Compressively Characterizing High-Dimensional Entangled States with Complementary, Random Filtering

Gregory A. Howland,[1,2,*] Samuel H. Knarr,[1] James Schneeloch,[1,2] Daniel J. Lum,[1] and John C. Howell[1]

[1]*Department of Physics and Astronomy, University of Rochester, 500 Wilson Boulevard, Rochester, New York 14627, USA*

[2]*Air Force Research Laboratory, 525 Brooks Road, Rome, New York 13441, USA*

(Received 23 September 2015; revised manuscript received 22 December 2015; published 12 May 2016)

The resources needed to conventionally characterize a quantum system are overwhelmingly large for high-dimensional systems. This obstacle may be overcome by abandoning traditional cornerstones of quantum measurement, such as general quantum states, strong projective measurement, and assumption-free characterization. Following this reasoning, we demonstrate an efficient technique for characterizing high-dimensional, spatial entanglement with one set of measurements. We recover sharp distributions with local, random filtering of the same ensemble in momentum followed by position—something the uncertainty principle forbids for projective measurements. Exploiting the expectation that entangled signals are highly correlated, we use fewer than 5000 measurements to characterize a 65,536-dimensional state. Finally, we use entropic inequalities to witness entanglement without a density matrix. Our method represents the sea change unfolding in quantum measurement, where methods influenced by the information theory and signal-processing communities replace unscalable, brute-force techniques—a progression previously followed by classical sensing.

DOI: 10.1103/PhysRevX.6.021018

Subject Areas: Optics, Quantum Physics, Quantum Information

I. INTRODUCTION

Practicing experimentalists most commonly perform quantum measurement in the context of state and parameter estimation [1]. While great historical emphasis has been placed on using measurement to probe the validity of quantum mechanics itself—where measurements must not only agree with quantum predictions but also rule out any competing explanations [2]—state estimation accepts quantum theory *a priori*. Here, measurements on identically prepared copies of a system are used to generate a model from which *testable* predictions can be made about future measurement statistics [3]. This point of view lifts the burden of validation, leading to simpler experiments and technologies.

Even so, quantum-state estimation remains a persistent obstacle for scaling quantum technologies. The familiar approach of quantum tomography (QT) scales at least quadratically poorly with added dimensions and exponentially poorly with added particles. QT in an N-dimensional Hilbert space requires of order N^2 measurements [4]—when N is a prime power, N projections are taken in each of $N+1$ mutually unbiased bases [5]. For example, tomography of a single-spin qubit ($N=2$) requires dividing the ensemble three ways, where expectation values of the $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ spin components are separately measured. For most nontrivial quantum systems, traditional, brute-force QT is unmanageable in the lab. In particular, continuous-variable degrees of freedom, such as transverse position and transverse momentum or energy and time, where $N \to \infty$, cannot be realistically characterized via QT [6].

Efforts to overcome the limitations of QT fall into three major categories. First, often only a subset of a system's behavior is of interest; e.g., if one only needs to predict a qubit's spin along one axis, information about the other two is irrelevant. The general tomographic density matrix can be discarded here in favor of simpler models [7]. A practical example is quantum key distribution (QKD), where only two (instead of order N) bases, such as energy and time, need to be characterized [8]. Many entanglement witnesses only require a small subset of possible measurements to confirm entanglement [9,10].

Second, one can leverage prior knowledge about a system. In standard tomography, maximum likelihood estimation is used to find a valid density matrix consistent with measurement data [11,12]—a simple assumption that quantum mechanics holds. Or, given a model of the physical system, one can begin with a prior distribution which is updated or parametrized in response to measurements, as in Bayesian inference [13,14].

One powerful presupposition is that a signal is structured, or *compressible*. For classical signals, this

*gregory.howland.3@us.af.mil

surprisingly broad assumption spawned the field of compressed sensing (CS) to tremendous multidisciplinary impact [15,16] with a strong presence in imaging [17–20]. In compressed sensing, signals are compressed during measurement so they can be sampled below the Nyquist limit [21]. Several recent efforts apply CS to quantum measurement to dramatic effect [22–26]—in some cases, reducing measurement times from years to hours [27]. For tomography, all protocols exploiting positivity are a form of compressed sensing [28].

Finally, one can choose measurements well suited to the model and prior knowledge. There is a compelling movement beyond traditional, projective measurements that localize quantum particles. Notably, there is weak measurement, where a system and measurement device are very weakly coupled, leaving the system nearly undisturbed [29]. With weak measurement, researchers have directly measured the quantum wave function [30], observed average trajectories of particles in the double-slit experiment [31], and performed tests of local realism [32]. More recently, we investigated partially projecting measurements that lie somewhere between weak and projective measurement. Using random, binary filtering in position followed by strong projections in momentum, we measured the sharp image and diffraction pattern of a transverse optical field without dividing the initial ensemble, a feat impossible for strong, projective measurements [33]. With nonprojective measurement, the conventional wisdom that incompatible variables must be separately investigated is discarded.

Guided by these principles, we demonstrate a novel approach for efficiently witnessing large-dimensional entanglement with a single set of measurements. We apply this technique to Einstein-Podolsky-Rosen (EPR) correlations in the spatial degrees of freedom of the biphoton state produced in spontaneous parametric down-conversion (SPDC), a system closely resembling the EPR *gedankenexperiment* [34,35]. Inspired by the random measurements used in CS, we show that random, local, partial projections in momentum followed by random, local, partial projections in position can be used to efficiently and accurately image EPR correlations in both domains. The ensemble is not split—position and momentum measurements are performed on the same photons. Remarkably, the measurement disturbance introduced by the momentum filtering manifests as a small amount of additive noise in the position distribution, which remains unbroadened. This allows the position and momentum measurements to be decoupled, and the joint probability distributions to be recovered in a 65,536-dimensional discretization of the infinite-dimensional Hilbert space. Our measurements do not violate the uncertainty principle; rather, they highlight the complex and subtle behavior of measurement disturbance given nonprojective measurements.

Exploiting our expectation that the distributions are highly correlated, we use compressive sensing optimization techniques to dramatically under-sample—we need fewer than 5000 measurements to obtain high-quality distributions. By comparing the conditional Shannon entropy in the position and momentum joint distributions, we witness high-dimensional entanglement and determine a quantum secret key rate for the joint system without needing a density matrix.

II. THEORY

A. Random, partially projective measurements of an EPR state

Consider a two-photon quantum state $|\psi\rangle$ encoded in the transverse-spatial degrees of freedom of the biphoton produced by SPDC. SPDC is a nonlinear-optical process, where a high-energy pump photon is converted into two lower-energy daughter photons, labeled signal and idler. Conservation of momentum dictates that the signal and idler momenta be anticorrelated for a plane-wave pump. Conservation of "birthplace," the notion that both photons originate from the same location in the crystal, dictates positive correlations in the daughters' transverse positions.

