


Production of photonic universal quantum gates enhanced by machine learning

Krishna Kumar Sabapathy, Haoyu Qi, Josh Izaac, and Christian Weedbrook
Xanadu, 777 Bay Street, Toronto, Ontario, Canada M5G 2C8

 (Received 8 October 2018; revised manuscript received 11 April 2019; published 17 July 2019)

We introduce photonic architectures for universal quantum computation. The first step is to produce a resource state which is a superposition of the first four Fock states with a probability greater than or equal to 10^{-2} , an increase by a factor of 10^4 over standard sequential photon-subtraction techniques. The resource state is produced with near-perfect fidelity from a quantum gadget that uses displaced squeezed vacuum states, interferometers, and photon-number-resolving detectors. The parameters of this gadget are trained using machine learning algorithms for variational circuits. We discuss in detail various aspects of the non-Gaussian state preparation resulting from the numerical experiments. We then propose a notion of resource farms where these gadgets are stacked in parallel, to increase the success probability further. We find a trade-off between the success probability of the farm, the error tolerance, and the number of gadgets. Using the resource states in conventional gate teleportation techniques, we can then implement weak tunable cubic phase gates. The numerical tools that have been developed could potentially be useful for other applications in photonics as well.

DOI: [10.1103/PhysRevA.100.012326](https://doi.org/10.1103/PhysRevA.100.012326)

I. INTRODUCTION

Continuous-variable systems are one of the leading candidates for realizing universal quantum computation. In particular, there has been considerable recent progress in theory and experiments, such as computation using temporally encoded modes [1–3] and frequency encoded domains [4], loop-based architectures [5–7], and bosonic-codes-based qubit or qumode encoding [8–13], and promising applications in quantum key distribution [14–16] and sensing [17,18].

Universal quantum computation in continuous-variable systems requires non-Gaussian gates generated by Hamiltonians that are beyond quadratic in the quadrature operators [19,20]. In this context the quadrature phase gates which are of the form $\Theta(\gamma) = \exp(i\gamma\hat{x}^n/\hbar)$ have played a very important role, especially the cubic phase gate corresponding to $n = 3$, which has the lowest-order non-Gaussian gate Hamiltonian among phase gates required for universal computation. Current methods to implement this gate are via gate teleportation where a resource state is prepared and used in a teleportation (measurement-based) circuit [21–28]. However, the methods of preparation of the resource state either require very high photon counting [21,29] or have a very low probability of success due to repeated photon subtractions, though some methods to improve the latter have been proposed recently [30–32].

We address this aspect of the resource state preparation wherein we constrain the resources to the simplest possible ones, namely, preparation of Gaussian multimode pure states followed by conditional photon detection measurements. For our purpose it suffices to consider only two-mode and three-mode architectures. We delegate the tuning of this multiparameter constrained variational circuit to a machine learning algorithm that trains the circuit to learn the required state preparation, inspired by recent works [33–38]. It turns out that we can prepare a learned state with near-perfect fidelity to the target state and with comparatively high probability, leading to a preparation efficiency greater than 1%, whereas

sequential photon-subtraction techniques have an efficiency of $\sim 10^{-4}\%$ (see Appendix A). We then estimate the resources required to build quantum resource farms as a possible route to near-deterministic state generation, depending on an error threshold, for implementing cubic phase gates.

The rest of the paper is structured as follows. In Sec. II we introduce the basic theory behind the gate teleportation method for implementing a weak cubic phase gate using the requisite resource state. In Sec. III we introduce architectures for quantum gadgets whose parameters are trained to generate the resource states using machine learning algorithms. We also provide in detail the physical interpretation of the numerical results. We then obtain in Sec. IV a trade-off between the number of these quantum gadgets in what we term a quantum resource farm and the total success probability of producing the state. We then analyze the effects of photon loss in the quantum gadgets on the output states in Sec. V. We summarize in Sec. VI. Explicit details of the numerical techniques are relegated to Appendix B.

II. GOTTESMAN-KITAEV-PRESKILL GATE TELEPORTATION

We now focus on the optical implementation of the lowest-order quadrature phase gate, namely, the cubic phase gate that we denote by $V(\gamma) = \exp(i\gamma\hat{x}^3/\hbar)$, where γ is the gate strength. All known methods to implement the cubic phase gate involve preparing a suitable resource state and using measurement-based techniques (see, for example, Table II of [22]). We use the teleportation technique of Gottesman *et al.* [21], but this can be translated into an adaptive gate teleportation presented using different gates and additional auxiliary squeezed states [26]. We take $\hbar = 2$ for the rest of the article.

For the cubic phase gate the resource state is the cubic phase state defined as $V(\gamma)|0\rangle_p$, which is nonphysical due to the zero-momentum ket $|0\rangle_p$. Therefore, as an approximation,

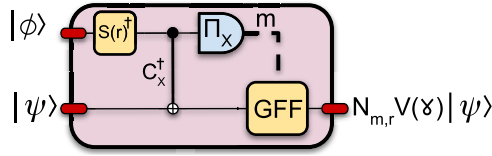


FIG. 1. The GKP teleportation for implementation of a cubic phase gate [21]. Here a state $|\psi\rangle$ is input to one mode along with a resource state $|\phi\rangle$ in the other. In addition, $\mathbf{S}(r)^\dagger$ denotes the squeezing operator, \mathbf{C}_x the controlled- X gate given by $\exp(-i\hat{x}_1\hat{p}_2/2)$, Π_x the x -homodyne measurement with outcome labeled m , \mathbf{GFF} the Gaussian feed-forward correction operator that needs to be applied, $\mathbf{N}_{m,r}$ the noise operator, and $\mathbf{V}(\gamma)$ the final cubic phase gate that is applied to the input state as shown in Eq. (4).

we consider $\mathbf{V}(\gamma)\mathbf{S}(r)^\dagger|0\rangle$, where $\mathbf{S}(r)$ is the standard single-mode squeezing gate given by $\mathbf{S}(r) = \exp[r(\hat{a}^2 - \hat{a}^{\dagger 2})/2]$. For large squeezing, this state has a large Fock support and hence would be difficult to synthesize directly. To get around this we commute the squeeze operator across the cubic phase gate to obtain $\mathbf{S}(r)^\dagger\mathbf{V}(\gamma')|0\rangle$. Let us further assume that $\gamma' \ll 1$; then we can expand the cubic phase gate to first order in gate strength to obtain $\mathbf{S}(r)^\dagger[1 + i\gamma'\hat{x}^3/2]|0\rangle$. Assuming that we can apply the on-line squeezing gate using methods such as measurement-based squeezing [39–41], we call the remaining terms the resource state [27,28] and it can be expanded in the Fock basis as

$$|\phi\rangle = \frac{1}{\sqrt{1+5|a|^2/2}} \left[|0\rangle + ia\sqrt{\frac{3}{2}}|1\rangle + ia|3\rangle \right], \quad (1)$$

where $a \in \mathbb{R}$.

