Error Correction Through Error Averaging

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

[1] *Centre for Quantum Computation and Communication Technology (CQC2T), The School of Mathematics and Physics, The University of Queensland, Australia.*

[2] *Centre for Quantum Computing & Intelligent Systems (QCIS), University of Technology Sydney, Australia.*

(Dated: June 14, 2017)

Given the current drive to witness the supremacy of quantum computing it is imperative to design implementable error correction protocols. In this paper we propose and investigate Error Averaging, a novel method of error detection and reduction. Error averaging improves on existing methods of error detection and correction by not relying on ancillary photons or complicated control and correction circuits [1]. Error averaging can be applied to any quantum computing architectures which satisfies two requirements, both of which are satisfied automatically by Linear Optical Quantum Computing (LOQC). Error averaging will be introduced first in a general context followed by a series of proof of principle examples. Two methods of Error Averaging are then compared to determine the most effective manner of implementation and probe the related error thresholds. It is hoped that by employing existing methods of protecting against losses that this system could be used as a method of error correction in its own right.

I. INTRODUCTION

Quantum technology and in particular Quantum computing is a very promising area of research with the potential to lead to some of the most significant technological advances in recent times. There are, however, still many hurdles before universal quantum computing is achieved. The focus of this paper will be on the hurdle of error correction. The Error Averaging protocol requires the input qubits to be placed in a spatial superposition and restricts the effect that any applied unitary has on the vacuum. Both of these requirements are satified by LOQC and as such this will form the basis for most of this paper.

Error Averaging in its current state forms a rudimentary form of error correction which currently acts to distil errors form the system in such a way that they can be post selected away. Figure 1 shows how Error Averaging by itself lowers the probability of success, but increases the probability of obtaining the correct result after post selecting on no errors being detected. The variance of the errors in the system have been shown to scale as $\frac{1}{N}$ where N represents the amount of correction or equivalently the number of copies of the system.

The next section introduces the details and results for a general set-up. This is done to both explain how Error Averaging is implemented and demonstrate its effect on an output state. Section III includes numerous proof of principle examples which serve to highlight the effects of Error Averaging with a focus on the scaling for the probability of success. LOQC will serve as the architecture for all these examples. The feasibility of Error Averaging will be considered including a detailed discussion of some of the underlying assumptions made in the examples below can be found in section VI followed by some concluding remarks.

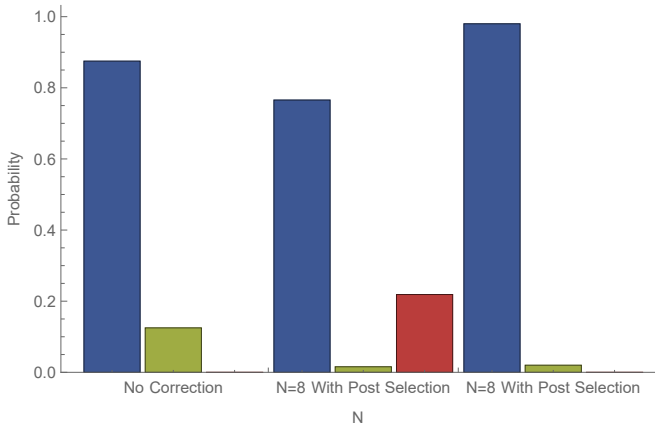

FIG. 1: Comparison of output probability distributions with and without error averaging. This also demonstrates the effect post selection has on the output distribution. The blue bars represent the probability of observing the photon in the correct output mode, green corresponds to observing the photon in the incorrect output mode and red 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 shifters variance $v = 0.5$ rad^2. This high variance is chosen.

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 a 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.

Consider a linear network whose elements are those of a Discrete Fourier Transform (DFT). That is, we have an 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 of the modes has been used to simplify this expression. 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 the N copies of a target unitary U. By this we mean that there is some variation between the copies but the intention was 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 as $U_1, U_2, \ldots, U_N$.

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 initialised in the vacuum state and post-select on the cases where no photons are present in the output of the redundant modes. This means that we only need 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 post-selected 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 analysing the scenaiors that arise from the multitude of choices for U_k that form expression.

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 matricies $\{U_1, U_2, \ldots, U_N\}$ 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 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. □

The question now is what forms can the mean average matrix M 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 $|M| \leq 1$.*

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

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

This is exactly the characteristic function of $p(\theta)$ evaluated at 1. The characteristic function is complex valued and has 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 $0 \leq c \leq 1$ 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 is 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. If $p(\theta)$ were a Gaussian distribution with mean zero and variance σ^2 then we can write

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

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

$$M = e^{i\mu T} e^{-\frac{\sigma^2}{2} T^2} \tag{15}$$

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

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

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 PDF $p_l(\alpha_l)$ which are all Gaussian with mean zero and small (but possibly different) variances so that they 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 independance of the distributed parameters, we can write a PDF for all parameters as $p(\alpha_1, \ldots, \alpha_{n^2}) = p_1(\alpha_1) \times \ldots \times p_{n^2}(\alpha_{n^2})$. The approximate mutual commutivity for this expansion means

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

or in other words, that the $\alpha_{kl} T_l$ are all small with high probability. With this we can write M as

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

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

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

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

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, we have the desired result. □

The requirement of $T_L^2 = I$ merely reflects a simplification where the generators are built from the Pauli matricies which is the constructions we will focus on in this paper. If this is not the case then it is possible to identify the hermitian operator $\prod_l e^{-\frac{\sigma_l^2 T_l^2}{2}}$ as a state dependant 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 be proportional to a unitary matrix. Furthermore, is not guaranteed that M satisfy the conditions for a normal matrix and hence cannot be unitary diagonalised. So it is unclear if this post-selected 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 example systems. These examples also serve as a verification of the range of validity of approximately commuting errors assumption. It can also be noted that Equation (7) can become the appropriate transformation for duality quantum computing by allowing the U_i to be arbitrary[2].