Strong correlations in incompatible observables are a signature of entanglement—in fact, the original EPR paradox was described using position and momentum [34]. EPR considers the ideal state

$$|\psi\rangle = \int dx_1 dx_2 \delta(x_1 - x_2)|x_1, x_2\rangle = \int dk_1 dk_2 \delta(k_1 + k_2)|k_1, k_2\rangle, \quad (1)$$

perfectly correlated in position and perfectly anticorrelated in momentum. Although the ideal EPR state is nonnormalizable and consequently impossible to realize in the lab, the biphoton state generated via SPDC is very similar [36,37].

EPR correlations are observed by measuring the joint probability distribution in position, $|\psi(x_1, x_2)|^2$, and in momentum $|\psi(k_1, k_2)|^2$. Because these domains of interest are known in advance, only these two distributions are needed—not a full density matrix. Spatial correlations are usually measured by jointly raster scanning single-element, photon-counting detectors through either the near field (position) or far field (momentum) [38]. This approach scales extremely poorly with increased single-particle dimensionality n—measurement time scales between n^3 and n^4. For a typical source, this could take upwards of one year for a modest $n = 32 \times 32$ pixel resolution [27].

To avoid dividing the ensemble, and to require many fewer measurements, we instead apply local, partially projective measurements in momentum followed by local, partially projective measurements in position, to the same photons. Our approach is illustrated in Fig. 1. The signal and idler photons from an EPR-like state $\psi(x_1, x_2)$ are

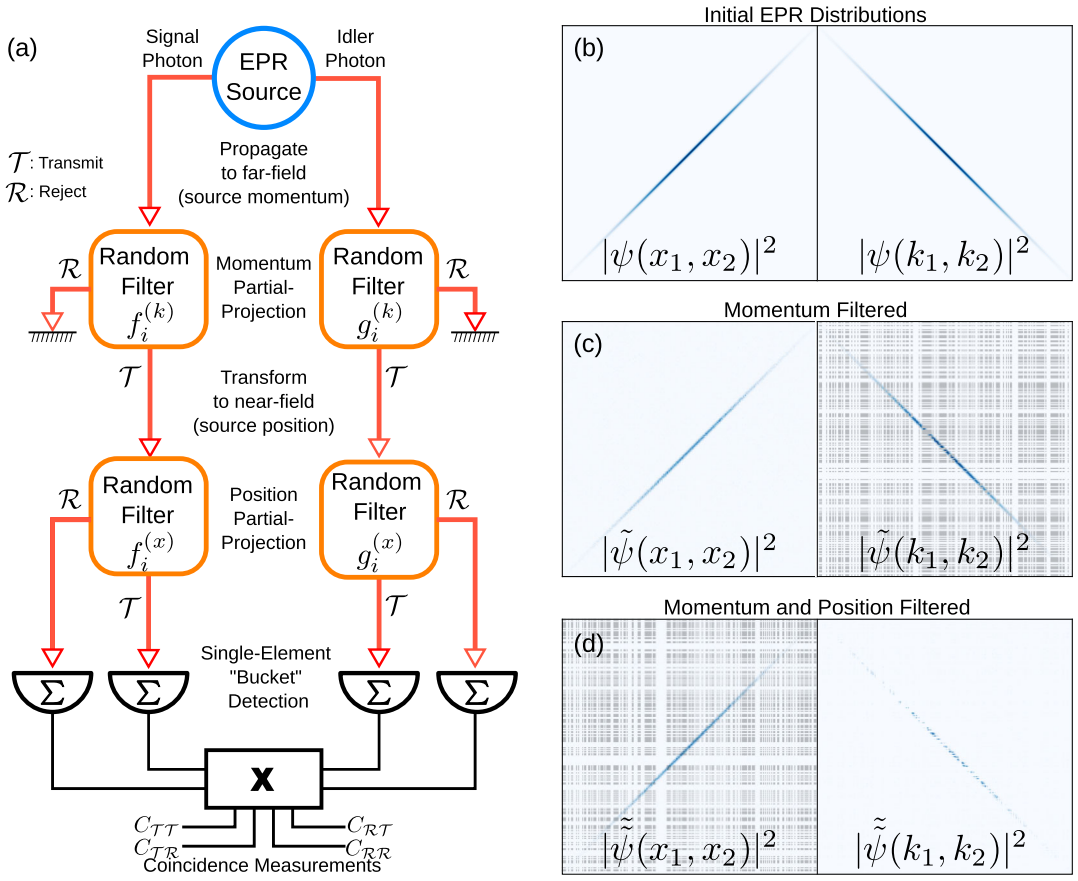

FIG. 1. Sequential, partial projections in position and momentum. The block diagram (a) describes a sequence of partially projective measurements on an EPR entangled source. (b–d) Simulated joint-position and joint-momentum distributions at each point in the experiment. Signal and idler photons from an EPR source (b) are separated and allowed to propagate to the far field (momentum). Here, they are subjected to random binary filtering by a pixelated mask (faded gray overlay). Each pixel in the mask either fully transmits ($\mathcal{T}$) or fully rejects ($\mathcal{R}$). The momentum-filtered fields (c) propagate through an optical system to an image plane of the source, where they are again filtered with random, binary filters (d). Single-element, photon-counting detectors are placed in the $\mathcal{T}$ and $\mathcal{R}$ ports of each filter and are connected to a coincidence circuit. The total number of coincident detection events between signal and idler channels gives a random projection of the momentum distribution. The relative distribution of coincident detections between the $\mathcal{T}$ and $\mathcal{R}$ modes (four possibilities) for the signal and idler photons gives a random projection of the position distribution up to a small noise floor injected by the momentum filtering.

separately allowed to propagate to the far field. Here, each photon is locally filtered by a random, binary mask $f_i^{(k)}(k_1)$ (signal) or $g_i^{(k)}(k_2)$ (idler), where subscript i refers to a particular pair of filters. Each local filter is an n-pixel, binary intensity mask, where individual pixels fully transmit ($\mathcal{T}$) or fully reject ($\mathcal{R}$) with equal probability. The momentum filtering enacts a significant partial projection of $|\psi\rangle$—on average, half of the local intensity and three-quarters of the joint intensity is rejected—so this is not a weak measurement.

