

Imperfect Quantum Photonic Neural Networks

Jacob Ewaniuk,* Jacques Carolan, Bhavin J. Shastri, and Nir Rotenberg*

Quantum photonic neural networks are variational photonic circuits that can be trained to implement high-fidelity quantum operations. However, work-to-date has assumed idealized components, including a perfect π Kerr nonlinearity. This work investigates the limitations of non-ideal quantum photonic neural networks that suffer from fabrication imperfections leading to unbalanced photon loss and imperfect routing, and weak nonlinearities, showing that they can learn to overcome most of these errors. Using the example of a Bell-state analyzer, the results demonstrate that there is an optimal network size, which balances imperfections versus the ability to compensate for lacking nonlinearities. With a sub-optimal $\pi/10$ effective Kerr nonlinearity, it is shown that a network fabricated with current state-of-the-art processes can achieve an unconditional fidelity of 0.905 that increases to 0.999999 if it is possible to precondition success on the detection of a photon in each logical photonic qubit. These results provide a guide to the construction of viable, brain-inspired quantum photonic devices for emerging quantum technologies.

1. Introduction

Quantum neural networks, brain-inspired quantum circuits, harness artificial intelligence to enhance quantum information processing. When driven with light, quantum photonic neural networks (QPNNs) leverage the strengths of mature photonic platforms,^[1] including multiplexing, low latency, and ultra-low operational powers already being exploited by conventional neural networks^[2] and linear-optical quantum processors.^[3] This allows QPNNs to perform quantum state tomography,^[4] act as quantum simulators,^[5,6] process^[7] or reduce the noise^[8] of quantum states, or speed up tasks normally carried out by

classical neural networks, such as image recognition^[9] and natural language processing.^[10]

An example of a QPNN circuit, a two-layer network trained to act as a Bell-state analyzer (BSA), is shown in Figure 1a. Here, the connectivity and activation function for the network are provided by linear, rectangular interferometer meshes (U)^[11] and single-site optical nonlinearities (Σ), respectively. A BSA can distinguish between, or create, all four highly entangled Bell-states, and the addition of this nonlinearity ideally increases the success probability of the circuit to unity^[7] from 0.5 as possible solely with linear optics in the absence of ancillary photons.^[12] The operation performed by the QPNN in this example is therefore crucial to entanglement swapping^[13] and hence provides a route toward a deterministic quantum repeater node,^[14] a vital component of a future quantum internet.^[15]

Quantum photonic circuits are not ideal, and here we report on the performance of imperfect QPNNs. Specifically, we consider how propagation losses and imperfect optical nonlinearities affect the fidelity of the QPNN, using the example of a BSA to benchmark our results. As shown in Figure 1b,c, we find that even networks with weak nonlinearities can vastly outperform those based on (ideal) linear optics. In Figure 1c, we observe that using state-of-the-art waveguide fabrication (as described in Section 2) and a perfect (π) two-photon nonlinearity, ten BSA nodes made from two-layer QPNNs can be applied in series with a success rate of 72%. Surprisingly, this rate is only decreased to 61%, for a much weaker ($\pi/4$) two-photon nonlinearity if a third (lossy) layer is added. Moreover, if each operation is conditioned on the measurement of two photons, the conditional success rate of ten nodes becomes 99.99999% and 99.9% for π and $\pi/4$, respectively. In what follows, we unravel the dependence on loss, effective nonlinearity, and network size, providing a methodology for the design of optimal QPNNs with non-ideal components.

2. Network Architecture and Nonidealities

The architecture of an imperfect QPNN is the same as that of an ideal network, and is thus designed to operate on dual-rail encoded photonic qubits.^[7] Each layer consists of a mesh of tunable Mach–Zehnder interferometers (MZIs) with two controllable phase shifters (ϕ, θ), as shown in the inset to Figure 1a. The interferometer mesh can be programmed to perform any

J. Ewaniuk, B. J. Shastri, N. Rotenberg
Centre for Nanophotonics
Department of Physics, Engineering Physics and Astronomy
Queen's University
64 Bader Lane, Kingston, Ontario K7L 3N6, Canada
E-mail: jacob.ewaniuk@queensu.ca; nir.rotenberg@queensu.ca

J. Carolan
Wolfson Institute for Biomedical Research
University College London
London WC1E 6BT, UK

B. J. Shastri
Vector Institute
Toronto, Ontario M5G 1M1, Canada

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/qute.202200125>

DOI: 10.1002/qute.202200125

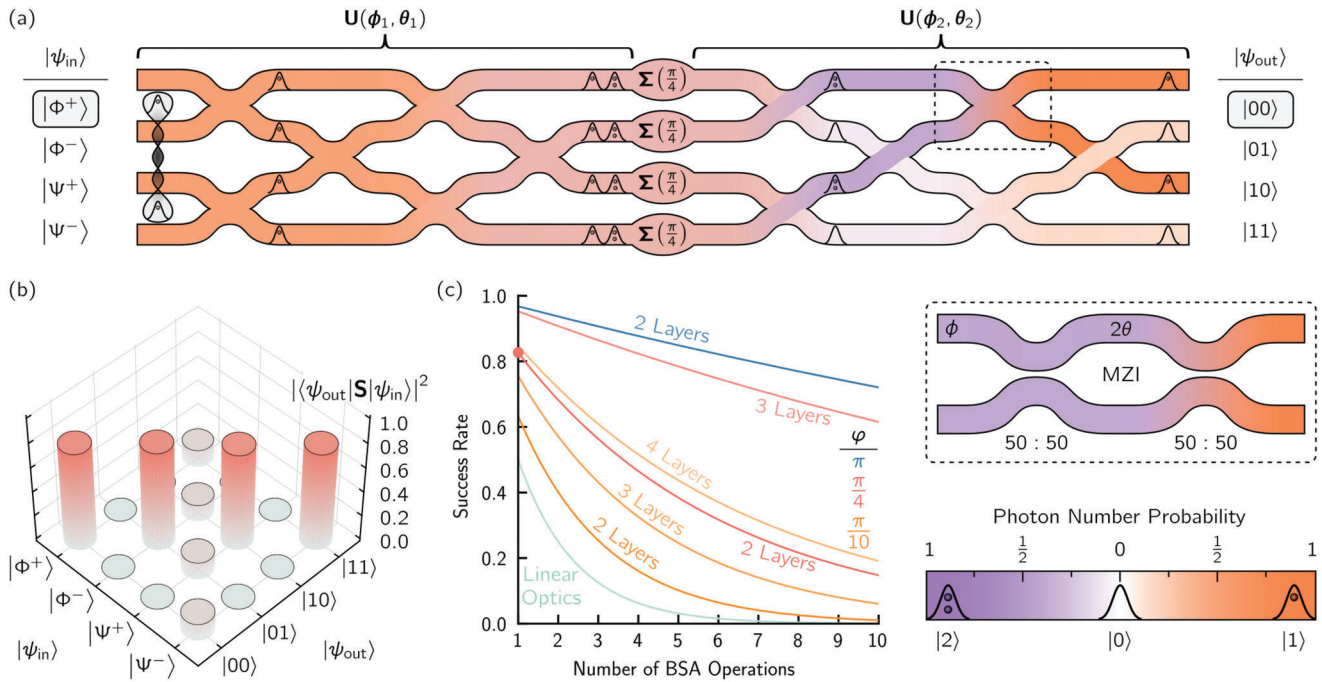


Figure 1. An imperfect QPNN-based BSA. a) An exemplary two-layer QPNN consisting of meshes (\mathbf{U}) of parameterized Mach–Zehnder interferometers (inset) separated by single-site nonlinearities (Σ). The network features two dual-rail encoded qubits, one where a single photon occupies the upper two spatial modes, the other in the lower two modes. Here, realistic losses (0.3 dB cm^{-1}) and errors in the 50 : 50 directional couplers (5.08%), as well as a weak $\pi/4$ nonlinear phase shift are assumed. The network was trained to act as a BSA according to the truth table shown, with a resultant unconditional fidelity of 0.825. As an example, the network is colored to portray the propagation of the photons through the network when the $|\Phi^+\rangle$ Bell-state is incident. The colors represent the probabilities that there are zero, one, or two photons in each spatial mode at each part of the network (colorbar), showing the evolution of the state as it propagates through the circuit. For this example, there is an 82.5% chance of measuring the correct $|00\rangle$ target state. b) Probabilities of measuring a state $|\psi_{out}\rangle$ when a state $|\psi_{in}\rangle$ is fed into the network shown in (a). c) Comparison between the success rates of ideal linear-optical and imperfect QPNN-based BSAs when up to ten are operated in series. The linear-optical BSA has a maximal unconditional fidelity of 0.5^[12] and is compared to imperfect QPNNs with varying amounts of layers and effective nonlinear phase shifts (ϕ), as explained in the main text. The pink marker highlights the success rate for the network shown in (a).

arbitrary linear unitary transformation $\mathbf{U}(\phi, \theta)$ on the spatial modes of the photons.^[11] Single-site nonlinearities, of strength ϕ , are placed between consecutive layers. In Section S1, Supporting Information, we provide further details on the construction of the system transfer function.

The components of linear photonic networks are not perfect, and various techniques have been developed to mitigate the effects of these imperfections. Specifically, both imperfect splitting ratios of the directional couplers (DCs) that form the MZIs and imperfectly calibrated phase shifters lead to errors that can be mitigated by optimizing the circuits after fabrication.^[16–18] In contrast, we account for these errors and those due to imbalanced photon loss or imperfect nonlinearities by training the variational parameters $\{\phi_i, \theta_i\}$ in situ, as would be done on-chip, post-fabrication.

We model a realistic linear mesh by allowing each element to uniquely suffer from slightly different imperfections, resulting in unbalanced, photon-path-dependent errors. We define the transmittance of each DC as t , randomly selected from a normal distribution with a mean of 50% and standard deviation of 5.08%, matching experimental results of a broadband DC fabricated for silicon-on-insulator (SOI) platforms.^[19] Likewise, propagation losses, where photons are scattered out of the circuit due

to, for example, surface roughness, or are absorbed by the waveguides, are parameterized by the fraction of light lost,

$$\alpha = 1 - 10^{-\frac{\alpha_{\text{WG}} \ell}{10}} \quad (1)$$

for an element of length ℓ and propagation losses in dB per unit length α_{WG} . These losses depend on the platform upon which the photonic circuit is constructed,^[20–26] which additionally determines the size of each photonic element. In our analysis, we randomly select α_{WG} values for each individual component from a normal distribution with a standard deviation of 6.67% of the mean, as is the case for current state-of-the-art photonic circuits built on SOI, which suffer from $\alpha_{\text{WG}} = 0.3 \pm 0.02 \text{ dB cm}^{-1}$ at 1550 nm.^[26] For this value of α_{WG} , the loss per MZI, of length $\ell_{\text{MZI}} = 287 \mu\text{m}$, is thus 0.009 dB on average. More information on the inclusion of fabrication imperfections, and all characteristic lengths, can be found in the Experimental Section and Section S1, Supporting Information.

In our architecture, as in previous realizations,^[7] a Kerr nonlinearity, resolved in the Fock basis as,

$$\Sigma(\phi) = \sum_n \exp \left[in(n-1) \frac{\phi}{2} \right] |n\rangle \langle n| \quad (2)$$

is assumed, where in the ideal case $\varphi = \pi$ such that a single photon passing through will experience no phase change while two photons will undergo a π phase change. To date, a π Kerr nonlinearity has yet to be observed at the single-photon level; however, this efficiency has been reached by other nonlinearities, such as those based on electromagnetically induced transparency^[27] or the saturation^[28] of atoms. While these enable neural networks capable of quantum state tomography^[29] or image recognition^[30] respectively, neither are scalable nor compatible with most quantum information processing applications since they lead to photon loss. It is, however, likely that in the near future, efficiencies approaching π will be demonstrated through the coherent chiral scattering of photons from single quantum emitters,^[31] by using integrated nanophotonic cavities designed to address this need,^[32–34] or with dispersion-engineered temporal trapping.^[35,36] Hence, we consider single-site Kerr nonlinearities but examine the performance of the QPNN also in the imperfect scenario where $\varphi \lesssim \pi$.

In sum, the total transfer function for an L -layer QPNN is given by

$$\mathbf{S} = \mathbf{U}(\boldsymbol{\phi}_L, \theta_L) \cdot \prod_{i=1}^{L-1} \boldsymbol{\Sigma}(\varphi) \cdot \mathbf{U}(\boldsymbol{\phi}_i, \theta_i) \quad (3)$$

This transfer function will act on any input state to produce an actual output state $|\psi_{\text{out,act}}^{(i)}\rangle = \mathbf{S}|\psi_{\text{in}}^{(i)}\rangle$, which is compared to the ideal output state $|\psi_{\text{out}}^{(i)}\rangle$ to determine the unconditional fidelity for that input–output pair

$$\mathcal{F}_i^{(\text{unc})} = \left| \langle \psi_{\text{out}}^{(i)} | \mathbf{S} | \psi_{\text{in}}^{(i)} \rangle \right|^2 \quad (4)$$

The complete unconditional fidelity, the chance that the network provides the correct output state for any given input state without preconditions, is then found by averaging over all K input–output pairs according to

$$\mathcal{F}^{(\text{unc})} = \frac{1}{K} \sum_{i=1}^K \mathcal{F}_i^{(\text{unc})} \quad (5)$$

Conversely, we can calculate the unconditional infidelity (i.e., network error) according to $\mathcal{C}^{(\text{unc})} = 1 - \mathcal{F}^{(\text{unc})}$, which we minimize in training the QPNN, using the local gradient-free BOBYQA nonlinear optimization algorithm,^[37] as available in the NLOpt library.^[38] Further details on network optimization are provided in the Experimental Section.

The success of the network may be conditioned on the detection of photons only in ports which abide by the dual-rail encoding scheme, as in this case we know that a logical output was produced. For the BSA shown in Figure 1a, this means that a single photon was detected in one of the top two modes and the other in the bottom two modes. We call this the conditional fidelity $\mathcal{F}^{(\text{con})}$, each i^{th} term of which is related to the unconditional fidelity by the probability that the network operation results in a computational basis state (that is, no photons are lost and each logical photonic qubit contains a single photon at the output) $\mathcal{P}^{(\text{cb})}$ by

$$\mathcal{F}_i^{(\text{unc})} = \mathcal{F}_i^{(\text{con})} \mathcal{P}_i^{(\text{cb})} \quad (6)$$

In the Experimental Section, we provide the expressions used to calculate these conditional measures. For a network operation that requires detection, such as the BSA and not, for example, the realization of a quantum logic gate for quantum computation, $\mathcal{P}^{(\text{cb})}$ gives the probability that the network successfully performed its task, while $\mathcal{F}^{(\text{con})}$ provides the quality of the result; in practice, one may optimize on either the conditional or unconditional infidelities, depending on the task under consideration. In the following, we train solely on the unconditional, however, we provide results from optimizing the conditional infidelity in Section S2, Supporting Information.

