Verifying energy-time entanglement with irregularly sampled correlations

James Schneeloch,[1,*] Christopher C. Tison,[1,†] Richard J. Birrittella,[2] Ian Brinkley,[3,4] Michael L. Fanto,[1,5] and Paul M. Alsing[1]

[1]*Air Force Research Laboratory, Information Directorate, Rome, New York 13441, USA*
[2]*Booz Allen Hamilton, 8283 Greensboro Drive, McLean, Virginia 22102, USA*
[3]*Center for Quantum Science and Engineering, University of California Los Angeles, Los Angeles, California 90095, USA*
[4]*UCLA Physics and Astronomy, PAB 4-437, 475 Portola Plaza, Los Angeles, California 90095, USA*
[5]*RIT Integrated Photonics Group, Rochester Institute of Technology, Rochester, New York 14623, USA*

(Received 9 December 2024; accepted 11 July 2025; published 28 July 2025)

Verifying entanglement with experimental measurements requires that we take the limitations of experimental techniques into account, while still proving that the data obtained could not have been generated from a classical source. In the energy-time degree of freedom, this challenge is exacerbated because realistic high-resolution frequency measurements are obtained as a function of light passing through arbitrary filters positioned at uneven intervals. In this work, we show how the data gathered from these kinds of measurements can be used to fully certify the degree of energy and timing correlations needed to certify energy-time entanglement without having to make special assumptions about the state or the measurement device. We accomplish this by showing how to construct a continuous-variable probability density from the data that can closely estimate, but never overestimate the correlations (and entanglement) actually present in the system, and note that these methods are applicable to all continuous-variable degrees of freedom (e.g., spatial, field quadratures, etc.). We illustrate the feasibility of these methods using frequency and timing correlations obtainable from photon pairs in spontaneous parametric down-conversion.

DOI: 10.1103/6nws-mlgs

I. INTRODUCTION

As a means to quantum networking, energy-time entanglement between photons is a valuable resource because it may be transmitted over long distances through single-mode optical fiber with minimal distortion and loss. The amount of energy-time entanglement a pair of photons can share is virtually unlimited, but the direct characterization of it through measuring energy and timing correlations has remained a persistent challenge. While correlations in energy and time can be used to both witness and quantify entanglement when they exceed a classical threshold (as has been done successfully for position and momentum [1]), the stringent energy and timing resolutions required have made this difficult to achieve directly.

To date, only one experiment [2] has succeeded in directly demonstrating energy-time entanglement in this fashion, and that was because the photon pair source used was extremely narrowband using four-wave mixing in rubidium vapor. Using the quantification methods in Ref. [3], this four-wave mixing experiment effectively verifies an entanglement of formation of at least 2.54 ebits per photon pair.[1] Other demonstrations of energy-time entanglement have relied on extra assumptions about the state, such as assuming that the bandwidth of the frequency sum $\sigma(\omega_A + \omega_B)$ must equal that of the pump light consumed to generate the photon pairs (such as was done for three-photon energy-time entanglement in Ref. [4]), which can be measured separately to high precision. Alternatively, by assuming factors about the physics of the photon pair generation process (e.g., that the quantum and classical related nonlinear optical processes have related efficiencies), one can also determine the energy correlations to a high precision using stimulated emission tomography [5]. To overcome the timing resolution limits of conventional photon detectors, there has also been work in optical gating of the signal/idler photons with ultrashort pulses [6] where the enhanced timing resolution comes from an additional frequency conversion process of the down-converted light that only occurs when the photons overlap with corresponding ultrashort pulses.

In many cases, the correlations used to characterize entanglement are estimated by fitting functions to the data, and using the parameters of the fitting functions to define the degrees of correlation. Although this approach becomes more effective as the fit improves, it still risks overestimating correlations or mistakenly identifying entanglement in borderline

*Contact author: james.schneeloch.1@afrl.af.mil
†Contact author: christopher.tison.2@afrl.af.mil

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

[1]With the reported value in Ref. [2] of $\sigma(t_A - t_B)\sigma(\omega_A + \omega_B) = 0.063 \pm 0.0044$, and where [3] shows that -1 minus the base-2 logarithm of this uncertainty product is a lower bound to the entanglement of formation in ebits, we obtain from their product a minimum value of 2.546 ± 0.101 ebits.

cases. In this work, we show how to use such data to witness and certify entanglement without this manner of loophole.

The general strategy to determine energy-time entanglement directly through energy and timing correlations is to prove that these correlations exceed a threshold that all separable (i.e., unentangled) states obey. For energy and arrival time (and other Fourier-conjugate pairs of observables), the most encompassing criterion thus far was developed by Walborn *et al.* in 2009 [7].

For Fourier-conjugate continuous-variable observables of two particles A and B, Walborn *et al.* developed an entropy-based criterion that all separable states (both pure and mixed) of AB obey. Adapted for Fourier conjugates energy $E = \hbar\omega$ and (arrival) time t the inequality is as follows:

$$h(t_A - t_B) + h(\omega_A + \omega_B) \geqslant \log(2\pi e). \tag{1}$$

Here, $h(x)$ is the continuous Shannon entropy [8] of probability density $\rho(x)$ obtained from the measurement outcomes of continuous observable $\hat{x}$, and all logarithms in this paper are assumed to be base-two since we measure entropy in bits. *Note*: Although time and frequency are not regarded as direct observables in quantum mechanics (unlike position and momentum), the amplitudes associated to measuring these statistics are connected by a Fourier transform in the same way that relates position-space and momentum-space wave functions, and consequently are constrained by the same entropic relations. If we consider a state of minimum mixedness (i.e., a pure state), then this inequality will witness entanglement for even the very smallest nonzero amounts of correlation (see Appendix C for a worked-out example of the double-Gaussian wave function).[2] For mixed states, this separability inequality applies equally well, owing to the Shannon entropy being a concave function. This makes this tool (1) exceptionally powerful at identifying energy-time entanglement in the presence of both energy and time correlations.

For our analysis, we will use a looser entanglement witness, also developed by Walborn *et al.* [9] (also applying to both pure and mixed states, and also derived from the same fundamental entropic uncertainty relation [10]) utilizing conditional quantum entropies as a measure of conditional uncertainty and correlation:

$$h(t_A|t_B) + h(\omega_A|\omega_B) \geqslant \log(\pi e), \tag{2}$$

where, for example, the conditional entropy $h(t_A|t_B)$ is equal to the difference between joint and marginal entropies: $h(t_A, t_B) - h(t_B)$. Inequality (2) is known as an Einstein-Podolsky-Rosen (EPR) steering [11] inequality because its violation not only witnesses entanglement, but also demonstrates that the correlations are strong enough that one could in principle use these systems to demonstrate the EPR paradox [12] (where in short the conditional entropies would add to a limit below what is required for single particles [10]). This steering inequality is looser than the separability criterion (1) because all states violating the steering inequality (2) must also violate the separability inequality (1), meaning that the set of states obeying the steering inequality contains the set that obey the separability inequality. For what follows, we note that when the statistics of $\omega_A + \omega_B$ may be obtained directly (as statistics of $t_A - t_B$ may be obtained through coincidence counting), the methods in this work could be adapted to the separability criterion (1) to witness entanglement in a wider variety of states. For the rest of this work, we focus on techniques to violate the looser (EPR steering) inequality, which relies on more conventional sampling of ω_A and ω_B to obtain a joint probability distribution from which entropies can be calculated.

Starting with Eq. (2) as our principal entanglement witness, we may witness entanglement (and also EPR-steering) whenever the conditional entropies $h(t_A|t_B)$ and $h(\omega_A|\omega_B)$ add to a value smaller than $\log(\pi e)$. Determining this experimentally is a challenge because one cannot obtain a completely faithful representation of the probability density $\rho(\omega_A, \omega_B)$ with a finite number of measurements carried out at finite resolution. To resolve this challenge, it was shown in Ref. [13] and subsequently in Refs. [1,14] how measurements of continuous observables carried out at finite resolution can still be used to place tight upper bounds to these conditional entropies. This is distinct from the case of constructing new uncertainty relations for discrete measurements of continuous observables (as reviewed in Ref. [15]).

Where entropy is a measure of mixedness [16], and subsequent mixing (i.e., majorization) operations cannot decrease the entropy or increase correlation, this paper is focused on how to obtain a conservative estimate of the correlations present from experimental data. This process follows multiple steps of increasing complexity:

(i) First, we review how the simple coarse-graining of a continuous-variable (CV) probability distribution $\rho(\omega_A, \omega_B)$ into pixels of finite width $\Delta\omega$ and transmission spectra $\Pi(\omega - \omega_n)$, equally spaced with center frequencies $\omega_n = \omega_0 + n\Delta\omega$ (what we call simple-top-hat coarse graining) creates a distribution $\bar{\rho}_\Pi(\omega_A, \omega_B)$ whose continuous conditional entropy cannot be lower than that of the underlying CV probability distribution $\rho(\omega_A, \omega_B)$.

(ii) Next, we consider that high-resolution frequency measurements are carried out by examining the fraction of biphotons that pass through a pair of tuneable-center frequency filters whose transmission spectra $f(\omega - \omega_n)$ differ from the top-hat distributions $\Pi(\omega - \omega_n)$ (generally looking more like Lorentzian functions). We prove that when the spacing $\Delta\omega$ between two adjacent filter positions (ω_n, ω_{n+1}) is small enough that a top-hat distribution of the same width would majorize these filter functions, then the resulting probability distribution $\bar{\rho}_f(\omega_A, \omega_B)$ has a conditional entropy no less than that obtained from $\bar{\rho}_\Pi(\omega_A, \omega_B)$ via the simple top-hat coarse graining, and subsequently, no less than that obtained from the underlying CV probability distribution $\rho(\omega_A, \omega_B)$.

