} \sum_{(i j) \in(12),(23),(13)}\left(\left|g_{+-}^{i j}-\tilde{g}_{+-}^{i j}\right|+\left|g_{-+}^{i j}-\tilde{g}_{-+}^{i j}\right|\right) \tag{11}
\end{align*}
$$

We estimate the bound for the ideal value $R_{3}^{Q}$ based on the measured value $\tilde{R}_{3}^{Q}$ and the estimated $\tilde{G}_{\epsilon \varepsilon}^{i j}$. We find

$$
\begin{equation*}
R_{3}^{Q} \geq 0.7964 \pm 0.0012 \tag{12}
\end{equation*}
$$

The uncertainty here is larger than that in $\tilde{R}_{3}^{Q}$ because of uncertainties in the $\tilde{g} s$ that contribute to the correction term in Eq. (11). Now, the appropriate point of comparison is the ideal noncontextual bound of 0.7767 , from the aimed-for $\eta=0.67$, because we are inferring the correlations from an ideal measurement with this $\eta$. The value of the bound in Eq. (12) implies a violation of this ideal bound by at least 16 standard deviations.

Note that as we assume the validity of quantum mechanics, there is no need to establish operational equivalences between the measured POVM elements in different contexts, as done in Ref. [25].


FIG. 3: The distance $D\left(E^{i}, \tilde{E}^{i(j)}\right)$ between the aimed-for POVM element $E^{i}$ for context $\left\{\mathcal{M}_{i}, \mathcal{M}_{j}\right\}$ and the estimated POVM elements from the experiment $\tilde{E}^{i(j)}$. Error bars indicate the statistical uncertainty, obtained from Monte-Carlo simulations assuming Poissonian photon-counting statistics.

## IV. DISCUSSION

Any realistic measurement necessarily has some nonvanishing amount of noise and therefore never achieves the ideal of sharpness. This provides a compelling reason to test contextuality applicable to unsharp measurements. Here we test the generalized noncontextuality inequality for the unsharp measurements of LSW [23]. For unsharp measurements that can be jointly performed, correlated noise could allow correlations to be generated by a non-
contextual hidden variable model. The LSW inequality takes such correlations into account by setting a higher bound. Thus a violation of the LSW inequality certifies nonclassicality that cannot be attributed to hidden variables associated with noise in the unsharp measurements.

Our experimental results show convincing violation of the LSW inequality with single-photon qubits. That is, it is a demonstration of contextuality for the simplest type of quantum system. It is also the first experiment to apply the LSW argument to rule out noncontextuality within quantum theory.

The experimental confirmation of quantum contextuality in its simple and fundamental form sheds new light on the contradiction between quantum mechanics and noncontextual realistic models. Furthermore, we realize joint POVMs of noisy spin- $\frac{1}{2}$ observables on a single-qubit system which is the key point to implement the unsharp measurements, paving the way for further developments such as the real time estimation [29], monitoring of the Rabi oscillations of a single qubit in a driving field 30] and understanding the relation between information gain and disturbance 31].

## Funding Information

NSFC (Nos. 11474049, 11674056 and 11674306), NSFJS (No. BK20160024), Scientific Research Foundation of Graduate School of Southeast University (No. YBJJ1623), Australian Research Council Centre of Excellence CE110001027 and grant (FQXi-RFP-610 1504) from the Foundational Questions Institute Fund (fqxi.org) at the Silicon Valley Community Foundation.

## Acknowledgments

We would like to thank R. Kunjwal for stimulating discussions, and three anonymous reviewers for helpful suggestions.