3. Correcting Imperfect Linear Interferometer Meshes

We begin to consider the effects of imperfections on QPNNs by holding the nonlinearity at the ideal value ($\varphi = \pi$) but introducing DC splitting ratio variations and unbalanced photon loss as described above. The resultant unconditional infidelity of a BSA for two to six-layer QPNNs with losses ranging from 0.001 to 3 dB cm^{−1}, as a function of training iterations, is shown in Figure 2a–c. Each case is repeated 50 times, resulting in plateaus of $\mathcal{C}^{(\text{unc})}$ that increase in value for increasing losses, as expected. Interestingly, for two-layer BSAs, we observe a large spread in the final $\mathcal{C}^{(\text{unc})}$, particularly for low-loss networks (cf., blue and purple curves in Figure 2a), indicating that the final performance of the QPNN is largely dictated by imbalance due to imperfect DCs. Adding more layers to the network, as in Figure 2b,c, reduces this spread, showing that larger QPNNs may learn to correct for these errors and more often reach optimal performance.

This is reflected in Figure 2d–f, which show the unconditional fidelity $\mathcal{F}^{(\text{unc})}$ as a function of waveguide loss for different sized networks. Here, we compare the in situ trained QPNNs of Figures 2a–c, denoted by the symbols, to the case where the network is trained offline. Offline training means that a perfect network was trained, then losses and DC errors were subsequently added to the solution. This was repeated 50 times for each α_{WG} , selecting different random imperfections at each repetition, with the mean given by the solid-black curve and standard deviation by the gray region. When trained in situ, the QPNN learns to overcome these imperfections as is seen by the convergence toward loss-limited performance (see Experimental Section for more information on network training and the loss limit). This is more apparent for larger networks, where the fidelity of those trained offline significantly reduces due to increased losses and DC errors, while those trained in situ maintain and even increase $\mathcal{F}^{(\text{unc})}$.

The balance between fabrication imperfections and network size, as a function of losses, is summarized in Figure 2g. Here, we observe that for state-of-the-art losses ($\alpha_{\text{WG}} = 0.3$ dB cm^{−1}) or worse, the unconditional fidelity decreases as expected when more layers are added to the network. When $\alpha_{\text{WG}} = 0.3$ dB cm^{−1}, $\mathcal{F}^{(\text{unc})} \geq 0.905$ (0.904), where the bracketed result is the lower bound of the 95% confidence interval, even for a six-layer QPNN, demonstrating that high-efficiency performance is possible on realistic state-of-the-art systems. Conversely, a more complex evolution is seen in Figure 2g for 0.01 dB cm^{−1} (0.0003 dB per MZI) losses or less, where there exists an optimal network size other than two layers. Losses at 0.01 dB cm^{−1} are similar to those of the silicon nitride platform for 1550 nm, reported as low as 0.007 dB

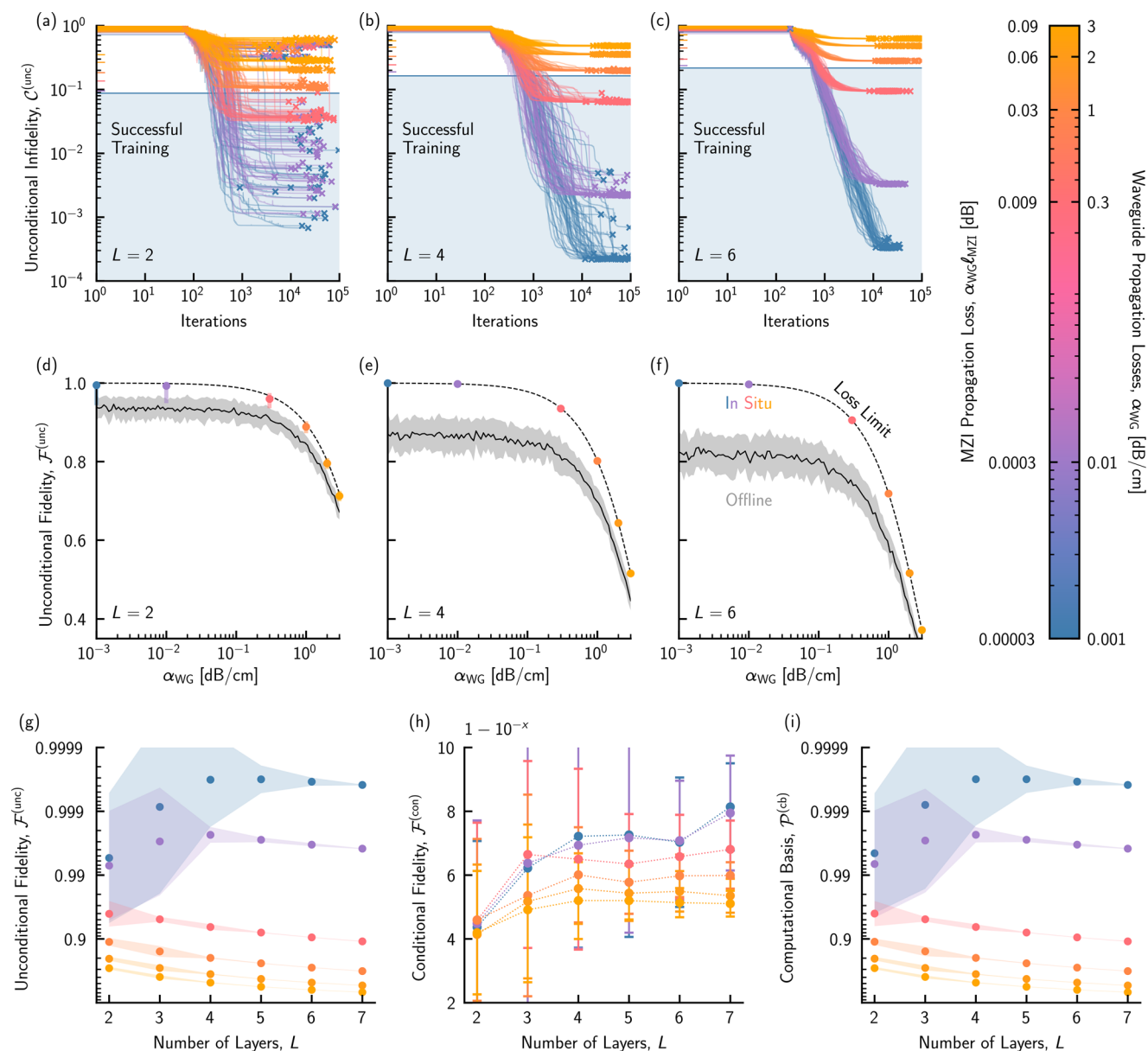


Figure 2. Performance of a QPNN-based BSA suffering from fabrication imperfections. The unconditional infidelity $C^{(unc)}$ of a) two, b) four, and c) six-layer networks are shown as a function of the training iteration for increasingly lossy networks. In each pane, the results of 50 optimization trials are displayed, with clear plateaus visible in $C^{(unc)}$ that increase with the losses. In each case, only trials that result in infidelity at or below those achieved by offline training (colored ticks in (a–c), shaded regions in (d–f)) are considered successful (shaded blue region shows an example for 0.001 dB cm^{-1}). The unconditional fidelity $F^{(unc)}$ of d) two, e) four, and f) six-layer networks are plotted with respect to the average losses α_{WG} , with colored symbols (shaded regions) corresponding to the mean (95% confidence interval) of a logarithmic normal distribution fitted to the successful trials of (a–c) (see Section S3, Supporting Information, for more details). These points are seen to lie on the (dashed) loss limit curve, where the performance of the network is only limited by uniform photon loss (assumes perfect DCs; see Experimental Section for additional details), in contrast to networks that are trained offline (solid black curves and shaded gray regions), demonstrating the ability of QPNNs to learn to overcome imperfections. g) Unconditional fidelity $F^{(unc)}$, h) conditional fidelity $F^{(con)}$, and i) computational basis probability $P^{(cb)}$, as a function of L for the QPNNs trained in situ, where the mean (symbols) and 95% confidence intervals (shaded regions in (g,i), error bars in (h)) are determined via the same method as (d–f). The losses per MZI, for each α_{WG} and $\ell_{MZI} = 287 \text{ }\mu\text{m}$, are displayed on the left side of the colorbar.

cm^{-1} ,^[39] but more typically near 0.01 dB cm^{-1} .^[40–42] In this low-loss case ($\alpha_{WG} = 0.01 \text{ dB cm}^{-1}$), $F^{(unc)}$ first increases from 0.993 (0.949) to 0.998 (0.997) by adding two layers to the base size, as the network is better able to account for imperfections. The un-

conditional fidelity then only slightly decreases to 0.996 (0.996) as the network grows to seven layers. That is, near-deterministic QPNN-based quantum elements such as BSAs will be realistic in the near-future as platform losses continue to decrease.

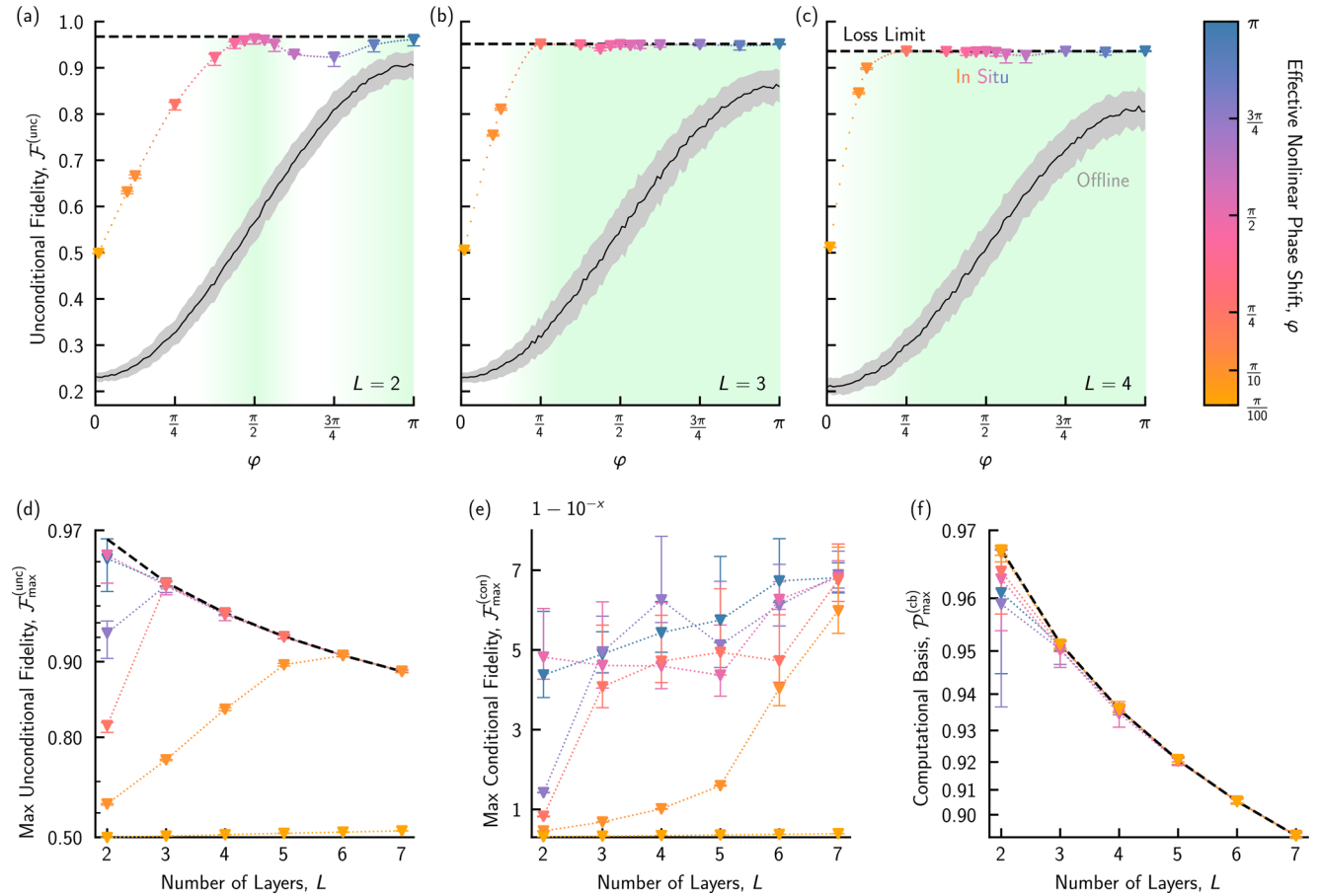


Figure 3. Performance of imperfect QPNN-based BSAs with sub-optimal ($\varphi \lesssim \pi$) nonlinearities and state-of-the-art ($\alpha_{\text{WG}} = 0.3 \text{ dB cm}^{-1}$) losses. The unconditional fidelity $\mathcal{F}^{(\text{unc})}$ of a) two, b) three, and c) four-layer networks is shown with respect to the effective nonlinear phase shift φ , showing both offline (solid black curves, shaded grey regions) and in situ (colored triangles) trained networks, and the loss limit (dashed line), as in Figure 2. In situ results include only the best-case training trials, where triangles (error bars) show the mean (95% confidence intervals) of a beta distribution fit to the maximal unconditional fidelity plateau (see Sections S3 and S4, Supporting Information, for statistical analysis details and an example of this plateau). Shaded green regions act as a visual aid to highlight the φ where near-loss-limited performance is achieved. The d) unconditional fidelity $\mathcal{F}^{(\text{unc})}$, e) conditional fidelity $\mathcal{F}^{(\text{con})}$, and f) computational basis probability $\mathcal{P}^{(\text{cb})}$ are plotted for each φ denoted on the colorbar, for networks of up to seven layers. All means (triangles) and 95% confidence intervals (error bars) were determined in the same manner as the best-case in situ results of (a–c). Connecting dotted lines serve only as a visual aid.

