

Boosting photon-number-resolved detection rates of transition-edge sensors by machine learning

ZHENGHAO LI,^{1,4,†}  MATTHEW J. H. KENDALL,^{1,†}  GERARD J. MACHADO,¹  RUIDI ZHU,^{1,2} EWAN MER,¹ HAO ZHAN,^{1,3} AONAN ZHANG,¹  SHANG YU,¹  IAN A. WALMSLEY,¹  AND RAJ B. PATEL^{1,5} 

¹Department of Physics, Imperial College London, Prince Consort Road, London, SW7 2AZ, UK

²Department of Applied Physics, Yale University, 15 Prospect Street, New Haven, Connecticut 06511, USA

³College of Engineering and Applied Sciences, Nanjing University, 163 Xianlin Road, Nanjing 210093, China

⁴zhenghao.li21@imperial.ac.uk

⁵raj.patel1@imperial.ac.uk

[†]These authors contributed equally to this work.

Received 4 December 2024; revised 7 April 2025; accepted 12 May 2025; published 29 May 2025

Transition-edge sensors (TESs) are very effective photon-number-resolving (PNR) detectors that have enabled many photonic quantum technologies. However, their relatively slow thermal recovery time severely limits their operation rate in experimental scenarios compared with leading non-PNR detectors. In this work, we develop an algorithmic approach that enables TESs to detect and accurately classify photon pulses without waiting for a full recovery time between detection events. We propose two machine-learning-based signal processing methods: one supervised learning method and one unsupervised clustering method. By benchmarking against data obtained using coherent states and squeezed states, we show that the methods extend the TES operation rate to 800 kHz, achieving at least a four-fold improvement, whilst maintaining accurate photon-number assignment up to at least five photons. Our algorithms will find utility in applications where high rates of PNR detection are required and in technologies that demand fast active feed-forward of PNR detection outcomes.

Published by Optica Publishing Group under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

<https://doi.org/10.1364/OPTICAQ.555325>

1. INTRODUCTION

Transition-edge sensors (TESs) are photon-number-resolving (PNR) detectors that represent a pinnacle of precision in modern optical detection technology. These superconducting sensors operate at the threshold between superconducting and normal conductive states, leveraging this transition phase to act as a micro-calorimeter that can measure the minute temperature changes induced by photon absorption [1,2]. Their photon-number-resolving capabilities, along with high efficiency, high dynamic range, and low noise characteristics are utilized in a wide array of scientific applications ranging from astrophysics [3–5], particle physics [6,7], and quantum physics. In the field of quantum information science in particular, the availability of PNR detectors has enabled advances in topics such as quantum metrology and sensing [8–11], bio-imaging [12], quantum computing [13–16], quantum communication [17–19], and quantum state preparation [20–25].

A key drawback of the TES technology is its slower operation rate compared with other prevalent detectors such as superconducting nanowire single-photon detectors (SNSPDs). The temperature response of a TES to a photon detection event has both a relatively long rise time, primarily limited

by the inductance of the electric read-out circuit, and subsequently, a long recovery time, which is limited by the weak electron–phonon coupling between the active device material and the substrate. While the TES is still able to produce an electronic response to subsequent photon absorption events during its post-detection recovery times, this takes place at an elevated base temperature due to the un-dissipated heat. The analog voltage signals, for the same photon number, would have different shapes depending on the influence of the preceding signals, thereby creating a much more difficult classification problem in signal processing. The presence of background optical and electronic noise further corrupts the signal shape and complicates the classification problem.

Consequently, methods that classify TES signals by a single attribute, such as pulse height or area [16,26,27], have poor accuracy if the detector is operated above its thermal recovery rate. For example, in a recent experiment that deployed TESs as PNR detectors for a quantum computational advantage demonstration, the detection rate was limited to 375 kHz despite the photon source operating at a 6 MHz repetition rate [16]. Active demultiplexing had to be deployed to reduce the clock rate, which adds additional experimental complexity and hardware

overheads, typically at the cost of photon throughput. For larger scale experiments, it is important to mitigate these effects to reach quantum advantage.

Other single-photon detectors, such as SNSPDs, provide a much higher detection rate but lack intrinsic PNR capability. Although various pseudo-PNR schemes have been developed based on electrical-signal analysis [28–30] or time-/spatial-multiplexing [31–33], their detection efficiencies still scale poorly with photon number compared with intrinsic PNR detectors [33]. As such, TESs remain the best candidate for applications that require accurate multi-photon detection.

Recent fabrication improvements have reduced the TES thermal recovery time by increasing the thermal conductance between the sensor and the substrate using a layer of gold. This pushes the TES operation rate to 1–1.5 MHz, but further speed-up with this approach is traded off against decreasing photon-number resolution [34,35].

Alternatively, further improvement to the detection rate can be attained by better algorithmic photon-number assignment to the noisy and overlapping TES signals at high repetition rates. Signal processing methods have been shown to boost the photon-number-resolving capability of quantum detectors [16,28–30], while machine learning has proven to be a potent tool for time-series classification in many industrial applications [36–38]. In this work, we utilize machine learning to solve the classification problem of detection with TES. We first propose a supervised learning method where a classifier is trained by low-repetition-rate traces and applied to predict the labels of high-repetition-rate signals. We also propose an unsupervised method using a hierarchical density-based clustering algorithm to classify the high-repetition-rate signals directly. We explain how principal component analysis (PCA) can be utilized to reduce the dimensionality of the data.

By benchmarking the methods with the detection of coherent states and squeezed states, we show that appropriate selection of algorithm along with its hyperparameters enables an 800 kHz detection rate with accurate photon-number assignment to at least five photons, which is quadruple the detectors' intrinsic thermal recovery rate and beats the state-of-the-art result in Ref. [16]. We also reconstruct the photon-number distribution of a two-mode squeezed vacuum state at 800 kHz with 99.2% fidelity to the same state at 100 kHz and generate high-fidelity heralded Fock states up to four photons, demonstrating the capability of our methods to support high-rate quantum optical experiments.

2. PNR DETECTION BY TES

We first offer a brief overview of our experimental arrangement for generating and detecting multi-photon states of light. Figure 1(a) shows the overall arrangement, which includes an attenuated laser diode for producing coherent states, and a mode-locked Ti:sapphire laser system and pulse picker for producing two-mode squeezed vacuum (TMSV) states via the parametric downconversion (PDC) process. Both can be produced at different repetition rates. The generated light, which is in the telecom C-band, is coupled into single-mode fiber and fed into a dilution refrigerator with a base temperature <130 mK where the TESs are mounted, as shown in Fig. 1(b).

The active device material in the TES is a superconducting thin film, such as tungsten for the detectors utilized in this

work. The material is temperature-biased near its superconducting critical temperature. A temperature rise, even as small as one caused by the absorption of a single photon, generates a steep and proportionate rise in the material's electrical resistance. This response is amplified by superconducting quantum interference devices (SQUIDs) [Fig. 1(c)] that also operate at cryogenic temperatures, and further amplified by low-noise room-temperature electronics. See Appendix A for further details regarding the experimental setup and the TESs.

The amplified output from the SQUIDs is digitized into a time-series of voltage values, which can be cut into segments that match the pulse repetition rate. Each segment, which is a vector of voltage values, defines a voltage trace corresponding to a single pulse detection event with examples given in Figs. 2(a)–2(c). If adequate time is allowed between two consecutive light pulses, the sensor's temperature and electrical resistance will recover to its baseline value before absorbing the next pulse. As a result, the voltage traces are well-separated in time and show a clear distinction between different bands of traces [Fig. 2(a)]. For our detectors, this requires the light pulses to have a repetition rate of no more than 200 kHz.

