

Error-mitigated photonic quantum circuit Born machineAlexia Salavrakos^{1,*}, Tigran Sedrakyan¹, James Mills^{1,2}, Shane Mansfield¹, and Rawad Mezher¹¹*Quandela, 7 Rue Léonard de Vinci, 91300 Massy, France*²*School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, Scotland*

(Received 9 October 2024; accepted 27 February 2025; published 18 March 2025)

In this Letter, we study quantum circuit Born machines (QCBMs) in the context of photonic quantum computing. QCBMs are a popular choice of quantum generative machine learning models, and we present a QCBM designed for linear optics. We show that a recently developed error mitigation technique called recycling mitigation greatly improves the training of QCBMs in realistic scenarios with photon loss, which is the primary source of noise in photonic systems. We demonstrate this through numerical simulations and through an experiment on a quantum photonic integrated processor. We expect our work to pave the way towards more demonstrations of error mitigation techniques tailored to photonic devices which can enhance the performance of a quantum algorithm.

DOI: [10.1103/PhysRevA.111.L030401](https://doi.org/10.1103/PhysRevA.111.L030401)

Introduction. Generative learning has captured much attention in the field of classical machine learning over the last decade, with the advent of deep generative models such as variational autoencoders (VAEs), generative adversarial networks (GANs), diffusion models, and most recently large autoregressive models [1–3]. Over the same period of time, the field of quantum machine learning has emerged and grown fast, with a recent focus on variational quantum algorithms [4], which are hybrid quantum-classical models that can be implemented on noisy intermediate-scale quantum (NISQ) devices [5]. In these algorithms, the model is implemented by a parametrized quantum circuit often called an ansatz, and the parameters are optimized through a classical procedure. While classification tasks have been studied extensively [6,7], considerable progress has also been achieved on quantum generative models [8]. Quantum GANs and quantum VAEs were introduced in Refs. [9,10] and Ref. [11] respectively, while energy-based models such as quantum Boltzmann machines were studied in Refs. [12,13].

Quantum circuit Born machines (QCBMs) were introduced as a circuit-based model [14,15], where the circuit ansatz prepares a state and the Born rule is naturally implemented by the measurements at the end of the circuit. These models are implementable on NISQ hardware, and a first experimental realization was obtained in Ref. [14] on an ion trap quantum processor. Since then, further demonstrations of QCBMs and QCBM-like models were realized on ion traps [16,17], as well as on superconducting platforms [18–20].

Each platform that is being considered for the development of a quantum computer has its strengths and weaknesses,

whose importance may vary according to the task at hand. Fast sampling, for instance, could be particularly interesting for generative learning, and is a benefit of photon-based devices. Quandela introduced in Ref. [21] a quantum processor called Ascella, made of a quantum-dot single-photon source supplying a universal linear optical network on a reconfigurable chip. A very similar processor with a larger chip called Altair was recently developed [22]. Photonic platforms are also known for demonstrations of quantum advantage [23,24] via Gaussian boson sampling [25].

While such devices are currently state of the art, they have not yet entered a regime where quantum error correction (QEC) can be applied at scale. In contrast to QEC protocols, which require a large overhead of qubits to be realized [26], quantum error mitigation (QEM) replaces this qubit overhead with a sampling overhead. More precisely, QEM requires more runs of a quantum algorithm, but allows this algorithm to be run on a NISQ device without the need for any ancillary qubits [27,28]. This makes QEM particularly useful for current and near-term quantum hardware. Photonic devices suffer significantly from a particular type of noise: erasure noise in the form of photon loss. In related work [29], a recycling mitigation technique is presented that deals with photon loss. This technique makes use of photonic output states that would normally be discarded in postselection, and provably outperforms postselection up to large sample sizes.

In this Letter, after introducing a QCBM scheme for linear optical circuits, we show that its training is significantly improved by the recycling mitigation technique in realistic scenarios with photon loss. We demonstrate several situations where a QCBM becomes trainable if recycling mitigation is applied, including a proof-of-concept experiment on our photonic quantum processor Altair. In the discussion, we consider the question of classical simulability, hardness, and the link with boson sampling [30,31].