The situation is even more promising if the success of the network is preconditioned on detection in the computational basis, as is shown in Figure 2h,i. Here we present $\mathcal{F}^{(\text{con})}$ and $\mathcal{P}^{(\text{cb})}$ for different sized QPNNs, and for differing α_{WG} . Even for the lossy networks, where $\alpha_{\text{WG}} = 2 \text{ dB cm}^{-1}$ (0.06 dB per MZI), $\mathcal{F}^{(\text{con})}$ remains above 0.9999 (0.9888) for all $L \leq 7$, while for state-of-the-art losses this conditional fidelity does not drop below 0.999999 (0.999784) for $3 \leq L \leq 7$, as we observe in Figure 2h. Although the networks are trained to optimize their unconditional fidelity, adjusting the variational parameters to account for photon-path-dependent errors in the process, these near-unity results demonstrate that the conditional fidelity is not significantly decreased for any network size (see results from optimizing $\mathcal{F}^{(\text{con})}$ in Section S2, Supporting Information). In fact, as shown in Figure 2i, it is mainly the rate at which the QPNN produces a logical output that is affected by an increase in network size, showing the potential of even lossy networks if fault-tolerant protocols are used.

4. Embracing Weak Nonlinear Interactions

Having studied the effect of fabrication errors on network performance, we now turn to the consequences of sub-optimal nonlinearities. Assuming state-of-the-art losses ($\alpha_{\text{WG}} = 0.3 \text{ dB cm}^{-1}$), we vary the effective nonlinear phase shift φ from the ideal π to $\pi/100$ and attempt to train QPNNs of different sizes to act as BSAs (see Section S4, Supporting Information for exemplary training traces, cf. Figure 2a–c). For each network size and effective nonlinearity, we attempt to train 200 QPNNs, showing the results in Figure 3. Figure 3a–c depicts the highly non-trivial dependence of $\mathcal{F}^{(\text{unc})}$ on the effective nonlinearity φ . When the QPNN is trained offline, $\mathcal{F}^{(\text{unc})}$ increases monotonically with φ , as would be the case for a quantum-optical Fredkin gate-based BSA^[43–45] (see Section S4, Supporting Information, for additional information). Conversely, a QPNN can be trained to account for the weak nonlinearity, in which case it can vastly

outperform this expectation. Considering a two-layer network (Figure 3a), we observe a strikingly different φ dependence when comparing the best-case in situ trained networks (triangles; see Section S3, Supporting Information, for statistical analysis information) to those trained offline. We observe that networks trained in situ can reach the loss limit with sub-optimal nonlinearities, in addition to fabrication imperfections. Specifically, we observe optimal performance of two-layer QPNNs when $\varphi = \pi/2$ in addition to π . Moreover, as can be seen in Figure 3a, near-optimal performance is reached for a domain of φ centered at $\pi/2$, providing a pathway to robust QPNN-based BSAs without the need for a perfect Kerr nonlinearity. It must be noted, however, that operating with weaker nonlinearities decreases the probability that the QPNN converges at the loss limit during training, as is shown in Section S4, Supporting Information.

For all φ , a QPNN trained in situ learns how to account for weak nonlinearities and thus approach the loss limit. These capabilities improve as redundancies are added via an increase in network size, as visually evident across Figure 3a–c and summarized in Figure 3d. By adding a single additional (lossy) layer, QPNNs were trained to within 1.13% of the unconditional fidelity achieved with the ideal nonlinearity, 0.951 (0.950) at $\varphi = \pi$, and within 1.15% of the loss limit, 0.952, for all $\varphi \geq \pi/4$. Even networks with nonlinearities as weak as $\varphi = \pi/10$ approach the loss limit at six layers, in contrast to the case of $\pi/100$ where $\mathcal{F}^{(\text{unc})}$ increases only to 0.528 (0.528) at seven layers from 0.499 (0.498) at two layers, essentially acting as a linear-optical BSA.^[12] Trainability also improves with increased network size, as it becomes easier for the QPNN to find optimal solutions (see Section S4, Supporting Information, for more details).

In Figure 3e,f, the conditional fidelity and computational basis probability are shown as a function of $L \leq 7$, for differing φ . In contrast to the case where photon losses were varied (c.f. Figure 2), we observe that the behavior of $\mathcal{F}^{(\text{con})}$ strongly depends on φ . While QPNNs with $\varphi = \pi$ and $\pi/2$ operate with $\mathcal{F}^{(\text{con})} \geq 0.9999$ (0.9998) for all $L \leq 7$, networks with nonlinearities near $\pi/4$ and $3\pi/4$ require at least three layers to reach this level, while at $\pi/10$, six layers are needed. For all nonlinearities and network sizes, $\mathcal{P}^{(\text{cb})}$ is within 0.009 (0.031) of loss-limited performance, as seen in Figure 3f, and as expected for a QPNN suffering from state-of-the-art losses (cf. Figure 2d–f,i). Altogether, this demonstrates that for each combination of fabrication imperfections and effective nonlinearity, there exists an optimal network size that maximizes $\mathcal{F}^{(\text{unc})}$. While adding layers will always tend to increase $\mathcal{F}^{(\text{con})}$, a balance must be struck with the exponential decrease in $\mathcal{P}^{(\text{cb})}$. In the Section S5, Supporting Information, we demonstrate a QPNN trained to generate Greenberger–Horne–Zeilinger states, and show that this remains true beyond the BSA application.

5. Discussion

We have shown that high-fidelity operation is possible in imperfect quantum photonic neural networks based on non-ideal Kerr nonlinearities and realistic linear interferometer meshes. Since propagation through these networks leads to inevitable photon loss, their unconditional fidelity ceiling tends to decrease with increasing size. While this loss limit is unavoidable, these networks are able to learn to manage additional errors from non-

uniform losses and directional coupler splitting ratio variations, often demonstrating increased fidelity with the addition of imperfect layers. Crucially, we have shown that weak nonlinearities, which are mere fractions of the ideal, are sufficient for near-optimal network performance. Even as these sub-optimal nonlinearities are realized,^[31,32] the desired phase change will likely be accompanied by wave-packet distortions,^[46,47] and although complex solutions based on dynamically-coupled cavities^[33,34] or temporal confinement^[35,36] have been proposed, it remains an open question if, instead, a QPNN may learn to overcome them in much the same way it does fabrication imperfections. Already in the work presented here, QPNNs offer a fascinating view of the intricate balance between loss, imperfect photon routing and weak nonlinearity, which we have unraveled to demonstrate how each combination leads to an optimal network geometry. Understanding and respecting this balance will be important, in the near future, as QPNNs are designed and fabricated.

It is now clear why QPNNs far outperform linear-optical networks. Even with a weak $\pi/4$ effective nonlinearity, they can learn to surpass the 0.5 unconditional fidelity possible with perfect linear optics,^[12] achieving $\mathcal{F}^{(\text{unc})} = 0.820$ (0.809) at two layers (see Figures 1b and 3a), which grows to 0.951 (0.949) with an additional layer (see Figure 3b). At six layers, loss-limited operation, $\mathcal{F}^{(\text{unc})} = 0.905$ (0.904), can be achieved with nonlinearities as weak as $\pi/10$. Returning to Figure 1c, which summarizes the success rate of operating N BSAs in series, as would be necessary to connect quantum repeater nodes by entanglement swapping,^[13,14] the performance benefits offered by QPNNs become more apparent. While ten consecutive perfect linear-optical BSAs have a success rate of just 0.1%, six-layer, $\pi/10$ nonlinearity QPNNs reach 36.9%, and three-layer, $\pi/4$ networks achieve 60.5%.

Preconditioning the success of each QPNN-based BSA on the detection of two photons in the correct ports, as would be the case for generating cluster states from fusion gates,^[48] allows the much higher conditional fidelities to be leveraged. While $\mathcal{F}^{(\text{con})}$ for a perfect linear-optical ten-BSA sequence remains at 1, imperfect QPNNs of $\pi/4$ (three layers) and $\pi/10$ (six layers) nonlinearities reach 0.999 (0.997) and 0.99999 (0.99996), respectively. Given that these conditional fidelities are all near-unity, the rather large variations to $\mathcal{F}^{(\text{unc})}$ seen above can be attributed to the operational rate of the circuits (cf. Equation (6)), which are 369× improved when the perfect linear-optical BSAs are replaced by even six-layer, $\pi/10$ QPNNs. Hence, imperfect QPNNs are likely to play a key role in emerging large-scale quantum technologies.

6. Experimental Section

Modeling Fabrication Imperfections: An ideal MZI, as displayed in the inset to Figure 1a, could be described by a 2×2 matrix,

$$\begin{aligned} \mathbf{T}^{(\text{ideal})} &= \frac{1}{2} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} e^{i2\theta} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix} \\ &= e^{i\theta} \begin{pmatrix} e^{i\phi} \cos \theta & -\sin \theta \\ e^{i\phi} \sin \theta & \cos \theta \end{pmatrix} \end{aligned} \quad (7)$$

as is commonly found in the literature,^[11,17] up to the arrangement of components specified here. To model a realistic MZI, two types of imperfections: photon loss due to propagation and an imperfect splitting ra-

tio of the nominally 50 : 50 DCs were included. Imperfect phase shifter calibration was neglected as a QPNN trained in situ would intrinsically learn the phase shifts that account for these errors. A photonic element of length ℓ introduced the probability α that a photon was lost via propagation through it. The characteristic lengths used in the simulations were $\ell_{\text{MZI}} = 287 \mu\text{m}$ and $\ell_{\text{PS}} = 50 \mu\text{m}$. By Equation (1), α depends on the propagation losses per unit length, α_{WG} , selected from a normal distribution with a width of 6.67%, corresponding to the state-of-the-art experimental results for SOI.^[26] For each MZI, an individual α was computed, then applied through multiplication by the 2×2 matrix

$$\begin{pmatrix} \sqrt{1-\alpha} & 0 \\ 0 & \sqrt{1-\alpha} \end{pmatrix} \quad (8)$$

In Section S1, Supporting Information, further details are given for the inclusion of these non-uniform losses, including how the characteristic length (ℓ) of each photonic element was determined, and how the lack of unitarity was dealt with in the simulations. Similarly, each imperfect DC had an individual transmittance t that was taken from a normal distribution centered at 0.5 with a standard deviation of 0.0508, matching experimental results of a broadband DC fabricated for SOI platforms.^[19] For a given t , the corresponding 2×2 transformation of the DC is,

$$\begin{pmatrix} \sqrt{t} & \pm i\sqrt{1-t} \\ \pm i\sqrt{1-t} & \sqrt{t} \end{pmatrix} \quad (9)$$

Altogether, these result in a 2×2 transformation describing a realistic MZI,

$$\begin{aligned} T^{(\text{real})} &= \begin{pmatrix} \sqrt{1-\alpha} & 0 \\ 0 & \sqrt{1-\alpha} \end{pmatrix} \begin{pmatrix} \sqrt{t_2} & -i\sqrt{1-t_2} \\ -i\sqrt{1-t_2} & \sqrt{t_2} \end{pmatrix} \begin{pmatrix} e^{i2\theta} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{t_1} & i\sqrt{1-t_1} \\ i\sqrt{1-t_1} & \sqrt{t_1} \end{pmatrix} \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}, \\ &= \sqrt{1-\alpha} \begin{pmatrix} \sqrt{t_1 t_2} e^{i2\theta} e^{i\phi} + \sqrt{(1-t_1)(1-t_2)} e^{i\phi} & i\sqrt{t_1(1-t_2)} e^{i2\theta} - i\sqrt{t_2(1-t_1)} \\ -i\sqrt{t_2(1-t_1)} e^{i2\theta} e^{i\phi} + i\sqrt{t_1(1-t_2)} e^{i\phi} & \sqrt{(1-t_1)(1-t_2)} e^{i2\theta} + \sqrt{t_1 t_2} \end{pmatrix} \end{aligned} \quad (10)$$

In Section S1, Supporting Information, the regimes in α_{WG} were analyzed, L where the imperfect DC splitting ratios are dominant, and vice versa.

Network Optimization and Training Processes: A QPNN was trained to perform a mapping between a set of K input–output state pairs $|\psi_{\text{in}}^{(i)}\rangle \rightarrow |\psi_{\text{out}}^{(i)}\rangle$. For the QPNN-based BSA, the training set was provided in the computational basis in Figure 1a. Since dual-rail encoding was applied, $|0\rangle$ ($|1\rangle$) in the computational basis was equivalent to $|10\rangle$ ($|01\rangle$) in the Fock basis for the two spatial modes that realize the photonic qubit.

The unconditional infidelity of the network, $C^{(\text{unc})} = 1 - F^{(\text{unc})}$ (see Equations (4) and (5) for $F^{(\text{unc})}$), was minimized to facilitate the optimization process. The variational parameters, $\{\phi_i, \theta_i\}$ for each layer in the network, were selected randomly from a uniform distribution upon initialization. If the parameters were instead selected to initialize each linear mesh as a Haar random unitary matrix,^[49] convergence speed might be improved.^[50] Then, the local, gradient-free BOBYQA nonlinear optimization algorithm^[37] (available from the NLOpt library^[38]) was applied until the absolute change in infidelity was less than some threshold chosen empirically based on the available computational resources. This algorithm constructed a quadratic approximation to the infidelity and thus does not require an analytical gradient. Gradient-free optimization was deemed pertinent since it was unlikely that the internal state of the network, as would be necessary for backpropagation methods, would be accessible during in situ training.^[7] It was, however, relevant to note that recent works had demonstrated the efficacy of performing gradient descent on average by perturbing all parameters toward a random direction,^[51] and of employing

physics-aware training where backpropagation was applied to controllable physical systems.^[52]

In contrast to in situ training, as described in the main text, offline training was conducted by training a QPNN with idealized components, then adding fabrication imperfections, and if necessary, adjusting the effective nonlinearity (cf. Section 4). Due to the loss and DC splitting ratio variations, such imperfections were added to an idealized solution in 50 (200) repetitions in Figure 2 (3), matching the number of in situ trials conducted. From these results, an in situ trial was deemed successful if it achieved an optimized unconditional infidelity at or below the worst-case of offline training (mean minus standard deviation). Only successful optimization trials were considered for further analysis. Similarly, the loss limit is computed by adding imperfections to an idealized solution, however, losses are assumed to be completely uniform at α_{WG} , and the DC splitting ratios are all 50 : 50.

All simulations were conducted on the Frontenac Platform computing cluster offered by the Centre for Advanced Computing at Queen's University. The accompanying code was written in Python (version 3.10.2) using Numpy (version 1.22.2) and NLOpt (version 2.6.1). Cython (version 0.29.30) was used to translate performance-sensitive operations to C to improve computation runtime. It was identified in Section S1, Supporting Information, where computational complexity arises when constructing the system transfer function.