2.1. Pulse Filtering

In a final step known as “pulse filtering,” a photon number is assigned to each light pulse by various signal processing methods. One method that shows high photon-number resolution is the “inner product” (IP) method, which filters the voltage traces by their overlap with a reference trace [26]. Each voltage trace, as a vector of voltage values, is denoted $\mathbf{v}_i$. We take the average trace to be a reference, $\bar{\mathbf{v}}$, and calculate the inner product between each trace with this reference trace, $\langle \mathbf{v}_i, \bar{\mathbf{v}} \rangle$. The histogram of these inner product values shows distinct peaks with high extinction ratio [Fig. 2(d)]. The traces can then be appropriately assigned a photon-number label according to the grouping of their inner products.

However, for detection of higher-repetition-rate light pulses, the voltage traces start to overlap in time. The extinction ratio of their inner-product histograms drastically reduces until no clear peaks can be seen, thereby rendering the method inaccurate or unusable [Figs. 2(b)–2(c) and 2(e)–2(f)]. Similar signal processing methods based on filtering the voltage traces by a single attribute, such as trace height or area under the curve [16,27], would encounter the same issue as the traces become increasingly similar at higher repetition rate.

2.2. Principal Component Analysis

Another way to visualize the clustering of TES voltage traces is by PCA [39,40]. The voltage traces are decomposed into a weighted sum of “principal components,” where each component extracts the largest variance from the data while being orthogonal to the previous components.

To extract the principal components from a set of signals, we express the full set of signals as a matrix, $\mathbf{V}$, where each row vector is a voltage trace, $\mathbf{v}_i$. The average trace, $\bar{\mathbf{v}}$, is subtracted from $\mathbf{V}$, and the mean-subtracted voltage traces form a new matrix, $\tilde{\mathbf{V}}$, which has singular value decomposition, $\tilde{\mathbf{V}} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{P}$ and $\mathbf{Q}$ are the matrices of left and right singular vectors and $\mathbf{\Lambda}$ the diagonal matrix of singular values. For PCA, matrix $\mathbf{Q}$ becomes the projection matrix, whose column vectors, $\mathbf{q}_j$, are the principal components. Each voltage trace can now be

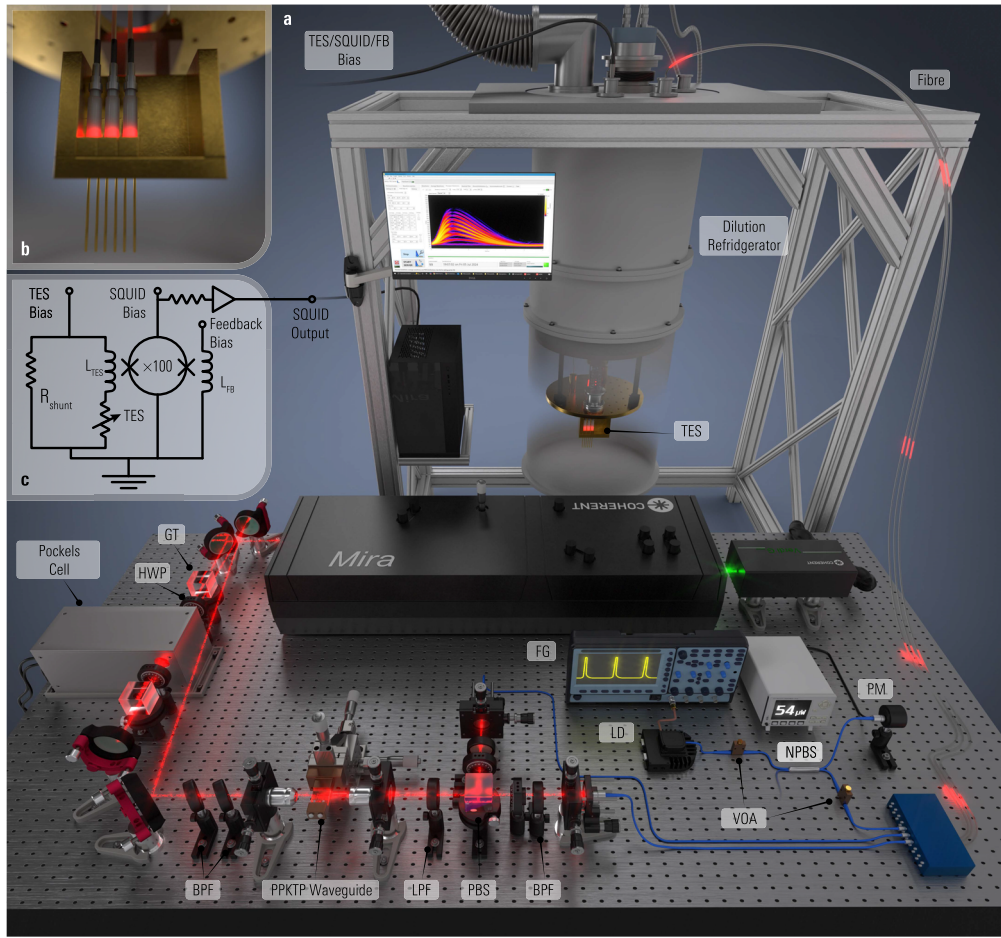


Fig. 1. (a) The quantum light sources and the PNRD apparatus. Coherent states at a desired repetition rate are generated by modulating a 1550 nm laser diode (LD) with a function generator (FG). Variable optical attenuators (VOA) and a 99 : 1 non-polarizing beam splitter (NPBS) split the output of the diode into two where one output is attenuated to the few-photon level and the other is measured on a power meter (PM). For squeezed state generation, a Ti:sapphire laser produces 150-fs-long pulses at 80 MHz which are double-passed through a pulse-picker consisting of a Pockels cell sandwiched between half-wave plates (HWP) and orthogonal Glan–Taylor polarizers (GT). The Pockels cell is driven by amplified radio-frequency signals from a function generator (not shown). The pump pulses are filtered to 780 ± 1 nm (full width at half maximum) using a pair of angle-tuned bandpass filters (BPF). The pulses are coupled into a 10-mm-long ppKTP waveguide that is phase-matched for type-II PDC. The residual pump is filtered using a long-pass filter (LPF) and a polarizing beam splitter (PBS) separates signal and idler pulses along separate paths, each containing a HWP and a BPF with a matched bandwidth to eliminate spectral correlations between signal and idler. Photons from either source are coupled into the TES, via a single-mode fiber, housed in a dilution refrigerator at 130 mK shown in (b). (c) Circuit diagram showing how the TES is coupled to the readout electronics consisting of an array of 100 SQUIDs. See Appendix A for more details.

expressed as a weighted sum of principal components,

$$\mathbf{v}_i = \sum_{j=1}^D F_{ij} \mathbf{q}_j + \bar{\mathbf{v}}, \quad (1)$$

where the weights, F_{ij} , known as “factor scores,” are elements of the matrix $\mathbf{F} = \mathbf{P}\mathbf{\Lambda}$. The summation upper limit, D , is the dimension of the signal, i.e., the length of $\mathbf{v}_i$ [39].

Due to the nature of PCA, the summation in Eq. (1) can be truncated to a low order while maintaining a good approximation to the raw signal [40]. This offers a way to reduce the dimensionality of the data from full voltage traces to points in N -dimensional space, where N is the number of principal components one chooses to preserve.

Figures 2(g)–2(i) and 2(j)–2(l) show an example with $N = 2$, which is already sufficient to distinguish photon-number clusters. At low repetition rate, the first-order component is just the rise and fall of the mean-subtracted voltage response to the current light pulse, and the first-order factor scores alone are enough to classify the traces into distinct clusters. Clusters with smaller first-order factor score values correspond to smaller photon numbers, as these are traces with smaller height. At higher repetition rate, however, the second-order factor scores, representing the weight of the tail of the preceding traces, become more important, though their absolute values are still an order of magnitude smaller than the first-order factor scores. The data still organize into clusters, albeit in a two-dimensional space that includes the second-order component and with much reduced visibility.