All measurements are subject to uncertainty relations, which imply unavoidable measurement disturbance. Conventional projective measurements, often associated with "wave-function collapse," localize a quantum state in one domain (e.g., momentum) at the cost of broadening it in a conjugate domain (e.g., position). Critically, however, random filtering *does not localize* the quantum state; it maps a small amount of momentum information onto the total intensity passing the filter. The measurement disturbance of nonprojective measurements is best understood via the entropic uncertainty principle

$$h(x) + h(k) \geq \log(\pi e), \tag{2}$$

where $h(*)$ is the Shannon entropy. The entropic uncertainty principle implies an information exclusion relation; the more information a measurement gives about the momentum distribution, the less information a subsequent measurement can give about the position distribution [39]. There are no restrictions, however, on how information loss manifests. In particular, a measurement in one domain need not broaden, or blur, the statistics in a complementary domain.

The joint amplitude passing the momentum filtering is $\tilde{\psi}(k_1,k_2)=\psi(k_1,k_2)f_i^{(k)}(k_1)g_i^{(k)}(k_2)$. To see the effect of the momentum filtering on the position distribution, we take a Fourier transform to find $\tilde{\psi}(x_1,x_2)=\mathcal{F}\{\tilde{\psi}(k_1,k_2)\}$, which is given by the convolution of the state and filter functions in the position domain: $\tilde{\psi}(x_1,x_2)=\psi(x_1,x_2)\star(f_i^{(k)}(x_1)g_i^{(k)}(x_2))$. At high resolution, the Fourier transform of an n-pixel, random binary pattern is approximately proportional to $\delta(x)+\sqrt{2/n}\phi(x)$, where values for $\phi(x)$ are taken from a unit variance, complex, Gaussian noise distribution—a sharp central peak riding a small noise floor [33] (see Ref. [40]).

Because convolution with a delta function returns the original function, the perturbed state's position distribution is the true distribution with some weak additive noise terms,

$$|\tilde{\psi}(x_1,x_2)|^2=\mathcal{N}|\psi(x_1,x_2)\star[(\delta(x_1)+\sqrt{2/N}\phi_i(x_1))(\delta(x_2)+\sqrt{2/N}\phi_i(x_2))]|^2. \tag{3}$$

Expanding this product in powers of $1/\sqrt{N}$, where $N=n^2$, yields

$$\begin{aligned}|\tilde{\psi}(x_1,x_2)|^2=\mathcal{N}\{&|\psi(x_1,x_2)|^2\\&+\sqrt{2/N}\mathbf{Re}[\psi^*(x_1,x_2)(\psi(x_1,x_2)\\&\star(\delta(x_1)\phi_2(x_2)+\delta(x_2)\phi_1(x_1)))]\\&+\mathcal{O}(1/N)+\cdots+\mathcal{O}(1/N^2)\},\end{aligned} \tag{4}$$

where $\mathcal{N}$ is a normalizing constant. Remarkably, disturbance from filtering adds only a small noise floor, at most a factor $\sqrt{2/N}$ weaker, without otherwise broadening the position distribution. This can be seen in Fig. 1(c), where the position distribution maintains tight correlations despite the effect of momentum filtering. A rigorous derivation of Eq. (4), including the effect of finite-width pixels, is given in Ref. [40].

Next, we again perform random filtering—this time in position—as seen in Fig. 1(d). The transmitted and rejected ports are directed to single-element "bucket" detectors that are not spatially resolving. Photon detection events are time correlated with a coincidence circuit.

Each coincidence measurement contains information about both position and momentum; these must be decoupled to fit a measurement model,

$$\begin{aligned}\boldsymbol{Y}^{(k)}&=\boldsymbol{A}\boldsymbol{K}+\boldsymbol{\Phi}^{(k)},\\ \boldsymbol{Y}^{(x)}&=\boldsymbol{B}\boldsymbol{X}+\boldsymbol{\Phi}^{(x)}+\boldsymbol{\Gamma}^{(x)}.\end{aligned} \tag{5}$$

Here, $\boldsymbol{K}$ and $\boldsymbol{X}$ are N-dimensional signal vectors representing $|\psi(k_1,k_2)|^2$ and $|\psi(x_1,x_2)|^2$, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are $M\times N$ sensing matrices. $\boldsymbol{Y}^{(k)}$ and $\boldsymbol{Y}^{(x)}$ are measurement vectors whose elements are the inner product of $\boldsymbol{X}$ or $\boldsymbol{K}$ onto the ith row (or sensing vector) of $\boldsymbol{A}$ or $\boldsymbol{B}$. Noise vectors $\boldsymbol{\Phi}$ represent additive measurement noise. Noise vector $\boldsymbol{\Gamma}^{(x)}$ represents the noise injected by momentum filtering.

Momentum information is encoded in the *total* coincidences between all detection modes. Each row of $\boldsymbol{A}$ is the Kronecker product of two, random single-particle sensing vectors $\boldsymbol{a}_i^{k_1}\otimes\boldsymbol{a}_i^{k_2}$ such that $\boldsymbol{A}_i=\boldsymbol{a}_i^{k_1}\otimes\boldsymbol{a}_i^{k_2}$, where, for example, $\boldsymbol{a}_i^{k_1}$ encodes $f_i^{(k)}(k_1)$.

Position information is encoded in the *relative* distribution of coincidences between signal and idler $\mathcal{T}$ and $\mathcal{R}$ modes. By adding coincidences between like modes ($\mathcal{TT}$ and $\mathcal{RR}$) and subtracting coincidences between differing modes ($\mathcal{TR}$ and $\mathcal{RT}$), the effect of momentum filtering is removed up to injected noise. Like momentum, the position-sensing vector is a Kronecker product of two local sensing vectors: $\boldsymbol{B}_i=\boldsymbol{b}_i^{(x_1)}\otimes\boldsymbol{b}_i^{(x_2)}$. However, because of the relative measurement, the local sensing matrices take values "1" for transmitting pixels and "−1" for rejecting pixels.

In our experiment, we use a slightly more sophisticated, but conceptually similar, approach (see Ref. [40]) that retains the transmission and rejection modes from both momentum and position. In this case, there are 16 possible correlation measurements that are combined to give either position or momentum information, and both $\boldsymbol{A}$ and $\boldsymbol{B}$ take values "1" and "−1."

B. Recovering the position and momentum distributions

To obtain the joint-position and joint-momentum distributions from our measurements, we turn to compressive sensing (CS). Here, we exploit our expectation that both distributions are highly correlated. Therefore, the distributions are sparse in their natural (position-pixel or momentum-pixel) representations—relatively few elements in each distribution have significant values. This allows us to dramatically under-sample so that $M\ll N$. In this case, there are many possible $\boldsymbol{X}$ and $\boldsymbol{K}$ consistent with the measurements. CS posits that the correct $\boldsymbol{X}$ and $\boldsymbol{K}$ are the sparsest distributions consistent with the measurements.