Constructions for the redundant encoding using the DFT implementation from the previous section use 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 [3]. Carolan et. al. [4] 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 non-systematic noise. Furthermore, redundantly encoding phase shifts are well characterised by the results presented above from Corollary 1. So we will focus on phase shift induced errors the analysis for 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 multi-valued nature of phase shifts becomes important. But initially we will focus on the limit where $v \ll \pi^2$.

The remainder to 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 analyse two key cases, the single photon and two photon performance. The former can be considered a demonstration of the classical wave nature of the probability distribution for a single photon. The latter considers Hong-Ou-Mandel [5] style quantum interference.

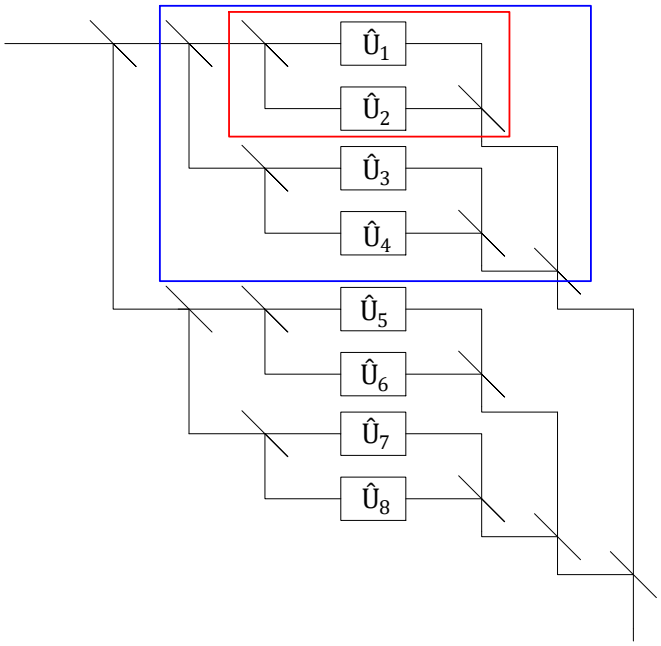

FIG. 2: Redundant encoding using 50:50 beam-splitters for $N = 8$. The boxes labeled U_i can be single or multi-mode. In the multi-mode case the encoding beam-splitter network is repeated for each mode input and output. Output modes that are post-selected on the vacuum are not shown here. The red box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

A. 1 photon inputs

The 1 photon network considered here is shown in Figure 3 for the $N = 2$ case.

The input single photon state is $|\phi\rangle = \hat{a}^\dagger |0\rangle$. After traversing the error averaged network, the resulting unnormalised output state conditional on all encoded modes

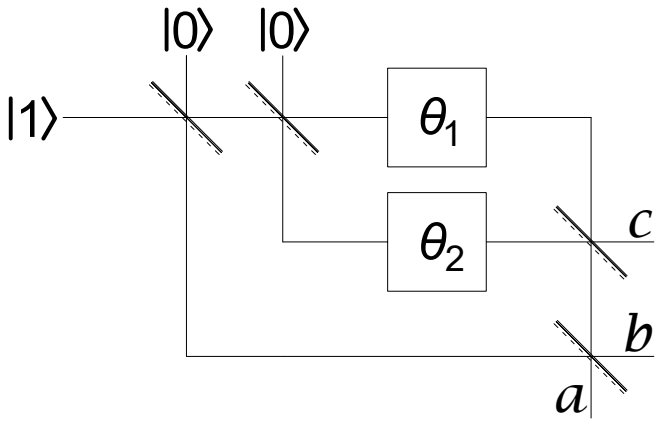

FIG. 3: Diagram of MZ based tunable beam-splitter with a redundantly encoded phase shift of $N = 2$. a and b label output modes and c labels an error detection mode. The input state shown is used for all single photon calculations. The phase shift elements are marked with θ_j and are random variables.

being vacuum is

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

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 conserves photon number, and we have post-selected 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 un-normalised state $|\psi\rangle$ the probability of observing the photon in the $\hat{a}$ and $\hat{b}$ modes without post-selection to be

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

and

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

where we have taken a first-order expansion in the phase shift variance v. Using the conservation of photon number, the vacuum post-selection is equivalent to the case when the photon exits back into the MZ interferometer. This means the probability of success is the sum of the probabilities of the $\hat{a}$ mode and $\hat{b}$ mode. This is

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

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

In the large N limit, this corresponds to the linear approximation of Equation 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 non-ideal interferometer performance. In this case the probability of observing the output in the correct mode without post-selection is

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

After post-selection this becomes

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

Figure 4 shows how these two quantities scale with N. In particular, it can be seen that after post-selection 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 thought the total quantity of errors added to the system increases, the effects of the combined errors on the interferometer is less. This can be understood by the process of redundantly splitting and recombining the state, tends drive noise into the post-selected ports by interference.

B. 2 photon inputs

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 behaviour of quantum interference when using the redundant encoding. For two photons we can compute the un-normalised the output state for the a and b modes is

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

where we have chosen $\theta = 0$ when computing this state. 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 behaviour, just the single photon results independently applied to the two input photons.