We now use this resource state in a Gottesman-Kitaev-Preskill (GKP) teleportation scheme as depicted in Fig. 1. The wave function of the squeezed resource state is

$$\tilde{\phi}(x) = \langle x|\mathbf{S}(r)^\dagger|\phi\rangle = \int dx' \phi(x') \langle x|\mathbf{S}(r)^\dagger|x'\rangle. \quad (2)$$

We note that $\langle x|\mathbf{S}(r)^\dagger|x'\rangle = e^{r/2}\langle x|e^r|x'\rangle$. So we have that $\tilde{\phi}(x) = e^{-r} \int dy \phi(e^{-r}y) e^{r/2} \delta(x-y) = e^{-r/2} \phi(e^{-r}x)$. Then the output state can be derived [similar to Eq. (8) of Ref. [22]] as

$$|\psi_{\text{out}}\rangle = N' \exp\left[-\frac{(\hat{x}+m)^2}{4e^{2r}}\right] \left[1 + i\frac{\gamma}{2}(\hat{x}+m)^3\right] |\psi_{\text{in}}\rangle, \quad (3)$$

$$\gamma = 2ae^{-3r}/\sqrt{6},$$

where N' is the normalization factor, m is the homodyne measurement outcome, and $\mathbf{C}_x = \exp(-i\hat{x}_1\hat{p}_2/2)$ is an entangling gate that can be implemented [2,41–44]. We now assume that $\gamma \ll 1$, allowing us to approximate the terms in the second set of square brackets as a first-order expansion in the gate strength of a cubic phase gate, resulting in

$$|\psi_{\text{out}}\rangle = N' \exp\left[-\frac{(\hat{x}+m)^2}{4e^{2r}}\right] \exp\left[\frac{i\gamma(\hat{x}+m)^3}{2}\right] |\psi_{\text{in}}\rangle.$$

Expanding the terms in the second operator and applying a Gaussian feed-forward $\mathbf{GFF}(m) = \exp[-i\gamma(3m\hat{x}^2 + 3m^2\hat{x} + m^3)/2]$, we obtain the final action on the input state to be

$$|\psi_{\text{out}}\rangle = N'\mathbf{N}(m,r)\mathbf{V}(\gamma)|\psi_{\text{in}}\rangle, \quad (4)$$

where $\mathbf{N}(m,r) = \exp[-(\hat{x}+m)^2/4e^{2r}]$ is the Gaussian damping noise operator that depends on the homodyne measurement outcome m .

So using the resource state $|\phi\rangle$, we can effect a transformation which is a weak cubic phase gate along with an unavoidable Gaussian noise factor. Note that the initial squeezing gate $\mathbf{S}(r)^\dagger$ not only reduces the strength of the final cubic phase gate but also negates the effect of the Gaussian noise operator as seen from Eq. (3).

III. MACHINE LEARNING FOR STATE PREPARATION

The resource state $|\phi\rangle$ can be prepared in the laboratory using standard sequential photon-subtraction and photon-addition techniques [45–49]. However, such a method is not scalable since the successful probability of three successive photon additions or subtractions is extremely low, due to the use of very-high-transmission beam splitters [45,50], as discussed in Appendix A.

To bypass this difficulty, we introduce very different architectures that use Gaussian states conditioned on non-Gaussian postselected photon-number-resolving (PNR) detectors, akin to Gaussian boson sampling circuits [51,52]. Early works considered conditional measurements on the outputs of beam splitters for state preparation [53–59].

Algorithm 1 StatePrep_3mode

```

function loss(x, para, handle):
    # x = {x_s, x_d, x_\theta} parameters for squeezing,
    # displacement and beam splitter array
    para = {a, m_1, m_2, cutoff_dim}
    |\psi_{in}\rangle \leftarrow \mathbf{S}(x_s)\mathbf{D}(x_d)|\mathbf{0}\rangle
    |\psi_{out}\rangle \leftarrow \mathbf{U}(x_\theta)|\psi_{in}\rangle
    normIn, normOut \leftarrow |\langle\psi_{in}|\psi_{in}\rangle|, |\langle\psi_{out}|\psi_{out}\rangle|
    penalty = 100 \times |1-normIn| + 100 \times |1-normOut|
    |\psi\rangle \leftarrow \langle m_1, m_2|\psi_{out}\rangle
    prob \leftarrow \langle\psi|\psi\rangle
    |\psi\rangle_L \leftarrow |\psi\rangle/prob
    fid = |\langle\psi_L|\phi_T\rangle|^2
    if handle == 'fid' then
        return -fid + penalty
    else if handle == 'fid_prob' then
        return -fid + prob + penalty
    end if
end function
procedure optimization(para, niter):
    # global exploration to optimize fidelity
    # basinhopping is a global search algorithm
    # further optimize the probability by local search
    initialize x_0 # randomly chosen with proper range
    x_1 \leftarrow basinhopping(x_0, loss, args=(para, 'fid'), niter)
    x_2 \leftarrow local_search(loss, x_1, args=(para, 'fid_prob'))
    save x_2
end procedure

```

Our circuits are depicted in Fig. 2 for both the two-mode and three-mode architectures. We then use machine learning algorithms and the STRAWBERRY FIELDS quantum simulator [60] to train these circuits against the target state given in Eq. (1), as presented in Algorithm 1, where $|\phi_T\rangle$ denotes the

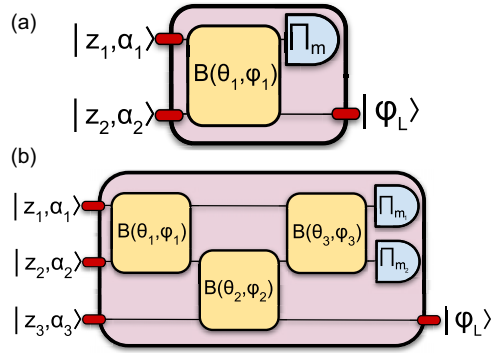


FIG. 2. Quantum gadgets of (a) two and (b) three modes. Here $\{z_i, \alpha_i\} \in \mathbb{C}$ denotes the input squeezing $S(z)$ and displacement $D(\alpha)$ gates applied to the vacuum states. In addition, $B(\theta, \phi) = \exp(e^{i\phi} a_1 a_2^\dagger - e^{-i\phi} a_1^\dagger a_2)$ denotes a beam splitter and Π_n a photon-number-resolving detector postselected to value n . Further, $|\phi_L\rangle$ is the output after training the circuit parameters.

target state and $|\psi_L\rangle$ denotes the learned or trained state. The algorithm executes a two-step optimization, where the circuit parameters are first trained to maximize the fidelity with the target state using `basinhopping`, which is a global search heuristic. The second step is then to perform a `local_search` starting from the global optimum found by `basinhopping`, to further increase the probability of producing the trained state. We choose the cutoff dimension of each mode in Fig. 2 to be 15 such that there is a large enough Hilbert space to be explored but without too much overhead. The PNR detectors in the two-mode case are set to $m = 2$ and in the three-mode case to $(m_1, m_2) = (1, 2)$, as argued in the following section. The details of all the numerical techniques that we use are explained in Appendix B.

A. Postselection of the PNR detectors

We provide numerical reasons for our choice of postselection of the PNR detectors. For the two-mode example we plot the fidelity and probability as a function of the postselected PNR measurement in Fig. 3(a) for a fixed value of $a = 0.3$. We find that while the fidelity is increasing with higher photon measurement, the probability drops rapidly. As a reasonable trade-off we pick $m = 2$, which gives an increase with respect to the fidelity with a slight cost to the probability of producing the state.

In Fig. 3(b) we compare the fidelity with respect to the target state parameter a for three settings, namely, (i) two-mode architecture with $m = 2$, (ii) three-mode architecture with $(m_1, m_2) = (1, 1)$, and (iii) three-mode architecture with $(m_1, m_2) = (1, 2)$. There are two important findings that we wish to highlight. The first is that we achieve near-perfect fidelity for the three-mode circuit for all values of $a \in [0.3, 1]$ when the PNR detectors are postselected to values $(m_1, m_2) = (1, 2)$. The second interesting numerical observation is that the performance of the two-mode architecture with $m = 2$ is extremely close to the three-mode circuit with $(m_1, m_2) = (1, 1)$. In both cases a common feature is that the sums of the values of postselected PNR detectors are the same. We anticipate that a deeper understanding is possible if we explore these aspects from the perspective of the resource theory of non-Gaussianity [61–65].

