Important Notice to Authors

No further publication processing will occur until we receive your response to this proof.

Attached is a PDF proof of your forthcoming article in PRA. Your article has 15 pages and the Accession Code is **AJ11649**.

Please note that as part of the production process, APS converts all articles, regardless of their original source, into standardized XML that in turn is used to create the PDF and online versions of the article as well as to populate third-party systems such as Portico, Crossref, and Web of Science. We share our authors' high expectations for the fidelity of the conversion into XML and for the accuracy and appearance of the final, formatted PDF. This process works exceptionally well for the vast majority of articles; however, please check carefully all key elements of your PDF proof, particularly any equations or tables.

Figures submitted electronically as separate files containing color appear in color in the online journal. However, all figures will appear as grayscale images in the print journal unless the color figure charges have been paid in advance, in accordance with our policy for color in print (https://journals.aps.org/authors/color-figures-print).

Specific Questions and Comments to Address for This Paper

1 Please check byline addresses.
2 Please check definition of PDF.
3 Please note that Phys. Rev. style discourages the use of one-sentence paragraphs.
4 Please check Fig. 5 caption.
5 Please check deletion of multiplication sign in equations following (42).
6 Please note that Fig. 7 has been cited before Fig. 6 in the text. Please verify.
7 Please check definition of GHZ.
8 Please check substitution of "IVB" and "IVD" for "3.2" and "3.4."
9 Please check substitution of Table V for VI here.
10 Please check substitution of "above" for "below." If this is not what you mean please clarify.

Open Funder Registry: Information about an article's funding sources is now submitted to Crossref to help you comply with current or future funding agency mandates. Crossref's Open Funder Registry (https://www.crossref.org/services/funder-registry/) is the definitive registry of funding agencies. Please ensure that your acknowledgments include all sources of funding for your article following any requirements of your funding sources. Where possible, please include grant and award ids. Please carefully check the following funder information we have already extracted from your article and ensure its accuracy and completeness:

Australian Research Council (AU), CE110001027
ARC (AU), FT160100397

Other Items to Check

- Please note that the original manuscript has been converted to XML prior to the creation of the PDF proof, as described above. Please carefully check all key elements of the paper, particularly the equations and tabular data.
- Title: Please check; be mindful that the title may have been changed during the peer-review process.
- Author list: Please make sure all authors are presented, in the appropriate order, and that all names are spelled correctly.
- Please make sure you have inserted a byline footnote containing the email address for the corresponding author, if desired. Please note that this is not inserted automatically by this journal.
- Affiliations: Please check to be sure the institution names are spelled correctly and attributed to the appropriate author(s).
- Receipt date: Please confirm accuracy.
- Acknowledgments: Please be sure to appropriately acknowledge all funding sources.
- Hyphenation: Please note hyphens may have been inserted in word pairs that function as adjectives when they occur before a noun, as in "x-ray diffraction," "4-mm-long gas cell," and "*R*-matrix theory." However, hyphens are deleted from word pairs when they are not used as adjectives before nouns, as in "emission by x rays," "was 4 mm in length," and "the *R* matrix is tested."

  Note also that Physical Review follows U.S. English guidelines in that hyphens are not used after prefixes or before suffixes: superresolution, quasiequilibrium, nanoprecipitates, resonancelike, clockwise.
- Please check that your figures are accurate and sized properly. Make sure all labeling is sufficiently legible. Figure quality in this proof is representative of the quality to be used in the online journal. To achieve manageable file size for online delivery, some compression and downsampling of figures may have occurred. Fine details may have become somewhat fuzzy, especially in color figures. The print journal uses files of higher resolution and therefore details may be sharper in print. Figures to be published in color online will appear in color on these proofs if viewed on a color monitor or printed on a color printer.
- Please check to ensure that reference titles are given as appropriate.

- Overall, please proofread the entire *formatted* article very carefully. The redlined PDF should be used as a guide to see changes that were made during copyediting. However, note that some changes to math and/or layout may not be indicated.

Ways to Respond

- ***Web:*** If you accessed this proof online, follow the instructions on the web page to submit corrections.
- ***Email:*** Send corrections to praproofs@aptaracorp.com
  Subject: **AJ11649** proof corrections
- ***Fax:*** Return this proof with corrections to +1.703.791.1217. Write **Attention:** PRA Project Manager and the Article ID, **AJ11649**, on the proof copy unless it is already printed on your proof printout.

PHYSICAL REVIEW A **00**, 002300 (2018)

Passive quantum error correction of linear optics networks through error averaging

Ryan J. Marshman,[1] Austin P. Lund,[1] Peter P. Rohde,[2,3,*] and Timothy C. Ralph[1]

[1]*Centre for Quantum Computation and Communication Technology, Department of Mathematics and Physics, The University of Queensland, St. Lucia, Queensland 4072, Australia*

[2]*Centre for Quantum Software and Information, Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway, New South Wales 2007, Australia*

[3]*Hearne Institute for Theoretical Physics and Department of Physics & Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA*

(Received 24 September 2017; published xxxxxx)

We propose and investigate a method of error detection and noise correction for bosonic linear networks using a method of unitary averaging. The proposed error averaging does not rely on ancillary photons or control and feedforward correction circuits, remaining entirely passive in its operation. We construct a general mathematical framework for this technique and then give a series of proof of principle examples including numerical analysis. Two methods for the construction of averaging are then compared to determine the most effective manner of implementation and probe the related error thresholds. Finally we discuss some of the potential uses of this scheme.

DOI: 10.1103/PhysRevA.00.002300

I. INTRODUCTION

The evolution of a multimode bosonic quantum state in a linear network can be simply described by a linear set of equations relating input and output bosonic modes. These types of interactions are of interest as they are simple to arrange for most experiments involving electromagnetism but nevertheless are useful and have interesting quantum information applications.

Linear networks are not universal for quantum information processing on their own. However, they can be made universal using postselection and feedforward methods with a polynomial overhead in the number of photons [1–3]. More recently they have been shown to deterministically generate quantum statistics that cannot be efficiently computed using classical computing resources alone (i.e., the boson-sampling problem) [4]. They also form the basis for optical quantum walks, for which numerous applications have been described, and have been subject to widespread experimental demonstration [5–9].

Linear networks for quantum optics experiments have traditionally been implemented using bulk optical devices [3]. However, efforts to build integrated optical circuits have meant that the size of the networks has the potential to be made orders of magnitude smaller and consequently there is a great potential for their complexity to increase [10].

In theoretical proposals for optical quantum information tasks using linear networks, it is often assumed that it is possible to configure an arbitrary linear network rapidly and with high precision. This paper considers the second of these requirements by studying the effects of imprecision in configuring linear networks.

The model we consider assumes that large linear networks can be configured arbitrarily but with some additional noise. This may be due to experimental imprecision of defining linear network parameters which shot by shot results in fluctuations of the parameters around their mean values. We wish to concentrate on the effects due to linear network errors, so we assume ideal generation of Fock basis states and the ability to make ideal Fock basis detections. Furthermore, we also assume that the networks have no loss at any stage, whether in the injection of states, the outcoupling to detection devices, or the network itself.

We show that by redundantly encoding the network matrix describing a desired linear network, it is possible to generate an effect which tends towards the target network matrix when averaging over the redundant encoding. The averaging effect occurs in a nondeterministic manner and hence the transformation acts as a filter where noise is directed into outputs which are then postselected away (see Fig. 1). The central-limit theorem applies to the individual matrix elements of the averaged transformation and hence their variance decreases as $\frac{1}{N}$. The form of the average matrix and the distribution of the transformations on a finite number of averages depend on the details of the noise applied to the network encoding. The results presented in this paper analyze these details, showing conditions in which this technique may be of utility.

The next section introduces the averaging scheme and some mathematical details that apply in the most general case. Section III includes numerous proof of principle examples which serve to highlight the effects of error averaging with a focus on the behavior of the probability of success. Section IV studies two different ways of redundantly encoding a single-mode phase shift and the effects of the different encodings on the resultant error and probability of success. We then numerically analyze the averaging method for a four-mode operation in Sec. V. We discuss some of the consequence of these results as well as future directions in Sec. VI and

*dr.rohde@gmail.com; http://www.peterrohde.org

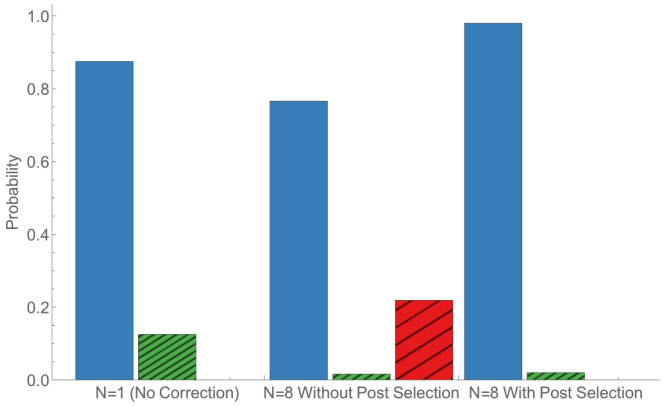

FIG. 1. Comparison of output probability distributions with and without error averaging. Here N corresponds to the number of redundant copies of the unitary being employed. The effect postselection has on the output distribution can be seen. The blue solid bars represent the probability of observing the photon in the correct output mode, green thin-striped bars correspond to observing the photon in the incorrect output mode, and the red thick-striped bar corresponds to observing the photon in any of the error detection modes. The probabilities are based on a single photon in a Mach-Zehnder interferometer with an individual phase shifter variance $v = 0.5$ rad^2.

draw comparisons between error averaging and standard error correction in Sec. VII before summarizing in Sec. VIII.

II. GENERAL UNITARY ERROR AVERAGING

Here we are concerned with the case of bosonic linear scattering networks. These are evolutions of a multimode bosonic field where the Heisenberg equations of motion for the annihilation operators of each mode can be written as a linear combination of all annihilation operators. That is, if $\mathcal{U}_U$ is a unitary operation on an m-mode system, then

$$\mathcal{U}_U a_i \mathcal{U}_U^\dagger = \sum_j U_{ij} a_j, \tag{1}$$

where, to preserve commutation relationships, U must be a unitary matrix. It is the network matrix U that we will focus on.