A photonic QCBM: Generative learning. Given a set of data instances X and labels Y , while a discriminative or classification model aims at predicting $P(Y|X)$, a generative model

*Contact author: alexia.salavrakos@quandela.com

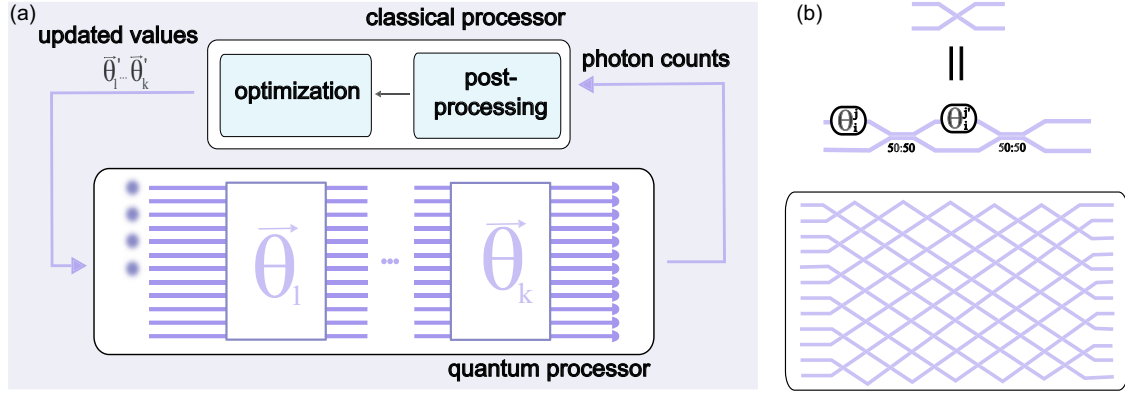


FIG. 1. (a) QCBM ansatz pictured for a circuit with $m = 12$ modes and $n = 4$ photons and input Fock state $|101010100000\rangle$. We allow the variational blocks parametrized by $\vec{\theta}_i$ to be repeated k times. The photon coincidence counts measured at the end of the circuit are sent to a classical processor. (b) One variational block is a universal interferometer. Each crossing between the optical modes corresponds to a Mach-Zender interferometer that is parametrized by two phases.

seeks to estimate the joint distribution $P(X, Y)$, or simply $P(X)$ if there are no labels. If the model has parameters $\vec{\theta}$, a loss function $\mathcal{L}$ is defined which should express how close the distribution of the generated samples is to the one of X . The loss is optimized so as to obtain $\vec{\theta}^* = \arg \min_{\vec{\theta}} \mathcal{L}(\theta|X)$. Once the model is trained, it can be used to generate new samples with similar properties to the training dataset.

Photonic architecture. We propose using an ansatz based on the universal interferometer design of Ref. [32], as shown in Fig. 1. We believe our choice to be well motivated, since the interferometer guarantees interactions between all modes, allows for arbitrary unitary transformation on the modes, and matches the scenario of boson sampling. Note that in Refs. [33,34], the authors present QCBMs with a different linear optical ansatz resembling the gate-based one of Ref. [15].

Our QCBM ansatz consists of a linear optical circuit with m modes and n photons. It receives a Fock state as input, $|\vec{n}_{\text{in}}\rangle = |n_1^{\text{in}}, \dots, n_m^{\text{in}}\rangle$, where n_i^{in} indicates the number of photons in mode i and $\sum_i n_i^{\text{in}} = n$. This state is transformed according to the parametrized universal interferometer which implements a unitary $U(\vec{\theta})$. Indeed, the scheme of Ref. [32] allows for the decomposition of an arbitrary unitary into a product of $T(\theta_i^j, \theta_i^{j+1})$ matrices. Each matrix corresponds to a Mach-Zender element between the j th and $(j+1)$ th modes, as depicted in Fig. 1(b), and has the form

$$\begin{pmatrix} 1 & 0 & \dots & & \dots & 0 \\ 0 & 1 & & & & \vdots \\ \vdots & & e^{i\theta_i^j} \cos \frac{\theta_i^j}{2} & -\sin \frac{\theta_i^j}{2} & & \\ & & e^{i\theta_i^j} \sin \frac{\theta_i^j}{2} & \cos \frac{\theta_i^j}{2} & & \vdots \\ \vdots & & & & 1 & 0 \\ 0 & \dots & & \dots & 0 & 1 \end{pmatrix}.$$

