Adversarial autoencoder ensemble for fast and probabilistic reconstructions of few-shot photon correlation functions for solid-state quantum emitters

Andrew H. Proppe, Kin Long Kelvin Lee, Cristian L. Cortes, Mari Saif, David B. Berkinsky, Tara Sverko, Weiwei Sun, James Cassidy, Mikhail Zamkov, Taehyung Kim, Eunjoo Jang, Stephen K. Gray, Brett A. McGuire, and Moungi G. Bawendi

Phys. Rev. B **106**, 045425 — Published 29 July 2022

DOI: 10.1103/PhysRevB.106.045425

An Adversarial Autoencoder Ensemble for Fast and Probabilistic Reconstructions of Few-Shot Photon Correlation Functions for Solid-State Quantum Emitters

Andrew H. Proppe,[1,†] *Kin Long Kelvin Lee,*[1,2,†] *Cristian L. Cortes,*[2] *Mari Saif,*[1] *David B. Berkinsky,*[1] *Tara Sverko,*[1] *Weiwei Sun,*[1] *James Cassidy,*[3] *Mikhail Zamkov,*[3] *Taehyung Kim,*[4] *Eunjoo Jang,*[4] *Stephen K. Gray,*[2] *Brett A. McGuire,*[1,*] *Moungi G. Bawendi*[1,*]

[1] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2] Center for Nanoscale Materials, Argonne National Laboratory, Lemont, Illinois 60439, USA

[3] The Center for Photochemical Sciences and Department of Physics, Bowling Green State University, Bowling Green, Ohio 43403

[4] Samsung Advanced Institute of Technology, Samsung Electronics, Suwon-si, Gyeonggi-do 16678, Republic of Korea

[†] These authors contributed equally to this work.

Corresponding authors: mgb@mit.edu, brettmc@mit.edu

Abstract

Second-order photon correlation measurements ($g^{(2)}(\tau)$ functions) are widely used to classify single photon emission purity in quantum emitters, or to measure the multiexciton quantum yield of emitters that can simultaneously host multiple excitations – such as quantum dots – by evaluating the value of $g^{(2)}(\tau = 0)$. Accumulating enough photons to accurately calculate this value is time consuming and could be accelerated by fitting of few-shot photon correlations. Here, we develop an uncertainty-aware, deep Adversarial Autoencoder Ensemble (AAE) that reconstructs noise-free $g^{(2)}(\tau)$ functions from noise-dominated, few-shot inputs. The model is trained with simulated $g^{(2)}(\tau)$ functions that are facilely generated by Poisson sampling time bins. The AAE reconstructions are performed orders of magnitude faster, with reconstruction errors and estimates of $g^{(2)}(\tau = 0)$ that are lower in variance and similar in accuracy compared to Maximum Likelihood Estimation and Levenberg–Marquardt least-squares fitting approaches, for simulated and experimentally measured few-shot $g^{(2)}(\tau)$ functions (~100 two-photon events) of InP/ZnS/ZnSe and CdS/CdSe/CdS quantum

dots. The deep ensemble model comprises 8 individual autoencoders, allowing for probabilistic reconstructions of noise-free $g^{(2)}(\tau)$ functions, and we show that the predicted variance scales inversely with number of shots, with comparable uncertainties to computationally intensive Markov Chain Monte Carlo sampling. This work demonstrates the advantage of machine learning models to perform uncertainty-aware, fast, and accurate reconstructions of simple Poisson distributed photon correlation functions, allowing for on-the-fly reconstructions and accelerated materials characterization of solid-state quantum emitters.

I. Introduction

Correlating the intensities of photons impinging on two detectors in a Hanbury Brown-Twiss configuration (Fig. 1a) allows for the creation of a second-order intensity correlation function, $g^{(2)}(\tau)$ (hereafter referred to as $g^{(2)}$ for succinctness). This technique can be used to characterize photon statistics and determine whether or not the emissive species is a single emitter by evaluating the value of $g^{(2)}(\tau = 0)$ [1,2], with antibunching being associated with a single photon emitter. For systems that can host multiple excitations and therefore emit multiple photons following a single laser pulse, like semiconductor quantum dots (QDs, Fig. 1b), $g^{(2)}(\tau = 0)$ is instead indicative of its multiexciton quantum yield – which can be distinguished from emission by multiple particles by applying a time gate to the photon stream [1,3]. Photon correlation functions are also used to classify the indistinguishability of single photons and [4] the entanglement fidelity of photon pairs [5], in quantum networks [6–8], quantum spectroscopy [9,10], and enhancing resolution in quantum imaging [11–13].

The most commonly considered multiexciton state is the biexciton, which can recombine radiatively or nonradiatively (e.g. via Auger recombination) to the exciton state (Fig. 1c) [14]. Radiative recombination of both the biexciton and exciton creates a two-photon event triggered by a single laser pulse, contributing to the center peak at $g^{(2)}(\tau = 0)$ in the absence of spectral filtering. Two-photon events where each detection occurs following a different laser pulse instead contribute to side peaks at $g^{(2)}(\tau = n\Lambda)$, where n is the number of laser pulses separating the photons and Λ is the laser repetition rate. The ratio of the areas or heights of the center peak and first side peak gives the biexciton quantum yield (BXQY) [3]. Examples of $g^{(2)}$s for the two dots considered in this study – InP/ZnSe/ZnS and CdS/CdSe/CdS – are shown in Fig. 1d.

The BXQY of QDs is an important property for their usage in high-flux applications like light-emitting diodes or lasers that can benefit from high BXQY [15]. In the domain of quantum optics, antibunched single photon emission requires QDs with near-zero BXQY [16,17], whereas generation of polarization-entangled

photon pairs via the biexciton-exciton cascade requires near-unity BXQY [4,18]. InP/ZnSe/ZnS QDs are emergent materials that are highly emissive and Cd-free, and have been used to make LEDs with external quantum efficiencies exceeding 20% [19]. Auger recombination is one performance limiting factor in these devices [20], and so understanding multiexciton quantum yields of single QDs is imperative. CdS/CdSe/CdS QDs – also called colloidal quantum wells – are nanostructures that embed the emissive CdSe layer between and inner and outer shell of CdS, which increases the confinement volume of biexcitons and suppresses Auger recombination [21]. Cassidy et al. have recently shown that these CdS/CdSe/CdS dots exhibit the longest optical gain lifetime (>6 ns) reported for colloidal QDs to date [22].

For weakly emissive species or optical configurations with low collection efficiency, the accumulation of a sufficient number of two-photon events or shots to construct a clean $g^{(2)}$ can take up to several hours. This makes the characterization of numerous single emitters a time-consuming process, or prohibitive for emitters that easily degrade under photoexcitation [23]. Experimental throughput could be accelerated via a denoising or signal reconstruction approach, whereby rapidly accumulated, few-shot $g^{(2)}$ data could be used to reconstruct an estimate of the true underlying function (i.e. the well-accumulated $g^{(2)}$). This can be done using fitting approaches like the Levenberg–Marquardt (LM) algorithm or Maximum Likelihood Estimation (MLE). Recently, Cortes et al. demonstrated that using an alternative Bayesian maximum *a posteriori* approach with a Poisson likelihood offered higher accuracy and lower variance for fitting $g^{(2)}$ of single QDs [23].