The error model we are considering can then be stated using this construction. We consider the matrix U to be randomly selected from any distribution over unitary matrices. Though this definition is quite broad, we are particularly interested in nonuniform (or more precisely non-Haar) distributions that describe a realization of constructing particular optical elements. For example, one might consider a large number of beam splitters which have, on average, a reflectivity of 50%. However, a randomly chosen beam splitter's reflectivity varies slightly from this average value in a way that is specified by some probability distribution. Alternatively, one could imagine an integrated optical circuit where the inputs have some small random fluctuations. Each time this integrated circuit is utilized, the underlying parameters will be different. The distributions of unitary matrices we initially consider will be arbitrary, but later we will concentrate on the case of distributions near the identity with small variances and Gaussian profiles in the coefficients for the differential generators of the unitary group.

This model is not a universal description of error processes that could occur. In particular, errors such as loss, mode mismatch, and nonlinearities are not included. It is possible that there is some component of these processes that may be reducible under our technique. However, it is also possible that their effects be exacerbated. For this paper, we will focus only on the errors in describing the unitary network matrix and consider other error processes to be absent.

We will now describe the error-averaging technique. Consider a linear network whose elements are those of a discrete Fourier transform (DFT). That is, we have a Heisenberg-style evolution between mode annihilation operators of the form

$$a_{j,r} \to \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega^{rk} a_{j,k}, \tag{2}$$

where $\omega = e^{-i2\pi/N}$ and zero indexing has been used, that is, $k = 0$ corresponds to the first mode. The first subscript for the annihilation operator denotes the input mode and the second describes a quantity of redundancy N, which we explain shortly.

We then act on the N copies of a target unitary U. By this we mean that there is some variation between the copies, but the intention is to implement the unitary U. This can be described by the transformation

$$a_{j,r} \to \sum_{l=0}^{m-1} (U_r)_{lj} a_{l,r}, \tag{3}$$

where N noisy copies of U are made, denoted here by $U_1, U_2, \ldots, U_N$, where we assume an independent-error model across the redundancies.

After this the DFT matrix is applied again. This results in the overall transformation

$$a_{j,r} \to \frac{1}{N} \sum_{l=0}^{m-1} \sum_{k,k'=0}^{N-1} (U_{k'})_{lj} \omega^{(r+k)k'} a_{l,k}. \tag{4}$$

We consider the case where all redundant modes are initialized in the vacuum state and postselect on the cases where no photons are present in the output of the redundant modes. This means that we only need to consider the parts of this transformation expression where the second subscript of the annihilation operator is zero. In this case we have

$$a_{j,0} \to \frac{1}{N} \sum_{l=0}^{m-1} \sum_{k'=0}^{N-1} (U_{k'})_{lj} a_{l,0} = \sum_{l=0}^{m-1} (M_N)_{lj} a_{l,0}, \tag{5}$$

where M_N is a matrix defined by

$$M_N = \frac{1}{N} \sum_k U_k. \tag{6}$$

This matrix is then the effective linear network matrix for the postselected system. It includes information about the probability of success and so in general it will be not unitary. The remainder of this paper is directed towards analyzing the scenarios that arise from the multitude of choices for U_k that form the expression.

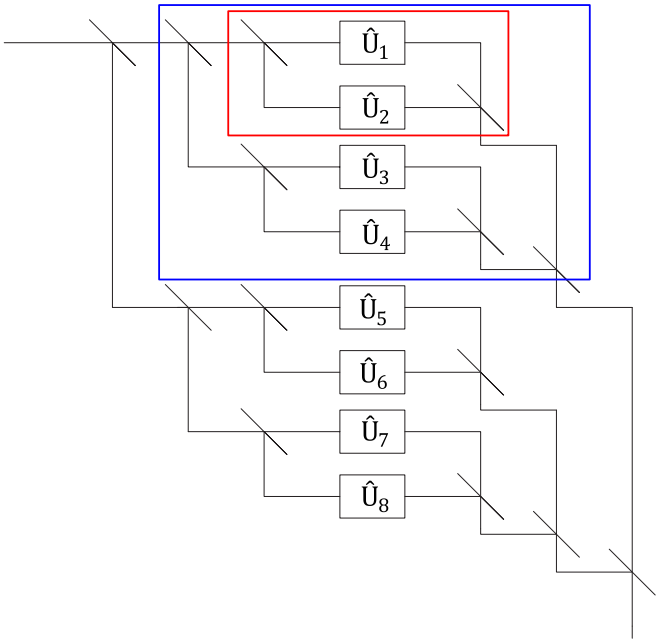

FIG. 2. Redundant encoding using 50:50 beam splitters for $N = 8$. The boxes labeled $\hat{U}_i$ can be single mode or multimode. In the multimode case the encoding beam-splitter network is repeated for each mode input and output. Output modes that are postselected on the vacuum are not shown here. The inner red box shows an $N = 2$ level encoding and the outer blue box shows $N = 4$. Further nesting of this arrangement can achieve any N being a power of 2.

Using the DFT matrix for redundantly encoding and decoding has been used here to illustrate in a compact way the generic approach. However, it is not necessary to perform these operations in exactly this way. The encoding and decoding will give the desired properties provided the transformation of Eq. (5) is generated. Figure 2 shows how this encoding can be realized using 50:50 beam splitters for eight redundant copies of the system. This approach gives a method of encoding which is recursive and would allow the components implementing the encoding to themselves be redundantly encoded. Given this, although the encoding beam splitters are assumed perfect, this requirement can potentially be relaxed. See Sec. V for more detail.

For the main theorem of our work we consider a general linear network described by a unitary network matrix U with any dimensionality.

Theorem 1. Given N linear networks described by unitary matrices $\{U_1, U_2, \ldots, U_N\}$ that are random with independent and identically distributed statistics such that for all $i \in 1, \ldots, N$, $\langle U_i \rangle = M$, then the random variable

$$M_N = \frac{1}{N} \sum_{i=1}^{N} U_i \tag{7}$$

is a matrix with mean value M and whose matrix elements have variance scaling as $O(1/N)$.

Proof. Our aim in the proof is to use the central-limit theorem. Consider the matrix element r,s of M_N. This is a random variable

$$(M_N)_{rs} = \frac{1}{N} \sum_{i=1}^{N} (U_i)_{rs}. \tag{8}$$

As the matrix elements $(U_i)_{rs}$ are constructed from unitary matrices, their magnitude is bounded by 1. Given this finite domain, the real and imaginary parts have maximum variance and covariance of 1 (though these extremal values are not simultaneously achievable). Given this bounded variance, we can use the central-limit theorem to conclude that the matrix element $(M_N)_{rs}$ is a random variable with mean value M_{rs}. The variance of the real or imaginary part of $(M_N)_{rs}$ is then upper bounded by $1/N$ as per the central-limit theorem. ■

Here we are explicitly taking the matrices U_i to be unitary. This fixes our error model to consist of random unitary errors. Theorem 1 also holds for any set of independent, random, and bound matrices U_i which might allow this technique to be effective for other types of errors. However, this will not give protection against loss errors. The situation to have in mind with regard to the systems considered in this paper are reprogrammable linear optical devices which, due to thermal noise or some other source of random noise, fail to exactly reproduce the desired target unitary.

The question now is what forms the mean average matrix M, as defined in Theorem 1, can take. First we consider the trivial case where the unitary matrices are 1×1 dimensional.

Corollary 1. If each $\{U_1, \ldots, U_N\}$ are 1×1 dimensional, then M is a complex number with magnitude $|M| \leqslant 1$.

Proof. Write $U_k = e^{i\theta_k}$, where $p(\theta)$ is the probability density function for each of the angles θ_k. From Theorem 1 we need to compute the mean value

$$M = \int_{-\pi}^{\pi} e^{i\theta} p(\theta) d\theta. \tag{9}$$

This is exactly the characteristic function of $p(\theta)$ evaluated at 1. The characteristic function is complex valued and has a bounded magnitude of 1, which is the desired result. ■

By Corollary 1 it can be concluded that for the (1×1)-dimensional case we can write $M = cU$, where $0 \leqslant c \leqslant 1$ and $U = e^{i\theta}$ has magnitude 1.

Next consider higher-dimensional matrices whose distribution is generated by a single parameter. In this case, for any Hermitian matrix T, which can be thought of as an infinitesimal generator from the $u(n)$ Lie algebra, we have

$$M = \int e^{i\theta T} p(\theta) d\theta. \tag{10}$$

We can make a change of variables in θ so that the distribution is changed to one that has mean zero

$$M = \int e^{i(\mu+\theta')T} p(\mu + \theta') d\theta' \tag{11}$$

$$= e^{i\mu T} \int e^{i\theta' T} \bar{p}(\theta') d\theta', \tag{12}$$

where $\bar{p}(\theta) = p(\mu + \theta)$ so that it has mean value zero. By expanding the matrix exponential this expression can be

written as

$$M = \sum_n \frac{(iT)^n}{n!} \int \theta^n p(\theta) d\theta, \tag{13}$$

which now relates to the moments of the underlying distribution in θ. Assuming $p(\theta)$ is a Gaussian distribution with mean zero and variance σ^2, then we can write

$$M = \sum_{n\in\text{even}} \frac{(iT)^n}{n!} (n-1)!!\sigma^n, \tag{14}$$

where $n!! = n(n-2)(n-4)\cdots$ is the double factorial. This series can be written back in the form of a matrix exponential, and by reintroducing the mean value we have

$$M = e^{i\mu T} e^{-(\sigma^2/2)T^2}. \tag{15}$$

If $T^2 = I$, which would be the case when choosing a Pauli matrix for T, then this expression would simplify to

$$M = U e^{-\sigma^2/2}, \tag{16}$$

where U is the unitary generated by the average parameter for $p(\theta)$. The decaying exponential for the magnitude depends only on the variation in the distribution of θ.

In the full parameter case, provided the target unitary U again commutes with all errors, a similar result can be found as discussed in Corollary 2.

Corollary 2. If $\{U_1,\ldots,U_N\}$ are random n-dimensional unitaries such that $U_k = U\exp\{i\sum_l \alpha_{kl}T_l\}$ with n^2 generators T_l that are all Hermitian and satisfy $T_l^2 = I$, the parameters α_{kl} distributed independently with probability density function (PDF) $p_l(\alpha_l)$ which are all Gaussian with mean zero and small (but possibly different) variances so that all U_k approximately commute with each other, then $M = cU$, where $0 < c < 1$ and U is a unitary matrix.

