

# Error Correction Through Error Averaging

Ryan J. Marshman,<sup>1,\*</sup> Austin Lund,<sup>1</sup> Peter Rohde,<sup>2</sup> and Timothy Ralph<sup>1</sup>

<sup>1</sup>*Centre for Quantum Computation and Communication Technology (CQC2T), The University of Queensland, Australia.*

<sup>2</sup>*Centre for Quantum Computing & Intelligent Systems (QCIS), University of Technology Sydney, Australia.*

(Dated: March 24, 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 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 improves on existing methods of error detection and correction by not relying on ancillary photons or complicated control and correction circuits [? ]. 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 satisfied by LOQC and as such this will form the basis for most of this paper. In particular Jacques Carolan and colleagues [? ] used linear optics to demonstrate a device capable of realising all possible linear optical computational protocols although was limited by the size of the circuit. Given the success of this device it was chosen to form the test bench for Error Averaging.

Error Averaging in its current state forms a rudimentary form of error correction which currently acts to distill errors from 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 size of the error 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.

In the next section details and results of most general set-up will be discussed. This is done to both explain Error Averaging as well as highlight the necessity of the two requirements on its implementation. Section III includes numerous proof of principle examples which serve to highlight the effects of Error Averaging. LOQC will serve as the architecture for all these examples. The fea-

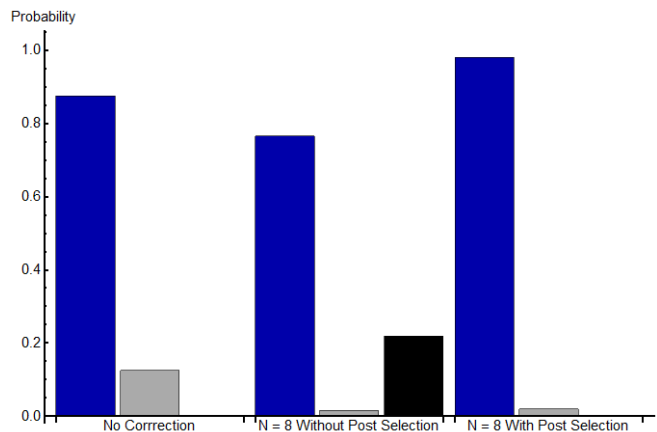


FIG. 1. Comparison of output probability distributions with and without error averaging. This also demonstrates the effect post selection has on the output distribution. The blue bars represent the probability of observing the photon in the correct output mode, grey corresponds to observing the photon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

sibility 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 IV followed by any concluding remarks.

## II. GENERAL RESULTS

First let us consider a very general system. This will serve as a basis to explore the properties and effect of error averaging.

\* ryan.marshman@uqconnect.edu.au

### A. Theorem 1

Given some target transformation  $|\Psi\rangle = \hat{U}|\psi\rangle$  to be implemented by some other transformation  $\hat{U}_{error}$  where  $\hat{U}_{error}$  is in some sense close to  $\hat{U}$ ,  $N$  levels of redundancy can be employed to give an effective transformation  $\hat{U}_{effective}$  given by

$$\hat{U}_{effective} = \frac{1}{N} \sum_{i=1}^N \hat{U}_i \quad (1)$$

### B. Proof

Define the total state space as a direct sum of  $N$  copies of the original transformation's state space  $\{|\phi\rangle\}$ . The input state  $|\psi\rangle$  is then

$$|\psi\rangle \rightarrow |\psi\rangle_1 \oplus \left( \bigoplus_{i=2}^N |0\rangle_i \right) \equiv |\psi\rangle_{in} \quad (2)$$

that is, the total input state  $|\psi\rangle_{in}$  corresponds to the original input state  $|\psi\rangle$  in the first subspace and the vacuum state in all others.

Define the  $N$  copies of the transformation as

$$\hat{U}_i = \left( \hat{U}_{error} \right)_i \oplus \left( \bigoplus_{j=2}^N \mathbb{I}_j \right) \quad (3)$$

with the constraints that

$$\hat{U}_i \bigoplus_{j=1}^N |\psi\rangle_j = \left( \hat{U}_i |\psi\rangle_i \right)_i \oplus \left( \bigoplus_{j \neq i} |\psi\rangle_j \right) \quad (4)$$

$$\hat{U}_i |0\rangle_i = |0\rangle_i \quad (5)$$

Define  $\hat{U}_{tot} = \bigotimes_{i=1}^N \hat{U}_i$

Define the splitting operator,  $\hat{S}$  as an operator which mixes the subspaces such that

$$\hat{S} |0\rangle_i = |0\rangle \quad (6)$$

$$\hat{S} |\phi\rangle_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N (-1)^{f(i)} |\phi\rangle_i \oplus \left( \bigoplus_{j \neq i} |0\rangle_j \right) \quad (7)$$

where  $f(i) \in \{0, 1\}$ .

Define the recombining operator,  $\hat{R}$  as a similar subspace mixing operator which acts as

$$\hat{R} |0\rangle_i = |0\rangle \quad (8)$$

$$\hat{R} |\phi\rangle_i = \frac{1}{\sqrt{N}} \sum_{j=1}^N g(i, j) |\phi\rangle_j \oplus \left( \bigoplus_{k \neq j} |0\rangle_k \right) \quad (9)$$

where  $g(i, j) \in \{-1, 0, 1\}$  and  $g(i, 1) = (-1)^{f(i)}$ .

