Optics Letters

Efficient compressive and Bayesian characterization of biphoton frequency spectra

E. M. Simmerman,[1] H.-H. Lu,[2] A. M. Weiner,[2] and J. M. Lukens[1,*]

[1]Quantum Information Science Group, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
[2]School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA
*Corresponding author: lukensjm@ornl.gov

Received 13 March 2020; revised 16 April 2020; accepted 21 April 2020; posted 21 April 2020 (Doc. ID 392694); published 15 May 2020

Frequency-bin qudits constitute a promising tool for quantum information processing, but their high dimensionality can make for tedious characterization measurements. Here we introduce and compare compressive sensing and Bayesian mean estimation for recovering the spectral correlations of entangled photon pairs. Using a conventional compressive sensing algorithm, we reconstruct joint spectra with up to a 26-fold reduction in measurement time compared to the equivalent raster scan. Applying a custom Bayesian model to the same data, we then additionally realize reliable and consistent quantification of uncertainty. These efficient methods of biphoton characterization should advance our ability to use the high degree of parallelism and complexity afforded by frequency-bin encoding. © 2020 Optical Society of America

https://doi.org/10.1364/OL.392694

Given the sparsity of many signals of interest in the real world, compressive sensing (CS) [1] has emerged as a powerful, general technique for reconstructing a signal from significantly fewer measurements than required by traditional sampling methods. In the context of quantum information, CS has been utilized for quantum state reconstruction [2,3], process tomography [4], and ghost imaging [5,6], and has proven to be an effective tool for efficient characterization of high-dimensional quantum states [7–9], especially in the spatial degree of freedom (DoF). Bayesian mean estimation (BME) is more general, in the sense that neither system sparsity nor a particular class of measurements are required for the procedure's validity, though this knowledge can nonetheless be neatly accounted for in the prior distribution. This is an advantage of Bayesian methods, along with return of confidence intervals commensurate with the data gathered [10]. BME has been explored experimentally, in single- [11] and two-qubit polarization quantum state tomography [12,13], and identification of low-light-level sources [14].

Recently, the frequency DoF has developed into a promising platform for photonic quantum information processing (QIP) [15–17], yet despite the naturally high-dimensional nature of this Hilbert space, CS techniques have yet to be leveraged to characterize spectral properties and, while Bayesian methods have been applied to recover density matrices [18] and mode transformations [19] in frequency-bin QIP, they have not been explored for extracting high-dimensional biphoton frequency correlations. In this work, we retrieve the frequency correlations of quantum states with CS and BME for the first time.

We consider biphotons generated by spontaneous parametric downconversion, which can exhibit strong frequency anticorrelations and are a common source of information carriers in frequency-domain QIP. Biphoton correlations can be assessed by measuring the joint spectral intensity (JSI). The conventional raster scan method of obtaining the JSI for an N-dimensional Hilbert space ($\sqrt{N}$ dimensions per photon) requires N coincidence measurements between two narrow spectral passbands, one for each photon. One way to improve on this approach is to leverage dispersion in a time-of-flight spectrometer, converting from spectral to temporal correlations and using time-tagging to map out the JSI [8,20–22]. However, this requires a timing reference to determine absolute—not just relative—frequency, a condition which is not satisfied by, e.g., a free-running continuous-wave–pumped biphoton source. Our approach, valid for asynchronous sources, is to perform measurements over many passbands at once. We use Fourier-transform pulse shaping [23] to apply a code of spectral filters to the frequency bins of each photon, with transmission values taken from random binary, random gray-level, or Hadamard codes.

In particular, the use of Hadamard codes is well established in classical spectroscopy, where measuring linear combinations of frequency bands can enable significantly higher signal-to-noise ratios (SNRs) in background-limited environments [24]—features which have been explored in measuring correlations of frequency-entangled photons as well [25], though without recovering the underlying probability distribution. Likewise, the CS and Bayesian methods we consider here enjoy an SNR improvement from measuring many bins at once. CS goes one step further, reducing the total number of measurements needed by exploiting the sparsity anticipated for our biphoton system. BME takes a slightly different perspective on the problem: whereas CS seeks to exceed a threshold of measurements needed to find the underlying signal, BME instead is formulated to return a credible estimate given any collection of measurements.