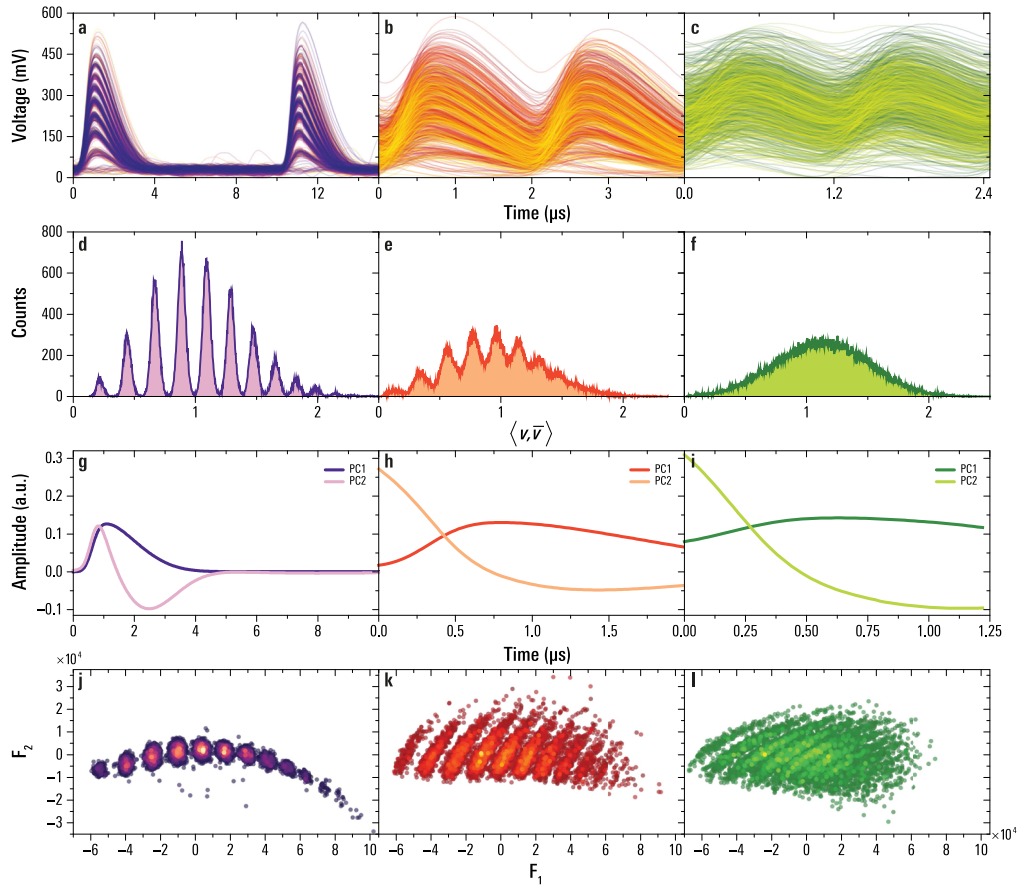


Fig. 2. TES and PCA data for coherent states measured at rates (a,d,g,j) 100 kHz, (b,e,h,k) 500 kHz, and (c,f,i,l) 800 kHz, generated by the light source in Fig. 1 with equal average power per pulse. (a)–(c) Raw SQUID signal traces after amplification by low-noise room temperature electronics. The traces are stacked to visualize their clustering into photon-number groups. (d)–(f) The inner-product histograms that can be used to classify the voltage traces. (g)–(i) The leading two principal components for the mean-subtracted voltage traces. (j)–(l) A density scatter plot of the voltage traces, where coordinates are given by their first two factor scores from PCA. A lighter color represents a higher density of points.

3. PULSE FILTERING BY MACHINE LEARNING

3.1. Supervised Learning Method

For detection of higher-repetition-rate light pulses, we adopt a supervised learning method for pulse filtering. In this method, we first calibrate the detector with well-separated pulses at 100 kHz, whose photon-number labels can be accurately assigned by the IP method. Next, their voltage traces are overlapped to emulate the detector's response to light pulses at the target repetition rate, producing a labeled training dataset for a machine-learning classifier. The trained classifier is then applied to subsequent measurements at the target repetition rate.

We use the k -nearest neighbor (KNN) classifier from the scikit-learn Python package as an example [41,42]. The classifier stores the instances of the training data, and unknown data are classified by a simple majority vote of the K nearest neighbors. We choose K to be five and select nearest neighbors by the Euclidean distance between two vectors.

Figures 3(a) and 3(c) show an example of detecting coherent-state light pulses at 500 kHz and 800 kHz, respectively. The KNN classifier successfully classifies the light pulses into photon-number clusters up to 13 photons for 500 kHz and 12 photons for 800 kHz. The assignment error, which can be visualized by the points that cross over into a neighboring cluster with a

different color, increases at higher rate but remains low. Furthermore, the overall photon-number distribution is close to the reference distribution obtained by running the IP method on light pulses from the same laser diode operated at 100 kHz with equal average power per pulse.

We also tested other supervised classifiers, such as random forest (RF) [42,43], support vector machine (SVM) [42,44,45], gradient-boosted decision tree (GBDT) [46] and a convolutional neural network (CNN) model for time-series classification that consists of two convolutional layers and one global average pooling operation [37,47–49]. With the same training procedure, we found that the KNN classifier has the least processing time with comparable accuracy. For the coherent states at 800 kHz shown in Fig. 3, which recorded 3.82 photons per pulse on average, it took 0.79 s to train on 100,000 training traces and 6.8 s to predict 100,000 unknown traces, averaging at 68 μ s per trace. The training and prediction are performed on entire voltage traces. Further speedup can be obtained by reducing the dimensionality of the data by PCA and by parallelization on batched traces.

The primary limitation of the supervised method is its reliance on high-quality training data. This stipulates low noise on the 100 kHz calibration data and easy switching of the repetition rate from low to high, ideally while maintaining the same mean photon number per pulse.