## Supplemental Documents

See Supplement 1 for supporting content.
[1] J. S. Bell, Physics 1, 195 (1964).
[2] S. Kochen and E. Specker, J. Math. Mech. 17, 59 (1967).
[3] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. 81, 5039 (1998).
[4] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.C. Guo, Phys. Rev. Lett. 90, 250401 (2003).
[5] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, Nature 460, 494 (2009).
[6] E. Amselem, M. Rådmark, M. Bourennane, and A. Cabello, Phys. Rev. Lett. 103, 160405 (2009).
[7] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa, Phys. Rev. Lett. 103, 040403 (2009).
[8] O. Moussa, C. A. Ryan, D. G. Cory, and R. Laflamme, Phys. Rev. Lett. 104, 160501 (2010).
[9] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Nature 474,

490 (2011).
[10] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. 101, 020403 (2008).
[11] A. Peres, J. Phys. A: Math. Gen. 24, L175 (1991).
[12] J. Bub, Found. Phys. 26, 787 (1996).
[13] P. Badziag, I. Bengtsson, A. Cabello, and I. Pitowsky, Phys. Rev. Lett. 103, 050401 (2009).
[14] S. Yu and C. Oh, Phys. Rev. Lett. 108, 030402 (2012).
[15] A. Cabello, E. Amselem, K. Blanchfield, M. Bourennane, and I. Bengtsson, Phys. Rev. A 85, 032108 (2012).
[16] C. Zu, Y.-X. Wang, D.-L. Deng, X.-Y. Chang, K. Liu, P.Y. Hou, H.-X. Yang, and L.-M. Duan, Phys. Rev. Lett. 109, 150401 (2012).
[17] X. Zhang, M. Um, J. Zhang, S. An, Y. Wang, D.-l. Deng, C. Shen, L.-M. Duan, and K. Kim, Phys. Rev. Lett. 110, 070401 (2013).
[18] P. Busch, Phys. Rev. Lett. 91, 120403 (2003).
[19] A. Cabello, Phys. Rev. Lett. 90, 190401 (2003).
[20] P. Aravind, Phys. Rev. A 68, 052104 (2003).
[21] A. A. Méthot, Int. J. Quantum Inf. 5, 353 (2007).
[22] R. W. Spekkens, Found. Phys. 44, 1125 (2014).
[23] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. 506, 1 (2011), see also erratum http://dx.doi.org/10.1016/j.physrep.2016.12.001.
[24] R. Kunjwal and S. Ghosh, Phys. Rev. A 89, 042118 (2014).
[25] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, Nature Communications 7, 11780 (2016), URL http://dx.doi.org/10.1038/ncomms11780.
[26] J. Thorn, M. Neel, V. Donato, G. Bergreen, R. Davies, and M. Beck, Am. J. Phys. 72, 1210 (2004).
[27] J. Fiurášek, Phys. Rev. A 64, 024102 (2001).
[28] G. M. D'Ariano, M. G. Paris, and M. F. Sacchi, Adv. Imaging Electron Phys. 128, 206 (2003).
[29] B. A. Chase and J. Geremia, Phys. Rev. A 79, 022314 (2009).
[30] T. Konrad and H. Uys, Phys. Rev. A 85, 012102 (2012).
[31] H. M. Wiseman, Nature 470, 178 (2011).

## Experimental generalized contextuality with single-photon qubits: supplementary material

In the supplementary document, we provide the details of the experiment and data analysis. The raw probabilities to demonstrate the joint measurability of positive operator-valued measures (POVMs) are also provided.

## I. IMPLEMENTATION OF THE ELEMENTS OF THE JOINT POVM $G$

Each element of the POVM can be written as $G_{\epsilon \varepsilon}^{i j}=\lambda_{\epsilon \varepsilon}\left|\xi_{\epsilon \varepsilon}^{i j}\right\rangle\left\langle\xi_{\epsilon \varepsilon}^{i j}\right|$, with $\epsilon, \varepsilon \in\{+1,-1\}, i \neq j \in\{1,2,3\}$,

$$
\begin{equation*}
\lambda_{++}=\lambda_{--}=\frac{\left(2-\eta^{2}\right)}{4}, \lambda_{+-}=\lambda_{-+}=\frac{\left(2+\eta^{2}\right)}{4} \tag{S1}
\end{equation*}
$$

