Algorithm for finding clusters with a known distribution and its application to photon-number resolution using a superconducting transition-edge sensor

Zachary H. Levine,[1,*] Thomas Gerrits,[2] Alan L. Migdall,[1] Daniel V. Samarov,[1] Brice Calkins,[2] Adriana E. Lita,[2] and Sae Woo Nam[2]

[1]*National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA*
[2]*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*
**Corresponding author: zlevine@nist.gov*

Received May 8, 2012; accepted June 21, 2012;
posted June 29, 2012 (Doc. ID 168200); published July 20, 2012

Improving photon-number resolution of single-photon sensitive detectors is important for many applications, as is extending the range of such detectors. Here we seek improved resolution for a particular superconducting transition-edge sensor (TES) through better processing of the TES output waveforms. With that aim, two algorithms to extract number resolution from TES output waveforms are compared. The comparison is done by processing waveform data sets from a TES illuminated at nine illumination levels by a pulsed laser at 1550 nm. The algorithms are used to sort the individual output waveforms and then create clusters associated with individual photon numbers. The first uses a dot product with the waveform mean (for each illumination level), while the second uses K-means clustering modified to include knowledge of the Poisson distribution. The first algorithm is shown to distinguish adjacent peaks associated with photon numbers up to 19, whereas the second algorithm distinguishes photon numbers up to 23, using the same data.

OCIS codes: 040.5570, 120.5630.

1. INTRODUCTION

National metrology institutes have shown recent interest in using single photons as the basis of radiometric standards [1–5] to complement existing thermal and synchrotron sources. Such metrology efforts require the linking of single-photon measurements to radiometric standards at higher power levels, and extending photon counting and photon-number resolution to higher numbers is a step toward that goal. Such an extension of photon-number resolution to higher numbers for one type of photon-number-resolving detector is the motivation of the work presented here. A further and important step, but not one discussed here, is characterizing the uncertainties from the photon-number determination of the detector output for larger incident photon numbers.

Photon-number determination for number-resolving superconducting transition-edge sensors (TES) is generally accomplished through the use of an amplitude threshold that distinguishes between output waveforms of adjacent photon numbers. This method allows real-time knowledge of the observed photon number, which is of particular importance for applications such as quantum information processing, quantum repeaters, and quantum (and even classical) communications, where delays in determining the photon number add significantly to the already formidable implementation challenges [6–10]. However, it requires careful setting of the output amplitude threshold, as misidentification of photon numbers can occur due to improperly chosen levels [11]. Even with an optimal threshold, noise on the detector output waveforms also contributes to misidentification of individual photon numbers. Such noise is always present to some degree with TES detection and is often quantified in terms of energy resolution of the TES. When measuring an ensemble of a monochromatic input light state, a histogram of the ensemble allows for the identification of individual peaks that are largely due to the absorption of a specific photon number [12]. Although these peaks may overlap, under the assumption of white noise and hence a Gaussian distribution of individual photon-number output, one can still determine the photon probability distribution of the input state by fitting the individual peaks of the ensemble histogram. Hence, a characterization of the input state is possible. It is important, though, to note the limitations of this Gaussian assumption. With well-resolved histogram peaks and clearly defined Gaussian shapes, this assumption is likely to be valid for some range from the center of the peaks, but in the regions where the peaks significantly overlap, the Gaussian assumption cannot be used with much confidence, as there are a number of physical phenomena, such as changes in where the optical energy is absorbed, that can lead to nonlinear and non-Gaussian behavior. As a result, the uncertainty of a particular photon-number determination can be hard to correctly quantify and easy to underestimate.

To avoid the pitfalls associated with assuming an overall Gaussian noise when it cannot be verified, except near the center of the peak, we use of the term “visibility” to gauge how well a detector and an analysis of its output waveforms can distinguish adjacent photon-number peaks. This photon-number visibility, as we shall see, is defined similarly to fringe visibility in terms of the maximums and minimums of an oscillatory shape. We feel that such a term is less likely to be misinterpreted with respect to photon-number uncertainty. While the TES is often used in conjunction with classical

states, the detection and characterization of squeezed light [13] is just one example of an application where better photon-number determination and better understanding of what is and is not known about the uncertainties associated with that determination is particularly important.

To motivate our focus on the TES and to put its performance in context, we compare it to some other photon-number-resolving detector technologies. TESs have been used to distinguish photon peaks ranging from 0 to 15 photons at a wavelength of 1550 nm [14]. Using an alternate technology, the charge integration photon detector, individual photon peaks up to 10 have been demonstrated at wavelengths of both 1530 and 1310 nm [15]. Another technology, the Multi-Pixel Photon Counter, resolved individual photon-number peaks for optical pulses with an average of 96 photons using a 2500 pixel detector [16]. Recent reviews of the field of single-photon detectors [17,18] show that of all optical single-photon detector technologies, the TES is unique in its capability of energy-resolving the absorbed light and, for monochromatic light, in its photon-number resolution. In addition, the TES's near unity detection efficiencies put the detector in a class by itself [19,20]. It is only the slow thermal response that detracts from the otherwise extraordinary performance of this device.

In a parallel study, the output of a TES was studied while illuminating it with coherent state laser pulses at a 1 kHz repetition rate. TES output waveforms were collected for a large number of pulses (20 480), and such data sets were taken at 45 different illumination levels ranging from a mean of 2.00 photons per pulse to a mean of 6.3 million photons per pulse. The waveforms were digitized using 8192 time steps of 100 ns each. In the study here, we use the data of the lowest nine mean photon numbers (from the highest nine attenuator settings), which range from 2.00 to 31.6.