Define the post selection operator,  $\hat{P}$  to be employed at the end of the error averaging procedure such that is post selects on all but the first subspace being in the vacuum state, that is

$$\hat{P} = \mathbb{I}_1 \oplus \left( \bigoplus_{i=2}^N |0\rangle_i \langle 0|_i \right) \quad (10)$$

With this error averaging can be defined as  $\hat{P} \hat{R} \hat{U}_{tot} \hat{S} |\phi\rangle_{in}$ . Its action on the input state will therefore be

$$\hat{P} \hat{R} \hat{U}_{tot} \hat{S} |\psi\rangle_{in} \quad (11)$$

$$= \frac{\hat{P} \hat{R} \hat{U}_{tot}}{\sqrt{N}} \sum_{i=1}^N (-1)^{f(i)} |\psi\rangle_i \oplus \left( \bigoplus_{j=2}^N |0\rangle_j \right) \quad (12)$$

$$= \frac{\hat{P} \hat{R}}{\sqrt{N}} \sum_{i=1}^N (-1)^{f(i)} \hat{U}_i |\psi\rangle_i \oplus \left( \bigoplus_{j \neq i} |0\rangle_j \right) \quad (13)$$

$$= \frac{\hat{P}}{N} \sum_{k=1}^N \sum_{i=1}^N g(i, k) (-1)^{f(i)} \left( \hat{U}_i |\psi\rangle_i \right)_k \oplus \left( \bigoplus_{j \neq k} |0\rangle_j \right) \quad (14)$$

$$= \frac{1}{N} \sum_{i=1}^N g(i, 1) (-1)^{f(i)} \left( \hat{U}_i |\psi\rangle_i \right)_1 \oplus \left( \bigoplus_{j=2}^N |0\rangle_j \right) \quad (15)$$

$$= \frac{1}{N} \sum_{i=1}^N \left( (-1)^{f(i)} \right)^2 \left( \hat{U}_i |\psi\rangle_i \right)_1 \oplus \left( \bigoplus_{j=2}^N |0\rangle_j \right) \quad (16)$$

Thus

$$\hat{P} \hat{R} \hat{U}_{tot} \hat{S} |\psi\rangle_1 \oplus \left( \bigoplus_{i=2}^N |0\rangle_i \right) \quad (17)$$

$$= \frac{1}{N} \sum_{i=1}^N \left( \hat{U}_i |\psi\rangle_i \right)_1 \oplus \left( \bigoplus_{j=2}^N |0\rangle_j \right) \quad (18)$$

$$\hat{U}_{effective} |\psi\rangle_1 = \frac{1}{N} \left( \sum_{i=1}^N \hat{U}_i \right) |\psi\rangle_1 \quad (19)$$

Where  $\hat{U}_{effective} = \hat{P} \hat{R} \hat{U}_{tot} \hat{S}$ .

### C. Theorem 2

Error averaging applied to  $N$  unitaries with small, random unitary errors each with fixed, finite variance  $\nu$  results in an effective unitary with variance  $\frac{\nu}{N}$ .

### D. Proof

By Theorem II A error averaging applies an effective unitary of  $\hat{U}_{effective} = \frac{1}{N} \sum_{j=1}^N \hat{U}_j$ . Given  $\hat{U}_j$  differ from

the ideal  $\hat{U} = e^{i \sum_{j=1}^{n^2-1} \alpha_j \hat{\sigma}_j}$  by small, random errors,  $\hat{U}_j$  can be written

$$\hat{U}_j = e^{i \sum_{k=1}^{n^2-1} (\alpha_k + \delta_k^j) \hat{\sigma}_k} \quad (20)$$

where  $\sigma_k$  are a complete set of  $n^2 - 1$  generators of  $\mathfrak{su}(n)$  and  $\delta_k^j$  are the Gaussian random variables with variance  $\nu$  which dictate the size and direction of errors in  $\hat{U}_j$ . Thus

$$\hat{U}_{effective} = \frac{1}{N} \sum_{j=1}^N e^{i \sum_{k=1}^{n^2-1} (\alpha_k + \delta_k^j) \hat{\sigma}_k} \quad (21)$$

$$= \frac{1}{N} \sum_{j=1}^N \left( 1 + i \sum_{k=1}^{n^2-1} (\alpha_k + \delta_k^j) \hat{\sigma}_k \right) + O \left( \left( i \sum_{k=1}^{n^2-1} (\alpha_k + \delta_k^j) \hat{\sigma}_k \right)^2 \right) \quad (22)$$

$$= \frac{1}{N} \left( N + i \sum_{k=1}^{n^2-1} \left( N \alpha_k \hat{\sigma}_k + \sum_{j=1}^N \delta_k^j \hat{\sigma}_k \right) \right) + O \left( \left( i \sum_{k=1}^{n^2-1} (\alpha_k + \delta_k^j) \hat{\sigma}_k \right)^2 \right) \quad (23)$$

$$= e^{i \sum_{k=1}^{n^2-1} (\alpha_k + \frac{1}{N} \sum_{j=1}^N \delta_k^j) \hat{\sigma}_k} \quad (24)$$

$$= e^{i \sum_{k=1}^{n^2-1} \alpha_k \hat{\sigma}_k} e^{i \sum_{k=1}^{n^2-1} (\frac{1}{N} \sum_{j=1}^N \delta_k^j) \hat{\sigma}_k} \quad (25)$$

$$= \hat{U} e^{i \sum_{k=1}^{n^2-1} (\frac{1}{N} \sum_{j=1}^N \delta_k^j) \hat{\sigma}_k} \quad (26)$$

Which by the central limit theorem tends towards a transformation  $\hat{U}$  with Gaussian random errors which have mean 0 and variance  $\frac{\nu}{N}$ . In general an actual system, particularly the error balanced systems considered in this paper, does not necessarily result in the variance of each generator being equal. However this would not change the key result here, that the variance scales as  $\frac{1}{N}$

### E. A General System

Consider an  $n$  mode system implementing some unitary transformation  $\hat{U}$  on an input state  $|\phi\rangle$  such that

$$|\psi'\rangle = \hat{U} |\phi\rangle \quad (27)$$

where  $|\psi'\rangle$  differs from the ideal output of  $|\psi\rangle$  due to errors in  $\hat{U}$ . While this unitary is not completely general as Error Averaging requires that  $\hat{U} |0\rangle = |0\rangle$ . This form the first of the two requirements on the computational architecture.

To implement error averaging on this system  $N$  copies of  $\hat{U}$  are made, denoted here as  $\hat{U}_1, \hat{U}_2, \dots, \hat{U}_N$ . The input state is then placed in an even superposition of  $N$  spacial modes, each of which implement one of the copies of  $\hat{U}$ . This forms the second formal requirement on the implementation, specifically the input particles must be able to be placed in a spacial superposition. These spacial modes are then recombined by applying the splitting mechanism

used to encode the superposition in reverse. Errors can then be filtered out by post selecting on all extra modes left over from the splitting process being in the vacuum state. Figure 2 shows how this can be implemented using linear optics. Figure 3 highlights the location of the error detection modes and photo-detectors.