There are $m(m-1)/2$ Mach-Zender elements per universal interferometer as depicted in Fig. 1(b), and thus $m(m-1)$ phases or parameters per block. While a sequence of unitaries can always be rewritten as one single unitary and thus implemented with one block, we consider it interesting to explore

whether this parametrization structure can impact training performance, along with other parametrization choices such as setting all θ_i^j to a fixed value and optimizing only over the θ_i^{j+1} . In practice, however, we obtained the best numerical results for $k = 1$.

At the output of the circuit, the state is measured by photon detectors. Different Fock states $|\vec{n}_{\text{out}}\rangle = |n_1^{\text{out}}, \dots, n_m^{\text{out}}\rangle$ are detected as arrangements of photons in the output modes, that we denote $s = (n_1^{\text{out}}, \dots, n_m^{\text{out}})$. With photon loss, the n_i^{out} may not sum to n . Detecting all possible output states is not straightforward in practice because photon number resolving detectors are not readily available with current technology. To match experimental settings, we make the simplifying choice of using threshold detectors in most of our simulations, thus only allowing n_i^{out} to be 0 or 1 (“click” or “no click”). We refer to the Supplemental Material [35] for more details on no-collision regimes and recycling mitigation.

After detection, outputs s are mapped through classical postprocessing to the space of the data that the model aims to generate. For instance, if we consider simple one-dimensional numerical data X set between x_{min} and x_{max} , we proceed by binning the range of values $[x_{\text{min}}, x_{\text{max}}]$, then defining a mapping between each possible output and each bin. If the data are discrete, there is no need for the binning, and a certain output s^* can directly be interpreted as a value x^* .

Training strategy. The model is trained as a variational algorithm, which means that a classical optimization procedure is applied to the model parameters $\vec{\theta}$, as shown in Fig. 1. In many of our simulations, we use the Kullback-Leibler (KL) divergence [36] as a loss function. We recall the definition of the KL divergence, given two probability distributions Q and P defined on the same sample space $\mathcal{X}$:

$$\text{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log_2 \left(\frac{P(x)}{Q(x)} \right). \quad (1)$$

We also use the total variational distance (TVD) which can be written on discrete domains as [37]

$$\text{TVD}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|. \quad (2)$$

Both the KL and the TVD measure how different Q is from P . The lower the value of the metric, the more similar are the distributions, and a value of 0 can be reached if the two distributions are identical. This is not always desirable as it may indicate overfitting or memorization of the model, and slightly higher values could imply better generalization properties as pointed out in Ref. [38].

In generative learning, additional metrics are often used outside of the loss function to evaluate new model samples, sometimes with respect to a test set. For the generation of images, examples include the Fréchet inception distance [39] or the structural similarity index measure [40]. We consider this to be outside the scope of this Letter, as we focus on the impact of recycling mitigation on loss function behavior.

Other popular choices of loss functions for quantum models include the maximum mean discrepancy (MMD) [41] or the Sinkhorn divergence studied in Ref. [42]. We refer to the last section of this Letter for a discussion around the choice of the loss function, and to the Supplemental Material [35] for more details on model training.

Recycling mitigation. Recycling mitigation is a technique for mitigating the effects of uniform photon loss in linear optical quantum circuits consisting of single photons, multimode linear optical interferometers, and single-photon detectors [29]. At the heart of recycling mitigation are the *recycled probabilities*, constructed from the statistics of the linear optical circuit affected by photon loss, that is, statistics where the number of detected photons at the output of the circuit is less than at the input. The standard method in near-term photonic quantum algorithms is to apply postselection, i.e., discard the statistics where photon loss occurs, and only use the statistics where no loss occurred for the computation.

