

Error mitigation on a near-term quantum photonic device

Daiqin Su¹, Robert Israel¹, Kunal Sharma², Haoyu Qi¹, Ish Dhand¹, and Kamil Brádler¹

¹Xanadu, Toronto, Ontario, M5G 2C8, Canada

²Hearne Institute for Theoretical Physics and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA USA

Photon loss is destructive to the performance of quantum photonic devices and therefore suppressing the effects of photon loss is paramount to photonic quantum technologies. We present two schemes to mitigate the effects of photon loss for a Gaussian Boson Sampling device, in particular, to improve the estimation of the sampling probabilities. Instead of using error correction codes which are expensive in terms of their hardware resource overhead, our schemes require only a small amount of hardware modifications or even no modification. Our loss-suppression techniques rely either on collecting additional measurement data or on classical post-processing once the measurement data is obtained. We show that with a moderate cost of classical post processing, the effects of photon loss can be significantly suppressed for a certain amount of loss. The proposed schemes are thus a key enabler for applications of near-term photonic quantum devices.

1 Introduction

Error is the main hindrance for large scale quantum computation. Quantum error correction codes are introduced to correct the errors and allow for fault-tolerant quantum computation. However, the conditions for fault-tolerant quantum computation are extremely stringent, requiring very low-error gates and a large amount of physical qubits, e.g., about a thousand or more physical qubits are needed to construct a single fault-tolerant logical qubit [1]. Currently, the state of the art technologies allow for building quantum devices consisting of about fifty noisy physical qubits, which is very far away from the

Daiqin Su: sudaiqin@gmail.com

number required for fault-tolerant computation. It is expected that in the intermediate future quantum devices with hundreds of noisy physical qubits are available [2]. Efforts have been made to discover algorithms that are compatible with these noisy quantum devices and are capable of demonstrating the advantages of quantum computing. Examples include the Random Circuit Sampling [3, 4, 5], IQP [6, 7, 8], Boson Sampling [9], Gaussian Boson Sampling (GBS) [10, 11, 12], Variational Quantum Eigensolver [13], and the Quantum Approximate Optimization Algorithm [14, 15], etc.

To improve the performance of these near-term quantum devices, several error mitigation techniques have been developed to suppress the noise. One of the promising error mitigation techniques exploits extrapolation to approximately estimate the expectation values of some observables of a noise-free circuit by using the measured expectation values from noisy quantum circuits with various error rates [16, 17]. A proof of principle experiment has been performed in a superconducting device and the accuracy of the variational eigensolver has been significantly improved [18]. Another error mitigation technique that is capable of improving the estimation of expectation values is called quasi-probability decomposition technique, in which a noise-free circuit is simulated by a collection of randomly selected noisy circuits following a particular quasi-probability distribution [16, 19]. This technique has also been tested in various experimental platforms [20, 21]. Instead of estimating the expectation values, there are error mitigation techniques that can partially recover the error-free quantum states, though more resources are required. One of those examples is to measure the symmetries of the quantum systems and project it back into the error-free subspace [22, 23, 24, 25]. Another method is to employ variational quantum algorithms to mitigate errors in the state prepara-

arXiv:2008.06670v4 [quant-ph] 1 May 2021

tion circuits [26]. Moreover, inherent noise resilience of quantum algorithms can lead to error mitigation [27, 28, 29]. Other error mitigation techniques have also been developed to correct measurement errors [30, 31, 32, 33], to mitigate decoherence [34], and for analog quantum simulation [35]. Furthermore, learning-based algorithms have also been designed to mitigate errors [36, 37, 38].

All previously mentioned error mitigation techniques are specifically tailored for quantum systems with a finite Hilbert space dimension. It is an open question whether these methods are applicable for computing models based on infinite-dimensional systems. Here we focus on the GBS device as a promising near-term photonic infinite-dimensional platform. It is believed that GBS can be used to demonstrate computational advantages of quantum computer over classical computer. Various algorithms based on the GBS device have also been proposed, for example, sampling the dense subgraphs [39], distinguishing non-isomorphic graphs [40] and quantifying the similarity of graphs [41, 42]. These applications are sensitive to the imperfections of the GBS device, within which photon loss is the main source of errors for a photonic implementation. Schemes to mitigate the effect of photon loss are thus paramount for the applications of the GBS devices. However, no error mitigation scheme for photonic systems has been developed to date.

In this work, we propose two schemes to mitigate the effect of photon loss in a GBS device. The first scheme exploits the Richardson extrapolation technique and is tailored to extrapolate the probability of a photon number pattern or a collection of patterns for a loss-free GBS circuit. To perform the extrapolation, one has to vary the loss of the circuit and measure the probabilities of click patterns for every loss value. This requires only a small modification of the circuit but one needs to perform several experiments, depending on the number of chosen loss values. The second scheme estimates the probability of a click pattern of a loss-free circuit by linearly combining the probabilities of click patterns with higher total photon numbers in the presence of loss. This requires no modifications of the circuits since one does not need to vary the loss value as the first scheme. Classical post-processing is required to compute the linear combination coefficients.

This paper is organized as follows. In Sec. 2 we briefly introduce the GBS device and some of its applications. In Sec. 3 we discuss the error mitigation scheme based on the Richardson extrapolation technique. We detail a standard version and an improved version of the extrapolation technique, with the latter has a better performance for large photon loss. We discuss the second scheme, the loss cancellation method, in Sec. 4. In Sec. 5, we test our error mitigation schemes in an eight-mode circuit and compare their performances. Finally, we summarize our discussion in Sec. 6, followed by the Appendix.

2 Overview of GBS device

A GBS device consists of three parts: the input state, linear passive interferometer, and photon detectors. The input is usually chosen as a product state of pure single-mode Gaussian states, each of which is fully characterized by its displacement and squeezing. The linear passive interferometer implements a unitary transformation to the input state and produces a multimode Gaussian state in the output. In each output mode a photon detector counts the number of photons.

Consider a GBS device with M input/output modes. The output multimode Gaussian state is characterized by a $2M$ -component vector \mathbf{d} and a $2M \times 2M$ covariance matrix σ [43]. Assume that $\mathbf{n} = [n_1, n_2, \dots, n_M]$ represents a certain measurement pattern of photons (a click pattern), where n_j is the detected photon number in the j -th mode. In the case of no displacement, namely $\mathbf{d} = \mathbf{0}$, the probability of detecting a click pattern \mathbf{n} is given by [10]

$$P(\mathbf{n}) = \frac{1}{\mathbf{n}! \sqrt{\det \sigma_Q}} \prod_{k=1}^M \left(\frac{\partial^2}{\partial \alpha_k \partial \alpha_k^*} \right)^{n_k} \times \exp \left(\frac{1}{2} \boldsymbol{\alpha}_v^\top A \boldsymbol{\alpha}_v \right) \Big|_{\boldsymbol{\alpha}_v = \mathbf{0}}, \quad (1)$$

where $\mathbf{n}! = n_1! n_2! \dots n_M!$, $\sigma_Q = \sigma + \mathbb{I}_{2M}/2$, $\boldsymbol{\alpha}_v = (\alpha_1, \dots, \alpha_M, \alpha_1^*, \dots, \alpha_M^*)^\top$, and the matrix A is given by

$$A = X_{2M}(\mathbb{I}_{2M} - \sigma_Q^{-1}), \quad \text{with } X_{2M} = \begin{pmatrix} 0 & \mathbb{I}_M \\ \mathbb{I}_M & 0 \end{pmatrix}.$$

The expression for probability in Eq. (1) is valid for both pure and mixed Gaussian states.

It is believed that a GBS device with a sufficiently large number of modes and input photons can generate a photon number probability distribution that is hard to sample from using a classical computer [10]. The GBS device can also be used to solve graph-related problems by mapping the adjacency matrix of a graph to the covariance matrix of the GBS output state [44]. The properties of the graph are thus encoded into the photon number probability distribution. Such applications include sampling the dense subgraphs [39], distinguishing non-isomorphic graph [40, 42], quantifying similarity of two graphs [41], and other potential graph-related applications to be discovered.

A realistic GBS device is noisy, suffering from experimental imperfections like the photon loss, thus limiting its computational power [45]. Given that fault tolerant error correction will not be accessible in the near future, it is paramount to develop some error mitigation techniques with least amount of hardware modifications to suppress the effects of photon loss and improve the performance of the GBS device. In the following sections, we focus on algorithms that involve estimating sampling probabilities of a click pattern or a collection of click patterns, and propose two schemes to improve the estimation of the probability for a loss-free GBS device.

3 Suppress photon loss via extrapolation

The first method of suppressing errors in GBS is inspired from the extrapolation procedure introduced in Ref. [16]. Similar to Ref. [16], our method relies on performing multiple measurements and extrapolating to obtain the desired outcome. However, while Ref. [16] focuses on decreasing an overall multiplicative factor in the system Hamiltonian thus leading to an effectively higher error rate, we exploit the possibility of directly changing the error rate in photonic systems. Another novelty of our work is an improved method that provides enhanced performance for GBS probabilities by removing the poles of the extrapolation function before extrapolating as described in Sec. 3.4.

3.1 General formalism

Consider a quantum circuit with an ideal input state and gates, the density operator of the output state before detection is $\hat{\rho}_0$. The expectation value of an observable \hat{O} is $O_0 = \text{tr}(\hat{O}\hat{\rho}_0)$, where "tr" represents the trace. If the quantum circuit is imperfect, for example, the gate has an error rate characterized by a small parameter ϵ , then the output density operator becomes $\hat{\rho}_\epsilon$ and the expectation value becomes $O(\epsilon) = \text{tr}(\hat{O}\hat{\rho}_\epsilon)$. Both $\hat{\rho}_\epsilon$ and $O(\epsilon)$ deviate from $\hat{\rho}_0$ and O_0 , respectively, but approach to them in the limit of $\epsilon \rightarrow 0$. We thus can perform a series expansion of $\hat{\rho}_\epsilon$ as

$$\hat{\rho}_\epsilon = \hat{\rho}_0 + \sum_{k=1}^{\infty} \hat{\rho}_k \epsilon^k, \quad (2)$$

and similarly for $O(\epsilon)$,

$$O(\epsilon) = O_0 + \sum_{k=1}^{\infty} O_k \epsilon^k, \quad (3)$$

where $\hat{\rho}_k$ are operators and O_k are real numbers.

Suppose one can vary the error rate in a controllable way, in particular, to increase the error rate on purpose. Denote the error rates as $\epsilon_j = \epsilon c_j$, with $j = 0, 1, 2, \dots, m$, and we choose $c_0 = 1$ and $c_j > 1$ for $j \neq 0$. For each error rate ϵ_j , one performs the experiment and measures the expectation value $O(\epsilon_j)$. By linearly combining $(m+1)$ measured expectation values, one arrives at an estimation of the expectation value O_0 as

$$\tilde{O}(\epsilon) = \sum_{j=0}^m \gamma_j O(\epsilon_j) = \sum_{j=0}^m \gamma_j O(\epsilon c_j). \quad (4)$$

The coefficients γ_j are appropriately chosen such that O_k for $k = 1, 2, \dots, m$ are cancelled, giving

$$\tilde{O}(\epsilon) = O_0 + \mathcal{O}(\epsilon^{m+1}), \quad (5)$$

a better estimation of the loss-free expectation value when ϵ is small. Taking this requirement into account, we find that γ_j satisfy a linear system of equations

$$\sum_{j=0}^m \gamma_j = 1, \quad \sum_{j=0}^m \gamma_j c_j^k = 0, \quad k = 1, 2, \dots, m, \quad (6)$$

and the solution can be found as

$$\gamma_j = (-1)^m \prod_{k \neq j} \frac{c_k}{c_j - c_k}. \quad (7)$$

This extrapolation technique has been applied to qubit systems.