Sparse $\boldsymbol{X}$ and $\boldsymbol{K}$ are found by solving a pair of optimization problems

$$\begin{aligned}\min_{\boldsymbol{K}}\frac{\mu_k}{2}\|\boldsymbol{Y}^{(k)}-\boldsymbol{A}\boldsymbol{K}\|_2^2+TV(\boldsymbol{K}),\\ \min_{\boldsymbol{X}}\frac{\mu_x}{2}\|\boldsymbol{Y}^{(x)}-\boldsymbol{B}\boldsymbol{X}\|_2^2+TV(\boldsymbol{X}),\end{aligned} \tag{6}$$

where $\|*\|_2^2$ is the ℓ_2 (Euclidean) norm and μ are weighting constants. The first penalty is a least-squares term that ensures the result is consistent with measured data. The second penalty $TV(*)$ is the signal's total variation (TV), which is the ℓ_1 norm of the discrete gradient

$$TV(\boldsymbol{X}) = \sum_{\text{adj}i,j} |\boldsymbol{X}_i - \boldsymbol{X}_j|, \tag{7}$$

where i, j run over pairs of adjacent elements in the signal. The TV regularization promotes structured, sparse signals over noisy, uncorrelated signals. Total variation minimization has been extremely successful for compressed sensing and denoising in the context of imaging [41–43]. In many cases, a signal can be recovered from M as low as a few percent of N. For a more complete introduction to compressive sensing, see the excellent tutorials by Baraniuk [44] and Candès and Wakin [45].

Total variation minimization is also extremely effective for denoising signals [46]. Normally, this helps to mitigates environmental and photon-counting shot noise ($\boldsymbol{\Phi}$), but in our case, it also largely removes the filtering measurement disturbance $\boldsymbol{\Gamma}$. With strong measurements, e.g., raster scanning a pinhole aperture, one requires deconvolution techniques to obtain a similar effect. Not only is deconvolution far more challenging than denoising, it can never recover high-frequency content beyond the aperture size.

CS measurements are most effective in a representation that is incoherent, or maximally unbiased, with respect to the sparse representations (in our case, position or momentum). Fortunately, random projections perfectly suit this criteria, leading to the surprising conclusion that random measurement is actually preferable. Random matrices are overwhelmingly likely to be restricted isometries that preserve the relative distance between sparse signals, ensuring that solving Eq. (6) returns the true signal instead of a sparse but otherwise incorrect result [47]. Not only do random filters extract information in complementary domains, they are the among the best measurements for leveraging CS.

One might reasonably ask if our technique employs circular reasoning—assuming the distributions are highly correlated in order to then measure their correlations. This is not the case. The initial assumption is a compressibility assumption; relative to all possible distributions, our distributions are expected to be sparse in the natural pixel basis. We do not know exactly how sparse the distributions will be, or which elements will be significant. However, the vast majority of possible distributions are just unstructured noise—these are the outcomes we are initially rejecting.

The assumption is similar to assuming that a digital photograph can be effectively compressed by the JPEG standard [48]. A natural photographic scene contains more low-spatial-frequency content than high-spatial-frequency content and contains objects with well-defined edges and recognizable shapes—regardless of the specific scene.

III. EXPERIMENT

Our experimental setup is shown in Fig. 2. An EPR-like state at 810 nm is generated by pumping a 1-mm-thick BiBO crystal oriented for type-I collinear SPDC with a 405-nm pump laser. The generated fields propagate to a spatial light modulator (SLM) in the focal plane of a 125-mm lens. Because the phase-only SLM only retards one polarization, it can perform per-pixel polarization rotation. These polarization rotations are converted to intensity modulations with a half-wave plate and a polarizing beam splitter. Random

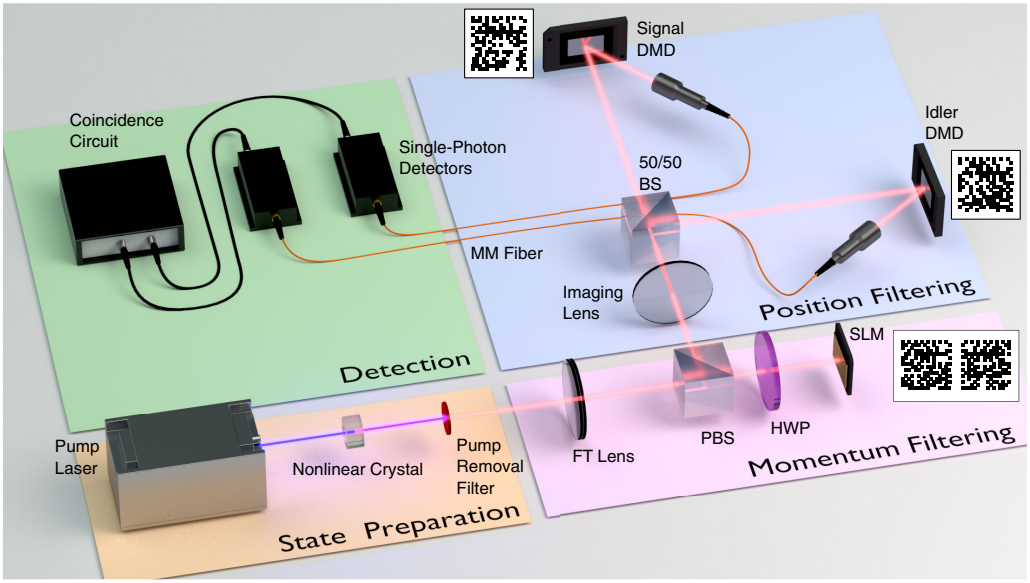

FIG. 2. Experimental setup. A two-photon, EPR-like state is generated by pumping a nonlinear crystal for type-1 SPDC. Random, binary patterns placed on a SLM in a Fourier plane of the crystal and on DMDs in an image plane of the crystal implement a sequence of random, partially projecting measurements. Example patterns are shown next to the SLM and DMDs; note the separate patterns for signal and idler photons on the SLM. Coincident detection events between single-photon detectors for signal and idler photons give information about *both* the joint-position and joint-momentum distributions of the two-photon state.

masks that cause zero or π polarization rotations perform the momentum filtering. We exploit the negative correlations in the momentum state to assign signal and idler particles to the left and right halves of the SLM, respectively.