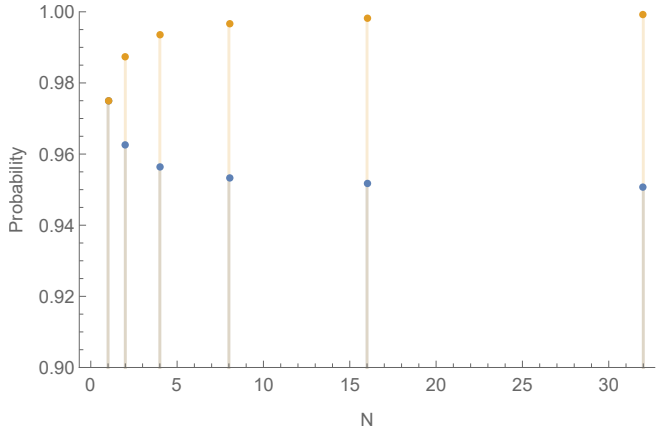

FIG. 4: Probability of single photon being detected at the a output port as shown in Figure 3 as a function of the redundant encoding size $N = 2^n$. Here the phase shifts are sampled from a distribution with mean value 0 and variance $v = 0.1$ rad^2. The lower orange values correspond to the case without post-selection and blue upper values corresponds to post-selected probability conditional on the photon not exiting the added redundant encoding modes. Eq. (27) predicts an asymptote of 0.95 without post-selection and Eq. (28) predicts an asymptote of 1 with post-selection. The asymptotic behaviour is consistent with the plotted data.

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 *(TODO: CHECK THESE EQUATIONS)*

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

where the approximation is assuming v small. Post-selection will increase this to

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

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, the large N limit of this equation the result matches the prediction of Equation 16. Also, the probability of success scales as $\frac{1}{N}$ which is the same as for the single photon input case.

IV. COMPARISON BETWEEN AVERAGING TECHNIQUES

So far we have presented two specific procedures for implementing a redundantly encoded phase shift with linear elements. There exists many more combinations of possible ways to do this. In this section we will look at combinations of redundant encodings using the concatenated beam-splitter method. In particular we will study two different methods, which we will refer to as averaging at the end and averaging each step. Fig. 5 shows schematically these two configurations as well as a baseline comparison. The system analysed 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 (Figure 5(b)). The method of averaging each step involves a redundancy of N for each of the M applied phase shifts (Figure 5(c)). By averaging each component individually (Fig. 5(c)), significantly more encoding beam-splitters are required, however we will show this leads to more stability in the output state. In the low resource and low error limit these two methods also appear to be equivalent. We can then infer that in the case when the two methods do yield equivalent results, the approximately commuting error assumption of Corollary 2

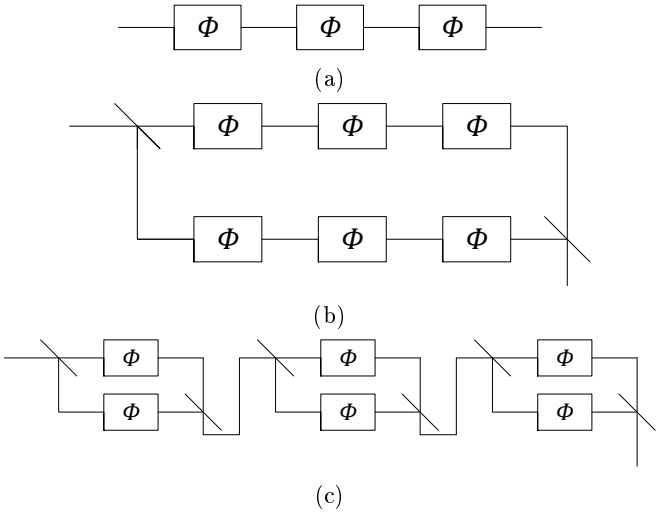

FIG. 5: Three methods of applying three phase shifts, each marked with a in series. (a) Three phase shifts with no error averaging. (b) Three phase shifts when averaging across the system. (c) Three phase shifts when averaging across each phase shifter individually. Averaging across the system will in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

does hold. In particular the low error limit allows the approximation that the generators of each error commute with one another. However this approximately commuting assumption will fail to hold in general when the size of the errors becomes too large. The effect of this can lead to significant differences in the effectiveness of the two arrangements as will be discussed in this section.

Following an approach motivated by the previous 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 allow for the errors here to be compared with those modelled in sections 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 it 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 post-selection. As however the probability of success is 1 no post selection 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 for the quantity of error when no averaging is applied, E_{noAv}, is

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

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 Figure 5(b). Preceding 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

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

This result is similar to the what was found in previous sections, see Eq.26 and Eq.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 post selection. 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

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

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

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

Therefore, the error measure, E_{aveEnd}, will be

$$E_{aveEnd} \approx \left[1 - \frac{1}{4}\left(Mv - \frac{M^2v^2}{4} + \left(1-\frac{1}{N}\right)\left(Mv - \frac{1}{2}M^2v^2\right)\right)\right] \times \left[1 - \left(1-\frac{1}{N}\right)\left(\frac{Mv}{2} - \frac{1}{4}M^2v^2\right)\right]^{-1}$$

C. Averaging Across Each Phase Shifter Individually

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

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

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

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

and an error of

$$E_{aveStep} \approx \left[\frac{3}{4} - \frac{Mv}{4} + \frac{M^2v^2}{16} + \frac{1}{4}\left(1 - \left(v - \frac{v^2}{2}\right)\left(1 - \frac{1}{N}\right)\right)^M\right] \times \left[\frac{1}{2} + \frac{1}{2}\left(1 - \left(v - \frac{v^2}{2}\right)\left(1 - \frac{1}{N}\right)\right)^M\right]^{-1}$$

Importantly, for both this case as well as when averaging each step, if only the first order approximation is used and $M = 1$ then the error matches the error found in section III A.