Figure 2 also demonstrates that the encoding used to implement error averaging is naturally self correcting until the outermost layer, with the inner layers being error averaged by the outer layers.

Considering a system with  $N = 2$ , that is, an averaged system with only two copies of the intended transformation. This system's transformation can be defined by taking  $u_i$  as the  $n \times n$  matrix representation of  $\hat{U}_i$  and setting

$$U_1 = \begin{bmatrix} u_1 & 0 \\ 0 & \mathbb{I} \end{bmatrix} \quad (28)$$

and

$$U_2 = \begin{bmatrix} \mathbb{I} & 0 \\ 0 & u_2 \end{bmatrix} \quad (29)$$

Next the splitting and recombining processes,  $S$ , is defined as

$$S = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbb{I} & -\mathbb{I} \\ \mathbb{I} & \mathbb{I} \end{bmatrix} \quad (30)$$

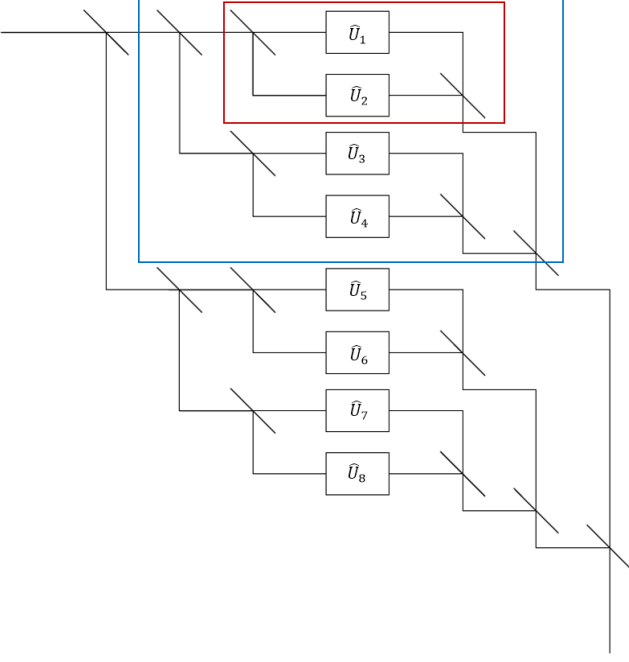


FIG. 2. Diagram showing the innermost components marked with  $\hat{U}_i$  being error averaged eight times using fourteen beam splitters. Note that for clarity the error detection modes are not included. See figure 3 for an example explicitly including the error detection modes. The inner fixed beam splitters can be considered error averaged four times by taking the component of the system in the red box as a smaller system which is error averaged. Similarly the system in blue box which is effectively being averaged twice.

It is worth noting that this is equivalent to the action of a 50:50 beam splitter mixing the  $j^{th}$  mode with the  $(j+n)^{th}$  mode. Now defining the input state

$$|\phi\rangle = |\phi_1, \dots, \phi_n\rangle \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} = \vec{\phi} \quad (31)$$

which is our desired input state  $|\phi\rangle$  spread across the first  $n$  input modes. The remaining input modes are set to the vacuum state. Similarly we will define the output state to be  $|\psi\rangle$  where

$$|\psi\rangle = |\psi_1, \dots, \psi_n\rangle \equiv \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{bmatrix} = \vec{\psi} \quad (32)$$

Half of the output states are then post selected on being in the vacuum state, leaving a post selected output state of  $|\psi\rangle \otimes |0\rangle$ . This implies the total input vector is

$$\begin{bmatrix} \vec{\phi} \\ \vec{0} \end{bmatrix} \quad (33)$$

and total output vector is

$$\begin{bmatrix} \vec{0} \\ \vec{\psi} \end{bmatrix} \quad (34)$$

where  $\vec{0}$  is a vector of  $n$  zeros. Now that everything is defined the input state can be mapped to the output state

$$\begin{aligned} \begin{bmatrix} \vec{0} \\ \vec{\psi} \end{bmatrix} &= BS \ U_1 \ U_2 \ BS \ \begin{bmatrix} \vec{\phi} \\ \vec{0} \end{bmatrix} \\ \begin{bmatrix} \vec{0} \\ \vec{\psi} \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} (u_1 - u_2) \vec{\phi} \\ (u_1 + u_2) \vec{\phi} \end{bmatrix} \end{aligned} \quad (35)$$

Therefore

$$\vec{\psi} = \frac{1}{2} (u_1 + u_2) \vec{\phi} \quad (36)$$

It is worth making clear that the result  $\vec{0} = \frac{1}{2} (u_1 - u_2) \vec{\phi}$  is entirely due to post selection. Hence it is best to consider Eq.36 as the only result from Eq.35. This shows that the above process does indeed take the average of the two error laden unitary transformations and applies that. What is more, we can easily see that by applying error averaging to each of these unitary transformations that in general, when error averaged  $N$  times the systems transformation will become

$$\vec{\psi} = \frac{1}{N} \left( \sum_{i=1}^N u_i \right) \vec{\phi} \quad (37)$$

This exemplifies the chosen name of error averaging, this can also be written as  $\vec{\psi} = A \vec{\phi}$  where  $A = \frac{1}{N} \left( \sum_{i=1}^N u_i \right)$  is an average of  $N$  unitary transformations. Importantly, provided the errors are unbiased they to should be averaged to zero, particularly in the limit of large  $N$ . This averaging filters the error into the post selection modes and post selection provides the actual error reduction mechanism.

One final useful quantity to determine is the probability of success. We can defined this probability to be the mod squared of the amplitude of the 'error free' component of the state. As such, the probability of success will be the length of  $\vec{\psi}$  which is simply

$$P(\text{success}) = \frac{1}{N^2} \left| \left( \sum_{i=1}^N u_i \right) \vec{\phi} \right|^2 \quad (38)$$

It can also be noted that Equation 37 can become the appropriate transformation for duality quantum computing by allowing the  $u_i$  to be arbitrary [? ].