Such conventional approaches rely on assigning a model to fit to the data, which imparts bias onto the predicted result, and model selection can be difficult for noisy data. Machine learning (ML) approaches can remove the need for model selection, by instead incorporating multiple different possible models into the training data. ML approaches have also been shown to achieve higher accuracies with fewer measurements (in our case, photons collected) compared to conventional fitting, accelerating experimental throughput. For example, convolution neural networks have been used to classify quantum emitters based on sparse $g^{(2)}$ data of diamond nitrogen vacancies, realizing a ~100-fold speedup compared to LM fitting [24]. Neural networks have also been employed to distinguish thermal and coherent light sources using few measurements with low mean photon numbers [25], to identify nonclassicality of quantum light sources [26,27], and to perform reconstructions in quantum state tomography [28], also achieving orders of magnitude in speedup [28]. Machine and statistical learning approaches are increasingly being used to accelerate other types of spectroscopic experiments like ultrafast transient absorption [29], two-dimensional electronic spectroscopy [30], and vibrational spectroscopy [31].

In this work, we sought to further expand upon such advances in statistical and machine learning by developing a probabilistic neural network model that could be used for signal reconstruction of the entire

$g^{(2)}$ using few-shot data as an input, and evaluating its performance against conventional methods like LM and MLE.

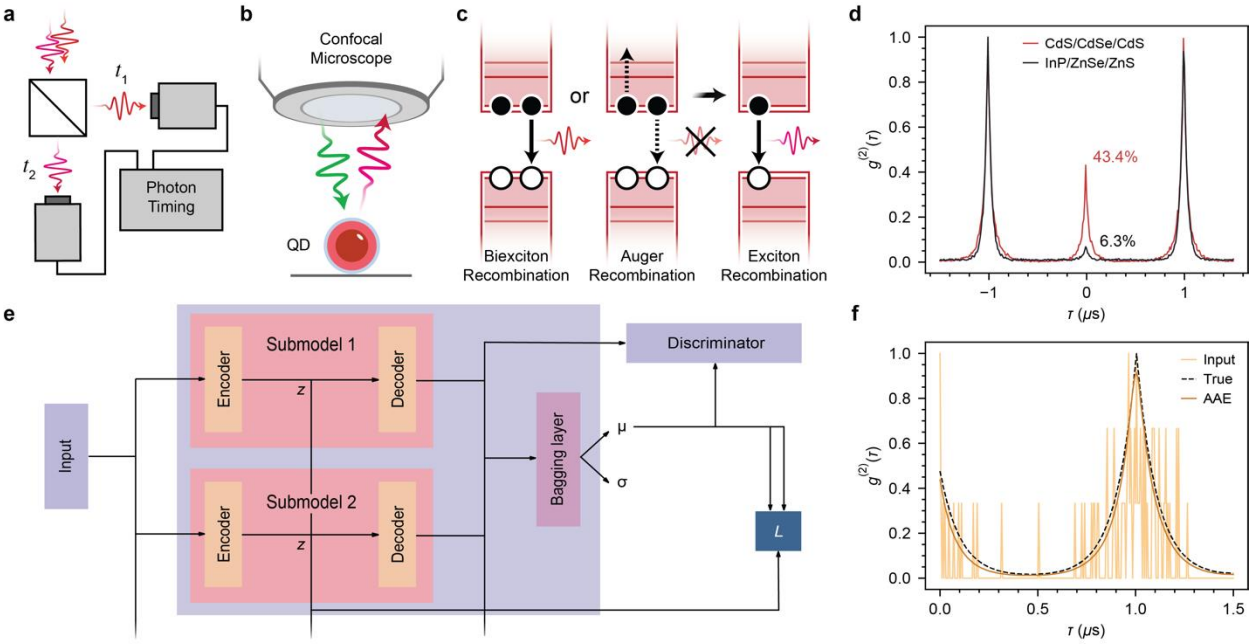

Figure 1. (a) A Hanbury Brown-Twiss optical setup, where photons emitted from the sample are passed through a 50:50 beamsplitter towards two single photon detectors in order construct $g^{(2)}$. **(b)** Illustration of confocal microscope used to probe single QDs. A detailed description of the full optical setup can be found in the Supplemental Material [32]. **(c)** Biexcitons can recombine radiatively or nonradiatively (Auger) before emission of the remaining exciton, which determines the BXQY (area of the $g^{(2)}(\tau = 0)$ peak). **(d)** Experimental $g^{(2)}$s for single CdS/CdSe/CdS (BXQY = 43.4%) and InP/ZnSe/ZnS (BXQY = 6.3%) dots. **(e)** General architecture of the Adversarial Autoencoder Ensemble (AAE) network used in this work. Only the first two of eight submodels are shown, with the black lines extending down towards the remaining six submodels. The output of each individual submodel is sent to the discriminator and a bagging layer to obtain the mean (μ) and variance (σ) of the output. **(f)** Simulated few-shot $g^{(2)}$ input with a $g^{(2)}(\tau = 0)$ peak of ~0.5, true underlying function, and AAE reconstruction. The parameters used to generate this data were randomly sampled from the ranges given in Table S1. Negative and positive time bins are combined to create functions with fewer numbers of bins to reduce network size.

II. Adversarial Autoencoder Ensemble Model

At the core of our model is a denoising autoencoder architecture that takes noisy photon correlation spectra as input, projects it onto a latent, low-dimensional vector representation, and reconstructs the noise-free $g^{(2)}$ [33]. Here, the encoding layers can be a fully-connected multilayer perceptron (MLP) or a deep convolutional neural network; in either case, each layer comprises weight multiplication, followed by batch normalization and application of a non-linearity.

An individual autoencoder is deterministic, meaning that passing a given input through the network will always result in the same output. As our application often involves noise-dominated data, it is advantageous to consider model architectures that can reconstruct a multitude of *reasonable* reconstructions, and so we

sought to render our model probabilistic rather than deterministic. Whereas conventional Bayesian estimation requires stochastic sampling to determine uncertainty in modeling parameters, probabilistic deep learning approaches offer performant and accurate means to infer uncertainty without the need for computationally intensive Monte Carlo simulations. Ensemble models in particular are conceptually well-grounded, and in the case of deep ensembles, comparatively straightforward to implement and train for uncertainty estimation [34]. In this approach, an ensemble model is constructed that consists of a set of equivalent submodels, where each submodel is randomly initialized and trained independently to reconstruct the $g^{(2)}$ from the noisy input. The random initialization introduces sufficient variance to each of the submodels comprising the ensemble, without the need to subdivide the training data, as is commonly done in the deep learning literature. These independent reconstructions are then aggregated to yield a mean and variance of photon counts for each bin, i.e. the model prediction and associated uncertainty measure. Examples of the individual submodel reconstructions are shown later in this work.