Conditional Measures: As for the unconditional fidelity, the conditional fidelity could be found by projecting the actual output state, $|\psi_{\text{out,act}}^{(i)}\rangle = S|\psi_{\text{in}}^{(i)}\rangle$, onto the computational basis, CB, and finding its overlap with the ideal output $|\psi_{\text{out}}^{(i)}\rangle$. Averaging over all K input–output pairs, this is written as,

$$F^{(\text{con})} = \frac{1}{K} \sum_{i=1}^K \left| \langle \psi_{\text{out}}^{(i)} | A^{(i)} S | \psi_{\text{in}}^{(i)} \rangle \right|^2 \quad (11)$$

where,

$$A^{(i)} = \left[\sum_{|x\rangle \in \text{CB}} \left| \langle x | S | \psi_{\text{in}}^{(i)} \rangle \right|^2 \right]^{-\frac{1}{2}} \quad (12)$$

normalizes the i^{th} $S|\psi_{\text{in}}^{(i)}\rangle$ to the computational basis. Similarly, the probability of measuring an output in the computational basis is

$$P^{(\text{cb})} = \frac{1}{K} \sum_{i=1}^K \sum_{|x\rangle \in \text{CB}} \left| \langle x | S | \psi_{\text{in}}^{(i)} \rangle \right|^2 \quad (13)$$

The i^{th} terms of Equations (11) and (13) could be multiplied to yield Equation (4), which follows simply from the fact that the i^{th} term of Equation (13) can be expressed as $(A^{(i)})^{-2}$.

Supporting Information

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

Acknowledgements

This research was supported by the Vector Scholarship in Artificial Intelligence, provided through the Vector Institute. The authors thank N. R. H. Pedersen for his insights into linear meshes, and gratefully acknowledge support by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Foundation for Innovation (CFI), and Queen's University.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

N.R. and J.C. conceived the project, which they developed along with J.E. J.E. was responsible for designing and performing all simulations and analysis, with supervision from B.S. and N.R. All authors discussed the results and shared in the writing and editing responsibilities for the manuscript.

Data Availability Statement

The data that support the findings of this study are openly available in the Borealis repository of the Queen's University Dataverse at <https://doi.org/10.5683/SP3/HRY0UO>, reference number 1.

Keywords

machine learning, neural networks, photonics, quantum information, quantum optics

Received: September 28, 2022

Revised: January 9, 2023

Published online:

- [1] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, S. Lloyd, *Phys. Rev. Res.* **2019**, 1, 033063.
- [2] B. J. Shastri, A. N. Tait, T. F. de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, P. R. Pruchal, *Nat. Photonics* **2021**, 15, 102.
- [3] J. Wang, S. Paesani, Y. Ding, R. Santagati, P. Skrzypczyk, A. Salavrakos, J. Tura, R. Augusiak, L. Mančinska, D. Bacco, D. Bonneau, J. W. Silverstone, Q. Gong, A. Acín, K. Rottwitz, L. K. Oxenløwe, J. L. O'Brien, A. Laing, M. G. Thompson, *Science* **2018**, 360, 285.
- [4] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, G. Carleo, *Nat. Phys.* **2018**, 14, 447.
- [5] A. Aspuru-Guzik, P. Walther, *Nat. Phys.* **2012**, 8, 285.
- [6] C. Sparrow, E. Martín-López, N. Maraviglia, A. Neville, C. Harrold, J. Carolan, Y. N. Joglekar, T. Hashimoto, N. Matsuda, J. L. O'Brien, D. P. Tew, A. Laing, *Nature* **2018**, 557, 660.
- [7] G. R. Steinbrecher, J. P. Olson, D. Englund, J. Carolan, *npj Quantum Inf.* **2019**, 5, 60.
- [8] D. Bondarenko, P. Feldmann, *Phys. Rev. Lett.* **2020**, 124, 130502.
- [9] R. Parthasarathy, R. T. Bhowmik, *IEEE Access* **2021**, 9, 103337.
- [10] R. D. Sipio, J.-H. Huang, S. Y.-C. Chen, S. Mangini, M. Worring, in *ICASSP 2022—2022 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Piscataway, NJ **2022**, p. 8612.
- [11] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, I. A. Walmsley, *Optica* **2016**, 3, 1460.
- [12] J. Calsamiglia, N. Lütkenhaus, *Appl. Phys. B* **2001**, 72, 67.
- [13] M. Żukowski, A. Zeilinger, M. A. Horne, A. K. Ekert, *Phys. Rev. Lett.* **1993**, 71, 4287.
- [14] K. Azuma, K. Tamaki, H.-K. Lo, *Nat. Commun.* **2015**, 6, 6797.
- [15] H. J. Kimble, *Nature* **2008**, 453, 1023.
- [16] D. A. B. Miller, *Optica* **2015**, 2, 747.
- [17] J. Mower, N. C. Harris, G. R. Steinbrecher, Y. Lahini, D. Englund, *Phys. Rev. A* **2015**, 92, 032322.
- [18] R. Hamerly, S. Bandyopadhyay, D. Englund, *Phys. Rev. Appl.* **2022**, 18, 024019.
- [19] Z. Lu, D. Celo, P. Dumais, E. Bernier, L. Chrostowski, in *2015 IEEE 12th Int. Conf. on Group IV Photonics (GFP)*, IEEE, Piscataway, NJ **2015**, p. 57.
- [20] R. Ding, T. Baehr-Jones, T. Pinguet, J. Li, N. C. Harris, M. Streshinsky, L. He, A. Novack, A. E.-J. Lim, T.-Y. Liow, S. H.-G. Teo, G.-Q. Lo, M. Hochberg, in *Optical Fiber Communication Conf., OFC/NFOEC*, IEEE, Piscataway, NJ **2012**, pp. 1–3.
- [21] K. Wörhoff, R. G. Heideman, A. Leinse, M. Hoekman, *Adv. Opt. Technol.* **2015**, 4, 189.
- [22] M. Melchiorri, N. Daldosso, F. Sbrana, L. Pavesi, G. Pucker, C. Kom-pocholis, P. Bellutti, A. Lui, *Appl. Phys. Lett.* **2005**, 86, 121111.
- [23] L. Cai, Y. Wang, H. Hu, *Opt. Lett.* **2015**, 40, 3013.
- [24] M. P. Nezhad, O. Bondarenko, M. Khajavikhan, A. Simic, Y. Fainman, *Opt. Express* **2011**, 19, 18827.
- [25] D. D'Agostino, G. Carnicella, C. Ciminelli, P. Thijs, P. J. Veldhoven, H. Ambrosius, M. Smit, *Opt. Express* **2015**, 23, 25143.
- [26] J. Cardenas, C. B. Poitras, J. T. Robinson, K. Preston, L. Chen, M. Lipson, *Opt. Express* **2009**, 17, 4752.
- [27] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, S. Du, *Optica* **2019**, 6, 1132.
- [28] X. Guo, T. D. Barrett, Z. M. Wang, A. I. Lvovsky, *Photonics Res.* **2021**, 9, B71.
- [29] Y. Zuo, C. Cao, N. Cao, X. Lai, B. Zeng, S. Du, *Adv. Photonics* **2022**, 4, 026004.
- [30] A. Ryou, J. Whitehead, M. Zhelyeznyakov, P. Anderson, C. Keskin, M. Bajcsy, A. Majumdar, *Photonics Res.* **2021**, 9, B128.
- [31] P. Lodahl, S. Mahmoodian, S. Stobbe, A. Rauschenbeutel, P. Schneeweiss, J. Volz, H. Pichler, P. Zoller, *Nature* **2017**, 541, 473.
- [32] H. Choi, M. Heuck, D. Englund, *Phys. Rev. Lett.* **2017**, 118, 223605.
- [33] M. Heuck, K. Jacobs, D. R. Englund, *Phys. Rev. Lett.* **2020**, 124, 160501.
- [34] M. Heuck, K. Jacobs, D. R. Englund, *Phys. Rev. A* **2020**, 101, 042322.
- [35] R. Yanagimoto, E. Ng, M. Jankowski, H. Mabuchi, R. Hamerly, *Optica* **2022**, 9, 1289.
- [36] I. Babushkin, O. Melchert, U. Morgner, A. Demircan, *Res. Square* **2022**, <https://doi.org/10.21203/rs.3.rs-1982423/v1>.
- [37] M. J. D. Powell, *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives*, Technical Report NA06, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge **2009**.
- [38] S. G. Johnson, The NLOpt nonlinear-optimization package, <http://github.com/stevengj/nlopt> (accessed: June 2021).
- [39] J. F. Bauters, M. J. R. Heck, D. D. John, M.-C. Tien, W. Li, J. S. Barton, D. J. Blumenthal, J. E. Bowers, A. Leinse, R. G. Heideman, in *2011 37th European Conf. and Exhibition on Optical Communication*, IEEE, Piscataway, NJ **2011**, pp. 1–3.
- [40] M. J. Shaw, J. Guo, G. A. Vawter, S. Habermehl, C. T. Sullivan, in *Micro-machining Technology for Micro-Optics and Nano-Optics III* (Eds: E. G. Johnson, G. P. Nordin, T. J. Suleski), SPIE, Bellingham, WA **2005**, p. 109.
- [41] D. J. Blumenthal, R. Heideman, D. Geuzebroek, A. Leinse, C. Roelof-fzen, *Proc. IEEE* **2018**, 106, 2209.
- [42] J. Liu, G. Huang, R. N. Wang, J. He, A. S. Raja, T. Liu, N. J. Engelsen, T. J. Kippenberg, *Nat. Commun.* **2021**, 12, 2236.
- [43] G. J. Milburn, *Phys. Rev. Lett.* **1989**, 62, 2124.

- [44] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge **2011**.
- [45] C. C. Gerry, P. L. Knight, *Introductory Quantum Optics*, Cambridge University Press, Cambridge **2004**.
- [46] J. H. Shapiro, *Phys. Rev. A* **2006**, 73, 062305.
- [47] J. Gea-Banacloche, *Phys. Rev. A* **2010**, 81, 043823.
- [48] D. E. Browne, T. Rudolph, *Phys. Rev. Lett.* **2005**, 95, 010501.
- [49] N. J. Russell, L. Chakhmakhchyan, J. L. O'Brien, A. Laing, *New J. Phys.* **2017**, 19, 033007.
- [50] S. Pai, B. Bartlett, O. Solgaard, D. A. B. Miller, *Phys. Rev. Appl.* **2019**, 11, 064044.
- [51] S. Bandyopadhyay, A. Sludds, S. Krastanov, R. Hamerly, N. Harris, D. Bunandar, M. Streshinsky, M. Hochberg, D. Englund, *arXiv:2208.01623*, **2022**.
- [52] L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, P. L. McMahon, *Nature* **2022**, 601, 549.
- [53] J. Ewaniuk, J. Carolan, B. J. Shastri, N. Rotenberg, *arXiv:2208.06571*, **2022**. <https://doi.org/10.5683/SP3/HRY0UO>.

ADVANCED QUANTUM TECHNOLOGIES

Supporting Information

for *Adv. Quantum Technol.*, DOI 10.1002/qute.202200125

Imperfect Quantum Photonic Neural Networks

*Jacob Ewaniuk**, *Jacques Carolan*, *Bhavin J. Shastri* and *Nir Rotenberg**

Supplementary Information for “Imperfect quantum photonic neural networks”

Jacob Ewaniuk,^{1,*} Jacques Carolan,² Bhavin J. Shastri,^{1,3} and Nir Rotenberg¹

¹*Centre for Nanophotonics, Department of Physics,
Engineering Physics & Astronomy, 64 Bader Lane,*

Queen’s University, Kingston, Ontario, Canada K7L 3N6

²*Wolfson Institute for Biomedical Research, University College London, London WC1E 6BT, UK*

³*Vector Institute, Toronto, Ontario, Canada, M5G 1M1*

(Dated: December 20, 2022)

S1. CONSTRUCTING THE SYSTEM TRANSFER FUNCTION FOR AN IMPERFECT QUANTUM PHOTONIC NEURAL NETWORK

The quantum photonic neural network (QPNN) architecture, as illustrated in Fig. 1a of the main manuscript, can be described via operation in the discrete $N = \binom{n+m-1}{n}$ -dimensional Fock basis as generated by the m spatial modes of n photons [S1]. Photonic qubits are dual-rail encoded such that $n = 2$ photons and $m = 4$ modes are required to establish the two-qubit inputs of the QPNN-based Bell-state analyzer (BSA). Explicitly, the set of two-photon Fock basis states is expressed as,

$$\{|2000\rangle, |1100\rangle, |1010\rangle, |1001\rangle, |0200\rangle, |0110\rangle, |0101\rangle, |0020\rangle, |0011\rangle, |0002\rangle\}, \quad (\text{S1})$$

where the subset,

$$\{|1010\rangle, |1001\rangle, |0110\rangle, |0101\rangle\}, \quad (\text{S2})$$

represents the computational basis such that $|10\rangle = |0\rangle_{\text{logic}}$, $|01\rangle = |1\rangle_{\text{logic}}$.

The i^{th} layer of a QPNN consists of a linear unitary transformation $\tilde{\mathbf{U}}(\phi_i, \theta_i)$ that is modelled according to the encoding scheme of Ref. [S2]. This scheme involves a rectangular mesh of Mach-Zehnder interferometers (MZIs), each corresponding to a unitary transformation $\mathbf{T}_{p,q}(\phi, \theta)$ on modes p, q , and a set of single-mode phase shifts at the output of the mesh as described by the diagonal unitary \mathbf{D} . In terms of these transformations, $\tilde{\mathbf{U}}(\phi_i, \theta_i)$ is defined as,

$$\tilde{\mathbf{U}}(\phi_i, \theta_i) = \mathbf{D} \prod_{(p,q) \in R} \mathbf{T}_{p,q}(\phi, \theta), \quad (\text{S3})$$

where R is a sequence of the $m(m-1)/2$ two-mode transformations, and ϕ, θ are elements of the corresponding vectors ϕ_i, θ_i that are selected according to the sequence [S2]. This $m \times m$ unitary transformation correctly describes the photonic circuit at layer i , however, state propagation is resolved into the N -dimensional basis of Eq. S1 in the simulations, where $N \neq m : n > 1$. Thus, a multi-photon unitary transformation, $\mathbf{U} = \Phi(\tilde{\mathbf{U}})$, is applied as defined by Eq. 6 of Ref. [S3]. This transformation requires the calculation of N^2 permanents of $n \times n$ matrices, making it by far the most computationally demanding process in the simulations [S1]. Between consecutive QPNN layers, single-site Kerr nonlinearities are assumed. These components generate an $N \times N$ diagonal unitary of the form,

$$\Sigma(\varphi) = \sum_n \exp\left[in(n-1)\frac{\varphi}{2}\right] |n\rangle \langle n|, \quad (\text{S4})$$

where φ is the effective nonlinear phase shift [S1]. By left-multiplying each transformation from the input to the output of the network architecture (c.f. Fig. 1a of the main manuscript), the system transfer function is defined as,

$$\mathbf{S} = \mathbf{U}(\phi_L, \theta_L) \cdot \prod_{i=1}^{L-1} \Sigma(\varphi) \cdot \mathbf{U}(\phi_i, \theta_i), \quad (\text{S5})$$

for a QPNN of L layers.