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 behaviour 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 $M \approx 6$ onwards. This could either be suggesting an issue with the quality of the second order approximations, as seen in Figure 7 or that there is some more fundamental point at which there is a clear benefit to averaging each component individually. The next consideration was 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. A first and second order analytical solutions is shown as well as a statistical model of the probability of success where the phase value of each phase shifter was sampled from a Gaussian random distribution. The probability of success was then calculated based on these random samples. The top four graphs were plotted for a low value of the variance on the individual phase shifters. This was done so that the behaviour of when the first and second order approximations fail to match the actual simulated result can be more 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 behaviour in their overall trends although the variation between the

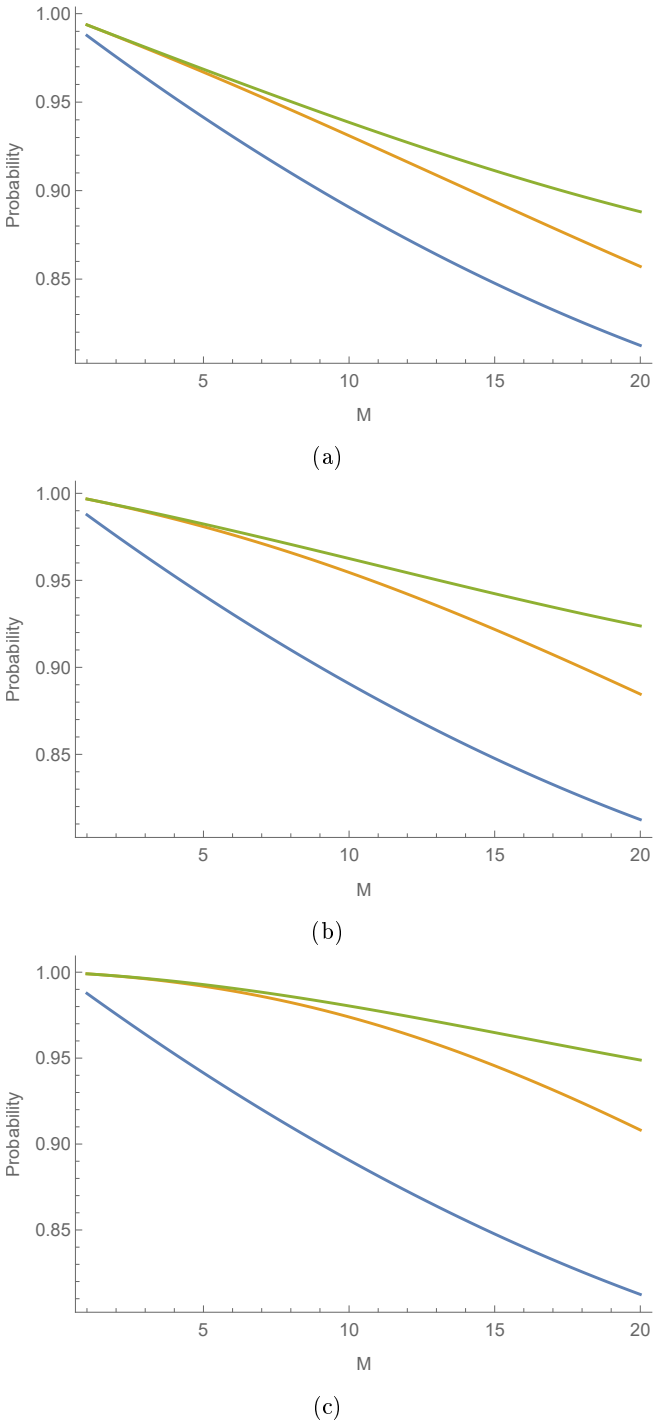

FIG. 6: Probability of obtaining the correct result as measured by the Mach-Zehnder interferometer set up as a function of the number of phase components M. Here a probability of 1 corresponds to no error and the smaller the probability the larger the error. The blue line represents the no Error Averaging applied result, the orange line corresponds to the error when averaging across the entire system and the 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) $N = 2$, in (b) $N = 4$ and in (c) $N = 16$.

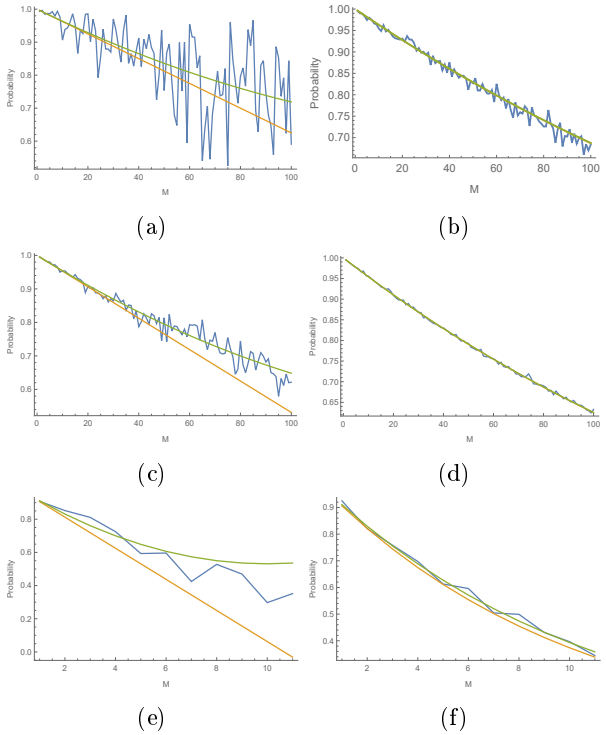

FIG. 7: Probability of success as a function of the number of phase components M for averaging across the system (left) and averaging across each component (right). Blue line is the statistical model, orange is the first order approximation and green is the second order approximation of the analytical value. a) and b) individual phase shifter variance is 0.005 rad^2 with $N = 4$, c) and d) individual phase shifter variance is 0.005 rad^2 with $N = 16$ and e) and f) individual phase shifter variance is 0.1 rad^2 with $N = 16$.

statistical simulation and the analytical solutions is significantly greater when averaging across the entire system than when averaging each component individually. 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 fail 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. This can most clearly be seen in Figure 7(e) where the first order approximation fails almost instantly and the second order approximation soon after while in Figure 7(f) both first and second order approximations both follow the statistical model closely. It is again observed that as N increases, the probability of success goes down. What is also shown is that the variation in the statistical simulation also goes down suggesting greater amount of averaging reduces the variability in both the average case and any given sample of the applied phase.