Throughout the course of model development, it became evident however that the submodels failed to produce a *smooth* $g^{(2)}$ on their own: the mean of the submodels could produce accurate $g^{(2)}$ reconstructions, but the individual submodels could take on unphysical forms. For example, some of the submodels would predict negative counts in some time bins. To overcome this, we regularized the outputs of each submodel by including a self-supervised metric to minimize the distance between embeddings (i.e. each submodel encodes similar information), and by involving a discriminator model – which follows the same architecture as the encoder – that learns to distinguish between real $g^{(2)}$ functions and those produced by the model. This is done in the same fashion as those found in generative adversarial models [35], which forces each submodel decoder to reconstruct $g^{(2)}$ functions that are indistinguishable from analytic ones (the clean functions that are generated using the equation described below).

With the incorporation of the discriminator and adversarial loss, we thus named our model as an Adversarial Autoencoder Ensemble (AAE). A schematic of the model architecture is shown in Fig. 1e. For training, the input consists of both the analytic (true) $g^{(2)}$, and the same function after Poisson sampling to create noise. The noisy part of the input is sent to each of the autoencoder submodels (only two of eight submodels are shown in Fig. 1e for clarity). The outputs of each submodel are sent to the discriminator, which enforces the submodels to produce reconstructed outputs that resemble the true $g^{(2)}$. The mean of all the reconstructed outputs are compared with the true $g^{(2)}$ to obtain the reconstruction loss. For inference of the simulated and experimental data in the remaining sections of this work, only the noisy $g^{(2)}$ input is sent through the network, returning a mean reconstructed output (μ) and variance (σ).

While the process could be substantially simplified if one were to infer the $g^{(2)}$ parameters and reconstruct them using analytic expressions (as in the LM and MLE approaches), the deep ensemble approach removes

the need for model selection during fitting (i.e. which model is most likely attributable to the data) which adds another layer of complexity for noisy data. Instead, we can expose our network to multiple possible models during data generation and training.

Finally, in order to improve the accuracy of the models, and to a certain extent improve the variance in the ensemble, we employed a cyclical annealing learning rate schedule: each successive cycle decreases model bias, as well as ensuring the submodel trajectories diverge sufficiently from their initialized points [36]. The deep learning stack was implemented using PyTorch [37] and PyTorch Lightning [38], with weight updates performed using the Adam optimizer [39]; code and values for the data generation, hyperparameters, training, and evaluation pipeline can be found in the Github repository. Training was performed on an Nvidia 3080 GPU, while model evaluation and inference was performed on CPUs (2.6 GHz 6-Core Intel Core i7, 32GB RAM). Fig. S1 [32] displays the training and validation loss for the ensemble, submodels, and encodings versus number of epochs.

III. Training data generation

We begin by generating training data for our network by simulating $g^{(2)}$s that are similar to those we measured experimentally for two different types of QDs in this study, InP/ZnSe/ZnS and CdS/CdSe/CdS. These QDs exhibit lifetimes ranging from ~10 - 100 ns, and in the case of the CdS/CdSe/CdS, a single QD can exhibit two lifetimes. Here we consider the case where the QD is excited periodically by a pulsed laser. Accordingly, we use the following equation to generate our data, modified from the ansatz used by Cortes et al. to include two exponential lifetimes and amplitudes [23]:

$$g^{(2)}(\tau) = y_0 + a_1\left(R \cdot e^{-\left|\tau/t_1\right|} + \sum_{n\neq 0} e^{-\left|(\tau - n\Lambda)/t_1\right|}\right) + a_2\left(R \cdot e^{-\left|\tau/t_2\right|} + \sum_{n\neq 0} e^{-\left|(\tau - n\Lambda)/t_2\right|}\right) \quad (1)$$

where y_0 is a constant flat background to capture correlations from random dark counts, a_1 and a_2 control the relative amplitude of the double-sided exponential peaks with lifetimes t_1 and t_2, R is the coefficient of the peak at $\tau = 0$, representing the ratio of center-to-side peak areas or heights (BXQY), and Λ is the repetition period of the pulsed excitation laser. This equation describes a series of periodic peaks, with the center peak with amplitude $(a_1 + a_2)R$ at $\tau = 0$, and the side peaks with amplitude $(a_1 + a_2)$ at $\tau = n\Lambda$, where n is an integer corresponding to the number of laser pulses that have passed between the first and second photon detection events. The equation allows us to include an arbitrary number of side peaks. For the sake of simplicity and proof-of-concept, the majority of our analysis will focus on datasets containing only the center peak and the first side peak. This is mainly to allow the benefit of a smaller number of time bins to represent the data, which correspondingly greatly reduces the overall size of the neural network. To further reduce the number of bins required, we combine the $g^{(2)}$ bins at equivalent positive and negative times. $g^{(2)}$s are typically reported as shown in Fig. 1d, where τ spans negative and positive times according

to which detector is labelled 'start' and 'stop'. These designations are artificial, and the data loses no physical meaning by folding the $g^{(2)}$ data to the positive part of the τ axis, reducing the number of bins required by half. Fig 1f shows an example of this folded $g^{(2)}$, and we use this form throughout the rest of this work. The experimental and simulated $g^{(2)}$s used in this work are constructed with 301 individual time bins, a laser repetition period Λ of 1 μs, and τ spanning 0 to 1.5 μs, giving bin size of ~5 ns. We found that changing the number of time bins between 300 and 1000 did not significantly affect the accuracy of the fits, and so kept the lowest number to minimize neural network size. An odd number of bins was preferable to be able to more consistently observe the sharply peaked double-sided exponentials.

The parameters used to generate training data with eq. 1 were taken from 10 evenly spaced values between upper and lower bounds for each variable, whose values can be found in Table S1. This ensures that each combination of these parameters is seen once by the model. For generating training data, we set $a_2 = (1 - a_1)$ in order to minimize redundancies. For 6 parameters gridded with 10 points each, this results in 10^6 simulated $g^{(2)}$s in our training set. These clean functions have Poissonian noise added while batches are being loaded for the network training, by drawing a sample from a Poisson distribution for each individual time bin. To simulate different levels of signal accumulation, the clean functions (before Poisson sampling) are scaled by random factor between 1 and 1000: higher amplitudes will result in less uncertainty for the Poisson sampling at each time bin, simulating longer accumulation times in actual experiments. After adding noise, the functions are renormalized before being input into the fitting algorithms and the AAE.