The principal goal of this paper is the development of an algorithm that is capable of classifying the detector output waveforms into groups with single-photon-number resolution to as high a photon number and the best photon peak visibility as possible. We also discuss how the shapes of the waveforms vary with photon number and how that impacts the classification processes.

2. ALGORITHMS TO OBTAIN PHOTON-RESOLVED RESPONSE CURVES

Our intent is to classify the TES's output waveforms in terms of the actual number of photons n absorbed by the TES from a given pulse of an ensemble defined by an average number of detected photons $\langle n\rangle$. Note that the experimenter has control over only the mean value $\langle n\rangle$, but not over n, and the response of the TES depends on n, not on $\langle n\rangle$. Thus, to classify an individual waveform, we first generate a family of output waveforms for each input photon number. We do this by sorting our large sample of waveforms generated from a source with a particular statistical distribution of n with known mean $\langle n\rangle$. Then we break those sorted waveforms into groups associated with a particular photon number and create a family of waveforms $\bar{V}_n(t)$ from averages of the waveforms in each group. (Here, V is a voltage, and t is time from the start of the pulse.)

To begin classifying the waveforms, we make several assumptions. First, the optical source generates photons according to a Poisson distribution and any subsequent losses reduce the mean, but retain the Poisson form. This will be true whether the losses are in delivering the light to the detector or within the detector itself. Thus, the numbers of detected photons n follow a Poisson distribution with detected mean $\langle n\rangle$

$$P_n(\langle n\rangle) = e^{-\langle n\rangle}\frac{\langle n\rangle^n}{n!}. \tag{1}$$

We define an average waveform for n to be $\bar{V}_n(t)$. Second, we assume $\bar{V}_0(t) = 0$; i.e., if no photons enter the detector, the mean response is some baseline voltage that is defined to be 0. Third, the mean response for all n obeys $\bar{V}_n(0) = 0$, is single peaked, and returns asymptotically to 0 at large time. Finally, the mean response for a given n may be described as an excitation above some baseline, and if $n > n'$, $\bar{V}_n(t) \geq \bar{V}_{n'}(t)$ for all t.

The assumptions encompass the notion that more photons lead to a bigger response. But individual waveforms are subject to noise, so there is not necessarily a unique way to say a specific output waveform is larger than another. Hence, we explored several options for ordering the waveforms, including using their integrals, root-mean-squared (RMS) values, and projection onto a fixed function.

A. Dot-Product Algorithm for Ordering the Waveforms

One effective method we found was the following: we first averaged all the waveforms taken at a given average photon number $\langle n\rangle$ to find

$$\hat{V}_{\langle n\rangle}(t) = \frac{1}{M}\sum_{i=1}^{M} V_i(t), \tag{2}$$

where M is the total number of waveforms taken at that average photon number. Then, the dot product of each waveform is found with that average waveform via

$$n_i^{(\mathrm{eff})} = \langle n\rangle \frac{\sum_t V_i(t)\hat{V}_{\langle n\rangle}(t)}{\sum_t [\hat{V}_{\langle n\rangle}(t)]^2}. \tag{3}$$

By dotting each waveform with a function that depends on $\langle n\rangle$, we avoided the loss of sensitivity that occurred when we chose a single $\langle n\rangle$-independent function. The normalization yields a rough approximation to the photon number n.

The waveforms were ordered by their values of $n^{(\mathrm{eff})}$. Then the ordered list of waveforms was divided into clusters of waveforms. The number of waveforms in each cluster was chosen to be equal to $MP_n(\langle n\rangle)$ rounded to integers that sum to M, where $P_n(\langle n\rangle)$ is the probability of having n photons according to the Poisson distribution with mean $\langle n\rangle$, given in Eq. (1). Thus, if the $n = 0$ cluster is to have, say, three members, the three waveforms with the lowest values of $n^{(\mathrm{eff})}$ are assigned to it. The $n = 1$ might get the next, say, seven waveforms, and so on. We refer to this algorithm as the "dot-product method."

B. The Poisson-Influenced *K*-Means Algorithm (PIKA)

While the dot-product method was reasonably effective, we wanted to see if a better algorithm could be found. First, we considered K-means clustering [21], a popular and efficient method of classifying high-dimensional data into a few clusters. A typical application for K-means clustering is to classify a spectral image obtained from an airplane or satellite

into terrain types (e.g., water, forest, road, etc.) In our TES application, the clusters are the waveforms for a given incident photon number n, and the cluster means are the average waveform within a cluster. A high-dimensional space arises from the many discrete time points in each waveform.

In the standard K-means algorithm [22], the data are grouped into K distinct clusters where K is an integer. The clusters minimize the sum of the squared deviation of each waveform from the mean of the waveforms in the cluster. Although it is not always recognized, the K-means algorithm may be viewed as the minimization of an objective function [21]. The form of the objective function we used is negative of the logarithm of the likelihood of an isotropic, multivariate Gaussian with no correlations between variables. This is a convenient form, as it converts a Gaussian to a paraboloid [23]. The K-means objective function, adapted to the present variables, is

$$O_K = \sum_{n=n_0}^{n_0+K-1} \sum_{i\in C_n} \frac{1}{N_t} \sum_t [V_i(t) - \bar{V}_n(t)]^2. \tag{4}$$

As before, n is the photon number characterizing the cluster. Also, i is an index over the waveforms, t is a discrete time variable, N_t is the number of time points in each waveform, C_n is the set of indices of waveforms in cluster n, and $\bar{V}_n(t)$ is the mean of all the waveforms in cluster C_n. The range of n is bounded by n_0 and $n_0 + K - 1$, the smallest and largest numbers of photons expected to have been observed at least once; this consideration sets the number of clusters K. The cluster means are defined by

$$\bar{V}_n(t) = \frac{1}{m_n} \sum_{i\in C_n} V_i(t), \tag{5}$$

where m_n is the number of waveforms in cluster n.