(iii) Third, we consider the more general case where these filter functions not only differ from the top-hat function, but also from one another, where they are only approximately regularly spaced, and only approximately equal in shape to one another due to experimental imperfections in the

[2]Indeed, one can lower-bound the amount of entanglement directly as the difference (when positive) between the measured correlations, and the mixedness of the state described by its quantum entropy (as shown in Eq. (8) in Ref. [16]). For pure states (with zero quantum entropy) any nonzero degree of measured correlations places a corresponding nonzero minimum to the amount of entanglement.

measurement apparatus. Modulo a shift of frequency center, we show that the extent to which these filter functions $f_n(\omega - \omega_n)$ are equal to one another, one can determine and apply a correction factor to the conditional entropy to ensure that the subsequent statistics never underestimate the entropy, nor overestimate entanglement.

The rest of the paper is organized as follows. For background, we first describe the concept of majorization in Sec. II for both discrete and continuous probability distributions and discuss its consequences in the estimation of entropy and correlations. We then detail in Sec. III the preceding measurement strategies, showing how to obtain conservative estimates of the entropies and correlations of the underlying quantum observables. To conclude this work, we illustrate in Sec. IV the effectiveness of our techniques by showing how typical data obtained from measuring energy and timing correlations between photon pairs generated in type-I spontaneous parametric down-conversion (SPDC) in periodically poled potassium titanyl phosphate (PPKTP) could be used to verify entanglement.

II. MAJORIZATION, ENTROPY, AND ENTANGLEMENT

Majorization of discrete probability distributions is a well-used tool to order different probability distributions by mixedness (when possible). Given two probability distributions $\vec{p}$ and $\vec{q}$ (whose components sum to unity), we say that $\vec{p}$ *majorizes* $\vec{q}$ (denoted $\vec{p} \succ \vec{q}$), when the sum of the n largest elements of $\vec{p}$ is greater than or equal to the sum of the n largest elements of $\vec{q}$ for all n from 1 to the dimension of $\vec{p}$ and $\vec{q}$.[3]

When $\vec{p} \succ \vec{q}$, one can obtain $\vec{q}$ from $\vec{p}$ through a series of permutations and mixing operations where pairs of outcomes with unequal probabilities are brought closer to their arithmetic mean. All measures of mixedness $f(\vec{p})$[4] (such as entropy, among others) will agree that if $\vec{p} \succ \vec{q}$, then $f(\vec{p}) \leqslant f(\vec{q})$. For a more thorough discussion of majorization, and its generalization to continuous probability densities, we recommend the Ref. [17].

Majorization may be used to witness entanglement by violating the majorization criterion [18], which witnesses entanglement whenever a joint state is less mixed than its marginal subsystems (by any measure of mixedness) [16]. However, this requires knowledge of the eigenspectrum of the joint and marginal density matrices, which becomes prohibitively difficult to acquire at high dimension. Instead, we may also use majorization to witness entanglement by violating entanglement criteria that place limits on the relative mixedness of the classical probability distributions of measured observables.

To connect majorization with the conservative estimation of correlations (and subsequently entanglement), we use the finding in the Supplemental Material of Ref. [1], which shows that for any two measurement probability distributions $\vec{p}$ and $\vec{q}$ (joint, marginal, or otherwise), the relative entropy between them (denoted as $\mathscr{D}(\vec{p}||\vec{q})$) cannot increase following a majorization operation (where the same operation is applied to both $\vec{p}$ and $\vec{q}$). Because the joint entropy, marginal entropy, conditional entropy, and mutual information can each be expressed in terms of relative entropies between pairs of probability distributions [8], one can show that majorization operations cannot decrease entropies of any kind, nor increase the mutual information. This is useful because if we can use experimental data to determine probability distributions for time and frequency that are definitely majorized by the true underlying distributions, then one obtains upper limits to the entropies in the left-hand side of inequality (2), which when totalling to less than $\log(\pi e)$, will witness energy-time entanglement.

III. COARSE GRAINING AT DIFFERENT LEVELS OF COMPLEXITY

In general, our task is to determine probability distributions consistent with measurement data that are majorized by the unknown underlying probability distributions to prove that the underlying correlations must exceed an entanglement-witnessing threshold. There are multiple strategies to do this, each of which is progressively more challenging.

A. Case 0: Simple top-hat coarse graining $\Pi(\omega - \omega_n)$

Top-hat coarse-graining involves partitioning a continuous space of outcomes (e.g., frequency) into discrete intervals, and using the discrete probabilities corresponding to the fraction of events occurring within each interval to create a probability density $\bar{\rho}_\Pi(\omega_A, \omega_B)$ majorized by the underlying statistics $\rho(\omega_A, \omega_B)$. This works because the total probability occurring within an interval is equal to integrating the mean probability density over that same interval. Where replacing the underlying probability density with its mean values over each interval is a form of mixing/majorization, simple top-hat coarse-graining yields probability distributions that are majorized by the underlying observable statistics. As a result, the entropies and correlations we obtain from $\bar{\rho}_\Pi(\omega_A, \omega_B)$ can be used to certify entanglement without overestimating it. For example, we have

$$H_\Pi(\Omega_A|\Omega_B) + \log(\Delta\omega_A) = h_\Pi(\omega_A|\omega_B) \geqslant h(\omega_A|\omega_B), \quad (3)$$

where $H_\Pi(\Omega_A|\Omega_B)$ is the conditional entropy of the discrete measurement probabilities of photon A having frequency within the bin Ω_A, given that photon B is known to be within frequency bin Ω_B and the measurement probabilities are obtained by integrating over simple top-hat functions denoted by Π. In addition, $\Delta\omega_A$ is the size of the frequency bins of ω_A, and $h_\Pi(\omega_A|\omega_B)$ is the continuous conditional entropy of the discretized probability densities $\bar{\rho}_\Pi(\omega_A, \omega_B)$ and $\bar{\rho}_\Pi(\omega_B)$. For a sufficiently high resolution (i.e., sufficiently small values of $\Delta\omega_A$ and $\Delta\omega_B$), the value of $H(\Omega_A|\Omega_B) + \log(\Delta\omega_A)$ will converge (from above) to the continuous-conditional entropy $h(\omega_A|\omega_B)$.

For frequency measurements, these distributions may be obtained by measuring the total power reflected off of a

[3]Note: Because their components sum to unity, majorization and weak majorization between two probability vectors are equivalent.

[4]We define for a probability distribution $\vec{p}$ that a measure of mixedness $f(\vec{p})$ is a Schur-concave function whose minimum value is obtained when one element of $\vec{p}$ equals unity with all others being zero. This minimum value is most often set equal to zero.

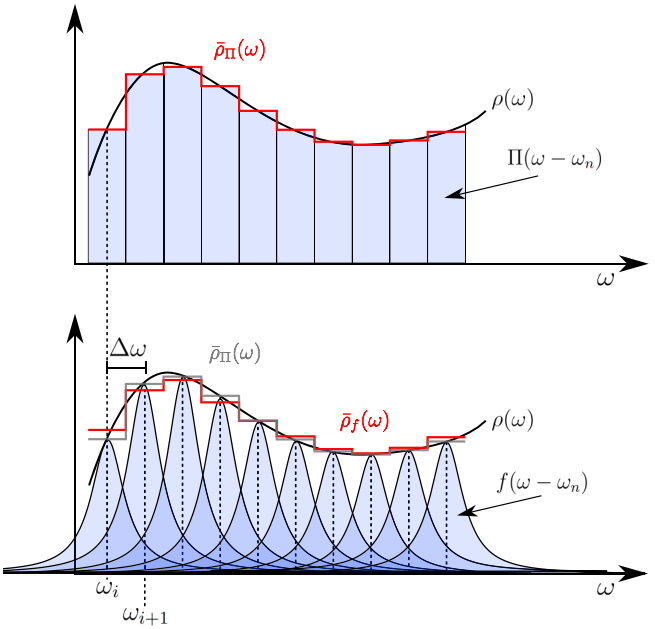


FIG. 1. Diagram showing difference between top-hat coarse graining, and sampling with a regularly spaced Lorentzian filter. The area under the shaded curves or rectangles roughly corresponds to the discrete probability associated to that frequency spacing. When the width of the Lorentzian filters $f(\omega-\omega_n)$ is wide enough that it is majorized by the corresponding top-hat filter $\Pi(\omega-\omega_n)$, then the corresponding discretized probability density $\bar{\rho}_f(\omega)$ is guaranteed to majorize $\rho(\omega)$.

diffraction grating in discrete intervals, though this becomes technically challenging at the required frequency resolutions and ranges. For timing measurements, this is more straightforward, though one is still limited by the timing resolution of one's photon counters. In practice, detector jitter and tagger jitter may be dominant sources of uncertainty in the timing measurements.

B. Case 1: Filter-shifted coarse graining $f(\omega-\omega_n)$

Where simple coarse graining measurements may be impractical, it is still possible to acquire frequency information by examining how much light passes through a given frequency filter as a function of its center frequency.