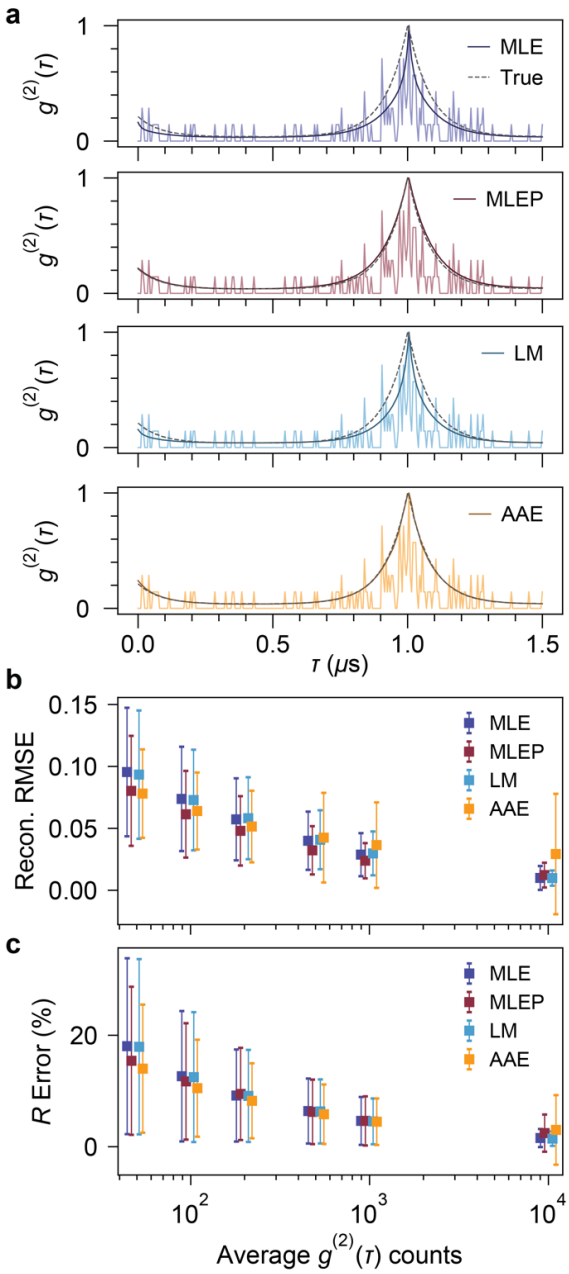

Figure 2. (a) Fits to a single simulated, few-shot (~100 counts) $g^{(2)}$ input from the set of 1000 using Maximum Likelihood Estimation (MLE), MLE with a Poisson likelihood (MLEP), Levenberg-Marquardt (LM), and the Adversarial Autoencoder Ensemble (AAE). **(b)** Root Mean Squared Error of the function reconstruction (Recon. RMSE) and **(c)** absolute error of center-to-side peak area ratios (*R* Error) versus average number of $g^{(2)}$ counts. Error bars indicate the standard deviation, σ, of the computed errors.

IV. Reconstructions of simulated $g^{(2)}$s

We first evaluated the performance of the AAE against three other fitting methods – Maximum Likelihood Estimation with a Normal Likelihood (MLE), MLE with a Poisson likelihood (MLEP), and Levenberg–Marquardt (LM) – on sets of simulated $g^{(2)}$s (see Methods section for details about MLE/LM fitting and

data generation). We generate sets of 1000 $g^{(2)}$s with different fixed levels of accumulation, which are quantified by the total number of counts in the histogram, i.e. the total number of two-photon events in this time range. We chose 1000 to obtain good statistical averages, and to avoid long computation times of the least-squares fitting algorithms (which are much slower than the AAE reconstruction). These datasets are generated using variables randomly sampled from a uniform distribution bounded between the same range as used for training the neural network. Whereas the AAE was trained on a parameter grid of 10 different values per parameter, these simulated $g^{(2)}$s use parameters with values that can be in between grid points, meaning that the AAE here is tested on variations of $g^{(2)}$s not in the training set. The tolerance for the total number of counts when generated each $g^{(2)}$ was ±10% of the target number. Fig. 2a shows an example of reconstructions of sparse data with ~100 counts using the four different approaches under study. All approaches work well, and we see that the AAE is clearly able to accurately reconstruct the clean $g^{(2)}$ using inputs equivalent to few-shots of integration.

We evaluate the average normalized Root-Mean-Square Error (RMSE) of the reconstructed functions versus the true function for all $g^{(2)}$s in the dataset (Recon. Error). We also evaluate the average absolute error in the ratio, R, of the areas of the center peak and side peak. These average errors and the standard deviation of the errors, σ, are plotted versus the average number of counts in Fig. 2b. For $g^{(2)}$s with an average number of counts $\leq$ 1000, the AAE is able to generate approximations with lower overall reconstruction error and better estimates of R with less variance. The average values for the reconstruction and R error from each of the approaches are all within 1σ of each other, and so we conclude that the AAE overall exhibits comparable, but not necessarily superior, accuracies. These values are tabulated in Table S2.

The lower variance of the AAE versus the fitting algorithms is further exemplified by performing linear regression of the true versus predicted R. These plots are shown in Fig. 3 along with r^2 values. Here we show plots for the dataset with ~100 counts on average (see Fig. S2 [32] for similar plots with ~50 and ~10000 counts). Of the four approaches, the AAE has the highest r^2 value, and also avoids a clustering effect suffered by the three other approaches whereby a significant number of $g^{(2)}$s with $R > 0.5$ are erroneously predicted to have $R \approx 1.0$.

The AAE also exhibits a high computational speedup and performs reconstructions much faster, requiring an average time of ~0.35 ms per $g^{(2)}$ reconstruction (17.7 μs on a GPU), whereas MLE, MLEP and LM require ~260, ~280, and ~60 ms, respectively. This corresponds to a computational speedup of approximately 743 (14689), 800 (15819), and 171 (3389) for AAE reconstructions performed on a CPU (GPU) compared to MLE, MLEP, and LM. These comparisons represent preliminary estimates of the computational speedup, since the optimizers can be further refined. Details of how the optimization was

performed with MLE and LM can be found in the Supplemental Material [32], and in the Github repository. We note here the distinction between computational speedup, which is the time it takes for an input to be reconstructed into a denoised output, and the experimental speedup, which is the amount of time required to accumulate enough shots to obtain reconstructions with a sufficiently high (low) accuracy (error).

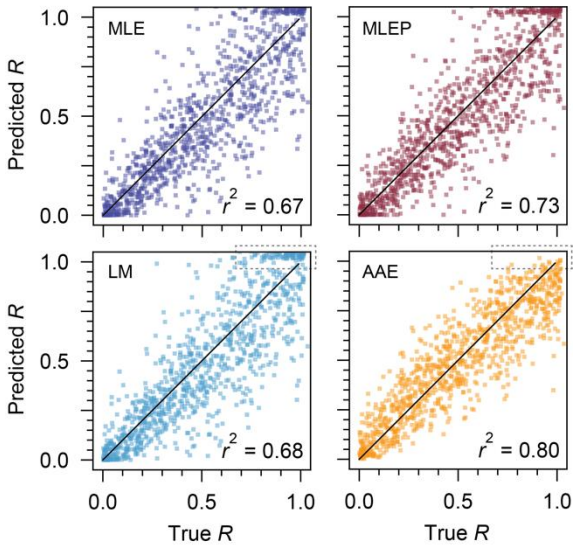

Figure 3. True versus predicted peak area ratios and r^2 values for a dataset with ~100 counts in each histogram, with the solid black line representing ideal correlation (i.e. true equals predicted). The dashed grey box highlights the lack of clustering for the AAE.