The K-means objective function failed to find reasonable clusters for our data. From a random starting point, certain known clusters (e.g., $n = 0$ response) did not emerge if the data sets were large. The results were better using an initial ordering from the dot-product method; however, under iteration, there was a tendency for large clusters to split into two, leading to large deviations from the Poisson distribution and uneven spacings of the cluster-mean waveforms.

To address the situation, we decided to combine the K-means objective function with our prior knowledge of the cluster sizes, namely the Poisson distribution of photon numbers with a known mean. The mean of the Poisson distribution is known to $\pm 1\%$ (expanded uncertainties calculated from the standard deviation of the mean with coverage factor $k = 2$ are used here and throughout), because it was found for the data set with the lowest optical input to be $\langle n\rangle = 2.00 \pm 0.02$ and scaled for the other values based on the known attenuation of the optical filters used in the experiment [24]).

We create a new objective function by adding a regularization term leading to

$$O_{\mathrm{KPC}} = \frac{1}{2\sigma^2} O_K + O_{\mathrm{PC}}, \tag{6}$$

where σ is a constant and

$$O_{\mathrm{PC}} = -\ln \mathcal{L}(m_{n_0}, \ldots, m_{n_0+K-1}; \langle n\rangle). \tag{7}$$

Here, $\mathcal{L}$ is the likelihood of having $M = \sum_{n=n_0}^{n_0+K-1} m_n$ waveforms apportioned among the K clusters with m_n waveforms assigned to each n-photon cluster. The objective function O_{KPC} has been used in the context of gene expression as implemented in the PoissonL algorithm [25]. However, our method for minimizing the objective function, which is described in Appendix A, differs. In practice, we have $M = 20\,480$, if all available data with a given mean photon number $\langle n\rangle$ are used. We choose to write the constant relating O_K and $-\ln \mathcal{L}$ in the form $1/(2\sigma^2)$ to emphasize the connection of the K-means objective function to the negative-log-likelihood of the normal distribution [23]. This makes it possible to estimate σ, rather than simply tuning a regularization parameter with otherwise unknown properties. Of course, as $\sigma \to 0$, O_{KPC} reduces to the O_K times a constant, whereas for $\sigma \to \infty$, the term $\ln \mathcal{L}$ dominates, so the number of waveforms assigned to each cluster approaches the Poisson mean values (up to discretization). In practice, we chose a single $\sigma = 14.6$ mV for all $\langle n\rangle$ based on the standard deviation of waveforms assigned to clusters at small n. Later, when higher n clusters became available to us, we found that σ may decrease as n increases, but the effect is modest, and only the original parameter is reported.

The likelihood of observing cluster frequencies $m_{n_0}, \ldots, m_{n_0+K-1}$ depends on two factors: $\mathcal{L}_P$, the likelihood from the Poisson distribution, and $\mathcal{L}_C$, a combinatorial factor to account for the number of ways M waveforms could have been grouped into K groups of with frequencies $m_{n_0}, \ldots, m_{n_0+K-1}$. Of course,

$$\mathcal{L} = \mathcal{L}_{\mathrm{P}} \mathcal{L}_{\mathrm{C}}. \tag{8}$$

To write $\mathcal{L}_P$, the Poisson distribution of Eq. (1) is used. The likelihood that the laser would yield a particular photon-number sequence is

$$\mathcal{L}_P = \prod_{n=n_0}^{n_0+K-1} \left(\frac{e^{-\langle n\rangle} \langle n\rangle^n}{n!}\right)^{m_n}. \tag{9}$$

The combinatorial factor is

$$\mathcal{L}_{\mathrm{C}} = M! \left(\prod_{n=n_0}^{n_0+K-1} m_n!\right)^{-1}, \tag{10}$$

the number of permutations of the arrival sequences leading to the same cluster frequencies. (As an example, the three arrival sequences 011, 101, and 110 all yield frequencies of (1,2) for $n = (0, 1)$ photons, respectively, but the only way to achieve frequencies of (0,3) is to have the arrival sequence 111.)

With the objective function chosen and the initial clusters assigned, it remains to minimize the objective function. Here, we implemented two strategies for optimizing O_{KPC}, known as the greedy algorithm [26] and simulated annealing [27]. We found the simulated annealing offered only a very modest improvement over the results of the greedy algorithm, suggesting that the objective function does not have important local minima, which is a pitfall of the greedy algorithm. Results are given for the greedy algorithm.

The relationship between the dot-product algorithm and the Poisson-influenced K-means algorithm is shown in Fig. 1. In plain language, the dot-product method generates a set of mean waveform curves that simply scale with n. In an alternative representation, the dot-product method orders waveforms on a line through the hyperspace consisting of a voltage for each of a finite set of time points. The Poisson-assisted K-means method finds a piecewise linear representation of an ideal detector response curve in the same hyperspace. The closest approach between the curve and the line will occur near $n = \langle n \rangle$, where one would expect the best determination of the mean cluster waveform to be. Two waveforms are shown schematically as points on the graph, where for the case shown, the ordering of the waveforms is seen to reverse, depending upon which algorithm result is chosen.

3. NOISE AND BACKGROUND REDUCTION

The data used in this work consisted of sets of 20 480 voltage waveforms recorded in 8192 times steps of 100 ns for each of nine illumination levels. The voltages were discretized in steps of 2^{-14} V ≈ 61 μV with about 0.2 V of the available 4 V range being used in practice. For this study, except for determining the background rate, it was sufficient to retain the first 20 μs, as the pulses were quiescent thereafter.