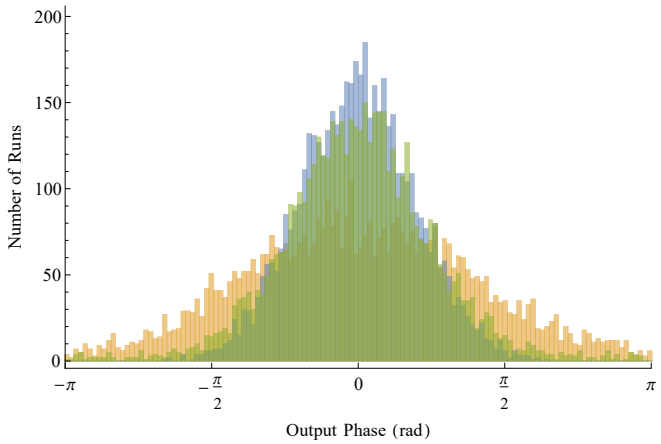

FIG. 8: Top: Total applied phase over 5000 runs for no averaging (orange), averaging across the entire system (green) and averaging each phase shifter individually (blue). Bottom: Histogram of the total applied phases. Each individual phase shifter has a variance of 0.1 rad^2 and each system has 15 phase shifters in series. The two error averaged circuits are each averaged 4 times.

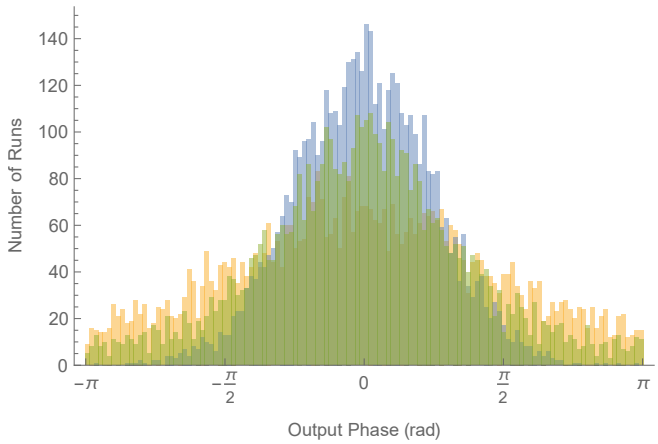

FIG. 9: Top: Total applied phase over 5000 runs for no averaging (orange), averaging across the entire system (green) and averaging each phase shifter individually (blue). Bottom: Histogram of the total applied phase. Each individual phase shifter has a variance of 0.3 rad^2 and each system has 8 phase shifters in series. The two error averaged circuits are each averaged 4 times.

E. Statistical Modelling of the Applied Phase

To better understand the behaviour of the three phase applying systems, the total applied phase was modelled using *Mathematica* with phase values chosen from a Gaussian random distribution with mean 0 and variance v. This now corresponds to Equation (9) with $\theta = \prod_i \theta_i$. This was repeated 5000 times and the results are shown in Figure 8 and Figure 9. This again shows a difference between averaging across the entire system and averaging at each step. In the range of applied phases is smaller when each phase shifter is corrected individually, an indication that averaging each step is more effective. By comparing Figure 8 with Figure 6 at $M = 15$ we can infer that the difference between the two error correction methods seen in Figure 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 the individual applied phases are uncorrelated they are expected to simply add, such that the variance without any Error Averaging is expected to simply 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 can similarly be expected to simply 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 behaviour 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 $-\pi < \theta \leq \pi$. 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 the maximum variance will be given by

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

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 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 suggest 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 50000 data points for each value of v. Figure 11 demonstrated that again the two error correction methods initially are equivalent. Once

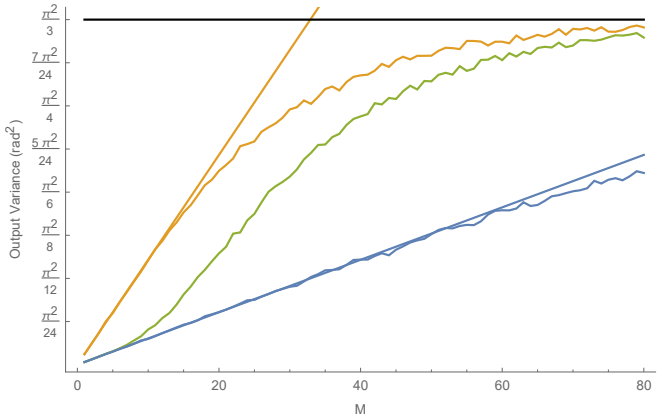

FIG. 10: Variance in the total applied phase without Error Averaging (orange), when averaging across the entire system (green) and when averaging each component individually (blue), all plotted as a function of the number of phase shifters in series (M). The variance of a single phase shifter is 0.1 rad^2 and the two error averaged systems averaged 4 times. The predicted, linear variance without any averaging (orange) and with averaging (blue) is also shown. These 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 in black.

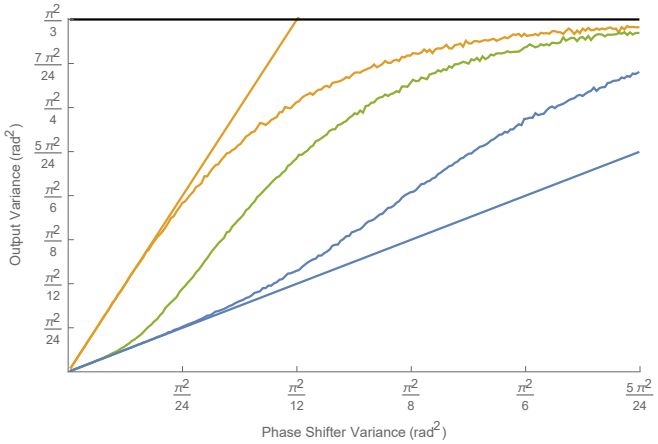

FIG. 11: Variance in the total applied phase without Error Averaging (orange), when averaging across the entire system (green) and when averaging each component individually (blue), all plotted as a function of the variance in each individual phase shifter. Each system applied 4 phase shifters in series, that is $M = 4$, and the two error averaged systems averaged 4 times, that is $N = 4$. The predicted variance without any averaging (orange) and with averaging (blue) is also shown along with the maximum allowable variance.

more averaging across the system departs from the linear regime and now we can clearly see that so too does the averaging each step.