While these simulated $g^{(2)}$s were limited to only the center peak and the first side peak, extending the correlation time range to more side peaks will account for a greater number of two-photon events within the same acquisition time. We trained another AAE network that takes as an input $g^{(2)}$ data ranging from 0 to 3.5 μs with 901 time bins, which results in correlations up to a third side peak and has approximately the same bin spacing as the datasets spanning 0 to 1.5 μs with only one side peak (Fig. S3 [32]). As expected, simulated $g^{(2)}$ data with 3 side peaks have approximately 3-fold more counts than the same function with 1 side peak, given the same simulated acquisition time. Comparing the AAE versus MLE, MLEP, and LM for 3-peak $g^{(2)}$ datasets with 150, 300, 600, 1500, and 3000 average counts (corresponding to 50, 100, 200, 500, and 1000 for 1-peak datasets), we observe similar trends in reconstruction and R error as in Fig. 2b, but with overall slightly lower errors owing to the increased number of counts.

In practice, this 3-peak network should be more useful in rapid reconstructions of experimental $g^{(2)}$ data, but we reserve the remainder of our analysis on 1-peak data for the sake of simplicity and computational efficiency. In principle, more peaks and time bins could be included in the model to accumulate more shots

in each $g^{(2)}$ without increasing acquisition time, but this greatly increases the size of the model and training data, and so was not pursued in this work.

V. Reconstructions of experimental $g^{(2)}$s

We next tested the performance of the AAE on experimental $g^{(2)}$s collected from four different types of emitters: a single InP/ZnSe/ZnS dot, a single CdS/CdSe/CdS dot, an InP/ZnSe/ZnS aggregate, and a CdS/CdSe/CdS aggregate. This combination of single QDs and aggregates allows us to test the model on a variety of different forms of the $g^{(2)}$ with different center peak heights – which vary with single QD BXQY and when multiple QDs are being probed simultaneously in the confocal area. The QDs under investigation are determined to be single or multiple emissive species by applying a variable time-gate on the emitted photon stream (Fig. S4 [32]). [1]

The fully accumulated $g^{(2)}$ (>10000 counts) for each QD or aggregate are shown in Fig. 4a, along with fits using eq. 1, which we use as the 'true' function. This is in contrast with the simulated data, where the true distribution is known and used to generate the data. Here, using a fit derived from MLE to generate the 'true' distribution, and then comparing MLE and LM least-squares fits with the 'true' distribution, may bias the reconstruction error in favour of these algorithms over the AAE. We next use the full photon stream to generate 1000 partially accumulated $g^{(2)}$s with a target number of counts. This is achieved by correlating a subsection of the photon stream, the length of which is determined by dividing the target number of counts by the total counts of the full $g^{(2)}$ and multiplying this fraction by the length of the full photon stream array. The subsections of the photon stream are randomly chosen and correlated until they return a $g^{(2)}$ with the target number of counts (with a ±10% tolerance).

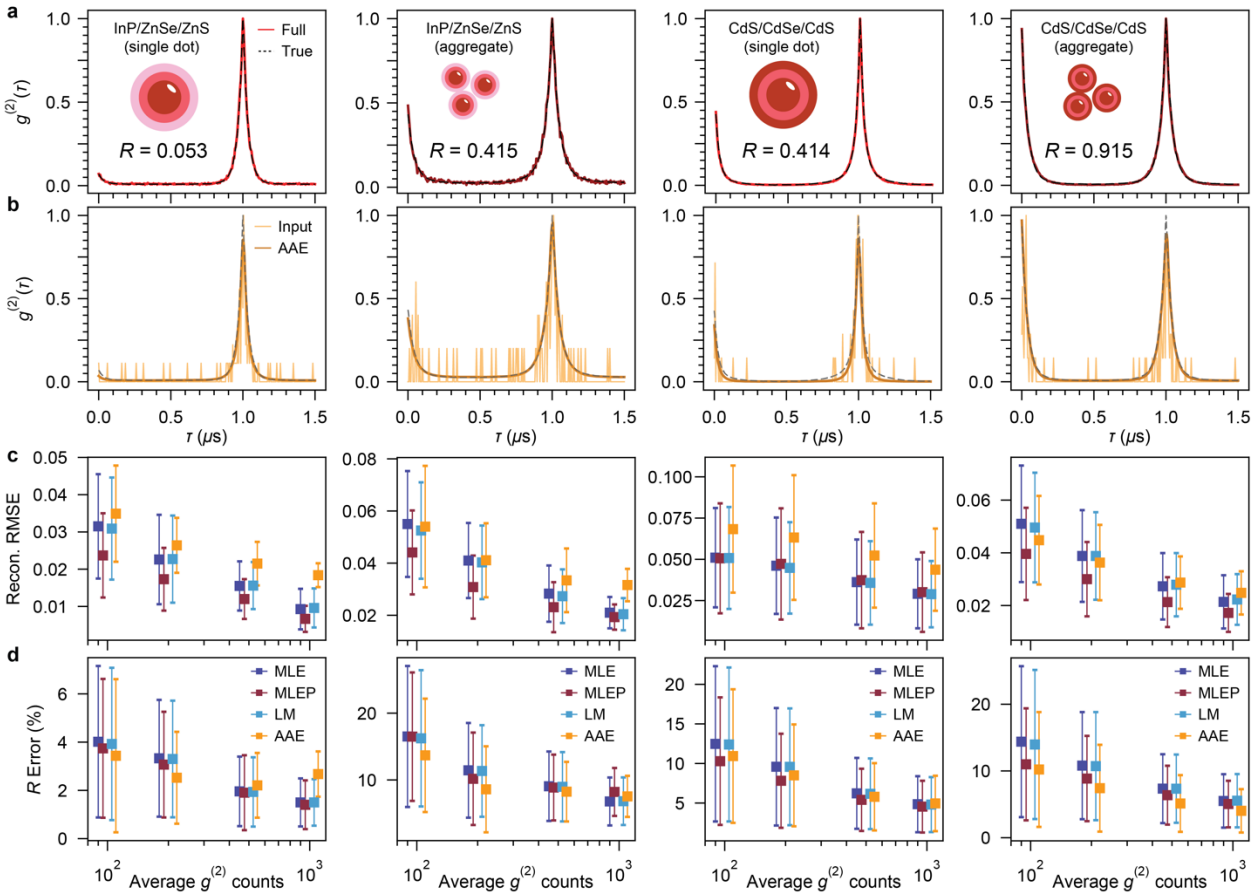

Figure 4. (a) Full experimental $g^{(2)}$s with >10,000 counts and true distribution fit using MLE with eq. 1. The $g^{(2)}(0)$ values for each sample are given beside the center peak, and the remaining fitted parameters for each sample can be found in Table S3. **(b)** Experimental $g^{(2)}$ input with ~100 counts and AAE reconstruction, **(c)** reconstruction RMSE and **(d)** R error, with error bars indicating $\pm\sigma$, versus average $g^{(2)}$ counts for an InP single dot, an InP aggregate, a CdSe single dot and a CdSe aggregate.