and

$$
\begin{aligned}
& \left|\xi_{+-}^{12}\right\rangle=\binom{\sqrt{3} \eta+i \sqrt{4-8 \eta^{2}+\eta^{4}}}{-2+3 \eta-\eta^{2}} / \sqrt{8-12 \eta+8 \eta^{2}-6 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{-+}^{12}\right\rangle=\binom{\sqrt{3} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2+3 \eta+\eta^{2}} / \sqrt{8+12 \eta+8 \eta^{2}+6 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{++}^{12}\right\rangle=\binom{\sqrt{3} \eta+i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2-\eta-\eta^{2}} / \sqrt{8-4 \eta-8 \eta^{2}+2 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{--}^{12}\right\rangle=\binom{\sqrt{3} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{-2-\eta+\eta^{2}} / \sqrt{8+4 \eta-8 \eta^{2}-2 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{+-}^{13}\right\rangle=\binom{\sqrt{3} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2-3 \eta+\eta^{2}} / \sqrt{8-12 \eta+8 \eta^{2}-6 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{-+}^{13}\right\rangle=\binom{-\sqrt{3} \eta+i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2+3 \eta+\eta^{2}} / \sqrt{8+12 \eta+8 \eta^{2}+6 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{++}^{13}\right\rangle=\binom{\sqrt{3} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{-2+\eta+\eta^{2}} / \sqrt{8-4 \eta-8 \eta^{2}+2 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{--}^{13}\right\rangle=\binom{-\sqrt{3} \eta+i \sqrt{4-8 \eta^{2}+\eta^{4}}}{-2-\eta+\eta^{2}} / \sqrt{8+4 \eta-8 \eta^{2}-2 \eta^{3}+2 \eta^{4}}, \\
& \left|\xi_{+-}^{23}\right\rangle=\binom{\sqrt{12} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2+\eta^{2}} / \sqrt{2}\left(2+\eta^{2}\right), \\
& \left|\xi_{-+}^{23}\right\rangle=\binom{-\sqrt{12} \eta-i \sqrt{4-8 \eta^{2}+\eta^{4}}}{2+\eta^{2}} / \sqrt{2}\left(2+\eta^{2}\right), \\
& \left|\xi_{++}^{23}\right\rangle=\binom{-i \sqrt{4-4 \eta-4 \eta^{2}+2 \eta^{3}+\eta^{4}}}{-\sqrt{4+4 \eta-4 \eta^{2}-2 \eta^{3}+\eta^{4}}} / \sqrt{2}\left(-2+\eta^{2}\right),
\end{aligned}
$$

$$
\left|\xi_{--}^{23}\right\rangle=\binom{-i \sqrt{4+4 \eta-4 \eta^{2}-2 \eta^{3}+\eta^{4}}}{-\sqrt{4-4 \eta-4 \eta^{2}+2 \eta^{3}+\eta^{4}}} / \sqrt{2}\left(-2+\eta^{2}\right)
$$

Each element can be implemented by a single-qubit rotation followed by a two-outcome measurement. In the first step, to realize the element $G_{+-}^{i j}$ the single-qubit rotation is designed as

$$
\begin{equation*}
U_{+-}=|0\rangle\left\langle\phi_{+-}^{\prime}\right|+|1\rangle\left\langle\phi_{+-}^{\prime \perp}\right|,\left|\phi_{+-}^{\prime}\right\rangle=\left|\xi_{+-}^{i j \perp}\right\rangle . \tag{S2}
\end{equation*}
$$

The POVM element $\left\{P_{+-}^{0}, P_{+-}^{1}\right\}$ takes the form

$$
\begin{equation*}
P_{+-}^{0}=|0\rangle\langle 0|+\left(1-\chi_{+-}\right)|1\rangle\langle 1|, P_{+-}^{1}=\chi_{+-}|1\rangle\langle 1|, \chi_{+-}=\lambda_{+-} . \tag{S3}
\end{equation*}
$$