3.2 Extrapolate sampling probability

We now apply the above technique to suppress the photon loss of a GBS device and obtain a better estimation of the sampling probability. To illustrate the method, we first consider the uniform loss case, namely, the photon loss in each mode is the same and is characterized by a single parameter ϵ . This is a good approximation to a realistic linear interferometer if it is implemented using the Clements' decomposition [46] (see Appendix A for more details).

Assume that the covariance matrix of the output state without photon loss is σ_0 and that with photon loss is σ_ϵ . They are related via the action of a lossy bosonic channel,

$$\sigma_\epsilon = (1 - \epsilon)\sigma_0 + \frac{1}{2}\epsilon\mathbb{I}_{2M}, \quad (8)$$

where \mathbb{I}_{2M} is the identity matrix. We thus have

$$\begin{aligned} \sigma_Q(\epsilon) &= \sigma_\epsilon + \frac{1}{2}\mathbb{I}_{2M} \\ &= \sigma_0 + \frac{1}{2}\mathbb{I}_{2M} - \frac{1}{2}\epsilon(2\sigma_0 - \mathbb{I}_{2M}), \end{aligned} \quad (9)$$

and

$$\begin{aligned} [\sigma_Q(\epsilon)]^{-1} &= 2(2\sigma_0 + \mathbb{I}_{2M})^{-1} \left[\mathbb{I}_{2M} + \sum_{k=1}^{\infty} (2\sigma_0 - \mathbb{I}_{2M})^k (2\sigma_0 + \mathbb{I}_{2M})^{-k} \epsilon^k \right]. \end{aligned}$$

By substituting this into Eq. (2), we obtain a series expansion of the A matrix as

$$A(\epsilon) = A_0 + \sum_{k=1}^{\infty} A_k \epsilon^k, \quad (10)$$

where A_0 is the matrix corresponding to no photon loss state and the coefficients A_k is given by

$$A_k = -2X_{2M}(2\sigma_0 - \mathbb{I}_{2M})^k (2\sigma_0 + \mathbb{I}_{2M})^{-k-1}. \quad (11)$$

The probability, $P(\mathbf{n}; \epsilon)$, of a click pattern \mathbf{n} with photon loss ϵ can be obtained by replacing the A matrix and σ_Q in Eq. (1) by $A(\epsilon)$ and $\sigma_Q(\epsilon)$, respectively. Since both $A(\epsilon)$ and $\sigma_Q(\epsilon)$ have series expansions with respect to ϵ , we thus can find a series expansion for $P(\mathbf{n}; \epsilon)$ as

$$P(\mathbf{n}; \epsilon) = P_0(\mathbf{n}) + \sum_{k=1}^{\infty} P_k(\mathbf{n})\epsilon^k, \quad (12)$$

where $P_0(\mathbf{n})$ corresponds to the sampling probability without photon loss and is the quantity

that we want to estimate, and $P_k(\mathbf{n})$ is a complicated expression and its explicit form is not relevant here.

Now we apply the general formalism developed in Sec. 3.1 to derive a better estimation of $P_0(\mathbf{n})$. Assume that one can choose different loss values, ϵ_j , and estimate the corresponding probability, $P(\mathbf{n}; \epsilon_j)$, in the experiment. Similarly, define $\epsilon_j = \epsilon c_j$, with $j = 0, 1, 2, \dots, m$, and we choose $c_0 = 1$ and $c_j > 1$ for $j \neq 0$. Then we get a better estimation of the probability $P_0(\mathbf{n})$ by defining

$$\tilde{P}(\mathbf{n}, \epsilon) = \sum_{j=0}^m \gamma_j P(\mathbf{n}; \epsilon_j) = \sum_{j=0}^m \gamma_j P(\mathbf{n}; \epsilon c_j), \quad (13)$$

which further simplifies to

$$\tilde{P}(\mathbf{n}; \epsilon) = P_0(\mathbf{n}) + \mathcal{O}(\epsilon^{m+1}), \quad (14)$$

for γ_j as in Eq. (7) and for small values of ϵ . Different values of loss are indeed possible in the experiment: specifically, loss can be programmably increased in experiment using M additional tunable beam splitters that have one output port discarded. Experimentally, a tunable beam splitter is implemented by a Mach-Zehnder interferometer consisting of two static 50:50 beam splitters and two phase shifters [47] (see also Appendix A for more details). Its transmission coefficient can be tuned by varying the phases, e.g., via changing the temperature of the device [48]. Suppose the transmission coefficient of each additional beam splitter is η_{ad} , then the overall transmission of the circuit is modified to be $(1 - \epsilon)\eta_{\text{ad}}$. This implies the loss value is changed to $\epsilon' = 1 - (1 - \epsilon)\eta_{\text{ad}} = \epsilon[\eta_{\text{ad}} + (1 - \eta_{\text{ad}})/\epsilon]$. One can appropriately choose the value of η_{ad} to attain a target loss value ϵ_j . A potential challenge is that the detection probability of the needed click patterns decreases and thus it requires a lot of data collection.

3.3 A two-mode squeezed vacuum example

To showcase the extrapolation technique, we consider a simple example: to mitigate photon loss in a two-mode squeezed vacuum (TMSV) state. A TMSV state is defined as

$$|\chi\rangle_{\text{TMSV}} = \sqrt{1 - \chi^2} \sum_{n=0}^{\infty} \chi^n |n\rangle |n\rangle, \quad (15)$$

where $\chi = \tanh r$ and r is the squeezing parameter [49]. When the TMSV state is detected by

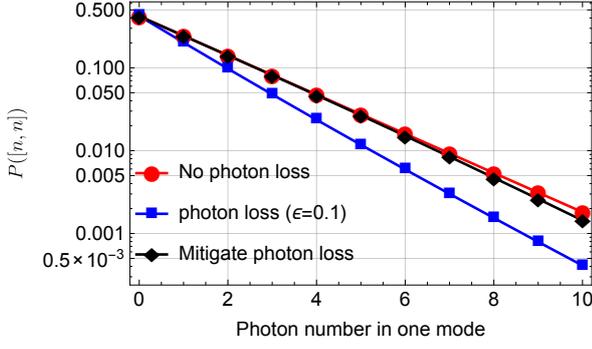


Figure 1: Error mitigation to a lossy TMSV state with squeezing parameter $r = 1.0$. The red circles represent probabilities for a pure TMSV state, the blue squares represent probabilities for a lossy TMSV state and the black rhombuses represent the extrapolated probabilities with photon loss $\epsilon = 0.1$.

two photon-number-resolving (PNR) detectors, the only possible detected photon number patterns are $\mathbf{n} = [n, n]$ and the probability is

$$P_0([n, n]) = (1 - \chi^2)\chi^{2n} = \frac{(\tanh r)^{2n}}{\cosh^2 r}. \quad (16)$$

The measurement probability $P_0(n, n)$ for $r = 1.0$ is plotted in Fig. 1 (red circle). Now a pure lossy channel with transmissivity $\eta = 1 - \epsilon$ is added to each output mode of the TMSV state, resulting in a mixed two-mode Gaussian state. The photon number distribution will be modified, which we denote as $P(n, n; \epsilon)$. We then plot $P(n, n; \epsilon)$ for $r = 1.0$ and $\epsilon = 0.1$ in Fig. 1 (blue square), which has a big deviation from the no photon loss case.

Now we apply the extrapolation technique to mitigate the photon loss. We choose $m = 4$ and $\mathbf{c} = (1.0, 1.2, 1.4, 1.6, 1.8)$. From Eq. (7), we find $\boldsymbol{\gamma} = (126, -420, 540, -315, 70)$. The new approximation to the ideal case can be evaluated directly from Eq. (13). We plot $\tilde{P}(n, n; \epsilon)$ for $r = 1.0$ and $\epsilon = 0.1$ in Fig. 1 (black rhombus). We see that $\tilde{P}(n, n; \epsilon)$ is much closer to $P_0(n, n)$ as compared to $P(n, n; \epsilon)$, showing that the extrapolation technique significantly mitigate the effect of photon loss. After the error mitigation, the effective photon loss is approximately $\epsilon' \approx 0.01$.

3.4 Improved extrapolation of sampling probability

We have shown that the extrapolation technique works very well for low photon loss. However, when the amount of photon loss and input squeezing increase, the extrapolated result becomes less

and less accurate. We find that by slightly modifying the previous extrapolation procedure, one can obtain a better estimation of the probability for large squeezing and relatively high photon loss.

By using the expression of $\sigma_Q(\epsilon)$ in Eq. (9), one can rewrite the A matrix with photon loss as

$$A(\epsilon) = (1 - \epsilon)X_{2M}(2\sigma_0 - \mathbb{I}_{2M}) \times [(2\sigma_0 + \mathbb{I}_{2M}) - (2\sigma_0 - \mathbb{I}_{2M})\epsilon]^{-1}. \quad (17)$$

There exists a unitary matrix U diagonalizes the the covariance matrix σ_0 as

$$\sigma_0 = \frac{1}{2}U \bigoplus_{k=1}^M \begin{pmatrix} e^{2r_k} & 0 \\ 0 & e^{-2r_k} \end{pmatrix} U^\dagger, \quad (18)$$

where we assume σ_0 is a pure-state covariance matrix and r_k are the squeezing parameters of the input squeezed vacuum states. Here U represents the transformation of the linear interferometer, and is independent of the input squeezing and the photon loss. Then $A(\epsilon)$ can be rewritten as

$$\begin{aligned} A(\epsilon) &= (1 - \epsilon)X_{2M}U \bigoplus_{k=1}^M \begin{pmatrix} \frac{\tanh r_k}{1 - \epsilon \tanh r_k} & 0 \\ 0 & -\frac{\tanh r_k}{1 + \epsilon \tanh r_k} \end{pmatrix} U^\dagger \\ &\equiv \frac{1 - \epsilon}{\mathcal{P}(\epsilon, \tilde{\mathbf{r}})} R(\epsilon, \mathbf{r}), \end{aligned} \quad (19)$$

where $\mathbf{r} = (r_1, r_2, \dots, r_M)^\top$ and $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{N_\lambda})^\top$, with $\{\tilde{r}_j\}_{j \in \{1, \dots, N_\lambda\}}$ denotes a set of nonzero different squeezing parameters in the set $\{r_k\}_{k \in \{1, \dots, M\}}$, and we defined

$$\begin{aligned} \mathcal{P}(\epsilon, \tilde{\mathbf{r}}) &= \prod_{j=1}^{N_\lambda} (1 - \epsilon^2 \tanh^2 \tilde{r}_j), \\ R(\epsilon, \mathbf{r}) &= \mathcal{P}(\epsilon, \tilde{\mathbf{r}})X_{2M} \\ &\quad \times U \bigoplus_{k=1}^M \begin{pmatrix} \frac{\tanh r_k}{1 - \epsilon \tanh r_k} & 0 \\ 0 & -\frac{\tanh r_k}{1 + \epsilon \tanh r_k} \end{pmatrix} U^\dagger. \end{aligned}$$

It is evident that $\mathcal{P}(\epsilon, \tilde{\mathbf{r}})$ is a polynomial of order $2N_\lambda$, and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}})/(1 \pm \epsilon \tanh r_k)$ is a polynomial of order $2N_\lambda - 1$ if $r_k \neq 0$ and of order $2N_\lambda$ if $r_k = 0$. Since X_{2M} is a constant matrix and S (represents the linear interferometer) is independent of photon loss ϵ and input squeezing r_j , so every entry of the matrix $R(\epsilon, \mathbf{r})$ is a polynomial of ϵ of order at most $2N_\lambda$ (when at least one of the r_j is zero), or $2N_\lambda - 1$ (when all r_j are not zero).