### III. IMPLEMENTATION

Now that the basic idea is understood it is necessary to consider how error averaging can actually be implemented. The first of the two key requirements is that the fundamental particle used as the computational qubit can be placed in a spacial superposition. The second requirement is that any employed unitary  $\hat{U}$  does not act on the vacuum, that is  $\hat{U}|0\rangle = |0\rangle$ . Although these seem restrictive there is a natural implementation which satisfies both of these requirements, namely LOQC. A spacial superposition can be created easily using beam splitters and any linear optical unitary cannot, by definition, effect the vacuum.

Within the LOQC architecture, the following discussion will be agnostic about the specific encoding method by, as above, working in the Fock basis. In this way both single and dual rail encoding can be corrected and as such will generally not be considered in all further discussions.

The following sections will discuss a series of example systems implementing error averaging with a particular focus on phase errors. First note that two of the controllable and fundamental components in LOQC are phase shifters and beam splitters. A paper by Jacques Carolan et. al. [?] demonstrated how this can be reduced to controllable phase shifters and fixed beam splitters. The basic idea is to replace each beam splitter with a Mach Zehnder (MZ) interferometer consisting of a controllable phase shifter in one arm and two fixed 50 : 50 beam splitters. Although this idea is not particularly new they showed that such a system is able to implement many quantum logic gates and linear optical protocols with a high fidelity. In this way we also only need to consider

phase errors to correct a general linear optical quantum computing system.

It is this system which inspired the basic form of our error averaging example systems. Unless otherwise specified the 50 : 50 beam splitters are assumed to be perfect and the phase shifters are taken to be the source of all errors. The applied phase shift was taken to be of the form  $e^{i(\theta+\delta)}$  where, for the identity,  $\theta = 0$  and  $\delta$  is the random error variable, chosen from some fixed distribution. This distribution is taken to be Gaussian with mean 0 and variance  $v$ . We then error correct, or perhaps more accurately, error reduce by concatenating small MZ interferometers inside the phase implementing arm of the total system. This intuitively has the result of averaging the net result of the applied phase and its error, thus reducing the entire error of the system. This system will be discussed for a one and two photon input with  $N$  internal phase shifters where  $N = 2^n, n \in \mathbb{N}$ . As the entire system is implementing a beam splitter it has two desired input and output modes. However as  $N$  increases so too does the actual number of input and output modes. All of these extra input modes are effectively ignored by being set to the vacuum state  $|0\rangle$ . The extra output ports all have photo-detectors which will be post selected on being zero.

#### A. 1 photon N arbitrary

Consider the single photon input state  $|\phi\rangle = \hat{a}^\dagger |0\rangle$  into a single MZ beam splitter with  $N$  levels of redundancy attempting to apply some rotation angle  $\theta$ . This system is shown in Figure 3 for the  $N = 2$  case. The resulting output state is

$$|\psi\rangle = \left( \left( \frac{e^{i\theta}}{2} \left\{ \frac{1}{N} \sum_{j=1}^N e^{i\delta_j} \right\} + \frac{1}{2} \right) \hat{a}^\dagger + \left( \frac{e^{i\theta}}{2} \left\{ \frac{1}{N} \sum_{j=1}^N e^{i\delta_j} \right\} - \frac{1}{2} \right) \hat{b}^\dagger \right) |0\rangle. \quad (39)$$

where  $\theta_j = \theta + \delta_j$ .

This output is effectively truncated so as to ignore all output modes which are post selected to be in the vacuum state. This gives the probability of observing the photon in the  $\hat{a}$  and  $\hat{b}$  modes without post selection as

$$\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle = \cos^2(\theta/2) + \frac{v}{4N} - \frac{v \cos^2(\theta/2)}{2} \quad (40)$$

and

$$\langle\psi|\hat{b}^\dagger\hat{b}|\psi\rangle = \sin^2(\theta/2) + \frac{v}{4N} - \frac{v \sin^2(\theta/2)}{2}. \quad (41)$$

It is also important to note the probability of success, in this case it is given by

$$P(\text{success}) = \langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle + \langle\psi|\hat{b}^\dagger\hat{b}|\psi\rangle \quad (42)$$

$$= 1 + \frac{v}{2N} - \frac{v}{2} \quad (43)$$

To further illustrate the effect of error averaging when combined with post selection, consider the case when  $\theta = 0$ , the probability of observing the correct result without post selection is

$$\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle = 1 - \frac{(2N-1)v}{4N}. \quad (44)$$

However after post selection this becomes

$$\langle\psi|\hat{a}^\dagger\hat{a}|\psi\rangle = 1 - \frac{v}{4N}. \quad (45)$$

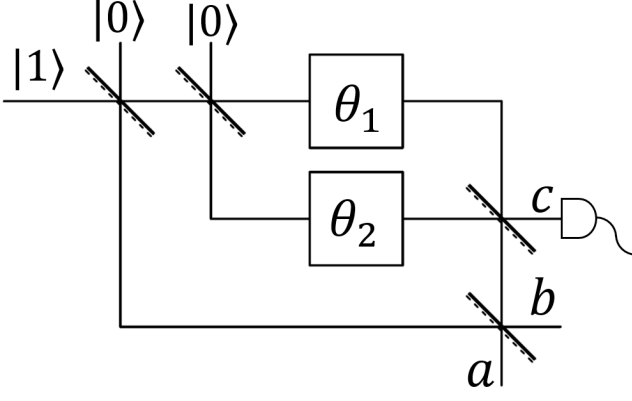


FIG. 3. Diagram of interferometer with  $N = 2$  that is, two phase shifter. Note that the  $a$  and  $b$  output modes and  $c$  error detection mode are marked along with the input state used for all single photon calculations. The phase shifters are marked with  $\theta_i$  and are the source of the error in the applied phase shift.

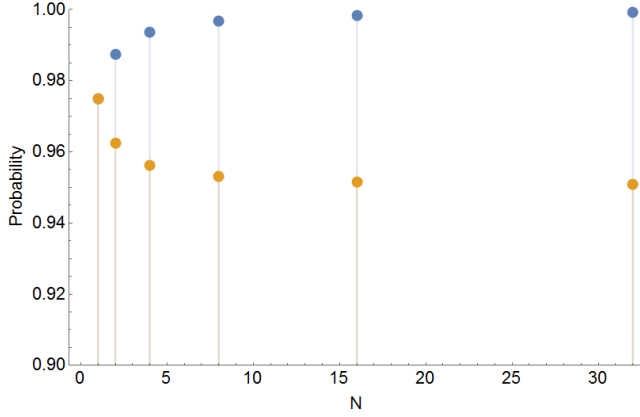


FIG. 4. Probability of observing the photon number at the  $a$  output port as shown in Figure 3. Or equivalently expected value of the photon number at the  $a$  output port. Note that this is equivalent to the probability of the photon coming out in the correct mode with and without post selection. Here  $v = 0.1 \text{ rad}^2$ , the orange values correspond to with post selection and blue corresponds to without post selection.