This is suggestive of the existence of some threshold for the amount of error in a system before Error Averaging fails to be beneficial. It also suggests this threshold might correspond to when the errors no longer approximately commute. 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. 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 $M \leq 5$ which is precisely what is seen in Figure 10. These two thresholds obviously do not apply to a general error however it is hoped that with 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 maximised 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 complicated system was also investigated. In particular, a four mode system with each method of Error Averaging was considered. The set up consisted of a four mode system with four beam splitters, each implemented as above, that is each being its own MZ interferometer as shown in Figure 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 section III A and by concatenating the entire system in an interferometer as shown in Figure 13. All results were found using *Mathemtatica* 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 Appendix A

After determining the output probability distribution with and without post selection it was found that the two correction methods produced equivalent results. The $1/N$ scaling in error after post selection, as seen above, was also observed, suggesting this pattern holds for an arbitrary number of modes and arbitrary system. The

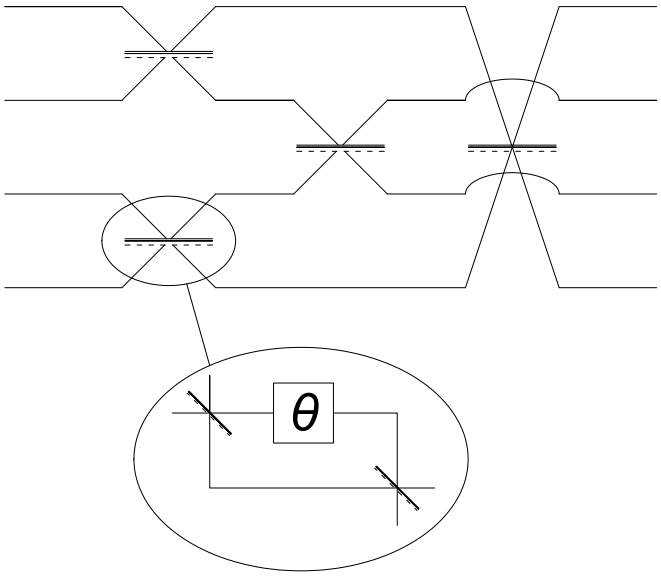

FIG. 12: Diagram of the four mode linear optical network which forms the basis of the four mode set-ups.

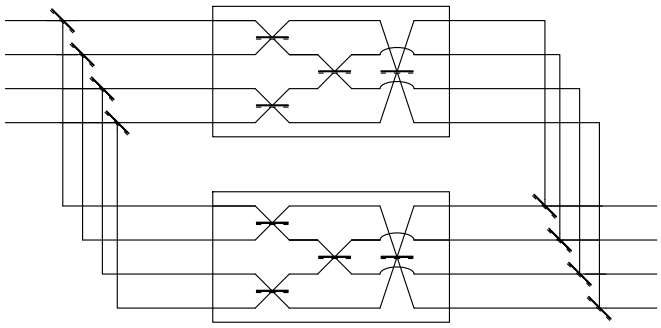

FIG. 13: Diagram of the four mode linear optical network averaged across the system once. A two mode system averaged in this fashion could be considered as an Error Averaged dual rail single qubit unitary transformation.

following section is based on the results of these simulations.

1. *Generalising 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 as when $N = 1$:

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

The probability of an incorrect state is therefore trivially:

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

At this stage there are no error ports and so $P_i(wrong) + P_i(correct)$ represents the probability of success as previously defined. Note that the subscript indexes the corresponding size of the system. Explicitly $i+1$ corresponds to a system with twice as much averaging as i. So averaging once changes these values to:

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

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

And so on, so in general:

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

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

The probability of getting the correct state with post selection will then be

$$\begin{aligned} &P_n\left(correct \,|\text{post selection}\right) \\ &= \left(1 - \frac{a_1}{2^{n-1}b_1}v\right) \xrightarrow[n\to\infty]{} 1 \end{aligned} \tag{54}$$

The probability of success was found to be

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

Hence we get a general bound on the probability of success as

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

The result can be understood to be the first order approximation to Equation 21. The $\frac{a_1+1}{b_1}$ coefficient does not quite match what might be expected from Equation 21 however it can be understood 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 Figure 2 where, depending on which components are considered to be the system, it is averaged either 8, 4 or 2 times.

VI. SCALING OF ERROR AVERAGING

The previous sections have analysed this protocol in the case of small fixed mode numbers. We will now discuss the more general situation involving larger numbers of modes. For these larger systems we are particularly interested in understanding how to minimise the number of encoding resources and while ensuring the system remains in the well understood small error regime of Corollary 2.

The number of possible configurations of linear optical elements for a general m-mode system is large. We will therefore consider a constrained set of configurations in the analysis presented here. We construct the m-mode unitary using beam-splitters and phase shifters such that each input mode is connected via some path to each output mode. Also, these are laid out so that the same number of beam-splitters and phase shifters are encountered by every path. The first requirement ensures that non-trivial interactions occur between all modes. Note, that even though every mode will be mixed with every other mode, the system is insufficient to implement an arbitrary unitary transformation of the input. The second requirement, that each path has the same number of beam splitters and phase shifters to ensure the system's error profile is completely symmetric.