In dealing with experimental data, it is common to apply a Wiener filter before further analysis [23]. In the case of a TES, the Wiener filter has been shown not to be optimal by a factor of about 2 [28]. However, the Wiener filter is simpler to implement and was chosen for this work. The power spectrum had a noise floor similar to the example in [23]. Following that as a guide, the signal to noise ratio was fit to $|S(\nu)|^2/|N(\nu)|^2 = 104 \exp(-\nu/\nu_0)$, with $\nu_0 = 137$ kHz. These relations were used in the Wiener filter formula [23]

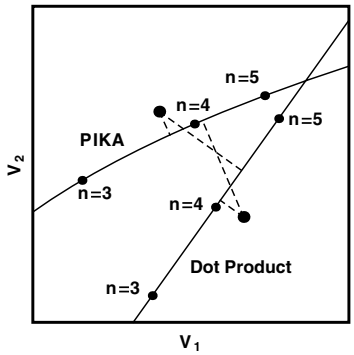

Fig. 1. Schematic representation of the ordering of waveforms in the two algorithms. A waveform may be viewed as a point in a high-dimensional space. Specifically, a waveform discretized at times $t_1, \ldots, t_{N_t}$ may be considered to be the point $(V(t_1), \ldots, V(t_{N_t})) = (V_1, \ldots, V_{N_t})$. Just two dimensions of this high-dimensional space are shown here. The evolution of the waveform with photon number is seen for the dot-product algorithm in the straight line resulting from Eq. (2) and for the Poisson-influenced K-means algorithm (PIKA) in the curved line representing Eq. (13). The values of $n = 3$, 4, or 5 are shown for both cases (small dots). Each large dot represents a particular waveform $V_i(t)$ to be ordered. These waveforms are projected onto the line and the curve to show how they would be ordered by the two algorithms. One feature of the actual curves represented here is the uniform spacing of n for the dot-product method and the decreasing spacing for PIKA.

$$\Phi(\nu) = \frac{|S(\nu)|^2}{|S(\nu)|^2 + |N(\nu)|^2}. \tag{11}$$

A spike in all the waveforms at 2.268 MHz was attributed to electrical pickup. While no special attempt was made to remove the spike, it was, however, strongly attenuated by the filter.

Each waveform was Fourier transformed from the time domain to the frequency domain. The transform was filtered, i.e., multiplied by $\Phi(\nu)$. After applying the Wiener filter, rapid fluctuations in the data were largely removed. The upper half of the frequencies yielded negligible amounts of power, so these were discarded before the back transform. As a result of halving the number of frequencies, the time step increased from 100 to 200 ns, leaving 100 time points. The summations over t in this paper refer to this domain.

For background studies, the final 4000 time steps (equal to 800 μs) were considered. A total of 10 240 observations were considered for a total of 8.192 s. The background portion of the waveforms fluctuated in a band around some constant value. However, certain pulses brought the background out of the band. These were attributed to blackbody photons. In the total observation time, 37 049 such events were identified yielding a rate of 4520 s^{-1}. The ratio of double height peaks to single height peaks is 0.020, indicating a photon arrival rate of 4570 s^{-1}, consistent with statistically independent arrival times.

For the four data sets with the lowest illumination ($\langle n \rangle = 2.00$ to $\langle n \rangle = 5.68$), it was possible to identify and remove most of the blackbody photons [29]. For the two lowest illumination data sets, such removal was necessary to form plausible cluster-mean waveforms. However, for the larger data sets, it became increasingly difficult to identify the blackbody photons. Fortunately, for the $\langle n \rangle = 5.68$ set and beyond, the parameters of interest were not affected by the inclusion of blackbody photons, so no attempt at background removal was made for the next five data sets, $\langle n \rangle = 7.99$ to $\langle n \rangle = 31.6$.

4. RESULTS

For data sets $\langle n \rangle = 2.00$ and $\langle n \rangle = 2.83$, the clusters for different n are well separated, and the dot-product method as well as the minimization of the objective function O_{PC} led to the same result. Cluster means are presented for these cases in Fig. 2(a). The figure shows the cluster means for various n are nearly independent of $\langle n \rangle$. While we expect this on physical grounds (as the TES responds to the actual photon number absorbed in a single light pulse not the average of many pulses), it is a useful cross-validation that both algorithms obtain common results from two different sets of observations.

We are interested in resolving the photon number. To this end, each waveform was assigned an interpolated photon number that takes advantage of our knowledge of $\bar{V}_n(t)$, which is more detailed knowledge than $n^{(\mathrm{eff})}$ derived from $\hat{V}_{\langle n \rangle}(t)$ in Eq. (3). To assign the photon numbers, the photon number n of the cluster of which the waveform is a member is noted. Then, the cluster second closest (in the sense of finding the n' that minimizes $\int \mathrm{d}t(V_i(t) - \bar{V}_{n'}(t))^2$) for a given waveform $V_i(t)$ is found. In practice, this is always cluster $n-1$ or $n+1$. We call the value of the second closest cluster n'. The waveform is fit as a linear combination of the cluster means