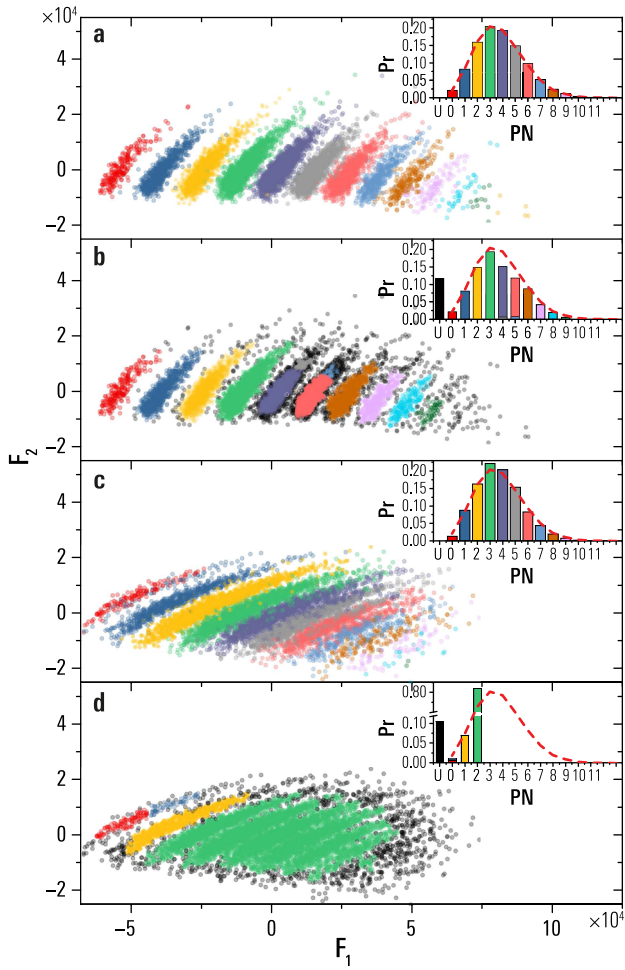


Fig. 3. Classification of coherent states at 500 kHz [which are shown in Figs. 2(b), 2(e), 2(h) and 2(k)] by (a) the KNN algorithm and (b) the HDBSCAN algorithm. A subset of the traces are density-scatter plotted according to their PCA decomposition, and each label assignment is represented by a different color. In KNN, each label corresponds to a unique photon number (PN). In HDBSCAN, sub-clusters might be identified within a photon-number cluster (e.g., purple and grey for PN=4, and pink and light-blue for PN = 5). Unclassified traces are assigned the label “U” and plotted in black. The insets in the panels plot the predicted probability (Pr) of a particular photon-number outcome. The classification of coherent states at 800 kHz [Figs. 2(c), 2(f), 2(i) and 2(l)] by (c) the KNN algorithm and (d) the HDBSCAN algorithm are also shown. Red dashed curves in inset figures show the reference PN distribution, classified by the IP method for 100 kHz coherent states with equal average power per pulse [Figs. 2(a), 2(d), 2(g) and 2(j)].

3.2. Unsupervised Clustering Method

To remove the requirement for training, we also propose an unsupervised method based on a spatial clustering algorithm, hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [50,51]. We reduce the voltage traces to points in a two-dimensional plane by performing PCA and preserving only the first two factor scores. The HDBSCAN algorithm is applied to identify clusters which are areas of high density. To reduce processing time, the clustering model is first trained on a subset of data, then held fixed and applied to predict the cluster label for the rest of the data. For example, it took 1.6 s

to train the clustering model on 20,000 traces from the squeezed states shown in Section 4.2 and subsequently, on average, 22 μ s to predict an unknown trace.

Figure 3(b) shows the clustering results of HDBSCAN on the same coherent states in Section 3.1 at 500 kHz. The main advantage of HDBSCAN is a lower assignment error, as there is little to no crossover between different clusters. The trade-off is that many data points in low-density areas, sandwiched by two high-density clusters, cannot be classified into either cluster and are labeled as noise instead. This skews the overall photon number distribution. Occasionally, the algorithm identifies extraneous sub-clusters inside a photon-number cluster. These sub-clusters can be manually grouped together in an extra post-processing step.

At 800 kHz, the visibility between different photon-number clusters in the PCA plane is greatly reduced [Fig. 3(d)]. As a result, the algorithm fails to identify any photon number clusters from two photons upwards despite careful tuning of the model’s hyperparameters that define the minimum cluster size, the minimum number of neighbors to a core point, and the minimum distance between distinct clusters.

4. BENCHMARKING

4.1. Detection of Coherent States

4.1.1. Total Variation Distance

To benchmark the performance of our proposed signal processing methods, we first apply them to the detection of coherent state light pulses. The light source is a highly attenuated, modulated, laser diode whose repetition rate can be freely controlled by a function generator. When switching between different rates, we maintain the same pre-attenuation average power per pulse through adjustments of the duty cycle, so as to preserve the photon number distribution.

Figures 4(a)–4(c) show how the photon number distribution measured by each method deviates from the reference distribution at different repetition rates and pulse power. The reference distribution is measured by the IP method at 100 kHz repetition rate. We use the total variation distance (TVD) as the distance metric between two distributions, defined by $TVD(p, q) = \sum_i \frac{1}{2} |p_i - q_i|$. The results confirm the improvement by the supervised machine learning methods over the IP method at higher repetition rates up to at least 800 kHz. The unsupervised clustering method by HDBSCAN, expectedly, has poor accuracy in the photon-number distribution, which is due to the fact that the method classifies a number of traces as “noise” and fails altogether to separate clusters for higher photon numbers at 800 kHz.

4.1.2. Detector POVM Reconstruction

A set of coherent states with varying amplitudes, $\{|\alpha_j\rangle\}$, also forms a tomographically complete set of probe states that allows us to reconstruct the POVM of the phase-insensitive TES, which can be expressed as [27,40]

$$\left\{ \hat{\Pi}_n : \hat{\Pi}_n = \sum_{m=0}^{+\infty} \theta_{nm} |m\rangle\langle m| \right\}, \quad (2)$$

where $|m\rangle$ is an m -photon Fock state. In practice, the summation in Eq. (2) is truncated to some large photon number, M . The scalars θ_{nm} , represent the conditional probability that

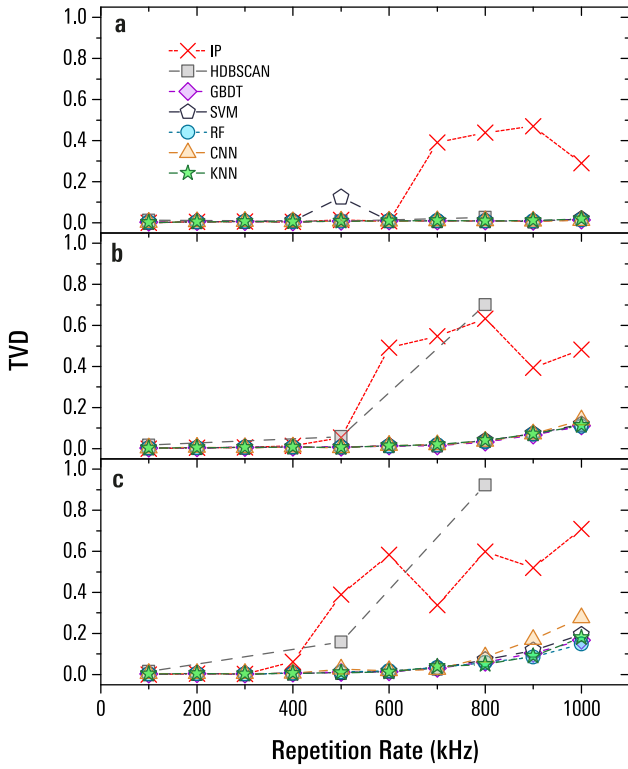


Fig. 4. Detection of coherent state with varying mean photon number per pulse: (a) $\mu = 0.86$; (b) $\mu = 3.82$; (c) $\mu = 5.29$. In each case, the TVD is calculated against the reference distribution measured by the IP method at 100 kHz.

the detector reports n -photons given an m -photon input and form a confusion matrix Θ that illustrates the accuracy of the detector.

To reconstruct the detector's POVM in Eq. (2), we first calibrate the intensities of the probe states, $|\alpha_j|^2$. The light from the laser diode is split on a calibrated beam splitter, with one output being monitored by an independent power meter, and the other output being measured by the TES after a calibrated two-stage fixed attenuation. The amplitude is adjusted by a variable optical attenuator before the beam splitter. Assuming perfect Poissonian behavior, the ground-truth distributions of the coherent state can be calculated. One can then reconstruct the confusion matrix Θ by a least-square minimization procedure between the TES-measured distributions and the ground-truth distributions multiplied by a parameterized version of Eq. (2) [27,40]. The optimization routine was performed using the CVXPY Python package [52,53].

Assuming a perfect power meter, the mean photon number detected by the IP method at 100 kHz gives a detection efficiency of 93.3% at the 1550 nm wavelength (see Supplement 1). Realistically, however, this efficiency estimate will also include inaccuracy in the power meter reading and calibration errors. To benchmark the accuracy of the signal processing methods with minimal error carried over from the power meter's and the detector's intrinsic behavior, we use the reconstructed POVM by the IP method at 100 kHz as a reference, and compare it with the reconstructed POVMs at higher repetition rates.