* jacob.ewaniuk@queensu.ca

Additional Details on the Fabrication Imperfection Model

Waveguide propagation losses are proportional to waveguide length [S4]. Thus, the characteristic lengths of the components in the QPNN architecture were identified as displayed in Fig. S1a. The single-site Kerr nonlinearities

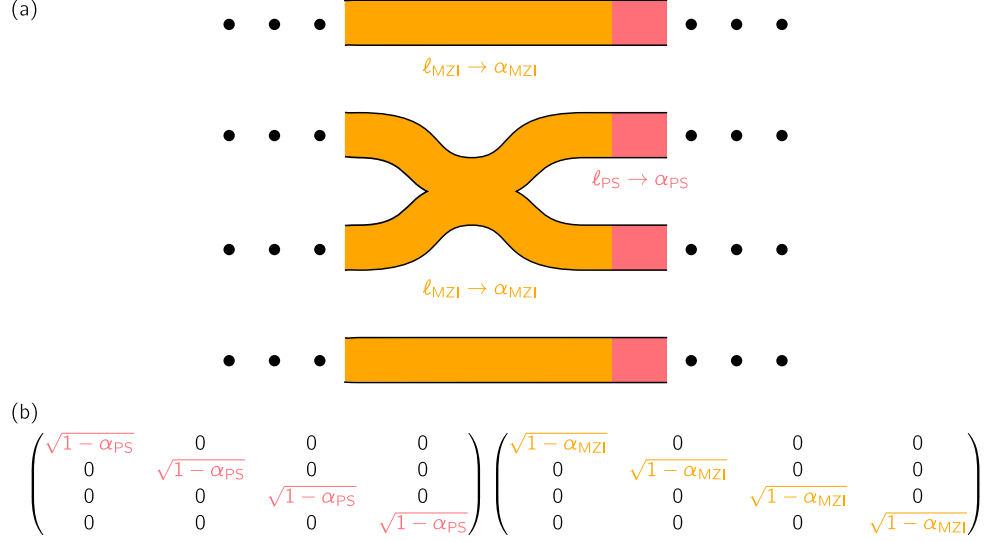


FIG. S1. (a) Diagrammatic representation of the characteristic lengths identified for the propagation loss model of an imperfect QPNN. Each component of length ℓ contributes the probability α of losing a photon. The flat sections adjacent to a MZI must share its length for the circuit to be balanced. (b) Matrices, with color-coded elements, constructed to apply losses in the simulations for the components shown in (a).

were assumed to be much shorter in length than the elements of a MZI mesh such that their losses were deemed negligible. To determine the characteristic lengths, we considered components fabricated for 1550 nm operation in silicon-on-insulator (SOI) that do not contribute excess loss. Specifically, the chosen broadband directional coupler (DC), with 5.08% splitting ratio variations as discussed in the main text, has a waveguide length of 93.342 μm when two 10 μm bends (input and output) are added to the design of Ref. [S5]. The phase shifters were modelled after the 50 μm -long thermo-optic indium tin oxide design of Ref. [S6], where < 0.01 dB insertion losses were reported (negligible excess loss). Since each MZI has two phase shifters and two DCs, the characteristic lengths are defined as $\ell_{\text{MZI}} = 287 \mu\text{m}$ and $\ell_{\text{PS}} = 50 \mu\text{m}$ respectively. These lengths are in agreement with those reported as typical in Ref. [S4] for SOI thermo-optic phase shifters and DCs when bend radii are chosen such that excess losses are negligible. It is necessary to reiterate that all fabrication imperfections are treated on a component-by-component basis in the simulations. This is shown diagrammatically in Fig. S2, where each DC throughout the network has an individual splitting ratio (see Fig. S2a) and each MZI, output phase shifter contributes a unique amount of loss (see Fig. S2b).

When losses are included ($\alpha_{\text{WG}} \neq 0$) in the QPNN simulations, the MZI transformation (c.f. Eq. 10 in the main manuscript) becomes non-unitary, and each linear transformation $\mathbf{U}(\phi_i, \theta_i)$ follows. When applied to a normalized state resolved in the Fock basis, the resultant state lacks normalization. At the output of the network, the ‘missing’ coefficient that would yield a normalized state can be regarded as attached to some state $|\epsilon_{\text{loss}}\rangle$, appended to the basis of Eq. S1, that accounts for all potential output states where $n < 2$ across the four spatial modes of interest. For our purposes, it is sufficient that the loss model accurately reduces the probability of measuring a state in the ideal Fock basis (Eq. S1), and thus that of measuring the target output state, without fully describing $|\epsilon_{\text{loss}}\rangle$.

The DC splitting ratio variations are independent of propagation losses. Thus, there exists a regime where the losses are small enough such that the splitting ratio variations become the dominant imperfection. This is evident in the main manuscript (Fig. 2g) when considering the increase to the unconditional fidelity $\mathcal{F}^{(\text{unc})}$ from $L = 2$ to $L = 4$ for $\alpha_{\text{WG}} \leq 0.01$ dB/cm. It is more explicitly shown in Fig. S3, where $\mathcal{F}^{(\text{unc})}$ is plotted for QPNNs of 2 to 6 layers. The shaded blue region highlights the regime where the DC splitting ratio variations are dominant. In this regime, an increase to the size of the QPNN does not yield a decrease in $\mathcal{F}^{(\text{unc})}$ for those trained *in situ*. In fact, the unconditional fidelity often increases due to the additional parameters that can be used to learn how to optimize around the fabrication imperfections. This further explains why $\mathcal{F}^{(\text{unc})}$ decreases for QPNNs trained offline, regardless of α_{WG} . Conversely, in the loss-dominant regime (shaded green region), additional layers only hinder the performance of the QPNN, given that it has ideal nonlinearities ($\varphi = \pi$) in this case, regardless of *in situ* or offline training.

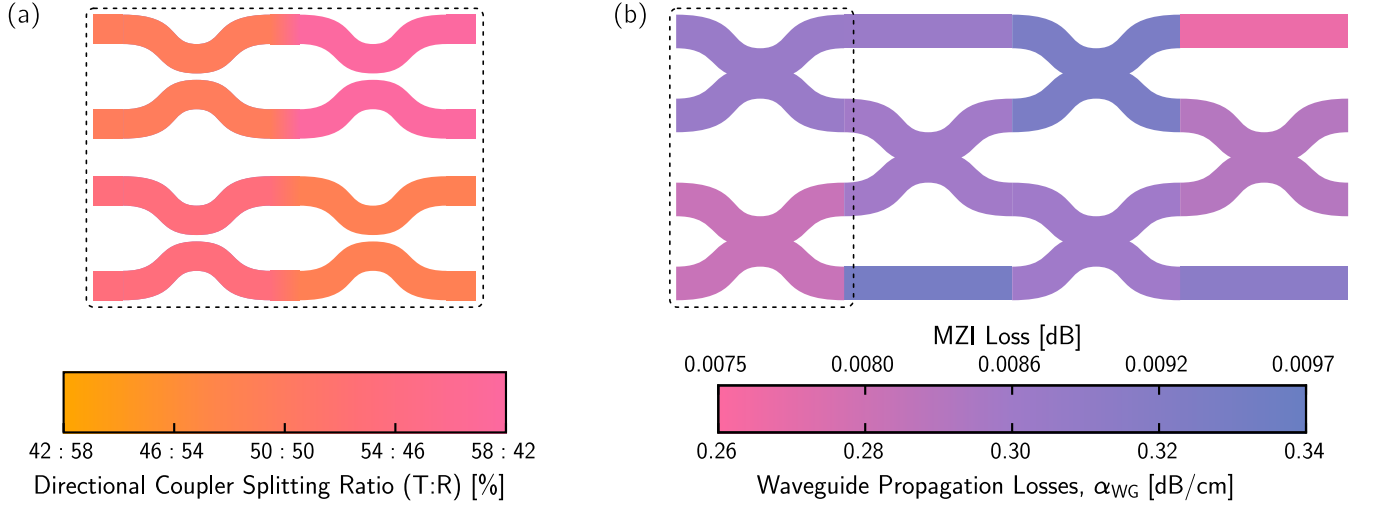


FIG. S2. Diagrammatic representation of varying fabrication imperfections throughout a QPNN, where the dotted box shows that DCs are contained within each MZI of a given linear mesh. (a) Example directional coupler splitting ratios selected randomly from a normal distribution with a mean (standard deviation) of 50% (5.08%). (b) Example waveguide propagation losses selected randomly from a normal distribution with a mean (standard deviation) of 0.3 dB/cm (0.02 dB/cm). The colorbar also displays the loss per MZI, calculated as $\alpha_{WG} \ell_{MZI}$ where $\ell_{MZI} = 287 \mu\text{m}$.

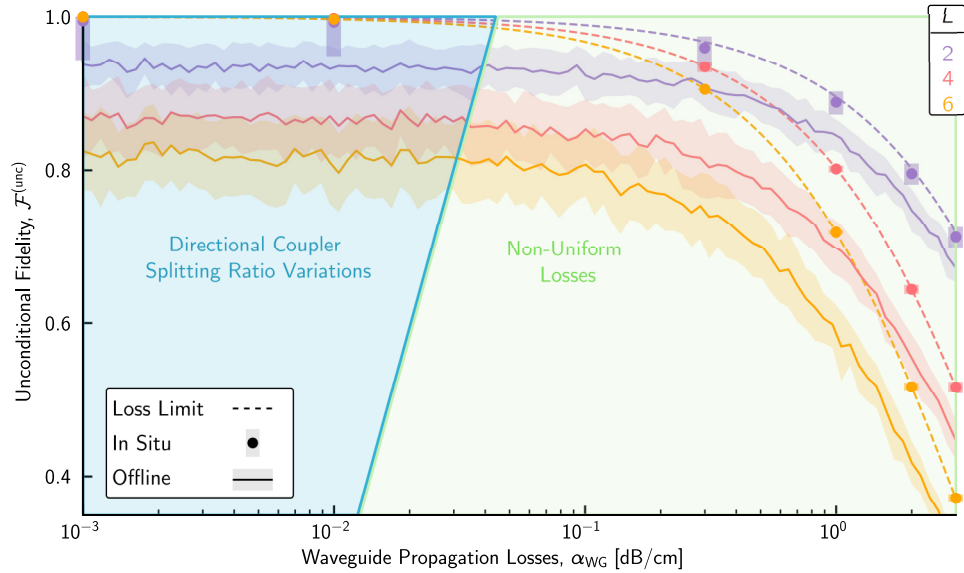


FIG. S3. Identifying dominant QPNN fabrication imperfections. The unconditional fidelity is plotted for QPNNs of 2, 4, and 6 layers, with ideal nonlinearities ($\varphi = \pi$), as a function of losses. Both *in situ* and offline-trained QPNNs feature loss and splitting ratio variations that match the model applied in the main manuscript. The shaded blue and green regions were added as a visual aid to show the domains in α_{WG} where DC splitting ratio variations and non-uniform losses dominate, respectively.

S2. OPTIMIZING THE CONDITIONAL FIDELITY

In Fig. S4a, we provide training traces for imperfect 2-layer QPNNs with weak nonlinearities ($\varphi \lesssim \pi$), assuming state-of-the-art losses ($\alpha_{\text{WG}} = 0.03$ dB/cm), when minimizing the conditional infidelity, $1 - \mathcal{F}^{(\text{con})}$, rather than the unconditional. All training procedures were chosen to match those explained in the main manuscript. Thus, the

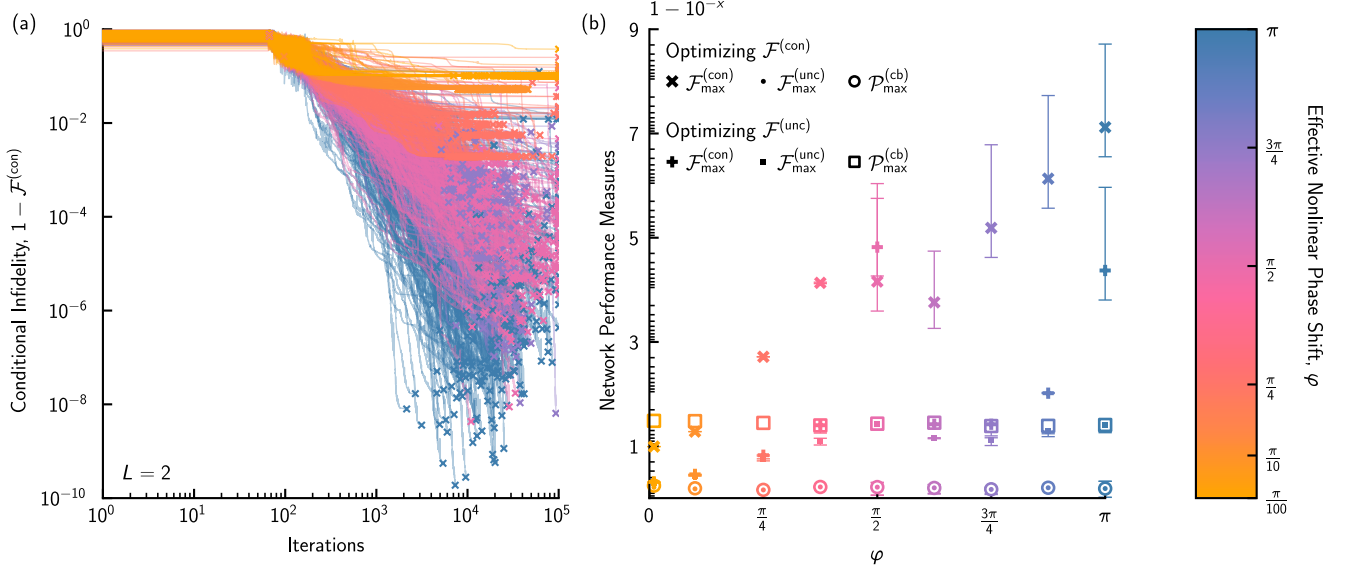


FIG. S4. Training imperfect 2-layer QPNNs with state-of-the-art losses ($\alpha_{\text{WG}} = 0.03$ dB/cm) and weak nonlinearities to optimize conditional fidelity $\mathcal{F}^{(\text{con})}$. (a) Training traces for QPNNs, with effective nonlinear phase shifts as labelled in the colorbar, when the conditional infidelity is minimized. (b) Conditional fidelities $\mathcal{F}^{(\text{con})}$, unconditional fidelities $\mathcal{F}^{(\text{unc})}$, and computational basis probabilities $\mathcal{P}^{(\text{cb})}$ for QPNNs of varying effective nonlinear phase shifts φ when trained to optimize $\mathcal{F}^{(\text{con})}$ and $\mathcal{F}^{(\text{unc})}$ respectively. Error bars show the 95% confidence intervals of a beta distribution (see Sec. S3 for more details).

absence of plateaus for $\varphi \geq \pi/2$ may be attributable to sub-optimal training parameters. The QPNN performance measures in this case are compared to those achieved when minimizing the unconditional infidelity, as in the main manuscript, in Fig. S4b. In general, when a QPNN is trained to minimize its conditional infidelity, it is able to achieve multiple order-of-magnitude improvements at the expense of significantly reduced operational efficiency. Taking $\varphi = 3\pi/4$ as an example, the conditional fidelity increases from 0.851 (0.846) to 0.998 (0.998) while the probability of yielding a logical output decreases from 0.964 (0.957) to 0.312 (0.312), correlating to a decrease in unconditional fidelity from 0.820 (0.809) to 0.311 (0.311), where the bracketed results show the lower bounds of 95% confidence intervals. Overall, if an application prioritizes $\mathcal{F}^{(\text{con})}$, imperfect QPNNs have the versatility to meet this desire.