From Eq. (10) we find

$$\begin{aligned} \det\{[\sigma_Q(\epsilon)]^{-1/2}\} &= \prod_{k=1}^M \frac{1}{\cosh r_k} \frac{1}{\sqrt{1 - \epsilon^2 \tanh^2 r_k}} \\ &= \left(\prod_{k=1}^M \frac{1}{\cosh r_k} \right) \frac{1}{\mathcal{Q}(\epsilon, \mathbf{r})}, \end{aligned} \quad (20)$$

where $\mathcal{Q}(\epsilon, \mathbf{r}) \equiv \prod_{k=1}^M \sqrt{1 - \epsilon^2 \tanh^2 r_k}$. By substituting $A(\epsilon)$ and $\sigma_Q(\epsilon)$ into Eq. (1), we find the probability of measuring a click pattern \mathbf{n} in the presence of photon loss can be written as

$$P(\mathbf{n}; \epsilon) = \frac{1}{\mathbf{n}!} \left(\prod_{k=1}^M \frac{1}{\cosh r_k} \right) \frac{(1 - \epsilon)^N}{\mathcal{Q}(\epsilon, \mathbf{r}) \mathcal{P}^N(\epsilon, \tilde{\mathbf{r}})} \mathbb{P}(\epsilon), \quad (21)$$

where $N = \sum_{j=1}^M n_j$ is the total detected photon number, $\mathbb{P}(\epsilon)$ is a polynomial of ϵ of order at most $2NN_\lambda$ (when at least one of the r_j is zero), or $N(2N_\lambda - 1)$ (when all r_j are not zero).

Notice that $\mathcal{Q}(\epsilon, \mathbf{r}) \rightarrow 0$ and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}}) \rightarrow 0$ when $\epsilon \rightarrow 1$ and $r_k \rightarrow \infty$, namely, both $\mathcal{Q}(\epsilon, \mathbf{r})$ and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}})$ are close to zero in the large squeezing and photon loss regime. Therefore, $\mathcal{Q}(\epsilon, \mathbf{r})$ and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}})$ contribute to the ‘‘singular’’ part of the probability expression $P(\mathbf{n}; \epsilon)$, and the corresponding poles are given by $\epsilon = \coth r_k > 1$. The presence of these poles limits the accuracy of the first extrapolation procedure. To obtain a better extrapolation accuracy, we can remove these poles by multiplying $P(\mathbf{n}; \epsilon)$ with $\mathcal{Q}(\epsilon, \mathbf{r}) \mathcal{P}^N(\epsilon, \tilde{\mathbf{r}})$, and then perform the linear combination. That is to say, the new estimation for the probability without photon loss is

$$\bar{P}(\mathbf{n}, \epsilon) = \sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon) \mathcal{Q}(c_j \epsilon, \mathbf{r}) \mathcal{P}^N(c_j \epsilon, \tilde{\mathbf{r}}). \quad (22)$$

Notice that $\mathcal{Q}(\epsilon, \mathbf{r}) \rightarrow 1$ and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}}) \rightarrow 1$ when $\epsilon \rightarrow 0$, so the estimation $\bar{P}(\mathbf{n}, \epsilon)$ approaches to $P_0(\mathbf{n})$ in the limit of $\epsilon \rightarrow 0$.

To showcase the performance of the improved extrapolation technique, we consider mitigating the photon loss of a TMSV state as in Sec. 3.3. Table 1 compares the results of extrapolation and improved extrapolation for low photon loss $\epsilon = 0.2$ and relatively high photon loss $\epsilon = 0.5$ cases. We see that for low photon loss, they both give very good approximations to the exact sampling probabilities. While for high photon loss,

the improved extrapolation gives better results for click patterns with low total photon number. In particular, the probabilities for click patterns $[0, 0]$ and $[1, 1]$ can always be exactly extrapolated in the improved extrapolation. This can be understood as follows. After removing the poles, the right hand side of Eq. (21) is simply a polynomial, so in principle the probability $P_0(\mathbf{n})$ can be extrapolated exactly if sufficient loss values are chosen, namely, m is at least the same as the order of the polynomial. For the TMSV state example, the two input squeezing parameters are the same, so $N_\lambda = 1$. By taking into account the factor $(1 - \epsilon)^N$, the right hand side of Eq. (21) is a polynomial with order $2N$ after removing the poles. Therefore, by taking $m = 4$, one can exactly extrapolate probabilities of click patterns with total photon number less than three.

From the experiment perspective, removing the poles is only possible when one can fully control the photon loss and know the input squeezing parameters accurately. This requires a good calibration of the GBS device in a priori.

3.5 Extrapolation precision analysis

We have showed that the extrapolation technique works quite well by simply choosing several loss values, e.g., $m = 4$. Better results can be obtained by increasing the number of loss values, namely, to increase the number of experiments. For improved extrapolation technique, one can in principle extrapolate the exact value of the sampling probability by increasing the number of loss values. However, we show that this is challenging in practice. Specifically, the required measurement accuracy should increase exponentially in order to get a good extrapolated probability when the number of loss values increases, resulting in exponential increase of running time for the experiment.

To estimate some quantities, like the sampling probability, there is always an uncertainty due to a limited number of samples, or due to the experimental imperfections. We now consider how the statistic uncertainty and experimental imperfections affect the extrapolation results. From Eq. (13) we can see that the uncertainty in $P(\mathbf{n}; c_j \epsilon)$ results in uncertainty in $\bar{P}(\mathbf{n}; \epsilon)$. Assume that $\hat{P}(\mathbf{n}; c_j \epsilon)$ is an estimator of the sampling probability of click pattern \mathbf{n} with photon loss $c_j \epsilon$ and $P(\mathbf{n}; c_j \epsilon)$ is considered as its mean

Table 1: Comparing the performance of the improved extrapolation and the normal extrapolation technique for a TMSV state. The first column is the click pattern and the second column gives the probability for a pure TMSV state. The third and fourth main columns compare results for loss $\epsilon = 0.2$ and $\epsilon = 0.5$, respectively. The symbol "extrap." stands for extrapolation, and "imp. extrap." stands for improved extrapolation.

$[n, n]$	exact	$\epsilon = 0.2$		$\epsilon = 0.5$	
		extrap.	imp. extrap.	extrap.	imp. extrap.
[0, 0]	0.4200	0.4202	0.4200	0.8406	0.4200
[1, 1]	0.2436	0.2429	0.2436	0.3125	0.2436
[2, 2]	0.1413	0.1387	0.1400	0.1597	0.1140
[3, 3]	0.0820	0.0770	0.0781	0.0308	0.0701
[4, 4]	0.0475	0.0415	0.0421	0.0128	0.0317
[5, 5]	0.0276	0.0218	0.0222	0.0102	0.0115
[6, 6]	0.0160	0.0114	0.0116	0.0068	0.0037

value, then

$$\hat{P}(\mathbf{n}; c_j \epsilon) = P(\mathbf{n}; c_j \epsilon)(1 + \hat{X}_j), \quad (23)$$

where \hat{X}_j is a random variable with zero mean and its variance characterizes the relative uncertainty of the measured probability. Here we assume that the precision analysis is performed for multiplicative precision, where the variance of \hat{X}_j does not depend on the pattern \mathbf{n} of photons. A more useful analysis would model the actual error in an experiment and would need to account for inaccuracies in the modelling of the experiment. Such an analysis is left for future work.

From Eq. (23), the relative fluctuation of the extrapolated probability $\tilde{P}(\mathbf{n}; \epsilon)$ is

$$\begin{aligned} \hat{Y}(\mathbf{n}; \epsilon) &\equiv \frac{\hat{P}(\mathbf{n}; \epsilon) - \tilde{P}(\mathbf{n}; \epsilon)}{\tilde{P}(\mathbf{n}; \epsilon)} \\ &= \frac{\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon) \hat{X}_j}{\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon)}, \end{aligned} \quad (24)$$

where $\hat{P}(\mathbf{n}; \epsilon)$ is obtained by replacing $P(\mathbf{n}; c_j \epsilon)$ in Eq. (13) by $\hat{P}(\mathbf{n}; c_j \epsilon)$. It is straightforward to show that the variance of $\hat{Y}(\mathbf{n}; \epsilon)$ is given by

$$\text{Var}(\hat{Y}) = \frac{\sum_{j=0}^m \gamma_j^2 P^2(\mathbf{n}; c_j \epsilon) V_j}{[\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon)]^2}, \quad (25)$$

where V_j is the variance of \hat{X}_j and we have assumed that \hat{X}_j are independent random variables. Denote the minimum variance of $\{\hat{X}_j\}_{j=0}^m$ as V_{\min} and the minimum nonzero $P(\mathbf{n}; c_j \epsilon)$ as P_{\min} ; and the corresponding maximums as V_{\max} and P_{\max} . Then we can derive a lower bound and an upper

bound for $\text{Var}(\hat{Y})$,

$$\begin{aligned} \text{Var}(\hat{Y}) &\geq \frac{P_{\min}^2 V_{\min}}{[\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon)]^2} \sum_{j=0}^m \gamma_j^2 \\ &= C_1 \Gamma_2 V_{\min}, \\ \text{Var}(\hat{Y}) &\leq \frac{P_{\max}^2 V_{\max}}{[\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon)]^2} \sum_{j=0}^m \gamma_j^2 \\ &= C_2 \Gamma_2 V_{\max}, \end{aligned} \quad (26)$$

where C_1 and C_2 are approximately constants because $\sum_{j=0}^m \gamma_j P(\mathbf{n}; c_j \epsilon)$ approaches $P_0(n)$, and $P_{\min}(P_{\max})$ becomes constant for large values of m . Moreover, $\Gamma_2 = \sum_{j=0}^m \gamma_j^2$ increases exponentially as the number of loss values, m , increases, which follows from Eq. (7). Therefore, $\text{Var}(\hat{Y})$ increases exponentially if V_{\min} is fixed, which implies that it becomes exponentially hard to estimate the loss-free sampling probability.