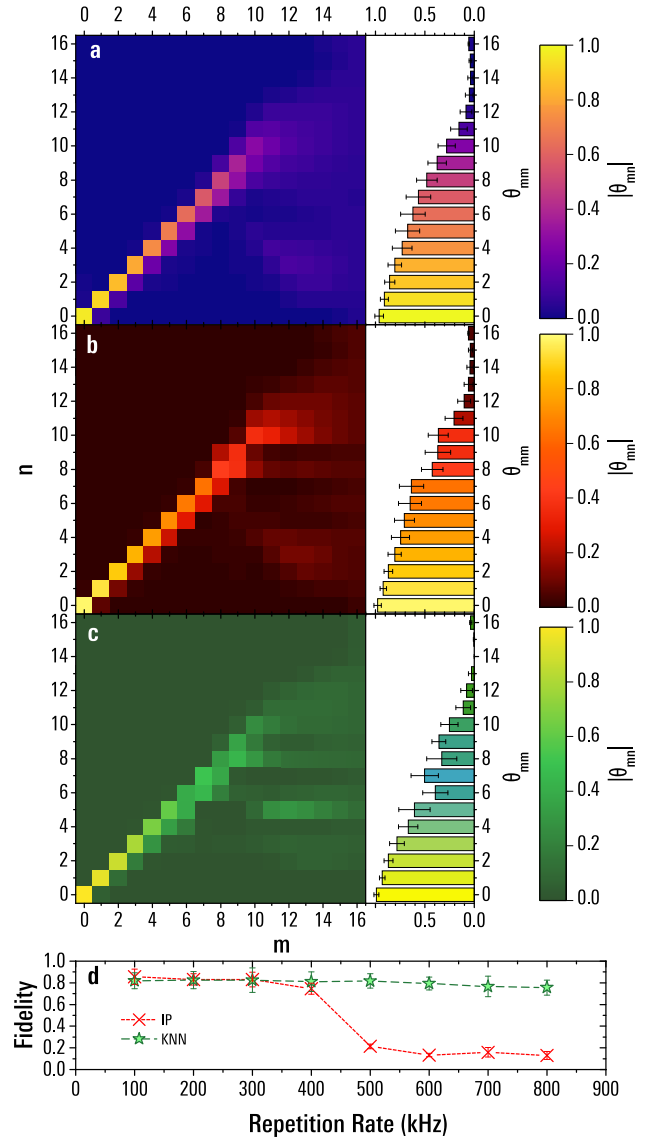


Fig. 5. Reconstructed confusion matrix, Θ by the IP and KNN method at (a) 100 kHz, (b) 500 kHz, and (c) 800 kHz. Bar plots of the diagonal elements are also plotted. The error bars are estimated by bootstrapping over the uncertainties of estimating the probe state intensities (4.6%, see Appendix A) and Poissonian photon counting errors. (d) The fidelity of the Θ reconstructions, as defined in Eq. (3). The reference POVM, $\Theta^{(\text{ref})}$, is the reconstruction result by the IP method at 100 kHz, averaged over the aforementioned bootstrapping procedure.

We define the average fidelity between a reconstructed Θ and the reference $\Theta^{(\text{ref})}$ by [54]

$$F = \frac{1}{N} \sum_n \frac{\left(\sum_m \sqrt{\theta_{nm} \theta_{nm}^{(\text{ref})}} \right)^2}{\left(\sum_m \theta_{nm} \right) \left(\sum_m \theta_{nm}^{(\text{ref})} \right)}, \quad (3)$$

where the reported and actual photon numbers are truncated to N and M , respectively. The average fidelity with $N = M = 16$ shows that the KNN algorithm's performance at up to 800 kHz remains on par with the detector's performance at low repetition rates [Fig. 5(d)].

Looking more closely at the reconstructed diagonal terms of Θ , Figs. 5(a)–5(c) (bar plots) show that the KNN algorithm preserves the detection accuracy for up to at least seven photons at 500 kHz and five photons at 800 kHz. A caveat to this is that this detection accuracy will also depend on the state of the input light. For example, while the reconstructed POVM at 500 kHz showed a drop in accuracy for eight-photon detection [Fig. 5(b)], this is an averaged result across the probe states. For probe states with lower mean photon number, the “tail” influence from preceding traces is on average smaller, reducing the difficulty of the classification problem. For example, for the light pulses shown in Fig. 3(a) which average at 3.82 photons per pulse, the KNN algorithm was still able to label traces accurately up to 13 photons at 500 kHz.

4.2. Detection of Squeezed States

Squeezed states are reduced quantum uncertainty states that have proved to be an invaluable resource in quantum information and sensing. The ability of performing PNR detection on squeezed states of light has enabled key applications in topics such as photonic quantum computing [16], sub-shot-noise quantum sensing [10], and highly non-classical state generation [20,25]. With distinctly different photon-number distributions from the coherent state, detection of squeezed states may pose a different classification task for our algorithms, which we benchmark in this section.

Our squeezed-light source is based on a periodically poled potassium titanyl phosphate (ppKTP) waveguide pumped by a pulsed laser at 780 ± 1 nm wavelength. The waveguide produces orthogonally polarized signal and idler photons at 1560 nm via the type-II PDC process. The generated pairs exhibit photon-number correlation, forming a TMSV state described by $|TMSV\rangle = \sqrt{1 - |\lambda|^2} \sum_{n=0}^{\infty} \lambda^n |nn\rangle$, where λ is the parameter that determines the mean photon number per pulse and depends on the pump field and the ppKTP crystal. The repetition rate of the pump light is 80 MHz but can be modulated by a Pockels cell, which acts as a voltage-controlled pulse picker. We first collect the 800 kHz data. Then we reduce the repetition rate of the pump light to 100 kHz while maintaining the same average pump power per pulse, and measure the photon number distribution by the IP method, which is used as the reference distribution.

The signal and idler pulses are split into two spatial modes and coupled into the TESs. In each mode, the TES records a thermal state distribution. As a result, the PCA result has a different density distribution from that for a coherent state [Fig. 6(a)]. The factor scores for the second-order component, which represents the weight of the tail of the preceding traces, are concentrated in lower values, since the thermal distribution is dominated by lower photon numbers. This reduces the difficulty for our unsupervised clustering method. Even at 800 kHz repetition rate, the HDBSCAN algorithm was able to identify clusters corresponding to photon numbers up to 10 [Fig. 6(b)], which is a drastic improvement from the detection of coherent states.

The algorithm still leaves more than 10% of the traces unclassified, resulting in an inaccurate reconstruction of the thermal distribution. However, among the traces that the algorithm was able to classify, the assignment error is low, which is illustrated by the low crossovers between the different clusters in Fig. 6(b).