The case of simple-coarse graining can be contextualized as recording the light passing through a top-hat filter $\Pi(\omega-\omega_n)$, positioned at frequency intervals equal to its width $\Delta\omega$ (so that we have central frequencies $\omega_n=\omega_0+n\Delta\omega$). The top-hat coarse-grained probability density $\bar{\rho}_\Pi(\omega)$ is equal to an approximation of $\rho(\omega)$ but where within each interval of size $\Delta\omega$, $\rho(\omega)$ is replaced by its mean value over the same interval. Because the light passing though a top-hat filter is (up to a constant factor) equal to the amount of light incident on a discrete bin in frequency probability space, this method is already known to provide the statistics needed to verify entanglement. See Fig. 1 for diagram.

The complexity arises when we consider filters with shapes other than a top-hat function. In particular, narrow-linewidth frequency filters (such as from a Fabry-Perot etalon) used for higher-resolution frequency measurements have an approximately Lorentzian rather than top-hat transmission spectrum. In what follows, we prove when for a fixed (albeit arbitrary) filter profile $f(\omega-\omega_n)$, sampled at equal intervals, we can also obtain the statistics needed to verify entanglement.

If we consider equally spaced normalized frequency filters given by $f(\omega-\omega_n)$ (such that it integrates to unity), then whenever the peak of $f(\omega-\omega_n)$ is lower than $1/\Delta\omega$ (i.e., the peak of the normalized top-hat filter $\Pi(\omega-\omega_n)$), the resulting coarse-grained probability distribution $\bar{\rho}_f(\omega_A,\omega_B)$ will be majorized by what can be obtained with simple top-hat coarse graining $\bar{\rho}_\Pi(\omega_A,\omega_B)$.

To illustrate this, we provide the following quirk of majorization. One probability distribution majorizes another if and only if the other distribution can be expressed as the result of applying a doubly-stochastic operator on the first. For distributions $f(\omega-\omega_n)$ and $\Pi(\omega-\omega_n)$ and any central frequency ω_n, we can assert

$$f(\omega-\omega_n)\prec\Pi(\omega-\omega_n) \tag{4a}$$

$$\text{if and only if}\quad f(\omega-\omega_n)=\int d\omega' p(\omega,\omega')\Pi(\omega'-\omega_n), \tag{4b}$$

where $p(\omega,\omega')$ is a doubly-stochastic function (such that its range is nonnegative, and integrating over either ω or ω' gives unity). Moreover, because $p(\omega,\omega')$ is independent of ω_n, it may be taken to be a convolution kernel such that $f(\omega-\omega_n)$ is the convolution of $\Pi(\omega-\omega_n)$ and an unknown but fixed distribution $p(\omega)$.

With the majorization condition defined, the discrete measurement probabilities $P(\Pi_n)$ and $P(f_n)$ of continuous probability density $\rho(\omega)$ are given by

$$P(\Pi_n)=\mathcal{N}_\Pi\int d\omega\rho(\omega)\Pi(\omega-\omega_n), \tag{5a}$$

$$P(f_n)=\mathcal{N}_f\int d\omega\rho(\omega)f(\omega-\omega_n), \tag{5b}$$

and the discretized probability densities $\bar{\rho}_\Pi(\omega)$ and $\bar{\rho}_f(\omega)$ are given by

$$\bar{\rho}_\Pi(\omega)=\sum_n P(\Pi_n)\Pi(\omega-\omega_n), \tag{6a}$$

$$\bar{\rho}_f(\omega)=\sum_n P(f_n)\Pi(\omega-\omega_n), \tag{6b}$$

where the normalization constants $\mathcal{N}_\Pi$ and $\mathcal{N}_f$ are defined such that the discrete probabilities summed over n add to unity.

To begin, we examine the discrete probabilities $P(f_n)$ from which $\bar{\rho}_f(\omega)$ is constructed on the assumption that $f(\omega-\omega_n)\prec\Pi(\omega-\omega_n)$:

$$P(f_n)=\mathcal{N}_f\int d\omega\rho(\omega)f(\omega-\omega_n) \tag{7a}$$

$$=\mathcal{N}_f\int d\omega\rho(\omega)\left(\int d\omega' p(\omega,\omega')\Pi(\omega'-\omega_n)\right) \tag{7b}$$

$$=\mathcal{N}_f\iint d\omega' d\omega\ \rho(\omega)p(\omega,\omega')\Pi(\omega'-\omega_n) \tag{7c}$$

$$= \mathcal{N}_f \int d\omega' \left(\int d\omega \rho(\omega) p(\omega, \omega') \right) \Pi(\omega' - \omega_n) \quad (7d)$$

$$= \mathcal{N}_f \int d\omega' \tilde{\rho}(\omega') \Pi(\omega' - \omega_n). \quad (7e)$$

The first simplification comes from the majorization relationship allowing us to substitute $f(\omega - \omega_n)$ for the doubly-stochastic integral over $\Pi(\omega - \omega_n)$. By regrouping terms, we can see that these probabilities $P(f_n)$ can be expressed as a simple-top-hat coarse graining over a new probability density $\tilde{\rho}(\omega)$, also expressed as a doubly-stochastic integral over $\rho(\omega)$. From this we understand first, that $\tilde{\rho}(\omega) \prec \rho(\omega)$, and that $\bar{\rho}_f(\omega) \prec \tilde{\rho}(\omega)$. Therefore, we can conclude that $\bar{\rho}_f(\omega) \prec \rho(\omega)$, and the statistics from these frequency measurements can be used to place conservative bounds on the entropies used in the entanglement witness (2) so as not to falsely witness it. Moreover, since $P(f_n)$ is related to $\tilde{\rho}(\omega)$ by the same doubly-stochastic linear transformation that generates $P(\Pi_n)$ from $\rho(\omega)$, we also have that $P(f_n) \prec P(\Pi_n)$. In brief, for any doubly-stochastic linear operator M such that $\rho' = M(\rho)$ and $\sigma' = M(\sigma)$, one may prove that if $\rho \succ \sigma$, then $\rho' \succ \sigma'$. See Appendix B for argument.

In general, we may state that if $f(\omega - \omega_n) \prec \Pi(\omega - \omega_n)$, then

$$H_f(\Omega_A|\Omega_B) + \log(\Delta\omega_A) \geqslant h(\omega_A|\omega_B), \quad (8)$$

where $H_f(\Omega_A|\Omega_B)$ is the discrete conditional entropy obtained from the measurement probabilities gathered from light passing through filters $f(\omega - \omega_n)$ instead of $\Pi(\omega - \omega_n)$.

To utilize these techniques, we need to know the widest frequency spacing $\Delta\omega$ in which the filter functions will still be majorized by the corresponding top-hat distributions of width $\Delta\omega$. For a top-hat filter, it is straightforward to understand that any frequency spacing smaller than the filter width $\Delta\omega$ will satisfy this condition. For a Gaussian filter of Full-width at half-maximum F_G, the maximum spacing of $\Delta\omega$ will be $F_G\sqrt{\pi/(4\ln(2))}$, or about $1.06F_G$. For a Lorentzian filter of full-width at half-maximum F_L, the maximum spacing of $\Delta\omega$ will be $(\pi/2)F_L$, or about $1.57F_L$. See Appendix C for a full discussion of these sampling conditions.

Filters intermediate between Gaussian and Lorentzian, such as a Voigt profile resulting from a Lorentzian filter with Gaussian jitter sampled over time will have a correspondingly similar sampling spacing requirements to the Lorentzian and Gaussian filter cases. In general, we may expect that a filter whose width is larger than the spacing $\Delta\omega$ between sampled frequencies will be one that meets our sampling requirements.

C. Case 2: Sampling with changing filters at changing intervals

Thus far, we have a viable strategy for conservatively characterizing frequency correlations between photon pairs on the assumption that we assemble a coarse-grained probability distribution by shifting a pair of filters to different pairs of frequencies (such that the spacing between adjacent filter positions is generally a good deal smaller than its linewidth to satisfy the majorization relation). To make the preceding methods more fully general, we have to consider that the filter $f(\omega - \omega_n)$ will generally change shape when set to different central frequencies (however slightly). In what follows, we document how bounding the difference between the fluctuating frequency filter $f_n(\omega - \omega_n)$ and an empirical ideal $f(\omega - \omega_n)$ can be used to establish a conservative estimate of correlations even when the filtering process is not a static one.

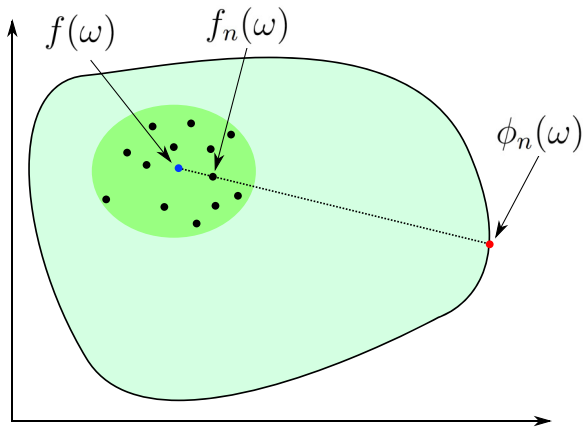


FIG. 2. Diagram showing convex set normalized filter functions and how the individual filter functions $f_n(\omega)$ relate to the target filter function $f(\omega)$ and extremal filter function $\phi_n(\omega)$. In particular, $f(\omega)$ is taken as the arithmetic mean of $f_n(\omega)$ over all values of n.

