1. Introduction

Integrated quantum photonics has recently emerged as one of the key enablers for the quantum information science and technology (QIST). Typically, quantum photonic circuits are realized using nonlinear sources of single photons that operate probabilistically and do not allow the generation of large multi-photon states. As a result, the number of photonic qubits in quantum circuits is currently limited to 20. Alternatively, solid-state quantum emitters have recently reached near-ideal single-photon characteristics. Successful implementation of quantum photonic circuits based on quantum emitters depends crucially on the selection of these emitters from a large inhomogeneous set. It requires efficient identification of bright, stable single-photon emitters with fast emission rates, high quantum yield, and narrow optical linewidth. Since a universal platform for integrated quantum photonics is lacking, hybrid integration methods are being actively explored.

Recently, deterministic atomic force microscopy-assisted assembly has been utilized to realize hybrid nano-structures for ultrafast emission of single-photons at room temperature. As a rule, micromanipulation methods are needed to integrate devices across different material platforms. With the growing interest in scalable realization and rapid prototyping of quantum devices, high-speed, robust classification of “single” or “not-single” emitters becomes of utmost significance.

For optical quantum emitter characterization, photon antibunching measurements using second-order auto-correlation have long been used to measure the single-photon purity of the emission. Such experiments are typically implemented using a Hanbury-Brown-Twiss (HBT) interferometer, composed of a beamsplitter directing the emitted light to two single-photon detectors connected to a correlation board. The correlation board collects pairs of clicks generated by different detectors (detector 1, then detector 2 or vice versa) and bins these co-detection events according to the time delay $\tau$ between the detectors' clicks. Depending on which detector clicks first, the delay is either considered positive or negative. The distribution of co-detections as a function of $\tau$ is described by the second-order autocorrelation function of the emission $g^{(2)}(\tau)$. This distribution is normalized so that $g^{(2)}(\infty) = 1$, indicating that there is no correlation between the detection events spaced by an infinite time interval. In practice, the number of co-detection events collected with this method tends to 0 as $\tau$ increases. Therefore, the maximum $\tau$ in the measured distribution is kept at a finite value that is much smaller than the inverse photon detection rate, but much larger than the time constants of the emitter photodynamics.

Commonly, the metric for the single-photon purity of the emission is the value of $g^{(2)}(0)$, indicating the normalized number of coincidental detections. The measurement of single-photon purity is of paramount importance for characterizing quantum emitters. Besides describing the quality of quantum emission,
Figure 1. Rapid binary classification of quantum emitters enabled by machine learning. a) Fluorescence from diamond nitrogen vacancy (NV) sources (single centers and center ensembles) is analyzed by the Hanbury-Brown-Twiss (HBT) autocorrelator. The emitters are modeled as three-level quantum optical systems. b) Examples of autocorrelation datasets with different acquisition times: 1 s, 1 min, and 1 hr. The Levenberg–Marquardt (L–M) fitted complete datasets yield the actual values of $g^{(2)}(0)$ used for training and assessment of the classification accuracy. $N$ is an average co-detection counts per bin. c) Classifier network training. Sparse data collected from HBT measurement are used as a training set. The $g^{(2)}(0)$ values retrieved from L–M fit of a complete dataset from each corresponding emission source are supplied to the classifier in training as the ground truth. d) Machine learning-assisted rapid single photon emitter detection. The trained model classifies the previously unencountered emission sources as “single” and “not-single” emitters based on the sparse autocorrelation data.

...it contains information about the fluorescence lifetime and can be used to correct other data such as optical saturation curves and photon indistinguishability. Figure 1b shows examples of autocorrelation datasets with different acquisition times. In practice, the co-detection rate is proportional to the square of source intensity, which is very low for single-photon emitters. At the same time, the conventional statistical classification with the Levenberg–Marquardt (L–M) fit requires a sufficiently populated histogram to reliably retrieve the value of $g^{(2)}(0)$ (see Experimental Section, Supporting Information). As a result, determining $g^{(2)}(0)$ with sufficient accuracy is highly time-consuming, thus hindering the development of scalable techniques for photonic QIST device assembly and prototyping. Faced with the challenge of rapidly identify emitters with high single-photon purity from a large set, one may, however, define a heuristic threshold for the value of $g^{(2)}(0)$ and only strive to characterize the emitters that meet this threshold fully.