In practice, the strategy is to choose a number of loss values such that a sufficient good approximation to $P_0(\mathbf{n})$ is obtained while it is still tractable to measure the probabilities for various photon loss. As an example, we discuss the effect of uncertainty of estimating the probability of a certain click pattern. Estimating a probability can be achieved by sampling a process many times and counting the number of success event. Assume that N_{succ} and N are the number of success tries and the total number of tries, respectively. The probability estimated by $\hat{p} = N_{\text{succ}}/N$ is a random variable and its variance is given by

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{N}, \quad (27)$$

where p is the mean value of \hat{p} . The variance of

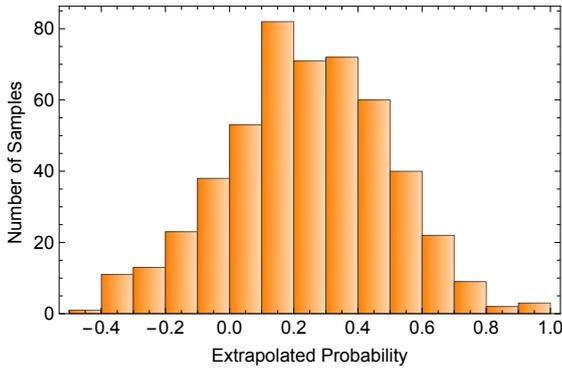


Figure 2: Distribution of extrapolated probability for click pattern $[1, 1]$ due to the uncertainty of probability measurement. Here 500 samples of size $N = 10^5$ are collected. The mean and standard deviation of the extrapolated probability are 0.2373 and 0.2531, respectively.

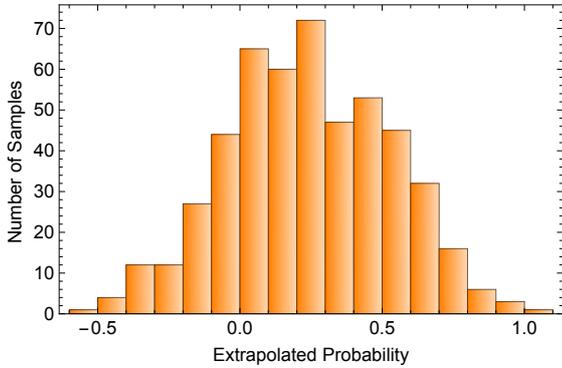


Figure 3: Distribution of extrapolated probability for click pattern $[1, 1]$ due to the uncertainty of loss values. The fluctuation of the loss value is assumed to follow a normal distribution with standard deviation 0.01. The total number of samples is 500, and the mean and standard deviation of the extrapolated probability is 0.2425 and 0.2873, respectively.

the relative error of \hat{p} is

$$\text{Var}\left(\frac{\hat{p}}{p}\right) = \frac{1-p}{Np}. \quad (28)$$

If this variance is treated as V_{\max} , then it is straightforward to find the required total number of tries,

$$N = \frac{1-p}{p} \left[\text{Var}\left(\frac{\hat{p}}{p}\right) \right]^{-1} \leq \frac{1-p}{p} \frac{C_2 \Gamma_2}{\text{Var}(\hat{Y})}. \quad (29)$$

This gives an estimation of the required total number of tries in order to achieve a certain accuracy of the extrapolated probability, and also the required running time to collect data if the sampling rate is known.

For the TMSV state example, we consider the effect of probability measurement uncertainty

and the fluctuation of loss values, and the results are shown in Figs. 2 and 3, respectively. The probability we want to extrapolate is $P_0([1, 1]) = 0.2436$. We choose the photon loss as $\epsilon = 0.3$ and $\mathbf{c} = (1.0, 1.3, 1.6, 1.9, 2.2)$, and obtain the extrapolated result $\tilde{P}([1, 1]) = 0.2367$ assuming no fluctuations. In the first case, the relative probability measurement uncertainty, as defined in Eq. (23), comes from the finite number of collecting tries. Here we collect 500 samples of size $N = 10^5$. We can see from Fig. 2 that the extrapolated probability deviates from $\tilde{P}([1, 1])$ and the variance is so big such that one sometimes gets negative extrapolated probabilities. However, the peak of the distribution is still around 0.2 and the mean value can be calculated to be 0.2373, which is close to the $\tilde{P}([1, 1])$, and the standard deviation is 0.2531. The effect of the loss fluctuation is similar, see Fig. 3. We assume the loss fluctuation follows a normal distribution with standard deviation 0.01. The extrapolated probability also follows a Gaussian-like distribution with big variance, and one sometimes gets negative values. However, the peak of the distribution is still around 0.2 and the mean value can be calculated to be 0.2112, which is close to the $\tilde{P}([1, 1])$, and the standard deviation is 0.2873. This example shows that the extrapolation amplifies the variance, as indicated by Eq. (26). However, it is still practical to get good results by choosing an appropriate number of loss values.

3.6 Nonzero displacement and nonuniform loss

We have shown that the extrapolation technique works for a GBS device with squeezed vacuum states as inputs. More general input states consist of displacements, for example in the algorithm to simulate the molecular vibronic spectra in a GBS device [50]. Here we show that in the presence of displacements the probability can also be expanded as a series with respect to the loss ϵ , therefore the extrapolation technique still works. We further show that the ‘singular’ part (the poles) in the probability expression is independent of the displacements, therefore the improved extrapolation technique also applies. For details see Appendix B.

We have studied the simplified case where the photon loss in each mode are the same, which can be considered as a good approximation if the linear interferometer is implemented using the

Clements' decomposition [46]. However, in the realistic implementation the photon loss is not uniform, and the situation becomes even worse if the interferometer is implemented using the Reck's decomposition [47]. We now propose an extrapolation technique for a GBS device with nonuniform loss.

A universal M -mode linear interferometer consists of $N = M(M-1)/2$ beam splitters. Assume that each beam splitter is lossy and is characterized by two lossy channels. We order the N beam splitters and label them using integers k , and the loss parameters of the corresponding two lossy channels as ϵ_{ka} and ϵ_{kb} . The probability of detecting a click pattern \mathbf{n} is a function of all loss parameters, and is denoted as $P(\mathbf{n}; \boldsymbol{\epsilon})$ with $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_a, \boldsymbol{\epsilon}_b)$ and $\boldsymbol{\epsilon}_a = (\epsilon_{1a}, \epsilon_{2a}, \dots, \epsilon_{Na})$, $\boldsymbol{\epsilon}_b = (\epsilon_{1b}, \epsilon_{2b}, \dots, \epsilon_{Nb})$. Define $P^{(ia)}(\mathbf{n})$ ($P^{(ib)}(\mathbf{n})$) as the probability by changing only one loss parameter from ϵ_{ia} (ϵ_{ib}) to $c_{ia}\epsilon_{ia}$ ($c_{ib}\epsilon_{ib}$), with c_{ia} and c_{ib} greater than one. We then obtain an approximation of $P_0(\mathbf{n})$ to the second order of the photon loss,

$$\tilde{P}(\mathbf{n}; \boldsymbol{\epsilon}) = P(\mathbf{n}; \boldsymbol{\epsilon}) - \sum_{\mu} \sum_{k=1}^N \frac{P^{(k\mu)}(\mathbf{n}) - P(\mathbf{n}; \boldsymbol{\epsilon})}{c_{k\mu} - 1}, \quad (30)$$

where $\mu = \{a, b\}$.

In the experiment, one first measures $P(\mathbf{n}; \boldsymbol{\epsilon})$ with loss $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\epsilon}_b$; then change one of the loss parameters while keeping other parameters unchanged to measure $P^{(ia)}(\mathbf{n})$ and $P^{(ib)}(\mathbf{n})$, which requires $M(M-1)$ repeats of experiments. By combining all these measurement results we get a better approximation of the probability $P_0(\mathbf{n})$ via Eq. (30). This method is advantageous for a large circuit when the photon loss of each beam splitter is small. To precisely control the loss of each beam splitter could be challenging for an interferometer integrated on a chip, but is realizable for architectures based on delay loops.

4 Loss Cancellation

We now discuss another scheme to mitigate the effect of photon loss in a GBS device. It is specifically tailored for photon number detection and is particularly suitable for platforms where photon number detection is available. One of the advantages of this scheme is that it requires no

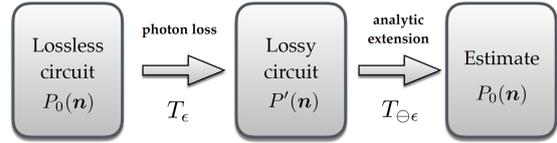


Figure 4: Procedure to cancel photon loss. In the first step, the probabilities of a lossy device are calculated using the probabilities of a loss-free device. This corresponds to a physical process and is accomplished by applying the operator T_ϵ . In the second step, the probabilities of a loss-free device are inferred using the probabilities of a lossy device. This is accomplished by applying the operator $T_{\ominus\epsilon}$ but does not correspond to a physical process.

hardware modifications of the GBS device. The only computational cost is to calculate a set of coefficients, which we will discuss in details.

4.1 The general procedure

Consider a noisy M -mode device with same photon loss in each mode, characterized by a single parameter ϵ . Equivalently, this noisy GBS device can be modelled as placing M beam splitters, with the same transmissivity $\eta = 1 - \epsilon$, after a lossless GBS device. The beam splitters take a Fock state \mathbf{n} (corresponding to an M -tuple of nonnegative integers n_j) to $\mathbf{n}' \leq \mathbf{n}$ ($n'_j \leq n_j$ for all j) with conditional probabilities

$$P_\epsilon(\mathbf{n}'|\mathbf{n}) = \prod_{j=1}^M \binom{n_j}{n'_j} \epsilon^{n_j - n'_j} (1 - \epsilon)^{n'_j}. \quad (31)$$

Then given probabilities for the states in the lossless GBS device, we get probabilities for the states in a lossy GBS device. This can be expressed in terms of a linear operator T_ϵ on measures on \mathbb{N}^M . Define $P_0(\mathbf{n})$ as the probabilities for Fock states \mathbf{n} in the lossless GBS device, then the probabilities in the lossy GBS device are

$$P'(\mathbf{n}'; \epsilon) = T_\epsilon(P_0)(\mathbf{n}') = \sum_{\mathbf{n} \geq \mathbf{n}'} P_\epsilon(\mathbf{n}'|\mathbf{n}) P_0(\mathbf{n}). \quad (32)$$

This involves an infinite series, but convergence is clear (for $0 < \epsilon < 1$) since $\sum_{\mathbf{n}} P_0(\mathbf{n}) = 1$ and $0 \leq P_\epsilon(\mathbf{n}'|\mathbf{n}) \leq 1$. To compute $T_\epsilon(P_0)(\mathbf{n}')$ for a given final Fock state \mathbf{n}' , we do not need to compute conditional probabilities for all \mathbf{n} , but only for those that could lead to \mathbf{n}' , i.e., with

$\mathbf{n} \geq \mathbf{n}'$. If P_0 has finite support, we only need to include click patterns up to the largest $|\mathbf{n}|$ (the total number of photons of click pattern \mathbf{n}) for which $P_0(\mathbf{n}) \neq 0$. Even if the support is infinite, in a practical computation we might impose a finite cutoff.

What we have just described, evaluating the probabilities of a lossy GBS device from the probabilities of a loss-free GBS device, is straightforward. However, what we want is the inverse: using the probabilities of a lossy GBS device to infer the probabilities of a loss-free GBS device. This can be accomplished by performing an analytic continuation to the transformation (32).