For a given output s of a linear optical circuit, as defined above, the recycled probability $p_R(s)$ is of the form

$$p_R(s) = p_1 p_{\text{id}}(s) + (1 - p_1) I(s), \quad (3)$$

where $0 \leq p_1 \leq 1$, $p_{\text{id}}(s)$ is the ideal probability of outcome s in the absence of photon loss, and $I(s)$ is the *interference term*, which is an artifact of constructing recycled probabilities from probabilities affected by photon loss. Recycling mitigation consists of applying certain classical postprocessing techniques, detailed in Ref. [29], to the set of recycled probabilities $\{p_R(s)\}$, resulting in a set of mitigated probabilities $\{p_{\text{mit}}(s)\}$, which are approximations of the ideal probabilities $\{p_{\text{id}}(s)\}$.

On an intuitive level, recycling mitigation uses lossy photon statistics in order to get estimates of the ideal (lossless) probabilities. Crucially, the goal of recycling mitigation is to beat postselection, in the sense of getting lower statistical errors on the estimated ideal probabilities. Postselection is an unbiased estimator of the ideal probability with the only error being a statistical error $\epsilon_{\text{stat,post}}$. Conversely, recycling mitigation results in a *biased* estimator of the ideal probability, with a bias error ϵ_{bias} as well as a statistical error $\epsilon_{\text{stat,rec}}$. Because recycling mitigation uses the (more probable) lossy statistics to construct its estimators, then $\epsilon_{\text{stat,rec}} < \epsilon_{\text{stat,post}}$ by construction. Furthermore, it was shown in Ref. [29] that the bias error is small enough such that the inequality

$$\epsilon_{\text{stat,rec}} + \epsilon_{\text{bias}} \leq \epsilon_{\text{stat,post}}, \quad (4)$$

holds for up to a very large number of samples, and for a generic choice of interferometer. This makes recycling mitigation a powerful error mitigation technique applicable to near-term photonic quantum hardware.

Numerical simulations. PERCEVAL [43] is a software platform for discrete-variable photonic quantum computing, which we use in our simulations. We denote by N the number of initial samples for a given simulation, and by η the loss parameter, i.e., the probability that one photon is lost during the computation. Note that, when working with n photons, the probability of observing n -photon coincidence counts at the output of the circuit scales as $(1 - \eta)^n$.

Given the same resources, we typically compare the following cases: (1) the ideal lossless case; (2) the lossy case, where for typical values of η , only a small fraction of N are detected as n -photon counts, and where the n' -photon counts with $n' < n$ are discarded through postselection; and (3) the same lossy case where recycling mitigation is applied, which means that information from the n' -photon counts is taken into account to train the QCBM, as explained in the previous section.

We summarize our results in Fig. 2. We consider in Figs. 2(a) and 2(b) the problem of learning a bimodal distribution consisting of two superimposed Gaussian distributions with different means, as in Ref. [44]. This distribution can be generated straightforwardly in PYTHON. We then consider financial foreign exchange data in Fig. 2(c)—as in Ref. [19], we collected daily log returns of the EUR/USD currency pair over the last 20 years to form the dataset. There, we use the TVD as a loss function as we found it more suitable for a sparser distribution. In Fig. 2(d), we compare the distributions of the generated and target data.

We observe that for given N and η , recycling mitigation significantly improves the training of lossy QCBMs and brings the value of the loss function close to the lossless one, while the lossy unmitigated case remains mostly untrainable. While the steadily decreasing trend of the loss function shows that the model is learning, we note that there is room for improvement in the model optimization, which would be interesting to study in future work. In particular, the importance of initialization parameters and the variance in the training process can be observed in Fig. 2(b). Note that simulation results are displayed for loss parameter $\eta = 0.8$, which is within reach for near-term hardware [21]. Our results remain valid for higher loss regimes and values of η , as we show in the next section.

Experiment. For our experimental setup, we use the Altair quantum photonic integrated processor. We implement the same scenario as in Fig. 2(a), although the main difference is that we find ourselves in a high loss regime: η is estimated to be around 0.96. Additionally, the device is affected by other types of noise, such as photon distinguishability [45,46], which would require a different kind of error mitigation [47]. Further technical details on the setup can be found in the Supplemental Material [35].