Theoretically, the autocorrelation at zero delay from an ensemble of $n$ identical, equally contributing emitters is equal to $1 - 1/n$, yielding $g^{(2)}(0) = 0$ for a single emitter. In practice, single-photon purity is always finite due to experimental imperfections, such as background radiation and detector dark counts, and thus the measured $g^{(2)}(0)$ is always strictly positive. Moreover, emitters may not be identical or equally contributing. These uncertainties may lead to any value of $g^{(2)}(0)$ between 0 and 1, thus further blurring the boundary between the “single-photon” and “classical” emission regimes. These regimes are still heuristically defined by $0 < g^{(2)}(0) < G$ and $g^{(2)}(0) \geq G$, respectively, where $G$ is a threshold level, chosen based on the user’s desired degree of discrimination. This level may account for the known contribution of background counts to the autocorrelation histogram. In the rest of the paper, unless explicitly mentioned, we consider $G = 0.5$ as our classification threshold. In the experimental section, we study the scalability of the proposed ML classification schemes for different values of G. Despite its definite necessity, high-speed, accurate emitter classification based on single-photon purity has so far remained elusive.

In the last decades, various machine learning (ML) methods have attracted significant interest in the optics community. Recently, ML algorithms have also been applied to quantum photonics. Combining the Bayesian phase estimation with Hamiltonian Learning techniques for analyzing large datasets...
from nitrogen vacancy (NV) centers in bulk diamond allowed for magnetic field measurements with extreme sensitivity at room temperature.\(^{[35]}\) Hamiltonian Learning was adopted for the characterization of different quantum systems,\(^{[36]}\) including the characterization of electron spin states in diamond NV centers.\(^{[37]}\) The development of autonomous adaptive feedback schemes incorporates decision mechanisms into quantum measurements.\(^{[38,39]}\) Yet, one of the most powerful applications of ML algorithms relates to object classification problems that encompass most quantum optics measurements, including emitter classification. Machine learning can dramatically speed up quantum measurements, thus transforming the area of quantum photonic testing, assembly, and prototyping.

In this work, we implemented supervised ML algorithms to classify single versus not-single photon emitters using the sparse autocorrelation data. We have applied four different supervised ML classifiers (support vector classification (SVC),\(^{[40]}\) gradient boosting classifier (GBC),\(^{[41]}\) voting classifier (VC),\(^{[42]}\) and convolutional neural network (CNN))\(^{[43]}\) to map autocorrelation measurement data into “single”/“not-single” categories. First, autocorrelation data was acquired on a set of emitters with an integration time of several minutes, allowing an accurate retrieval of \(g^{(2)}(0)\) using the standard L–M fitting. For these datasets that we refer to as “complete”, the L–M fitted \(g^{(2)}(0)\) values yield an error of about 0.03 and can be regarded as the ground truth. These complete datasets were further sectioned into “sparse” datasets with an integration time of 1 s. The classifier was trained (Figure 1c) on sparse datasets. Subsequently, sparse datasets from previously unencountered NVs were fed into the trained classifier, and the classification accuracy was statistically assessed (Figure 1d). Both the training of the ML classifier and the estimation of classification accuracy were based on the ground truth \(g^{(2)}(0)\) values retrieved from the L–M fits of the complete datasets. We show that ML-assisted classification remains highly accurate even for sparse data, whereas the conventional L–M fit applied to the same sparse data fails and performs no better than a random guess.

2. Results

2.1. Theoretical Framework

To determine the best ML-assisted algorithm for the sparse data classification, we first performed a numerical experiment. In the numerical experiment, the emulated autocorrelation data was fed into different types of ML algorithms targeting their classification based on a \(g^{(2)}(0)\) threshold value of 0.5. The autocorrelation data were obtained by emulating results of an HBT experiment with photophysical emitter parameters similar to those featured by the NV centers under investigation. Each emitter was modeled as a three-level system with an excited state (E) radiatively coupled to a ground state (Gr), and a metastable state (M) non-radiatively coupled to both E and Gr (Figure 1a). We simulated the autocorrelation experiment by counting co-detection events from two virtual “detectors” and binning these events according to the time delay between the two “detector” clicks in each co-detection event. An elementary numerical experiment was repeated until obtaining the desired number of such events (see Experimental Section, Supporting Information). To generate autocorrelation histograms with a range of \(g^{(2)}(0)\) values, we assumed that the emission was produced partly by a single quantum emitter and partly by photonic background presenting no autocorrelation features. To emulate the emission with a certain input level of antibunching \(g^{(2)}(0)\), we set the intensity fraction of the photonic background in the total emission to be \(r_{bg} = 1 - \sqrt{1 - g^{(2)}(0)}\). More details on the numerical experiment and the underlying model are provided in Section S1, Supporting Information.