S3. STATISTICAL ANALYSIS

In all imperfect QPNN simulations, there are trials where the optimization process reaches a local minimum and converges to a solution that can be deemed a failure. To eliminate these trials from the analysis of the network performance measures, successful training thresholds were computed for each combination of network size (L), losses (α_{WG}), and effective nonlinearity (φ), examples of which are displayed in Figs. 2a-c of the main manuscript and Figs. S8a-c. A given trial is deemed to have successfully trained the QPNN if the optimized unconditional fidelity $\mathcal{F}^{(\text{unc})}$ is greater than the lower bound (mean minus standard deviation) of the corresponding offline training result (c.f. lower shaded grey regions in Figs. 2d-f, 3a-c of the main manuscript). When considering imperfect QPNNs with weak nonlinearities (c.f. Fig. 3 of the main manuscript, Figs. S4, S9, S11), trials were often separated into a set of plateaus in unconditional infidelity, as visually evident in Fig. S8a. In these cases, the trials reaching the minimum plateau were further isolated for analysis, denoting the corresponding performance measures with the subscript ‘max’ (minimum infidelity corresponds to maximum fidelity). Once trials were appropriately isolated, the raw data was fit to a matching probability distribution to compute the results plotted in Figs. 2d-i, 3a-f of the main manuscript.

Fitting the Propagation Loss Results

In Fig. S5, we provide the raw data for the successfully-trained 2-layer imperfect QPNNs of ideal nonlinearities ($\varphi = \pi$) and 0.01 dB/cm losses, as an example. The offline training data is compared with the *in situ* in Fig. S5a, showing the general increase in unconditional fidelity that is achieved as the network learns to account for fabrication imperfections, thus approaching the loss limit. The *in situ* unconditional fidelity (Fig. S5a), conditional fidelity (Fig. S5b), and computational basis probability (Fig. S5c) data were respectively fit with logarithmic normal distributions using SciPy (version 1.8.1). The mean (95% confidence intervals) of each of these distributions provide the points (error bars or shaded regions) plotted in Fig. 2 of the main manuscript.

Fitting the Weak Nonlinearity Results

In Fig. S6, we provide the raw data for the successfully-trained 3-layer imperfect QPNNs of effective nonlinearity $\varphi = 3\pi/4$ and state-of-the-art 0.3 dB/cm losses, as an example. The offline training data is compared with the *in situ* in Fig. S6a, showing the general increase in unconditional fidelity that is achieved as the network learns to account for weak nonlinearities and fabrication imperfections, thus approaching the loss limit. There are clearly separated plateaus in unconditional fidelity (separate peaks in the blue all *in situ* histogram), as can also be viewed in terms of infidelity Fig. S8b. Only trials within the maximum unconditional fidelity plateau were further analyzed by fitting respective beta distributions to the unconditional fidelities (inset to Fig. S6a), conditional fidelities (Fig. S5b), and computational basis probabilities (Fig. S5c) using SciPy (version 1.8.1). The mean (95% confidence intervals) of each of these distributions provide the points (error bars or shaded regions) plotted in Fig. 3 of the main manuscript.

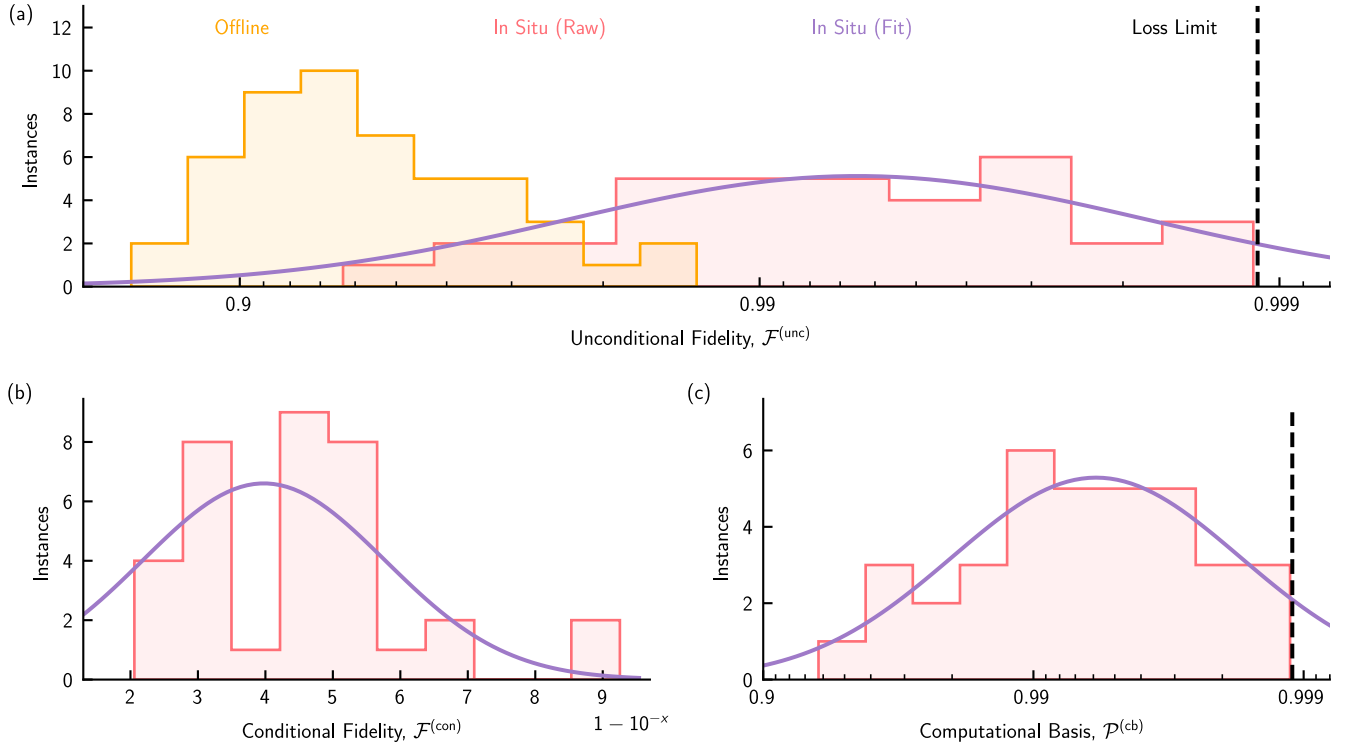


FIG. S5. Data analysis for successfully-trained 2-layer imperfect QPNNs of ideal nonlinearities ($\varphi = \pi$) and 0.01 dB/cm losses. (a) Histograms of the unconditional fidelities $\mathcal{F}^{(\text{unc})}$ for QPNNs trained offline (orange) and *in situ* (pink). The *in situ* data was fit with a logarithmic normal distribution (purple line), and the loss limit (dashed black line, also in (c)) is shown for comparison. The (b) conditional fidelities $\mathcal{F}^{(\text{con})}$ and (c) computational basis probabilities $\mathcal{P}^{(\text{cb})}$ for QPNNs trained *in situ* are given as histograms (pink) with respective logarithmic normal distribution fits (purple lines).

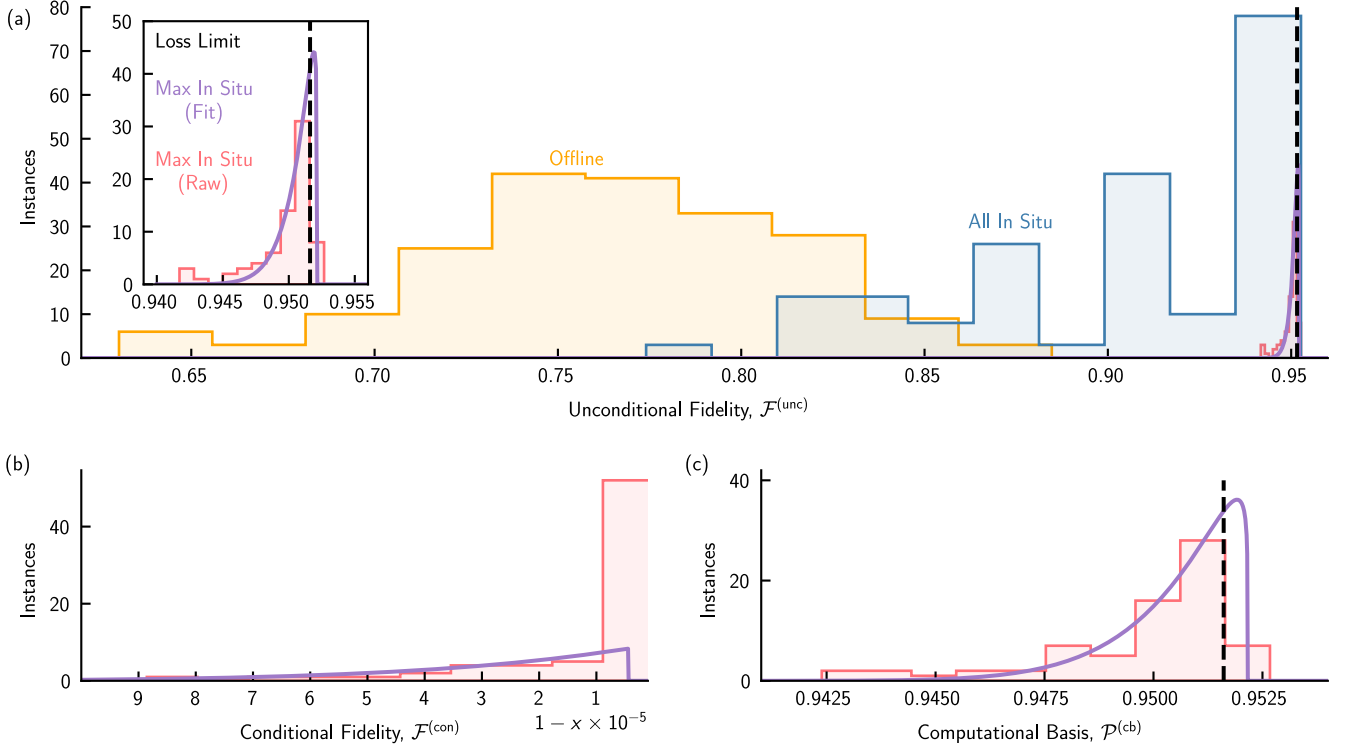


FIG. S6. Data analysis for successfully-trained 3-layer imperfect QPNNs of effective nonlinearity $\varphi = 3\pi/4$ and 0.3 dB/cm losses. (a) Histograms of the unconditional fidelities $\mathcal{F}^{(\text{unc})}$ for QPNNs trained offline (orange) and *in situ* (blue). A separate histogram (pink) is displayed that only contains the *in situ* results for QPNNs trained to the minimum plateau in unconditional infidelity (c.f. Fig. S8a), and the inset provides an expanded view of it. This histogram was fit with a beta distribution (purple line), and the loss limit (dashed black line, also in (c)) is shown for comparison. The (b) conditional fidelities $\mathcal{F}^{(\text{con})}$ and (c) computational basis probabilities $\mathcal{P}^{(\text{cb})}$ for QPNNs trained *in situ* are given as histograms (pink) with respective beta distribution fits (purple lines).

S4. TRAINING NETWORKS WITH WEAK NONLINEARITIES

In Fig. S7, we provide an alternate version of Fig. 3 from the main text that includes the average unconditional fidelities of all successful training trials (of 200 total), for each network size L and effective nonlinearity φ , across panels (a) to (c). Here, a trade-off between the network's trainability and its effective nonlinearity begins to become evident. While 2-layer, $\pi/2$ -nonlinearity QPNNs (see Fig. S7a) achieve loss-limited performance, it is difficult to attain this best-case result in a given training trial, as reflected by the much lower average unconditional fidelity. While in general, it is apparent that networks with effective nonlinearities less than the ideal tend to be more difficult to train, there exists a non-trivial relationship between the trainability, effective nonlinearity, and network size. Specifically, for 3-layer networks (see Fig. S7b), the additional layer improves the trainability of networks with $\varphi \sim \pi/2$ more drastically than for other φ , as can be seen by the larger increase in average unconditional fidelity. In fact, this increase suggests that the trade-off between trainability and effective nonlinearity is essentially removed for networks with $\varphi \sim \pi/2$ when a single additional layer is added.

In Figs. S8a-c, we provide exemplary training traces for imperfect QPNNs of 2 to 4 layers with state-of-the-art losses ($\alpha_{\text{WG}} = 0.3$ dB/cm) and weak Kerr nonlinearities ($\varphi \lesssim \pi$). For each effective nonlinear phase shift φ , 200 optimization trials were conducted. The trainability of these QPNNs is further examined in Figs. S8d-f, where the percentage of trials that reach a unconditional fidelity threshold is plotted over a domain of $\mathcal{F}^{(\text{unc})}$. Here, the difficulty of training QPNNs, with $\varphi < \pi$, to loss-limited performance is evident. QPNNs of 2 layers (Fig. S8d) can reach optimal performance with $\varphi = \pi/2$ in addition to π , however, the desired unconditional fidelity is only reached in 5.0% of trials for the former, compared to 39.5% of trials for the latter. By increasing the network size, the trainability improves in general, for all φ , at the cost of a decreased loss limit.