We first show that the transformation T_ϵ form a semigroup. Suppose we apply two sets of beam splitters to a loss-free device, with photon losses ϵ and μ , respectively. The two sets of beam splitters can be combined into one set of beam splitters with transmissivity $(1 - \epsilon)(1 - \mu)$, namely, with photon losses $\epsilon + \mu - \epsilon\mu$. This implies that the operators T_ϵ form a semigroup, with $T_\epsilon T_\mu = T_{\epsilon \oplus \mu}$, where $\epsilon \oplus \mu = \epsilon + \mu - \epsilon\mu$. Now the operators T_ϵ can be defined formally for arbitrary complex numbers ϵ , not just those in $[0, 1]$: at least if P is a complex measure of finite support, $T_\epsilon(P)$ will be a complex measure with finite support, and $\sum_{\mathbf{n}} T_\epsilon(P)(\mathbf{n}) = 1$, and by analytic continuation the formula $T_\epsilon T_\mu = T_{\epsilon \oplus \mu}$ is still true. The formulas are all the same, although the interpretation of Eq. (31) as a conditional probability is no longer there.

We can use this to go backwards, inferring P_0 from $P' = T_\epsilon(P_0)$. By setting $\epsilon \oplus \mu = 0$, we find $\mu = \epsilon/(\epsilon - 1)$, which we denote as $\ominus\epsilon$. Thus $T_{\ominus\epsilon}T_\epsilon$ is formally equal to the identity map, so that applying $T_{\ominus\epsilon}$ formally brings the probabilities $P' = T_\epsilon(P_0)$ back to the loss-free probabilities P_0 . This whole procedure is schematically shown in Fig. 4. In the calculation, one only needs to substitute $\ominus\epsilon = \epsilon/(\epsilon - 1)$ for ϵ :

$$\begin{aligned} P_0(\mathbf{m}) &= T_{\ominus\epsilon}(P')(\mathbf{m}) = \sum_{\mathbf{n} \geq \mathbf{m}} P_{\ominus\epsilon}(\mathbf{m} | \mathbf{n}) P'(\mathbf{n}) \\ &= \sum_{\mathbf{n} \geq \mathbf{m}} \left[\prod_{j=1}^M \binom{n_j}{m_j} \left(-\frac{1}{\epsilon}\right)^{m_j} \left(\frac{\epsilon}{\epsilon - 1}\right)^{n_j} \right] P'(\mathbf{n}). \end{aligned} \quad (33)$$

We call this procedure loss cancellation.

Our formal result Eq. (33) involves an infinite series. The convergence of this infinite series is

not guaranteed in the whole parameter regime. Of course in practice we can only consider finitely many terms, so we might impose a cutoff, but unless the series converges, the result of this computation might not be a good approximation to the loss-free probabilities. See Appendix C for a proof that under appropriate conditions, for sufficiently low photon loss the series converges and gives good approximation to the loss-free probabilities. In an actual experiment, the overall photon loss ϵ has to be determined first. This can be done by injecting coherent lights into the circuit and measuring the output coherent lights [51]. After the overall photon loss ϵ being determined, one runs the experiment many times and collects enough data to estimate the probability $P'(\mathbf{n})$. Finally, the probability without photon loss $P_0(\mathbf{m})$ is calculated using Eq. (33).

4.2 Test for a two-mode squeezed vacuum

The two-mode squeezed vacuum state is given by Eq. (15), from which it is clear that the probabilities for click patterns $[i, i]$ are given by

$$P_0([i, i]) = \chi^{2i}(1 - \chi^2), \quad (34)$$

and zero for other patterns. With uniform loss ϵ we have $P' = T_\epsilon(P_0)$, where for $i \leq j$,

$$\begin{aligned} &P'([i, j]; \epsilon) \\ &= (1 - \chi^2) \sum_{k=j}^{\infty} \binom{k}{i} \binom{k}{j} \chi^{2k} \epsilon^{2k-i-j} (1 - \epsilon)^{i+j} \\ &= \binom{j}{i} \chi^{2j} \epsilon^{j-i} (1 - \chi^2) (1 - \epsilon)^{i+j} \\ &\quad \times {}_2F_1(j + 1, j + 1; j - i + 1; (\epsilon\chi)^2), \end{aligned}$$

and by symmetry we have $P'([j, i]; \epsilon) = P'([i, j]; \epsilon)$. Since $0 < \epsilon < 1$ and $0 < \chi < 1$, the series converges and can be written using a hypergeometric function.

We now apply the loss cancellation procedure to estimate the probability $P_0([1, 1])$. Two cases are considered: in one case we choose the cutoff photon number as $n_{\max} = 7$, which means only click patterns $[n_1, n_2]$ with $n_1 \geq 1, n_2 \geq 1$ and $n_1 + n_2 \leq 7$ are considered, and the resulting approximation is denoted as $\tilde{P}_7([1, 1])$; while in the other case we choose the cutoff photon number as $n_{\max} = 10$ and the resulting approximation is denoted as $\tilde{P}_{10}([1, 1])$. We find that the Maclaurin series of $\tilde{P}_7([1, 1])$ in ϵ begins

$$\chi^2(1 - \chi^2) - 16\chi^8(1 - \chi^2)\epsilon^6 + O(\epsilon^8), \quad (35)$$

and that for $\tilde{P}_{10}([1, 1])$ begins

$$\chi^2(1 - \chi^2) + 324(\chi^{12} - \chi^{14})\epsilon^{10} + O(\epsilon^{12}), \quad (36)$$

where $\chi^2(1 - \chi^2)$ is the value of the actual $P_0([1, 1])$. Because the next nonzero coefficients are ϵ^6 (for $n_{\max} = 7$) and ϵ^{10} (for $n_{\max} = 10$) respectively, this should give very good results when ϵ is small.

Table 2 shows the estimated result for $P_0([1, 1])$ using loss cancellation for different input squeezing and cutoff photon number. For $r = 0.5$ and $n_{\max} = 7$, the loss cancellation gives very good approximations for photon loss up to $\epsilon = 0.5$, and the estimation fails for high photon loss like $\epsilon > 0.6$. Increasing the cutoff photon number to $n_{\max} = 10$ slightly improves the result and pushes the boundary to about $\epsilon = 0.6$, but still fails for photon loss $\epsilon > 0.7$. For $r = 1.0$ and $n_{\max} = 7$, good approximation is achieved for photon loss $\epsilon \leq 0.3$ and the result becomes meaningless for $\epsilon \geq 0.5$. We discuss this result in detail in Appendix C. For $r = 0.5$, our series should converge for all $\epsilon < 1$, so for any ϵ we should be able to attain good results by taking a sufficiently large cutoff. However, for $r = 1.0$ our convergence result only works for $\epsilon < 1/(2\chi) \approx 0.6565$, and for ϵ greater than that an increased cutoff would be useless (see Appendix C for more details).

4.3 Using empirical data

In a typical application, we will use an empirical distribution \hat{P} from measured data as an approximation to P' , and compute our approximation to $P_0(\mathbf{n})$ as $T_\mu(\hat{P})(\mathbf{n})$. Suppose the empirical data is obtained from a sample of size N , the estimator $\hat{P}(\mathbf{n})$ has mean $P'(\mathbf{n})$, variance $P'(\mathbf{n})[1 - P'(\mathbf{n})]/N$, and covariances

$$\text{Cov}(\hat{P}(\mathbf{n}), \hat{P}(\mathbf{n}')) = -P'(\mathbf{n})P'(\mathbf{n}')/N \quad (37)$$

for $\mathbf{n} \neq \mathbf{n}'$. If we write

$$T_\mu(\hat{P})(\mathbf{n}') = \sum_{\mathbf{n}} P_\mu(\mathbf{n}'|\mathbf{n})\hat{P}(\mathbf{n}),$$

then this has mean $T_\mu(P')(\mathbf{n}') = P_0(\mathbf{n}')$ and variance

$$\begin{aligned} & \frac{\sum_{\mathbf{n}} P_\mu(\mathbf{n}'|\mathbf{n})^2 P'(\mathbf{n})}{N} - \frac{[\sum_{\mathbf{n}} P_\mu(\mathbf{n}'|\mathbf{n})P'(\mathbf{n})]^2}{N} \\ &= \frac{\sum_{\mathbf{n}} P_\mu(\mathbf{n}'|\mathbf{n})^2 P'(\mathbf{n})}{N} - \frac{P_0(\mathbf{n}')^2}{N}. \end{aligned} \quad (38)$$

It should be noted that even when the infinite series for $T_\mu(P')(\mathbf{n})$ converges, the variance might not. Practically speaking, this means that "outliers" could have a large influence on the variance. Rather than use all the click patterns that appear in our sample, it may be better to impose a fixed cutoff. This means our estimator will no longer be unbiased, but it may be more stable.

For several different values of ϵ with $\chi = \tanh(1/2)$, we took 100 samples of size 10^5 from the distribution P_0 , added loss ϵ by letting each photon survive or disappears with probabilities $1 - \epsilon$ and ϵ , and then took the estimator $T_{\ominus\epsilon}(\hat{P})([1, 1])$. Recall that the correct value is 0.167948. The results are shown in Table 3.

As a function of μ , $T_\mu(P)(\mathbf{n})$ is an analytic function which we can expand in a power series around $\mu = 0$, the coefficients involving probabilities of various click patterns. The terms involving higher powers of μ will contain click patterns with more extra photons, which will have low probability of being observed but may have large coefficients. In a simulation, a few of these "outlier" click patterns will often be observed, and this can have a bad effect on the accuracy of our estimate. It can be better to only use a limited number of coefficients. The estimator will no longer be unbiased, but the variance may decrease significantly. To illustrate this we consider the example of a TMSV state as before. To fourth order of μ , the series for $T_\mu(P)([1, 1])$ reads

$$\begin{aligned} & T_\mu(P)([1, 1]) \\ &= P([1, 1]) + \{-2P([1, 1]) + 2P([2, 1]) + 2P([1, 2])\}\mu + \{P([1, 1]) - 4P([2, 1]) - 4P([1, 2]) \\ & \quad + 3P([3, 1]) + 4P([2, 2]) + 3P([1, 3])\}\mu^2 + \{2P([2, 1]) + 2P([1, 2]) - 6P([3, 1]) - 8P([2, 2]) \\ & \quad - 6P([1, 3]) + 4P([4, 1]) + 6P([3, 2]) + 6P([2, 3]) + 4P([1, 4])\}\mu^3 + \{3P([3, 1]) + 4P([2, 2]) \\ & \quad + 3P([1, 3]) - 8P([4, 1]) - 12P([3, 2]) - 12P([2, 3]) - 8P([1, 4]) + 5P([5, 1]) + 8P([4, 2]) \\ & \quad + 9P([3, 3]) + 8P([2, 4]) + 5P([1, 5])\}\mu^4. \end{aligned} \quad (39)$$

Table 2: Approximate the probability $P_0([1, 1])$ using the loss cancellation procedure for cutoff photon number $n_{\max} = 7$ and $n_{\max} = 10$. The first column lists the loss values and the first row gives the lossless probability $P_0([1, 1])$. The second column gives the estimated probability $\tilde{P}_7([1, 1])$ for $n_{\max} = 7$, with two subcolumns corresponding to input squeezing $r = 1/2$ and $r = 1$, respectively. The third column gives the estimated probability $\tilde{P}_{10}([1, 1])$ for $n_{\max} = 10$.