If we relax the assumption that the filter functions $f_n(\omega - \omega_n)$ have an identical profile (modulo a shift), then the doubly-stochastic operator will itself be a function of n:

$$f_n(\omega - \omega_n) = \int d\omega' p_n(\omega, \omega') \Pi(\omega' - \omega_n), \quad (9a)$$

$$P(f_n) = \mathcal{N}_f \int d\omega' \left(\int d\omega \rho(\omega) p_n(\omega, \omega') \right) \Pi(\omega' - \omega_n), \quad (9b)$$

$$= \mathcal{N}_f \int d\omega' \bar{\rho}_n(\omega') \Pi(\omega' - \omega_n), \quad (9c)$$

where here, the distributions $\bar{\rho}_n(\omega)$ each are majorized by $\rho(\omega)$, but it remains to be proven whether the statistics of $P(f_n)$ in this situation can be used to conservatively estimate entanglement.

Let us consider the situation where all of these filters are "nearly" equivalent, such that the normalized filter functions $f_n(\omega - \omega_n)$ are all nearly equal (for the same center frequency) to a single filter function $f(\omega - \omega_n)$. Accounting for this same central frequency, we would say $f_n(\omega) \approx f(\omega)$. Because the set of normalized filter functions is convex, every filter function $f_n(\omega)$ can be expressed as a weighted mixture of a fixed filter function $f(\omega)$, and an extremal filter function $\phi_n(\omega)$:

$$f_n(\omega) = w_n f(\omega) + (1 - w_n)\phi_n(\omega). \quad (10)$$

This weighted average may in turn be interpreted as the distribution obtained when drawing from either $f(\omega)$ or $\phi_n(\omega)$ with probabilities w_n and $1 - w_n$, respectively.

To the extent that all of the filter functions $f_n(\omega)$ are nearly equal to one another, we can select a target filter function $f(\omega)$ such the weights w_n are on average nearly unity. See Fig. 2 for diagram. In particular, by taking $f(\omega)$ to be the arithmetic mean of all $f_n(\omega)$, this target function is one whose mean square distance to all functions $f_n(\omega)$ is minimized.

Within the convex set of normalized filter functions, we define the extremal filter functions $\phi_n(\omega)$ relative to $f(\omega)$ and $f_n(\omega)$ such that a line starting at $f(\omega)$ and passing through $f_n(\omega)$ will intersect the boundary of the set at $\phi_n(\omega)$. See Fig. 2 for diagram.

Using this weighted decomposition of $f_n(\omega)$, and the fact that the simple top-hat coarse graining is a linear operation, one can decompose the discrete probability distribution $P(f_n)$ similarly using the same set of weights:

$$P(f_n) = w_n P(f_n|f) + (1 - w_n)P(f_n|\phi_n). \tag{11}$$

Here, $P(f_n|f)$ can be associated with the probability of passing through filter f_n given it is drawn from the target function $f(\omega)$, and $P(f_n|\phi_n)$ is similarly defined, relative to the extremal filter function $\phi_n(\omega)$. Just as $f_n(\omega)$ is equivalent to the average distribution obtained when selecting from either $f(\omega)$ or $\phi_n(\omega)$, the corresponding discrete probabilities $P(f_n)$ follow the same distribution decomposition.

Extending this analysis to two dimensions (to measure joint statistics for biphotons passing through pairs of filters), we have

$$P(f_m, f_n) = w_{mn}P(f_m, f_n|p) + (1 - w_{mn})P(f_m, f_n|\phi_{mn}). \tag{12}$$

In what follows, we shall show how a minimum weight w_0, defining the relative closeness of different filter functions $f_n(\omega)$, can be used as a correction factor between the conditional entropies $H(f_A|f_B)$ (obtained through measurement) and $H_f(\Omega_A|\Omega_B)$ (assuming a constant filter function) so that its resulting value remains a conservative estimation.

Let $P(G = 1|m, n)$ be the probability that the statistics are drawn from the target distribution when employing filter types (m, n) (also given by weight w_{mn}), and let $P(G = 0|m, n)$ be the corresponding probability that this does not happen. Then the joint frequency measurement distribution $P(f_A, f_B|m, n)$ (also given as $P(f_m, f_n)$) when using filter types m and n can be broken up according to these weighted probabilities:

$$\begin{aligned} P(f_A, f_B|m, n) = {} & P(G = 1|m, n)P(f_A, f_B|G = 1, m, n) \\ & + P(G = 0|m, n)P(f_A, f_B|G = 0, m, n). \end{aligned} \tag{13}$$

To reiterate, the event $G = 1$ is the event that the data is drawn from the target distribution. As a consequence, we note that

$$H(f_A|f_B, G = 1) = H_f(\Omega_A|\Omega_B). \tag{14}$$

From this, we can make the following relation regarding the conditional entropy $H(f_A|f_B)$:

$$\begin{aligned} H(f_A|f_B) &\geqslant \sum_{m,n} P(m, n)H(f_A|f_B, G, m, n) \\ &= \sum_{m,n} P(m, n)[P(G = 0|m, n)H(f_A|f_B, G = 0, m, n) + P(G = 1|m, n)H(f_A|f_B, G = 1, m, n)] \\ &\geqslant \sum_{m,n} P(m, n)P(G = 1|m, n)H(f_A|f_B, G = 1, m, n) \\ &= \left(\sum_{m,n} P(m, n)P(G = 1|m, n)\right) H_f(\Omega_A|\Omega_B) \\ &\geqslant (\min_{m,n}(P(G = 1|m, n)))H_f(\Omega_A|\Omega_B). \end{aligned} \tag{15}$$

The simplification between the second and third expressions comes from the nonnegativity of the entropy. The simplification from the third to the fourth expression comes from the statement that the filter function specified by $G = 1$ is independent of values m and n. Between the fourth and fifth expression, we have used the fact that the mean value of a function must be greater than or equal to its minimum value.

To condense notation, we will define this minimum in parentheses as the constant w_0:

$$w_0 \equiv (\min_{m,n}(P(G = 1|m, n))). \tag{16}$$

It is in this way that when w_0 is close to unity, we may conservatively put a relatively tight upper bound on $h(\omega_A|\omega_B)$:

$$\begin{aligned} \frac{H(f_A|f_B)}{w_0} + \log(\Delta\omega_A) &\geqslant H_f(\Omega_A|\Omega_B) + \log(\Delta\omega_A) \\ &\geqslant h(\omega_A|\omega_B). \end{aligned} \tag{17}$$

Thus, by determining a minimum value of w_0, one may witness energy-time entanglement so long as the frequency spacing is sufficiently narrow relative to the filter linewidth.

As the filter functions $f_n(\omega)$ can be measured independently, we can explicitly determine w_0 by considering their proximity to the mean filter function, or some other common centroid. Given how the filter functions $f_n(\omega)$ are expressible as the weighted average (10) between a target filter function $f(\omega)$ and extremal filter function $\phi_n(\omega)$, the filter function $f_n(\omega)$ is always intermediate between $f(\omega)$ and $\phi_n(\omega)$ for all values of ω (see Fig. 2 for diagram). These quantities (in addition to w_n) are all nonnegative, which together gives us the inequality

$$w_n \leqslant \frac{f_n(\omega)}{f(\omega)} \tag{18}$$

for all values of ω. Then, because $\phi_n(\omega)$ is on the boundary of the set of possible probability density functions, there must

be at least one value ω_0 for which $\phi_n(\omega_0) = 0$, but where $f(\omega_0) \neq 0$ and $f_n(\omega_0) \neq 0$. Our inequality is thus tight, and we have a formula for w_n in terms of determined or measured quantities:

$$w_n = \min_{\omega} \left(\frac{f_n(\omega)}{f(\omega)} \right). \tag{19}$$

Once this weight is determined for each filter function, the final value for w_0 will just be the minimum among these weights.

Note: For these weights to be usable, the filter function cannot decay too rapidly (i.e., exponentially or faster) because for finite changes in the parameters of rapidly decaying filter functions, there will be values of ω for which the ratio/weight w_n will be arbitrarily close to zero. Fortunately, the most common frequency filters have Lorentzian spectra (decaying more slowly), which may be well-utilized in this style of measurement.

As one example, we can consider a target Lorentzian filter function $f(\omega)$ of FWHM a, and the empirical filter function $f_n(\omega)$ of FWHM$a(1 + \epsilon_n)$ (for $\epsilon_n \geqslant 0$). In this case, w_n would be $1/(1 + \epsilon_n)$, and for small ϵ_n, would differ from unity by approximately ϵ_n as well.

For a Lorentzian filter function with a constant FWHM of a, but a small varying central position differing from x_n by amount $\delta_n a$, we have

$$w_n = 1 + \frac{\delta_n^2}{2} - \delta_n \sqrt{1 + \frac{\delta_n^2}{4}}, \tag{20}$$

which for small deviations, approaches

$$w_n \approx 1 - \delta_n. \tag{21}$$

Thus, for Lorentzian functions, this measurement strategy can be expected to work well so long as the deviations of the functions from the target are a relatively small fraction of the FWHM.