The choice of $\left|\phi_{+-}^{\prime}\right\rangle$ guarantees if $P_{+-}^{1}$ clicks the initial state is projected onto the eigenstate $\left|\xi_{+-}^{i j}\right\rangle$ and let the component of the state $\left|\xi_{+-}^{i j \perp}\right\rangle$ which is orthogonal to $\left|\xi_{+-}^{i j}\right\rangle$ all pass through for the next measurement. Therefore the state after the first rotation is $U_{+-}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right| U_{+-}^{\dagger}$ for any input $\left|\phi_{0}\right\rangle$. The first detector $\left(P_{+-}^{1}\right)$ clicks with the probability $\operatorname{Tr}\left(P_{+-}^{1} U_{+-}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right| U_{+-}^{\dagger}\right)=\lambda_{+-}\left\langle\xi_{+-}^{i j} \mid \phi_{0}\right\rangle\left\langle\phi_{0} \mid \xi_{+-}^{i j}\right\rangle=\operatorname{Tr}\left(G_{+-}^{i j}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right)$. Thus we implement the element of POVM $G_{+-}^{i j}$. The other state without click is

$$
\begin{equation*}
\sqrt{P_{+-}^{0}} U_{+-}\left|\phi_{0}\right\rangle=\left\langle\xi_{+-}^{i j \perp} \mid \phi_{0}\right\rangle|0\rangle+\sqrt{1-\chi_{+-}}\left\langle\xi_{+-}^{i j} \mid \phi_{0}\right\rangle|1\rangle \tag{S4}
\end{equation*}
$$

In the second step, to implement the element $G_{-+}^{i j}$, the single-qubit rotation is designed as

$$
\begin{equation*}
U_{-+}=|0\rangle\left\langle\phi_{-+}^{\prime}\right|+|1\rangle\left\langle\phi_{-+}^{\prime \perp}\right| \tag{S5}
\end{equation*}
$$

where

$$
\begin{equation*}
\left|\phi_{-+}^{\prime}\right\rangle=\frac{1}{\mathcal{N}_{-+}}\left(\left\langle\xi_{+-}^{i j \perp} \mid \xi_{-+}^{i j \perp}\right\rangle|0\rangle+\sqrt{1-\chi_{+-}}\left\langle\xi_{+-}^{i j} \mid \xi_{-+}^{i j \perp}\right\rangle|1\rangle\right) \tag{S6}
\end{equation*}
$$

with the normalization factor $\mathcal{N}_{-+}$. Compared Eqs. (S5) and (S7), one can see that we only replace $\left|\phi_{0}\right\rangle$ in Eq. (S5) by $\left|\xi_{-+}^{i j \perp}\right\rangle$ in Eq. (S7). The second POVM element takes the form $\left\{P_{-+}^{0}, P_{-+}^{1}\right\}$, where

$$
\begin{equation*}
P_{-+}^{0}=|0\rangle\langle 0|+\left(1-\chi_{-+}\right)|1\rangle\langle 1|, P_{-+}^{1}=\chi_{-+}|1\rangle\langle 1|, \chi_{-+}=\frac{\lambda_{-+} \mathcal{N}_{-+}^{2}}{1-\lambda_{+-}} \tag{S7}
\end{equation*}
$$

The parameter $\chi_{-+}$is chosen such that after the projector is performed, the probability of the click of $P_{-+}^{1}$ is that of the measurement $G_{-+}^{i j}$ performing on the initial state $\left|\phi_{0}\right\rangle$. The state after the second qubit rotation is $U_{-+} P_{+-}^{0} U_{+-}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right| U_{+-}^{\dagger} U_{-+}^{\dagger}$. The second detector ( $P_{-+}^{1}$ ) clicks with the probability $\operatorname{Tr}\left(P_{-+}^{1} U_{-+} P_{+-}^{0} U_{+-}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right| U_{+-}^{\dagger} U_{-+}^{\dagger}\right)=\lambda_{-+}\left\langle\xi_{-+}^{i j} \mid \phi_{0}\right\rangle\left\langle\phi_{0} \mid \xi_{+-}^{i j}\right\rangle=\operatorname{Tr}\left(G_{-+}\left|\phi_{0}\right\rangle\left\langle\phi_{0}\right|\right)$. Thus we implement the element $G_{-+}^{i j}$.