To create such a system we choose and even number of modes m arranged in rows. Then we arrange the beam splitters in columns with $b = \frac{m}{2}$ beam splitters in each column. They are arranged such that, starting with the top most mode, the first column mixes nearest neighbour modes, the second column mixes the next nearest neighbour modes and each subsequent column mixes the next closest mode which it is yet to be mixed with as demonstrated in figure 14. This gives a depth of c columns where $c = \lceil \log_2(m) \rceil$.

As stated above, this arrangement is not suitable for implementing an arbitrary unitary transformation. This is because to $\mathfrak{U}(\mathrm{m})$ has dimensionality m^2 and so one requires a universal decomposition to have at least this many elements with parameters to vary (for such a decomposition see [3]). Our arrangement has

$$b \times c = \frac{m}{2} \log_2(m) \tag{56}$$

beam-splitters of arbitrary reflectivity and so, the number of free parameters will be only $m \log_2(m)$. However, this arrangement does satisfy the two properties that we required.

Using such a system we can either average across the system or average each component individually to protect against errors.

In the analysis to this point we have made the assumption that perfect, fixed $50:50$ beam splitters can be used to generate the redundancy in the input modes.

The following discussion ignores the probability of success of Error Averaging, which will lead to an exponential overhead in the number of times the device must be run before obtaining a successful outcome. To address this Error Averaging will require a loss recovery code [7] however this is outside the scope the current paper.

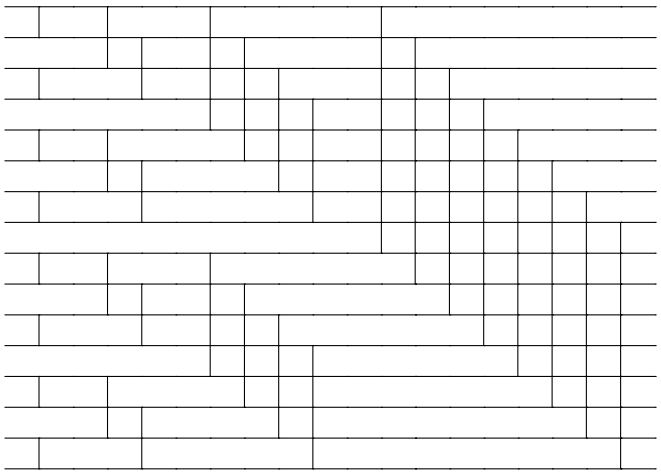

FIG. 14: Example single photon, sixteen mode system. Each vertical line represents a tunable beam splitter which would need to include appropriate phase shifters to improve generality. These were left of . Such a system allows every mode mixed with every other mode in a unique way with equal depth.

It is worth noting that in the paper *Universal Linear Optics* a single, controllable, beam-splitter was implemented using an tunable Mach-Zehnder interferometer with fidelity $\mathcal{F} = 0.992 \pm 0.008$ [4]. This suggests that currently it is possible to achieve high quality, low error, tunable linear optics components.

One necessary and perhaps sufficient condition for this assumption to be valid is for the number of error causing components in the system to be far greater than the number of perfect components required. This would imply that even if the fixed beam splitters are no better than the variable ones being implemented, Error Averaging will still dramatically reduce the total error in the system.

If such a system is corrected using Error Averaging and the minimum number of correction components then the method of correcting across the entire system would be used. To average a single mode N times, $2(N-1)$ perfect beam splitters are required. So the precise resource cost will be dependent on the circuits specific set-up. Following on from the previously described set up, we know we can mix m modes in a unique way such that every path has the same number of beam splitters and phase shifters with a circuit depth of $\log_2(m)$.

Aaronson and Arkhipov, inspired by a boson sampling type system with n photons in the top n modes, have shown [8] that for this system to be adequate to implement an arbitrary unitary operation it must be chained together n times. Each block will have m modes, a depth of $\log_2(m)$ and therefore $\frac{m}{2}\log_2(m)$ beam splitters. This then means that our system will have $\frac{mn}{2}\log_2(m)$ error

laden components and require a minimum of $2(N-1)m$ correction components to allow for averaging N times across the entire system. Given that averaging across the entire system falls out of the linear regime very quickly as the depth increases it would likely be insufficient to correct our system. If however each component has Error Averaging applied to it individually then there would need to be at most $mn\log_2(m)(N-1)$ perfect $50:50$ beam splitters, more than the original size of the circuit. Note this can result in correcting components which cannot possibly be involved in the result. For example if only the top one or two modes have photon inputs then only $2(m-1)(N-1)+m(n-1)\log_2(m)(N-1)$ correction components are necessary. The upper bound on the number of correction components, $mn\log_2(m)(N-1)$, is more general, and so will be used for the remainder of this discussion. This suggests something part way between these two cases is necessary. If a system was split up into x sections, each of which is averaged N times, than $2xm(N-1)$ correction components are necessary. This then gives the very loose inequality of

$$\begin{aligned} 2xm(N-1) &\ll \frac{mn}{2}\log_2(m) \\ 4x(N-1) &\ll n\log_2(m) \end{aligned} \tag{57}$$

for the assumption that an averaging circuit can be made using relatively perfect beam splitters to be valid. The requirement of Error Averaging to be in the linear regime will also impose the condition that $x \to n\log_2(m)$ as the error in each component gets bigger. There is the possibility that this is not quite so bad however as most of the perfect beam splitters are actually error averaged themselves as shown in Figure 2. This could allow greater freedom in what is meant by a perfect beam splitter. However even if only the outermost beam splitters are to be considered, and all inner beam splitters are said to be corrected themselves, there will still require between $2m$ and $mn\log_2(m)$ perfect fixed beam splitters.