This low assignment error manifests in the preparation of heralded Fock states. The photon-number correlation in a TMSV

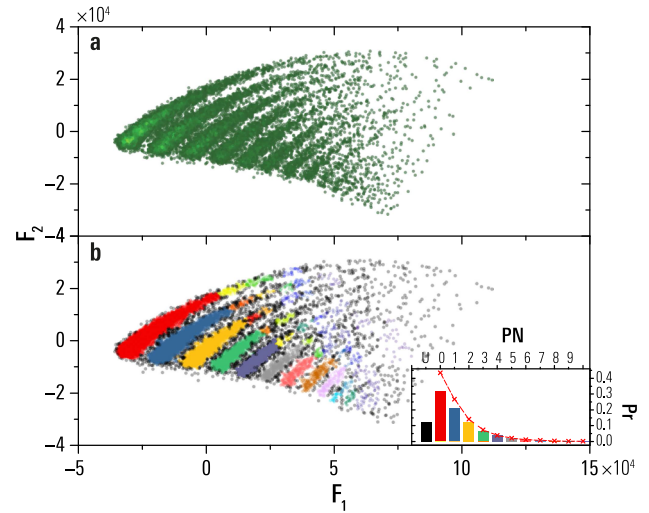


Fig. 6. (a) Density scatter plot of the PCA factor scores for light pulses in a thermal state. (b) The classification result by the HDBSCAN algorithm visualized on the PCA scatter plot. Each cluster is plotted with a different color. The inset shows the photon-number distribution of the thermal state, where the color of each bar is the color of the corresponding cluster in (b). Sub-clusters that have the same photon-number label are grouped together. The red-dashed curve is the distribution reconstructed by the inner product method at 100 kHz for the same average pump power per pulse.

state allows for the preparation of an n -photon Fock state in the signal mode by heralding on the idler mode. For up to $n = 4$, the heralded Fock state detected by the HDBSCAN algorithm produces the correct photon number with the highest probability. The spread of probabilities in unwanted photon numbers is due to photon loss between the PDC source and the detector, rather than due to the detection scheme. After discarding the unclassified traces and renormalizing the remaining probabilities, the heralded photon number distributions at 800 kHz by the HDBSCAN algorithm are almost identical to the reference distributions at 100 kHz [Figs. 7(a)–7(d)]. The overall fidelity between the two-mode photon number distribution at 800 kHz and at 100 kHz [Figs. 7(e) and 7(f)] is $F = (\sum_i \sqrt{p_i q_i})^2 = 99.2\%$.

The supervised KNN classification algorithm was also tested on the detection of the TMSV state and reported 97.6% fidelity at 800 kHz, which is worse than the unsupervised HDBSCAN algorithm. This is likely due to a lower signal-to-noise ratio by the Pockels cell, as compared with the laser diode for coherent state detection, which reduces the quality of the training data and reduces the algorithm’s ability to differentiate low-photon-number traces from background noise.

5. CONCLUSIONS

In this work, we reported two machine-learning-based signal processing methods for TESs that can extend their PNR capabilities beyond the limitations of their sensors’ thermal recovery time: one supervised learning method and one unsupervised clustering method. Testing with coherent states, the first method enabled our TES to operate at up to 800 kHz with little-to-no drop in accuracy for the detection of up to five photons, beating the state-of-the-art signal processing method in Ref. [16]. This is achieved by constructing a training data set from 100 kHz voltage traces, where photon numbers can be accurately assigned

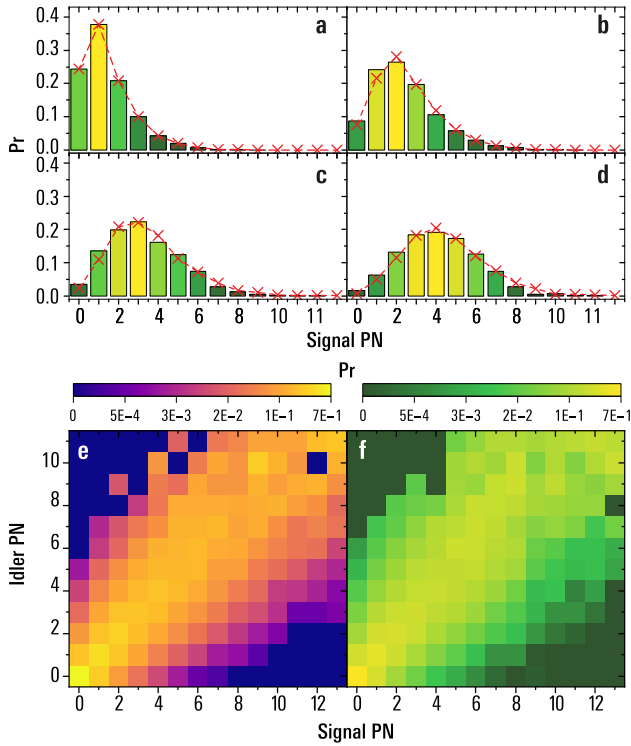


Fig. 7. Photon-number distribution of the signal mode whilst heralding on (a) one, (b) two, (c) three, and (d) four photons in the idler mode. The red dashed curve and crosses represents the reference distribution obtained by the IP method at 100 kHz for the same average pump power per pulse. The two-mode photon-number distribution reconstructed by (e) the inner product method at 100 kHz and (f) the HDBSCAN algorithm at 800 kHz. All data have been renormalized after discarding the unclassified traces.

via the IP method, and overlapping the traces to emulate a particular detection rate. This is verified against the tomographic reconstruction of the TES's POVM at 100 kHz where the TES voltage traces are well-separated in time.

The unsupervised method removes the requirement for training and clusters the raw voltage traces according to their PCA factor scores. The weakness of this density-based clustering method is that traces that do not obviously belong to any cluster might be mislabeled as noise. However, in the traces that the algorithm is able to label, the method has very high precision and low assignment error. This is showcased in the algorithm's ability to prepare heralded Fock states of up to four photons at 800 kHz from a TMSV state of light.

Our proposed methods provide the algorithmic approach to speed up PNR detection with a TES. The hardware-agnostic nature means that they can be implemented in conjunction with any existing or future fabrication improvements, and deliver the last-mile for speeding up quantum experiments that require PNR capability. Notable examples include applications where operations are conditioned on PNR measurements, requiring fast feedforward of PNR outcomes, such as non-Gaussian quantum state preparation [20–25]. In other fields of research that utilize the TES, such as bio-imaging [12], higher detection rates may benefit dynamic studies like live-cell imaging or real-time biological processes where fast image acquisition with low signal-to-noise are essential. Overcoming the slow recovery time might also be beneficial for detecting particle bursts

and in studies of the cosmic microwave background [55] where fine-temporal resolution of events is desired.

The accuracy of our algorithms may be limited by the ambient conditions. One limitation is by the background noise level in the TES signal traces. While the intrinsic dark count rate of TES is close to zero thanks to its low operating temperature (<130 mK) [56], stray background light, electrical noise, secondary cosmic ray particles, and naturally occurring radioactive decays may contribute to an elevated background noise level [35,57]. This can negatively impact the accuracy of the inner-product method and therefore the training step of our supervised learning method. In addition to suppressing background noise by spectral filtering and stray light shielding, PCA is also effective in suppressing extraneous absorption events that are out-of-sync with the laser's trigger signal.

Our algorithms also remain subject to the inherent trade-off between detection rate and photon-number resolution. As the detection rate continues to increase, the temporal features of the signal traces become increasingly difficult to differentiate, resulting in a drop in the accuracy of our algorithms. Beyond a certain rate, accumulated heat from consecutive photon absorptions without sufficient thermal dissipation will saturate the sensor and diminish its PNR capability.

Finally, the machine-learning algorithms presented here can be further optimized by fine-tuning their parameters to specific experimental conditions. Beyond the methods tested in this work, broader machine learning techniques for time-series classification can be further explored in future work in the context of photon detection [36–38].

During preparation of this manuscript, we became aware of a new study [58] demonstrating two non-linear dimensionality reduction techniques which preserve TES signal structures better than the trace area method and the PCA method. If deployed in the training stage of our supervised learning algorithm, we believe the non-linear techniques can further improve the classification accuracy of the training traces. However, these techniques require computation of the full data set, and thus cannot be directly deployed in the prediction stage or in our unsupervised clustering algorithm when real-time detection is required. Nevertheless, this leaves open a promising direction for future research in improving parameterized approaches that better extract and cluster the features of temporally overlapping TES signals.