The other state without click is

$$
\begin{align*}
& \sqrt{P_{-+}^{0}} U_{-+} \sqrt{P_{+-}^{0}} U_{+-}\left|\phi_{0}\right\rangle  \tag{S8}\\
& =\left(\left\langle\phi_{-+}^{\prime} \mid 0\right\rangle\left\langle\xi_{+-}^{i j \perp} \mid \phi_{0}\right\rangle+\sqrt{1-\chi_{+-}}\left\langle\phi_{-+}^{\prime} \mid 1\right\rangle\left\langle\xi_{+-}^{i j} \mid \phi_{0}\right\rangle\right)|0\rangle \\
& +\sqrt{1-\chi-+}\left(\left\langle\phi_{-+}^{\prime \perp} \mid 0\right\rangle\left\langle\xi_{+-}^{i j \perp} \mid \phi_{0}\right\rangle+\sqrt{1-\chi_{+-}}\left\langle\phi_{-+}^{\prime \perp} \mid 1\right\rangle\left\langle\xi_{+-}^{i j} \mid \phi_{0}\right\rangle\right)|1\rangle
\end{align*}
$$

In the third step, to implement $G_{++}^{i j}$ the single-qubit rotation is designed as

$$
\begin{equation*}
U_{++}=|0\rangle\left\langle\phi_{++}^{\prime}\right|+|1\rangle\left\langle\phi_{++}^{\perp}\right| \tag{S9}
\end{equation*}
$$

where

$$
\begin{align*}
\left|\phi_{++}^{\prime}\right\rangle= & \frac{1}{\mathcal{N}_{++}}\left[\left(\left\langle\phi_{-+}^{\prime} \mid 0\right\rangle\left\langle\xi_{+-}^{i j \perp} \mid \xi_{++}^{i j \perp}\right\rangle+\sqrt{1-\chi_{+-}}\left\langle\phi_{-+}^{\prime} \mid 1\right\rangle\left\langle\xi_{+-}^{i j} \mid \xi_{++}^{i j \perp}\right\rangle\right)|0\rangle\right.  \tag{S10}\\
& \left.+\sqrt{1-\chi_{-+}}\left(\left\langle\phi_{-+}^{\prime \perp} \mid 0\right\rangle\left\langle\xi_{+-}^{i j \perp} \mid \xi_{++}^{i j \perp}\right\rangle+\sqrt{1-\chi_{+-}}\left\langle\phi_{-+}^{\prime \perp} \mid 1\right\rangle\left\langle\xi_{+-}^{i j} \mid \xi_{++}^{i j \perp}\right\rangle\right)|1\rangle\right]
\end{align*}
$$

with the normalization factor $\mathcal{N}_{++}$. Comparing Eqs. (S9) and (S11), one can find that we replace $\left|\phi_{0}\right\rangle$ in Eq. (S9) by $\left|\xi_{++}^{i j \perp}\right\rangle$ in Eq. (S11). The third POVM element takes the form $\left\{P_{++}^{0}, P_{++}^{1}\right\}$, where

$$
\begin{equation*}
P_{++}^{0}=|0\rangle\langle 0|, P_{++}^{1}=|1\rangle\langle 1| . \tag{S11}
\end{equation*}
$$

With this setup when the input state is $\left|\xi_{++}^{i j \perp}\right\rangle$ the detector $\left(P_{++}^{1}\right)$ never clicks. Thus the measurement corresponding to click of $P_{++}^{1}$ is proportional to $\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|$. We now prove the measurement we implement is exactly $G_{++}^{i j}$.

