Supplementary Information for "Quantum optical neural networks"

Gregory R. Steinbrecher[1], Jonathan P. Olson[2], Dirk Englund[1], and Jacques Carolan[1*]
[1] *Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA 02139 and*
[2] *Zapata Computing Inc., 501 Massachusetts Ave., Cambridge, MA, USA 02139*

S1. FULL TRAINING DATA

In Fig. S1(a-c) we present training curves for all experiments not otherwise shown in the main manuscript, including (a) Ising Hamiltonian results, (b) Bose-Hubbard Hamiltonian results, (c) one-way quantum repeater results. In Fig. S1(d) we plot the median error as a function of layer depth for the benchmarking results in the main manuscript. Notably, for the GHZ generator (green) — which experiences a highly non-monotonic behavior when the 'success probability' is plotted (defined as the proportion of runs with an error $< 10^{-4}$) — the monotonic behavior is recovered. For the CNOT and Bell state separator there is a spike in the median error at 6 layers due to the large variance in the final error.

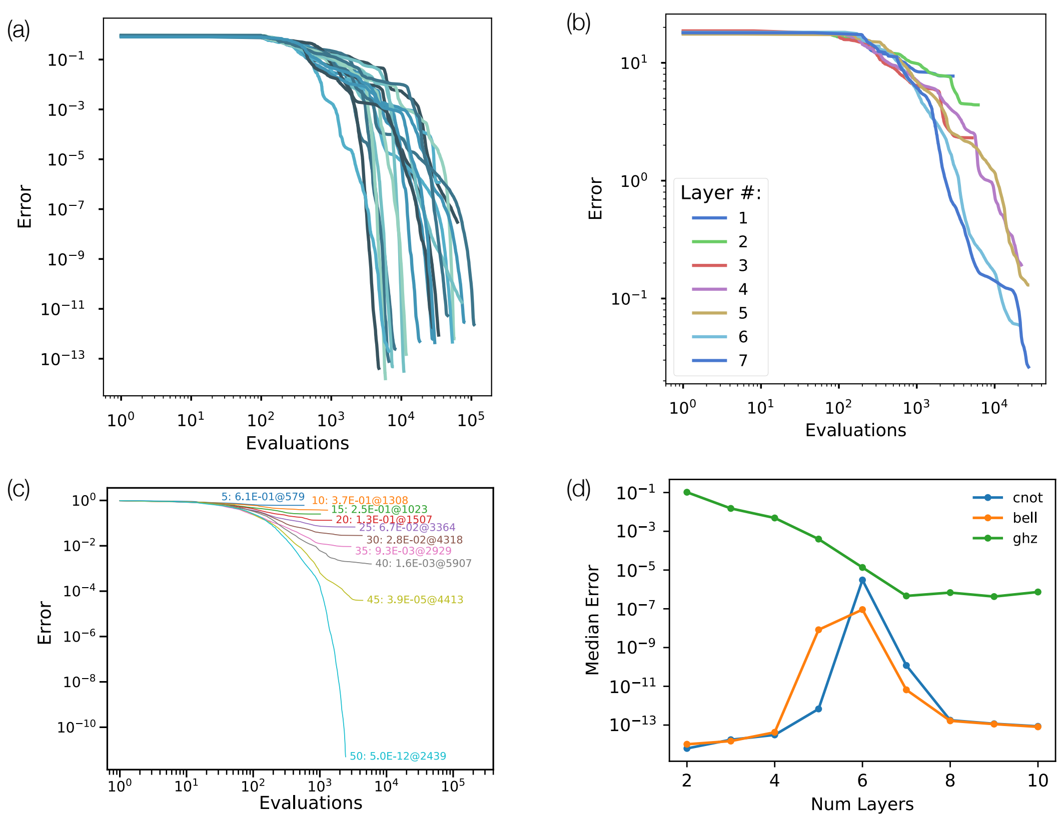

FIG. S1. **Training data.** Error vs. evaluation number curves for all (a) Ising Hamiltonian simulations (as shown in Fig. 3(a) of the main manuscript), (b) all Bose-Hubbard Hamiltonian simulations [as shown in Fig. 3(b)] and (c) all one-way quantum repeater experiments [Fig. 6(b)]. (d) Plots the median error vs layer depth for the benchmarking results shown in Fig. 2 of the main manuscript.