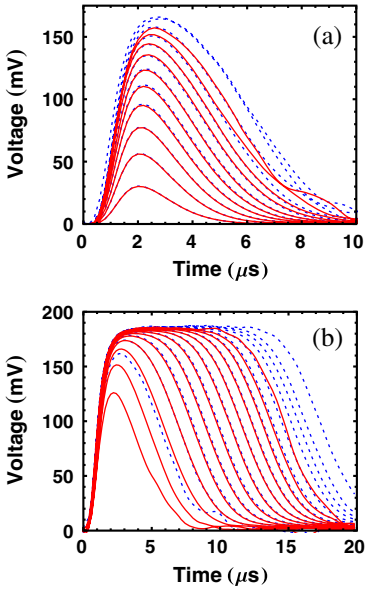

Fig. 2. (Color online) (a) Cluster waveform means for $0 \le n \le 12$ derived from the two lowest illumination data sets $\langle n \rangle = 2.00$ (solid red) and $\langle n \rangle = 2.83$ (dotted blue). The $n = 0$ response is approximately zero and is not very visible in the plot. (b) Waveform cluster means are given as derived from the $\langle n \rangle = 22.6$ (solid red) and $\langle n \rangle = 31.6$ (dashed blue) with $n = 6, 9, \ldots, 42$ and $n = 12, 15, \ldots, 54$, respectively. The solid red curves nearly obscure the dashed blue curves for $0 \le n \le 9$ in (a) and for many n's in (b), showing that the cluster means derived from independent data sets are nearly identical except where the number of waveforms in the data set is very few.

$$V_i(t) \approx \alpha_i \bar{V}_n(t) + (1 - \alpha_i)\bar{V}_{n'}(t), \tag{12}$$

where α_i is a constant. The value of α_i that minimizes the RMS deviations is found. The interpolated photon number is given by

$$n_i^{(\mathrm{eff})} = \alpha_i n + (1 - \alpha_i)n'. \tag{13}$$

The strong clustering can be seen by plotting a histogram of the cluster means such as Fig. 3 for the $\langle n \rangle = 2.00$ data set.

The largest illumination level for which some single-photon-number resolution could be found corresponded to $\langle n \rangle = 31.6$. The waveforms for these clusters are presented in Fig. 2(b); cluster means from $\langle n \rangle = 22.6$ are also shown to indicate that the method produces very similar curves as a function of individual photon number n-independent of the average photon number $\langle n \rangle$, as was the case in Fig. 2(a).

To illustrate the photon-number-resolving power, we classify the waveforms according to the dot-product method and the Poisson-influenced K-means algorithm (Figs. 3 and 4) and then quantify the visibility of the photon-number peaks (Fig. 5) using the Michelson definition of $(\max - \min)/(\max + \min)$, where the minima are taken at half-integer photon numbers and the maxima the average of the two surrounding values. Whereas some photon-number peak visibility persists (defined as a nonzero value that exceeds its uncertainty) up through $n = 19$ using the dot-product method, peak visibility persists through $n = 23$ for PIKA, showing its improvement in photon-number-resolving power. This is the key result of this paper.

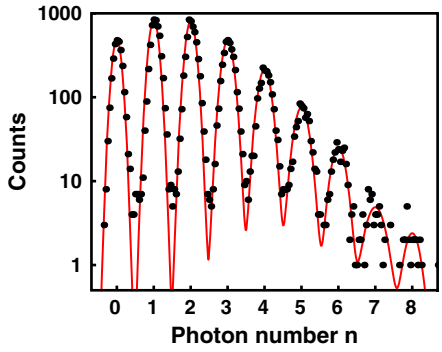

Fig. 3. (Color online) Counts of waveforms per bin (of width 0.05 photon) for 19 091 pulses (derived from 20 480 pulses after rejection of observations with blackbody photons) with an average of $\langle n \rangle = 2.00$ photons per pulse (blue points). The histogram was fit to a series of Gaussians centered on the integers leaving only the amplitude and width of each Gaussian as parameters (red line). Effective photon numbers $n^{(\mathrm{eff})}$ are found for each waveform using the Poisson-influenced K-means algorithm. Results for the dot-product algorithm (not shown) are very similar for this case.

Description of the cluster means. The waveforms grow in magnitude and width through $n = 15$ or so, and then grow principally in width for larger n. The full widths at half maximum (FWHM) as a function of the photon number n are presented in Fig. 6(a). A roughly linear behavior for the FWHM is seen for the photon numbers studied, although sublinear behavior occurs as the average photon number increases [24].

The saturation of the peak cluster means is not absolute as a cursory glance at Fig. 2(b) might suggest. The peaks continue to grow, as shown in Fig. 6(b). The two regimes in the plot represent the superconducting regime at low n

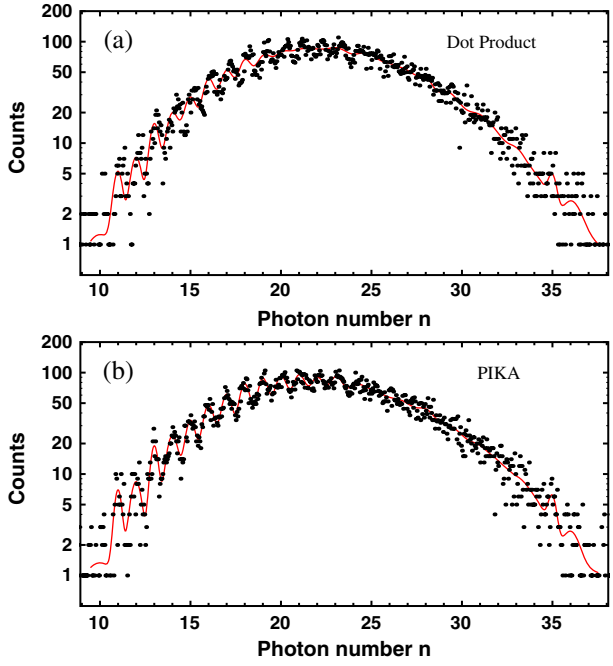

Fig. 4. (Color online) Histogram of waveforms versus (a) effective photon number found using the dot-product algorithm or (b) interpolated photon number found using the Poisson-influenced K-means algorithm, all for 20 480 pulses, an average of $\langle n \rangle = 22.6$ photons per pulse, and a bin width of 0.05 of a photon. The histograms were fit as in Fig. 3 (red line).