The experimental setup is depicted in Fig. 1(a). Broadband biphotons were generated by pumping a periodically poled lithium niobate (PPLN) ridge waveguide (AdvR) with a continuous-wave 780 nm laser under type-0 configuration.

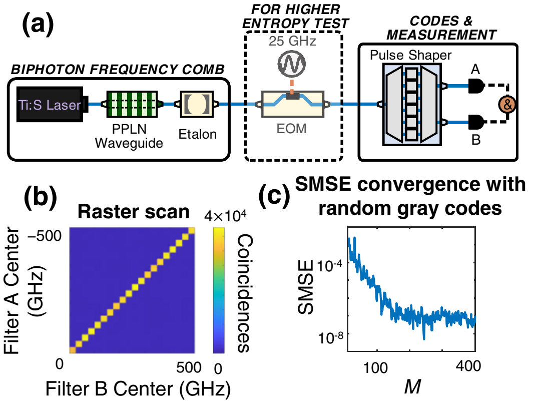

Fig. 1. (a) Experimental setup. (b) Raster scan of lower entropy state. (c) SMSE number-of-measurement convergence of LASSO reconstructions using random gray value codes.

A 25 GHz spaced etalon was used to produce a biphoton frequency comb. We used a pulse shaper (Finisar) to selectively attenuate 20 frequency bins on either side of the center frequency and block all others, giving JSI dimensionality $N = 400$ (limited in general not by the reconstruction methods but by the phase-matching bandwidth of the PPLN and the acceptance bandwidth and resolution of the pulse shaper). Signal and idler bins were sent to separate superconducting nanowire single-photon detectors (Quantum Opus) and a time tagger (PicoQuant) ascertained coincidences. We compare reconstructed JSIs obtained from length-20 random and Hadamard codes applied to each photon with 0.5 s integration time to raster scans obtained in 5 s, maintaining the same average singles counts (both random and Hadamard codes average a transmissivity of 50% over all bins). We first tested a relatively sparse, highly anticorrelated JSI, and then added an electro-optic phase modulator (EOM) driven with 25 GHz RF signal to produce spectral sidebands.

Given the results of an experiment, conventional CS techniques associate the measured coincidences (y_i) with a linear function of the joint spectral bins passed in measurement setting i (a length-N vector $\mathbf{x}_i$ given by the Kronecker product of the individual codes applied to each photon [7]), weighted by the length-N vector $\boldsymbol{\beta}$. Each element of $\boldsymbol{\beta}$ is defined as the biphoton flux within a particular bin-pair of the JSI. From this, the N-dimensional probability distribution follows as $\mathbf{p} = \frac{\boldsymbol{\beta}}{\sum_i \beta_i}$. For M measurements, $\boldsymbol{\beta}$ can be estimated by solving the linear system $\mathbf{y} = A\boldsymbol{\beta}$, where A is an $M \times N$ matrix with rows $\mathbf{x}_i^T$.

In order to incorporate knowledge of sparsity, we use the CS least absolute shrinkage and selection operator (LASSO) method [26] to find an estimate of $\boldsymbol{\beta}$, by solving

$$\boldsymbol{\beta}_{\mathrm{LASSO}} = \min_{\boldsymbol{\beta}} \left\{ \frac{1}{2M} \sum_{i=1}^{M} \left(y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta} \right)^2 + \lambda \sum_{j=1}^{N} |\beta_j| \right\}, \tag{1}$$

with respect to $\boldsymbol{\beta}$ and β_0 (an intercept). The absolute value accounts for LASSO's domain over all real numbers—although as needed in our case, only positive β_i are returned. The first term favors solutions that minimize error with respect to the measurement results, while the second term enforces sparsity. We used MATLAB's built-in LASSO algorithm to perform the minimization. In order to select a sensible value for the weight parameter λ, 10-fold cross validation was used, which partitions the measurements into training and validation sets and computes the mean squared error (MSE) of each solution against the test data. Multiple λ values were tested, and the largest within one standard error of the MSE-minimizing value was chosen.