*carolanj@mit.edu

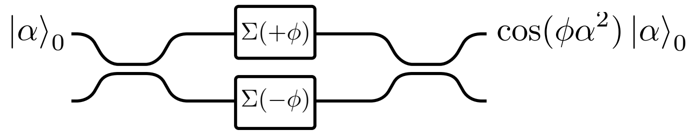

FIG. S2. **Nonlinear activation function.** To demonstrate the QONN is capable of implementing a nonlinear activation function, we consider two single mode Kerr interactions of opposite phase in an Mach-Zehnder interferometer configuration.

S2. CLASSICAL OPTICAL NEURAL NETWORKS

In this section we show that our QONN architecture is also capable of implementing classical optical neural networks [S1] and therefore suitable for classical inference tasks. First we note that while neural networks typically use an arbitrary matrix of real numbers for the linear transform (as opposed to a unitary matrix which the QONN implements), a n-dimensional non-unitary operation can always be embedded across a $2n$-mode optical circuit [S2]. Moreover, recent work has shown that unitary matrices are also suitable for classical neural networks [S3, S4], avoiding the well known vanishing (or exploding) gradient problem.

Next we show that the Kerr interaction is also suitable as the nonlinear activation function for optical neural networks by demonstrating a nonlinear response in the average photon number of an incident coherent state. To see this consider the configuration shown in Fig. S2 with a single mode Kerr nonlinearity

$$\Sigma\left(\phi\right) = \sum_{n=0}^{\infty} e^{in(n-1)\phi/2} \left|n\right\rangle\left\langle n\right| \tag{S1}$$

in each arm of a Mach-Zehnder interferometer, with phase shifts of equal strength but opposite sign. Such a setup can be achieved via atomic systems.

Injecting a coherent state $|\psi\rangle = |\alpha\rangle_0$ with average photon number $|\alpha|^2$ into the top arm of the interferometer (where the subscript 0 (1) indexes the top (bottom) mode) transforms to

$$|\psi\rangle = |\alpha/\sqrt{2}\rangle_0\, |\alpha/\sqrt{2}\rangle_1\,. \tag{S2}$$

After passing through the Kerr region the state becomes

$$|\psi\rangle \approx |e^{i\phi\alpha^2}\alpha/\sqrt{2}\rangle_0\, |e^{-i\phi\alpha^2}\alpha/\sqrt{2}\rangle_1\,. \tag{S3}$$

After the final beamsplitter the state of light in the first mode becomes

$$\begin{aligned} |\psi\rangle &\approx 1/2(e^{i\phi\alpha^2} + e^{-i\phi\alpha^2})\,|\alpha\rangle_0 \\ &\approx \cos(\phi\alpha^2)\,|\alpha\rangle_0\,. \end{aligned} \tag{S4}$$

The average photon number $|\alpha|^2$ therefore varies nonlinearly as $|\cos(\phi\alpha^2)\alpha|^2$. While neural networks typically use only handful of nonlinear activation functions (such as ReLU or Sigmiod), many other functions are also suitable [S5] including sinusoidal functions [S6]. In conclusion, while the QONN naturally performs inference on quantum optical data, it also capable of implementing classical neural networks and therefore capable of performing inference on classical optical data.

S3. REDUCED STRENGTH NONLINEARITIES

In the main paper, we only address nonlinearities of strength π, as those seemed to train most efficiently for the problems we addressed. However, as a full π nonlinearity is a daunting experimental task, the question of what can be accomplished with smaller nonlinearity strengths is of interest. Here, we present the results of two classes of numerical simulations. First, we assume that we have nonlinearities of strength $\frac{\pi}{N}$ and construct networks with N nonlinear layers, such that the total nonlinearity seen by each photon through the whole network is equal to π. Second, we look at the maximum fidelity achievable with a single layer of strength less than π. For both of these we use the CNOT gate as our benchmark of choice.

First, in Figure S3, we plot the training error of a CNOT gate vs. function evaluation where each layer has $\phi = \pi/N$ where N is the total number of layers. That is, each photon passing through the system eventually "sees" a total π of nonlinearity, but that nonlinearity is split over more and more sites. Such a system might be useful in cases where