Proof. We will extend the proof of Corollary 1 to the n-dimensional case. From the independence of the distributed parameters, we can write a PDF for all parameters as $p(\alpha_1,\ldots,\alpha_{n^2}) = p_1(\alpha_1)\times\cdots\times p_{n^2}(\alpha_{n^2})$. The approximate mutual commutativity for this expansion means that

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} [\alpha_{kl}T_l, \alpha_{km}T_m] p_l(\alpha_l) p_m(\alpha_m) d\alpha_l d\alpha_m \approx 0 \quad \forall l,m, \tag{17}$$

that is to say, the $\alpha_{kl}T_l$ are all small with high probability. With this we can write M as

$$M = U\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\left\{i\sum_l \alpha_l T_l\right\} p_1(\alpha_1)\cdots p_{n^2}(\alpha_{n^2}) d\alpha_1 d\alpha_2\cdots d\alpha_{n^2} \tag{18}$$

$$\approx U\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\prod_l \exp\{i\alpha_l T_l\} p_1(\alpha_1)\cdots p_{n^2}(\alpha_{n^2}) d\alpha_1 d\alpha_2\cdots d\alpha_{n^2} \tag{19}$$

$$= U\int_{-\pi}^{\pi}\exp\{i\alpha_1 T_1\} p_1(\alpha_1) d\alpha_1 \int_{-\pi}^{\pi}\exp\{i\alpha_2 T_2\} p_2(\alpha_2) d\alpha_2\cdots\int_{-\pi}^{\pi}\exp\{i\alpha_{n^2} T_{n^2}\} p_{n^2}(\alpha_{n^2}) d\alpha_{n^2} \tag{20}$$

$$\approx U\prod_l e^{-\sigma_l^2 T_l^2/2}, \tag{21}$$

where σ_l^2 is the variance of p_l and the final approximation is assuming the distribution is small so that the bounds of the integration do not matter. Using the $T_l^2 = I$ requirement on the generators, the final product of exponentials can be identified with the value c and we have the desired result. ■

The requirement of $T_L^2 = I$ merely reflects a simplification where the generators are built from the Pauli matrices, which are the constructions we will focus on in this paper. If this is not the case, then it is possible to identify the Hermitian operator $\prod_l e^{-\sigma_l^2 T_l^2/2}$ as a state-dependent decay in the amplitude of the operator.

Finding expressions for the matrix M outside of the situations just outlined is an open problem. In the most general case, M is not proportional to a unitary matrix. Furthermore, it is not guaranteed that M will satisfy the conditions for a normal matrix and hence cannot be unitary diagonalized. So it is unclear if in general this postselected regime has any connection to unitary quantum evolution at all. Nevertheless, we will begin to examine situations which approach this domain through decompositions into single-parameter problems and using numerical computations.

III. IMPLEMENTATION

This section demonstrates how error averaging can be implemented for various prototypical optical systems. These examples also serve as a verification of the range of validity of the assumption of approximately commuting errors. It can also be noted that Eq. (7) can become the appropriate transformation for duality quantum computing by allowing the U_i to be arbitrary [11].

Constructions for the redundant encoding using the DFT implementation from the preceding section are useful mathematically but may be inconvenient to implement in practice. The transformation of Eq. (5) can also be achieved using an array of beam splitters as shown in Fig. 2. This beam-splitter array has the desirable property of being generated by a recursive pattern. As shown by the bounding rectangles in Fig. 2, the outer and inner layers share the same basic structure.

All linear networks can be generated by arranging networks of beam splitters and phase shifts [12]. Carolan *et al.* [13] have experimentally probed a linear network where all possible networks can be generated using controllable phase shifts and unvarying beam splitters. In their experimental implementation

they demonstrated the ability to implement many quantum logic gates and linear optical protocols with a high fidelity. Following this same methodology, one can generate controllable beam splitters using a Mach-Zehnder (MZ) interferometer consisting of a controllable phase shift in one arm and two fixed 50:50 beam splitters.

Within this type of architecture, the controllable phase shift is the key source of nonsystematic noise. Furthermore, redundantly encoding phase shifts are well characterized by the results presented above from Corollary 1. So we will focus on phase-shift-induced errors for the analysis of this section and the next. The model we will use assumes 50:50 beam splitters which are fixed and phase shifts that vary and are the source of all noise.

The noise in a controllable phase shift can be written as $e^{i(\theta+\delta)}$, where θ is a real number representing the phase shift to be applied and δ is a zero-mean random variable representing the error. For the identity operation $\theta = 0$. We will assume the distribution for δ to be Gaussian with variance v. For values of v that are comparable to π^2, the multivalued nature of phase shifts becomes important; however, initially we will focus on the limit where $v \ll \pi^2$.

The remainder of this section considers the above implementation of a tunable beam splitter as a MZ interferometer with the phase shift being error averaged. The error averaging will be performed using the concatenated beam-splitter network, hence $N = 2^n, n \in \mathbb{N}$, and all beam splitters used in this system will be fixed and with a splitting ratio of 50:50. We will analyze two key cases, the single-photon and two-photon performance. The former involves the classical wave nature of the probability distribution for a single photon. The latter includes Hong-Ou-Mandel [14] -style quantum interference.

A. One-photon input

The one-photon network considered here is shown in Fig. 3 both without any correction [Fig. 3(a)] and for the $N = 2$ case [Fig. 3(b)]. The input single-photon state is $|\phi\rangle = \hat{a}^\dagger|0\rangle$. After traversing the error-averaged network, the resulting unnormalized output state conditional on all encoded modes being vacuum is

$$|\psi\rangle = \left[\left(\frac{e^{i\theta}}{2}\left\{\frac{1}{N}\sum_{j=1}^{N} e^{i\delta_j}\right\} + \frac{1}{2}\right)\hat{a}^\dagger + \left(\frac{e^{i\theta}}{2}\left\{\frac{1}{N}\sum_{j=1}^{N} e^{i\delta_j}\right\} - \frac{1}{2}\right)\hat{b}^\dagger\right]|0\rangle\,, \qquad (22)$$

which is consistent with Theorem 1. Here $\theta_j = \theta + \delta_j$, with θ a constant and δ_j a random variable.

As linear networks conserve photon number and we have postselected the cases where energy exits via the redundant encoding modes, we know that the output state always contains one and only one photon. The probability that the photon is measured in a particular mode can therefore be equated to the average photon number in that mode. Using this we can calculate from the unnormalized state $|\psi\rangle$ the probability of observing the photon in the $\hat{a}$ and $\hat{b}$ modes without postselection to be

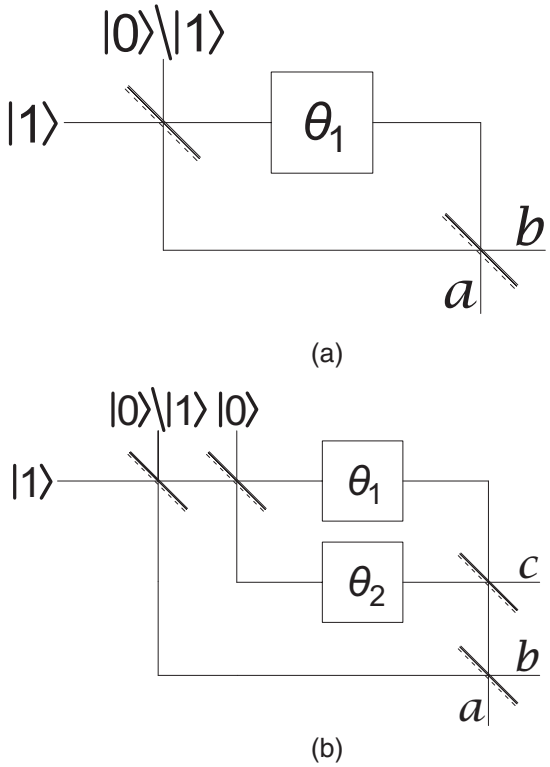

FIG. 3. Diagram of MZ-based tunable beam splitter: (a) uncorrected beam splitter implemented via a MZ interferometer and (b) such a beam splitter corrected by redundantly encoding the phase shift, here for $N = 2$; a and b label output modes and c labels an error detection mode. The input state shown is used for both one- and two-photon calculations. The phase-shift elements are marked with θ_j and are random variables.

$$\langle\psi|\,\hat{a}^\dagger\hat{a}\,|\psi\rangle \approx \cos^2(\theta/2) + \frac{v}{4N} - \frac{v\cos^2(\theta/2)}{2} \qquad (23)$$

and

$$\langle\psi|\hat{b}^\dagger\hat{b}|\psi\rangle \approx \sin^2(\theta/2) + \frac{v}{4N} - \frac{v\sin^2(\theta/2)}{2}, \qquad (24)$$

where we have taken a first-order expansion in the phase-shift variance v. The probability of success is the sum of the probabilities of the $\hat{a}$ mode and $\hat{b}$ mode. This is

$$P(\text{success}) = \langle\psi|\,\hat{a}^\dagger\hat{a}\,|\psi\rangle + \langle\psi|\,\hat{b}^\dagger\hat{b}\,|\psi\rangle \qquad (25)$$

$$\approx 1 + \frac{v}{2N} - \frac{v}{2}. \qquad (26)$$

In the large-N limit, this corresponds to the linear approximation of Eq. (16), where we can identify $P(\text{success}) = c$.

Without any noise, choosing $\theta = 0$ results in complete interference and the input single-photon state will be transferred to a single output. Any deviations from this are attributed to nonideal interferometer performance. In this case the probability of observing the output in the correct mode without postselection is

$$\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle \approx 1 - \frac{(2N-1)v}{4N}. \qquad (27)$$

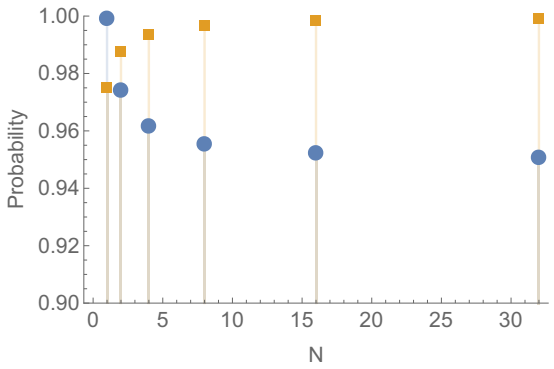

FIG. 4. Probability of a single photon being detected at the a output port as shown in Fig. 3 as a function of the redundant encoding size $N = 2^n$. Here the phase shifts are sampled from a distribution with mean value 0 and variance $v = 0.1$ rad^2. The blue circular values give the probability of success and the square orange values correspond to the probability of obtaining the correct result conditional on the photon not exiting the added redundant encoding modes, that is, with postselection. Equation (26) predicts an asymptote of 0.95 without postselection and Eq. (28) predicts an asymptote of 1 with postselection. The asymptotic behavior is consistent with the plotted data.