We first use LASSO to reconstruct a relatively sparse JSI, the raster scan of which appears in Fig. 1(b). The ratio of diagonal to off-diagonal coincidences is ~40. If nothing is known about the expected JSI, one can assess the number of codes necessary for CS reconstruction by calculating a serial MSE (SMSE) $\frac{1}{N}\sum_{j=1}^{N}(p_j - \bar{p}_j)^2$ between each normalized reconstructed JSI $\mathbf{p}$ (from M codes) and the mean $\bar{\mathbf{p}}$ of the previous several ($M-5$ to $M-1$ in our case), continuing to add measurements until the SMSE stabilizes. SMSE LASSO convergence calculated for $M = 10$ to $M = N$ is shown in Fig. 1(c). This approach highlights how one can determine convergence without explicitly comparing the recovered result to theory.

If one *can* predict the form of the JSI, however—as in the present case—convergence can also be assessed by comparing the reconstruction at various M s to the ideal case with probability vector $\mathbf{q}$, where, because of the broadband nature of our source, all antidiagonal bin-pairs are taken to have equal probability and all others are zero. We use the Bhattacharyya coefficient [27] as a metric of overlap for this purpose, defined as $B_c = \sum_{i=1}^{N}\sqrt{p_i q_i}$. B_c LASSO convergence is shown in Fig. 2(a). A transition occurs around $M = 160$, beyond which additional measurements produce minimal improvement. Representative reconstructed JSIs for Hadamard and random binary codes are shown in Figs. 2(b) and 2(c), respectively. Each Hadamard sequence has "1" in the first column, so that one spectral bin is passed for both signal and idler photons in every measurement, preventing meaningful extraction of information about the corresponding joint frequency bins in the reconstruction. Thus, these bins were removed in the Hadamard reconstructions below, making their JSIs 19×19. (Consequently, conventional Hadamard spectroscopy omits the first row and column of the matrix when defining code sequences [24].)

In order to compare the number of measurements required for convergence with that anticipated from CS theory, we can quantify the JSI sparsity by the effective number of states, $K = 2^{H(\mathbf{p})}$, with $H(\mathbf{p})$ the information entropy $H(\mathbf{p}) = -\sum_{j=1}^{N} p_j \log_2 p_j$. Experimentally, this number falls between that of an ideal uniform anticorrelated JSI ($K = 20$) and the maximum uncorrelated case of $K = N = 400$. The shoulder at $M = 160$ [Fig. 2(a)] corresponds to more measurements than the ~60 expected for a $K = 20$ sparse system, based on the scaling $K \ln \frac{N}{K}$ [1,7]. However, computing the effective dimensionality K from the raster scan itself gives $K = 97$, an increase that can be attributed to the off-diagonal background counts. Interestingly, $K = 97$ predicts a required number of measurements of ~140, much closer to our observations. The 10-fold improvement in the acquisition time (at the same average single-photon flux), combined with the advantage of fewer measurements, results in a 26-fold reduction in the measurement time compared to the raster scan. Each reconstruction took less than 2 min on a personal laptop.

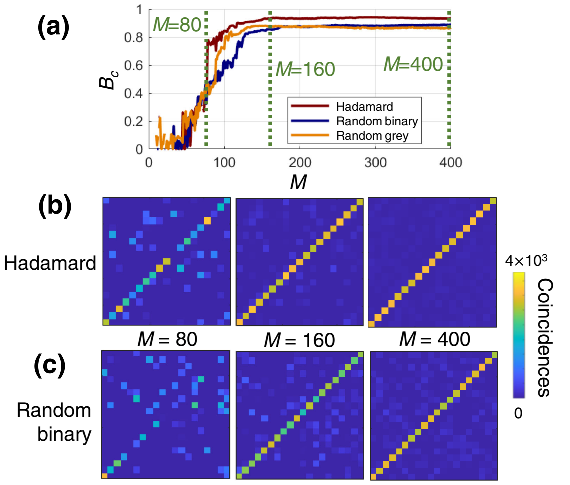

Fig. 2. LASSO reconstruction. (a) Number-of-measurement convergence. Overlap coefficients (B_c) computed with respect to the ideal distribution $\mathbf{q}$. (b)–(c) Representative reconstructions using (b) Hadamard and (c) random binary codes.