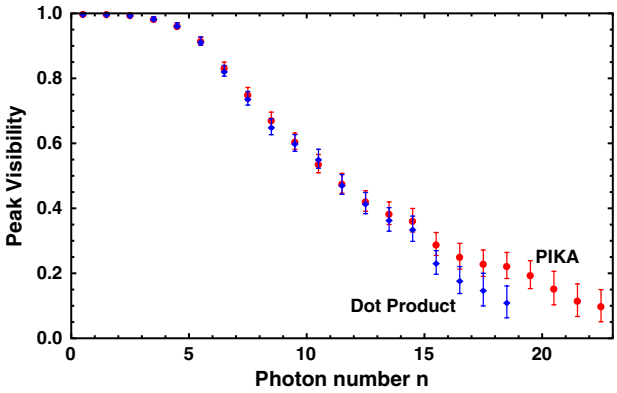

Fig. 5. (Color online) Peak visibility of the dot-product algorithm (blue diamonds) and PIKA (red circles) fits of the counts in Figs. 3 and 4 (along with others not shown), calculated using $(\max - \min)/(\max + \min)$ with minima taken at half-integer photon numbers and the maxima the average of the two surrounding values. The visibility uncertainties were obtained by assuming Poisson statistics for the counts using a Monte Carlo technique to sample variations in the fitted curve. Uncertainties are two standard deviations, and are purely statistical.

and the transition to normal conduction near $n = 15$. The trailing edges of each cluster mean are nearly identical, as illustrated in Fig. 7. The $n = 25$ cluster mean shifted in time by 3.2 μs is seen to have a nearly identical falling edge with the $n = 37$ cluster. (This behavior is typical of the cluster means.) The near equality of the falling edges is consistent with the idea that the detector's return from being driven normal from the heat of many photons should not depend on how long it spent in the normal state.

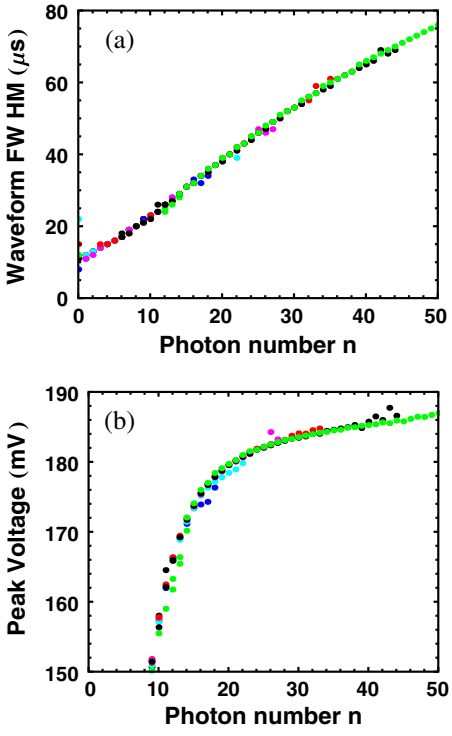

Fig. 6. (Color online) Full width at half maximum (a) and the peak voltage (b) of the waveform cluster means as a function of photon number. Curves are derived from the data sets with $\langle n \rangle = 2.00$, 2.84, 4.02, 5.68, 7.99, 11.3, 16.0, 22.6, and 31.6 (shown as different color points).

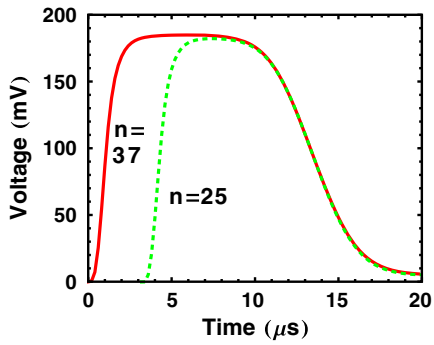

Fig. 7. (Color online) Waveform cluster mean corresponding to a photon number of $n = 37$ (red, solid) is compared to the $n = 25$ cluster mean (green, dotted) with the latter curve shifted right by 3.2 μs (taken from the $\langle n \rangle = 31.6$ data set). This common falling edge shape being independent of photon number is typical behavior for the cluster means.

5. CONCLUDING REMARKS

We have demonstrated, for one particular TES, the ability to distinguish photon-number peaks with up to 23 ($\lambda = 1550$ nm) photons—the highest value of which we are aware. In addition, we have quantified the degree of photon-number resolution in terms of visibility between adjacent peaks.

Individual TESs need to be calibrated. The method here provides a way to assist with such calibration. For small average photon numbers $\langle n \rangle$ (say, $\langle n \rangle < 8$), the agreement of the advanced algorithm with the simpler one indicates that sorting by a fixed dot product is sufficient. As seen in Fig. 5, both methods lead to nearly identical peak visibility values of 0.8 or more. However, the agreement of the two methods does increase the confidence in the results of the simpler method.

For larger $\langle n \rangle$, the clusters are less pronounced. Here, the Poisson-influenced K-means algorithm shows improved clustering over the dot-product algorithm (which is its starting guess) as evidenced by the additional peak visibility of about 0.1 for $n \geq 17$. We showed that the advanced algorithm outperformed the simpler ordering algorithm. Specifically, the maximum photon peak number visible in the histograms increased from 19 photons to 23 photons using the same data. This increased knowledge of the responses to individual photon numbers may also be helpful in thermal modeling of a detector [24].