All methods and code developed in this paper are publicly accessible in Ref. [59] and we hope our paper can provide a gateway to further algorithmic improvements on TES signal processing.

APPENDIX A: METHODS

1. Experimental Setup

The experimental setup is illustrated in Fig. 1(a). The light source for generating coherent states is based on a 1550 nm wavelength pigtailed fiber laser diode driven by a function generator producing a radio frequency signal consisting of 70 ns wide square pulses. The output light is reduced to the few-photon level by three stages of fiber-based attenuation: A variable optical attenuator, a non-polarizing beam splitter with 99 : 1 splitting ratio, and two optical attenuators (only one shown) connected in series with 30 dB attenuation each. The beam splitter output with the larger intensity is monitored by a Newport 918-IG-OD3R power

meter. The amplitudes of the coherent states are varied by the first variable optical attenuator. The repetition rate of the signal is controlled by the function generator. When changing between different repetition rates, the amplitude of the radio frequency signal is adjusted such that the power meter records the same average power per pulse to compensate for the differences in rise and fall times of the square pulses.

For the tomography routine, the amplitudes of the probe states $\{|\alpha_j\rangle\}$ need to be precisely known. Calibration of the asymmetric beam splitter and the attenuation scheme is performed by the same Newport 918-IG-OD3R power meter, which specifies 2.6% reading uncertainty and 0.5% linearity. Combining the uncertainties on beam splitter and attenuator calibrations along with the power meter readings, there is an overall 4.6% uncertainty on our estimation of the probe state intensities. However, since the power meter has not been recalibrated since purchase, the calibration uncertainty could be higher.

For Section 4.2, a PDC source was used to generate the TMSV state. The PDC source consists of a ppKTP waveguide pumped by a femtosecond mode-locked Ti:sapphire laser (Coherent Mira-HP, pumped by a Verdi G18). The pump pulses are filtered to 780 ± 1 nm (full width at half maximum) using a pair of angle-tuned bandpass filters. The Type-II PDC process generates orthogonally polarized signal and idler modes in the same spatial mode, which are spatially separated by a PBS and directed through single-mode optical fibers into two TES detectors. To enhance the spectral purity of the TMSV state, a bandpass filter is placed in front of each fiber coupler.

The inherent repetition rate of the Ti:sapphire pump laser is 80 MHz. Pulse-picking is performed by double-passing the pump light through a Pockels cell containing a rubidium titanyl phosphate (RTP) crystal. The Pockels cell acts as a fast, switchable, HWP rotating the polarization of the incident pump pulse by $\frac{\pi}{2}$. Sandwiched between orthogonal GT polarizers and HWPs, the Pockels cell reduces the effective repetition rate of the Ti:sapphire laser system to ≤ 1 MHz, with 40 dB extinction.

2. TES Operation

The active material in the TES is a tungsten thin film sandwiched between a multi-layer optical structure to optimize absorption efficiency at 1550 nm wavelength. For this type of structure, above 95% detection efficiency was reported by NIST [1]. Our measurement showed a 93.3% efficiency (see Supplement 1), though this figure was calculated against an uncalibrated power meter (Newport 918-IG-OD3R). As shown in Figs. 7(a) and 7(b), the TESs are housed inside an Oxford Instruments Triton 200 dilution refrigerator whose base temperature is maintained below 130 mK.

Each individual TES is connected in series with an inductor and in parallel with a shunt resistor [Fig. 1(c)]. The voltage-biased sensors are maintained at the material's superconducting transition phase by the negative electro-thermal feedback effect [60]. Upon absorbing photons, the sensor steeply rises in temperature, as well as resistance. This reduces the current flowing through the inductor, causing a change in magnetic flux that is amplified by a SQUID array. The SQUIDs output a microvolt-level voltage signal that is further amplified to millivolt-level by low-noise amplifiers (4 nV/ $\sqrt{\text{Hz}}$ root-mean-square) at room temperature. To increase the signal-to-noise ratio, a 1 MHz low-pass filter is used to filter out background noise. Finally, the voltage traces are collected by a 14-bit waveform digitizer

(AlazarTech ATS9440) for processing on an AMD Threadripper 7970X-based workstation.

Funding. Engineering and Physical Sciences Research Council Hub in Quantum Computing and Simulation (project T001062); UK Research and Innovation Future Leaders Fellowship (project MR/W011794/1); National Research Council Canada (project QSP 062-2); EU Horizon 2020 Marie Skłodowska-Curie Innovation Training Network (project no. 956071, 'AppQInfo'); UK Research and Innovation Guarantee Postdoctoral Fellowship (project EP/Y029631/1); ORCA Computing.

Acknowledgments. We thank the National Institute of Standards and Technology (NIST) for providing the TESs and SQUIDs used in this work, and Adriana Lita for useful discussions regarding the TES recovery process. Z.L. acknowledges partial funding support from ORCA Computing.

Disclosures. The authors declare no conflicts of interest.

Data availability. Methods and code developed in this paper are publicly accessible in Ref. [59].

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," *Opt. Express* **16**, 3032 (2008).
2. A. E. Lita, B. Calkins, L. A. Pellouchoud, *et al.*, *Advanced Photon Counting Techniques IV* (SPIE, 2010), p. 76810D.
3. R. W. Romani, A. J. Miller, B. Cabrera, *et al.*, "First astronomical application of a cryogenic transition edge sensor spectrophotometer," *Astrophys. J.* **521**, L153–L156 (1999).
4. W. S. Holland, D. Bintley, E. L. Chapin, *et al.*, "SCUBA-2: the 10 000 pixel bolometer camera on the James Clerk Maxwell Telescope," *Mon. Not. R. Astron. Soc.* **430**, 2513–2533 (2013).
5. L. Gottardi and K. Nagayashi, "A review of X-ray microcalorimeters based on superconducting transition edge sensors for astrophysics and particle physics," *Appl. Sci.* **11**, 3793 (2021).
6. K. Irwin and G. Hilton, *Cryogenic Particle Detection* (Springer, 2005), pp. 63–150.
7. J. N. Ullom and D. A. Bennett, "Review of superconducting transition-edge sensors for x-ray and gamma-ray spectroscopy," *Supercond. Sci. Technol.* **28**, 084003 (2015).
8. G.-Y. Xiang, B. L. Higgins, D. Berry, *et al.*, "Entanglement-enhanced measurement of a completely unknown optical phase," *Nat. Photonics* **5**, 43–47 (2011).
9. A. Datta, L. Zhang, N. Thomas-Peter, *et al.*, "Quantum metrology with imperfect states and detectors," *Phys. Rev. A* **83**, 063836 (2011).
10. G. S. Thekkadath, M. E. Mycroft, B. A. Bell, *et al.*, "Quantum-enhanced interferometry with large heralded photon-number states," *npj Quantum Inf.* **6**, 89 (2020).
11. C. You, M. Hong, P. Bierhorst, *et al.*, "Scalable multiphoton quantum metrology with neither pre- nor post-selected measurements," *Appl. Phys. Rev.* **8**, 041406 (2021).
12. K. Niwa, K. Hattori, and D. Fukuda, "Few-photon spectral confocal microscopy for cell imaging using superconducting transition edge sensor," *Front. Bioeng. Biotechnol.* **9**, 1 (2021).
13. S. Aaronson and A. Arkhipov, *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, 2011), pp. 333–342.
14. C. S. Hamilton, R. Kruse, L. Sansoni, *et al.*, "Gaussian boson sampling," *Phys. Rev. Lett.* **119**, 170501 (2017).
15. R. Kruse, C. S. Hamilton, L. Sansoni, *et al.*, "Detailed study of Gaussian boson sampling," *Phys. Rev. A* **100**, 032326 (2019).
16. L. S. Madsen, F. Laudenbach, M. F. Askarani, *et al.*, "Quantum computational advantage with a programmable photonic processor," *Nature* **606**, 75–81 (2022).
17. D. H. Smith, G. Gillett, M. P. de Almeida, *et al.*, "Conclusive quantum steering with superconducting transition-edge sensors," *Nat. Commun.* **3**, 625–626 (2012).