ϵ	$n_{\max} = 7$		$n_{\max} = 10$	
	$r = 1/2$	$r = 1$	$r = 1/2$	$r = 1$
0.0	0.167948	0.243596	0.167948	0.243596
0.1	0.167948	0.243595	0.167948	0.243596
0.2	0.167946	0.243502	0.167948	0.243597
0.3	0.167914	0.241527	0.167948	0.243697
0.4	0.167678	0.218252	0.167953	0.247736
0.5	0.166384	0.008163	0.168027	0.351743
0.6	0.160535	-1.698578	0.168753	2.555229
0.7	0.137057	-15.634539	0.174541	47.943868
0.8	0.049440	-142.109725	0.215083	1100.091815

Table 3: Approximation of the probability $P_0([1, 1])$ using empirical data. The first column lists the loss values. The second column ("No.1") gives results of the approximation to $P_0([1, 1])$ using the normal loss cancellation, with two subcolumns corresponding to the mean and standard deviation, respectively. The third column ("No.2") gives results using the series expansion introduced in Eq. (39). The fourth column ("No.3") gives results by taking into account the form of the probability expression in Eq. (40). We choose the input squeezing as $r = 1/2$ and collect 100 samples of size 10^5 .

ϵ	No.1		No.2		No.3	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
0.2	0.167905	0.001430	0.167823	0.001562	0.168155	0.001387
0.5	0.166660	0.006316	0.167998	0.004721	0.167166	0.005806
0.6	0.158845	0.021262	0.170362	0.011843	0.167394	0.011189
0.7	0.224839	0.065665	0.189705	0.034023	0.164793	0.034124
0.8	-0.015223	0.199697	0.249027	0.116481	0.166464	0.085374

Now by replacing μ and P in Eq. (39) by $\ominus\epsilon$ and \hat{P} , respectively, we obtain an estimator $T_{\ominus\epsilon}(\hat{P})([1, 1])$ for $P_0([1, 1])$. With $r = 1/2$ and 100 samples of size 10^5 , we get results with a significant improvement, see Table 3.

Further improvements are possible if we take advantage of knowledge of the form of $T_\nu(P_0)$. In the two-mode example we know (see Sec. 3.4) that $T_\nu(P_0)([1, 1])$ should have the form

$$T_\nu(P_0)([1, 1]) = \frac{A(\nu)}{(1 - \nu^2\chi^2)^3}, \quad (40)$$

where $A(\nu)$ is a polynomial, and in fact we know $A(\nu)$ has degree ≤ 4 . Note that $\nu = \epsilon \oplus \mu = \epsilon + \mu - \epsilon\mu$ so $\mu = (\nu - \epsilon)/(1 - \epsilon)$.

$$\begin{aligned} A(\nu) &= (1 - \nu^2\chi^2)^3 T_\nu(P_0)([1, 1]) \\ &= (1 - \nu^2\chi^2)^3 T_\mu(P)([1, 1]), \end{aligned} \quad (41)$$

where $P = T_\epsilon(P_0)$. If we expand the right hand side in a power series in $\nu - \epsilon$, since the left hand side is a polynomial of degree ≤ 4 the terms in higher powers on the right should be 0. We can take that series to order 4, evaluate at $\nu = 0$ using the empirical distribution \hat{P} instead of P , and the result should be a good approximation of $A(0) = P_0([1, 1])$. We tried this for $r = 1/2$ with 100 samples of size 10^5 , with quite good results, see Table 3. For $\epsilon = 0.9$, the mean and standard deviation are 0.159427 and 0.368437, respectively, which are not wildly off the mark.

5 Eight-mode example

We have discussed two schemes to mitigate the effect of photon loss and showcased their performance for a two-mode squeezed state. An impor-

Table 4: Mitigate photon loss for the orbit probability $P_0([00001111])$. The first column lists the loss values and the second column gives probabilities without doing any error mitigation. The third and fourth columns show probabilities after performing extrapolation and the improved extrapolation, respectively. The last column lists probabilities with the loss cancellation scheme.

Photon loss	No mitigation	Extrapolation	Improved extrapolation	Loss cancellation
$\epsilon = 0.0$	0.058419	0.058419	0.058419	0.058419
$\epsilon = 0.1$	0.040659	0.058371	0.058406	0.058582
$\epsilon = 0.2$	0.030128	0.057676	0.058019	0.058756
$\epsilon = 0.3$	0.023141	0.055638	0.055843	0.056712
$\epsilon = 0.4$	0.017815	0.053008	0.050619	0.040986
$\epsilon = 0.5$	0.013216	0.052349	0.045251	-0.042469
$\epsilon = 0.6$	0.008965	0.039900	0.050386	-0.446576
$\epsilon = 0.7$	0.005086	-0.189112	0.081490	-2.413260

tant question is whether these error mitigation techniques can be applied to a large GBS device, which is more relevant to practical applications. When the circuit size increases, an immediate issue arises as that the size of the Hilbert space for a fixed total photon number increases, so the probability of detecting a single click pattern decreases. It is not practical to estimate a tiny probability using a GBS device. It is thus necessary to consider the probability of a collection of click patterns, the coarse grained probability. One of the useful coarse grained probabilities is the orbit probability [40], which is critical in solving the graph isomorphism and graph similarity problems. An orbit O_n is defined as a collection of click patterns including all permutations of the click pattern n . The orbit probability P_{O_n} is the sum of all click-pattern probabilities inside an orbit O_n . Although we introduce the mitigation schemes by estimating the probability of a single click pattern, we show here that they can also be used to estimate the orbit probability, as well as other coarse grained probabilities.

Consider orbits with at most one photon in each mode up to 8 total photons: vacuum, [1], [11], [111], [1111], [11111], [111111], [1111111] and [11111111], where we have omitted "0" in the click pattern to simplify the notation and used a single click pattern to represent an orbit. The orbit probabilities in the absence of photon loss are (0.453, 0., 0.283, 0., 0.058, 0., 0.0044, 0., 0.00011). In the presence of photon loss, the photons

We consider a book graph of size eight, with adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \quad (42)$$

The adjacency matrix A has to be doubled and rescaled to be $cA \oplus A$, so that it can be encoded into an 8-mode pure Gaussian state σ_0 [40], which is generated by injecting eight pure single-mode squeezed vacuum states into a loss-free interferometer. Here we choose $c = 0.25$ so that the maximum input squeezing is about 7.25 dB, which is accessible for current quantum optics experiments. When the circuit is lossy, the generated state is different from σ_0 . We now use the proposed error mitigation schemes to estimate the orbit probabilities for a loss-free circuit.

tend to populate toward lower photon number orbits. We now apply the extrapolation technique to estimate the loss-free orbit probabilities. To perform the extrapolation, we choose five loss values and the vector \mathbf{c} is chosen as $\mathbf{c} = (1.0, 1.1, 1.2, 1.3, 1.4)$. The results are shown in Fig. 5. We can see that for low photon loss, both the extrapolation and the improved extrapolation work very well, giving good ap-

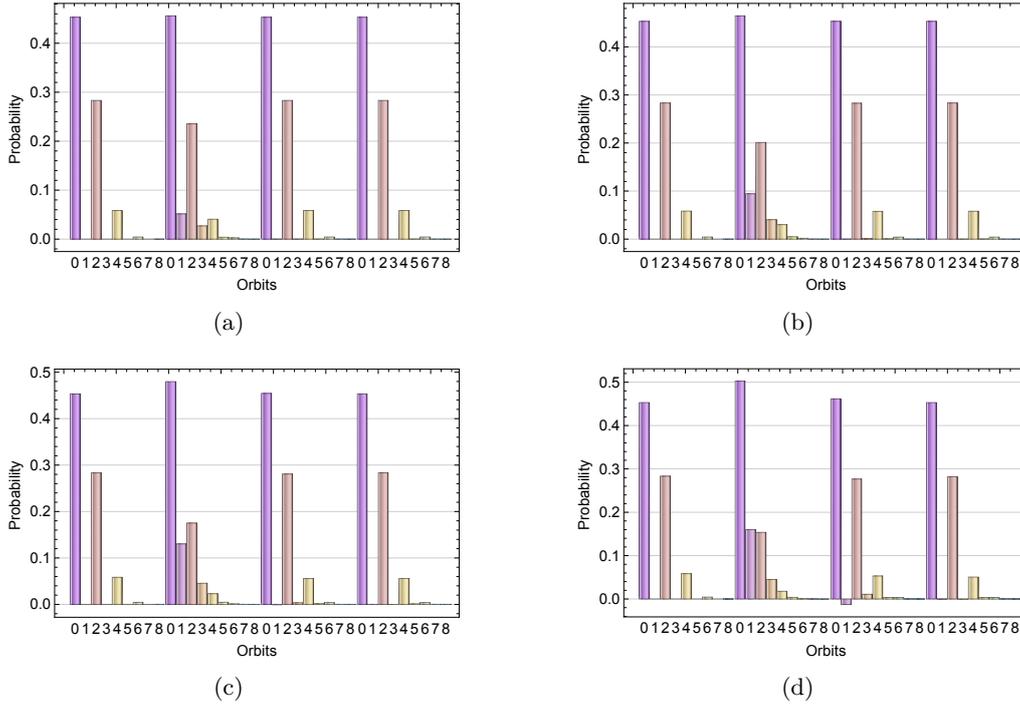


Figure 5: Estimate orbit probabilities using extrapolation technique. In each subfigure there are four histograms, they represent the orbit probabilities without photon loss, with photon loss, with extrapolation and with the improved extrapolation, respectively. The loss values for (a), (b), (c) and (d) are $\epsilon = 0.1, 0.2, 0.3$ and 0.4 , respectively.

proximations for most orbit probabilities. When the photon loss increases, some estimated orbit probabilities from extrapolation lost accuracy, e.g., the probability of the orbit with one photon becomes negative. However, for the improved extrapolation with poles removed we still obtain a good approximation, showing its advantage for high photon loss. We also use the loss cancellation method to estimate the orbit probability $P_0([00001111])$. With cutoff photon number $n_{\max} = 7$, we find that good approximation can be obtained for photon loss $\epsilon \leq 0.3$, see Table 4. This demonstrates that the proposed error mitigation techniques still work for a bigger circuit for reasonable amount of photon loss.

6 Summary

We have proposed two schemes to mitigate the effect of photon loss in a GBS device. The first scheme is based on the extrapolation technique and requires a small modification of the GBS circuit: to increase the photon loss of the circuit. The second scheme requires no modifications of the circuit and thus is hardware efficient. One only needs to measure the probabilities of a

lossy circuit and then linearly combine them in an appropriate way. The computational cost is to calculate the linear combination coefficients. We tested these error mitigation techniques in a two-mode and an eight-mode GBS circuits, and showed that they work extremely well for low photon loss, and also give fairly good approximations for relatively high photon loss.

In realistic experiments, the accuracy of the measured probabilities is limited by the finite number of samples and experimental imperfections. We show that the extrapolated probability is sensitive to the measured probabilities, and sometimes one gets apparently meaningless results like negative probabilities. This can be overcome by performing multiple experiments and taking the mean value as the estimate for the loss-free probability. The requirement of multiple experiments should be considered as the classical computational cost.

While the procedure that we have described so far is focused on GBS devices, similar ideas can also be applied to other near-term photonic architectures such as Boson Sampling. Boson Sampling differs from GBS in that expressions for click-pattern probabilities are related to perma-

nents of transformation matrices rather than to Hafnians as is the case in GBS. Devising a concrete procedure that accounts for this difference and mitigated Boson Sampling probabilities is an open problem. Another important research direction is to devise methods for error mitigation in sampling problems, which may require recovering the quantum states instead of recovering the expectation value of observables.