We present our results in Figs. 2(e) and 2(f). All runs were obtained with the same initialization, for a fairer comparison. In Fig. 2(e), we observe as before a large gap between the unmitigated and mitigated cases. In Fig. 2(f), we plot one mitigated experimental training instance along with

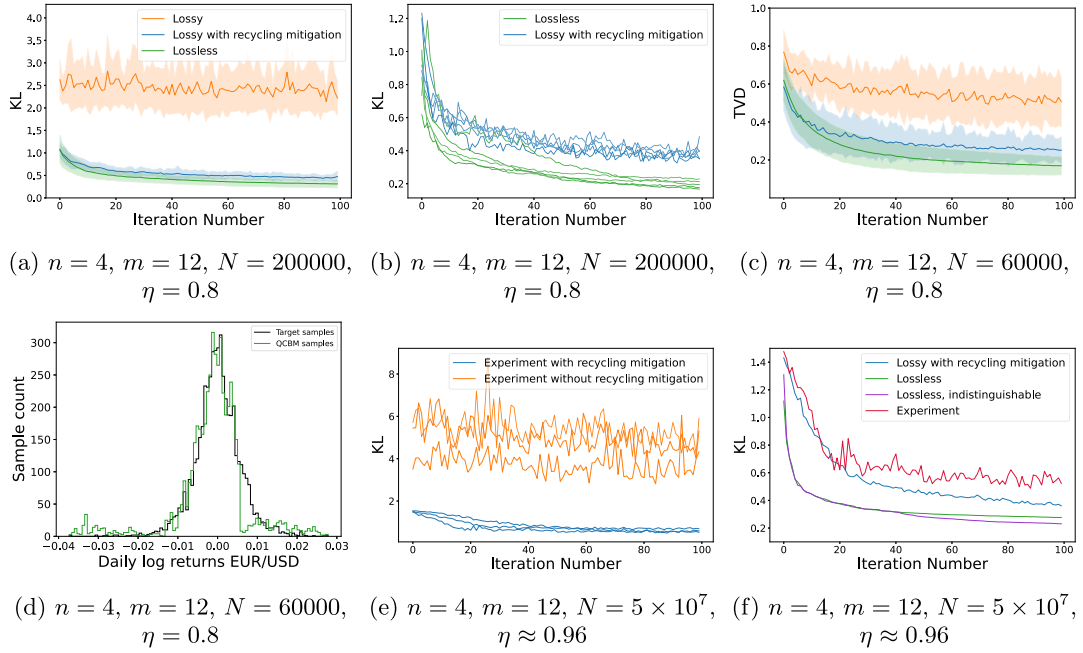


FIG. 2. (a) Value of the KL loss function for the mixed Gaussian case, averaged over 20 training instances. The shaded area depicts a standard deviation from the mean. Each training instance has different initialization parameters. (b) We highlight the five best training instances out of the 20 simulations. (c) Value of the TVD loss function for the financial foreign exchange dataset. (d) Comparison between the target samples from the financial dataset and the samples generated by one of the trained (lossless) QCBM models. (e) Experimental results on Altair for the mixed Gaussian case: three training instances with recycling mitigation, and three without. (f) Comparison between best mitigated experimental instance and corresponding simulations. All but one curve are affected by photon distinguishability.

corresponding simulations. We note the effect of distinguishability, as the fully noiseless simulation reaches a lower value of the loss.

While the experimental data follow the simulation, there remains a difference in the final loss function value which we can attribute to the following factors: There are variations due to the stochasticity of the optimizer; in practice, losses are not fully uniform throughout the setup; and there may be inaccuracies in the transpilation, which is the process of applying adequate voltages to the chip to implement the right phase values, as explained in Ref. [48].

As an alternative strategy for dealing with losses, we could have increased the initial number of samples N enough to make the unmitigated case trainable. However, this would significantly increase the time required for each experiment.

Discussion. When considering quantum advantage of models based on linear optical circuits, it is natural to look for a connection with boson sampling. In this context, we note that generative models may be better candidates for a potential quantum advantage than classifiers obtained from a similar circuit, such as the ones in Refs. [21,49]. Indeed, as Aaronson and Arkhipov pointed out in their original article [30], if the circuit is used for probability estimation instead of sampling, there is a polynomial-time (in the input size) classical algorithm by Gurvits which approximates the permanent of any submatrix of a unitary matrix to within polynomially small additive error [50,51]. This means that estimating the permanent from a polynomial (in input size) number of experiments does not yield an exponential advantage over classical algorithms estimating the permanent, by standard arguments from statistics [52].