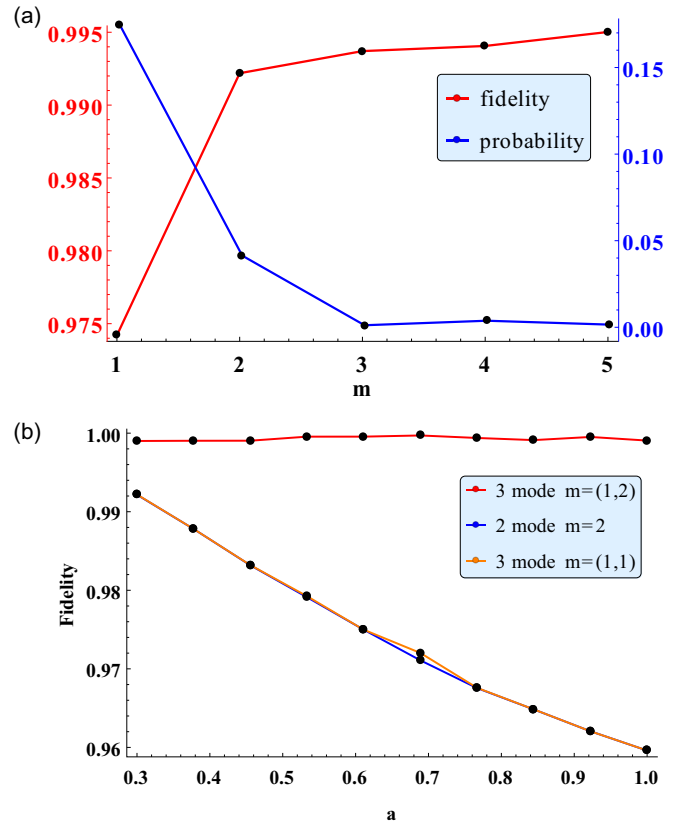


FIG. 3. (a) Fidelity (rising red line) of the trained state to the target state and probability (falling blue line) of preparing the trained state as a function of postselection of the PNR detectors in the two-mode circuit of Fig. 2. Here we use `basinhopping` with `niter = 20` (see Appendix B) without further optimizing the probability. We make the choice of $m = 2$, which has a reasonable trade-off between fidelity and probability, both of which we would require to be high. (b) Comparison of the optimal fidelities for $m = (1, 1)$ (bottom, upper orange line) and for $m = (1, 2)$ (top, red line) for the output state of the three-mode circuit, benchmarked with the two-mode circuit with measurement $m = 2$ (bottom, lower blue line), for various values of the target state parameter a . The two bottom curves corresponding to the two-mode case with $m = 2$ and the three-mode case with $m = (1, 1)$ differ only slightly and for $a = 0.77$. We find that the measurement setting of $m = (1, 1)$ has a very poor performance compared to the one with $m = (1, 2)$. For the three-mode case with $m = (1, 2)$ and $m = (1, 1)$, we use `niter = 40` and `niter = 80`, respectively.

B. Trained parameters for two-mode and three-mode circuits

The final optimization values for the two-mode circuit parameters and the three-mode circuit parameters are listed in Tables I and II, respectively. We find that the required squeezing values r are all less than or equal to 5.1 dB for the two-mode case and less than or equal to 6.7 dB for the three-mode case, which is a very experimentally accessible value. Further, unlike the two-mode case, it turns out that taking the displacements to be real provided better fidelity results for the three-mode case. We wish to highlight that the number of steps for which we run the optimization algorithms is fixed for all values of state parameter a for the sake of

TABLE I. Optimal circuit parameters for the two-mode architecture with PNR $m = 2$ [Fig. 2(a)]. The state parameter a of Eq. (1) is varied in equal steps in the range $[0.3, 1]$. Here $\{r_i, \phi_i^r\}$ ($\{d_i, \phi_i^d\}$) are the magnitude and phase for squeezing (displacement) applied to the i th vacuum mode; $\{\theta, \phi\}$ are the parameters for the beam splitter.

a	r_1	r_2	ϕ_1^r	ϕ_2^r	d_1	d_2	ϕ_1^d	ϕ_2^d	θ	ψ
0.3	0.37	-0.37	-3.08	0.59	-0.25	0.35	0.77	-0.98	-0.64	2.26
0.38	-0.43	-0.43	-1.72	-1.64	-0.36	-0.3	-2.05	-3.25	2.25	-1.93
0.46	-0.48	-0.48	-1.34	0.59	0.34	-0.39	-0.14	-0.93	2.49	2.85
0.53	-0.5	0.46	1.73	0.37	0.4	-0.29	-0.64	-1.2	0.97	-0.77
0.61	-0.54	-0.58	0.55	0.65	0.39	-0.41	0.75	2.28	-0.68	-1.26
0.67	-0.34	-0.48	3.1	0	-0.14	0.39	1.55	-1.57	0.61	-3.12
0.77	0.35	-0.51	0.85	0.01	0.14	-0.4	-1.14	1.57	0.63	2.72
0.84	0.28	-0.46	-0.57	0	0.07	0.38	4.42	-1.57	0.62	-2.86
0.92	-0.53	0.34	-1.84	0.02	-0.4	0.12	-2.49	7.9	-0.93	0.93
1	-0.23	0.43	-0.97	3.14	0.01	-0.38	-0.47	-1.57	3.76	-1.09

reproducibility of the numerical results. For certain values of a an improvement in probability cannot be ruled out.

C. Physical comparison of the two-mode and three-mode cases

The trained and target states are both pure and we have that the fidelity is equivalent to the Wigner overlap [66], i.e.,

$$F(\phi_1, \phi_2) = \langle \phi_1 | \phi_2 \rangle^2 = \int dx dp W(x, p; \phi_1) W(x, p; \phi_2).$$

Since the negativities of the Wigner function are crucial, we define by W_- the overlap of the negative region of the Wigner functions of the output and target state, each negative region being renormalized to 1. For the parameter $a \in [0.3, 1]$ we have that the gate strength $\gamma \in [0.0122, 0.0407]$ [by Eq. (3)] if the initial squeezing gate in Fig. 1 has $r = 1$.

We now use the optimization values of the two-mode and three-mode cases in Tables I and II to plot the fidelity, the overlap of the negative region of the Wigner function between the trained state $|\phi_L\rangle$ and the target state $|\phi\rangle$, and the probability of producing the trained state in Fig. 4. We find that using the three-mode architecture, the trained state is extremely close to the target state, and so is the overlap of the negative region of the Wigner function [Fig. 4(b)]. On the other hand, the two-mode circuit performs better than the three-mode

circuit in terms of the probability of producing the trained state, but there is a substantial drop in the corresponding W_- .

To provide a visual comparison of the two-mode and three-mode performances, we plot the Wigner functions for the trained states for specific values of the target state parameter a in Fig. 5. In each plot, the insets at the top right corner are contour plots of the Wigner function of the target state. In Fig. 5(a) we consider the three-mode gadget with $a = 0.3$. We find a very good match between the trained state and the target state. In Figs. 5(b) and 5(c) we consider $a = 0.61$, where Fig. 5(b) corresponds to the two-mode gadget and Fig. 5(c) to the three-mode gadget. We find an important difference that while the three-mode gadget produces a trained state that is close to the target state, the two-mode gadget produces a state with one of the Wigner negative dips missing. This leads to a reduced target state fidelity, as already mentioned in Fig. 4.