The signal and idler fields are routed to separate digital micromirror devices (DMDs) via a 500-mm lens and a 50/50 beam splitter; the DMDs are placed in a crystal image plane with 4X magnification. A DMD is a two-dimensional array of individually addressable mirrors, each of which can be oriented to direct light towards or away from a detector. These correspond to the transmit and reject ports in Fig. 1. Random patterns placed on the DMDs implement the position filtering. The light is coupled with 10X microscope objectives into multimode fibers which are connected to avalanche photodiodes operating in geiger (photon-counting) mode. A correlator records coincident detection events between filtered signal and idler photons.

Single-particle sensing matrices $\boldsymbol{a}^{(k_1)}, \boldsymbol{a}^{(k_2)}, \boldsymbol{b}^{(x_1)}$, and $\boldsymbol{b}^{(x_2)}$ are generated by taking M rows from randomly permuted $n \times n$ Hadamard matrices. This allows the repeated calculations of $\boldsymbol{AK}$ and $\boldsymbol{BX}$ performed by the solver to use a fast Hadamard transform, decreasing computational requirements [49]. Because we only collect transmitted modes from both position and momentum filters, we require 16 separate measurements to collect all coincident combinations of transmission and rejection for the four filters (described in Ref. [40]). This is not required, in principle, if one has eight detectors. The solver we use for Eq. (6) is TVAL3 [50]. The full measurement and reconstruction recipe we follow is similar to that described in Ref. [49].

Note that our choice of a single-momentum SLM and two position DMDs was due to available equipment. One would ideally use four SLMs to implement completely separate position and momentum filtering for both the signal and idler fields. The SLM is preferred for filtering because of its high ($> 90\%$) diffraction efficiency in contrast to the lower ($\approx 20\%$) diffraction efficiency for the DMDs.

IV. RESULTS

A. Signal recovery

Sample recovered joint signals for position and momentum are given in Fig. 3 as returned directly by the solver. The single-particle resolution was $n = 16 \times 16$ pixels, so the joint signal has dimensionality $N = n^2 = 65,536$. For the sample image, $M = 4439$ random projections were used corresponding to M less than $0.07N$. Positive correlations in position and negative correlations in momentum between signal and idler particles are clearly seen. The gaps visible on the diagonal are an artifact of row-wise reshaping to one dimension—these regions are physically outside the marginal beam width.

B. Reconstruction noise

Unfortunately, the images shown in Fig. 3 do not represent valid probability distributions due to the presence of weak, zero-mean, additive noise shown in Fig. 4. Note that solving the objective function, Eq. (6), does not strictly recover a valid probability distribution as it allows negative values. We found that current, established solvers such as TVAL3 performed better without such additional constraints—improved, quantum-specific solvers are a topic of future research.

Figure 4(a) shows slices of the joint-position reconstruction along the signal axis, where each curve corresponds to a particular idler pixel. Zooming in on a region with no signal in Fig. 4(b), we observe the noise.

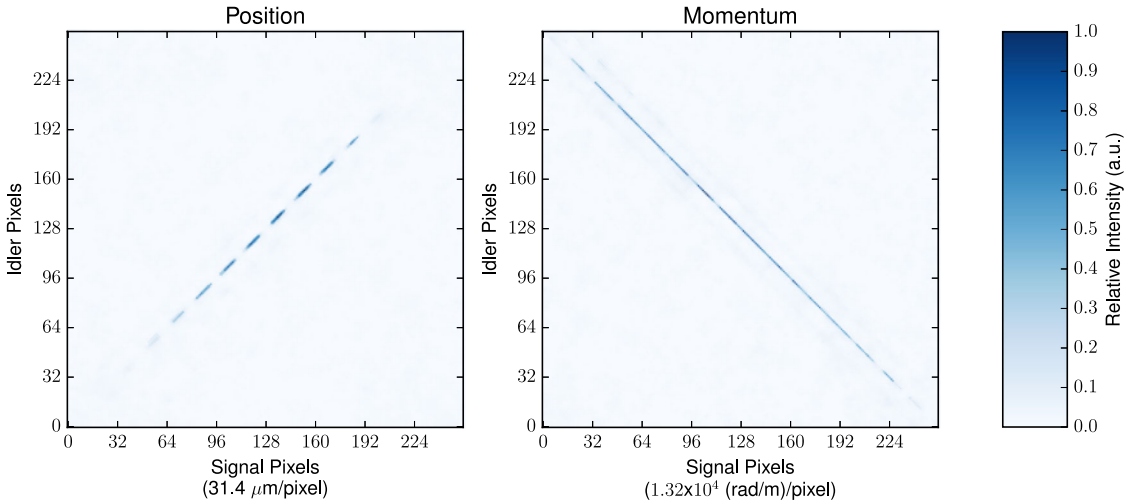

FIG. 3. Representative recovered joint-distributions in position and in momentum for a 16×16 pixel ($N = 256 \times 256$) discretization. Only $M = 4439$ measurements were needed, about $0.07N$. Gaps along the position diagonal occur because of reshaping to one dimension—these regions were outside the marginal width. Position and momentum units refer to the transverse plane at the nonlinear crystal ($z = 0$).

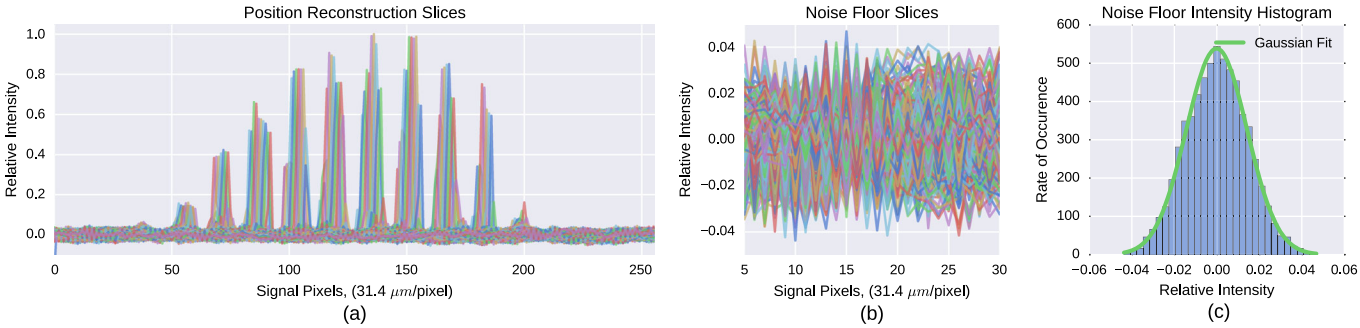

FIG. 4. Reconstruction noise. (a) One-dimensional slices along the signal axis of the joint-position reconstruction from Fig. 3 reveal the presence of zero-mean, additive Gaussian noise. The presence of negative values strongly suggests that this noise's form is nonphysical; the reconstruction process maps measurement uncertainty into this noise. A close-up of a noise-only region (signal pixels 5 to 30, all idler pixel spectra) is shown in (b). A histogram of outcomes (c) for the region shown in (b) demonstrates that the noise follows Gaussian statistics with zero mean and standard deviation 0.014. To obtain a valid probability distribution, values below a chosen threshold can be set to zero and the distribution normalized.