The theoretical model included two variable parameters: The ground truth value of \(g^{(2)}(0)\), and the average co-detection counts per bin \(N\). The latter parameter was defined as the total number of co-detection events in the dataset divided by \(N_{\text{ buena}}\). We generated forty thousand datasets, with \(g^{(2)}(0)\) spanning the interval between 0.1 and 0.9 with a step of 0.04 and \(N\) spanning from 0.5 to 10 with a step of 0.5. One hundred datasets were generated for each combination of \(g^{(2)}(0)\) and \(N\). 70% of all the datasets were used for training, and the remaining 30% served to test performance of the classifiers. The accuracy of “single” versus “not-single” emitter classification is defined as

\[
\text{accuracy} = \frac{\text{correct}}{N_{\text{correct}}} \times 100\%
\]

We applied 4 different supervised machine learning classification techniques to the generated datasets: i) SVC, ii) GBC, iii) VC (a combination of logistic regression\(^{[44]}\) and k-nearest neighbors (k-NN) algorithm), and iv) CNN based binary classifier. The performance and detailed description of each model is provided in Section S2, Supporting Information. Here, we outline the operation and results of the two classification algorithms, CNN and voting classifier (VC), that showed the best performance.

The CNN binary classifier consists of one input layer, three hidden convolutional layers, one max-pooling layer followed by two fully connected layers. The input layer has a dimension of \(N_{\text{ buena}}\), the same as the dimension of the autocorrelation dataset. The three hidden layers of the CNN classifier are extracting the main features of this dataset, while the preceding two fully connected layers perform binary classification of the dataset based on these main features. The training process was performed with the stochastic gradient descent optimization of the weight using the dataset and corresponding ground truth values (“single” or “not-single” based on \(g^{(2)}(0)\) values). More details on the CNN performance and “decision making” process can be found in Section S5, Supporting Information.

Along with the CNN, we adapted the VC method built on logistic regression (LG) and k-NN algorithm. The main idea of the VC is to average the results of several pre-trained different classifiers. The voting classification is realized in two steps. First, the LG and k-NN classifiers are independently trained on the same ensemble of autocorrelation datasets. Then, both trained models are applied to the test ensemble of datasets. The voting classification is obtained by weighted averaging of the two outputs with 2 to 1 weight ratio between the LG and k-NN classifiers.

Figure 2a shows the accuracy map of the CNN-based classifier plotted as a function of \(g^{(2)}(0)\) and \(N\). For comparison, we also plotted the classification accuracy of the L–M fit for the same ensemble of datasets (Figure 2b). The black curve shows the contour line corresponding to 75% accuracy. The CNN-based classifier
enables over 75% accuracy within a larger parameter space area, whereas the L–M fit i) performs poorly for the datasets with small $N$ and ii) breaks down completely in the region of $N < 3$. The asymmetry of the L–M fitting accuracy in this region comes from the large uncertainty in the normalization of sparse datasets, which skews the fit towards larger values. The classification performance of the L–M fit and CNN-based method is further analyzed in Figure 2c. The figure presents the cross-sections of the accuracy maps of Figure 2a,b for constant values of $N = 0.5, 5$ and 9. In nearly all the cases, CNN outperforms L–M fitting. For $N = 0.5$, L–M fitting leads to unbalanced accuracy distribution between “single” and “not-single” classes resulting in <50% mean accuracy, while the CNN classifier ensures 70% accuracy in this region. A trough in accuracy consistently occurs close to the decision boundary between two classes, corresponding to $g^{(2)}(0) = 0.5$. With increasing $N$, the accuracy trough becomes expectedly narrower, and in the limit of high $N$ both techniques (CNN and L–M) asymptotically arrive at 100% classification accuracy (Section S2, Supporting Information). The observed accuracy drop happens due to the similarity of the sparse autocorrelation histograms near the decision boundary region for the cases with lower $N$ values, which leads to the reduced performance of the ML classifiers. Figure 2d compares the classification accuracy of the L–M fit, CNN and VC methods, averaged over all of the $g^{(2)}(0)$ values as a function of $N$. The CNN-based classifier strongly overperforms the L–M fit and features slightly better performance in comparison with the VC method.