D. Generating random states

As a final numerical experiment we try to target a random state for each value of Fock state cutoff as outputs of the three-mode architecture with the same postselected PNR detectors $(m_1, m_2) = (1, 2)$. The target state is now of the form $|\phi_T\rangle = \sum_{n=0}^{n_c} c_n |n\rangle$, where n_c is the maximum Fock support or cutoff and the $\{c_n\}$ are randomly chosen coefficients that are normalized to unity. In Fig. 6 we plot the average fidelity of

TABLE II. Optimal circuit parameters for the three-mode architecture with PNR ($m_1 = 1, m_2 = 2$) in Fig. 2(b). Here $\{r_i, \phi_i^r\}$ are the magnitude and phase for squeezing applied to the i th vacuum mode; d_i are the real displacements applied to the i th vacuum mode. It turns out that taking the displacements to be real gives rise to more stable solutions. In addition, $\{\theta_i, \phi_i\}$ correspond to the parameters for the i th beam splitter.

a	r_1	r_2	r_3	ϕ_1^r	ϕ_2^r	ϕ_3^r	d_1	d_2	d_3	θ_1	θ_2	θ_3	ϕ_1	ϕ_2	ϕ_3
0.3	-0.27	0.65	0.66	-4.14	0.53	-1.94	0.38	-0.19	-0.47	2.29	-4.19	3.5	2.77	0.59	-0.2
0.38	-0.53	-0.75	-0.55	7.19	10.16	10.73	0.51	-0.03	0.53	2.3	-2.04	-4.11	-0.29	-1.99	-0.89
0.46	-0.75	-0.54	0.27	1.23	-5.31	1.53	-0.04	0.38	0.63	0.7	1.97	-0.88	-1.62	0.72	-1.79
0.53	0.71	0.67	-0.42	-2.07	0.06	-3.79	-0.02	0.34	0.02	-1.57	0.68	2.5	0.53	-4.51	0.72
0.61	0.72	0.65	-0.45	0.23	0.49	-3.81	0	-0.33	-0.01	-1.57	-2.46	0.63	-1.81	-0.22	6.42
0.67	0.52	0.56	0.74	0.09	3.35	-3.26	0.64	-0.33	0.02	-3.31	2.29	2.63	-3.33	1.41	-6.29
0.77	-0.55	0.73	0.03	1.14	0.98	-3.03	0.34	0.11	0.51	-2.3	-1.96	0.69	-3.22	4.32	-4.07
0.84	0.74	0.54	-0.48	0.5	-5.29	2.73	-0.01	-0.36	0.01	-1.52	0.7	-3.8	-0.95	-1.06	-1.24
0.92	-0.54	0.49	0.72	-0.44	2.36	3.14	0.15	-0.48	0.25	-1.3	2.27	0.67	-2.13	3.94	1.99
1	0.66	-0.38	-0.76	3.36	-0.73	0.4	-0.05	-0.82	0	1.56	-2.32	0.61	-3.67	-0.97	-0.71

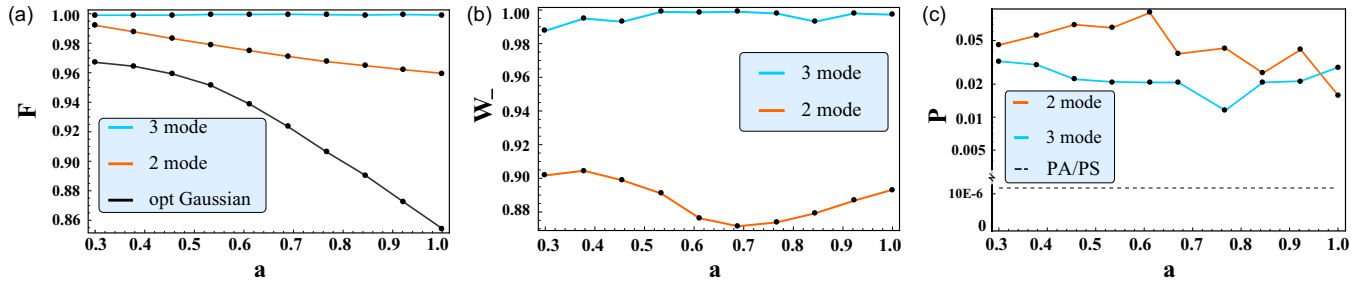


FIG. 4. (a) Optimal fidelity F of the machine learned state $|\phi_L\rangle$ in three-mode (top blue line) and two-mode (middle red line) architectures and the closest Gaussian state (bottom black line) with respect to the target state $|\phi\rangle$ in Eq. (1). (b) Overlap of the negative region of the Wigner function W_- between $|\phi_L\rangle$ and $|\phi\rangle$ for the three-mode (top blue line) and the two-mode (bottom orange line) cases. (c) Probability P of the optimal output state for the two-mode (top orange line) and three-mode (middle blue line) architectures compared with three consecutive photon-additions or subtractions (PA/PS) with a probability $\sim 10^{-6}$ (bottom dashed line). We see that for fidelity and Wigner negativity, the three-mode case performs best with near-perfect fidelity at the cost of a drop in probability when compared with the two-mode case.

the trained state to the target random state for various values of the cutoff dimension of the random state. As expected, the average fidelity is monotonically decreasing with an increase in the output cutoff dimension of the target state. However, for the cutoff Fock state value $n = 3$ we find that the average fidelity is extremely close to 1.

IV. QUANTUM RESOURCE FARMS TO IMPROVE SUCCESS PROBABILITY

The resource state preparation is probabilistic since it is conditioned on a particular postselected measurement outcome as shown in Fig. 2. To improve the success probability

of the resource state preparation, we propose an optical setup which we call a quantum resource farm. We line up identical copies of the state preparation gadget in parallel and connect all outputs to a single wire or sink as shown in Fig. 7. We then find a trade-off between the success probability and the number of gadgets in the resource farm, when allowing for a small error.

The model of our resource farm has n parallel quantum gadgets and let p be the probability of success of producing the resource state from each one of them. Let ϵ be the error which captures the degree of determinism of the entire farm. The probability of producing the required resource state from the farm is denoted by $P(F)$ and is given by the union of

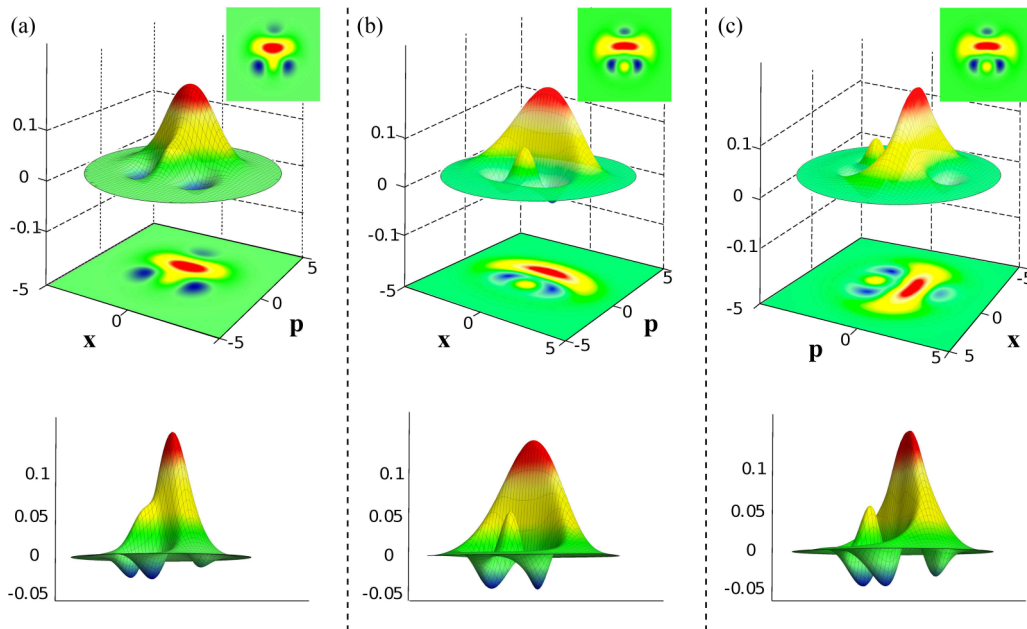


FIG. 5. Wigner function of the trained state for three cases (a) $a = 0.3$ with the three-mode architecture, (b) $a = 0.61$ with the two-mode circuit, and (c) $a = 0.61$ with the three-mode architecture. The insets are the Wigner contour plot for the target states for the corresponding values of a . We see that the trained state using the three-mode circuit is indistinguishable from the target state. We find that for the two-mode case (b) with $a = 0.61$, one Wigner negative region is missing with respect to the three-mode case (c), thereby leading to lesser fidelity to the target state.