IV. SUITABILITY OF MAJORIZATION-BASED SAMPLING USING TYPICAL EXPERIMENTAL PARAMETERS

The general experiment we consider requires a source of energy-time entangled photon pairs, but is not subject to any one physical process of generating them. For ease of illustration, we consider the process of SPDC to be what generates these entangled biphotons, but it is not the only kind of process that can do so. For our experiment, we consider a single-mode pump laser shining through a $\chi^{(2)}$-nonlinear-optical medium (typically a noncentrosymmetric crystal). Inside this medium, provided the conditions are met for simultaneous energy and momentum conservation (i.e., phase matching) events will occur in which single pump photons are annihilated to create signal-idler biphotons. These photon pairs can exhibit high levels of energy-time entanglement because their point of origin is nearly simultaneous (so that $\sigma(t_A - t_B)$ will be small) implying large timing correlations; and their energies must sum to the energy of the annihilated pump photon (so that $\sigma(\omega_A + \omega_B)$ is small for narrow pump bandwidths) implying large energy correlations. As mentioned earlier in this work, when these correlations together exceed a critical threshold (2), we may directly witness the biphoton's energy-time entanglement.

A. Timing correlation acquisition

In this subsection, we will show how the timing correlations of photon pairs are measured experimentally using coincidence counting. In short, coincidence counting allows us to obtain an empirical probability distribution of the arrival time difference $t_A - t_B$ by compiling a histogram of these timing differences between two distinct photon detectors. The distribution of $(t_A - t_B)$ may be used to verify a maximum value for the entropy $h(t_A - t_B)$, which is in turn an upper limit for the conditional entropy $h(t_A|t_B)$, which allows us to use this coincidence counting data for our entanglement witnessing criteria.

To record the number of photon pairs being generated with single-photon detectors, we split the biphotons so that one photon goes into one arm, and the other photon goes into another arm toward respective single-photon detectors. When the first detector registers a photon, a photon correlator used to measure coincidence counts starts a timer, which is stopped only when the second detector registers an event. As photons hit these detectors, a histogram of these time intervals builds up such that pairs of photons generated simultaneously will have a characteristic peak in this coincidence histogram associated to the difference in time each photon takes to travel from its source to its detector. As these timing measurements are digital, the coincidence peak covers a range of time bins, and one may use this discrete data of the empirical probability distribution of $(t_A - t_B)$ to conservatively bound the entropies used in witnessing energy-time entanglement via simple top-hat coarse graining.

After finding the coincidence window, and subtracting off the relatively small background, we may normalize the resulting histogram quantized into bins of size Δt into an empirical discrete probability distribution, and use it to conservatively bound the continuous entropy $h(t_A - t_B)$ with the corresponding entropy of its discrete approximation:

$$h(t_A|t_B) \leqslant h(t_A - t_B) \leqslant H(T_A - T_B) + \log(\Delta t). \tag{22}$$

Here, Δt is set by the time bin size chosen to make the coincidence histogram (in this case, 1 ps).

Beyond the time-bin size, this method of measuring $t_A - t_B$ is limited in precision by the jitter time of each detector, which corresponds to the uncertainty between when a photon is absorbed, and when the detector generates an electrical pulse corresponding to a single photon count. With detector jitter times at around 20 ps, and assuming their jitters are Gaussian-distributed and statistically independent, this means that if the photon pairs were completely simultaneous, then the minimum width of the coincidence histogram would be 28.28 ps, or the jitter time multiplied by $\sqrt{2}$.

Independent of technical noise, there is still a fundamental nonzero time uncertainty (of $t_A - t_B$) of the photon pairs as determined by the physics of SPDC, and in particular, by the phase matching. For a type-0 or type-I degenerate SPDC source crystal of length L_z, we may estimate [19] this uncer-

tainty as

$$\sigma(t_A - t_B) \approx \sqrt{\frac{9L_z\kappa_1}{10}}, \tag{23}$$

which for a 20 mm crystal of PPKTP pumped at 775 nm, (where κ_1, the group velocity dispersion constant is determined from Sellmeier equations in Ref. [20] to be 292 fs^2/mm) may be estimated to be of the order 72.5 fs, which assuming a Gaussian distribution corresponds to a FWHM of about 171 fs. This is well-below the jitter time uncertainty of 2.828×10^4 fs, so these pairs may be regarded as nearly simultaneous.

While we would use the empirical probabilities to experimentally bound the continuous entropy $h(t_A - t_B)$ by its discrete approximation, we can also use the given parameters to estimate how much entanglement may be witnessed for these biphotons. As a Gaussian distribution has maximum entropy for a given mean and variance, we may upper-bound the entropy $h(t_A - t_B)$ as

$$h(t_A - t_B) \leqslant \frac{1}{2}\log(2\pi e\sigma(t_A - t_B)^2) \tag{24}$$

$$\approx \frac{1}{2}\log\left(\frac{\pi e(\text{FWHM}(t_A - t_B))^2}{4\ln(2)}\right). \tag{25}$$

Using our example $\text{FWHM}(t_A - t_B)$ of 424 ps, and the timing resolution was 1 ps, we find the continuous time entropy to be approximately -30.324 bits. This defines the threshold of energy correlations we must surpass to demonstrate energy-time entanglement.

B. Energy correlation acquisition

1. Limit to beat

Given our maximal value of -30.324 bits for $h(t_A - t_B)$, and the uncertainty bound of $\log(\pi e) \approx 3.0942$ bits, this means that $h(\omega_A|\omega_B)$ must be less than 33.42 bits to witness entanglement, corresponding to a $\sigma(\omega_A + \omega_B)$ less than $2\pi(442.0\,\text{MHz})$, or when centered at 1550 nm, a linewidth (FWHM) of 8.8 pm.

The joint frequency distribution of the biphotons from SPDC has an uncertainty $\sigma(\omega_A + \omega_B)$[5] dominated by the pump bandwidth, which may be arbitrarily narrow. Consequently, our ability to demonstrate entanglement will be dominated by the resolutions we may achieve in different frequency measurement setups.

2. Direct quantization into frequency bins via diffraction or dispersion

The most straightforward way to perform these joint frequency measurements is to split the signal and idler photons into separate arms, use a spatially dispersive element such as diffraction grating or prism in concert with other optics including a digital micromirror device (DMD) array, to reflect portions of the joint frequency spectrum toward or away from photon detectors from which coincidence counts could be recorded. With this manner of acquisition and the previously acquired timing correlations, we would need frequency bins small enough to resolve the linewidth. Where the discrete approximation to the continuous entropy is an upper bound to it, $H(\Omega_A|\Omega_B) + \log(\Delta\omega) \geqslant h(\omega_A|\omega_B)$, we need this discrete entropy approximation to total a value less than 33.42 bits. Where the linewidth (FWHM) of the filter places an upper limit on acceptable frequency spacing $\Delta\omega$, we would need frequency bins significantly smaller than 1 GHz, or 8 pm (ideally by an order of magnitude) to begin witnessing entanglement. In what follows, we describe the technological limitations of different optical frequency measurement techniques.

Diffraction gratings are rated by their resolving power and diffraction efficiency. Where the diffraction efficiency describes the fraction of input light that exits the first diffracted mode, the resolving power is given from the Rayleigh criterion of being able to just barely distinguish two spectral lines in the far field. A typical diffraction grating adapted to the telecom band might have 600 grooves per mm and be able to have a 10 mm diameter beam incident upon it without clipping. Together, this corresponds to a resolving power of about one part in 6000 for the beam of light. If our central frequency is 193.4 THz (i.e., 1550 nm), then one might expect to be able to resolve the difference between two spectral lines 32 GHz apart. Any finer spectral features would effectively be blurred to approximately this size. As the maximum frequency bin size needed to witness energy-time entanglement (with the same timing correlation data) is likely an order of magnitude below 1 GHz, any apparatus utilizing a diffraction grating will have to be impractically large to have a sufficient resolving power (e.g., a 320–3200 mm wide diffraction grating uniformly illuminated by a single beam).

Prisms may be rated by their resolving power as well [21].[6] For typical equilateral triangle dispersive prisms, their resolving power is competitive with what can be accomplished with a diffraction grating, though this may be greatly enhanced by using a material with very high dispersion (by as much as three orders of magnitude [22]), though this dispersion often exists only at a narrow range of frequencies such as between two atomic transitions in a triangular rubidium vapor cell. Unlike diffraction gratings, ordinary prisms may work over a broader range of wavelengths, but this comes at the expense of having to keep track of how the refractive index varies over a large range of wavelengths (which may be far from a linear function over such a range).

Where both diffraction gratings and prisms rely on wavelength-dependent beam deflection to make spectral measurements possible, there is a fundamental tradeoff between the minimum resolvable difference between two wavelengths, and the maximum range of wavelengths over which data can be acquired, as these correspond to minimum and maximum observable angular deflections. This is also true for additional methods relying on angular dispersion, such as virtually imaged phased arrays [23,24] which have much improved frequency resolution albeit over a much narrower range.

[5]Where $\sigma(x)$ denotes the standard deviation of random variable x.

[6]For an equilateral prism of side length b, the Rayleigh resolving power of a prism (surrounded by vacuum) may be given by the formula $\Delta\lambda/\lambda \approx -bn(n - n_g)/\lambda$.

These arrays can be concatenated with other optical dispersive elements to achieve high resolutions over a comparatively wide bandwidth [25,26], though it remains to be shown how this method of measurement can be best used to conservatively estimate entanglement. Fortunately, there are alternative methods to separate different frequency bands of light simple enough that their analysis toward realizing energy-time entanglement is straightforward. In what follows, we consider the feasibility of acquiring the necessary frequency statistics by scanning tunable narrow-linewidth frequency filters.