Another important aspect of the classification is accuracy distribution between classes since imbalanced classification leads to biased predictions and misleading classification accuracies. Figure 2e shows the accuracy distribution between “single” and “not-single” classes for the CNN-based classifier, VC method, and L–M fits. The CNN-based classifier exhibits a more balanced accuracy distribution between “single” and “not-single” classes in comparison with the VC method, while the L–M fitting shows imbalanced performance for both sparse data region ($N < 3$) and region with $5 < N < 9$.

For $N < 3$ the L–M fit has a low mean accuracy of $57 \pm 3.7\%$ (53 $\pm$ 5% for “single”, 61 $\pm$ 1.3% for “not-single” emitters). The CNN classifier has $\approx 73\% \pm 3\%$ mean accuracy rate (71 $\pm$ 3% for “single”, 75 $\pm$ 3.1% for “not-single” emitters), while the VC has a 74% $\pm$ 2.4% (70% $\pm$ 2% and 76.8% $\pm$ 2.9% for “single”/“not-single” emitters, respectively). Along with the accuracy score additional performance metrics of classification, such as precision, recall, F1-score, and confusion matrices are analyzed (Section S4, Supporting Information). Our analysis clearly shows that on a broad range of emitter parameters, ML-based approaches perform significantly better than the L–M fit. The performance difference is especially striking for the sparsest datasets, which is particularly important for speeding up the single-photon emitter characterization.

### 2.2. Experimental Emitter Classification

Having selected the best performing network architectures, we test both the CNN and VC classifiers for emission classification from physical nanodiamond-based NV centers. In the
experiment, we used nanodiamonds (Figure 3a inset) randomly dispersed on a coverslip glass substrate with at least some of them containing single or multiple NV centers. Two avalanche detectors (D1, D2) featuring about 30 ps jitter were used for single-photon autocorrelation measurements. Time-correlated photon counting was performed by a correlation card with a 4 ps internal jitter. The total histogram span was set to 500 ns, and the co-detection events were collected into 215 equally sized time bins. The first 171 bins collected co-detections with a negative delay (D2 clicks first), while the remaining 43 bins corresponded to a positive delay (D1 clicks first). Thus, the maximum absolute observable delay in a co-detection event was \( t_{\text{max}} \approx 400 \) ns. Figure 3a summarizes the experimental setup.

Photon autocorrelation measurements were performed on a set of 40 emitters. For each emitter, the sparse datasets were compiled into a complete dataset, from which the ground truth \( g^{(2)}(0) \) value was attained using the L–M fitting algorithm. These fitted values of \( g^{(2)}(0) \) on complete datasets were appended to each corresponding sparse dataset as a label and used as the ground truth for the training/testing purposes. The L–M fitting uncertainty for the complete datasets varied from \( \pm 0.01 \) to \( \pm 0.05 \) (with an average of \( \pm 0.03 \)).

In total, 9416 sparse physical datasets were collected from 15 “single” and 25 “not-single” emitters (as determined by L–M fit of their complete datasets). The number of datasets acquired for each emitter is represented by the diameter of a corresponding circle; the circles are then plotted as a function of their \( N \) and \( g^{(2)}(0) \) in Figure 3b. Most of the experimental datasets exhibit \( N < 5 \) and fall in the range where a significant improvement in classification accuracy is expected from ML-assisted methods. More details regarding the experimental datasets (dataset size, ground truth values, the fitting errors for each emitter) are given in Section S7, Supporting Information.