After postselection this becomes

$$\frac{\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle}{P(\text{success})} \approx 1 - \frac{v}{4N}. \tag{28}$$

Figure 4 shows how these two quantities scale with N. In particular, it can be seen that after postselection the likelihood of the photon exiting the interferometer in the correct mode can be made arbitrarily close to unity by increasing N. Also, while the probability of success decreases for increasing N it asymptotes to a constant value. This implies that as N increases, even though the total quantity of errors added to the system increases, the effects of the combined errors on the interferometer is less. This result is also not dependent on the value chosen for θ. Explicitly as the intended phase shift can be factored out in Eq. (22) similarly to the result shown in Eq. (16), the effects of errors and our error correction can be considered separately from the transformation being applied.

B. Two-photon input

The single-photon interference effects in linear networks can be explained using classical wave interference. Now we will consider two-photon interference to demonstrate the behavior of quantum interference when using the redundant encoding. As such, here $|1,1\rangle$ is used as the input state. Again, a diagram of the explicit setup with and without the redundant encoding can be seen in Fig. 3. For two photons the unnormalized output state for the a and b modes is

$$\begin{aligned}|\psi\rangle = \frac{1}{2}&\left\{1 + \frac{1}{N^2}\left(\sum_{j=1}^{N}\sum_{k=1}^{N} e^{i(\delta_j+\delta_k)}\right)\right\}|1,1\rangle \\ &+ \frac{\sqrt{2}}{4}\left\{\frac{1}{N^2}\left(\sum_{j=1}^{N}\sum_{k=1}^{N} e^{i(\delta_j+\delta_k)}\right) - 1\right\} \\ &\times (|2,0\rangle + |0,2\rangle),\end{aligned} \tag{29}$$

where we have chosen $\theta = 0$ when computing this state. This is done, as above, to simplify the form of the equations and does not change the effect of the redundant encoding on the errors. Because of this choice, the action of the interferometer on the input state should be the identity operation and hence $|1,1\rangle$ is the desired output state. Note that we could have chosen the input state to be $|2,0\rangle$, but this would not necessarily show any new behavior, just the single-photon results independently applied to the two input photons.

We can again write probabilities as expectation values of occupation number. Using the form of Eq. (29), the ideal output is achieved when

$$\langle\psi|\,\hat{a}^\dagger\hat{a}\hat{b}^\dagger\hat{b}\,|\psi\rangle = 1. \tag{30}$$

This expectation value for the state including the phase-shift noise is

$$\begin{aligned}\langle\psi|\hat{a}^\dagger\hat{a}\hat{b}^\dagger\hat{b}|\psi\rangle &= \left\langle\left|\frac{1}{2}\left\{1 + \frac{1}{N^2}\left(\sum_{j=1}^{N}\sum_{k=1}^{N} e^{i(\delta_j+\delta_k)}\right)\right\}\right|^2\right\rangle \\ &= \left\langle\frac{1}{4}\left[1 + \frac{2}{N^2}\left(\sum_{j=1}^{N}\sum_{k=1}^{N}\cos(\delta_j+\delta_k)\right)\right]\right\rangle + \frac{1}{4}\left\langle\frac{1}{N^4}\left(\sum_{j}^{N} e^{-2i\delta_j} + \sum_{j=1}^{N}\sum_{k\neq j}^{N} e^{-i(\delta_j+\delta_k)}\right)\right. \\ &\quad \left.\times\left(\sum_{l}^{N} e^{2i\delta_l} + \sum_{l=1}^{N}\sum_{m\neq l}^{N} e^{i(\delta_l+\delta_m)}\right)\right\rangle \\ &\approx 1 - v,\end{aligned} \tag{31}$$

where the approximation is assuming v small. Postselection will increase this to

$$\begin{aligned}P(\text{coincidence}) &= \frac{\langle\psi|\hat{a}^\dagger\hat{a}\hat{b}^\dagger\hat{b}|\psi\rangle}{\langle\psi|\hat{a}^\dagger\hat{a}\hat{b}^\dagger\hat{b}|\psi\rangle + 0.5\langle\psi|\hat{a}^\dagger\hat{a}^\dagger\hat{a}\hat{a}|\psi\rangle + 0.5\langle\psi|\hat{b}^\dagger\hat{b}^\dagger\hat{b}\hat{b}|\psi\rangle} \\ &\approx 1 - \frac{v}{2N},\end{aligned} \tag{32}$$

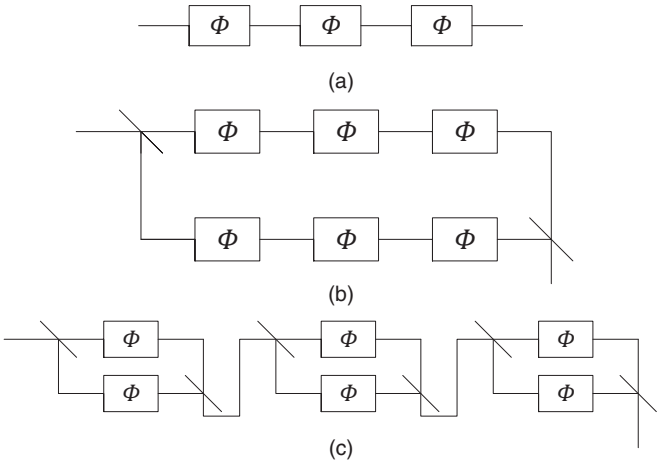

FIG. 5. Three methods of applying three phase shifts, each marked in series: (a) three phase shifts with no error averaging, (b) three phase shifts when averaging across the system, and (c) three phase shifts when averaging across each phase shifter individually. Averaging across the system will in general require far fewer encoding resources.

where $P(\text{coincidence})$ is the probability the photons exit modes a and b individually and the binomial approximation has been used to keep only variance terms to first order. Finally, the probability of success, which is the probability no photons exit the encoding modes, is

$$P(\text{success}) \approx 1 - v + \frac{v}{2N}. \tag{33}$$

Again, in the large-N limit of this equation the result matches the prediction (16). Also, the probability of success has a $\frac{1}{N}$ scaling which is the same as for the single-photon input case.

In this section we have demonstrated how redundantly encoding variable components can reduce the resulting variance within a system for the simple but highly important case of a single beam splitter. We have also shown that the results match what is expected from the couple of solved exact cases discussed in Sec. II. In the following section we will give some more complex examples to give clearer insight into how this redundant encoding might best be applied and its effect in the situations where the mathematical machinery introduced earlier is not easily solvable.

IV. COMPARISON BETWEEN AVERAGING TECHNIQUES

In this section we will study two different methods, which we will refer to as averaging at the end and averaging each step. To illustrate the two approaches we consider a simple system of phase shifters. Figure 5 shows schematically these two configurations as well as a baseline comparison. The system analyzed is applying a single-mode phase shift generated by M sequential phase shifters. Averaging across the entire system applies the M phases and redundantly encoding this N times [Fig. 5(b)]. The method of averaging each step involves a redundancy of N for each of the M applied phase shifts [Fig. 5(c)]. When averaging each component individually, significantly more encoding beam splitters are required, however we will show that this leads to more stability in the output state for larger errors. In the low error limit, however, these two methods yield equivalent results. Because of this, the difference is clearer when results are taken to the higher order and as such, in the following section, all approximations will be taken to the second order in the variance as opposed to the first order as done above.

Following an approach motivated by the preceding section, the applied phase shifters were placed in one arm of a MZ interferometer. The applied phase shift was chosen to have mean zero with a Gaussian random noise with a variance v. This choice allows for the errors here to be compared with those modeled in Secs. III A and III B. The probability of success is now defined as the photon-number expectation value evaluated at the end of the phase-applying systems, while the strength of the error is defined as the photon-number expectation value of the MZ interferometer system at the expected output given no photon exited any of the redundant encoding modes.

A. No averaging

Starting with the baseline comparison case where no error averaging is used [Fig. 5(a)], the output state for a single photon going through M phase shifters will be

$$|\psi\rangle = \left(\prod_{k=1}^{M} e^{i\delta_k}\right)|1\rangle. \tag{34}$$

As there is no path for the photon to exit the system, the probability of success is always 1.

To quantify the error the phase-applying system was then inserted into a MZ interferometer, giving a total output state $|\Psi\rangle$. As the mean phase shift is zero, the error is manifest in the photon expectation value at the correct output mode after postselection. As, however, the probability of success is 1, no postselection occurs here. The output state from the MZ interferometer is

$$|\Psi\rangle = \frac{1}{2}\left(\prod_{k=1}^{M} e^{i\delta_k} + 1\right)\hat{a}^\dagger|0\rangle + \left(\prod_{k=1}^{M} e^{i\delta_k} - 1\right)\hat{b}^\dagger|0\rangle. \tag{35}$$

The measure of the quantity of error which was used to compare the three situations chosen is the probability of observing the correct result conditional on the photon not being detected in an error mode, or $P(\text{correct})$. For no averaging this is

$$\begin{aligned} P(\text{correct}) &= \langle\Psi|\hat{n}_a|\Psi\rangle \\ &= \frac{1}{2}\langle 1 + \cos(\alpha)\rangle \\ &\approx 1 - \frac{Mv}{4} + \frac{M^2v^2}{16}, \end{aligned} \tag{36}$$

where $\alpha = \sum_{k=1}^{M}\delta_k$ and Gaussian statistics have been used to write higher-order moments in terms of the variance.

B. Averaging across the entire phase system

We now consider averaging across the whole system, as shown in Fig. 5(b). Proceeding as before, the state for a single photon after passing through M phase shifters in series which

is being averaged across N times will simply be

$$|\psi\rangle = \frac{1}{N}\sum_{j=1}^{N}\left(\prod_{k=1}^{M} e^{i\delta_{j,k}}\right)|1\rangle\,. \tag{37}$$

The probability of success is thus

$$\begin{aligned} P(\text{success}) &= \langle\psi|\psi\rangle \\ &\approx \left[1-\left(1-\frac{1}{N}\right)\left(Mv-\frac{1}{2}M^2v^2\right)\right]. \end{aligned} \tag{38}$$

This result is similar to the what was found in previous sections [see Eqs. (26) and (33)], with the probability of success asymptotically approaching some fixed value for large N.