18. M. Giustina, M. A. M. Versteegh, S. Wengerowsky, *et al.*, "Significant-loop-hole-free test of Bell's theorem with entangled photons," *Phys. Rev. Lett.* **115**, 250401 (2015).
19. M. E. Mycroft, T. McDermott, A. Buraczewski, *et al.*, "Proposal for the distribution of multiphoton entanglement with optimal rate-distance scaling," *Phys. Rev. A* **107**, 012607 (2023).
20. T. Gerrits, S. Glancy, T. S. Clement, *et al.*, "Generation of optical coherent-state superpositions by number-resolved photon subtraction from the squeezed vacuum," *Phys. Rev. A* **82**, 031802 (2010).
21. N. Namekata, Y. Takahashi, G. Fujii, *et al.*, "Non-Gaussian operation based on photon subtraction using a photon-number-resolving detector at a telecommunications wavelength," *Nat. Photonics* **4**, 655–660 (2010).
22. G. S. Thekkadath, B. A. Bell, I. A. Walmsley, *et al.*, "Engineering Schrödinger cat states with a photonic even-parity detector," *Quantum* **4**, 239 (2020).
23. O. S. Magaña-Loaiza, R. d. J. León-Montiel, A. Perez-Leija, *et al.*, "Multiphoton quantum-state engineering using conditional measurements," *npj Quantum Inf* **5**, 80 (2019).
24. I. Tzitrin, J. E. Bourassa, N. C. Menicucci, *et al.*, "Progress towards practical qubit computation using approximate Gottesman-Kitaev-Preskill codes," *Phys. Rev. A* **101**, 032315 (2020).
25. M. Endo, R. He, T. Sonoyama, *et al.*, "Non-Gaussian quantum state generation by multi-photon subtraction at the telecommunication wavelength," *Opt. Express* **31**, 12865–12879 (2023).
26. Z. H. Levine, T. Gerrits, A. L. Migdall, *et al.*, "Algorithm for finding clusters with a known distribution and its application to photon-number resolution using a superconducting transition-edge sensor," *J. Opt. Soc. Am. B* **29**, 2066 (2012).
27. L. A. Morais, T. Weinhold, M. P. de Almeida, *et al.*, "Precisely determining photon-number in real time," *Quantum* **8**, 1355 (2024).
28. C. Cahall, K. L. Nicolich, N. T. Islam, *et al.*, "Multi-photon detection using a conventional superconducting nanowire single-photon detector," *Optica* **4**, 1534–1535 (2017).
29. S. Sempere-Llagostera, G. S. Thekkadath, R. B. Patel, *et al.*, "Reducing $g^{(2)}(0)$ of a parametric down-conversion source via photon-number resolution with superconducting nanowire detectors," *Opt. Express* **30**, 3138 (2022).
30. T. Schapeler, N. Lamberty, T. Hummel, *et al.*, "Electrical trace analysis of superconducting nanowire photon-number-resolving detectors," *Phys. Rev. Appl.* **22**, 014024 (2024).
31. Y.-H. Deng, Y.-C. Gu, H.-L. Liu, *et al.*, "Gaussian boson sampling with pseudo-photon-number-resolving detectors and quantum computational advantage," *Phys. Rev. Lett.* **131**, 150601 (2023).
32. W. Zhang, J. Huang, C. Zhang, *et al.*, "A 16-pixel interleaved superconducting nanowire single-photon detector array with a maximum count rate exceeding 1.5 GHz," *IEEE Trans. Appl. Supercond.* **29**, 1–7 (2019).
33. L. Stasi, T. Taher, G. V. Resta, *et al.*, "Enhanced detection rate and high photon-number efficiencies with a scalable parallel SNSPD," *ACS Photonics* **12**, 320 (2025).
34. R. Hummatov, A. E. Lita, T. Farrahi, *et al.*, "Fast transition-edge sensors suitable for photonic quantum computing," *J. Appl. Phys.* **133**, 234502 (2023).
35. C. Pepe, "Development of superconducting single-particle detector transition-edge sensor," Ph.D. thesis (Politecnico di Torino, 2024).
36. A. Bagnall, J. Lines, A. Bostrom, *et al.*, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min. Knowl. Discov.* **31**, 606–660 (2017).
37. H. Ismail Fawaz, G. Forestier, J. Weber, *et al.*, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.* **33**, 917–963 (2019).
38. M. Middlehurst, P. Schäfer, and A. Bagnall, "Bake off redux: a review and experimental evaluation of recent time series classification algorithms," *Data Min. Knowl. Discov.* **38**, 1958–2031 (2024).
39. H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Stat.* **2**, 433–459 (2010).
40. P. C. Humphreys, B. J. Metcalf, T. Gerrits, *et al.*, "Tomography of photon-number resolving continuous-output detectors," *New J. Phys.* **17**, 103044 (2015).
41. T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory* **13**, 21–27 (1967).
42. F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
43. L. Breiman, "Random Forests," *Mach. Learn.* **45**, 5–32 (2001).
44. R.-E. Fan, K.-W. Chang, C.-J. Hsieh, *et al.*, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.* **9**, 1871–1874 (2008).
45. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *J. Mach. Learn. Res.* **2**, 1–27 (2011).
46. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16 (ACM, 2016), pp. 785–794.
47. M. Abadi, A. Agarwal, P. Barham, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," Software available from tensorflow.org (2015).
48. M. Lin, Q. Chen, and S. Yan, "Network In Network," *arXiv* (2013).
49. Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2017), pp. 1578–1585.
50. R. J. G. B. Campello, D. Moulavi, and J. Sander, *Advances in Knowledge Discovery and Data Mining* (Springer, 2013), pp. 160–172.
51. L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *The J. Open Source Softw.* **2**, 205 (2017).
52. S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.* **17**, 1–5 (2016).
53. A. Agrawal, R. Verschuere, S. Diamond, *et al.*, "A rewriting system for convex optimization problems," *J. Control. Decis.* **5**, 42–60 (2018).
54. L. Zhang, H. B. Coldenstrodt-Ronge, A. Datta, *et al.*, "Mapping coherence in measurement via full quantum tomography of a hybrid optical detector," *Nat. Photonics* **6**, 364–368 (2012).
55. D. Dutcher, S. M. Duff, J. C. Groh, *et al.*, "The Simons Observatory: large-scale characterization of 90/150 GHz TES detector modules," *J. Low Temp. Phys.* **214**, 247–255 (2024).
56. A. J. Miller, S. W. Nam, J. M. Martinis, *et al.*, "Demonstration of a low-noise near-infrared photon counter with multiphoton discrimination," *Appl. Phys. Lett.* **83**, 791–793 (2003).
57. L. Manenti, C. Pepe, I. Sarnoff, *et al.*, "Dark counts in optical superconducting transition-edge sensors for rare-event searches," *Phys. Rev. Appl.* **22**, 024051 (2024).
58. N. Dalbec-Constant, G. Thekkadath, D. England, *et al.*, "Accurate unsupervised photon counting from transition edge sensor signals," *arXiv* (2024).
59. Z. Li and M. Kendall, "High_speed_TES_ML," https://github.com/apprenticeadi/High_speed_TES_ML (2024).
60. K. D. Irwin, "An application of electrothermal feedback for high resolution cryogenic particle detection," *Appl. Phys. Lett.* **66**, 1998–2000 (1995).