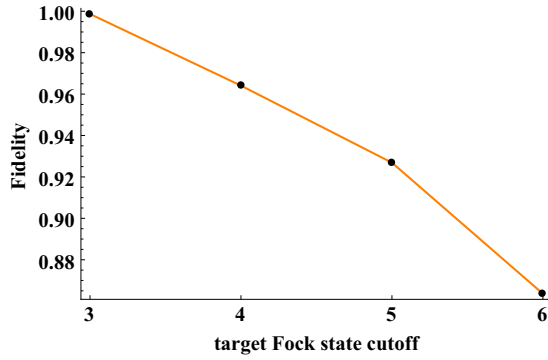


FIG. 6. Plot of the average fidelity of the trained state to a target random state $|\phi_T\rangle$ for various values of the Fock cutoff dimension n_c , using the three-mode gadget with $(m_1, m_2) = (1, 2)$. We sampled 100 random states for each value of target cutoff dimension. We find that using the three-mode gadget, any random superposition of up to three photon Fock states can be generated with near-perfect fidelity for this choice of postselected PNR detection pattern.

events of any of the constituent gadgets preparing this state. Even in the case where multiple resource states are simultaneously prepared, we continue to count it as a useful event. Thus, this success probability is identical to the complementary event probability that none of the gadgets produce the required resource state:

$$P(F) = 1 - (1 - p)^n. \quad (5)$$

Allowing an error ϵ for the failure of the state preparation from the farm, we set $P(F) \geq 1 - \epsilon$. Therefore, the minimum required number of gadgets n_{\min} is given by

$$n_{\min} = \left\lceil \frac{\log_2 \epsilon}{\log_2(1 - p)} \right\rceil, \quad (6)$$

where $\lceil y \rceil$ denotes the smallest integer greater than or equal to y . If we assume an ideal GKP teleportation using this resource state, then the effective cubic phase gate can be implemented at the same rate as the resource state preparation. If we use a smaller number of gadgets than n_{\min} , the rate at which the cubic phase gate can be implemented will be proportionally reduced.

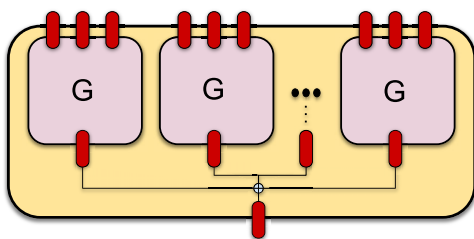


FIG. 7. Quantum resource farm that consists of copies of quantum gadgets G of Fig. 2 (three-mode case shown here) placed in a parallel manner where the individual outputs are collected to one output. There is a trade-off between the rate of production of the state and the number of gadgets used.

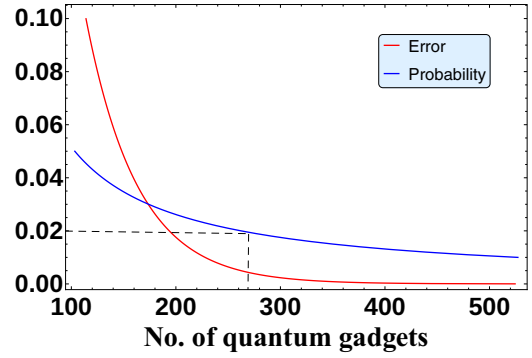


FIG. 8. Error (top red line) for a probability fixed at 0.02 and probability (bottom blue line) for error fixed at 0.005 as a function of the number of quantum gadgets in the resource farm. As an example, if we consider $p = 0.02$ and $\epsilon = 0.005$, i.e., the gadget produces the resource state with efficiency of 2% and if we want the farm to have a success rate of 99.5%, then we require ~ 260 gadgets (dotted lines) in the farm.

Examples

The top (red) line in Fig. 8 shows the variation of error ϵ as a function of the number of quantum gadgets n_{\min} for a fixed probability of $p = 0.02$ of state preparation for a gadget. The bottom (blue) line shows the variation of p vs n_{\min} for $\epsilon = 5 \times 10^{-3}$. For $p = 0.02$ (i.e., 2% efficiency) and $P(F) = 0.995$ ($\epsilon = 0.005$), $n_{\min} \sim 260$ (shown by dotted lines).

So we see that there is a trade-off between the number of gadgets used and the success probability as captured by Eq. (5). If one requires near determinism using this method to increase success probability, we see that the resource cost is challenging for the example mentioned in the preceding paragraph. However, the analysis provides us with a way to estimate the resources required to increase the success probability.

V. NOISE ANALYSIS IN A QUANTUM GADGET

In the ideal case the state in Eq. (1) is produced with near-perfect fidelity and a probability greater than 1%, as shown earlier in Fig. 4. However, in practical scenarios there are various sources of imperfections in the optical implementation. Of these, we focus on the state preparation stage where squeezed vacua are produced and in the measurement stage where a subset of modes is subjected to a photon-number-resolving detection. We model these imperfections by a photon-loss channel as described in the following section.

We now consider in Fig. 9 a noisy version of the three-mode gadget used to produce weak cubic phase states. The source and measurement losses are modeled using the pure-loss channel denoted by $L(\eta)$ with transmission coefficient $\eta \in [0, 1]$, a completely positive trace-preserving bosonic Gaussian channel with well-known Kraus operators [9,67]

$$A_k(\eta) = \left(\frac{1 - \eta}{\eta} \right)^{k/2} \frac{a^k}{\sqrt{n!}} (\sqrt{\eta})^{a^\dagger a}, \quad (7)$$

$$L(\eta)[\rho] = \sum_{k=0}^{\infty} A_k(\eta) \rho A_k(\eta)^\dagger.$$

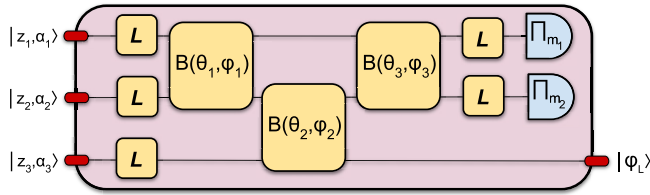


FIG. 9. Addition of source and detection losses to the three-mode gadget presented in Fig. 2. The loss is denoted by L and is modeled by a pure-loss bosonic channel $L(\eta)$ where $\eta \in [0, 1]$ is the transmission coefficient.

For simplicity, we set the detection efficiency equal to 96% and we take the noise at the source to be identical in the three modes. We vary this noise and consider the corresponding relation to properties of the output state.