To determine the size of the error, the phase-applying system was again inserted into one arm of a MZ interferometer, giving a total output state $|\Psi\rangle$. The error is then given by the photon expectation value in the correct output mode with postselection. The output state is

$$|\Psi\rangle = \frac{1}{2}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\prod_{k=1}^{M} e^{i\delta_{j,k}}\right)+1\right]\hat{a}^\dagger|0\rangle \tag{39}$$

$$+\left[\frac{1}{N}\sum_{j=1}^{N}\left(\prod_{k=1}^{M} e^{i\delta_{j,k}}\right)-1\right]\hat{b}^\dagger|0\rangle \tag{40}$$

So the photon-number expectation value for the expected output from the interferometer will be

$$\begin{aligned} &\langle\Psi|\,\hat{n}_a\,|\Psi\rangle \\ &\approx 1-\frac{1}{4}\left[Mv-\frac{M^2v^2}{4}+\left(1-\frac{1}{N}\right)\left(Mv-\frac{1}{2}M^2v^2\right)\right]. \end{aligned} \tag{41}$$

Similarly for the incorrect output port, the photon-number expectation value will be

$$\langle\Psi|\,\hat{n}_b\,|\Psi\rangle = \frac{1}{4}\left\langle 1+\langle\psi|\psi\rangle-\frac{2}{N}\sum_{j=1}^{N}\cos(\alpha_j)\right\rangle. \tag{42}$$

Therefore, our error measure, the conditional probability of observing the correct result, will now be

$$\begin{aligned} P(\text{correct}) \approx \Bigg\{1-\frac{1}{4}\Bigg[Mv-\frac{M^2v^2}{4}+\left(1-\frac{1}{N}\right) \\ \times\left(Mv-\frac{1}{2}M^2v^2\right)\Bigg]\Bigg\}\Bigg[1-\left(1-\frac{1}{N}\right) \\ \times\left(\frac{Mv}{2}-\frac{1}{4}M^2v^2\right)\Bigg]^{-1}. \end{aligned}$$

C. Averaging across each phase shifter individually

If each phase shifter is averaged individually, as shown in Fig. 5(c), then the state for a single photon after passing through the phase-applying system will be

$$|\psi\rangle = \frac{1}{N}\prod_{k=1}^{M}\left(\sum_{j=1}^{N} e^{i\delta_{j,k}}\right)|1\rangle\,. \tag{43}$$

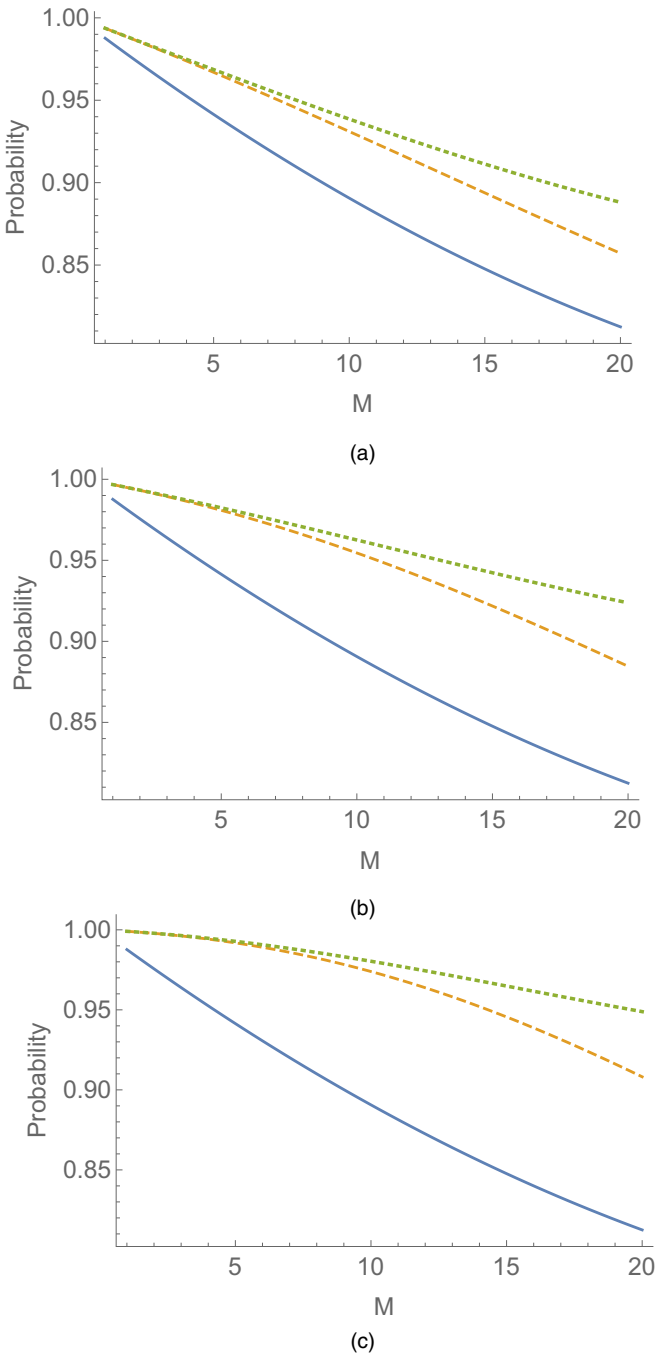

FIG. 6. Probability of obtaining the correct result as measured by the Mach-Zehnder interferometer setup as a function of the number of phase components M. Here a probability of 1 corresponds to no error and the smaller the probability the larger the error. The lower solid blue line represents the no error averaging applied result, the middle dashed orange line corresponds to the error when averaging across the entire system, and the upper dotted green line is the error when each component is averaged across individually. All three graphs were created with the variance of the error in a single phase shifter being 0.005 rad^2 and for (a) $N = 2$, (b) $N = 4$, and (c) $N = 16$.

Reproducing the above calculations with this state yields a probability of success of

$$P(\text{success}) \approx \left[1-\left(v-\frac{v^2}{2}\right)\left(1-\frac{1}{N}\right)\right]^M \tag{44}$$

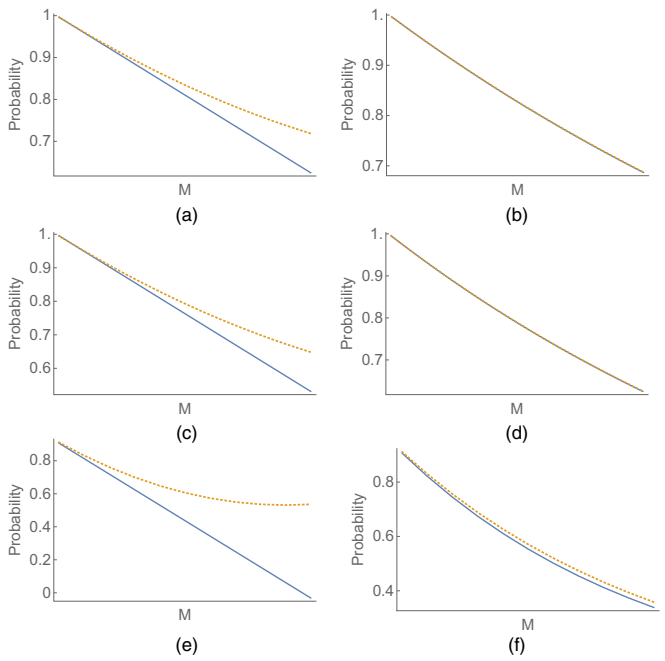

FIG. 7. Probability of success as a function of the number of phase components M for averaging (a), (c), and (e) across the system and (b), (d), and (f) across each component. The solid lower blue line is the first-order approximation and the dotted upper orange line is the second-order approximation of the analytical value. The individual phase-shifter variance is (a) and (b) 0.005 rad^2 with $N = 4$, (c) and (d) 0.005 rad^2 with $N = 16$, and (e) and (f) 0.1 rad^2 with $N = 16$.

and a conditional probability of observing the correct result of

$$P(\text{correct}) \approx \left\{\frac{3}{4} - \frac{Mv}{4} + \frac{M^2v^2}{16} + \frac{1}{4}\left[1 - \left(v - \frac{v^2}{2}\right)\right.\right.$$
$$\left.\times\left(1 - \frac{1}{N}\right)\right]^M\Bigg\}\left\{\frac{1}{2} + \frac{1}{2}\left[1 - \left(v - \frac{v^2}{2}\right)\right.\right.$$
$$\left.\left.\times\left(1 - \frac{1}{N}\right)\right]^M\right\}^{-1}.$$

Importantly, both for this case and when averaging each step, if only the first-order approximation is used and $M = 1$ then the error matches the error found in Sec. III A. However, we see that with the second-order terms included the two results diverge from one another. This can be seen most clearly in Fig. 7.

D. Summary of errors and probabilities

Figure 6 shows how the error, as measured by looking at expected photon-number values in the output port of a MZ interferometer, varies as the number of phase components increases as well as how the error changes with increasing error averaging N. The behavior as N increases is as expected with the error close to disappearing for low M, that is, a small number of phase shifters in series, and $N = 16$. Interestingly, a difference between the two error-averaging methods can be seen from $M \approx 6$ onward. This could be suggesting either an issue with the quality of the second-order approximations, as shown in Fig. 7, or that there is some more fundamental point at which there is a clear benefit to averaging each component individually. The next consideration is how the probability of success changes with M.

Figure 7 shows how the probability of success changes as the number of phase shifters in a series increases when averaging across the entire system as well as when averaging across each component individually. The effect of varying the amount of averaging is also shown for both the first- and second-order analytical solutions. The graphs in Figs. 7(a)–7(d) were plotted for a low value of the variance on the individual phase shifters. This was done so that the behavior when the first- and second-order approximations diverge can be clearly seen.

As the total number of components increases with both increasing M and N, the probability of success decreases; however, it does so at a decreasing rate, which is important for scaling to large systems. The two methods of error averaging also show very similar behavior in their overall trends, although the variation between the first- and second-order approximations in the two encoding methods diverges. This is suggestive of a manifestation of the Zeno effect, whereby continuously correcting produces less variation than doing the same amount of correction at the end. First- and second-order solutions in the averaging over the entire system case diverge very early when compared with those for averaging every step. Interestingly, it appears that the first-order analytical approximation is suitable when averaging each component individually even for larger or equivalently higher error systems. This can most clearly be seen in Fig. 7(e), where the first-order approximation diverges from the second-order approximation almost instantly, while in Fig. 7(f) the first- and second-order approximations both follow each other closely. It is again observed that as N increases, the probability of success goes down. This could also be suggesting that the variation in the statistical simulation is also reduced, implying that a greater amount of averaging reduces the variability in any given sample of the applied phase.