Figure 4 shows how these two results scale with  $N$ . In particular, it can be seen that with post selection  $\langle \hat{a}^\dagger \hat{a} \rangle$  can be made arbitrarily close to unity by increasing  $N$ . Also it is seen that while the probability of success decreases for increasing  $N$  when between 1 to 16, above  $N = 16$  the penalty asymptotes to some fixed value. This implies that as  $N$  increases the total amount of error generated in the system increases, however it does so non-linearly such that there is some maximum amount of error which can be introduced this way. Meanwhile the process of averaging and interfering the state ensures that this error tends to be found in the error ports resulting in the state in the output modes approaching perfection.

### B. 2 photons $N$ arbitrary

Given the highly quantum nature of two photon inference it was also be discussed. This also serves to demonstrate that error averaging is applicable to a wide variety of systems.

For two photons we have a slightly more complicated but similar general form for the output state as shown in Eq.46, again only considering the  $a$  and  $b$  output modes, but not considering the re-normalisation which occurs with post selection.

$$\begin{aligned}
 |\psi\rangle = & \frac{1}{2} \left\{ 1 + \frac{1}{N^2} \left( \sum_{j=1}^N \sum_{k=1}^N e^{i(\delta_j + \delta_k)} \right) \right\} \hat{a}^\dagger \hat{b}^\dagger |0\rangle \\
 & + \frac{\sqrt{2}}{4} \left\{ \frac{1}{N^2} \left( \sum_{j=1}^N \sum_{k=1}^N e^{i(\delta_j + \delta_k)} \right) - 1 \right\} (|2, 0\rangle + |0, 2\rangle)
 \end{aligned} \tag{46}$$

Here the system is attempting to implement the identity on the input state  $\hat{a}^\dagger \hat{b}^\dagger |0\rangle = |1, 1\rangle$ . This state was chosen as opposed to  $|2, 0\rangle$  as this would not necessarily

show any new behaviour, just a scaling of the single photon results. A perfect output is equivalent to having

$$\langle \psi | \hat{a}^\dagger \hat{a} \hat{b}^\dagger \hat{b} | \psi \rangle = 1 \quad (47)$$

where, again,  $|\psi\rangle_{out}$  does include and re-normalisation due to post selection. What would will actually be achieved is

$$\begin{aligned} \langle \psi | \hat{a}^\dagger \hat{a} \hat{b}^\dagger \hat{b} | \psi \rangle &= \left\langle \left| \frac{1}{2} \left\{ 1 + \frac{1}{N^2} \left( \sum_{j=1}^N \sum_{k=1}^N e^{i(\delta_j + \delta_k)} \right) \right\} \right|^2 \right\rangle \\ &= \left\langle \frac{1}{4} \left( 1 + \frac{2}{N^2} \left( \sum_{j=1}^N \sum_{k=1}^N \cos(\delta_j + \delta_k) \right) \right) \right\rangle \\ &\quad + \frac{1}{4} \left\langle \frac{1}{N^4} \left( \sum_j e^{-2i\delta_j} + \sum_{j=1}^N \sum_{k \neq j}^N e^{-i(\delta_j + \delta_k)} \right) \right\rangle \\ &\quad \times \left\langle \left( \sum_l e^{2i\delta_l} + \sum_{l=1}^N \sum_{m \neq l}^N e^{i(\delta_l + \delta_m)} \right) \right\rangle \\ &\approx 1 - v \end{aligned} \quad (48)$$

where, a second order Taylor approximation was used along with the moments of Gaussian random variables. After post selection this is improved to become what we refer to as the probability of observing a coincidence measurement,  $P(\text{coincidence})$ , which is, to a first order approximation

$$\begin{aligned} P(\text{coincidence}) &= \frac{\langle \psi | \hat{a}^\dagger \hat{a} \hat{b}^\dagger \hat{b} | \psi \rangle}{\langle \psi | \hat{a}^\dagger \hat{a} \hat{b}^\dagger \hat{b} | \psi \rangle + \langle \psi | \hat{a}^\dagger \hat{a} \hat{a} \hat{a} | \psi \rangle + \langle \psi | \hat{b}^\dagger \hat{b} \hat{b} \hat{b} | \psi \rangle} \\ &\approx 1 - \frac{v}{2N} \end{aligned} \quad (49)$$

where the binomial approximation was also used.

Finally the probability of success, when defined as the probability of not needing to post selected a single instance is

$$P(\text{success}) \approx 1 - v + \frac{v}{2N}. \quad (50)$$

### C. Comparison between averaging at end and averaging each step

With the basic procedure of error averaging in mind, it can be seen that there arises two different methods of implementation, which will be referred to as averaging at the end and averaging each step. Figure 5 shows how these two methods differ from one another and from an uncorrected system. It is seen that by averaging each component individually, significantly more resources are required, however we will show this leads to more stable results. In the small system and low error limit these two methods also appear to be directly equivalent. There can be a significant differences in the effectiveness of these two implementations as discussed in section IV D. Phase Applying Systems.

To compare the effect of averaging across the system and averaging each step a single Mach-Zehnder interferometer as a Hong-Ou-Mandel (HOM) like system was considered. The HOM system demonstrates two photon interference whereby the two photon state  $|1, 1\rangle$  is fed into a system and never observed as an output due to quantum interference [? ]. This was chosen as it generally serves as a standard test bench for quantum optics because it depending on the quality of optical components and how well the two individual photons can be distinguished the probability of observing the coincidence state,  $|1, 1\rangle$ , varies. The standard coincidence input

$$|\psi\rangle = |1, 1\rangle \quad (51)$$

was considered and the probability of observing this state in the output as a function of the applied angle  $\theta$  determined. By considering three different systems, a system with no error averaging (included for comparison), the standard beam splitter with minimal error averaging and our new system of averaging across the system. Both averaged circuits have two phase shifters and so give a valid way of testing the new system. Figure 3.6 shows the diagrammatic set up of all three systems. Table I shows the results for the three HOM set-ups.