This noise contains both measurement uncertainty and solver artifacts. Potential noise sources include shot-noise, long-term drift in the pump laser, stray light, and crystal temperature instability. Figure 4(c) gives a histogram of the noise shown in Fig. 4(b), which follows Gaussian statistics. An appropriate model for signals returned by the solver is therefore

$$\boldsymbol{X}^{(r)} = \boldsymbol{X} + \boldsymbol{G}^{(x)}, \tag{8}$$

$$\boldsymbol{K}^{(r)} = \boldsymbol{K} + \boldsymbol{G}^{(k)}, \tag{9}$$

where $\boldsymbol{X}^{(r)}$ and $\boldsymbol{K}^{(r)}$ refer to the signals returned by the solver and $\boldsymbol{G}^{(x)}$ and $\boldsymbol{G}^{(k)}$ are additive, zero-mean Gaussian noise.

The simplest way to obtain valid probability distributions is to threshold values below a small percentage of the maximum value to zero. As seen in Fig. 4(b), any threshold below 5% removes the uniform noise floor without removing any signal peaks. This approach is similar to the common technique of subtracting dark counts from data in coincidence measurements and other noise-suppression techniques.

C. Witnessing entanglement

To witness and quantify entanglement, we violate an entropic steering inequality [51–53] (see Ref. [40]); all classically correlated states satisfy

$$H(X_1|X_2) + H(K_1|K_2) \geq 2\log\left(\frac{\pi e}{\Delta_x \Delta_k}\right), \tag{10}$$

where $H(X_1|X_2)$ and $H(K_1|K_2)$ are the conditional, discrete Shannon entropies of the respective position and momentum joint distributions. Here, Δ_k (Δ_x) is the width in momentum (position) sampled by a single-pattern pixel on the SLM (DMD) in the transverse plane of the nonlinear crystal. For position Δ_x, this is found by dividing the physical width of a pattern pixel on the DMD by the magnification of the imaging system. For momentum, the physical width of a SLM pattern pixel p_k is related to Δ_k via the Fourier-transforming property of a lens, so $\Delta_k = p_k 2\pi/(\lambda f)$, where λ is the wavelength of light and f is the lens focal length.

The entropic steering inequality is powerful because it is computed directly from measured probability distributions and does not require a density matrix. Remarkably, despite being a function of discrete distributions, it witnesses continuous-variable entanglement. Moreover, the amount the inequality is violated corresponds to a secret key rate for quantum key distribution [37,54].

The conditional entropies in position and momentum for our experimental results are given in Fig. 5 as a function of measurement number. Different curves correspond to increased levels of thresholding, setting values below a percentage of the maximum value to 0. A sharp transition from poor reconstruction to good reconstruction is clearly demonstrated by dramatic drops in the conditional entropies around $M = 2000$. This transition is characteristic of compressed sensing as the number of measurements becomes sufficient to accurately reconstruct the signal [55]—strongly suggesting we made enough measurements. For too-small M, reconstructions fail spectacularly and return unstructured noise. For a k-sparse signal (k out of N elements have significant intensity), the required number of measurements scales as $ck\log(N/k)$, where c is a near-unity constant [21]. For M beyond the transition, one is sampling above the information rate. Traditionally, one is concerned with sampling at or beyond the Nyquist rate, where $M = N$.

In momentum, the conditional entropy drops to nearly zero; in position, it drops to less than 2 bits. The position entropy likely levels off because of slight pixel misalignment between the two-position DMDs. Physically, this indicates that a particular signal position pixel is correlated to about four idler pixels, whereas a particular signal-momentum pixel is only correlated to one idler pixel.

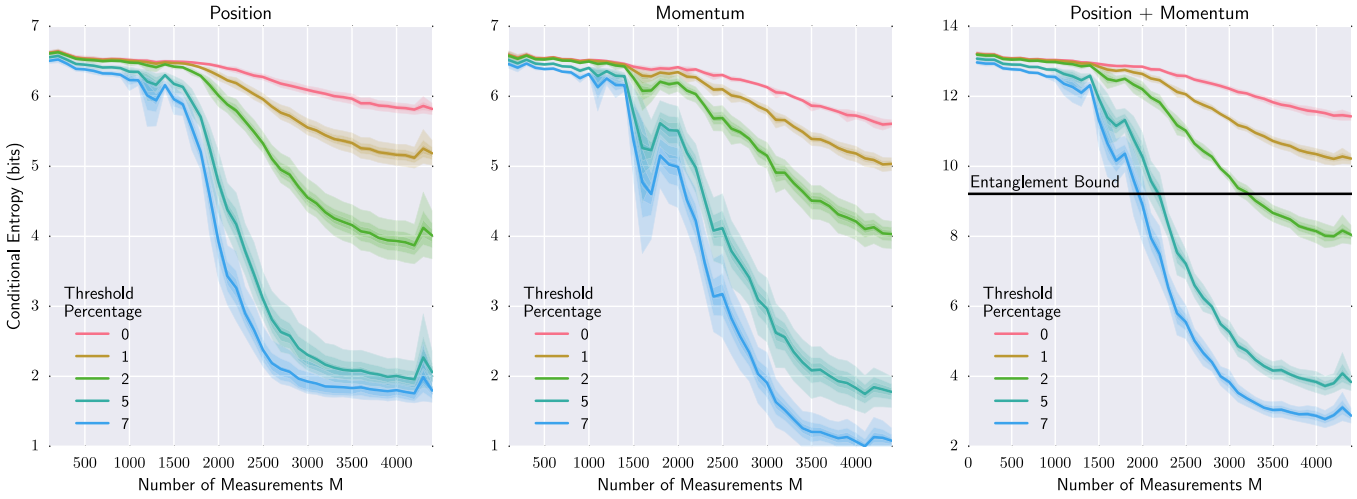

FIG. 5. Conditional entropy versus measurement number. A sharp transition from high to low conditional entropy is seen as the number of measurements increases. Note that $N = 256^2$, so $M = 2,000$ is only $0.03N$. Different curves correspond to different levels of thresholding to remove the noise floor. Bold lines indicate an average over nine trials. Faded lines enclose up to 4 standard deviations about the mean. When the conditional entropy sum is below the bound, the state is entangled.

The steering inequality is violated with as little as 2% thresholding, and by over 6 bits for thresholding beyond 7%.