The analysis of dataset size influence on the performance of CNN and VC classifiers reveals that the CNN based approach requires a larger amount of training data in comparison with VC method (Section S3, Supporting Information). Moreover, for balanced training of the VC and CNN classifiers, it is necessary to have an equal amount of datasets for both classes. Since the “single” emitter class has smaller amount of dataset, the datasets for this class were upsampled to 30% using the bootstrapping method during the training phase.

To assess the classification accuracy of a given emitter, we trained the network on all the datasets, except for those belonging to that given emitter. The datasets corresponding to the emitter of interest were then used for classification testing. That approach provided significant data augmentation and allowed for better training of the classifiers.

Figure 3c shows the accuracy comparison between the CNN, VC, and L–M fitting methods as a function of \( g^{(2)}(0) \). For all the

---

**Figure 3.** Classification of physical autocorrelation data. a) Schematics of the HBT interferometer. Labels: DM—dichroic mirror; LPF—long-pass filter; BS—beam splitter; D1/D2—detectors. Insets show typical photoluminescence map of nanodiamonds on glass (left) and the schematics of the NV center in nanodiamond (right). b) Distribution of datasets as a function of \( g^{(2)}(0) \) and \( N \). Each circle corresponds to an ensemble of datasets for a given emitter and therefore it shares the same values of \( g^{(2)}(0) \) and \( N \). The circle radius is proportional to the number of sparse datasets acquired for the corresponding emitter. c) Comparison of classification accuracy by the CNN (blue bars), VC (brown bars), and L–M fitting (yellow bars) methods for each emitter. \( g^{(2)}(0) \) for each emitter are displayed on the horizontal axis. Bar plot on the right shows mean accuracy of three methods averaged over all emitters classification. d) Accuracy as a function of average count per bin values of CNN (black squares), voting classifier (blue markers), and L–M fitting (red markers).
emitters CNN and VC classifiers expectedly outperform the L–M fitting approach. Noting that the drop in the accuracy of CNN and VC classification happens around the decision boundary region \( g^{(2)}(0) \approx 0.5 \). However, the accuracy trough is not as pronounced as expected from the numerical experiment. This effect can be explained by the fact that emitters with \( g^{(2)}(0) \) close to 0.5 have relatively high emission rates \( N > 4.4 \) (Section S7, Supporting Information), which allows a more accurate classification. For the “single” (“not-single”) emitters the CNN classifier ensures 88.9% ± 1.2% (97.6% ± 2.0%), VC approach yields 82.6% ± 1.5% (96.5% ± 2.2%) accuracy, while the L–M fit only attains 44.6% ± 4.0% (58.0% ± 4.5%), which is no better than a random guess within statistical uncertainty. The bar plot on the right shows the mean accuracy of all three methods averaged over all emitters.

Figure 4 shows the accuracy distribution for all 40 emitters for \( G = 0.3 \) and \( G = 0.4 \). Blue, orange, and yellow bars correspond to the CNN, VC classifiers, and the L–M fit, respectively. The accuracy dip of the ML-assisted classifiers stays expectedly aligned with the decision boundaries. In the case of \( G = 0.3 \), the CNN and VC classification schemes ensure 97% and 94% accuracy correspondingly, while in the case of \( G = 0.4 \) the accuracies of both ML methods are equal to 98% and 95%, respectively. The L–M fit features 55% accuracy in both cases that is comparable to the case of \( G = 0.5 \). Additional performance metrics of the CNN and VC classifiers for both decision boundary cases (\( G = 0.3 \) and \( G = 0.4 \)) can be found in Section S4, Supporting Information.

3. Discussion

Photon correlation measurements are at the heart of many quantum optical experiments. These measurements require collecting statistics over relatively rare co-detection events and therefore are inherently slow. Obtaining the desired information from a limited number of such co-detection events would be
critical for characterizing and assembling quantum optical systems pitting the way to scalable integrated quantum circuits.

We have implemented machine learning classification methods for rapidly identifying single emitters based on the value of the photon autocorrelation at zero delay. We analyzed sparse autocorrelation data and performed binary statistical classification (“single”/“not-single” emitter) on sparse datasets that cannot be characterized by a conventional fitting method. We demonstrated that by applying ML-assisted classifiers on the sparse datasets acquired within 1 s it was possible to achieve over 90% accuracy. The conventional fitting method applied to the same sparse datasets performed no better than a random guess and required a two orders of magnitude longer collection time to reach an accuracy of 95%. Our method can be further improved by taking into account the background fluorescence from the glass substrate. In our experiment, this background fluorescence accounted on average for 7% of the total fluorescence intensity. Correcting the complete histograms in the training set for this background emission could lead to a better prediction of intrinsic emitter behavior.