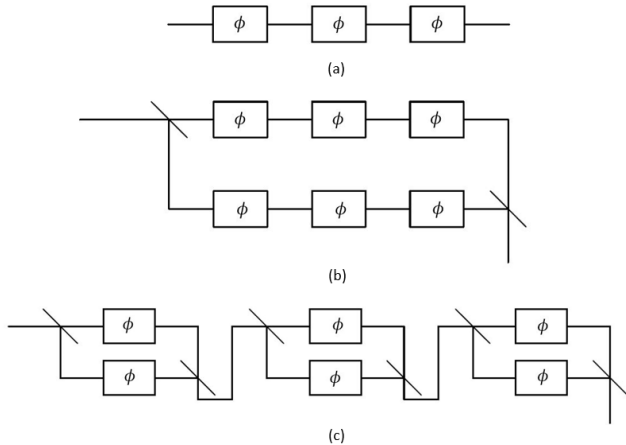


FIG. 5. Three methods of applying three phase shifters, each marked with a in series. a) three phase shifter system with no error average, b) shows the three phase shifter system when averaging across the system and c) shows the three phase shifter set-up when averaging across each phase shifter individually. It can be clearly seen that averaging across the system will in general require far fewer encoding resources.

What can be seen is that we again get a benefit from error averaging with post selection. Interestingly this also suggest averaging across the system has the same result as averaging the beam splitter. The results show the same pattern as the one and two photon systems above whereby averaging simply halves the error term. This is promising for the averaging across the system method of correction. This method is of interest as it could potentially protect against errors in the fixed beam splitters as well as phase shifters as it is averaging across all but the outer most beam splitters. This is significant as the main benefit of this method is seen for large  $N$  in which case there are a significant number of fixed beam splitters which have so far been considered perfect. It is hoped that the requirement of these to all be perfect can then be relaxed allowing these results to hold more generally and be more applicable to currently manufacturable systems. Note that by allowing an arbitrary phase shift this result again serves to verify that simply considering identity transformations is suitable to characterise the error when averaging across the beam splitter or across the system.

For error averaging to be more useful, it will need to protect against more than simply phase errors. As previously discussed, it is possible to turn beam splitter transformation errors into phase errors by simply using MZ interferometers as the beam splitters and so we can expect error averaging to reduce errors in these beam splitters too. As it has been shown that any linear unitary transformation can be implemented using beam splitters and phase shifters in some large semi-complicated arrangement [?], error averaging might be suitable for any linear optical circuit.

To gain a better understanding of how useful this method of error averaging actually is, a series of more complicated systems were then considered. In particular, we decided to explore a four mode system. Rather than just averaging each individual beam splitter, we sought to determine if it is possible to average across an entire systems. If this new method has a similar error profile as the original set-up it would be far more useful. This is because it would require far fewer perfect beam splitters for it to be implemented.

The set up being considered is a four mode system consisting of four beam splitters, each implemented as above, that is each being its own MZ interferometer as shown in Figure 6. 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 again implement the identity and error reduction was then applied as in the Hong-Ou-Mandel systems discussed above. All results were found using *Mathematica* to sample from the appropriate transformation matrix representing the system and then compute the photon number expectation values and probabilities. Figure 7 is a diagram of the 4 mode system averaged across the system once.

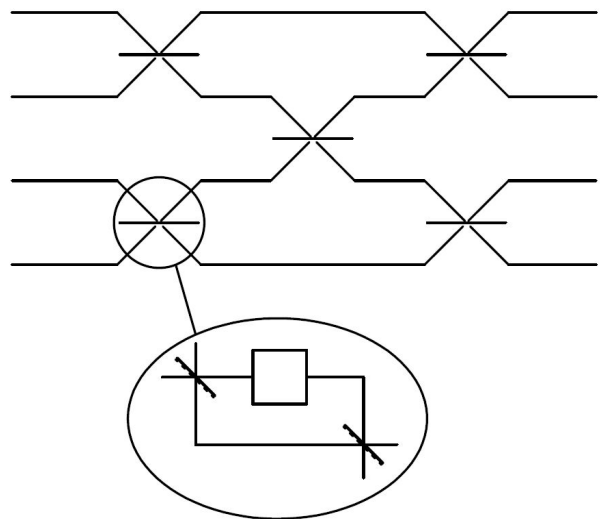


FIG. 6. Diagram of the four mode linear optical network which forms the basis of the four mode set-ups.

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. All results presented so far have a specific pattern as discussed in the following section



TABLE I. Hong-Ou-Mandel systems output results. The top line shows the probability of the coincidence result while the bottom line shows that same probability once all error ports have been post selected on. Once again  $v$  is the variance of the beam splitter error and  $\theta$  is the intended applied phase shift.

	No Averaging	Averaging in the beam splitter	Averaging across the system
$P(\text{DesiredOutput})$	$\frac{1}{2}(1 + \cos(2\theta)) - v \cos(2\theta)$	$\frac{1}{2}(1 + \cos(2\theta)) - \frac{v}{4}(1 + 3 \cos(2\theta))$	$\frac{1}{2}(1 + \cos(2\theta)) - \frac{v}{4}(1 + 3 \cos(2\theta))$
$P(\text{DesiredOutput})\text{with post selection}$	$\frac{1}{2}(1 + \cos(2\theta)) - v \cos(2\theta)$	$\frac{1}{2}(1 + \cos(2\theta)) - \frac{v}{2} \cos(2\theta)$	$\frac{1}{2}(1 + \cos(2\theta)) - \frac{v}{2} \cos(2\theta)$

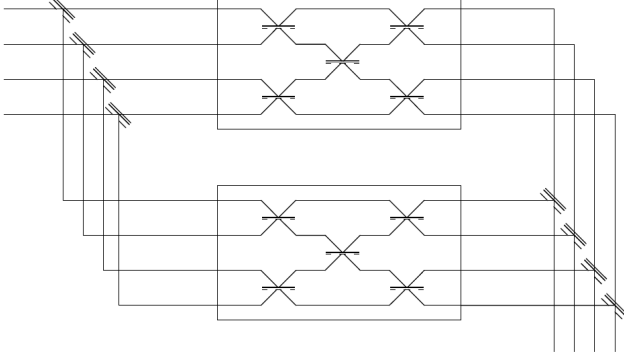


FIG. 7. Diagram of the four mode linear optical network averaged across the system once. A two mode system averaged in this fashion could be considered as an Error Averaged dual rail single qubit unitary transformation.