3. Requirements of scanning filters

For our frequency measurement, the last situation we consider is scanning a pair of tuneable Fabry-Perot Etalons. Where the pump spectrum may be arbitrarily narrow, we would be limited to the width of the transmission peaks of these etalon filters, which may be described by a Lorentzian function. With a Lorentzian filter profile (so that $h(\omega_1 + \omega_2) = \log(2\pi \mathrm{FWHM}(\omega_1 + \omega_2))$), we find that the maximum value for the $\mathrm{FWHM}(\omega_1 + \omega_2)$ is $2\pi \times 0.29$ GHz. For scale, at 1550 nm, a fluctuation of 1.25 GHz is a fluctuation of 0.01 nm, so these resolutions, while challenging to achieve, are obtainable with current technology.

In the case where we are able to scan sufficiently narrow filters, at sufficiently high resolution, it is worth considering not just how wide the filter spacing $\Delta\omega$ can be while still satisfying our majorization sampling condition, but also how narrow $\Delta\omega$ must be so that the discretely approximated entropies obtained will closely correspond with the values of the true underlying distribution. In Appendix C, we work out these conditions in one dimension where the underlying distribution is a one-dimensional Gaussian, which may be sampled by either top-hat filters, Gaussian filters, or Lorentzian filters.

Overall, one may show that for the discrete approximation to the continuous entropy of an underlying Gaussian distribution of standard deviation σ_g will overestimate the entropy by less than 0.1 bits for $\Delta\omega \leqslant \{1.336\sigma_g, 0.967\sigma_g, 0.100\sigma_g\}$, for the top-hat, optimum Gaussian, and optimum Lorentzian filters, respectively (see Appendix C for more thorough discussion). As a rule of thumb, we may say the top-hat and Gaussian filters are adequate when the spacing is narrower than $0.95\sigma_g$, but for the optimum lorentzian filters, one wants an order-of-magnitude smaller spacing of about $0.1\sigma_g$.

The most dramatic advantage in using tunable frequency filters over other frequency sampling techniques (not counting higher resolution) is that one can perform a continuous scan of an arbitrary curve in the joint frequency space to trace out lines over which $\omega_A + \omega_B$ is a constant, forming a direct measurement of the probability distribution of $\omega_A + \omega_B$ whose entropy may be calculated and used to more sensitively witness energy-time entanglement by violating Eq. (1) instead of Eq. (2). When the correlations are sufficiently strong that we can violate the entanglement witness by a nonnegligible margin, we may use the methods in Sec. III C to compensate for the intrinsic variability in the filter function scanned at different frequencies, due to technical, electronic, and mechanical noise in the measurement process.

V. DISCUSSION: RELATIVE NECESSITY OF DIFFERENT-CASE SAMPLING APPROACHES

At this point, we have discussed three classes of methods of increasing complexity to discretely sample continuous observables to conservatively bound their continuous entropies and in so doing, conservatively witness continuous-variable entanglement. At sufficiently high sensitivity, all experimental measurements are subject to fluctuations, so it may seem like the strongest, most complex and resource-intensive approach is always necessary to demonstrate energy-time entanglement, but this need not be the case in practical situations.

The joint frequency probability distributions obtained from counting the number of biphotons passing through a pair of frequency filters as a function of those filters' centers is subject to shot-noise (Poisson-distributed) uncertainty in the number of detected biphotons. Where these coincidence counting rates fluctuate, so do the corresponding conservative estimates of probabilities, entropies, correlation, and entanglement. As is done in the estimation of other physical parameters, we need not consider every conceivable influence that contributes to the experimental data if that effect can be shown to be sufficiently small.

Although the most conservative acquisition strategy (in which we measure all of the filter functions at all center frequencies to gauge their relative uniformity and apply a correction factor to the measured entropy) is valid in all situations, it also becomes significantly resource-intensive at high measurement resolution over a broad range. However, if the effect to the measured entropies from the variation/jitter of these filter functions cannot be resolved from the competing effect of shot noise, then the statistics we obtain will essentially be unperturbed, and we may opt for the simpler approach.

VI. CONCLUSION

Thus far, there has been no direct demonstration of energy-time entanglement through energy and time correlations except for extremely narrowband sources where techniques exist to scan the limited bandwidth directly at high resolution. Until now, the less standard techniques required for high-resolution frequency measurement made rigorous demonstrations of energy-time entanglement a puzzling endeavor. We have shown how these kinds of measurements may still be used to rigorously demonstrate entanglement, and explored the technological landscape required to achieve it experimentally. In demonstrating how both to directly and conservatively sample these correlations, we make it possible to utilize energy-time entanglement without making extra assumptions about the system or measurement device that might otherwise compromise their utility in quantum networking.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the Air Force Office of Scientific Research as well as insightful comments from Dr. Dylan Heberle and Dr. David Hucul. The views expressed are those of the authors and do not reflect the official guidance or position of the United States Government, the Department of Defense, or of the United States Air Force. The appearance of external hyperlinks does not constitute endorse-

ment by the United States Department of Defense (DoD) of the linked websites or of the information, products, or services contained therein. The DoD does not exercise any editorial, security, or other control over the information you may find at these locations.

APPENDIX A: PROOF THAT COARSE-GRAINING CANNOT INCREASE RELATIVE ENTROPY

To illustrate why continuous majorization is useful, we present an expanded version of the proof in the Supplemental Material of Ref. [1] that majorization cannot increase the relative entropy between two probability distributions.

Consider two discrete probability distributions P_1 and Q_1 each with n possible outcomes. The entropy of Q_1 relative to P_1, denoted $\mathscr{D}(P_1||Q_1)$ is given by

$$\mathscr{D}(P_1||Q_1) \equiv \sum_i P_{1i} \log\left(\frac{P_{1i}}{Q_{1i}}\right). \tag{A1}$$

Here and throughout the paper, we assume logarithms to be base 2, since we measure entropy in bits.

Let us assume two new probability distributions P_2 and Q_2 obtained from P_1 and Q_1 through the same permutation operation. Since the relative entropy is invariant under simultaneous permutations of outcomes, we have:

$$\mathscr{D}(P_1||Q_1) = \mathscr{D}(P_2||Q_2). \tag{A2}$$

The relative entropy is a jointly convex function [8]. Because of this, the majorization operation that comes from mixing permutations of a probability distribution cannot increase relative entropy:

$$\begin{aligned}\mathscr{D}(\lambda P_1 + (1-\lambda)P_2||\lambda Q_1 + (1-\lambda)Q_2) \\ \leqslant \lambda\mathscr{D}(P_1||Q_1) + (1-\lambda)\mathscr{D}(P_2||Q_2) \\ \leqslant \min\{\mathscr{D}(P_1||Q_1), \mathscr{D}(P_2||Q_2)\}.\end{aligned} \tag{A3}$$

Note that multiple forms of entropy can be expressed in terms of relative entropy:

$$\mathscr{D}(P(X_A)||P(U_A)) = \log(n_A) - H(X_A), \tag{A4}$$

$$\mathscr{D}(P(X_A, X_B)||P(U_A)P(U_B)) = \log(n_A n_B) - H(X_A, X_B), \tag{A5}$$

$$\mathscr{D}(P(X_A, X_B)||P(U_A)P(X_B)) = \log(n_A) - H(X_A|X_B), \tag{A6}$$

$$\mathscr{D}(P(X_A, X_B)||P(X_A)P(X_B)) = H(X_A : X_B) \tag{A7}$$

Here, $P(U_A)$ and $P(U_B)$ are the uniform probability distributions over the spaces defining the variables X_A and X_B, respectively. To prove that majorization cannot increase relative entropy, we use repeated applications of its joint convexity.

Let us define the mixed distributions $\bar{P} = \lambda P_1 + (1-\lambda)P_2$, and $\bar{Q}$ similarly for $\lambda \in [0, 1]$. By use of the joint convexity of relative entropy, we may show the following:

$$\begin{aligned}\mathscr{D}(P(X,Y)||P(X)P(Y)) &\geqslant \mathscr{D}(\bar{P}(X,Y)||P(X)P(Y)) \\ &\geqslant \mathscr{D}(\bar{P}(X,Y)||\bar{P}(X)P(Y)) \\ &\geqslant \mathscr{D}(\bar{P}(X,Y)||\bar{P}(X)\bar{P}(Y)),\end{aligned} \tag{A8}$$

which proves that majorization cannot increase relative entropy. Consequently, we have that majorization cannot decrease joint entropy $H(X_A, X_B)$, marginal entropy $H(X_A)$, or conditional entropy $H(X_A|X_B)$, and cannot increase mutual information $H(X_A : X_B)$ or even conditional mutual information $H(X_A : X_B|X_C)$, which can also be expressed in terms of relative entropy.

APPENDIX B: DERIVATION THAT FURTHER MIXING PRESERVES MAJORIZATION RELATIONSHIP

Let us assume two probability vectors $\vec{p}$ and $\vec{q}$ of equal dimension and that $\vec{p} \succ \vec{q}$. Furthermore, let us assume two other probability vectors $\vec{p}'$ and $\vec{q}'$ related to $\vec{p}$ and $\vec{q}$ through the same doubly-stochastic matrix $\mathbf{M}$, so that $\vec{p}' = \mathbf{M}\vec{p}$ and $\vec{q}' = \mathbf{M}\vec{q}$. In this Appendix, we shall prove that $\vec{p} \succ \vec{q}$ implies $\vec{p}' \succ \vec{q}'$.

Because $\vec{p} \succ \vec{q}$, there exists another doubly-stochastic matrix $\mathbf{D}$ such that $\vec{q} = \mathbf{D}\vec{p}$.