Fig. 4b shows examples of reconstructions of experimental $g^{(2)}$s for each QD or aggregate with ~100 total counts using the AAE. In all cases, the model is able to produce highly accurate reconstructions. The average reconstruction RMSE and R error for datasets with 100, 200, 500, and 1000 average counts are plotted in Fig. 4c and d. We observe that the AAE reconstructions are all within 1σ of MLE, MLEP, and LM in all cases, with slightly better performance for the CdS/CdSe/CdS aggregate sample and slightly worse for the CdS/CdSe/CdS single dot. It's possible that the poor performance for this particular system results from insufficient exposure to $g^{(2)}$s with this form in training. However, despite poorer reconstruction errors than for the simulated data, the AAE generally performs best for estimating R and having the lowest σ for the lowest number of counts. The only exception is again the CdSe/CdS/CdSe aggregate, where the AAE performs slightly worse than MLEP but better than MLE and LM. As with the simulated data, the reconstruction time is orders of magnitude faster with the AAE.

The reconstruction performance of the AAE using experimental $g^{(2)}$ inputs remains competitive with MLE and LM, and (within error) outperforms these algorithms in the estimation of R (BXQY), which confirms that the facilely simulated datasets – generating clean correlation functions and Poisson sampling at each time bin – serve as excellent training data for the model, suggesting that this methodology could be extended to other photon correlation spectroscopy or quantum optics experiments that rely on single photon counting with Poisson random noise. The poorer performance of the AAE on real experimental data in Fig. 4 compared to the simulated data in Fig. 2 is likely the result of two reasons: the first, as stated earlier, is that the 'true' distribution used to evaluate the reconstruction error is itself generated by a least-squares fit using MLE, which may bias the performance the MLE and LM algorithms over the AAE. The second possible reason is that the experimental data may contain subtle features that are absent in the training data. This limitation could be overcome by including few-shot experimental $g^{(2)}$s for QDs with varying lifetimes and BXQYs.

VI. Probabilistic outputs

We lastly explore the capability of the model to output uncertainties with its reconstructions. For conventional machine learning models, obtaining uncertainties in predictions or reconstructions is generally difficult to achieve; our AAE model, built from 8 individual submodels, simultaneously predicts the ensemble mean and variance. The ensemble approach thereby inherently provides a prediction of uncertainty, which is important for its use in practical scenarios. Plots of the mean (μ) and the 95% confidence interval (+/- 2σ) of the AAE output are shown in Fig. 5a for simulated $g^{(2)}$s with ~300 counts and ~10000 counts. For the probabilistic model to capture uncertainty correctly, the uncertainty in the output (denoted by the grey shaded area) should scale inversely with number of counts, which we observe. In other words, the ensemble reaches a consensus with increasing signal-to-noise. By inspecting the individual sub-models in Fig. 5b, we can see how each submodel reconstructs a good approximation of the $g^{(2)}$, and that the greatest variance appears to be in the estimated exponential lifetimes. The sub-models for the $g^{(2)}$ with higher counts are clearly more similar, resulting in lower variance.

We compare the uncertainties of our model with uncertainties computed on the same sets of data using Markov Chain Monte Carlo (MCMC) [40], implemented using the PyMC3 package for Python [41]. MCMC is a Bayesian approach that uses numerical sampling to estimate posterior distributions of parameters, conditional on the choice of prior and on the data likelihood, which may be intractable analytically [40,42]. It is therefore well suited to estimate uncertainty in parameters via their sampled posterior distributions. With PyMC3, we use a variant of MCMC called Hamiltonian Monte Carlo (HMC) using the No-U-Turn Sampler [43], which we use to generate posterior distributions for parameters and fits to $g^{(2)}$ data here.

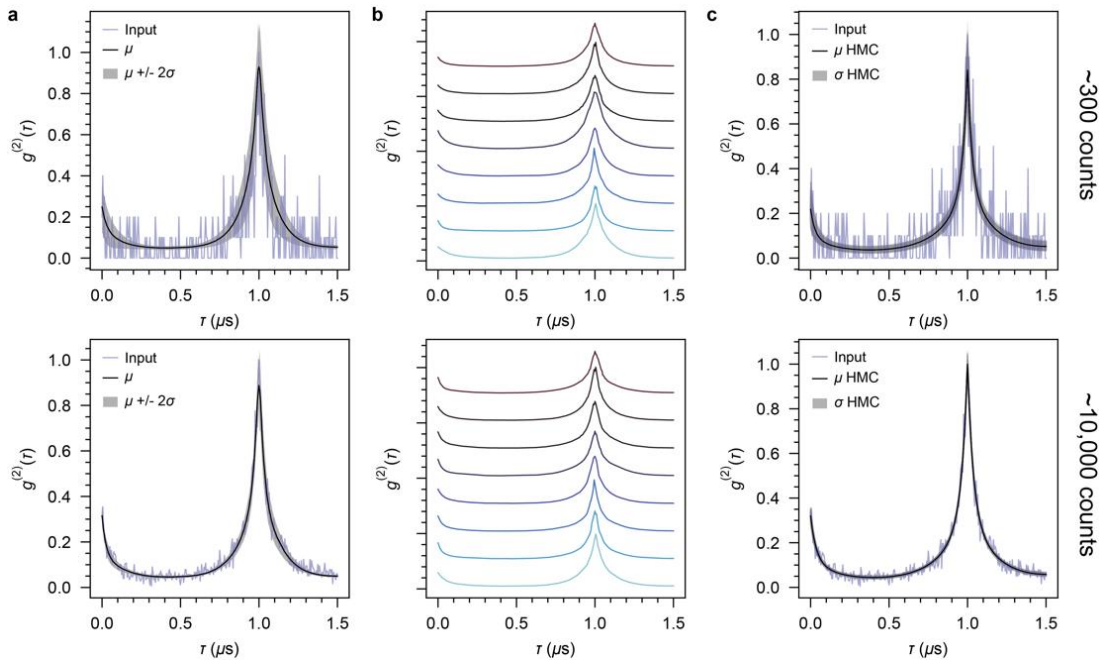

Figure 5. (a) Input data, AAE output average (μ), and two standard deviations around the average ($\mu \pm 2\sigma$), **(b)** outputs from the 8 individual models within the ensemble, normalized and offset for clarity, and **(c)** Hamiltonian Monte Carlo average fit (μ HMC) and 2000 fits drawn from the chain, plotted in gray with transparency (σ HMC).

To perform HMC, we first estimate the parameter means using MLE. Then for each parameter prior, we use a normal distribution with a mean equal to the result from MLE, and with a variance equal to 0.25 of the mean value. We initially attempted to use uniform priors with the same upper and lower bounds as used for generating training data, but were unable to obtain a converged HMC output. With normal priors, we found that 300 counts minimum were required for the HMC to converge and generate reasonable posterior distributions and covariances for each parameter (Fig. S5, S6 [32]), noting some covariance between the amplitude and lifetime parameters. The mean outputs of the converged HMC runs for the 300 and 10000 count $g^{(2)}$ data are plotted in Fig. 5c. We additionally plot 2000 fits from the converged chain to form the shaded gray area (each fit line is given some transparency), which is used to gauge the uncertainty of the fit from the HMC output [44]. Comparing the uncertainties from the AAE with HMC in Fig. 5 panels a and c, we see generally similar behavior, although the uncertainty in the HMC output appears to be lower; a likely consequence of overfitting, and/or over-constrained priors. We note here that quantitative agreement between the HMC and AAE uncertainties is not required: the former depends on our choice of prior distributions and parameters, and is used here to provide a qualitative guide into the correct behavior going from noisy to noise-less data.