E. Statistical modeling of the applied phase

Given the variability in the higher-order terms for larger errors it can be concluded that, in general, all orders need to be considered to fully understand the behavior of the corrected systems. As this is intractable and to better understand the behavior of the three phase-applying systems, the total applied phase was modeled numerically using *Mathematica* with phase values chosen from a Gaussian random distribution with mean 0 and variance v. This corresponds to Eq. (9) with $\theta = \prod_i \theta_i$. This was repeated 5000 times and the results are shown in Figs. 8 and 9. This again shows a difference between averaging across the entire system and averaging at each step. The variability of the total applied phase is smaller when each phase shifter is corrected individually, an indication that averaging each step is more effective. By comparing Fig. 8 with Fig. 6 at $M = 15$ we can infer that the difference between the two error-correction methods seen in Fig. 6 is not entirely due to the quality of the approximations used in each case.

The variance of the applied phase was estimated based on the statistical simulation of the total applied phase. Figure 10 shows this variance as a function of M, the number of phase shifters in a series. Given that the individual applied phases

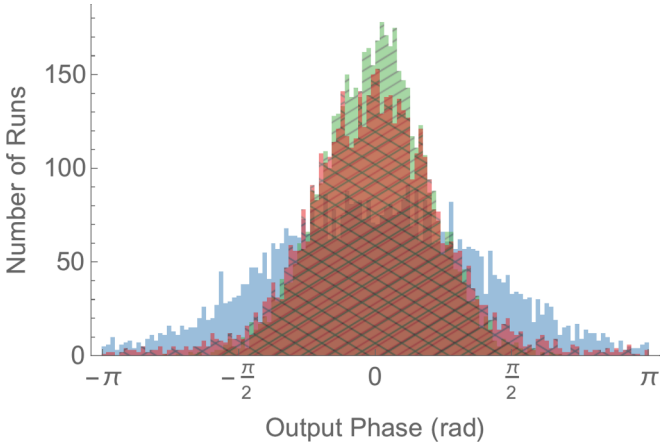

FIG. 8. Histogram of the total applied phases over 5000 runs for no averaging (solid blue region), averaging across the entire system (red cross-hatched region), and averaging each phase shifter individually (green striped region). Each individual phase shifter has a variance of 0.1 rad^2 and each system has 15 phase shifters in series. The two error-averaged circuits are each averaged four times.

are uncorrelated, they are expected to simply add such that the variance without any error averaging is expected to be

$$\text{total variance} = vM, \tag{45}$$

where M is the number of phase shifters and v is the variance in the individual phase shifter. The total variance when error averaging is similarly expected to be

$$\text{total variance} = \frac{vM}{N}, \tag{46}$$

where N is the number of times the system is averaged; again $N = 1$ implies no averaging. As the phase is an angle with a finite range, this behavior cannot hold for arbitrarily large vM. A completely random phase θ is still limited by the possible range of values, in our case chosen to be $-\pi < \theta \leqslant \pi$. If the value of θ is indeed completely random then one will expect a uniform probability distribution of $P(\theta) = \frac{1}{2\pi}$. This then implies that the maximum variance will be given by

$$\begin{aligned}\text{maximum variance} &= \int_{-\pi}^{\pi} \theta^2 P(\theta) d\theta \\ &= \frac{\pi^2}{3}.\end{aligned} \tag{47}$$

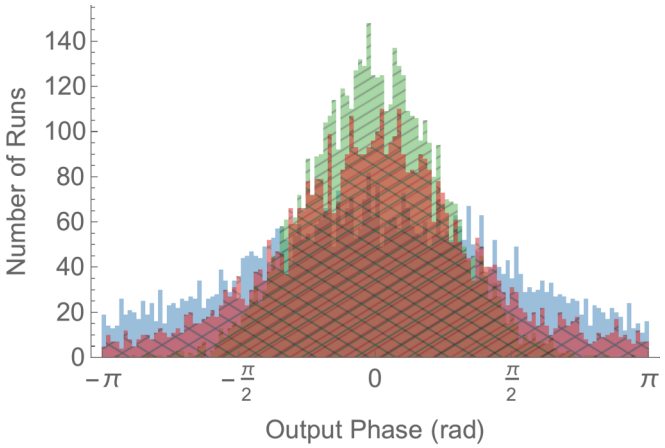

FIG. 9. Histogram of the total applied phase over 5000 runs for no averaging (solid blue region), averaging across the entire system (red cross-hatched region), and averaging each phase shifter individually (green striped region). Each individual phase shifter has a variance of 0.3 rad^2 and each system has eight phase shifters in series. The two error-averaged circuits are each averaged four times.

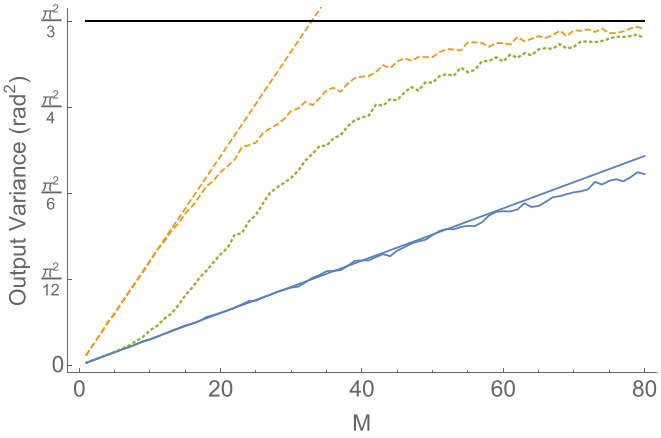

FIG. 10. Variance in the total applied phase without error averaging (upper dashed orange line), when averaging across the entire system (middle dotted green line) and when averaging each component individually (lower solid blue line), all plotted as a function of the number of phase shifters in series M. The variance of a single phase shifter is 0.1 rad^2 and the two error-averaged systems are averaged four times. The predicted linear variance without any averaging (dashed orange line) and with averaging (solid blue line) is also shown. These lines ignore the fact that the variance is actually the angular variance and so has some maximum allowable value given by Eq. (47), which is also shown as a horizontal black line.

Figure 10 shows that the two methods of error correction do indeed initially have the same effect. However, the averaging across the system method departs from this linear regime from approximately $M = 6$ after which it follows the general form of applying no correction. This suggests that there is some limit to the total variation in a system error averaging can handle. This is not unexpected due to the limited domain for a phase shift or beam splitter ratio. The fact that averaging across the entire system mirrors the no-averaging trend suggests that the positive effects of error averaging completely disappear in this regime. Averaging each step, however, does not appear to fall out of the linear regime. The lower slope at higher total output variance can be attributed to the variance approaching the maximum possible variance. To determine if and when the averaging each step method of error correction fails, this process was repeated as a function of the variance in a single beam splitter. The total variance was determined from 50 000 data points for each value of v. Figure 11 demonstrated that again the two error-correction methods initially are equivalent. Once more averaging across the system departs from the linear regime and now we can clearly see that so too does averaging each step.

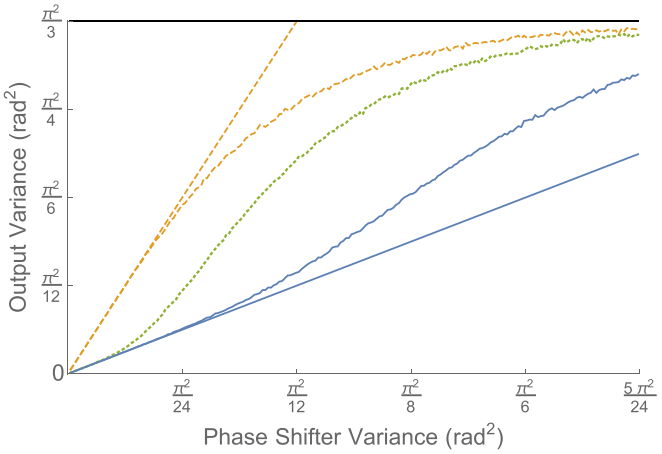

FIG. 11. Variance in the total applied phase without error averaging (upper dashed orange line), when averaging across the entire system (middle dotted green line), and when averaging each component individually (lower solid blue line), all plotted as a function of the variance in each individual phase shifter. Each system applied four phase shifters in series, that is, $M = 4$, and the two error-averaged systems are averaged four times, that is, $N = 4$. The predicted variance without any averaging (dashed orange line) and with averaging (solid blue line) is also shown along with the maximum allowable variance, which is shown as a horizontal black line.

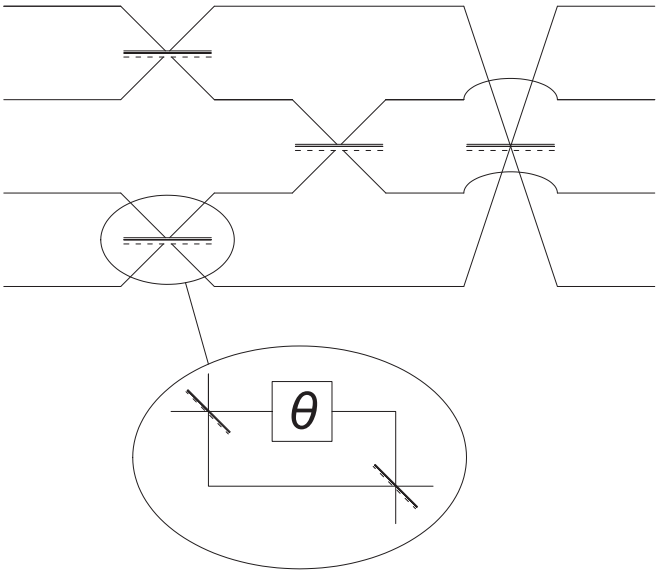

FIG. 12. Diagram of the four-mode linear optical network which forms the basis of the four-mode setups.