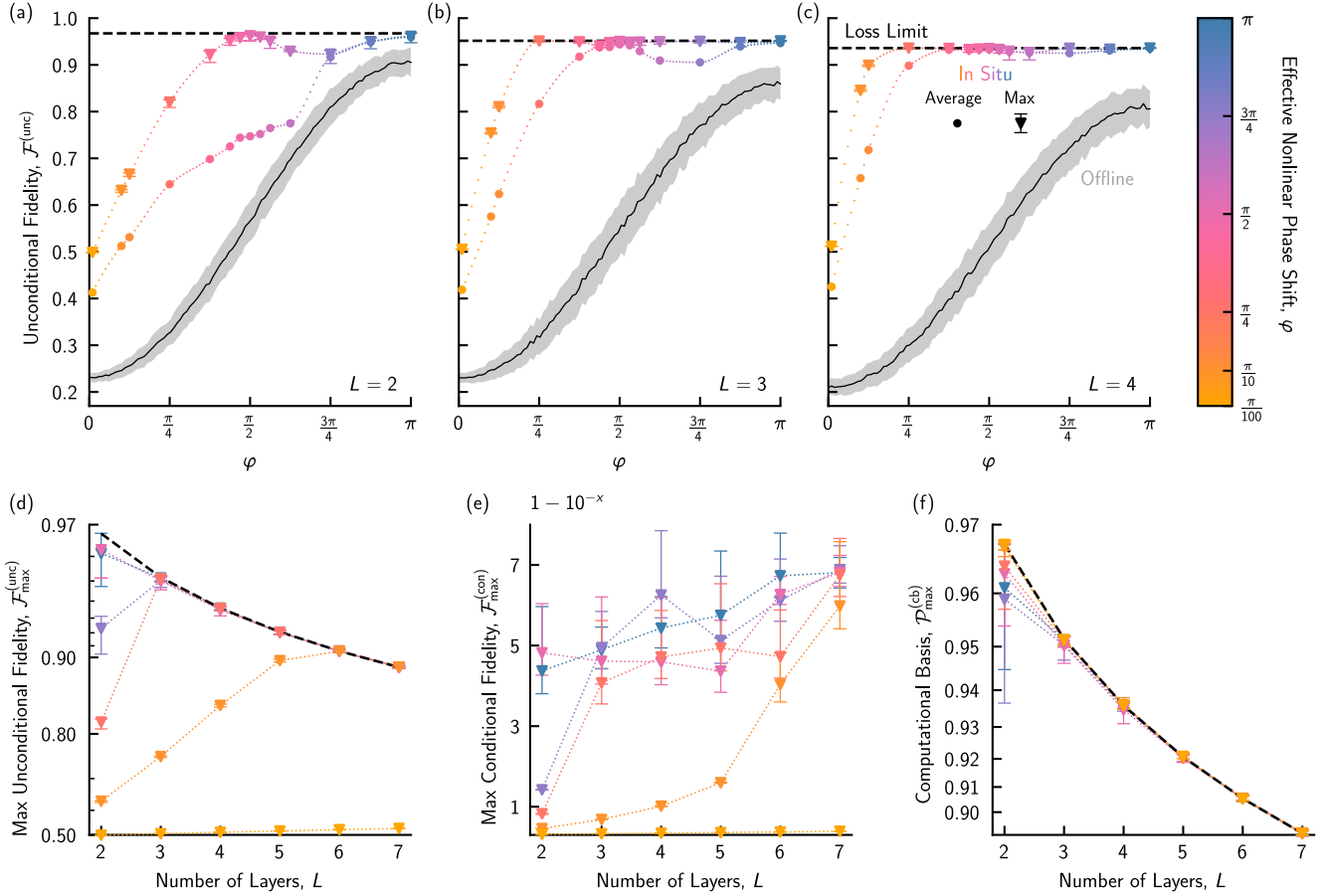


FIG. S7. Performance of realistic QPNN-based BSAs with sub-optimal ($\varphi \lesssim \pi$) nonlinearities and state-of-the-art ($\alpha_{\text{WG}} = 0.3$ dB/cm) losses. The unconditional fidelity $\mathcal{F}^{(\text{unc})}$ of (a) 2, (b) 3, and (c) 4-layer networks is shown with respect to the effective nonlinear phase shift φ , showing both offline (solid black curves, shaded grey regions) and *in situ* (colored curves, shaded regions) trained networks, and the loss limit (dashed line). *In situ* results include the average of all successfully-trained QPNNs (circles) and the best-case, where triangles (error bars) show the mean (95% confidence intervals) of a beta distribution fit to the maximal unconditional fidelity plateau (see the Supplementary Information S3, S4 for statistical analysis details and an example of this plateau). The (d) unconditional fidelity $\mathcal{F}^{(\text{unc})}$, (e) conditional fidelity $\mathcal{F}^{(\text{con})}$, and (f) computational basis probability $\mathcal{P}^{(\text{cb})}$ are plotted for each φ denoted on the colorbar, for networks of up to 7 layers. All means (triangles) and 95% confidence intervals (error bars) were determined in the same manner as the best-case *in situ* results of (a)-(c). Connecting dotted lines serve only as a visual aid.

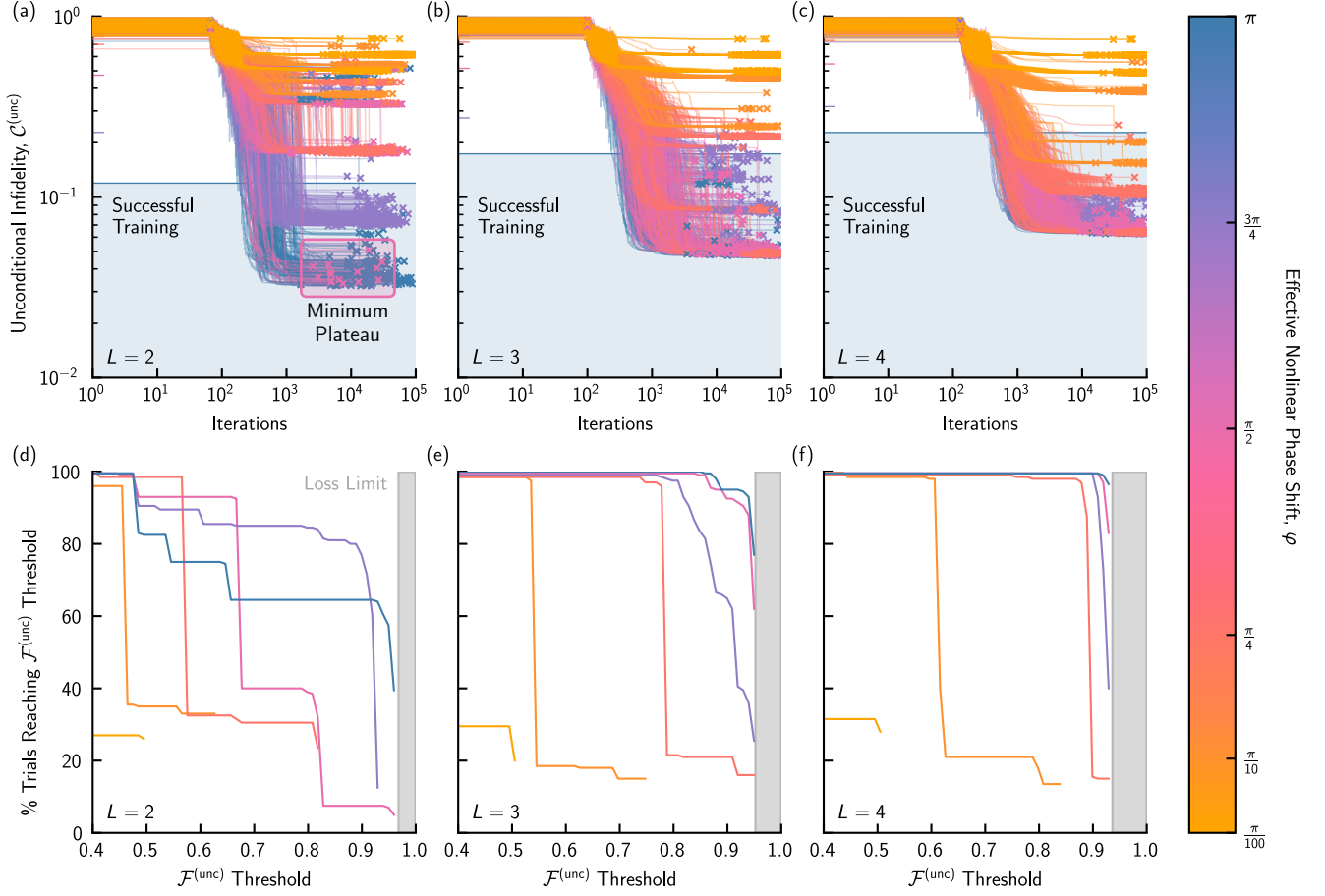


FIG. S8. Training data for imperfect QPNNs with state-of-the-art 0.3 dB/cm losses and weak nonlinearities ($\varphi \lesssim \pi$). The unconditional infidelity $\mathcal{C}^{(\text{unc})}$ of (a) 2, (b) 3, and (c) 4-layer networks are shown as a function of the training iteration for networks with effective nonlinearities as labelled in the colorbar. Colored ticks denote the thresholds computed to determine whether the QPNN was successfully-trained in a given trial, out of 200 trials each, with the shaded blue region displaying an example for $\varphi = \pi$. In (a), the minimum plateau in $\mathcal{C}^{(\text{unc})}$ for QPNNs of $\varphi = \pi/2$ is displayed as a visual aid. The trainability of (d) 2, (e) 3, and (f) 4-layer networks is examined by plotting the percentage of trials that achieve an unconditional fidelity threshold, over a domain of them. The shaded grey region depicts the $\mathcal{F}^{(\text{unc})}$ that cannot be achieved due to the loss limit.

Comparing Imperfect QPNNs with an Ideal Fredkin Gate-based BSA

A BSA can be constructed from a controlled-NOT (CNOT) gate followed by a Hadamard gate applied to the control qubit [S7]. With dual-rail encoded photonic qubits, the CNOT gate can be realized from a quantum-optical Fredkin gate [S8] which uses two 50:50 DCs (equivalent to Hadamard gates) and a nonlinear Kerr medium that connects the two photonic qubits as displayed in the inset to Fig. S9. However, it is only deterministic for an effective nonlinear phase shift $\varphi = \pi$ [S9]. Altogether, this circuit can be described by the unitary transformation,

$$\begin{aligned} \mathbf{U}_{\text{BSA}} &= (\mathbf{H} \otimes \mathbf{I}) \cdot (\mathbf{I} \otimes \mathbf{H}) \cdot \mathbf{\Sigma}(\varphi) \cdot (\mathbf{I} \otimes \mathbf{H}), \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \end{aligned} \quad (\text{S6})$$

in the computational basis of the two photonic qubits, where \mathbf{H} represents the Hadamard gate, \mathbf{I} is the identity, and $\mathbf{\Sigma}(\varphi)$ describes the transformation of effective nonlinearity φ conducted by the Kerr medium. With this BSA transformation, the unconditional fidelity of the Fredkin gate-based BSA is given by,

$$\mathcal{F}^{(\text{unc})} = \frac{1}{K} \sum_{i=1}^K \left| \langle \psi_{\text{out}}^{(i)} | \mathbf{U}_{\text{BSA}} | \psi_{\text{in}}^{(i)} \rangle \right|^2, \quad (\text{S7})$$

for the set of K input-output state pairs $|\psi_{\text{in}}^{(i)}\rangle \rightarrow |\psi_{\text{out}}^{(i)}\rangle$ that define the BSA operation. In this case, the unconditional fidelity is the same as the conditional since the model has considered ideal operation (i.e. no losses, DC splitting ratio variations) outside of a potentially sub-optimal φ . In Fig. S9, we compare the performance of a perfectly-fabricated Fredkin gate-based BSA to the unconditional fidelities of imperfect 2-layer QPNN-based BSAs with weak nonlinearities (c.f. Fig. 3a of the main manuscript), as a function of the effective nonlinear phase shift φ . The fidelity of the Fredkin gate-based BSA closely resembles the upper bound of QPNNs trained offline, yet is not hindered by the loss limit due to the absence of losses in the model described for it. Imperfect QPNNs trained *in situ* show improved performance for $\varphi \leq 3\pi/4$, reaffirming the learning capabilities demonstrated by the QPNN, even with weak nonlinearities. This comparison also shows why the non-monotonic relationship between $\mathcal{F}^{(\text{unc})}$ and φ for *in situ*-trained QPNNs, as described in the main manuscript, was unexpected.

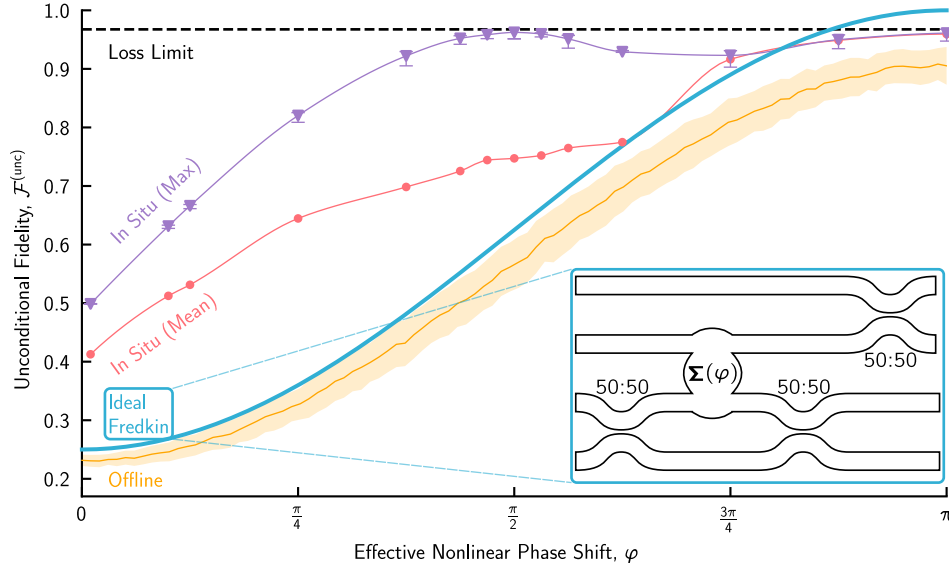


FIG. S9. Comparison between the respective unconditional fidelities of perfectly-fabricated quantum-optical Fredkin gate-based BSAs and imperfect 2-layer QPNN-based BSAs with state-of-the-art 0.3 dB/cm losses, as a function of effective nonlinear phase shift φ . The inset shows a photonic circuit diagram of the Fredkin gate-based BSA. Results are shown for QPNNs trained offline (orange line shows mean, shaded region shows standard deviation) and *in situ*, both as the mean of all trials (pink circles) and the results at the maximum unconditional fidelity plateau (purple triangles with error bars showing 95% confidence intervals), which are all constrained by the loss limit (dashed black line). Connecting lines for the *in situ* results serve as a visual aid.