Acknowledgement: We thank Mark Wilde, Seth Lloyd, Kang Tan, Dylan Mahler for insightful discussions.

References

- [1] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [2] J. Preskill, Quantum Computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [3] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nature Physics* **14**, 595 (2018).
- [4] S. Aaronson, and L. Chen, Complexity-theoretic foundations of quantum supremacy experiments, [arXiv:1612.05903](https://arxiv.org/abs/1612.05903).
- [5] F. Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [6] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467**, 459 (2011).
- [7] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting quantum computations, *Phys. Rev. Lett.* **117**, 080501 (2016).
- [8] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Achieving quantum supremacy with sparse and noisy commuting quantum computations, *Quantum* **1**, 8 (2017).
- [9] S. Aaronson, A. Arkhipov, The computational complexity of linear optics, *Proceedings of the forty-third annual ACM symposium on Theory of computing*, 333-342 (2011).
- [10] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, Christine, and I. Jex, Gaussian Boson Sampling, *Phys. Rev. Lett.* **119**, 170501 (2017).
- [11] S. Rahimi-Keshari, A. P. Lund, and T. C. Ralph, What Can Quantum Optics Say about Computational Complexity Theory?, *Phys. Rev. Lett.* **114**, 060501 (2015).
- [12] S. Rahimi-Keshari, T. C. Ralph, and C. M. Caves, Sufficient Conditions for Efficient Classical Simulation of Quantum Optics, *Phys. Rev. X* **6**, 021039 (2016).
- [13] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nature Communications* **5**, 4213 (2014).
- [14] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [15] E. Farhi, and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, [arXiv:1602.07674](https://arxiv.org/abs/1602.07674).
- [16] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [17] Y. Li, and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [18] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* **567**, 491 (2019).
- [19] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, *Phys. Rev. X* **8**, 031027 (2018).
- [20] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, H. and Li, Ying, Quantum computation with universal error mitigation on a superconducting quantum processor, *Science Advances* **5**, (2019).

- [21] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J. Zhang, and K. Kim, Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system, *Nature Communications* **11**, 1 (2020).
- [22] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O’Brien, Low-cost error mitigation by symmetry verification, *Phys. Rev. A* **98**, 062339 (2018).
- [23] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O’Brien, and L. DiCarlo, Experimental error mitigation via symmetry verification in a variational quantum eigensolver, *Phys. Rev. A* **100**, 010302(R) (2019).
- [24] S. McArdle, X. Yuan, and S. Benjamin, Error-Mitigated Digital Quantum Simulation, *Phys. Rev. Lett.* **122**, 180501 (2019).
- [25] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O’Brien, Low-cost error mitigation by symmetry verification, *Phys. Rev. A* **98**, 062339 (2018).
- [26] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, Variational quantum state eigensolver, [arXiv:2004.01372](https://arxiv.org/abs/2004.01372).
- [27] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New Journal of Physics* **18**, 023023 (2016).
- [28] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, Noise resilience of variational quantum compiling, *New Journal of Physics* **22**, 043006 (2020).
- [29] L. Cincio, K. Rudinger, M. Sarovar, and P. J. Coles, Machine learning of noise-resilient quantum circuits, *PRX Quantum* **2**, 010324 (2021).
- [30] Y. Chen, M. Farahzad, S. Yoo, and T. Wei, Detector tomography on IBM quantum computers and mitigation of an imperfect measurement, *Phys. Rev. A* **100**, 052315 (2019).
- [31] M. R. Geller, and M. Sun, Efficient correction of multiqubit measurement errors, [arXiv:2001.09980](https://arxiv.org/abs/2001.09980).
- [32] L. Funcke, T. Hartung, K. Jansen, S. Kühn, P. Stornati, and X. Wang, Measurement error mitigation in quantum computers through classical bit-flip correction, [arXiv:2007.03663](https://arxiv.org/abs/2007.03663).
- [33] H. Kwon, and J. Bae, A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms, *IEEE Transactions on Computers* (2020).
- [34] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, *Phys. Rev. A* **95**, 042308 (2017).
- [35] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, Mitigating Realistic Noise in Practical Noisy Intermediate-Scale Quantum Devices, *Phys. Rev. Applied* **15**, 034026 (2021).
- [36] A. Strikis, D. Qin, Y. Chen, B. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, [arXiv:2005.07601](https://arxiv.org/abs/2005.07601).
- [37] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, [arXiv:2005.10189](https://arxiv.org/abs/2005.10189).
- [38] A. Zlokapa, and A. Gheorghiu, A deep learning model for noise prediction on near-term quantum devices, [arXiv:2005.10811](https://arxiv.org/abs/2005.10811).
- [39] J. Arrazola, and T. R. Bromley, Using Gaussian Boson Sampling to Find Dense Subgraphs, *Phys. Rev. Lett.* **121**, 030503 (2018).
- [40] K. Brádler, S. Friedland, J. Izaac, N. Killoran, and D. Su, Graph isomorphism and Gaussian boson sampling, *Spec. Matrices* **9**, 166 (2021).
- [41] M. Schuld, K. Brádler, R. Israel, D. Su, and B. Gupt, Measuring the similarity of graphs with a Gaussian boson sampler, *Phys. Rev. A* **101**, 032314 (2020).
- [42] K. Brádler, R. Israel, M. Schuld, and D. Su, A duality at the heart of Gaussian boson sampling, [arXiv:1910.04022](https://arxiv.org/abs/1910.04022).
- [43] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
- [44] K. Brádler, P. Dallaire-Demers, P. Reben-trost, D. Su, and C. Weedbrook, Gaussian

- boson sampling for perfect matchings of arbitrary graphs, *Phys. Rev. A* **98**, 032310 (2018).
- [45] H. Qi, D. J. Brod, N. Quesada, and R. García-Patrón, Regimes of Classical Simulability for Noisy Gaussian Boson Sampling, *Phys. Rev. Lett.* **124**, 100502 (2020).
- [46] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, *Optica* **3**, 1460 (2016).
- [47] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental Realization of Any Discrete Unitary Operator, *Phys. Rev. Lett.* **73**, 58 (1994).
- [48] M. Jacques, A. Samani, E. El-Fiky, D. Patel, X. Zhenping, and D. V. Plant, Optimization of thermo-optic phase-shifter design and mitigation of thermal crosstalk on the SOI platform, *Opt. Express* **27**, 10456 (2019).
- [49] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, 2017).
- [50] J. Huh, G. G. Guerreschi, B. Peropadre, J. R. McClean, and A. Aspuru-Guzik, Boson sampling for molecular vibronic spectra, *Nature Photonics* **9**, 615 (2015).
- [51] S. Rahimi-Keshari, M. A. Broome, R. Fickler, A. Fedrizzi, T. C. Ralph, and A. G. White, Direct characterization of linear-optical networks, *Opt. Express* **21**, 13450 (2013).
- [52] V. Giovannetti, A. S. Holevo, and R. García-Patrón, A Solution of Gaussian Optimizer Conjecture for Quantum Channels, *Commun. Math. Phys.* **334**, 1553 (2015).
- [53] R. García-Patrón, J. Renema, and V. Shchepinov, Simulating boson sampling in lossy architectures, *Quantum* **3**, 169 (2019).
- [54] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Detailed study of Gaussian boson sampling, *Phys. Rev. A* **100**, 032326 (2019).

A Uniform loss approximation

In this appendix, we briefly review the physical implementation of a linear optics interferometer and how to model the photon loss, and demonstrate under what conditions the uniform loss approximation is valid. A reconfigurable linear optics interferometer consists of an array of tunable beam splitters and phase shifters. Practically, a tunable beam splitter is implemented by two static 50:50 beam splitters and two phase shifters. The transmission and reflection coefficients of the tunable beam splitter can be varied by changing the phases of the phase shifters [47]. Due to the imperfect implementation of the tunable beam splitter, photons may be lost when going through it. Here we assume that the loss rate of each of the two modes are the same. Mathematically, the lossy tunable beam splitter can be modelled by adding two beam splitters with reflection coefficient the same as the loss rate after a perfect tunable beam splitter. The photons may also be lost when travelling through the medium between adjacent tunable beam splitters, e.g., the fibre or waveguide. Similarly, this lossy channel can be modelled by a beam splitter, which can be effectively combined with the beam splitter that models the loss of the tunable beam splitter.

Any linear unitary matrix $U(N)$ can be decomposed into a product of a sequence of 2×2 unitary matrices, which correspond physically to the tunable beam splitters [47]. There are two main schemes to perform the decomposition: the Reck's scheme [47] and the Clements' scheme [46]. The former implements a linear unitary transformation by arranging the tunable beam splitters in a triangular configuration. The photon entering different mode will experience very different path length. Given that the loss rate of each tunable beam splitter is almost the same, the overall loss rate for each mode is quite different, resulting in nonuniform loss. The Clements' scheme implements a linear unitary transformation by arranging the tunable beam splitters in a rectangular configuration, in which the photon entering different mode will experience almost the same path length. If the loss rate of each tunable beam splitter is almost the same, then the overall loss rate for each mode is almost the same,

resulting in uniform loss.

Consider an M -mode linear interferometer. If the interferometer is perfect, then its input-output relation is given by $\hat{\mathbf{b}} = U\hat{\mathbf{a}}$, where $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_M)$ with \hat{a}_i the input annihilation operator, $\hat{\mathbf{b}} = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_M)$ with \hat{b}_i the output annihilation operator, and U is an $M \times M$ unitary matrix representing the transformation of the linear interferometer. If the linear interferometer is lossy, then the input-output relation has to be modified as [52, 53]

$$\hat{\mathbf{b}} = A\hat{\mathbf{a}} + \sqrt{I - AA^\dagger} \hat{\mathbf{e}}, \quad (43)$$

where $\hat{\mathbf{e}}$ represents the environmental modes and A is a complex matrix satisfying $AA^\dagger \leq I$. The matrix A can be decomposed as $A = V\hat{\lambda}W$, where V and W are unitary matrices, and $\hat{\lambda} = \text{diag}\{\sqrt{\eta_1}, \sqrt{\eta_2}, \dots, \sqrt{\eta_M}\}$ with $\eta_i \in [0, 1]$. The singular decomposition of A implies that the transformation of a lossy linear interferometer, Eq. (43), is mathematically equivalent to first applying a unitary transformation W , followed by M lossy channels with transmission coefficients η_i (or loss values $1 - \eta_i$), then applying another unitary transformation V .

In general, the singular eigenvalues of A are different and no further simplification can be made. However, when the photon entering each mode goes through almost the same number of tunable beam splitters and experiences almost the same path length, then the singular eigenvalues $\sqrt{\eta_i}$ are almost the same. This implies the diagonal matrix $\hat{\lambda}$ is close to an identity matrix, $\hat{\lambda} = \sqrt{\eta}I$, where η is the overall transmission coefficient of each mode. In this case the matrix $\hat{\lambda}$ commutes with V and the matrix A can be rewritten as $A = \hat{\lambda}\tilde{U}$, where $\tilde{U} = VW$ is a unitary transformation. This shows that a linear interferometer with uniform loss is equivalent to a perfect linear interferometer followed by M lossy channels with the same transmission coefficient.