Another important issue is to decide how any times the system needs to be averaged, that is, how N needs to grow with m and n. It has been shown that for the total systems error to be vanishingly small the error in each individual element in a linear optical system needs to be on the order of $\frac{1}{n}$ where n is the number of photons [9]. It has also been shown that it is necessary for the error to scale as $\frac{1}{n\log_2(m)}$ [8] in the case of a boson sampling style system with 1 photon in the first n modes of an m mode system. Combining these two results gives a necessary scaling of $\frac{1}{n^2\log_2(m)}$. All of our results so far suggest the error scales as $\frac{1}{N}$. By using this scaling we can estimate that $N \sim n^2\log_2(m)$ for the total error in the output to be preserved. Combining this with Eq. 57 then gives

$$4x\left(n^2\log_2(m)-1\right) \ll n\log_2(m) \tag{58}$$

which is clearly not satisfied. This suggests the best hope for the key assumption to be satisfied is if the inner perfect beam splitters are considered to be corrected. As this gives

$$\begin{aligned} 2xm &\ll \frac{mn}{2}\log_2(m) \\ \implies 4x &\ll n\log_2(m) \end{aligned} \tag{59}$$

there is the possibility that this can be satisfied. Even if this is reduced to $4x \approx n\log_2(m)$ there will still be some total reduction in error as the fixed beam splitters are likely to have less total error. It is possible to generalise this result by allowing photons to be inserted into any of the input modes. This will require our system with a depth of $\log_2(m)$ to be chained m times. This can be understood by considering the that an arbitrary unitary transformation requires that a photon from any of the input modes must be able to interact with all other photons. Given we are using beam splitters as our basic components this limits every interaction to two photon interferences. Each block of depth will allow any photon to interact with any one other photon and so for a photon to potentially interact with all other photons, we will require n of these blocks. We also need to consider the scaling to allow at least one photon in all m modes giving a scaling of $\frac{1}{mn\log_2(m)}$ to achieve a negligible amount of total error and giving a final rough bound of

$$4x \ll m\log_2(m) \tag{60}$$

where now $x \to m\log_2(m)$ as the error in each component gets bigger.

This is, in its current state a unsolved problem. It is hoped that this can be extended to a fault tolerant regime and thresholds can be found so that stronger conclusions on the applicability of Error Averaging can be made.

*(NOTE: I cannot figure out what the **key** messages are here.*

VII. CONCLUSION

We have shown how, given multiple noisy copies of a linear optical unitary network, Error Averaging can be use to implement an transformation that tends towards the average with some reduced variance at the cost of the probability of success. After post-selection Error Averaging forms a rudimentary error correction protocol as the Error Averaging appears to filter the noise from the unitary network encoding. 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 characterised solutions for the single parameter and small Gaussian error cases. We have also analytically determined the photon number expectation values in particular two mode systems for both one and two photon inputs, simulated the output expectation values in four mode systems for both one and two photon inputs and by simulated the variance in the output of a series 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 behaviour as averaging each phase provided the error within the system is small. This behaviour is conjectured to be explained by considering the errors as approximately commuting.

[1] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Physical Review Letters **77**, 198 (1996).

[2] G. L. Long, International Journal of Theoretical Physics **50**, 1305 (2011).

[3] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Physical Review Letters **73**, 58 (1994).

[4] 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).

[5] C. Hong, Z. Ou, and L. Mandel, Physical Review Letters **59**, 2044 (1987).

[6] Y. S. Patil, S. Chakram, and M. Vengalattore, Phys. Rev. Lett. **115**, 140402 (2015).

[7] T. C. Ralph and G. J. Pryde, Progress in optics **54**, 209 (2010).

[8] S. Aaronson and A. Arkhipov, in *Proceedings of the forty-third annual ACM symposium on Theory of computing* (ACM, 2011) pp. 333–342.

[9] A. Arkhipov, arXiv preprint arXiv:1412.2516 (2014).

Appendix A: Four Mode Numerical Results

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

Tables I and II show the output probabilities and correct result probability with post selection 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 sections 3.2 and 3.4 suggesting 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 post selection 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 however this is expected as adding two photons in the same mode will not necessarily lead to new interference effects being observed.

Tables VI and VI show the output probabilities and correct result probability with post selection 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 allowed a prediction of the error models for applications of Error Averaging, as discussed below.

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 post selection	$1-\frac{v}{2}$	$1-\frac{v}{4}$	$1-\frac{v}{8}$

TABLE I: Output probabilities for various levels of correcting the individual beam splitters in a 4 mode set-up 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 post selection	$1 - \frac{v}{2}$	$1 - \frac{v}{4}$	$1 - \frac{v}{8}$

TABLE II: Output probabilities for various levels of correcting the across the system in a 4 mode set-up 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 ($N = 2$)
$\|2,0,0,0\rangle$	$1 - v$	$1 - \frac{3v}{2}$
$\|0,2,0,0\rangle$	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 post selection	$1 - v$	$1 - \frac{v}{2}$

TABLE III: Output probabilities for various levels of correcting the individual beam splitters in a 4 mode set-up 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 post selection	$1 - v$	$1 - \frac{v}{2}$	$1 - \frac{v}{4}$

TABLE IV: Output probabilities for various levels of correcting the across the system in a 4 mode set-up 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 ($N = 4$)
$\|2,0,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,2,0,0\rangle$	$\frac{v}{2}$	$\frac{v}{4}$	$\frac{v}{8}$
$\|0,0,2,0\rangle$	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 post selection	$1-\frac{3v}{2}$	$1-\frac{3v}{4}$	$1-\frac{3v}{8}$

TABLE V: Output probabilities for various levels of correcting the individual beam splitters in a 4 mode set-up 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 post selection	$1-\frac{3v}{2}$	$1-\frac{3v}{4}$	$1-\frac{3v}{8}$

TABLE VI: Output probabilities for various levels of correcting the across the system in a 4 mode set-up given an input of the state $|1,1,0,0\rangle$ where v is the variance of the phase error.