In Fig. 10 we plot the fidelity, probability, and the minimum value of the Wigner function of the output state from the gadget. The fidelity is computed with respect to the target state (1) as a function of source loss where $\text{loss} = 1 - \eta_{\text{initial}}$, with η_{initial} the transmission at the source. For convenience, we set the loss at the detection part of the circuit equal to 4% and the target state parameter a to 0.3. The other parameters of the circuit are chosen to be the optimal values from the first row in Table II.

We find that there is a loss threshold ℓ_T beyond which all the Wigner negativity of the output state gets washed out. For the value of the target state parameter $a = 0.3$, we find that this threshold is at $\ell_T \sim 0.4$, i.e., at $\eta_{\text{initial}} = 0.6$. Any value of loss greater than 40% produces output Wigner functions that are positive everywhere. As expected, we find that both the fidelity and the probability of producing that state monotonically decrease up to the loss value ℓ_T . Beyond this value of loss the fidelity does increase a bit, but the dependence of the fidelity and probability is not of particular interest in this region [68] since the Wigner negativities have already been washed out. On the other hand, we also find that if the state preparation fidelity is required to be high (greater than 90%), then η_{initial} needs to be greater than 0.9, i.e., the source loss needs to be well below 0.1.

VI. CONCLUSION

We proposed a method for the production of target resource states using machine learning techniques. We first constructed quantum gadgets that produce very general Gaussian pure states in two and three modes which were conditioned on all but one of the modes on a postselected photon-number-resolving detector. We then tuned the parameters of the gadget to produce a trained state that is of almost perfect fidelity with the target state. Figure 6 indicates that this three-mode gadget and particular postselection could have a more universal property to produce a state with any choice of superposition in the first four Fock basis states.

We found that our architecture in Fig. 2 led to an increase in the probability (of the order of 10^4) of producing the target state compared to conventional sequential photon-addition or -subtraction methods. As a way to increase even this success probability, we proposed the notion of a quantum resource

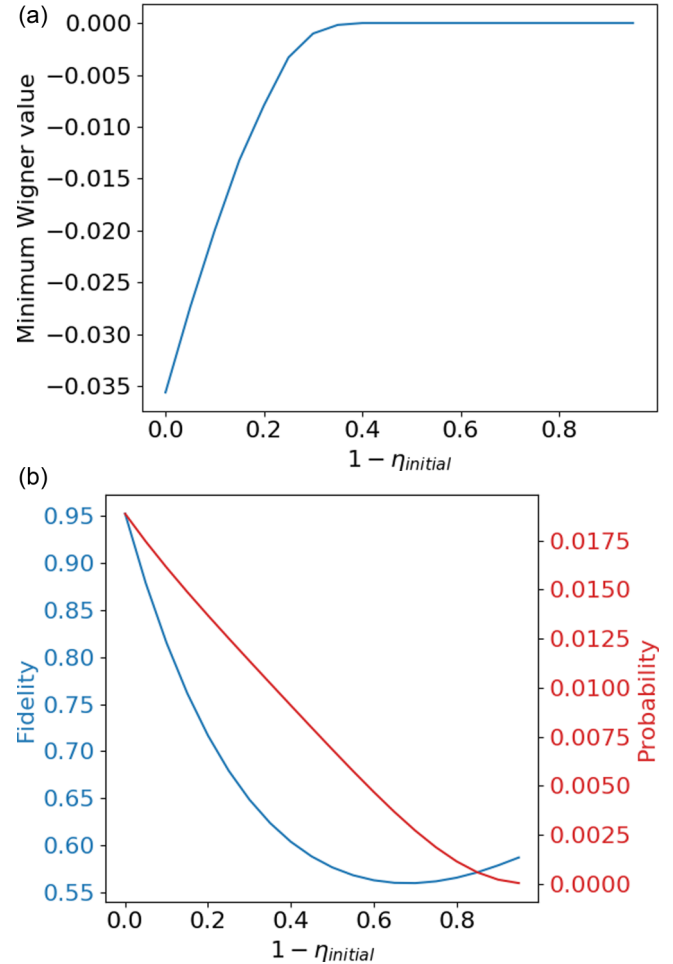


FIG. 10. Effect of noise on the preparation of target states in Eq. (1) using the optical circuit of Fig. 9. (a) Variation of the minimum value of the Wigner function of the output state as a function of the circuit loss $\ell = 1 - \eta_{\text{initial}}$. Here we set the measurement loss to a fixed value of 4% and vary only the preparation loss parameter. We find that for the output state parameter $a = 0.3$, the amount of loss the circuit can tolerate before all the Wigner negativities are washed out is approximately $\ell_T = 0.4$. (b) Variation of fidelity (bottom blue curve) of the output state with regard to the target state along with the probability (top red curve) of producing the state. The values of fidelity and probability monotonically decrease for increasing noise value up to the threshold value.

farm that connects such quantum gadgets in parallel. We then obtained the trade-off between the success probability and the number of required quantum gadgets. It turns out that the resource costs are high to achieve near determinism using this method as a means to circumvent quantum memories. Also, in this method, the rate of producing the resource states is limited by the measurement rate of the photon-number-resolving detectors.

We expect that our architecture for a quantum gadget can be realized in the near term and we anticipate substantial improvements in the future through technological advancements driven by demand. Also, the squeezing requirements for our state preparation (1) is within reasonable bounds of less than 7 dB. Having fewer gadgets in the resource farm would lead to

a lower production rate of these gates while not compromising on the fidelity. Also, photon loss plays an important role in the state preparation and needs to be accounted for as mentioned in Sec. V.

Our proposal is only an initial step in this direction and there is scope for improvements and optimization using tools such as machine learning. There are two ways to obtain higher gate strengths. One is to concatenate many weak cubic gates considered here and the other is to consider resource states that have a higher Fock support than those considered here. Our methods could prove useful in both these directions. Further, the final applied cubic phase gate using the resource state and gate-teleportation circuit tends to be noisy and requires additional considerations. Finally, the detailed numerical analysis also seems to suggest that the PNR detectors cannot be replaced by threshold detectors.

ACKNOWLEDGMENTS

We thank N. Killoran, K. Ramchandran, M. Schuld, K. Bradler, C. Myers, J. M. Arrazola, and B. Gupt for useful discussions. All numerical simulations were performed with open source software: The STRAWBERRYFIELDS quantum simulator [60] is available via [69]; Our PYTHON code for optimization built on STRAWBERRYFIELDS and SCIPY package [70] can be accessed via [71].

APPENDIX A: PHOTON SUBTRACTION

The standard photon-subtraction probability is obtained from a high-transmission beam splitter where $U(\theta) = \exp[\theta(\hat{a}^\dagger \hat{b} - \hat{a} \hat{b}^\dagger)]$, with $\theta \ll 1$. In this limit we can expand the unitary operator to the first order in gate strength to obtain $1 + \theta(\hat{a}^\dagger \hat{b} - \hat{a} \hat{b}^\dagger)$. For an arbitrary state and vacuum state incident on the beam splitter, we obtain $U(\theta)|\psi\rangle|0\rangle = N(|\psi\rangle|0\rangle - \theta\hat{a}|\psi\rangle|1\rangle)$, where $N = \sqrt{1 + \theta^2 \langle \hat{a}^\dagger \hat{a} \rangle_\psi}$. If we measure a single photon in the second mode, the probability of a photon subtraction on the input state is then given by $\theta^2 N^{-2}$. For beam splitters with around 98% transmission strength [45,50], this results in a probability $\sim 10^{-2}$. Resource states such as the one we consider require three successive photon subtractions, which results in a net probability of 10^{-6} . If we use this latter probability in Eq. (6) with $\epsilon \sim 10^{-3}$, we find that the required number of quantum gadgets is of the order of 10^5 , which is unfeasible.