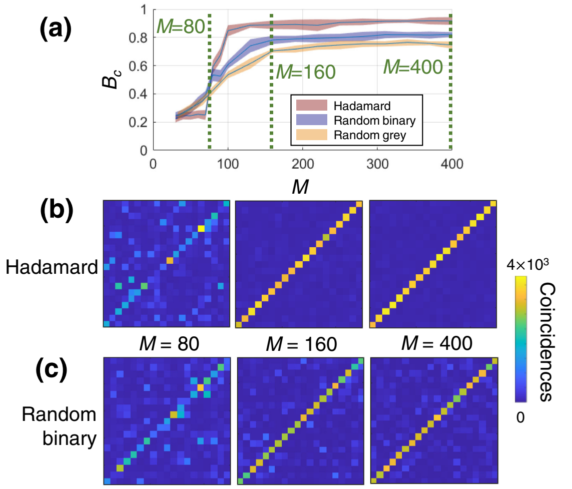

Fig. 3. Bayesian reconstruction. (a) Number-of-measurement convergence. Overlap coefficients (B_c) computed with respect to the ideal distribution $\mathbf{q}$. (b)–(c) Representative reconstructions using (b) Hadamard and (c) random binary codes.

As Eq. (1) corresponds to a convex optimization problem, it admits highly efficient numerical solutions; this ease of computation is an advantage of CS methods, but it comes at the cost of ambiguity in sparsity enforcement (signified by the parameter λ in the case of LASSO). Even when comparing different possible values, as by cross-validation, the final choice of λ is ultimately subjective. BME, on the other hand, "offers one answer to a well-posed problem" [28]. The estimator's initial model of the system (which can be made as uninformed as necessary to reflect actual knowledge) together with the data, uniquely determine the posterior distribution. As long as the probability model can be justified by physical principles and the resulting distribution is adequately sampled, the result of BME is unambiguous.

Specifically, Bayes' theorem gives the posterior probability distribution of parameters representing the state of a system as the normalized product of the likelihood function and prior distribution [28]. The likelihood is the probability of observing the collected data (the coincidence record $\mathbf{y}$) given a possible state of the system, according to some model, while the prior distribution represents initial knowledge of the state of the system. Again, we seek the vector $\boldsymbol{\beta}$ (biphoton fluxes), but we can now infer the underlying probability distribution $\mathbf{p} = \frac{\boldsymbol{\beta}}{\sum_i \beta_i}$ directly, by explicitly introducing a to-be-determined scaling parameter C such that $\boldsymbol{\beta} = C\mathbf{p}$. Then, for the same $\mathbf{x}_i$ codes as in Eq. (1), we can take as our likelihood

$$\mathcal{P}(\mathbf{y}|\mathbf{p}, C) \propto \prod_{i=1}^{N} e^{-C\mathbf{x}_i^T\mathbf{p}}\left(C\mathbf{x}_i^T\mathbf{p}\right)^{y_i}, \tag{2}$$

which models each observation according to a Poisson distribution of mean $C\mathbf{x}_i^T\mathbf{p}$. We assign a prior on C as a normal distribution of mean C_0 and standard deviation $0.1C_0$, where C_0 is set initially by averaging the coincidences (we found this variance sufficient for a uniform prior over all feasible values of C). For $\mathbf{p}$'s prior, we draw from an N-dimensional Dirichlet distribution $\mathrm{Dir}(\alpha)$ to enforce normalization and nonnegativity.

Given the collected data and an appropriate prior distribution $\mathcal{P}(\mathbf{p}, C)$, the BME result for any function $\phi(\mathbf{p}, C)$ is given by the mean over the posterior distribution $\mathcal{P}(\mathbf{p}, C|\mathbf{y}) \propto \mathcal{P}(\mathbf{y}|\mathbf{p}, C)\mathcal{P}(\mathbf{p}, C)$, i.e., $\langle\phi\rangle = \int d\mathbf{p}dC\mathcal{P}(\mathbf{p}, C|\mathbf{y})\phi(\mathbf{p}, C)$. Occasionally such an integral may be solved analytically [13], but more commonly a numerical sampling method is required. Here we use a preconditioned Crank–Nicolson Metropolis–Hastings algorithm (recently introduced in the context of quantum state tomography [29]), a Markov-chain Monte Carlo method, which iteratively generates samples from the posterior distribution, accepting new samples with a probability based on the evaluated likelihood-prior product at each point, and the proposal density function. The posterior distribution is sampled more heavily around local maxima, while allowing for jumps to lower-probability regions, effectively sampling the entire space.