This is suggestive of the existence of some threshold for the amount of error in a system before error averaging fails to be beneficial. A single phase shifter with variance v_1 is effectively equivalent to m phase shifters with individual variance $v_2 = \frac{v_1}{m}$ if the entire system is being averaged across. This allows the phase error threshold to be estimated at about 0.5 rad^2 when averaging across each element and $\frac{0.5}{m}$ rad^2 when averaging across a system of m phase shifters. Explicitly, this suggests a phase variance threshold of 0.5 rad^2 within the corrected unitary. If each individual beam splitter has a variance of 0.1 rad^2, averaging across the system would be expected to be in the linear regime when $M \leqslant 5$, which is precisely what is seen in Fig. 10. These two thresholds obviously do not apply to a general error; however, it is hoped that further study might reveal the values for such a threshold.

It can now be concluded that some hybrid method of error averaging would be most suitable in general. The entire system would need to be broken into x smaller systems, which are independently averaged across. The specific value of x would be such that the number of components in the system m is maximized while the total error within each subsystem is kept below the appropriate threshold.

V. FOUR-MODE IMPLEMENTATION COMPARISON

To gain a better understanding of how useful this method of error averaging actually is, a more complex system was also investigated along with both methods of redundant error correction, specifically, a four-mode system with four beam splitters, each implemented as above, that is, each being its own MZ interferometer as shown in Fig. 12. Three different input states were chosen: a single photon input in the top mode and the vacuum state at all other modes ($|1,0,0,0\rangle$); two photons, both in the top mode ($|2,0,0,0\rangle$); and two photons spread across the top two modes ($|1,1,0,0\rangle$). For simplicity, the system was chosen to target the identity and error reduction was then applied using both implementations, that is, by averaging each beam splitter as done in Sec. III A and by concatenating the entire system in an interferometer as shown in Fig. 13. All results were found using *Mathematica* to sample from the appropriate transformation matrix representing the system and then compute the second-order approximation of the photon-number expectation values and probabilities for $N = 1, 2$, and 4. All results can be found in the Appendix.

These results show that the two correction methods produced equivalent results under the approximations used. The $1/N$ scaling in error after postselection, as seen above, was also observed, suggesting that this pattern holds for higher numbers of modes and an arbitrary system as expected from Theorem 1. The analysis that follows is based on the trends observed from these simulations.

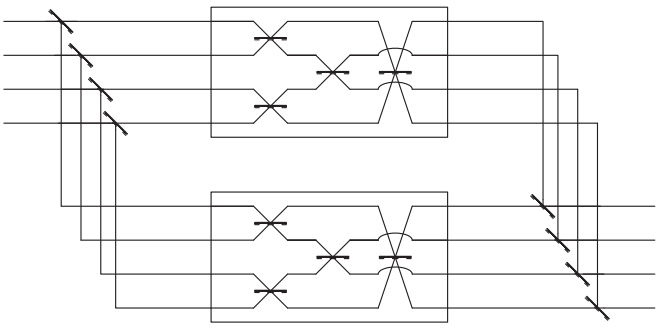

FIG. 13. Diagram of the four-mode linear optical network averaged across the system once. A two-mode system averaged in this fashion could be considered as an error-averaged dual-rail single-qubit unitary transformation.

Generalizing example results

Starting with no error reduction and given the correct output state $|\psi\rangle$, the following sequence can be defined. First defining the probability of obtaining the correct result when $N = 1$ takes the form

$$P_1(\text{correct}) = 1 - \frac{a_1}{b_1}v. \tag{48}$$

The probability of an incorrect state is therefore

$$P_1(\text{wrong}) = \frac{a_1}{b_1}v. \tag{49}$$

At this stage there are no error ports and so $P_i(\text{wrong}) + P_i(\text{correct})$ represents the probability of success as previously defined. Note that the subscript indexes the corresponding number of averaging rounds. Explicitly $i + 1$ corresponds to a system with twice as much averaging as i. So, using the same observed form for the probability, averaging once changes these values to

$$\begin{aligned} P_2(\text{correct}) &= 1 - \frac{2a_1 + 1}{2b_1}v \\ &\equiv 1 - \frac{a_2}{b_2}v, \end{aligned} \tag{50}$$

$$\begin{aligned} P_2(\text{wrong}) &= \frac{a_1}{2b_1}v \\ &\equiv \frac{a_2}{b_2}v, \end{aligned} \tag{51}$$

which can be further iterated. In general,

$$\begin{aligned} P_n(\text{correct}) &= 1 - \frac{2a_{n-1} + 1}{2b_{n-1}}v \\ &= 1 - \frac{2^{n-1}a_1 + (2^{n-1} - 1)}{2^{n-1}b_1}v, \end{aligned} \tag{52}$$

$$\begin{aligned} P_n(\text{wrong}) &= \frac{a_{n-1}}{2b_{n-1}}v \\ &= \frac{a_1}{2^{n-1}b_1}v. \end{aligned} \tag{53}$$

The probability of obtaining the correct state with postselection will then be

$$P_n(\text{correct}|\text{postselection}) = \left(1 - \frac{a_1}{2^{n-1}b_1}v\right) \xrightarrow[n\to\infty]{} 1. \tag{54}$$

The probability of success is

$$P_n(\text{success}) = 1 - \frac{(2^{n-1} - 1)(a_1 + 1)}{2^{n-1}b_1}v.$$

Hence we get an asymptotic expression for the probability of success to be

$$\lim_{n\to\infty} P_n(\text{success}) = 1 - \left(\frac{a_1 + 1}{b_1}\right)v. \tag{55}$$

The result can be understood to be the first-order approximation to Eq. (21). The $\frac{a_1+1}{b_1}$ coefficient does not quite match what might be expected from Eq. (21); however, one might only expect qualitative agreement, as there is not any clear isomorphic map between the parameters in the system and the error coefficients of the Lie algebra generators.

This result hints at the self-correcting nature of error averaging. By considering the inner corrected system with error-laden beam splitters as the initial step in the sequence, then each further step will be averaging across both the fixed beam splitters and the original error-laden system. This could allow some of the beam splitters to be corrected, making the base assumptions on the quality of the fixed beam splitters less restrictive. This is highlighted in Fig. 2 where, depending on which components are considered to be the system, it is averaged eight, four, or two times.

VI. DISCUSSION

The analysis presented in Secs. III–V concentrated on implementing a single-mode phase shift either on its own or as part of a Mach-Zehnder interferometer implementing a beam-splitter-type transformation. These could then be further used to build up higher-dimensional unitary transformations using any particular choice of decomposition, for which a specific decomposition was analyzed in Sec. V.

On the other hand, one may want to redundantly encode an entire unitary rather than just the phases defining the internal parameters. In this case we can use Eq. (21) and for simplicity consider the specific case with $T_l^2 = I$,

$$M = U\prod_l e^{-\sigma_l^2/2} = Ue^{-n^2\sigma^2/2}, \tag{56}$$

where the n^2 term appears as the product over all n^2 generators. This would result in an effective operator transformation for a k photon state as

$$\frac{a^{\dagger k}_i}{\sqrt{k!}} \to e^{-kn^2\sigma^2/2}\frac{1}{\sqrt{k!}}\left(\sum_j U_{ij}a^\dagger{}_i\right)^k. \tag{57}$$

The coefficient here represents a reduction in the amplitude should this transformation be applied to a state, and hence represents the probability of success. To achieve an $O(1)$ probability of success, then the operator noise must obey $\sigma = O(k^{-1/2})$ and $\sigma = O(n^{-1})$ as $k,n \to \infty$. These results are dependent on the assumptions and the desired performance, in terms of probability of success, will depend on the specific application. However, it must be kept in mind that no error correction has been performed, yet it is still possible to achieve a constant success probability with a reasonable scaling of the noise with respect to the network size.

Within the constructions presented here some optical elements utilized have been assumed ideal. In particular, the encoding beam splitters were assumed to have exactly 50:50 reflectivities. A more general consideration is that of the fault tolerance of this encoding, that is, if the ideal elements can be error corrected while maintaining the error-correcting power of the scheme. In this paper we have focused on merely the error-correcting power of different arrangements of phase shifts and how it varies across two choices of decomposition. However, there will be many and varied choices about how to implement fault-tolerant constructions with some better than others, in much the same way as is applicable for a discrete system in quantum computing implementations. Nevertheless, the fact that, under an approximation of small errors, the encoding

TABLE I. Output probabilities for various levels of correcting the individual beam splitters in a four-mode setup given an input of the state $|1,0,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging beam splitters once ($N = 2$)	Averaging beam splitters twice ($N = 4$)
$\|1,0,0,0\rangle$	$1 - \frac{v}{2}$	$1 - \frac{3v}{4}$	$1 - \frac{7v}{8}$
$\|0,1,0,0\rangle$	$\frac{v}{4}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|0,0,1,0\rangle$	0	0	0
$\|0,0,0,1\rangle$	$\frac{v}{4}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|1,0,0,0\rangle$ with postselection	$1 - \frac{v}{2}$	$1 - \frac{v}{4}$	$1 - \frac{v}{8}$

TABLE II. Output probabilities for various levels of correcting across the system in a four-mode setup given an input of the state $|1,0,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging across the system once ($N = 2$)	Averaging across the system twice ($N = 4$)
$\|1,0,0,0\rangle$	$1 - \frac{v}{2}$	$1 - \frac{3v}{4}$	$1 - \frac{7v}{8}$
$\|0,1,0,0\rangle$	$\frac{v}{4}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|0,0,1,0\rangle$	0	0	0
$\|0,0,0,1\rangle$	$\frac{v}{4}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|1,0,0,0\rangle$ with postselection	$1 - \frac{v}{2}$	$1 - \frac{v}{4}$	$1 - \frac{v}{8}$

tends to the ideal operation in the limit of large encoding sizes, it would be reasonable to expect that fault-tolerant constructions exist.

VII. COMPARISON WITH CONVENTIONAL QUANTUM ERROR CORRECTION

It is insightful to qualitatively discuss the parallels between conventional qubit quantum error detection and correction techniques, and our error averaging technique. The simplest code to see this parallel is by considering the three-qubit code, which is able to detect and correct at most a single physical bit-flip error on a threefold redundantly encoded logical qubit. In the three-qubit code the logical qubit is encoded via Greenberger-Horne-Zeilinger (GHZ) -type entanglement across the three physical qubits using two maximally entangling CNOT gates. Specifically, encoding implements the redundant mapping $\alpha\,|0\rangle + \beta\,|1\rangle \rightarrow \alpha\,|000\rangle + \beta\,|111\rangle$ in the logical basis. In our scheme, on the other hand, the redundant encoding takes the form of W-like entanglement, implemented via an optical fanout operation, where a single excitation in a single mode is mapped to a superposition of a single excitation across multiple modes. Specifically, the encoding is of the form $\hat{a}_1^\dagger \rightarrow \frac{1}{\sqrt{N}}(\hat{b}_1^\dagger + \cdots + \hat{b}_N^\dagger)$. This is qualitatively very distinct from the previous GHZ-type encoding, since GHZ states are maximally entangled states, whereas W states are not. Unlike GHZ states, which collapse onto a perfectly mixed $N - 1$ qubit state upon loss of just a single qubit, the loss of a single mode from a W state preserves most entanglement for large N. This leads us to speculate that this property of W states enables much of the structure of encoded states to be preserved upon localized errors. Indeed, for $N \gg 1$ we anticipate that the failure of a relatively small subset of the redundant operations will have little impact on the integrity of the entire encoded state, owing to this unique property of the structure of loss in W states.