#### D. 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  $N = 1$ :

$$P_1(\text{correct}) = 1 - \frac{a_1}{b_1}v \quad (52)$$

The probability of an incorrect state is therefore trivially:

$$P_1(\text{wrong}) = \frac{a_1}{b_1}v \quad (53)$$

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 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:

$$\begin{aligned} P_2(\text{correct}) &= 1 - \frac{2a_1 + 1}{2b_1}v \\ &\equiv 1 - \frac{a_2}{b_2}v \end{aligned} \quad (54)$$

$$\begin{aligned} P_2(\text{wrong}) &= \frac{a_1}{2b_1}v \\ &\equiv \frac{a_2}{b_2}v \end{aligned} \quad (55)$$

And so on, so in general:

$$\begin{aligned} P_n(\text{correct}) &= 1 - \frac{2a_{n-1} + 1}{2b_{n-1}}v \\ &= 1 - \frac{(2^{n-1}a_1 + (2^{n-1} - 1))}{2^{n-1}b_1}v \end{aligned} \quad (56)$$

$$\begin{aligned} P_n(\text{wrong}) &= \frac{a_{n-1}}{2b_{n-1}}v \\ &= \frac{a_1}{2^{n-1}b_1}v \end{aligned} \quad (57)$$

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

$$\begin{aligned} P_n(\text{correct} | \text{post selection}) &= \left(1 - \frac{a_1}{2^{n-1}b_1}v\right) \xrightarrow{n \rightarrow \infty} 1 \end{aligned} \quad (58)$$

The probability of success was found to be

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

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

$$\lim_{n \rightarrow \infty} P_n(\text{success}) = 1 - \left(\frac{a_1 + 1}{b_1}\right)v \quad (59)$$

This shows that these correction methods appear to have a very simple effect and may be equivalent. The asymptotic limit in the probability of success shown in Eq.59 is particularly promising as it again demonstrates that there is some limited price for correcting arbitrarily well as shown by Eq.58. This general pattern matches the result shown in Eq.45 with  $N = 2^{n-1}$  which is odd given it does not quite match Eq.49. More explicitly, the results from Table I suggest the error term in Eq.49 should be  $\frac{v}{N}$  which is twice what was found. This is despite being suitable in both the one and two photon cases for the four mode set-up. However as noted, Eq.49 holds for  $N \geq 4$ ,

so if we take  $N = 2$  as our initial system, the ‘initial’ error will taken to be half what it was and these results match exactly. Another potential explanation is that we are intending to implement the identity transformation results in the two photons effectively not interfering at all. It is therefore not all that surprising then that Eq.58 matches the previous single photon results, that is Eq.45, better than the two photon results, that is Eq.49.

This result may also allow the correction of errors in the inner most fixed beam splitters. If the once corrected system with error laden beam splitters is taken 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. As previously stated, averaging across the system also has a lower resource cost suggesting it is a far more preferable method of error correcting.

### E. Phase applying systems

In the previous sections error averaging has been explored by considering the average photon number. It is also necessary to consider the result in a ‘shot by shot’ case. To this end the output of a simple phase applying system is also considered. This will also serve to highlight the differences between the two methods of implementation.

The quantum Zeno effect can be summarised by the saying *a watched pot never boils* or, more scientifically, by repeatedly and regularly measuring a system, it will by necessity remain in an eigenstate of the measurement operator, even if that is not a stationary state of the system [? ]. In this section error averaging after each applied phase shift is compared to averaging across the entire system. To explore this consider a single channel containing  $m$  phasers. This was chosen to better characterise error averaging when applied to a series of error sources as any differences between the two methods of error correction will be more likely to be seen in larger systems. These systems were also used to determine if a quantum Zeno effect also known as the Turing paradox would be observed.

Figure 5 shows the set-up for the phase applying systems.

Following a very similar approach to the earlier sections the applied phase shifters were placed in one arm of a MZ interferometer and photon number expectation values will be calculated. For simplicity the applied phase shift was chosen to be zero with a Gaussian random noise of variance  $v$ . This was chosen such that the errors here match what was modelled in sections III A and III B. The probability of success is now defined as the photon num-

ber expectation value evaluated at the end of the phase applying systems while the strength of the error was initially defined as the photon number expectation value of the MZ interferometer system for the expected output had no error occurred. All approximate results are based of fourth order Taylor approximations with comparisons drawn to the second order approximations as used above.

#### 1. No Averaging

Starting with the baseline comparison case, seen in Figure 5a, where no error averaging is used, the output state for a single photon going through just the phase applying system will simply be

$$|\psi\rangle = \left( \prod_{k=1}^M e^{i\delta_k} \right) |1\rangle \quad (60)$$

The probability of success is then defined as

$$\begin{aligned} P(\text{success}) &= \langle \psi | \psi \rangle \\ &= \langle 1 | \left( \prod_{l=1}^M e^{-i\delta_l} \right) \left( \prod_{k=1}^M e^{i\delta_k} \right) | 1 \rangle \\ &= 1 \end{aligned} \quad (61)$$

which is unsurprising given there is no way, ignoring losses, for the photon to leave this particular system.

To determine a measure of the error the phase applying system was then inserted into a M.Z. interferometer giving a total output state  $|\Psi\rangle$ . The error is then determined by considering the photon expectation value in the correct output mode with post selection. As however the probability of success is unity no post selection will occur. So the output state is

$$|\Psi\rangle = \frac{1}{2} \left( \prod_{k=1}^M e^{i\delta_k} + 1 \right) \hat{a}^\dagger |0\rangle + \left( \prod_{k=1}^M e^{i\delta_k} - 1 \right) \hat{b}^\dagger |0\rangle \quad (62)$$

Now our error when no averaging is applied,  $E_{noAv}$ , will be

$$\begin{aligned} E_{noAve} &= \langle \Psi | \hat{n}_a | \Psi \rangle \\ &= \frac{1}{2} \langle 1 + \cos(\alpha) \rangle \\ &\approx 1 - \frac{Mv}{4} + \frac{M^2v}{16} \end{aligned} \quad (63)$$

$$\begin{aligned} E_{noAve} &= \langle \Psi | \hat{n}_a | \Psi \rangle \\ &\approx 1 - \frac{Mv}{4} + \frac{M^2v}{16} \end{aligned} \quad (64)$$

Where Gaussian statistics have been used to calculate the higher order moments.