APPENDIX B: NUMERICAL TECHNIQUES

We discuss two optimization functions that were used for training the circuit parameters. The landscape of our loss functions usually have several local minima, which makes it hard for standard local optimization methods because there is a very strong dependence on the initial conditions. The first global search function that we used is called `basinhopping`, which is described in Algorithm 2. `basinhopping` is a stochastic algorithm which attempts to find the global minimum of a smooth scalar function [72]. The implementation we used in our simulation is from the SCIPY package.

Algorithm 2 `basinhopping`

```

procedure basinhopping ( $x_0, f, args, niter, step\_size=1$ ):
   $x_{old} \leftarrow \text{local\_search}(\text{loss}, x_0)$ 
  for  $i$  in  $\text{range}(niter)$  do
     $\Delta \leftarrow \text{random}(0,1)$ 
     $x_{jump} \leftarrow x_{old} + step\_size \times \Delta$ 
     $x_{new} \leftarrow \text{local\_minimize}(x_{jump}, \text{loss}, args)$ 
    if  $\text{accept}(x_{new}, x_{old}) == \text{True}$  then
       $x_{old} \leftarrow x_{new}$ 
    end if
  end for
end procedure

```

The algorithm is iterative with each cycle composed of the following features.

- (1) Initialize the variables x_0 .
- (2) Perform `local_minimize` to minimize $f(x, args)$ starting from x_0 , to reach a local minimum we call x_{old} .
- (3) Randomly change the position of x_{old} with a tunable `step_size`.
- (4) Perform `local_minimize` starting from x_{old} to reach a local minimum we call x_{new} .
- (5) Perform an acceptance test $\text{accept}(x_{new}, x_{old})$: A simple rule could be that if $f(x_{new}, args) < f(x_{old}, args)$, $x_{old} = x_{new}$. However, the commonly used acceptance test is stochastic so as to maximize the likelihood of finding the global minimum. In the SCIPY implementation, the acceptance test used there is the Metropolis criterion of standard Monte Carlo algorithms [73], where the probability of acceptance is given by $\exp\{-[f(x_{old}, args) - f(x_{new}, args)]/T\}$. Here T is a fictitious temperate to control the degree of randomness.
- (6) Go back to step 2 and repeat this process `niter` times.

This global minimization method has been shown to be extremely efficient for a wide variety of problems in physics and chemistry. For a stochastic global heuristic there is no way to determine if the true global minimum has actually been found. In our simulation we set `niter` = 40, which is tuned to be able to have reproducible results. The `step_size` is set to be the default value from the SCIPY package. The algorithm for `local_minimize` can in principle be one of many options. In our simulation we choose sequential least-squares programming (SLSP) [74] which seems to be the fastest one for our task among the local search algorithms available in SCIPY. When there are no constraints on the variables, SLSP reduces to the well-known Newton method [74].

Remark. In principle, the `local_search` function mentioned in Algorithm 1 can be any local minimization algorithm and in our simulation we use the well-known BFGS algorithm, which is available from the SCIPY package. In principle, the `local_search` and the `local_minimize` algorithms can be the same; however, we use different names due to the role they play in our Algorithm 1.

Second optimization over probability

For our state preparation task, solely optimizing fidelity to the target state is insufficient. To have a better architecture the conditional probability of preparing the trained state needs to be sufficiently high as well. One possible strategy to achieve

both high fidelity and probability is to train the circuit to maximize fidelity and then use that point as a seed to further optimize the probability. This is exactly what we did for the three-mode case (see Algorithm 1), where we found that the second optimization over probability did little harm to the fidelity.

However, this is not the case for two-mode circuit where we found that second optimization quickly deteriorates the pretrained high fidelity. To tackle this problem, considering that the computation overhead for training two-mode circuit is moderate, we did a brute-force optimization over probability. That is, we repeat the `basinhopping` nbh times. We then pick out the global optimum, trained to optimize fidelity with the highest probability. In our simulation, we found that using $nbh = 20$ and $niter = 30$ is enough to obtain reproducible results. This procedure is summarized in Algorithm 3.

Algorithm 3 Optimization of probability for the two-mode architecture

```

procedure prob_opt ( $nbh$ ):
  initialize  $x\_list, prob\_list \leftarrow$  empty list
  for  $e$  in range( $nbh$ ) do
    initialize  $x_0$ 
     $x \leftarrow$  basinhopping( $x_0, loss, args, niter$ )
     $\_, prob, \_, \_ \leftarrow$  objective( $x$ )
     $x\_list.append(x)$ 
     $prob\_list.append(prob)$ 
  end for
  MaxIndex  $\leftarrow$  max( $prob\_ls$ )
  save  $x\_list[MaxIndex]$ 
end procedure

```

- [1] N. C. Menicucci, P. van Loock, M. Gu, C. Weedbrook, T. C. Ralph, and M. A. Nielsen, *Phys. Rev. Lett.* **97**, 110501 (2006).
- [2] S. Yokoyama, R. Ukai, S. C. Armstrong, J.-i. Yoshikawa, P. van Loock, and A. Furusawa, *Phys. Rev. A* **92**, 032304 (2015).
- [3] M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, and U. L. Andersen, [arXiv:1812.05358](https://arxiv.org/abs/1812.05358).
- [4] N. C. Menicucci, S. T. Flammia, and O. Pfister, *Phys. Rev. Lett.* **101**, 130501 (2008).
- [5] K. R. Motes, A. Gilchrist, J. P. Dowling, and P. P. Rohde, *Phys. Rev. Lett.* **113**, 120501 (2014).
- [6] H. Qi, L. G. Helt, D. Su, Z. Vernon, and K. Brádler, [arXiv:1812.07015](https://arxiv.org/abs/1812.07015).
- [7] Y. He, X. Ding, Z.-E. Su, H.-L. Huang, J. Qin, C. Wang, S. Unsleber, C. Chen, H. Wang, Y.-M. He *et al.*, *Phys. Rev. Lett.* **118**, 190501 (2017).
- [8] I. L. Chuang, D. W. Leung, and Y. Yamamoto, *Phys. Rev. A* **56**, 1114 (1997).
- [9] V. V. Albert, K. Noh, K. Duivendoorn, D. J. Young, R. T. Brierley, P. Reinhold, C. Vuillot, L. Li, C. Shen, S. M. Girvin, B. M. Terhal, and L. Jiang, *Phys. Rev. A* **97**, 032346 (2018).
- [10] M. H. Michael, M. Silveri, R. T. Brierley, V. V. Albert, J. Salmilehto, L. Jiang, and S. M. Girvin, *Phys. Rev. X* **6**, 031006 (2016).
- [11] M. Y. Niu, I. L. Chuang, and J. H. Shapiro, *Phys. Rev. A* **97**, 032323 (2018).
- [12] L. Li, C.-L. Zou, V. V. Albert, S. Muralidharan, S. M. Girvin, and L. Jiang, *Phys. Rev. Lett.* **119**, 030502 (2017).
- [13] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, *Nat. Commun.* **8**, 94 (2017).
- [14] R. Garcia-Patron and N. J. Cerf, *Phys. Rev. Lett.* **97**, 190503 (2006).
- [15] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, *Nat. Photon.* **7**, 378 (2013).
- [16] S. Pirandola, C. Ottaviani, G. Spedalieri, C. Weedbrook, S. L. Braunstein, S. Lloyd, T. Gehring, C. S. Jacobsen, and U. L. Andersen, *Nat. Photon.* **9**, 397 (2015).
- [17] V. Giovannetti, S. Lloyd, and L. Maccone, *Nat. Photon.* **5**, 222 (2011).
- [18] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, *Nat. Photon.* **12**, 724 (2018).
- [19] S. Lloyd and S. L. Braunstein, *Phys. Rev. Lett.* **82**, 1784 (1999).
- [20] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto, *Phys. Rev. Lett.* **88**, 097904 (2002).
- [21] D. Gottesman, A. Kitaev, and J. Preskill, *Phys. Rev. A* **64**, 012310 (2001).
- [22] K. K. Sabapathy and C. Weedbrook, *Phys. Rev. A* **97**, 062315 (2018).
- [23] F. Arzani, N. Treps, and G. Ferrini, *Phys. Rev. A* **95**, 052352 (2017).
- [24] K. Marshall, R. Pooser, G. Siopsis, and C. Weedbrook, *Phys. Rev. A* **91**, 032321 (2015).
- [25] P. Marek, R. Filip, H. Ogawa, A. Sakaguchi, S. Takeda, J.-i. Yoshikawa, and A. Furusawa, *Phys. Rev. A* **97**, 022329 (2018).
- [26] K. Miyata, H. Ogawa, P. Marek, R. Filip, H. Yonezawa, J.-i. Yoshikawa, and A. Furusawa, *Phys. Rev. A* **93**, 022301 (2016).
- [27] P. Marek, R. Filip, and A. Furusawa, *Phys. Rev. A* **84**, 053802 (2011).
- [28] M. Yukawa, K. Miyata, H. Yonezawa, P. Marek, R. Filip, and A. Furusawa, *Phys. Rev. A* **88**, 053816 (2013).
- [29] S. Ghose and B. C. Sanders, *J. Mod. Opt.* **54**, 855 (2007).
- [30] J. Honer, R. Löw, H. Weimer, T. Pfau, and H. P. Büchler, *Phys. Rev. Lett.* **107**, 093601 (2011).
- [31] P. Marek, J. Proszynski, and R. Filip, *Opt. Express* **26**, 29837 (2018).
- [32] S. Rosenblum, O. Bechler, I. Shomroni, Y. Lovsky, G. Guendelman, and B. Dayan, *Nat. Photon.* **10**, 19 (2016).
- [33] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, *Phys. Rev. Lett.* **116**, 090405 (2016).
- [34] H.-K. Lau, R. Pooser, G. Siopsis, and C. Weedbrook, *Phys. Rev. Lett.* **118**, 080501 (2017).
- [35] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, *Proc. Natl. Acad. Sci. USA* **115**, 1221 (2018).
- [36] J. Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, and X.-M. Jin, *Phys. Rev. Lett.* **120**, 240501 (2018).
- [37] L. Cincio, S. Yiğit, A. T. Sornborger, and P. J. Coles, *New J. Phys.* **20**, 113022 (2018).
- [38] J. M. Arrazola, T. R. Bromley, J. Izaac, C. R. Myers, K. Brádler, and N. Killoran, *Quantum Sci. Technol.* **4**, 024004 (2019).
- [39] Y. Miwa, J.-i. Yoshikawa, P. van Loock, and A. Furusawa, *Phys. Rev. A* **80**, 050303(R) (2009).