S5. GENERATING GREENBERGER-HORNE-ZEILINGER STATES

Here, we present results for imperfect QPNNs trained to generate Greenberger-Horne-Zeilinger (GHZ) states. The training set only uses one input-output state pair,

$$|\psi_{\text{in}}\rangle = |000\rangle \rightarrow |\psi_{\text{out}}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \quad (\text{S8})$$

as in Ref. [S1], where the states are presented in the computational basis such that there are three photonic qubits (and thus six spatial modes, see the inset to Fig. S10d for a QPNN-based GHZ generator diagram). All training procedures were chosen to match those explained in the main manuscript. Thus, the absence of plateaus for $\alpha_{\text{WG}} \leq 0.01$ dB/cm in Figs. S10a-c may be attributable to sub-optimal training parameters. Fig. S10 shows the results of 2 to 4-layer QPNN-based GHZ generators with perfect nonlinearities ($\varphi = \pi$) and varied losses α_{WG} (c.f. Fig. 2 of the main manuscript), trained in 50 trials for each set of parameters. Networks trained *in situ* approach the loss limit as discovered for the QPNN-based BSA. Similarly, the unconditional fidelity and computational basis probability may increase with additional layers for losses ≤ 0.01 dB/cm. Fig. S11 shows the results of 2 to 4-layer QPNN-based GHZ generators with state-of-the-art losses ($\alpha_{\text{WG}} = 0.3$ dB/cm) and weak nonlinearities ($\varphi \lesssim \pi$) (c.f. Fig. 3 of the main manuscript), trained in 200 trials for each set of parameters. Networks trained *in situ* reach improved unconditional fidelities over offline training, and tend toward the loss limit, as discovered for QPNN-based BSAs. However, the relationship between $\mathcal{F}^{(\text{unc})}$ and φ is solely monotonic. Overall, for a given amount of losses and effective nonlinear phase shift, there similarly exists an optimal network size for QPNN-based GHZ generators.

-
- [S1] Steinbrecher, G. R., Olson, J. P., Englund, D. & Carolan, J. Quantum optical neural networks. *npj Quantum Inf* **5**, 60 (2019). URL <https://doi.org/10.1038/s41534-019-0174-7>.
 - [S2] Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. & Walmsley, I. A. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016). URL <http://opg.optica.org/optica/abstract.cfm?URI=optica-3-12-1460>.
 - [S3] Aaronson, S. & Arkhipov, A. The Computational Complexity of Linear Optics. *arXiv:1011.3245* (2010).
 - [S4] Sala, A. *Integrated Photonics* (OP-TEC, 2016).
 - [S5] Lu, Z., Celo, D., Dumais, P., Bernier, E. & Chrostowski, L. Comparison of photonic 2×2 3-dB couplers for 220 nm silicon-on-insulator platforms. In *2015 IEEE 12th International Conference on Group IV Photonics (GFP)*, 57–58 (2015). URL <https://ieeexplore.ieee.org/abstract/document/7305944>.
 - [S6] Parra, J., Hurtado, J., Griol, A. & Sanchis, P. Ultra-low loss hybrid ITO/Si thermo-optic phase shifter with optimized power consumption. *Opt. Express* **28**, 9393–9404 (2020). URL <http://opg.optica.org/oe/abstract.cfm?URI=oe-28-7-9393>.
 - [S7] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, 2011).
 - [S8] Milburn, G. J. Quantum optical Fredkin gate. *Phys. Rev. Lett.* **62**, 2124–2127 (1989). URL <https://link.aps.org/doi/10.1103/PhysRevLett.62.2124>.
 - [S9] Gerry, C. C. & Knight, P. L. *Introductory Quantum Optics* (Cambridge University Press, 2004).

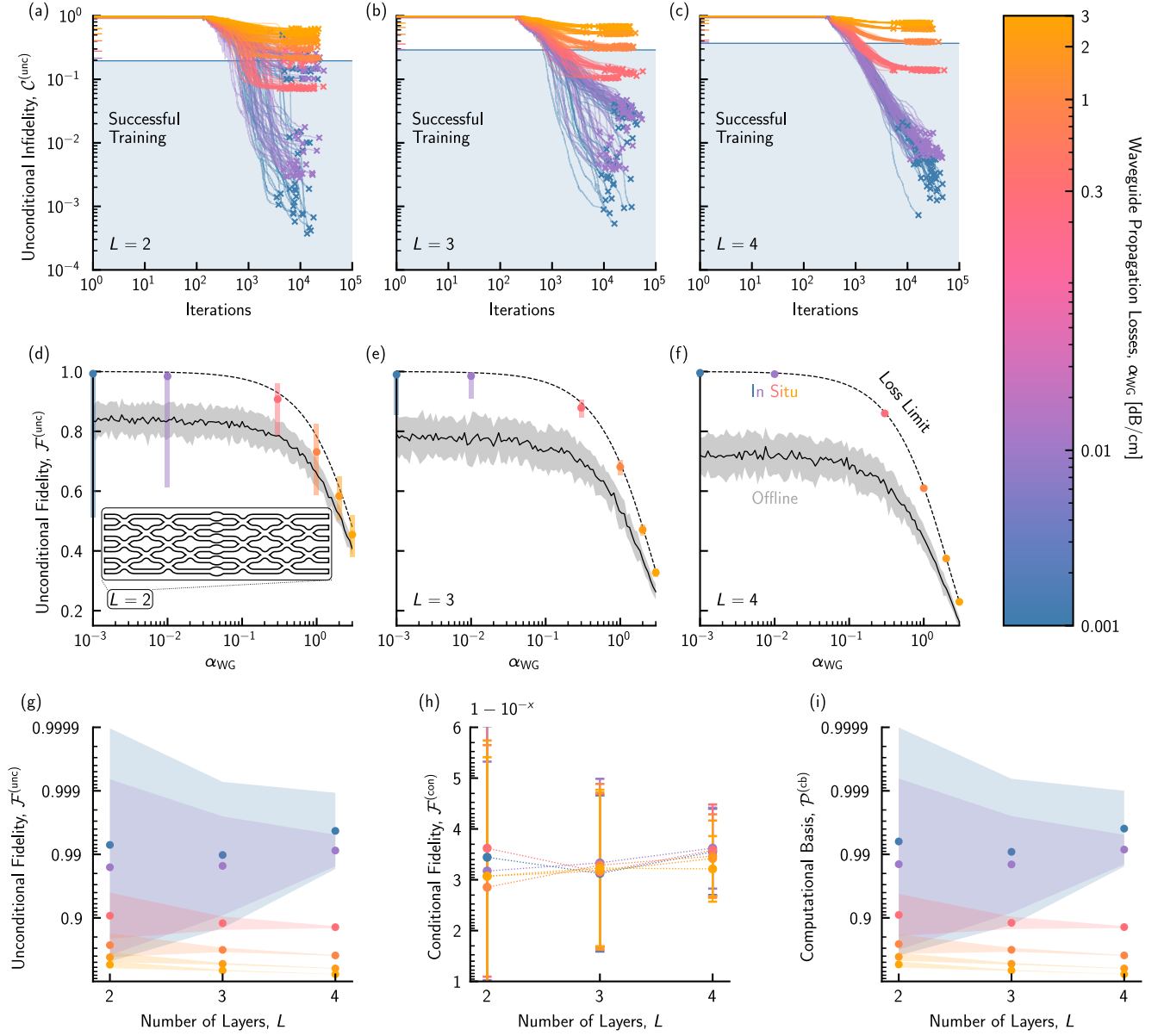


FIG. S10. Performance of a QPNN-based GHZ generator suffering from fabrication imperfections. The unconditional infidelity $\mathcal{C}^{(\text{unc})}$ of (a) 2, (b) 3, and (c) 4-layer networks are shown as a function of the training iteration for increasingly lossy networks. In each pane, the results of 50 optimization trials are displayed, with clear plateaus visible in $\mathcal{C}^{(\text{unc})}$ that increase with the losses. In each case, only trials that result in infidelity at or below those achieved by offline training (colored ticks in (a)-(c), shaded regions in (d)-(f)) are considered successful (shaded blue region shows an example for 0.001 dB/cm). The unconditional fidelity $\mathcal{F}^{(\text{unc})}$ of (d) 2, (e) 3, and (f) 4-layer networks are plotted with respect to the average losses α_{WG} , with colored symbols (shaded regions) corresponding to the mean (95% confidence interval) of a logarithmic normal distribution fitted to the successful trials of (a)-(c). These points are seen to lie on the (dashed) loss limit curve, where the performance of the network is only limited by uniform photon loss (assumes perfect DCs), in contrast to networks that are trained offline (solid black curves and shaded grey regions), demonstrating the ability of QPNNs to learn to overcome imperfections. (g) Unconditional fidelity $\mathcal{F}^{(\text{unc})}$, (h) conditional fidelity $\mathcal{F}^{(\text{con})}$, and (i) computational basis probability, $\mathcal{P}^{(\text{cb})}$, as a function of L for the QPNNs trained *in situ*, where the mean (symbols) and 95% confidence intervals (shaded regions in (g), (i), error bars in (h)) are determined via the same method as (d)-(f).

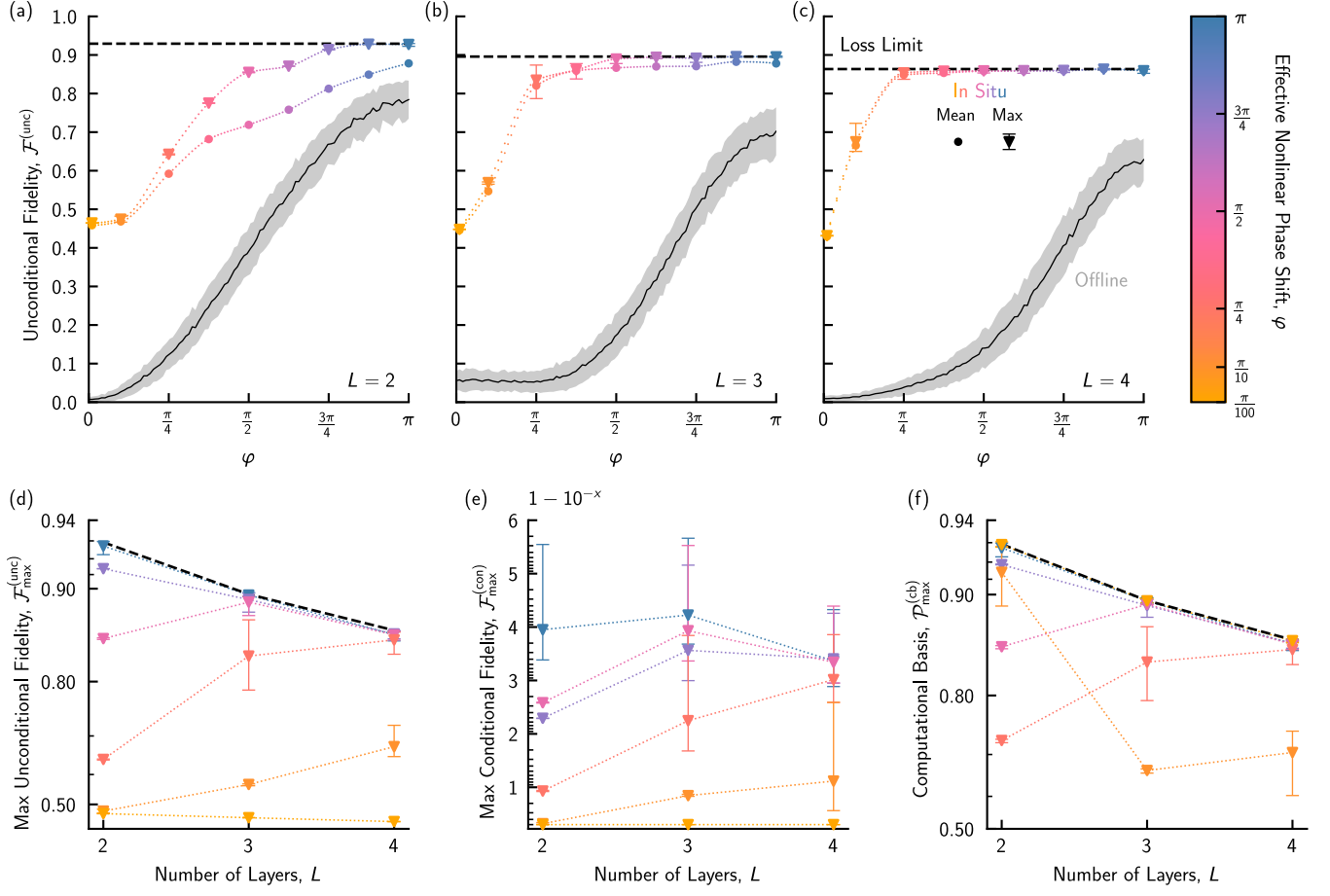


FIG. S11. Performance of imperfect QPNN-based GHZ generators with sub-optimal ($\varphi \lesssim \pi$) nonlinearities and state-of-the-art ($\alpha_{\text{WG}} = 0.3$ dB/cm) losses. The unconditional fidelity $\mathcal{F}^{(\text{unc})}$ of (a) 2, (b) 3, and (c) 4-layer networks is shown with respect to the effective nonlinear phase shift φ , showing both offline (black curves, shaded grey regions) and *in situ* (colored symbols) trained networks, and the loss limit (dashed line). *In situ* results include the average of all successfully-trained QPNNs (circles) and the best-case, where triangles (error bars) show the mean (95% confidence intervals) of a beta distribution fit to the maximal unconditional fidelity plateau. The (d) unconditional fidelity $\mathcal{F}^{(\text{unc})}$, (e) conditional fidelity $\mathcal{F}^{(\text{con})}$, and (f) computational basis probability $\mathcal{P}^{(\text{cb})}$ are plotted for each φ denoted on the colorbar, for networks of up to 4 layers. All means (triangles) and 95% confidence intervals (error bars) were determined in the same manner as the best-case *in situ* results of (a)-(c). Connecting dotted lines serve only as a visual aid.