The effect of thresholding for $M = 5000$ is given in Fig. 6. Figure 6(a) shows the conditional entropies for position, momentum, and their sum with the corresponding entanglement bound. Figure 6(b) gives the mutual information $I(X_1:X_2)$ and $I(K_1:K_2)$, where, for example,

$$I(X_1:X_2) = H(X_1) + H(X_2) - H(X_1, X_2). \tag{11}$$

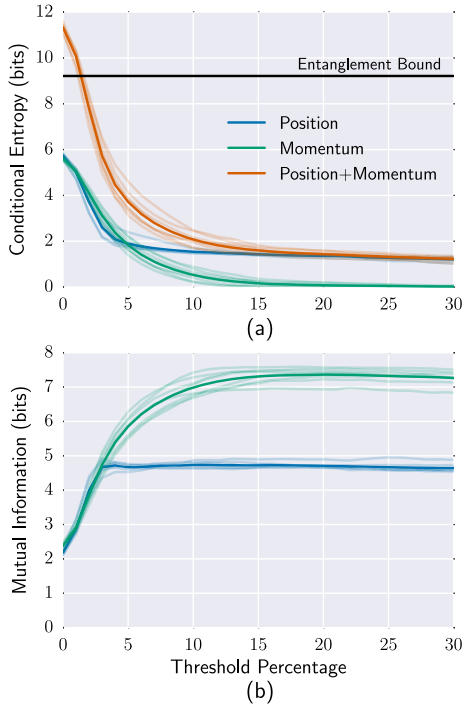

FIG. 6. Effect of thresholding. The effect of thresholding to remove weak background noise on the conditional entropy (a) and mutual information (b) is given. The bold line gives the average for nine trials; faded lines give the results from the individual trials. $M = 4439$ measurements were used. When the conditional entropy sum is below the bound, the state is entangled.

Here, $H(X_1, X_2)$ is the Shannon entropy of the joint distribution, and $H(X_1)$ and $H(X_2)$ are Shannon entropies of the marginal, single-particle distributions. From information theory, this mutual information provides a maximum bit rate for communication with joint-position or joint-momentum representations for this system [56]. The mutual information arises as a function of thresholding, indicating that thresholding is not trivially decreasing the conditional entropies and that the most likely joint outcomes are the most highly correlated. Again, the momentum mutual information is larger because of slight optical misalignments for position DMDs.

An important point is that the thresholded signal peaks still retain the additive Gaussian noise from the reconstruction process. Because of the data-processing inequality [56], this noise cannot decrease the conditional entropy and cannot increase the mutual information (this would be like arguing that a noisy channel is better for communication than its noiseless counterpart). Therefore, we conservatively underestimate our ability to violate the steering witness [see Eq. (10)].

V. CONCLUSION

We have demonstrated that local, random filtering in momentum followed by local, random filtering in position—of the same photons—can recover sharp, joint distributions for both observables. This is not possible with standard, projective measurements that localize photons in

either position or momentum. Using the expectation that the signals will be highly correlated allows us to use many fewer measurements than dimensions in the system via techniques of compressed sensing. We strongly emphasize that we have not violated any uncertainty relations; instead, we have chosen nonprojective measurements whose disturbance can easily be mitigated.

ACKNOWLEDGMENTS

This work was funded by Air Force Office of Scientific Research (AFOSR) Grant No. FA9550-13-1-0019 and AFOSR LRIR 14RI02COR. G. A. H. and J. S. acknowledge support from National Research Council Research Associate Programs. J. C. H. acknowledges support from Northrup Grumman. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL. G. A. H. conceived of the experiment and authored the manuscript with help from S. H. K., D. J. L., and J. S. S. H. K. and G. A. H. performed the experiment and analyzed the data. J. S. provided the theory on entanglement witnesses and Fourier transforms of random patterns. D. J. L. devised the scheme for using Hadamard matrices in the measurement and reconstruction process. The entire project was overseen by J. C. H.

G. A. H. and S. H. K. contributed equally to the work presented in this manuscript.

[1] M. Paris and J. Rehacek, Quantum State Estimation (Springer-Verlag, Berlin/Heidelberg, 2004), Vol. **649**.

[2] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, *Violation of Bell's Inequality under Strict Einstein Locality Conditions*, Phys. Rev. Lett. **81**, 5039 (1998).

[3] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, *Quantum Data Compression of a Qubit Ensemble*, Phys. Rev. Lett. **113**, 160504 (2014).

[4] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro, *Qudit Quantum-State Tomography*, Phys. Rev. A **66**, 012303 (2002).

[5] A. Klappenecker and M. Rötteler, *Constructions of Mutually Unbiased Bases*, in *Finite Fields and Applications* (Springer, Berlin/Heidelberg, 2004), pp. 137–144.

[6] A. I. Lvovsky and M. G. Raymer, *Continuous-Variable Optical Quantum-State Tomography*, Rev. Mod. Phys. **81**, 299 (2009).

[7] S. Aaronson, *The Learnability of Quantum States*, Proc. R. Soc. A **463**, 3089 (2007).

[8] I. Ali-Khan, C. J. Broadbent, and J. C. Howell, *Large-Alphabet Quantum Key Distribution Using Energy-Time Entangled Bipartite States*, Phys. Rev. Lett. **98**, 060503 (2007).

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

[10] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman, and M. D. Reid, *Experimental Criteria for Steering and the Einstein-Podolsky-Rosen Paradox*, Phys. Rev. A **80**, 032112 (2009).

[11] Z. Hradil, *Quantum-State Estimation*, Phys. Rev. A **55**, R1561 (1997).

[12] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Measurement of Qubits*, Phys. Rev. A **64**, 052312 (2001).

[13] C. M. Caves, C. A. Fuchs, and R. Schack, *Quantum Probabilities as Bayesian Probabilities*, Phys. Rev. A **65**, 022305 (2002).

[14] F. Huszár and N. M. T. Houlsby, *Adaptive Bayesian Quantum Tomography*, Phys. Rev. A **85**, 052120 (2012).

[15] D. L. Donoho, *Compressed Sensing*, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[16] I. Carron, *Compressive Sensing: The Big Picture*, http://sites.google.com/site/igorcarron2/cs.

[17] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, and R. G. Baraniuk, *Single-Pixel Imaging via Compressive Sampling*, IEEE Signal Process. Mag. **25**, 83 (2008).

[18] J. Romberg, *Imaging via Compressive Sampling [Introduction to Compressive Sampling and Recovery via Convex Programming]*, IEEE Signal Process. Mag. **25**, 14 (2008).

[19] B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. J. Padgett, *3D Computational Imaging with Single-Pixel Detectors*, Science **340**, 844 (2013).