Like conventional quantum error correction, we observed error threshold behavior in our analysis. That is, we are only able to improve the fidelity of a state if its initial fidelity is above an error-correction threshold. Below this threshold the error-correction technique fails to improve the state. Indeed, nonzero thresholds must necessarily apply so as not to violate the quantum no-cloning theorem.

We observed that a simple form of circuit concatenation, whereby the protocol is recursively embedded within itself to construct larger nested codes, enables higher degrees of error correction, asymptoting to some maximum. This is congruent with conventional codes, where code concatenation asymptotically improves error correction at the expense of increased physical resources to mediate the more complex encoding.

In our scheme the postselection upon detecting no photons in the designated failure modes is equivalent to syndrome measurement in traditional qubit codes. Successful postselection effectively projects the encoded state back into the code space, whereas failure heralds an unsuccessful syndrome extraction, thereby mapping the unitary error to a located loss error. The probability of detecting no photons in the failure modes can be associated with the error-detection probability in traditional codes and the respective conditional probability of measuring the correct output state with the error-correction probability.

An interesting open question is whether the structure of the redundant encoding we utilize in our protocol may be translated to other physical architectures or conventional qubit settings or rather whether it is very specific to photonic linear optics.

VIII. CONCLUSION

We have shown how, given multiple noisy copies of a linear optical unitary network, error averaging can be use to implement a transformation that tends towards the average with reduced variance at the cost of the probability of success. After postselection, error averaging forms a rudimentary error-correction protocol by filtering the noise from the redundant copies of the unitary network. For this to form a true error-correction protocol it will be necessary to introduce some sort of loss correction. The losses which will need to be corrected are unique, however, in that they are heralded and located, potentially simplifying the problem enormously. The variance in the transformations have been shown to scale as $\frac{1}{N}$, where N represents the number of redundant copies of the network. We have provided the mathematical basis necessary to determine the effect of error averaging on an arbitrary linear unitary and with fully characterized solutions for arbitrary single-parameter noise and multiple-parameter small Gaussian noise. We have also analytically determined the photon-number expectation values for two mode systems with both one- and two-photon inputs, numerically simulated the output expectation values in four-mode systems for both

TABLE III. Output probabilities for various levels of correcting the individual beam splitters in a four-mode setup given an input of the state $|2,0,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging beam splitters once ($N = 2$)
$\|2,0,0,0\rangle$	$1 - v$	$1 - \frac{3v}{2}$
$\|0,2,0,0\rangle$	0	0
$\|0,0,2,0\rangle$	0	0
$\|0,0,0,2\rangle$	0	0
$\|1,1,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$
$\|1,0,1,0\rangle$	0	0
$\|1,0,0,1\rangle$	$\frac{v}{2}$	$\frac{v}{4}$
$\|0,1,1,0\rangle$	0	0
$\|0,1,0,1\rangle$	0	0
$\|0,0,1,1\rangle$	0	0
$\|2,0,0,0\rangle$ with postselection	$1 - v$	$1 - \frac{v}{2}$

TABLE V. Output probabilities for various levels of correcting the individual beam splitters in a four-mode setup given an input of the state $|1,1,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging beam splitters once ($N = 2$)	Averaging beam splitters twice ($N = 4$)
$\|2,0,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,2,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,0,2,0\rangle$	0	0	0
$\|0,0,0,2\rangle$	0	0	0
$\|1,1,0,0\rangle$	$1 - \frac{3v}{2}$	$1 - \frac{7v}{4}$	$1 - \frac{15v}{8}$
$\|1,0,1,0\rangle$	$\frac{v}{2}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|1,0,0,1\rangle$	0	0	0
$\|0,1,1,0\rangle$	0	0	0
$\|0,1,0,1\rangle$	$\frac{v}{2}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|0,0,1,1\rangle$	0	0	0
$\|1,1,0,0\rangle$ with postselection	$1 - \frac{3v}{2}$	$1 - \frac{3v}{4}$	$1 - \frac{3v}{8}$

one- and two-photon inputs, and numerically simulated the variance for different arrangements of phase shifters.

Two methods of error averaging for phase shifts have been presented which appear to have similar effects under certain conditions. In particular, averaging after sequentially applying phases has the same behavior as averaging each phase provided the errors are small. This behavior is conjectured to be explained by considering the errors as approximately commuting.

ACKNOWLEDGMENTS

We thank Michael Bremner for motivating discussions. This research was funded by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (Project No. CE110001027). P.P.R. is funded by an ARC Future Fellowship (Project No. FT160100397).

APPENDIX: FOUR-MODE NUMERICAL RESULTS

This appendix contains all simulation results for the four-mode system discussed in Sec. V. All results are based on a second-order Taylor expansion with $v \ll 1$.

Tables I and II show the output probabilities and correct result probability with postselection for the single-photon input state $|1,0,0,0\rangle$. These show that, at least for a single photon the two correction methods are equivalent. We also see the halving of errors as seen in Secs. IV B and IV D, suggesting that this pattern may hold for a single photon with an arbitrary number of modes.

Tables III and IV show the output probabilities and correct result probability with postselection for the single-photon input state $|2,0,0,0\rangle$. What we see is, unsurprisingly, much the same as in the single-photon case with a heightened susceptibility to the error. This includes the halving pattern,

TABLE IV. Output probabilities for various levels of correcting across the system in a four-mode setup given an input of the state $|2,0,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging beam splitters once ($N = 2$)	Averaging beam splitters twice ($N = 4$)
$\|2,0,0,0\rangle$	$1 - v$	$1 - \frac{3v}{2}$	$1 - \frac{7v}{4}$
$\|0,2,0,0\rangle$	0	0	0
$\|0,0,2,0\rangle$	0	0	0
$\|0,0,0,2\rangle$	0	0	0
$\|1,1,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|1,0,1,0\rangle$	0	0	0
$\|1,0,0,1\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,1,1,0\rangle$	0	0	0
$\|0,1,0,1\rangle$	0	0	0
$\|0,0,1,1\rangle$	0	0	0
$\|2,0,0,0\rangle$ with postselection	$1 - v$	$1 - \frac{v}{2}$	$1 - \frac{v}{4}$

TABLE VI. Output probabilities for various levels of correcting across the system in a four-mode setup given an input of the state $|1,1,0,0\rangle$ where v is the variance of the phase error.

Output state	No error reduction	Averaging beam splitters once ($N = 2$)	Averaging beam splitters twice ($N = 4$)
$\|2,0,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,2,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,0,2,0\rangle$	0	0	0
$\|0,0,0,2\rangle$	0	0	0
$\|1,1,0,0\rangle$	$1 - \frac{3v}{2}$	$1 - \frac{7v}{4}$	$1 - \frac{15v}{8}$
$\|1,0,1,0\rangle$	$\frac{v}{2}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|1,0,0,1\rangle$	0	0	0
$\|0,1,1,0\rangle$	0	0	0
$\|0,1,0,1\rangle$	$\frac{v}{2}$	$\frac{v}{8}$	$\frac{v}{16}$
$\|0,0,1,1\rangle$	0	0	0
$\|1,1,0,0\rangle$ with postselection	$1 - \frac{3v}{2}$	$1 - \frac{3v}{4}$	$1 - \frac{3v}{8}$

however, this is expected, as adding two photons in the same mode will not necessarily lead to new interference effects being observed.

Tables V and VI show the output probabilities and correct result probability with postselection for the single-photon input state $|1,1,0,0\rangle$. It can once more be seen that the two methods of error correction appear to be equivalent. Now there is an underlying pattern clearly forming which appears to hold for arbitrary one- and two-photon inputs. This is important as it allows us to conclude about when it is most useful to use each type of correction. It also allows a prediction of the error models for applications of error averaging, as discussed above.

[1] E. Knill, R. Laflamme, and G. J. Milburn, Nature (London) **409**, 46 (2001).

[2] R. Raussendorf and H. J. Briegel, Phys. Rev. Lett. **86**, 5188 (2001).

[3] T. C. Ralph and G. J. Pryde, Prog. Opt. **54**, 209 (2010).

[4] S. Aaronson and A. Arkhipov, *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing* (ACM, New York, 2011), pp. 333–342.

[5] Y. Aharonov, L. Davidovich, and N. Zagury, Phys. Rev. A **48**, 1687 (1993).

[6] M. A. Broome, A. Fedrizzi, B. P. Lanyon, I. Kassal, A. Aspuru-Guzik, and A. G. White, Phys. Rev. Lett. **104**, 153602 (2010).

[7] A. Peruzzo, M. Lobino, J. C. F. Matthews, N. Matsuda, A. Politi, K. Poulios, X.-Q. Zhou, Y. Lahini, N. Ismail, K. Worhoff, Y. Bromberg, Y. Silberberg, M. G. Thompson, and J. L. OBrien, Science **329**, 1500 (2010).

[8] A. Schreiber, A. Gabris, P. P. Rohde, K. Laiho, M. Stefanak, V. Potocek, C. Hamilton, I. Jex, and C. Silberhorn, Science **336**, 55 (2012).

[9] A. Schreiber, K. N. Cassemiro, V. Potoček, A. Gábris, P. J. Mosley, E. Andersson, I. Jex, and C. Silberhorn, Phys. Rev. Lett. **104**, 050502 (2010).

[10] M. G. Thompson, A. Politi, J. C. Matthews, and J. L. O'Brien, IET Circ. Device. Syst. **5**, 94 (2011).

[11] G. L. Long, Int. J. Theor. Phys. **50**, 1305 (2011).

[12] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. **73**, 58 (1994).

[13] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, *et al.*, Science **349**, 711 (2015).

[14] C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).