However, since generative models such as QCBMs are based on sampling rather than permanent estimation, there remains potential for an exponential advantage. In Ref. [42], the authors clarified the distinction between classical hardness of simulating the circuits that make up the model, and what they defined as *quantum learning supremacy*: the ability of a quantum model to provably outperform all classical models in a certain task. Our model has the capability to achieve classical hardness, provided n and m have sufficiently high values, and losses are limited [53]. Indeed, by varying the values of $\bar{\theta}_k$, the model can span across all possible m -mode interferometers, and there is very strong evidence that sampling, either exactly or approximately, from the output distribution of generic m -mode interferometers is hard to do classically [30,54]. Quantum learning supremacy, however, remains an open question for this photonic QCBM as well as for the qubit-based one presented in Ref. [42].

While the above observations hold true when using a trained model to produce samples, they do not always apply to the training phase of the model. For instance, in our training, we exploited the output probability distribution of the circuit, which allowed us to apply recycling mitigation. The use of an explicit loss function [55] based on the output probability distribution then comes at no additional cost, so it made sense for us to use the KL divergence or the TVD.

Nevertheless, this particular approach to training retains potential, as it turns out that there is still a polynomial advantage in estimating probabilities given by the modulus squared of a permanent in a boson sampling experiment. Indeed, we show in the Supplemental Material [35] that statistics

collected from t independent and identically distributed runs of a boson sampler allow, with high probability, for estimating the modulus squared of the permanent of an arbitrary matrix to a smaller additive error, as compared to the additive error estimate obtained from Gurvits algorithm with a run time of $O(t)$. While these results hold for the lossless case only, this leads us to the open question: Could recycling mitigation make these claims robust to noise? Finally, with respect to the recent results of Ref. [56], we point out that they do not apply to our setup as we do not consider expectation values of Pauli operators, but a linear optical evolution.

In future work, it would be interesting to extend recycling mitigation techniques to work with implicit loss functions such as the MMD or the Sinkhorn divergence. These would allow for the training phase to be based on samples only, which would also be key to scaling the model [55]. Other future research directions include developing mitigation techniques for other types of photonic errors, as well as improving model optimization in the context of photonics, focusing on topics such as gradient evaluation or parameter initialization,

or considering circuit components beyond linear optics such as adaptivity and feedforward.

Conclusion. Our work shows that recycling mitigation can positively impact the training of quantum machine learning models on realistic photonic devices. Given a certain amount of resources, for several scenarios recycling mitigation makes QCBMs that would otherwise be untrainable trainable. Our work also proposes a straightforward design for a photonic-tailored QCBM, which we hope can be built upon, as we make our code available [57].

Acknowledgments. The authors would like to thank Eric Bertasi and Marion Fabre for their invaluable help improving the code, as well as Joseph Bowles and Nicolas Maring for fruitful discussions and feedback on the manuscript. This work has been co-funded by the European Commission as part of the EIC accelerator program under Grant Agreement No. 190188855 for the SEPOQC project; by the Horizon-CL4 program under Grant Agreement No. 101135288 for the EPIQUE project; and by the QuantERA project ResourceQ under Grant Agreement No. ANR-24-QUA2-007-003.