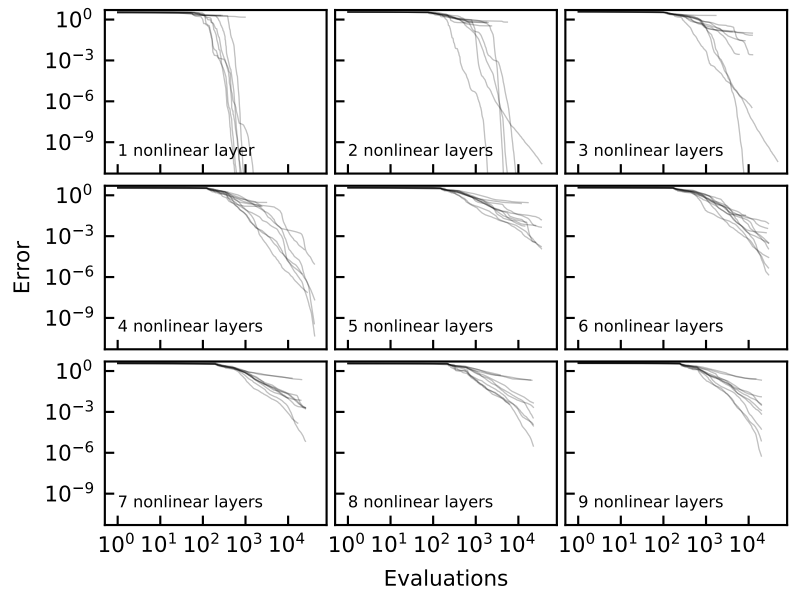

FIG. S3. **CNOT training with $\phi = \pi/N_{layers}$.** To investigate the effects of nonlinearities with $\phi < \pi$ we investigated the case where the total nonlinearity seen through the system is still π but each layer has $\phi < \pi$. For higher layer counts, we found the training to be slower—likely due to the higher number of free parameters—but with less variation in the final error compared with a smaller number of layers.

single nonlinearities do not reach $\phi = \pi$ but many are available. As can be seen, it is sometimes easier to train systems with a small number of layers, but other times the optimization finds a non-optimal local minimum. For higher layer counts, the optimization does not converge as rapidly with iteration number, but all of the optimization runs perform similarly (i.e. the optimizations do not find bad local minima, as they did for small layer counts).

Second, in Figure S4, we look at the performance of a CNOT gate when the available nonlinearities have $\phi < \pi$ and we only construct a single-layer system (i.e. two unitaries sandwiching a single nonlinear layer). In Figure S4 we plot the lowest possible error in the resulting CNOT gate over 250 simulations at each value of ϕ. We find that this is log-linear with respect to the fractional error in ϕ relative to π.

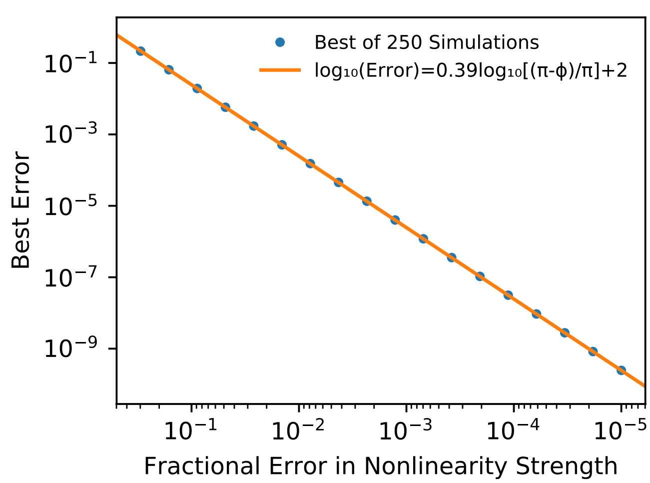

FIG. S4. **CNOT Fidelity impact of $\phi < \pi$.** To look at the case where we do not have access to a full π nonlinearity, we optimized 250 single-nonlinear-layer QONNs at each potential ϕ and plotted the best fidelity achieved vs. the infidelity in π. We found this relationship to be log-linear and plotted the best fit, rounded to two decimal places in both slope and y-intercept.

[S1] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Nat. Photon. **11**, 441 (2017).

[S2] D. A. B. Miller, Photon. Res., PRJ **1**, 1 (2013).

[S3] M. Arjovsky, A. Shah, and Y. Bengio, in *International Conference on Machine Learning* (2016) pp. 1120–1128.

[S4] L. Jing, Y. Shen, J. Peurifoy, S. Skirlo, Y. LeCun, and M. Tegmark, (2016), arXiv:1612.05231v3.

[S5] S. Eger, P. Youssef, and I. Gurevych, arXiv preprint arXiv:1901.02671 (2019).

[S6] G. Parascandolo, H. Huttunen, and T. Virtanen, (2016).