To ensure convergence, we increased the number of samples in the Markov chain until B_c values no longer varied. We found that tuning the α parameter in the prior to favor sparse solutions ($\alpha < 1$) made no considerable difference in reconstructions, confirming one prominent advantage of Bayesian methods—that the form of the prior becomes irrelevant if there is a sufficient amount of data. B_c BME convergence is shown in Fig. 3(a), along with standard deviation error bounds. Reconstructions with $M \sim 160$ took about 12 min each on the same laptop as used previously, considerably longer than LASSO. This is the chief disadvantage of Bayesian methods, making the SMSE approach to number-of-measurement convergence nonviable here. The counteracting advantage, however, is the return of appropriate error bounds for the legitimacy of each reconstruction. For BME, Hadamard codes produced notably better reconstructions than random codes, and number-of-measurement convergence occurred around $M = 160$, as with LASSO, and with similar B_c values at convergence. Thus, BME affords the same ∼26-fold reduction in *experimental* measurement time (though of course with significantly longer numerical analysis). Representative reconstructed JSIs for Hadamard and random binary codes are shown in Figs. 3(b) and 3(c), respectively, for $M = 80, 160$, and 400.

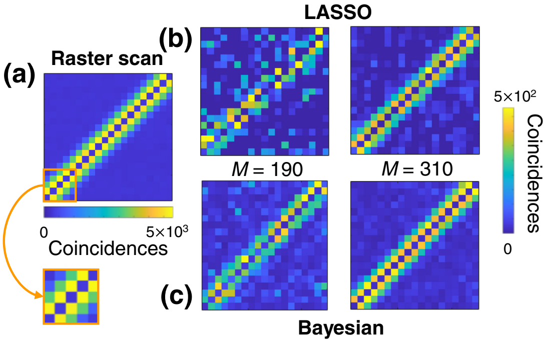

Fig. 4. Higher entropy state. (a) JSI raster scan. (b) LASSO and (c) BME reconstructions using Hadamard codes.

With the introduction of an EOM, more complicated JSIs can be explored as well [30], due to the presence of additional sidebands. Increasing the phase modulation amplitude to split the original diagonal into two main peaks, we measured the JSI with raster scan shown in Fig. 4(a), with a zoom-in highlighting the split interference pattern. At the chosen level of modulation, the effective number of states for an ideal input state is expected to reach $K = 89$, nearly a quarter of the total dimensionality—and, due to off-diagonal background, the number of states computed from the raster scan is even more, at $K = 205$—so the applicability of sparsity-based CS is questionable in this case. Nevertheless, though we did not observe a clear B_c convergence point, we did find reasonable reconstruction with $M \sim 300$ Hadamard codes. Figures 4(b) and 4(c) show representative LASSO and BME reconstructions, respectively. With a more complex JSI, BME visibly outperforms LASSO, especially at reconstructions with fewer measurements (as expected). While clearly noisier than the raster scan, the total measurement time is over 16 times shorter, indicating a practical advantage for rapid measurements in photon-starved environments.

In future work, thresholding [7] could be applied to enhance the contrast in recovered JSIs. Alternatively, because selecting a particular threshold in conventional CS is, like choosing λ, somewhat *ad hoc*, one may incorporate background into the likelihood [Eq. (2)] using a physically motivated model for accidentals—similar to the methods developed in Ref. [19]. Furthermore, it would be interesting to consider the Fourier dual of this experiment—temporal modulation with electro-optic intensity modulators driven with binary sequences—to characterize time-domain correlations, potentially beyond the resolution possible with single-photon detectors [31]. Such tests could be combined with spectral measurements as a witness of time-frequency entanglement. Hence, our Bayesian approach to JSI reconstruction should offer a flexible framework that can be specialized to a variety of situations in biphoton characterization.

Funding. U.S. Department of Energy, Office of Science; Office of Workforce Development for Teachers and Scientists (SULI) program; and Office of Advanced Scientific Computing Research, Early Career Research program.