[20] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, *First-Photon Imaging*, Science **343**, 58 (2014).

[21] E. J. Candes, J. Romberg, and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inf. Theory **52**, 489 (2006).

[22] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Quantum State Tomography via Compressed Sensing*, Phys. Rev. Lett. **105**, 150401 (2010).

[23] A Shabani, M. Mohseni, S. Lloyd, R. L. Kosut, and H. Rabitz, *Estimation of Many-Body Quantum Hamiltonians via Compressive Sensing*, Phys. Rev. A **84**, 012107 (2011).

[24] W.-T. Liu, T. Zhang, J.-Y. Liu, P.-X. Chen, and J.-M. Yuan, *Experimental Quantum State Tomography via Compressed Sampling*, Phys. Rev. Lett. **108**, 170403 (2012).

[25] F. Tonolini, S. Chan, M. Agnew, A. Lindsay, and J. Leach, *Reconstructing High-Dimensional Two-Photon Entangled States via Compressive Sensing*, Sci. Rep. **4**, 6542 (2014).

[26] C. Schwemmer, G. Tóth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, *Experimental Comparison of Efficient Tomography Schemes for a Six-Qubit State*, Phys. Rev. Lett. **113**, 040503 (2014).

[27] G. A. Howland and J. C. Howell, *Efficient High-Dimensional Entanglement Imaging with a Compressive-Sensing Double-Pixel Camera*, Phys. Rev. X **3**, 011013 (2013).

[28] A. Kalev, R. L. Kosut, and I. H. Deutsch, *Quantum Tomography Protocols with Positivity Are Compressed Sensing Protocols*, Npj Quantum Information **1**, 15018 (2015).

[29] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, *Colloquium: Understanding Quantum Weak Values: Basics and Applications*, Rev. Mod. Phys. **86**, 307 (2014).

[30] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, *Direct Measurement of the Quantum Wavefunction*, Nature **474**, 188 (2011).

[31] S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm, and A. M. Steinberg, *Observing the Average Trajectories of Single Photons in a Two-Slit Interferometer*, Science **332**, 1170 (2011).

[32] T. C. White, J. Y. Mutus, J. Dressel, J. Kelly, R. Barends, E. Jeffrey, D. Sank, A. Megrant, B. Campbell, Y. Chen *et al.*, *Violating the Bell-Leggett-Garg Inequality with Weak Measurement of an Entangled State*, arXiv:1504.02707.

[33] G. A. Howland, J. Schneeloch, D. J. Lum, and J. C. Howell, *Simultaneous Measurement of Complementary Observables with Compressive Sensing*, Phys. Rev. Lett. **112**, 253602 (2014).

[34] A. Einstein, B. Podolsky, and N. Rosen, *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*, Phys. Rev. **47**, 777 (1935).

[35] J. C. Howell, R. S. Bennink, S. J. Bentley, and R. W. Boyd, *Realization of the Einstein-Podolsky-Rosen Paradox Using Momentum- and Position-Entangled Photons from Spontaneous Parametric Down Conversion*, Phys. Rev. Lett. **92**, 210403 (2004).

[36] S. P. Walborn, C. H. Monken, S. Pádua, and P. H. Souto Ribeiro, *Spatial Correlations in Parametric Down-Conversion*, Phys. Rep. **495**, 87 (2010).

[37] J. Schneeloch and J. C. Howell, *Introduction to the Transverse Spatial Correlations in Spontaneous Parametric Down-Conversion through the Biphoton Birth Zone*, arXiv:1502.06996.

[38] M. N. O. Sullivan-Hale, I. A. Khan, R. W. Boyd, and J. C. Howell, *Pixel Entanglement: Experimental Realization of Optically Entangled $d = 3$ and $d = 6$ Qudits*, Phys. Rev. Lett. **94**, 220501 (2005).

[39] M. J. W. Hall, *Information Exclusion Principle for Complementary Observables*, Phys. Rev. Lett. **74**, 3307 (1995).

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.6.021018 for derivations regarding partial-projections, an expanded description of the measurement process, simulations exploring the effect of noise, and an introduction to EPR steering.

[41] X. Shu and N. Ahuja, *Hybrid Compressive Sampling via a New Total Variation TVl1*, in *Computer Vision–ECCV 2010* (Springer, Berlin, 2010), pp. 393–404.

[42] A. Chambolle and P.-L. Lions, *Image Recovery via Total Variation Minimization and Related Problems*, Numer. Math. **76**, 167 (1997).

[43] C. Li, Ph.D. thesis, Citeseer, 2009.

[44] R. G. Baraniuk, *Compressive Sensing [Lecture Notes]*, IEEE Signal Process. Mag. **24**, 118 (2007).

[45] E. J. Candè and M. B. Wakin, *An Introduction to Compressive Sampling*, IEEE Signal Process. Mag. **25**, 21 (2008).

[46] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D (Amsterdam) **60**, 259 (1992).

[47] E. J. Candès, *The Restricted Isometry Property and Its Implications for Compressed Sensing*, Comp. Rend. Math. **346**, 589 (2008).

[48] G. K. Wallace, *The JPEG Still Picture Compression Standard*, Commun. ACM **34**, 30 (1991).

[49] D. J. Lum, S. H. Knarr, and J. C. Howell, *Fast Hadamard Transforms for Compressive Sensing of Joint Systems: Measurement of a 3.2 Million-Dimensional Bi-Photon Probability Distribution*, Opt. Express **23**, 27636 (2015).

[50] C. Li, W. Yin, and Y. Zhang, *Users Guide for TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithms*, CAAM Report (2009).

[51] H. M. Wiseman, S. J. Jones, and A. C. Doherty, *Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox*, Phys. Rev. Lett. **98**, 140402 (2007).

[52] S. P. Walborn, A. Salles, R. M. Gomes, F. Toscano, and P. H. Souto Ribeiro, *Revealing Hidden Einstein-Podolsky-Rosen Nonlocality*, Phys. Rev. Lett. **106**, 130402 (2011).

[53] J. Schneeloch, C. J. Broadbent, S. P. Walborn, E. G. Cavalcanti, and J. C. Howell, *Einstein-Podolsky-Rosen Steering Inequalities from Entropic Uncertainty Relations*, Phys. Rev. A **87**, 062103 (2013).

[54] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, *One-Sided Device-Independent Quantum Key Distribution: Security, Feasibility, and the Connection with Steering*, Phys. Rev. A **85**, 010301 (2012).

[55] S. Ganguli and H. Sompolinsky, *Statistical Mechanics of Compressed Sensing*, Phys. Rev. Lett. **104**, 188701 (2010).

[56] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, Hoboken, 2012).