The Poisson-influenced K-means algorithm itself may see general application to clustering problems. We introduced a variant of the K-means algorithm that is aware of the cluster sizes through the Poisson distribution, although another distribution could be substituted. Within the K-means algorithm or for its generalization, the development in the appendix shows how to update the objective function in a time that is independent of the cluster size, and that takes into account the effect of the change of the cluster means when a particular waveform is shifted to another cluster.

APPENDIX A: OBJECTIVE FUNCTION UPDATES

The most computation-time consuming part of the optimization procedure is the update of the objective function in Eq. (6), moving a single waveform from one cluster to another. Here, we show how to do that in a time that is independent of the size of the clusters. Our implementation takes into account the effect of a waveform leaving a cluster on the cluster

means. Although the effect is usually small, keeping track of it avoids certain cases in which a waveform is transferred between clusters in an infinite loop.

UPDATE OF *K*-MEANS OBJECTIVE FUNCTION

Let us decompose the K-means objective functions as

$$O_K = \sum_{n=n_0}^{n_0+K-1} J_n, \tag{A1}$$

where

$$J_n = \sum_{i\in C_n} \sum_t [V_i(t) - \bar{V}_n(t)]^2. \tag{A2}$$

If we transfer a waveform j into the cluster $n = a$, the new cluster members are in the set $C_a^+ = C_a \cup \{j\}$. It may be shown that

$$J_a^+ = J_a + \frac{m_a}{m_a + 1} \sum_t [V_j(t) - \bar{V}_n(t)]^2. \tag{A3}$$

If we transfer a waveform j out of the cluster with $n = b$, the new cluster members $C_b^- = C_b - \{j\}$ obey

$$J_b^- = J_b - \frac{m_b - 1}{m_b} \sum_t [V_j(t) - \bar{V}_n(t)]^2. \tag{A4}$$

In both Eq. (A3) and Eq. (A4), $\bar{V}_n(t)$ refers to the original cluster, i.e., before the transfer of j. After a waveform is transferred, the cluster means are updated with

$$\bar{V}_n(t) = \frac{m_n \bar{V}_n(t) \pm V_j(t)}{m_n \pm 1} \tag{A5}$$

for $n = a$ (+ signs) or $n = b$ (− signs). In the spirit of K-means, we impose the rule that no cluster is allowed to become empty, so the denominator in Eq. (A5) never vanishes.

In our implementation, whenever the calculation requires the summed squared deviation of a waveform from a cluster mean, it is stored with a time stamp. Whenever a waveform is transferred, a time stamp of two cluster means is updated. Before a calculation of $\sum_t [V_j(t) - \bar{V}_n(t)]^2$, the code checks to see if it has evaluated the quantity since the last time the cluster mean was changed. This rule is a particular time saver toward the end of the optimization when few waveforms are transferred.

We emphasize that none of the formulas presented in this section require a summation over the waveforms in a cluster to move a single element.

UPDATE OF POISSON LOG-LIKELIHOOD

From Eq. (9),

$$\ln \mathcal{L}_P(m_{n_0}, \ldots, m_a, \ldots, m_b, \ldots, m_{n_0+K-1}; \mu) = -\mu M + \sum_{n=n_0}^{n_0+K-1} m_n [n \ln \mu - \ln(m_n!)]. \tag{A6}$$

Suppose a waveform is transferred to the cluster with $n = a$ from the cluster with $n = b$. The change in the objective function will contain a term

$$\begin{aligned} &\ln \mathcal{L}_P(m_{n_0}, \ldots, m_a + 1, \ldots, m_b - 1, \ldots, m_{n_0+K-1}; \mu) \\ &\quad - \ln \mathcal{L}_P(m_{n_0}, \ldots, m_a, \ldots, m_b, \ldots, m_{n_0+K-1}; \mu) \\ &\quad = (b - a) \ln \mu + (m_a + 1) \ln(m_a + 1) + \ln(m_a!) \\ &\quad\quad - (m_b) \ln(m_b) - \ln[(m_b - 1)!]. \end{aligned} \tag{A7}$$

Although the notation suggests $b > a$, the case $b < a$ is given by the same formula.

The treatment of the combinatorial factor, starting from Eq. (10), is similar. The argument is elementary, so only the result is stated here:

$$\Delta \ln \mathcal{L}_C = \frac{m_b}{m_a + 1}, \tag{A8}$$

where $\Delta \ln \mathcal{L}_C$ stands for the change in the combinatorial term analogous to the one given in Eq. (A7).

Our implementation in Mathematica [30] runs in about 4 min for the greedy algorithm for 20 480 waveforms with 100 time steps each. Simulated annealing runs a few times longer, depending on the annealing schedule.

ACKNOWLEDGMENTS

We are pleased to acknowledge assistance from Steven Conn and Adam Pintar.