Assume the POVM elements corresponding to clicks of $P_{++}^{1}$ and $P_{++}^{0}$ are $x\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|$ and $y|\psi\rangle\langle\psi|$. Due to the fact that we have realized $\lambda_{+-}\left|\xi_{+-}^{i j}\right\rangle\left\langle\xi_{+-}^{i j}\right|$ and $\lambda_{-+}\left|\xi_{-+}^{i j}\right\rangle\left\langle\xi_{-+}^{i j}\right|$, we have

$$
\begin{align*}
x\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|+y|\psi\rangle\langle\psi| & =\mathbb{1}-\lambda_{+-}\left|\xi_{+-}^{i j}\right\rangle\left\langle\xi_{+-}^{i j}\right|-\lambda_{-+}\left|\xi_{-+}^{i j}\right\rangle\left\langle\xi_{-+}^{i j}\right|  \tag{S12}\\
& =\lambda_{++}\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|-\lambda_{--}\left|\xi_{--}^{i j}\right\rangle\left\langle\xi_{--}^{i j}\right|
\end{align*}
$$

Tracing both sides of Eq. (S12) leads to

$$
\begin{equation*}
x+y=\lambda_{++}+\lambda_{--}, \tag{S13}
\end{equation*}
$$

and

$$
\begin{equation*}
\left\langle\xi_{++}^{i j \perp}\right|\left(x\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|+y|\psi\rangle\langle\psi|\right)\left|\xi_{++}^{i j \perp}\right\rangle=\left\langle\xi_{++}^{i j \perp}\right|\left(\lambda_{++}\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|-\lambda_{--}\left|\xi_{--}^{i j}\right\rangle\left\langle\xi_{--}^{i j}\right|\right)\left|\xi_{++}^{i j \perp}\right\rangle \tag{S14}
\end{equation*}
$$

leads to

$$
\begin{equation*}
y=\lambda_{--} \tag{S15}
\end{equation*}
$$

Then we also have

$$
\begin{equation*}
x=\lambda_{++} \tag{S16}
\end{equation*}
$$

Thus the POVM element corresponding to the click of $P_{++}^{1}$ is $G_{++}^{i j}=\lambda_{++}\left|\xi_{++}^{i j}\right\rangle\left\langle\xi_{++}^{i j}\right|$. The POVM element corresponding to the click of $P_{++}^{0}$ is $G_{--}^{i j}=\lambda_{--}\left|\xi_{--}^{i j}\right\rangle\left\langle\xi_{--}^{i j}\right|$.

## II. THE MEASUREMENT STAGE OF REALIZING JOINT POVMS

In the measurement stage, the single-qubit rotations can be realized by a combination of quarter-wave plates (QWPs) and half-wave plates (HWPs), so-called a sandwich-type QWP-HWP-QWP set, with certain setting angles depending on the parameters of the joint positive operator-valued measure (POVM). The setting angles of the wave plates (WPs) used to realize the corresponding elements of joint POVMs are shown in Table S1. We employ partially polarizing beam splitters (PPBSs) with specific transmission probabilities for vertical polarization $T_{V}$ and same transmission probability for horizontal polarization $T_{H}=1$. This allows us to implement the measurement $\langle V|$ on the reflected port of the PPBSs.

In the basis $\{|H\rangle,|V\rangle\}$, the single-qubit rotations realized by HWP and QWP are

$$
\begin{align*}
R_{H W P}\left(\theta_{H}\right) & =\left(\begin{array}{cc}
\cos 2 \theta_{H} & \sin 2 \theta_{H} \\
\sin 2 \theta_{H} & -\cos 2 \theta_{H}
\end{array}\right)  \tag{S17}\\
R_{Q W P}\left(\theta_{Q}\right) & =\left(\begin{array}{cc}
\cos ^{2} \theta_{Q}+i \sin ^{2} \theta_{Q} & (1-i) \sin \theta_{Q} \cos \theta_{Q} \\
(1-i) \sin \theta_{Q} \cos \theta_{Q} & \sin ^{2} \theta_{Q}+i \cos ^{2} \theta_{Q}
\end{array}\right)
\end{align*}
$$

respectively, where $\theta_{H}$ and $\theta_{Q}$ are the angles between the optic axes of HWP and QWP and horizontal direction.