Acknowledgment. We thank AdvR, Inc., for loaning the PPLN ridge waveguide and P. Lougovski for discussions. This research was performed in part at Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. E. J. Candès and M. B. Wakin, IEEE Signal Process. Mag. **25**, 21 (2008).
2. D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
3. M. Mirhosseini, O. S. Magaña Loaiza, S. M. Hashemi Rafsanjani, and R. W. Boyd, Phys. Rev. Lett. **113**, 090402 (2014).
4. A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Phys. Rev. Lett. **106**, 100401 (2011).
5. O. S. Magana-Loaiza, G. A. Howland, M. Malik, J. C. Howell, and R. W. Boyd, Appl. Phys. Lett. **102**, 231104 (2013).
6. P. A. Morris, R. S. Aspden, J. E. C. Bell, R. W. Boyd, and M. J. Padgett, Nat. Commun. **6**, 5913 (2015).
7. G. A. Howland and J. C. Howell, Phys. Rev. X **3**, 011013 (2013).
8. N. Montaut, O. S. Magaña-Loaiza, T. J. Bartley, V. B. Verma, S. W. Nam, R. P. Mirin, C. Silberhorn, and T. Gerrits, Optica **5**, 1418 (2018).
9. J. Schneeloch, C. C. Tison, M. L. Fanto, P. M. Alsing, and G. A. Howland, Nat. Commun. **10**, 2785 (2019).
10. R. Blume-Kohout, New J. Phys. **12**, 043034 (2010).
11. K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. Houlsby, F. Huszár, and S. P. Kulik, Phys. Rev. A **87**, 062122 (2013).
12. G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Phys. Rev. A **93**, 012103 (2016).
13. B. P. Williams and P. Lougovski, New J. Phys. **19**, 043003 (2017).
14. C. You, M. A. Quiroz-Juárez, A. Lambert, N. Bhusal, C. Dong, A. Perez-Leija, A. Javaid, R. J. León-Montiel, and O. S. Magana-Loaiza, arXiv:191104369 (2019).
15. J. M. Lukens and P. Lougovski, Optica **4**, 8 (2017).
16. M. Kues, C. Reimer, J. M. Lukens, W. J. Munro, A. M. Weiner, D. J. Moss, and R. Morandotti, Nat. Photonics **13**, 170 (2019).
17. H.-H. Lu, A. M. Weiner, P. Lougovski, and J. M. Lukens, IEEE Photon. Technol. Lett. **31**, 1858 (2019).
18. H.-H. Lu, J. M. Lukens, N. A. Peters, B. P. Williams, A. M. Weiner, and P. Lougovski, Optica **5**, 1455 (2018).
19. H.-H. Lu, J. M. Lukens, B. P. Williams, P. Imany, N. A. Peters, A. M. Weiner, and P. Lougovski, npj Quantum Inf. **5**, 24 (2019).
20. C. Chen, C. Bo, M. Y. Niu, F. Xu, Z. Zhang, J. H. Shapiro, and F. N. Wong, Opt. Express **25**, 7300 (2017).
21. A. O. C. Davis, P. M. Saulnier, M. Karpiński, and B. J. Smith, Opt. Express **25**, 12804 (2017).
22. A. O. C. Davis, V. Thiel, M. Karpiński, and B. J. Smith, Phys. Rev. A **98**, 023840 (2018).
23. A. M. Weiner, Rev. Sci. Instrum. **71**, 1929 (2000).
24. E. Nelson and M. Fredman, J. Opt. Soc. Am. **60**, 1664 (1970).
25. J. M. Lukens, A. Dezfooliyan, C. Langrock, M. M. Fejer, D. E. Leaird, and A. M. Weiner, Opt. Lett. **38**, 4652 (2013).
26. R. Tibshirani, J. Roy. Statist. Soc. B **58**, 267 (1996).
27. C. A. Fuchs and J. Van De Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).
28. D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University, 2003).
29. J. M. Lukens, K. J. H. Law, A. Jasra, and P. Lougovski, arXiv:2002.10354 (2020).
30. P. Imany, N. B. Lingaraju, M. S. Alshaykh, D. E. Leaird, and A. M. Weiner, arXiv:1911.04369 (2019).
31. J. M. Lukens, O. D. Odele, D. E. Leaird, and A. M. Weiner, Opt. Lett. **40**, 5331 (2015).

128.46.70.102 {ts '2021-06-18 08:42:44'}