REFERENCES AND NOTES

1. A. Migdall, "Differences explained in correlated-photon metrology techniques," Phys. Today **52**, 41–46 (1999).
2. A. P. Worsley, H. B. Coldenstrodt-Ronge, J. S. Lundeen, P. J. Mosley, B. J. Smith, G. Puentes, N. Thomas-Peter, and I. A. Walmsley, "Absolute efficiency estimation of photon-number-resolving detectors using twin beams," Opt. Express **17**, 4397–4411 (2009).
3. A. R. Beaumont, J. Y. Cheung, C. J. Chunnilall, J. Ireland, and M. G. White, "Providing reference standards and metrology for the few photon counting community," Nucl. Instrum. Methods Phys. Res. A **610**, 183–187 (2009).
4. J. C. Zwinkels, E. Ikonen, N. P. Fox, G. Ulm, and M. L. Rastello, "Photometry, radiometry and 'the candela': evolution in the classical and quantum world," Metrologia **47**, R15–R32 (2010).
5. R. Klein, R. Thornagel, and G. Ulm, "From single photons to milliwatt radiant power in electron storage rings as radiation sources with a high dynamic range," Metrologia **47**, R33–R40 (2010).
6. H. J. Briegel, W. Dur, J. I. Cirac, and P. Zoller, "Quantum repeaters: the role of imperfect local operations in quantum communication," Phys. Rev. Lett. **81**, 5932–5935 (1998).
7. E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics," Nature **409**, 46–52 (2001).
8. S. Dolinar, "An optimum receiver for the binary coherent state quantum channel," Massachusetts Institute of Technology Research Laboratory of Electronics Quarterly Progress Report, Vol. 111, 115–120 (1973).
9. C. Wittmann, M. Takeoka, K. N. Cassemiro, M. Sasaki, G. Leuchs, and U. L. Andersen, "Demonstration of near-optimal discrimination of optical coherent states," Phys. Rev. Lett. **101**, 210501 (2008).
10. F. E. Becerra, J. Fan, G. Baumgartner, S. V. Polyakov, J. Goldhar, J. T. Kosloski, and A. Migdall, "*M*-ary-state phase-shift-keying discrimination below the homodyne limit," Phys. Rev. A **84**, 062324 (2011).

11. S. V. Polyakov, A. Migdall, and S. W. Nam, "Real-time data-acquisition platform for pulsed measurements," in *Advances in Quantum Theory*, *AIP Conference Proceedings*, Vol. 1327 (Växjö, 2010), pp. 505–519.
12. While always keeping in mind that this correspondence is never 100% assured, we refer to these as photon-number peaks.
13. M. Mehmet, A. Ast, T. Eberle, S. Steinlechner, H. Vahlbruch, and R. Schnabel, "Squeezed light at 1550 nm with a quantum noise reduction of 12.3 dB," Opt. Express **19**, 25763–25772 (2011).
14. A. J. Miller, S. W. Nam, J. M. Martin, and A. V. Sergienko, "Demonstration of a low-noise near-infrared photon counter with multiphoton discrimination," Appl. Phys. Lett. **83**, 791–793 (2003).
15. M. Fujiwara and M. Sasaki, "Direct measurement of photon number statistics at telecom wavelengths using a charge integration photon detector," Appl. Opt. **46**, 3069–3074 (2007).
16. P. Buzhan, B. Dolgoshein, A. Ilyin, V. Kaplin, S. Klemin, R. Mirzoyan, E. Popova, and M. Teshima, "The cross-talk problem in SiPMs and their use as light sensors for imaging atmospheric Cherenkov telescopes," Nucl. Instrum. Methods Phys. Res. A **610**, 131–134 (2009).
17. R. H. Hadfield, "Single-photon detectors for optical quantum information applications," Nat. Photon. **3**, 696–705 (2009).
18. M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, "Single-photon source and detectors," Rev. Sci. Instrum. **82**, 071101 (2011).
19. D. Fukuda, G. Fujii, T. Numata, K. Amemiya, A. Yoshizawa, H. Tsuchida, H. Fujino, H. Ishii, T. Itatani, S. Inoue, and T. Zama, "Titanium-based transition-edge photon number resolving detector with 98 percent detection efficiency with index-matched small-gap fiber coupling," Opt. Express **19**, 870–875 (2011).
20. A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," Opt. Express **16**, 3032–3040 (2008).
21. C. Ding and X. He, "*K*-means clustering via principal component analysis," in *ICML Proceedings 21st International Conference on Machine Learning* (Assoc. Comp. Mach., 2004), pp. 1–9.
22. G. Hamerly and C. Elkan, "Alternatives to the *k*-means algorithm that find better clusterings," in *Proceedings of the Eleventh International Conference on Information and Knowledge Management* (ACM, 2002), CIKM '02, pp. 600–607.
23. W. H. Press, S. A. Teukolsky, W. T. Vettenberg, and B. P. Flannery, *Numerical Recipes*, 3rd ed. (Cambridge University, 2007), Sections 13.3 and 16.1.
24. T. Gerrits, B. Calkins, N. Tomlin, A. E. Lita, A. Migdall, S. W. Nam, and R. Mirin, "Extending single-photon optimized superconducting transition edge sensors beyond the single-photon counting regime," presented at the *Quantum Electronics and Laser Science Conference* (2012).
25. L. Cai, H. Y. Huang, S. Blackshaw, J. S. Liu, C. Cepko, and W. H. Wong, "Clustering analysis of SAGE data using a Poisson approach," Genome Biol. **5**, R51.1 (2004).
26. P. E. Black, "Greedy algorithm," in Dictionary of Algorithms and Data Structures [online], P. E. Black, ed., U. S. National Institute of Standards and Technology (2 February 2005) (accessed March 5, 2012). Available from http://www.nist.gov/dads/HTML/greedyalgo.html.
27. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).
28. D. J. Fixsen, S. H. Moseley, B. Cabrera, and E. Figueroa-Feliciano, "Pulse estimation in nonlinear detectors with nonstationary noise," Nucl. Instrum. Methods Phys. Res. A **520**, 555–558 (2004).
29. Curves were classified as containing background photons if the initial or final voltage was above 16.8 mV, or if there was a peak of at least 16.8 mV after a time of 3.4 μs. The number of curves removed was consistent with the background rate given above.
30. The mention of commercial products does not imply endorsement by the authors' institutions nor does it imply that they are the best available for the purpose.