-
- [1] D. P. Kingma and M. Welling, in *Proceedings of the 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014*, Conference Track Proceedings, edited by Y. Bengio and Y. LeCun, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
 - [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger (Curran Associates, Red Hook, NY, 2014), Vol. 27.
 - [3] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, in *Proceedings of the 32nd International Conference on Machine Learning*, edited by F. Bach and D. Blei, Proceedings of Machine Learning Research Vol. 37 (PMLR, Lille, France, 2015), pp. 2256–2265.
 - [4] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, *Nat. Rev. Phys.* **3**, 625 (2021).
 - [5] J. Preskill, *Quantum* **2**, 79 (2018).
 - [6] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, *Phys. Rev. A* **101**, 032308 (2020).
 - [7] E. Farhi and H. Neven, [arXiv:1802.06002](https://arxiv.org/abs/1802.06002).
 - [8] J. Tian, X. Sun, Y. Du, S. Zhao, Q. Liu, K. Zhang, W. Yi, W. Huang, C. Wang, X. Wu, M.-H. Hsieh, T. Liu, W. Yang, and D. Tao, [arXiv:2206.03066](https://arxiv.org/abs/2206.03066).
 - [9] S. Lloyd and C. Weedbrook, *Phys. Rev. Lett.* **121**, 040502 (2018).
 - [10] P.-L. Dallaire-Demers and N. Killoran, *Phys. Rev. A* **98**, 012324 (2018).
 - [11] A. Khoshaman, W. Vinci, B. Denis, E. Andriyash, H. Sadeghi, and M. H. Amin, *Quantum Sci. Technol.* **4**, 014001 (2018).
 - [12] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytsky, and R. Melko, *Phys. Rev. X* **8**, 021050 (2018).
 - [13] N. Wiebe and L. Wossnig, [arXiv:1905.09902](https://arxiv.org/abs/1905.09902).
 - [14] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, *npj Quantum Inf.* **5**, 45 (2019).
 - [15] J.-G. Liu and L. Wang, *Phys. Rev. A* **98**, 062324 (2018).
 - [16] D. Zhu, N. M. Linke, M. Benedetti, K. A. Landsman, N. H. Nguyen, C. H. Alderete, A. Perdomo-Ortiz, N. Korda, A. Garfoot, C. Brecque, L. Egan, O. Perdomo, and C. Monroe, *Sci. Adv.* **5**, eaaw9918 (2019).
 - [17] M. S. Rudolph, N. B. Toussaint, A. Katabarwa, S. Johri, B. Peropadre, and A. Perdomo-Ortiz, *Phys. Rev. X* **12**, 031010 (2022).
 - [18] K. E. Hamilton, E. F. Dumitrescu, and R. C. Pooser, *Phys. Rev. A* **99**, 062323 (2019).
 - [19] B. Coyle, M. Henderson, J. C. J. Le, N. Kumar, M. Painsi, and E. Kashefi, *Quantum Sci. Technol.* **6**, 024013 (2021).
 - [20] V. Leyton-Ortega, A. Perdomo-Ortiz, and O. Perdomo, *Quantum Mach. Intell.* **3**, 17 (2021).
 - [21] N. Maring, A. Fyrrillas, M. Pont, E. Ivanov, P. Stepanov, N. Margaria, W. Hease, A. Pishchagin, T. H. Au, S. Boissier, E. Bertasi, A. Baert, M. Valdivia, M. Billard, O. Acar, A. Briessell, R. Mezher, S. C. Wein, A. Salavrakos, P. Sinnott *et al.*, *Nat. Photon.* **18**, 603 (2024).
 - [22] QUANDELA Cloud (2022), <https://cloud.quandela.com>.
 - [23] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang *et al.*, *Science* **370**, 1460 (2020).
 - [24] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, *Nature (London)* **606**, 75 (2022).
 - [25] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, *Phys. Rev. Lett.* **119**, 170501 (2017).
 - [26] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, UK, 2013).
 - [27] S. Endo, S. C. Benjamin, and Y. Li, *Phys. Rev. X* **8**, 031027 (2018).
 - [28] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).