The Birkhoff-von Neumann theorem [17] states that doubly-stochastic matrices can be expressed as a weighted average of permutation matrices $\mathbf{P}_i$. So we have

$$\mathbf{D} = \sum_i v_i \mathbf{P}_i, \tag{B1a}$$

$$\mathbf{M} = \sum_i w_i \mathbf{P}_i. \tag{B1b}$$

Then we can express $\vec{q}'$ in terms of $\vec{p}$ and these permutations:

$$\begin{aligned}\vec{q}' &= \mathbf{M}\vec{q} = \mathbf{M}\mathbf{D}\vec{p} \\ &= \sum_i w_i \mathbf{P}_i \mathbf{D}\vec{p} = \sum_{i,j} w_i v_j \mathbf{P}_i \mathbf{P}_j \vec{p}.\end{aligned} \tag{B2}$$

For any two matrices (let alone permutation matrices) $\mathbf{P}_i$ and $\mathbf{P}_j$, there exist other matrices $\mathbf{L}_j$ such that $\mathbf{P}_i\mathbf{P}_j = \mathbf{L}_j\mathbf{P}_i$. Applying this to our expression for $\vec{q}'$, we find

$$\begin{aligned}\vec{q}' &= \sum_{i,j} w_i v_j \mathbf{L}_j \mathbf{P}_i \vec{p} = \sum_j v_j \mathbf{L}_j \sum_i w_i \mathbf{P}_i \vec{p} \\ &= \sum_j v_j \mathbf{L}_j \mathbf{M}\vec{p} = \sum_j v_j \mathbf{L}_j \vec{p}'.\end{aligned} \tag{B3}$$

Now, where $\mathbf{P}_i\mathbf{P}_j = \mathbf{L}_j\mathbf{P}_i$, we can solve for $\mathbf{L}_i$ to find $\mathbf{L}_j = \mathbf{P}_i\mathbf{P}_j\mathbf{P}_i^{-1}$. Since the product and inverse of permutation matrices are also permutation matrices, $\mathbf{L}_j$ are permutation matrices, which proves that $\vec{q}'$ is a weighted average of permutations of $\vec{p}'$ (i.e., is related by a doubly-stochastic matrix), and consequently, that $\vec{p}' \succ \vec{q}'$■.

APPENDIX C: COMPARISON OF SENSITIVITY OF DIFFERENT TECHNIQUES

1. Witnessable entanglement of double-Gaussian wave function

Given a double-Gaussian joint amplitude:

$$\begin{aligned}\psi(\omega_1, \omega_2) = \frac{1}{\sqrt{2\pi\sigma_+\sigma_-}} \mathrm{Exp}\left[-\frac{(\omega_1+\omega_2)^2}{8\sigma_+^2}\right] \\ \times \mathrm{Exp}\left[-\frac{(\omega_1-\omega_2)^2}{8\sigma_-^2}\right].\end{aligned} \tag{C1}$$

One may show that the degree of violation of Walborn's separability inequality is readily expressible in terms of the correlation ratio R:

$$\log(2\pi e) - h(\omega_1 + \omega_2) - h(t_1 - t_2) \tag{C2}$$

$$= \log\left(\frac{\sigma_+}{\sigma_-}\right) = \log(R). \tag{C3}$$

The degree of violation of Walborn's EPR steering inequality may similarly be simplified:

$$\log(\pi e) - h(\omega_1|\omega_2) - h(t_1|t_2) \tag{C4}$$

$$= \log\left(\frac{1}{2}\left(\frac{\sigma_+}{\sigma_-} + \frac{\sigma_-}{\sigma_+}\right)\right) = \log\left(\frac{1}{2}\left(R + \frac{1}{R}\right)\right). \tag{C5}$$

For both varieties of inequalities, any nonzero amount of correlation ($R > 0$) for this state will witness entanglement, though the amount of violation for the EPR-steering inequality is less than that for the separability inequality, approaching a constant difference of one bit.

2. Illustrating sensitivity of witness as a function of sampling resolution with different filters

In what follows, we will show under what conditions sampling with different kinds of filters will adequately characterize the entropy of a Gaussian distribution. We take this as a rule of thumb for understanding the resolutions required for joint frequency measurements of a double-Gaussian joint amplitude.

For reference, a Gaussian filter of full-width at half-maximum F_G will be adequately sampled for spacing $\Delta\omega \leqslant F_G\sqrt{\pi/(4\log(2))}$, which is about $1.06F_G$. For a Lorentzian filter of full-width at half-maximum F_L, adequate sampling with this filter occurs when $\Delta\omega \leqslant \pi F_L/2$, which is about $1.57F_L$.

a. Top-hat filter scanning of Gaussian distribution

To illustrate the resolution required for these filters to be practical in entanglement witnessing, we consider the simplified, one-dimensional case of characterizing the entropy of a Gaussian distribution $\rho(\omega)$ with unit variance using top-hat filters $\bar{\Pi}(\omega - \omega_i)$ of width $\Delta\omega$ (and bar overhead to note that we will be using nonnormalized filters of height unity). With this level of simplicity, we may obtain analytic results to inform our intuition.

First, the discrete probabilities are obtained by the following integral:

$$P(\Pi_n) = \int d\omega \rho(\omega)\bar{\Pi}(\omega - \omega_n), \tag{C6}$$

where the top-hat function is symmetric, this integral has the form of a convolution:

$$P(\Pi_n) = (\rho * \bar{\Pi})(\omega_n). \tag{C7}$$

Thus, we can obtain these discrete probabilities by sampling the convolution $(\rho * \bar{\Pi})(\omega_n)$ at regular intervals with spacing $\Delta\omega$. Given the functional forms

$$\rho(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{Exp}\left[-\frac{\omega^2}{2\sigma^2}\right]:$$

$$\bar{\Pi}(\omega) = \begin{cases} 1 & -\frac{\Delta\omega}{2} \leqslant \omega \leqslant \frac{\Delta\omega}{2} \\ 0 & \text{otherwise} \end{cases}, \tag{C8}$$

we may find

$$(\rho * \bar{\Pi})(\omega_n) = \frac{1}{2}\left(\mathrm{Erf}\left[\frac{\omega_n + \Delta\omega/2}{\sqrt{2}\sigma}\right] - \mathrm{Erf}\left[\frac{\omega_n - \Delta\omega/2}{\sqrt{2}\sigma}\right]\right), \tag{C9}$$

where $\sigma = 1$, and

$$\mathrm{Erf}(\omega) = \frac{2}{\sqrt{\pi}}\int_0^\omega dx\, e^{-x^2} \tag{C10}$$

is the error function. These probabilities extend over an infinite range, but decay rapidly enough that we may select a large range of integration to make the discrete probability distribution approximately finite-dimensional.

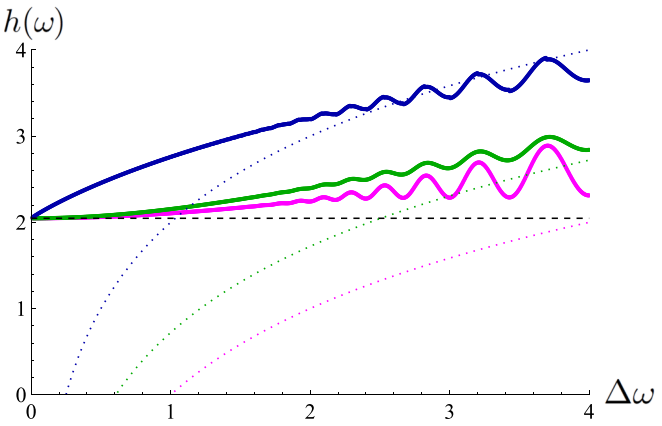


FIG. 3. Plot comparing entropy of a Gaussian with unit standard deviation (black dashed line) with: entropy of top-hat binning approximation (lower oscillating Magenta curve); entropy of Gaussian filter discrete sampling of minimum majorizing width (middle oscillating green curve); and entropy of Lorentzian filter discrete sampling of minimum majorizing width (top oscillating blue curve). The bottom dotted magenta curve corresponds to the entropy of a single top-hat distribution of width $\Delta\omega$, while the middle dashed green curve corresponds to the entropy of the majorizing Gaussian filter of minimum width relative to sampling spacing $\Delta\omega$, and the top dotted blue curve corresponds to the entropy of the Lorentzian filter of minimum width relative to spacing $\Delta\omega$.

In Fig. 3, we have plotted the discrete approximation to the continuous entropy $h_\Pi(\omega)$ (oscillating magenta curve) as would be obtained from the binned probability density in Eq. (6). For a Gaussian of unit standard deviation, we see that when the bin size $\Delta\omega$ is less than unity, the discrete approximation is close to the exact value of the entropy for the unit Gaussian of $(1/2)\log_2(2\pi e) \approx 2.0471$ bits (given by the black dashed line). In this plot, we set the range of ω of $(\rho * \bar{\Pi})(\omega)$ to be between -24 and 24 (i.e., six times the maximum value of $\Delta\omega$ in each direction). This is a great deal wider than what would be needed to capture the significant

values of the unit Gaussian, even when convolved with a top-hat function of maximum width $\Delta\omega = 4$. The discrete probability distribution is made from values of ω equally spaced into bins of size $\Delta\omega$. The number of bins is equal to the range (here, 60) divided by the bin width $\delta\omega$, and rounded down to the nearest integer. This step becomes important for larger values.