B Nonzero displacement

When the displacements are not zero, the probability of the click pattern \mathbf{n} is given by [54]

$$P(\mathbf{n}) = \frac{\exp(-\frac{1}{2}\mathbf{d}^\dagger \sigma_Q^{-1} \mathbf{d})}{\mathbf{n}! \sqrt{\det \sigma_Q}} \prod_{k=1}^M \left(\frac{\partial^2}{\partial \alpha_k \partial \alpha_k^*} \right)^{n_k} \exp\left(\frac{1}{2}\boldsymbol{\alpha}_v^\top A \boldsymbol{\alpha}_v + \mathbf{F}^\dagger \boldsymbol{\alpha}_v\right) \Big|_{\boldsymbol{\alpha}_v=\mathbf{0}}, \quad (44)$$

where $\mathbf{F} = \sigma_Q^{-1} \mathbf{d}$. In the presence of photon loss, the probability $P(\mathbf{n}; \epsilon)$ is obtained by replacing σ_Q and A in Eq. (44) by $\sigma_Q(\epsilon)$ and $A(\epsilon)$ given by Eqs. (9) and (10), respectively. Since both $A(\epsilon)$ and $\sigma_Q(\epsilon)$ have series expansions with respect to ϵ , a similar series expansion for $P(\mathbf{n}; \epsilon)$ like Eq. (12) can be obtained, thus the extrapolation applies.

By using the relation between A and σ_Q in Eq. (2), we find $[\sigma_Q(\epsilon)]^{-1} = \mathbb{I}_{2M} - X_{2M}A(\epsilon)$, and by further using the decomposition of $A(\epsilon)$ in Eq. (19), we have

$$\begin{aligned} [\sigma_Q(\epsilon)]^{-1} &= U \bigoplus_{k=1}^M \begin{pmatrix} \frac{1-\tanh(r_k)}{1-\epsilon \tanh(r_k)} & 0 \\ 0 & \frac{1+\tanh(r_k)}{1+\epsilon \tanh(r_k)} \end{pmatrix} U^\dagger \\ &= \frac{1}{\mathcal{P}(\epsilon, \tilde{\mathbf{r}})} U \bigoplus_{k=1}^M \begin{pmatrix} \frac{1-\tanh(r_k)}{1-\epsilon \tanh(r_k)} \mathcal{P}(\epsilon, \tilde{\mathbf{r}}) & 0 \\ 0 & \frac{1+\tanh(r_k)}{1+\epsilon \tanh(r_k)} \mathcal{P}(\epsilon, \tilde{\mathbf{r}}) \end{pmatrix} U^\dagger. \end{aligned} \quad (45)$$

By using Eq. (45) we can show that

$$\prod_{k=1}^M \left(\frac{\partial^2}{\partial \alpha_k \partial \alpha_k^*} \right)^{n_k} \exp\left(\frac{1}{2}\boldsymbol{\alpha}_v^\top A(\epsilon) \boldsymbol{\alpha}_v + \mathbf{F}^\dagger \boldsymbol{\alpha}_v\right) \Big|_{\boldsymbol{\alpha}_v=\mathbf{0}} = \frac{1}{\mathcal{P}^N(\epsilon, \tilde{\mathbf{r}})} \mathbb{P}_d(\epsilon), \quad (46)$$

where $\mathbb{P}_d(\epsilon)$ is a polynomial of ϵ . Therefore, the probability of measuring a click pattern \mathbf{n} in the presence of photon loss and displacements can be written as

$$P(\mathbf{n}; \epsilon) = \frac{\exp\left\{-\frac{1}{2}\mathbf{d}^\dagger [\sigma_Q(\epsilon)]^{-1} \mathbf{d}\right\}}{\mathbf{n}!} \left(\prod_{k=1}^M \frac{1}{\cosh r_k} \right) \frac{1}{\mathcal{Q}(\epsilon, \mathbf{r}) \mathcal{P}^N(\epsilon, \tilde{\mathbf{r}})} \mathbb{P}_d(\epsilon). \quad (47)$$

By using Eq. (45), the quantity $\mathbf{d}^\dagger[\sigma_Q(\epsilon)]^{-1}\mathbf{d}$ can be simplified as

$$\mathbf{d}^\dagger[\sigma_Q(\epsilon)]^{-1}\mathbf{d} = (\mathbf{d}^\dagger U) \bigoplus_{k=1}^M \begin{pmatrix} \frac{1-\tanh(r_k)}{1-\epsilon \tanh(r_k)} & 0 \\ 0 & \frac{1+\tanh(r_k)}{1+\epsilon \tanh(r_k)} \end{pmatrix} (U^\dagger \mathbf{d}) = \mathbf{d}_{\text{in}}^\dagger \bigoplus_{k=1}^M \begin{pmatrix} \frac{1-\tanh(r_k)}{1-\epsilon \tanh(r_k)} & 0 \\ 0 & \frac{1+\tanh(r_k)}{1+\epsilon \tanh(r_k)} \end{pmatrix} \mathbf{d}_{\text{in}},$$

where $\mathbf{d}_{\text{in}} = U^\dagger \mathbf{d}$ is the input displacement vector. Therefore the exponential term $\exp\{-\frac{1}{2}\mathbf{d}^\dagger[\sigma_Q(\epsilon)]^{-1}\mathbf{d}\}$ is known if the input squeezing parameters and displacements are known, and does not need to be extrapolated. The only term that is unknown is the polynomial $\mathbb{P}_d(\epsilon)$, which requires extrapolation. The improved extrapolation technique can be applied here since the poles are determined by $\mathcal{Q}(\epsilon \mathbf{r})$ and $\mathcal{P}(\epsilon, \tilde{\mathbf{r}})$ which can be removed by simply moving them to the left hand side of Eq. (47).

C Convergence of the loss cancellation procedure

In the Loss Cancellation procedure, $T_{\ominus\epsilon}(P')$ is expressed by an infinite series. We may ask how small $|\epsilon|$ must be to ensure that the series converges, i.e. that our approximations approach a limit as the cutoff goes to ∞ . This will depend on decay properties of the distribution P .

Definition The decay radius of the distribution P is the infimum of $u > 0$ such that there exists a constant $C(u)$ with $|P(\mathbf{n})| \leq C(u)u^{|\mathbf{n}|}$ for all \mathbf{n} .

Lemma Suppose P has decay radius $\leq t$, i.e. for any $u \in (t, 1)$ there exists a constant $C(u)$ such that $|P(\mathbf{n})| \leq C(u)u^{|\mathbf{n}|}$ for all \mathbf{n} . Then for any ϵ (not necessarily positive), the series for $T_\epsilon(P)$ converges if $|\epsilon| < 1/t$, and the result has decay radius $\leq \frac{1-|\epsilon|t}{1-|\epsilon|}$.

Proof. We have

$$|T_\epsilon(P)(\mathbf{n}')| \leq \sum_{k \geq |\mathbf{n}'|} L(\epsilon, k, \mathbf{n}'),$$

where

$$\begin{aligned} L(\epsilon, k, \mathbf{n}') &= \sum_{\mathbf{n}: \mathbf{n} \geq \mathbf{n}', |\mathbf{n}|=k} |\text{Prob}(\mathbf{n}'|\mathbf{n})P(\mathbf{n})| \\ &\leq \left(\sum_{\mathbf{n}: \mathbf{n} \geq \mathbf{n}', |\mathbf{n}|=k} \prod_{j=1}^m \binom{n_j}{n'_j} \right) |\epsilon|^{k-|\mathbf{n}'|} |1-\epsilon|^{|\mathbf{n}'|} C(u)u^k. \end{aligned}$$

We claim that

$$\sum_{\mathbf{n}: \mathbf{n} \geq \mathbf{n}', |\mathbf{n}|=k} \prod_{j=1}^m \binom{n_j}{n'_j} = \binom{k+m-1}{|\mathbf{n}'|+m-1}.$$

The claim can be proven by a “stars and bars” argument. Let $|\mathbf{n}'| = n$. The left side is the number of objects consisting of an m -tuple $\mathbf{n} \geq \mathbf{n}'$ and, for each $j \in \{1, \dots, m\}$, a subset of cardinality n'_j of $[1, \dots, n_j]$. The right side counts subsets of cardinality $n+m-1$ of $[1, \dots, k+m-1]$. These subsets can be placed in one-to-one correspondence with the objects on the left side as follows. If the subset is $[T_1, \dots, T_{n+m-1}]$, we designate $T_{n'_1+1}, T_{n'_1+n'_2+2}, \dots, T_{n'_1+\dots+n'_{m-1}+m-1}$ as “bars” b_1, \dots, b_{m-1} and the others as “stars”, so that there are n stars separated by bars into groups of n'_1, \dots, n'_m . We take $n_1 = b_1 - 1$, $n_2 = b_2 - b_1 - 1$, \dots , $n_m = k+m-1 - b_{m-1}$. The subset of cardinality n'_j of $[1, \dots, n_j]$ then consists of the j 'th group of “stars” translated to the left (if $j > 1$) by b_j .

Now, with $|\mathbf{n}'| = n$, we have

$$L(\epsilon, k, \mathbf{n}') \leq C \binom{k+m-1}{n+m-1} |\epsilon|^{k-n} |1-\epsilon|^n u^k.$$

Since

$$\binom{k+m-1}{n+m-1} \leq \frac{(k+m-1)^{n+m-1}}{(n+m-1)!},$$

the sum over k converges absolutely if $|\epsilon u| < 1$.

Now since

$$\sum_{k=n}^{\infty} \binom{k+m-1}{n+m-1} z^k = \sum_{j=0}^{\infty} \binom{j+n+m-1}{n+m-1} z^{n+j} = \frac{z^n}{(1-z)^{n+m+2}}$$

for $|z| < 1$, we get a bound

$$|T_{\epsilon}(P)(\mathbf{n}')| \leq C \frac{|1 - \epsilon|^{n+m} u^n}{(1 - |\epsilon|u)^{n+m-2}},$$

so that $T_{\epsilon}(P)$ has decay radius $\leq \frac{|1 - \epsilon|t}{1 - |\epsilon|t}$. □

Now we apply the lemma to loss cancellation. Suppose the lossless distribution P has decay radius t . If $0 < \epsilon < 1/t$, the distribution with loss $P' = T_{\epsilon}(P)$ has decay radius $\leq \frac{(1 - \epsilon)t}{1 - \epsilon t}$. Then with

$\ominus\epsilon = \epsilon/(\epsilon - 1)$, the series for $P = T_{\ominus\epsilon}(P')$ converges if $\left| (\ominus\epsilon) \frac{(1 - \epsilon)t}{1 - \epsilon t} \right| < 1$, and this is equivalent to

$$\epsilon < \frac{1}{2t}.$$

The two-mode example is exceptional in that the decay radius can easily be seen to be $\chi = \tanh(r)$. so we want $\epsilon < 1/(2\chi)$ to ensure convergence. For squeezing $r = 1/2$, $1/(2\chi) > 1.08$, so the series converges for all $\epsilon \in (0, 1)$. But for squeezing $r = 1$, $1/(2\chi) \approx 0.6565$.

In general it may be difficult to predict in advance the decay radius for P , but we may conjecture that it is typically finite and nonzero. The loss-cancellation procedure can then be expected to work very well if ϵ is sufficiently small, but very poorly when ϵ is too large.