VII. Conclusions

Overall, our deep ensemble AAE model is able to perform reconstructions of few-shot $g^{(2)}$s with lower variance and, within 1σ, commensurate reconstruction errors and estimates of BXQY when compared to MLE and LM fitting algorithms. The AAE model begins to perform more poorly than LM and MLE with increasing counts, which could possibly be resolved by using a training set that exposes the model to more $g^{(2)}$s with higher counts, for example to a level at which a target uncertainty in the estimate of R is achieved. Despite a computational speedup in reconstruction time, we were unable to achieve a meaningful speedup with respect to data acquisition time (e.g. to achieve a target uncertainty or error with the minimum number of accumulated shots), relative to LM and MLE fitting. We posit that this is due to the very simple form of the distributions that describe the data, which are easily fit by eq. 1 and require the estimation of only a few parameters. Nevertheless, it is promising that the AAE remains competitive with the fitting algorithms that are equipped with the same equation that describes the actual true distribution (for simulated data).

In comparison to MLE and LM which use eq. 1 as a model, it is important to emphasize that the AAE approach is able to reconstruct the $g^{(2)}$ without a model being explicitly assigned to the data. While the model here is trained only on simple data generated using the same analytical equation – which was suitable for testing on the experimental $g^{(2)}$ data we collected –other $g^{(2)}$ models or experimental data could easily be included into the training data that also account for photon bunching, non-exponential lifetimes, and time-ordering (i.e. in a biexciton-exciton cascade correlation measurements) [45]. Our methodology is also directly translatable to $g^{(2)}$s generated using continuous rather than pulsed excitation. As an aside, it is important to note that information theoretic arguments provide a lower bound for the minimum number of samples required to estimate a single parameter θ with uncertainty $\Delta\theta$. In practice, however, this requires knowledge of the model which is not always readily available. Moreover, the reconstruction can become computationally intractable (as in Bayesian MCMC) or is computationally difficult for a large parameter set (as in MLE and MLEP). Our approach overcomes this limitation by using training examples and the ML framework, while retaining the highly advantageous uncertainty prediction of these alternative approaches.

Furthermore, the model performs these reconstructions orders of magnitude faster than the other algorithms studied here, and generates probabilistic outputs with uncertainties similar to HMC. These capabilities make the AAE model suitable for on-the-fly signal reconstruction, whereby the accumulating $g^{(2)}$ data could be sent to the model for nearly instantaneous reconstruction and estimates of $g^{(2)}(\tau = 0)$ with corresponding uncertainties. It is worth noting that the MLE and MLEP approaches require solving a nonlinear and non-convex optimization problem in real-time. For a large number of parameters, this can become computationally intensive and can become a limiting factor for these types of approaches. Once the ML

predictor has been trained, it will require only a one-way evaluation of the neural network for its prediction (which can be further reduced through pruning), allowing for the fast performance we observed in our experiments. These properties of our model could make it useful in multi-pixel based $g^{(2)}$ reconstruction for quantum imaging and quantum spectroscopy applications [11–13]. This works serves as a foundation for probabilistic autoencoder models that can perform fast and accurate reconstructions of 1D photon correlation data, and is naturally extendable to other single molecule and quantum optics experiments.

Data availability

All the datasets that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All code used to analyze data and support the findings of this study are available from the corresponding author and is also made publicly available at the Github repository.

References

[1] B. D. Mangum, Y. Ghosh, J. A. Hollingsworth, and H. Htoon, *Disentangling the Effects of Clustering and Multi-Exciton Emission in Second-Order Photon Correlation Experiments*, Opt. Express **21**, 7419 (2013).

[2] B. Li, G. Zhang, C. Yang, Z. Li, R. Chen, C. Qin, Y. Gao, H. Huang, L. Xiao, and S. Jia, *Fast Recognition of Single Quantum Dots from High Multi-Exciton Emission and Clustering Effects*, Opt. Express **26**, 4674 (2018).

[3] G. Nair, J. Zhao, and M. G. Bawendi, *Biexciton Quantum Yield of Single Semiconductor Nanocrystals from Photon Statistics*, Nano Lett. **11**, 1136 (2011).

[4] M. Müller, S. Bounouar, K. D. Jöns, M. Glässl, and P. Michler, *On-Demand Generation of Indistinguishable Polarization-Entangled Photon Pairs*, Nat. Photonics **8**, 224 (2014).

[5] T. Müller, J. Skiba-Szymanska, A. B. Krysa, J. Huwer, M. Felle, M. Anderson, R. M. Stevenson, J. Heffernan, D. A. Ritchie, and A. J. Shields, *A Quantum Light-Emitting Diode for the Standard Telecom Window around 1,550 Nm*, Nat. Commun. **9**, 1 (2018).

[6] E. T. Khabiboulline, J. Borregaard, K. De Greve, and M. D. Lukin, *Optical Interferometry with Quantum Networks*, Phys. Rev. Lett. **123**, 070504 (2019).

[7] W. Van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello, and M. Mosca, *Optimal Phase Estimation in Quantum Networks*, J. Phys. Math. Theor. **40**, 7971 (2007).

[8] C. L. Cortes, P. Lefebvre, N. Lauk, M. J. Davis, N. Sinclair, S. K. Gray, and D. Oblak, *Sample-Efficient Adaptive Calibration of Quantum Networks Using Bayesian Optimization*, ArXiv Prepr. ArXiv210606113 (2021).