- [40] K. Miyata, H. Ogawa, P. Marek, R. Filip, H. Yonezawa, J.-i. Yoshikawa, and A. Furusawa, *Phys. Rev. A* **90**, 060302(R) (2014).
- [41] J.-i. Yoshikawa, Y. Miwa, A. Huck, U. L. Andersen, P. van Loock, and A. Furusawa, *Phys. Rev. Lett.* **101**, 250501 (2008).
- [42] R. Filip, P. Marek, and U. L. Andersen, *Phys. Rev. A* **71**, 042308 (2005).
- [43] R. Ukai, N. Iwata, Y. Shimokawa, S. C. Armstrong, A. Politi, J.-i. Yoshikawa, P. van Loock, and A. Furusawa, *Phys. Rev. Lett.* **106**, 240504 (2011).
- [44] R. Ukai, S. Yokoyama, J.-i. Yoshikawa, P. van Loock, and A. Furusawa, *Phys. Rev. Lett.* **107**, 250501 (2011).
- [45] M. Dakna, J. Clausen, L. Knöll, and D. G. Welsch, *Phys. Rev. A* **59**, 1658 (1999).
- [46] M. Dakna, J. Clausen, L. Knöll, and D.-G. Welsch, *Phys. Rev. A* **60**, 726 (1999).
- [47] M. Yukawa, K. Miyata, T. Mizuta, H. Yonezawa, P. Marek, R. Filip, and A. Furusawa, *Opt. Express* **21**, 5529 (2013).
- [48] J. Fiurášek, R. García-Patrón, and N. J. Cerf, *Phys. Rev. A* **72**, 033822 (2005).
- [49] F. Dell'Anno, S. De Siena, and F. Illuminati, *Phys. Rep.* **428**, 53 (2006).
- [50] T. Gerrits, S. Glancy, T. S. Clement, B. Calkins, A. E. Lita, A. J. Miller, A. L. Migdall, S. W. Nam, R. P. Mirin, and E. Knill, *Phys. Rev. A* **82**, 031802(R) (2010).
- [51] A. P. Lund, A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O'Brien, and T. C. Ralph, *Phys. Rev. Lett.* **113**, 100502 (2014).
- [52] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, *Phys. Rev. Lett.* **119**, 170501 (2017).
- [53] M. Ban, *J. Mod. Opt.* **43**, 1281 (1996).
- [54] M. Dakna, L. Knöll, and D.-G. Welsch, *Opt. Commun.* **145**, 309 (1998).
- [55] M. Dakna, L. Knöll, and D.-G. Welsch, *Eur. Phys. J. D* **3**, 295 (1998).
- [56] M. Dakna, T. Anhut, T. Opatrny, L. Knöll, and D.-G. Welsch, *Phys. Rev. A* **55**, 3184 (1997).
- [57] J. Clausen, M. Dakna, L. Knöll, and D. Welsch, *J. Opt. B* **1**, 332 (1999).
- [58] M. G. A. Paris, M. Cola, and R. Bonifacio, *Phys. Rev. A* **67**, 042104 (2003).
- [59] P. Knott, *New J. Phys.* **18**, 073033 (2016).
- [60] N. Killoran, J. Izaac, N. Quesada, V. Bergholm, M. Amy, and C. Weedbrook, *Quantum* **3**, 129 (2019).
- [61] K. K. Sabapathy and A. Winter, *Phys. Rev. A* **95**, 062309 (2017).
- [62] Q. Zhuang, P. W. Shor, and J. H. Shapiro, *Phys. Rev. A* **97**, 052317 (2018).
- [63] R. Takagi and Q. Zhuang, *Phys. Rev. A* **97**, 062337 (2018).
- [64] F. Albarelli, M. G. Genoni, M. G. A. Paris, and A. Ferraro, *Phys. Rev. A* **98**, 052350 (2018).
- [65] L. Lami, K. K. Sabapathy, and A. Winter, *New J. Phys.* **20**, 113012 (2018).
- [66] K. E. Cahill and R. J. Glauber, *Phys. Rev.* **177**, 1857 (1969).
- [67] J. S. Ivan, K. K. Sabapathy, and R. Simon, *Phys. Rev. A* **84**, 042311 (2011).
- [68] A. Mari and J. Eisert, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [69] <https://github.com/XanaduAI/strawberryfields>.
- [70] E. Jones, T. Oliphant, P. Peterson *et al.*, SciPy: Open source scientific tools for Python, (SciPy developers, 2001), available at <https://www.scipy.org/>.
- [71] <https://github.com/XanaduAI/constrained-quantum-learning>.
- [72] D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- [73] Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 6611 (1987).
- [74] J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects* (Springer Science + Business Media, Berlin, 2006).