The proposed approach has the potential to dramatically advance most quantum optical measurements that can be formulated as a binary or multi-class classification. It can also be straightforwardly extended into a multi-bin classification, providing predictive estimates of a(0) faster than any conventional fitting algorithm. In addition, our approach could also transform higher-order autocorrelation measurements because their datasets can be even more sparse for the same acquisition times. As an example application, our technique can speed up super-resolution microscopy based on single-photon autocorrelation measurements, that are currently limited by long image acquisition times. For the precise determination of the autocorrelation value at the zero delay however, one would need to employ regressive machine-learning techniques. The developed CNN classifier can be reshaped into the regression model by modifying the loss function of the fully connected layer accordingly. However, the precision of such a regression scheme and the required acquisition time of the autocorrelation measurement is to be investigated and will be addressed elsewhere.

Recently, quantum photonics circuits with over a 100 quantum emitters have been realized for scalable quantum communication. Heterogeneous pick-and-place integration allows to realize such circuits by fabricating emitter arrays on a separate material platform, then transferring and coupling them one-by-one to photonic chips. The proposed characterization method can readily be applied for large-scale probing of emitter arrays and rapid selection of the most promising candidates. It therefore represents a step towards the realization of quantum circuits containing thousands of emitters.

Supervised machine learning algorithms could be efficiently applied to characterize other properties of quantum emitters. For example, the NV center’s electron and nuclear spin states can be optically read out through spin-dependent fluorescence intensity. Single-shot readout of the electronic spin at room temperature is a highly-sought-after goal, which could be achieved using classification ML approaches. In combination with the recently proposed idea of plasmonic cavity-single photon emitter integration that significantly increases the photon count rate of the NV centers, our ML-assisted approach can open up a road for room temperature single-shot sensitive spin readout.

The proposed approach could also have a strong impact on single-molecule spectroscopy. The single-molecule optical signal is often unstable and could be lost for extended periods due to long-term shelving in non-radiative states. The use of ML-assisted signal processing can help with acquiring the necessary information from optically unstable emitters before the optical signal is lost.

To conclude, we demonstrated that supervised machine learning can significantly advance quantum measurements allowing for rapid single-photon source detection. Such an accelerated detection could enable scalable, rapid quantum device testing, assembly and prototyping, as well as real-time, precision metrology in quantum materials.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
Z.A.K. and S.I.B. contributed equally to this work. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Award DE-SC0017717 (S.I.B. and V.M.S.). DARPA/DSO Extreme Optics and Imaging (EXTREME) Program (HR00111720032, Z.A.K. and A.V.K.). This work was also supported in part by NSF EECS award 2015025.

Conflict of Interest
The authors declare no conflict of interest.

Author Contributions
Z.A.K., S.I.B., V.M.S. conceived the experiment. A.V.K., V.M.S, and A.B. supervised the project. Z.A.K. developed and wrote machine learning classification codes, performed neural network training, produced and optimized the classification results. Z.A.K and S.I.B. performed Monte-Carlo simulations of autocorrelation experiments. S.I.B. and T.I. performed HBT measurements on NV centers. All authors interpreted the data. Z.A.K., S.I.B., A.V.K., and A.B. wrote the manuscript.

Keywords
single photon sources, machine learning, quantum emitter classification

Received: June 22, 2020
Revised: August 2, 2020
Published online: September 2, 2020

[1] M. Simmons, npj Quantum Inf. 2015, 1, 15013.
[2] S. Bogdanov, M. Y. Shalaginov, A. Boltasseva, V. M. Shalaev, Opt. Mater. Express 2017, 7, 111.
[3] H. Wang, J. Qin, X. Ding, M.-C. Chen, S. Chen, X. You, Y.-M. He, X. Jiang, L. You, Z. Wang, C. Schneider, J. J. Renema, S. Höefling, C.-Y. Lu, J.-W. Pan, Phys. Rev. Lett. 2019, 123, 250503.