[9] K. E. Dorfman, F. Schlawin, and S. Mukamel, *Nonlinear Optical Signals and Spectroscopy with Quantum Light*, Rev. Mod. Phys. **88**, 045008 (2016).
[10] R. C. Fortenberry, *Quantum Astrochemical Spectroscopy*, Int. J. Quantum Chem. **117**, 81 (2017).
[11] F. Albarelli, M. Barbieri, M. G. Genoni, and I. Gianani, *A Perspective on Multiparameter Quantum Metrology: From Theoretical Tools to Applications in Quantum Imaging*, Phys. Lett. A **384**, 126311 (2020).
[12] A. Mikhalychev, B. Bessire, I. Karuseichyk, A. Sakovich, M. Unternährer, D. Lyakhov, D. L. Michels, A. Stefanov, and D. Mogilevtsev, *Efficiently Reconstructing Compound Objects by Quantum Imaging with Higher-Order Correlation Functions*, Commun. Phys. **2**, 1 (2019).
[13] E. Toninelli, P.-A. Moreau, T. Gregory, A. Mihalyi, M. Edgar, N. Radwell, and M. Padgett, *Resolution-Enhanced Quantum Imaging by Centroid Estimation of Biphotons*, Optica **6**, 347 (2019).
[14] C. Melnychuk and P. Guyot-Sionnest, *Multicarrier Dynamics in Quantum Dots*, Chem. Rev. **121**, 2325 (2021).
[15] M. Nasilowski, P. Spinicelli, G. Patriarche, and B. Dubertret, *Gradient CdSe/CdS Quantum Dots with Room Temperature Biexciton Unity Quantum Yield*, Nano Lett. **15**, 3953 (2015).
[16] H. Utzat et al., *Coherent Single-Photon Emission from Colloidal Lead Halide Perovskite Quantum Dots*, Science **363**, 1068 (2019).
[17] J. Claudon, J. Bleuse, N. S. Malik, M. Bazin, P. Jaffrennou, N. Gregersen, C. Sauvan, P. Lalanne, and J.-M. Gérard, *A Highly Efficient Single-Photon Source Based on a Quantum Dot in a Photonic Nanowire*, Nat. Photonics **4**, 174 (2010).
[18] R. Hafenbrak, S. M. Ulrich, P. Michler, L. Wang, A. Rastelli, and O. G. Schmidt, *Triggered Polarization-Entangled Photon Pairs from a Single Quantum Dot up to 30\hspace0.167emK*, New J. Phys. **9**, 315 (2007).
[19] Y.-H. Won, O. Cho, T. Kim, D.-Y. Chung, T. Kim, H. Chung, H. Jang, J. Lee, D. Kim, and E. Jang, *Highly Efficient and Stable InP/ZnSe/ZnS Quantum Dot Light-Emitting Diodes*, Nature **575**, 7784 (2019).
[20] T. Kim, Y.-H. Won, E. Jang, and D. Kim, *Negative Trion Auger Recombination in Highly Luminescent InP/ZnSe/ZnS Quantum Dots*, Nano Lett. **21**, 2111 (2021).
[21] N. Kholmicheva, D. S. Budkina, J. Cassidy, D. Porotnikov, D. Harankahage, A. Boddy, M. Galindo, D. Khon, A. N. Tarnovsky, and M. Zamkov, *Sustained Biexciton Populations in Nanoshell Quantum Dots*, ACS Photonics **6**, 1041 (2019).
[22] J. Cassidy et al., *Quantum Shells Boost the Optical Gain of Lasing Media*, ACS Nano (2022).
[23] C. L. Cortes, S. Adhikari, X. Ma, and S. K. Gray, *Accelerating Quantum Optics Experiments with Statistical Learning*, Appl. Phys. Lett. **116**, 184003 (2020).
[24] Z. A. Kudyshev, S. I. Bogdanov, T. Isacsson, A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, *Rapid Classification of Quantum Sources Enabled by Machine Learning*, Adv. Quantum Technol. **3**, 2000067 (2020).
[25] C. You, M. A. Quiroz-Juárez, A. Lambert, N. Bhusal, C. Dong, A. Perez-Leija, A. Javaid, R. de J. León-Montiel, and O. S. Magaña-Loaiza, *Identification of Light Sources Using Machine Learning*, Appl. Phys. Rev. **7**, 021404 (2020).
[26] V. Gebhart and M. Bohmann, *Neural-Network Approach for Identifying Nonclassicality from Click-Counting Data*, Phys. Rev. Res. **2**, 023150 (2020).
[27] V. Gebhart, M. Bohmann, K. Weiher, N. Biagi, A. Zavatta, M. Bellini, and E. Agudelo, *Identifying Nonclassicality from Experimental Data Using Artificial Neural Networks*, Phys. Rev. Res. **3**, 023229 (2021).
[28] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, *Quantum State Tomography with Conditional Generative Adversarial Networks*, Phys. Rev. Lett. **127**, 140502 (2021).
[29] S. Adhikari, C. L. Cortes, X. Wen, S. Panuganti, D. J. Gosztola, R. D. Schaller, G. P. Wiederrecht, and S. K. Gray, *Accelerating Ultrafast Spectroscopy with Compressive Sensing*, Phys. Rev. Appl. **15**, 024032 (2021).

[30] M. Rodríguez and T. Kramer, *Machine Learning of Two-Dimensional Spectroscopic Data*, Chem. Phys. **520**, 52 (2019).
[31] A. A. Kananenka, K. Yao, S. A. Corcelli, and J. Skinner, *Machine Learning for Vibrational Spectroscopic Maps*, J. Chem. Theory Comput. **15**, 6850 (2019).
[32] info supporting, *See Supplemental Material at [URL Will Be Inserted by Publisher] for [Give Brief Description of Material]*, (n.d.).
[33] D. P. Kingma and M. Welling, *An Introduction to Variational Autoencoders*, (2019).
[34] B. Lakshminarayanan, A. Pritzel, and C. Blundell, *Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 6405–6416.
[35] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative Adversarial Networks*, (2014).
[36] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, and L. Carin, *Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing*, (2019).
[37] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).
[38] W. Falcon et al., *PyTorchLightning/Pytorch-Lightning: 0.7.6 Release*, (2020).
[39] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (2014).
[40] D. van Ravenzwaaij, P. Cassey, and S. D. Brown, *A Simple Introduction to Markov Chain Monte–Carlo Sampling*, Psychon. Bull. Rev. **25**, 143 (2018).
[41] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, *Probabilistic Programming in Python Using PyMC3*, PeerJ Comput. Sci. **2**, e55 (2016).
[42] S. Brooks, *Markov Chain Monte Carlo Method and Its Application*, J. R. Stat. Soc. Ser. Stat. **47**, 69 (1998).
[43] M. D. Hoffman and A. Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, (2011).
[44] M. N. Ashner, S. W. Winslow, J. W. Swan, and W. A. Tisdale, *Markov Chain Monte Carlo Sampling for Target Analysis of Transient Absorption Spectra*, J. Phys. Chem. A **123**, 3893 (2019).
[45] R. M. Stevenson, R. J. Young, P. Atkinson, K. Cooper, D. A. Ritchie, and A. J. Shields, *A Semiconductor Source of Triggered Entangled Photon Pairs*, Nature **439**, 7073 (2006).

Acknowledgements

A. H. P (co-lead author, simulated data and acquired experimental data, performed all analyses and HMC modelling, and assisted with neural network model design and training) and W. S. (helped build the optical setup used to acquire data) were supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0021650. A. H. P. was additionally supported by a Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). K. L. K. L co-lead author, designed the neural network models, performed hyperparameter sweeps, and model training) was supported by B. A. M.'s startup package through MIT. D. B. B. (assisted in acquiring

experimental data) and T. S. (assisted in writing code for correlating experimental data) were supported by the Samsung Advanced Institute of Technology (SAIT). M. S. (assisted with HMC modelling) was supported by the National Institutes of Health through the Laser Biomedical Research Center (Grant # 2-P41-EB015871-31). J. C. and M. Z. (provided CdS/CdSe/CdS dots) were supported by the Award DE-SC0016872, U.S. Department of Energy, Office of Science. C. L. C. assisted in conceiving the project and designed the simulations. T. K. and E. J. (SAIT) provided InP/ZnSe/ZnS dots. S. K. G., B. A. M., and M. G. B. supervised the project. All authors read and commented on the manuscript. We thank A. Johnston for helpful discussions about convolutional neural networks during the model design. Work performed at the Center for Nanoscale Materials, a U.S. Department of Energy Office of Science User Facility, was supported by the U.S. DOE, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

Competing interests

The authors declare no competing interests.

Additional Information

Supplemental Materials: Supplemental Figures S1 – S6, and Supplemental Tables S1 – S3.