TABLE S1: The setting angles of WPs and the transmission probabilities for vertical polarization $T_{V}^{1(2)}$ of the two PPBSs for realization of joint POVMs.

| $G_{i j}$ | $\theta_{Q 1} / \theta_{Q 2}$ | $\theta_{H 1}$ | $\theta_{Q 3} / \theta_{Q 4}$ | $\theta_{H 2}$ | $\theta_{Q 5} / \theta_{Q 6}$ | $\theta_{H 3}$ | $T_{V}^{1}$ | $T_{V}^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $G_{12}$ | $-7.3^{\circ}$ | $32.9^{\circ}$ | $1.4^{\circ}$ | $-29.5^{\circ}$ | $22.4^{\circ}$ | $69.2^{\circ}$ | $0.3904(45)$ | $0.2897(50)$ |
| $G_{23}$ | $21.7^{\circ}$ | $60.2^{\circ}$ | $-4.9^{\circ}$ | $-36.5^{\circ}$ | $21.7^{\circ}$ | $60.2^{\circ}$ | $0.3904(45)$ | $0.2897(50)$ |
| $G_{13}$ | $7.3^{\circ}$ | $47.5^{\circ}$ | $-1.4^{\circ}$ | $-32.2^{\circ}$ | $-112.4^{\circ}$ | $110.8^{\circ}$ | $0.3904(45)$ | $0.2897(50)$ |

## III. RELATIVE DETECTION EFFICIENCIES OF DETECTORS

After applying joint POVM on the single-photon qubit, the photons are detected by single-photon avalanche photodiodes (APDs) on the reflected ports of the two PPBSs, and both reflected and transmitted ports of PBS, in coincidence with the trigger photons. The relative detection efficiencies of the detectors D1-D4 are measured and used to correct the coincidence counts.

To measure the relative efficiencies of the different detectors $\mathrm{D} 1, \mathrm{D} 2, \mathrm{D} 3$ and D 4 , we make a reasonable assumption that the total number of photons is fixed. We tune the setting angles of WPs to change the photon distribution. That is, for each time after we tune the setting angles of WPs, the normalized number of photons at each output port with respect to the total number of photons is changed, which can be read at the corresponding detector. The readout photon counts at each detector equal to the number of photons at each output port multiplied by the relative efficiency of the corresponding detector. After we tune the setting angles of WPs for four times, we have four linear equations with four variables (relative efficiencies of detectors) and then solve them to obtain the relative efficiencies of the detectors D1, D2, D3, and D4.

In our experiment, the relative efficiencies of the detectors D1, D2, D3, and D4 are 1, $0.9499 \pm 0.0070,0.9199 \pm 0.0069$ and $0.9801 \pm 0.0063$ respectively calculated from the experimental data. Thus we can use the relative efficiencies of the detectors to correct the photon counts in the measurement stage. For example, after the correction the photon counts at D2 which are used to calculate the probability of the photons being measured at D2 should be the readout photon counts divided by the relative efficiency of D2 $0.9499 \pm 0.0070$.