- [29] J. Mills and R. Mezher, [arXiv:2405.02278](#).
- [30] S. Aaronson and A. Arkhipov, in *STOC '11: Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (ACM, New York, 2011), pp. 333–342.
- [31] R. Berkowitz and P. Devlin, *Isr. J. Math.* **224**, 437 (2018).
- [32] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, *Optica* **3**, 1460 (2016).
- [33] S. Shankar and D. Towsley, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer, Berlin, 2022).
- [34] J. Shi, Y. Tang, Y. Lu, Y. Feng, R. Shi, and S. Zhang, *IEEE Trans. Knowl. Data Eng.* **35**, 1965 (2023).
- [35] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevA.111.L030401> for more information on recycling mitigation and the no-collision regime, for more details on the model training and hyperparameters, for a description of the experimental setup, and for two theorems on the approximation of output probabilities in a boson sampling experiment, which also includes Refs. [58–65].
- [36] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [37] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed. (American Mathematical Society, Providence, RI, 2017), p. 48.
- [38] K. Gili, M. Hibat-Allah, M. Mauri, C. Ballance, and A. Perdomo-Ortiz, *Quantum Sci. Technol.* **8**, 035021 (2023).
- [39] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Red Hook, NY, 2017), Vol. 30.
- [40] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *IEEE Trans. Image Process.* **13**, 600 (2004).
- [41] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, *J. Mach. Learn. Res.* **13**, 723 (2012).
- [42] B. Coyle, D. Mills, V. Danos, and E. Kashefi, *npj Quantum Inf.* **6**, 60 (2020).
- [43] N. Heurtel, A. Fyrillas, G. de Glinasty, R. L. Bihan, S. Malherbe, M. Pailhas, E. Bertasi, B. Bourdoncle, P.-E. Emeriau, R. Mezher, L. Music, N. Belabas, B. Valiron, P. Senellart, S. Mansfield, and J. Senellart, *Quantum* **7**, 931 (2023).
- [44] C. Zoufal, [arXiv:2111.12738](#).
- [45] V. S. Shchesnovich, *Phys. Rev. A* **91**, 013844 (2015).
- [46] M. C. Tichy, *Phys. Rev. A* **91**, 022316 (2015).
- [47] O. V. Borzenkova, G. I. Struchalin, I. V. Kondratyev, A. D. Moiseevskiy, N. N. Skryabin, I. V. Dyakonov, and S. S. Straupe, *Opt. Lett.* **49**, 4453 (2024).
- [48] A. Fyrillas, O. Faure, N. Maring, J. Senellart, and N. Belabas, *Optica* **11**, 427 (2024).
- [49] B. Y. Gan, D. Leykam, and D. G. Angelakis, *EPJ Quantum Technol.* **9**, 16 (2022).
- [50] L. Gurvits, in *Mathematical Foundations of Computer Science 2005*, edited by J. Jedrzejowicz and A. Szepietowski (Springer, Berlin, 2005), pp. 447–458.
- [51] S. Aaronson and T. Hance, [arXiv:1212.0025](#).
- [52] W. Hoeffding, *The Collected Works of Wassily Hoeffding* (Springer, Berlin, 2012).
- [53] S. Aaronson and D. J. Brod, *Phys. Rev. A* **93**, 012335 (2016).
- [54] A. Bouland, D. Brod, I. Datta, B. Fefferman, D. Grier, F. Hernandez, and M. Oszmaniec, [arXiv:2312.00286](#).
- [55] M. S. Rudolph, S. Lerch, S. Thanasilp *et al.*, *npj Quantum Inf.* **10**, 116 (2024).
- [56] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, [arXiv:2409.01706](#).
- [57] A. Salavrakos, T. Sedrakyan, J. Mills, S. Mansfield, and R. Mezher (2024), <https://github.com/Quandela/photon-qcbm>.
- [58] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Phys. Rev. A* **99**, 032331 (2019).
- [59] J. G. Vidal and D. O. Theis, [arXiv:1812.06323](#).
- [60] G. de Felice and C. Cortlett, [arXiv:2401.07997](#).
- [61] G. Facelli, D. D. Roberts, H. Wallner, A. Makarovskiy, Z. Holmes, and W. R. Clements, [arXiv:2409.16369](#).
- [62] A. Pappalardo, P.-E. Emeriau, G. de Felice, B. Ventura, H. Jaunin, R. Yeung, B. Coecke, and S. Mansfield, [arXiv:2410.02726](#).
- [63] J. C. Spall, *Johns Hopkins APL Tech. Dig.* **19** (4), 482 (1998).
- [64] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, *Nat. Commun.* **13**, 4919 (2022).
- [65] R. Mezher, A. F. Carvalho, and S. Mansfield, *Phys. Rev. A* **108**, 032405 (2023).