With the array of values of $(\rho * \bar{\Pi})(\omega)$ obtained over the prescribed range, we normalize to obtain the discrete probability distribution, calculate the Shannon entropy of this discrete probability distribution, and add to it $\log_2(\Delta\omega)$ to set it equal to the continuous entropy of the discrete approximation (oscillating magenta line). In this plot, we see oscillation in the entropy occurring and becoming more pronounced for larger values of $\Delta\omega$. This is because for relatively large values of $\Delta\omega$, this interval begins to coincide with the range of significant values of $(\rho * \bar{\Pi})(\omega)$ so that alternately one or two sampled points intersect this range as $\Delta\omega$ is swept over different values.

Each peak/valley in descending order of $\Delta\omega$ corresponds to a bin width commensurate with an increasing (integer) number of bins (i.e., $\Delta\omega \to \{48/12, 48/13, 48/14, \ldots\}$). The entropy minima correspond to a spacing such that the central bin is centered on the peak of $(\rho * \bar{\Pi})(\omega)$, while the entropy maxima correspond to spacings such that there is a pair of central bins whose centroid aligns with the peak of $(\rho * \bar{\Pi})(\omega)$. In either case, this approximate continuous entropy is bounded from below both by the entropy of the true underlying distribution (black dashed line) and the entropy of the top-hat filter function of width $\Delta\omega$, normalized as a probability distribution [dotted magenta line corresponding to $\log_2(\Delta\omega)$].

b. Gaussian filter scanning of Gaussian distribution

Just as the discrete probabilities obtained by scanning a top-hat filter over the unit Gaussian may be obtained analytically, we may find a similarly analytic set of probabilities when scanning with a Gaussian filter of width w. To obtain the highest quality data possible, we consider the minimum-width Gaussian filter such that if taken as a probability density, then it would still be majorized by a top-hat distribution of width $\Delta\omega$:

$$\frac{1}{\sqrt{2\pi w^2}} = \frac{1}{\Delta\omega}. \tag{C11}$$

This gives us the un-normalized filter function $\bar{f}_g(\omega)$:

$$\bar{f}_g(\omega) = \mathrm{Exp}\left[-\pi \frac{\omega^2}{\Delta\omega^2}\right]. \tag{C12}$$

The resulting convolution is then

$$(\rho * \bar{f}_g)(\omega_n) = \frac{\Delta\omega}{\sqrt{2\pi\sigma^2 + \Delta\omega^2}} \mathrm{Exp}\left[-\frac{\pi\omega_n^2}{2\pi\sigma^2 + \Delta\omega^2}\right]. \tag{C13}$$

The resulting approximate continuous entropy is given by the oscillating green curve in Fig. 3, where the theoretical minimum entropy of the Gaussian filter function is given by the corresponding green dotted curve. Unlike the case for simple top-hat coarse graining, the theoretical minimum entropy of the Gaussian filter function is not a minimal value to the discretely approximated entropy with Gaussian filters because the latter will converge to that of a top-hat function of appropriate width, which majorizes the corresponding Gaussian filter.

c. Lorentzian filter scanning of Gaussian distribution

As it well-describes many transmission linewidths, one can often use the lorentzian filter function $\bar{f}_L(\omega)$ (of minimum majorized width) when sampling frequencies:

$$\bar{f}_L(\omega) = \frac{1}{1 + \frac{\pi^2\omega^2}{\Delta\omega^2}}. \tag{C14}$$

The resulting convolution for this is a Voigt function:

$$(\rho * \bar{f}_L)(\omega_n) = \frac{\Delta\omega}{\sqrt{2\pi\sigma^2}} \mathrm{Re}\left[e^{-z^2(\omega_n)}(1 - \mathrm{Erf}(-iz(\omega_n)))\right] :$$
$$z(\omega_n) \equiv \frac{\pi\omega_n + i\Delta\omega}{\sqrt{2}\pi\sigma}, \tag{C15}$$

or more simply:

$$(\rho * \bar{f}_L)(\omega_n) = \frac{1}{2}\sqrt{\frac{\pi}{\tau}} U(x, \tau) : \quad x(\omega_n) \equiv \frac{\omega_n}{\Delta\omega} :$$
$$\tau(\omega_n) \equiv \frac{\pi^2\sigma^2}{2\Delta\omega^2}, \tag{C16}$$

where $U(x, \tau)$ is the standard Voigt function. For small values of τ, this approaches a Lorentzian function, though for all values of τ, this profile decays slowly like a Lorentzian. For $\sigma = 1$, and $\Delta\omega = 1$, integrating this profile from -5 to 5 gives a value of 96.2%, while for $\Delta\omega = 0.1$, we obtain 99.6%. Unlike with the rapid decay of the Gaussian, the effect of truncating the sampling distribution using Lorentzian filters will produce significant effects to the entropy estimation at all filter widths (illustrated by how the blue oscillating curve for the discrete approximation is not clearly bounded from above or below by the entropy of the Lorentzian filter (blue dotted curve).

[1] J. Schneeloch, C. C. Tison, M. L. Fanto, P. M. Alsing, and G. A. Howland, Quantifying entanglement in a 68-billion-dimensional quantum state space, Nat. Commun. **10**, 2785 (2019).

[2] Y. Mei, Y. Zhou, S. Zhang, J. Li, K. Liao, H. Yan, S.-L. Zhu, and S. Du, Einstein-Podolsky-Rosen energy-time entanglement of narrow-band biphotons, Phys. Rev. Lett. **124**, 010509 (2020).

[3] J. Schneeloch and G. A. Howland, Quantifying high-dimensional entanglement with Einstein-Podolsky-Rosen correlations, Phys. Rev. A **97**, 042338 (2018).

[4] L. K. Shalm, D. R. Hamel, Z. Yan, C. Simon, K. J. Resch, and T. Jennewein, Three-photon energy–time entanglement, Nat. Phys. **9**, 19 (2013).

[5] M. Liscidini and J. E. Sipe, Stimulated emission tomography, Phys. Rev. Lett. **111**, 193602 (2013).
[6] J.-P. W. MacLean, J. M. Donohue, and K. J. Resch, Direct characterization of ultrafast energy-time entangled photon pairs, Phys. Rev. Lett. **120**, 053601 (2018).
[7] S. P. Walborn, B. G. Taketani, A. Salles, F. Toscano, and R. L. de Matos Filho, Entropic entanglement criteria for continuous variables, Phys. Rev. Lett. **103**, 160505 (2009).
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley and Sons, New York, NY, 2006).
[9] S. P. Walborn, A. Salles, R. M. Gomes, F. Toscano, and P. H. S. Ribeiro, Revealing hidden Einstein-Podolsky-Rosen nonlocality, Phys. Rev. Lett. **106**, 130402 (2011).
[10] I. Białynicki-Birula and J. Mycielski, Uncertainty relations for information entropy in wave mechanics, Commun. Math. Phys. **44**, 129 (1975).
[11] 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).
[12] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935).
[13] J. Schneeloch, P. B. Dixon, G. A. Howland, C. J. Broadbent, and J. C. Howell, Violation of continuous-variable Einstein-Podolsky-Rosen steering with discrete measurements, Phys. Rev. Lett. **110**, 130407 (2013).
[14] J. Schneeloch, S. H. Knarr, G. A. Howland, and J. C. Howell, Demonstrating continuous variable Einstein–Podolsky–Rosen steering in spite of finite experimental capabilities using Fano steering bounds, J. Opt. Soc. Am. B **32**, A8 (2015).
[15] F. Toscano, D. S. Tasca, L. Rudnicki, and S. P. Walborn, Uncertainty relations for coarse–grained measurements: An overview, Entropy **20**, 454 (2018).
[16] J. Schneeloch, C. C. Tison, H. S. Jacinto, and P. M. Alsing, Negativity vs. purity and entropy in witnessing entanglement, Sci. Rep. **13**, 4601 (2023).
[17] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics (Springer, New York, NY, 2010).
[18] M. A. Nielsen and J. Kempe, Separable states are more disordered globally than locally, Phys. Rev. Lett. **86**, 5184 (2001).
[19] J. Schneeloch and J. C. Howell, Introduction to the transverse spatial correlations in spontaneous parametric down-conversion through the biphoton birth zone, J. Opt. **18**, 053501 (2016).
[20] K. Kato and E. Takaoka, Sellmeier and thermo-optic dispersion formulas for KTP, Appl. Opt. **41**, 5040 (2002).
[21] B. P. Ramsay, O. T. Koppius, and E. L. Cleveland, The prism and the theory of optical resolution, J. Opt. Soc. Am. **30**, 439 (1940).
[22] D. J. Starling, S. M. Bloch, P. K. Vudyasetu, J. S. Choi, B. Little, and J. C. Howell, Double Lorentzian atomic prism, Phys. Rev. A **86**, 023826 (2012).
[23] M. Shirasaki, Virtually imaged phased array as a wavelength demultiplexer (U.S. Patent 5 999 320, Dec. 7, 1999), https://patents.justia.com/patent/5999320#history.
[24] S. Xiao, A. M. Weiner, and C. Lin, A dispersion law for virtually imaged phased-array spectral dispersers based on paraxial wave theory, IEEE J. Quantum Electron. **40**, 420 (2004).
[25] S. Xiao and A. M. Weiner, 2-D wavelength demultiplexer with potential for $\geqslant$1000 channels in the C-band, Opt. Express **12**, 2895 (2004).
[26] X. Zhu, D. Lin, Z. Hao, L. Wang, and J. He, A VIPA spectrograph with ultra-high resolution and wavelength calibration for astronomical applications, Astron. J. **160**, 135 (2020).