## IV. JOINT MEASURABILITY OF POVMS.

We test the joint measurability of the constructed joint POVM $G_{i j}$. For different qubit states, we analyze the experimental results of the joint POVM and test whether the marginal condition $\sum_{X_{j(i)}} G_{X_{i}, X_{j}}^{i j}=E_{X_{i(j)}}^{i(j)}$ is satisfied. Without loss of generality, we choose the four states $|H\rangle,|V\rangle,|R\rangle=(|H\rangle+i|V\rangle) / \sqrt{2}$, and $|D\rangle=(|H\rangle+|V\rangle) / \sqrt{2}$ as states being measured. The results are shown in Table S2

The measured probabilities $\operatorname{Tr}\left(\sum_{X_{j}} G_{X_{i}, X_{j}}^{i j} \rho\right)$ are in agreement with the theoretical predictions of the probabilities $\operatorname{Tr}\left(E_{X_{i}}^{i} \rho\right)$ of the POVM element $E_{X_{i}}^{i}$ on the state $\rho \in\{|H\rangle\langle H|,|V\rangle\langle V|,|R\rangle\langle R|,|D\rangle\langle D|\}$, which proves the marginal condition is satisfied. Thus the constructed joint POVM $G_{i j}$ shows the joint measurability.

TABLE S2: Experimental results of $\sum_{X_{j}} G_{X_{i}, X_{j}}^{i j}$ on the four different single-qubit states compared to the theoretical predictions of the probabilities of $E_{X_{i}}^{i}$ on these states.

|  | H | $\|V\rangle$ | R $\rangle$ | D |
| :---: | :---: | :---: | :---: | :---: |
| $G_{++}^{12}+G_{+-}^{12}$ | 0.8357(13) | 0.1607(13) | 0.4990(21) | 0.4958(20) |
| $G_{++}^{13}+G_{+-}^{13}$ | 0.8353(13) | 0.1699(13) | 0.5003(22) | 0.5029(21) |
| $E_{+}^{1}$ (Theory) | 0.8350 | 0.1650 | 0.5000 | 0.5000 |
| $G_{-+}^{12}+G_{--}^{12}$ | 0.1642(13) | 0.8393(13) | 0.5009(21) | 0.5042(20) |
| $G_{-+}^{13+}+G_{--}^{13}$ | 0.1647 (13) | 0.8301(13) | $0.4997(22)$ | 0.4971(21) |
| $E_{-}^{1}$ (Theory) | 0.1650 | 0.8350 | 0.5000 | 0.5000 |
| $G_{++}^{12}+G_{-+}^{12}$ | 0.3155(20) | 0.6747(19) | 0.5013(22) | 0.7818(17) |
| $G_{++}^{23}+G_{+-}^{23}$ | 0.3179(19) | 0.6748(19) | 0.5001(21) | $0.7788(16)$ |
| $E_{+}^{2}$ (Theory) | 0.3325 | 0.6675 | 0.5000 | 0.7901 |
| $G_{+-}^{12}+G_{--}^{12}$ | 0.6845(20) | 0.3253(19) | 0.4987(22) | 0.2182(17) |
| $G_{-+}^{23}+G_{--}^{23}$ | 0.6821(19) | $0.3252(19)$ | 0.4999(21) | 0.2212(16) |
| $E_{-}^{2}$ (Theory) | 0.6675 | 0.3325 | 0.5000 | 0.2099 |
| $G_{++}^{13}+G_{-+}^{13}$ | 0.3454(21) | 0.6652(19) | 0.5008(22) | 0.2083(16) |
| $G_{++}^{23}+G_{-+}^{23}$ | $0.3448(20)$ | 0.6519(21) | 0.5020 (22) | 0.2102(17) |
| $E_{+}^{3}$ (Theory) | 0.3325 | 0.6675 | 0.5000 | 0.2099 |
| $G_{++-}^{13}+G_{--}^{13}$ | 0.6546(21) | 0.3348(19) | 0.4992(22) | 0.7917(16) |
| $G_{+-}^{23}+G_{--}^{23}$ | 0.6551(20) | 0.3481(21) | 0.4980 (22) | $0.7898(17)$ |
| $E_{-}^{3}$ (Theory) | 0.6675 | 0.3325 | 0.5000 | 0.7901 |


[^0]:    * Electronic address: yshzhang@ustc.edu.cn
    † Electronic address: h.wiseman@griffith.edu.au
    $\ddagger$ Electronic address: gnep.eux@gmail.com