## 2. Averaging Across the Entire Phase System

We now consider averaging across the whole system, as shown in Figure 5b. Preceding as before, the state for a single photon after passing through the phase applying system which is being averaged across will simply be

$$|\psi\rangle = \frac{1}{N} \sum_{j=1}^N \left( \prod_{k=1}^M e^{i\delta_{j,k}} \right) |1\rangle \quad (65)$$

The probability of success is then

$$\begin{aligned} P(\text{success}) &= \langle \psi | \psi \rangle \\ &\approx \left[ 1 - \left( 1 - \frac{1}{N} \right) \left( Mv - \frac{1}{2} M^2 v^2 \right) \right] \end{aligned} \quad (66)$$

This result is similar to the what was found in previous sections, see Eq.43 and Eq.50, with the probability of

---


$$\langle \Psi | \hat{n}_a | \Psi \rangle \approx 1 - \frac{1}{4} \left( Mv - \frac{M^2 v^2}{4} + \left( 1 - \frac{1}{N} \right) \left( Mv - \frac{1}{2} M^2 v^2 \right) \right) \quad (69)$$


---

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

$$\langle \Psi | \hat{n}_b | \Psi \rangle = \frac{1}{4} \left\langle 1 + \langle \psi | \psi \rangle - \frac{2}{N} \sum_{j=1}^N \cos(\alpha_j) \right\rangle \quad (70)$$

So the error,  $E_{aveEnd}$ , will be

---


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


---

## 3. Averaging Across Each Phase Shifter Individually

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

$$|\psi\rangle = \frac{1}{N} \prod_{k=1}^M \left( \sum_{j=1}^N e^{i\delta_{j,k}} \right) |1\rangle \quad (71)$$

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

$$P(\text{Success}) \approx \left( 1 - \left( v - \frac{v^2}{2} \right) \left( 1 - \frac{1}{N} \right) \right)^M \quad (72)$$

and an error of

---


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

success asymptotically approaching some fixed value for large  $N$ .

To determine the size of the error, the phase applying system was again inserted into the phase arm of a M.Z. 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

$$|\Psi\rangle = \frac{1}{2} \left( \frac{1}{N} \sum_{j=1}^N \left( \prod_{k=1}^M e^{i\delta_{j,k}} \right) + 1 \right) \hat{a}^\dagger |0\rangle \quad (67)$$

$$+ \left( \frac{1}{N} \sum_{j=1}^N \left( \prod_{k=1}^M e^{i\delta_{j,k}} \right) - 1 \right) \hat{b}^\dagger |0\rangle \quad (68)$$

So first the photon number expectation value for the correct output for this system will be

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. This is simply highlighting that error averaging scales in a very predictable manner.

#### 4. Summary of Errors and Probabilities

Figure 8 shows how the error, as measured by looking at expected photon number values in the output port of a Mach-Zehnder 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 the  $M \approx 6$  point onwards. This could either be suggesting an issue with the quality of the second order approximations, as seen in Figure 9 or that there is some more fundamental point at which it there is a clear benefit to averaging each component individually as opposed to averaging across the entire system. The next consideration was how the probability of success changes with  $M$ .

Figure 9 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 an unrealistically low value of the variance on the individual beam splitters. 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.

It is seen that as the total number of components increases 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 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 something along the lines 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

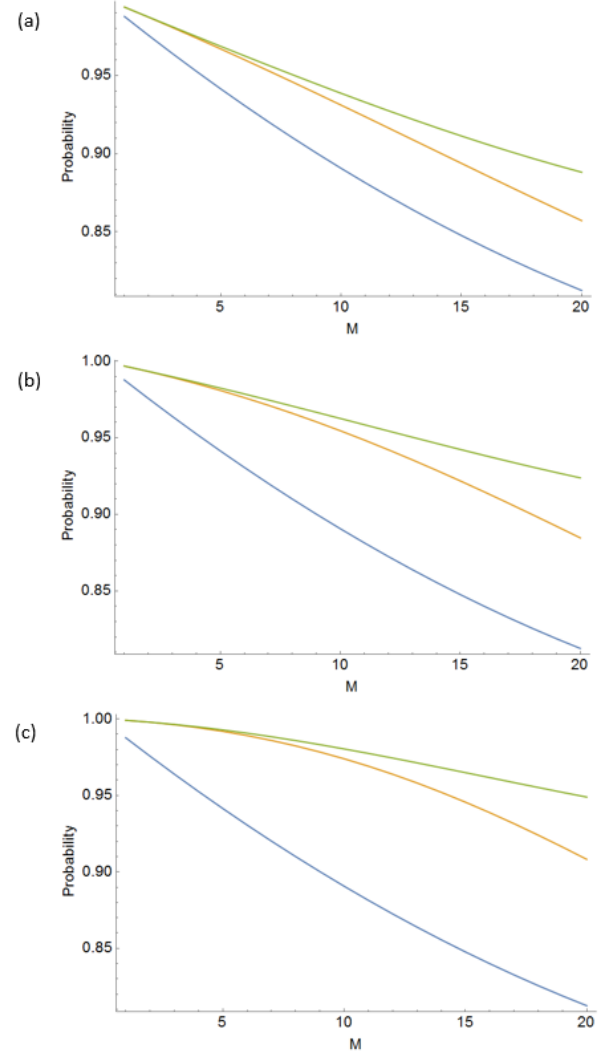


FIG. 8. Probability of obtaining the correct result as measured by the Mach Zehnder interferometer set up as a function of the number of phase components  $M$ . Here a probability of 1 corresponds to no error and the smaller the probability the larger the error. The blue line represents no error averaging applied, the orange line is the error when averaging across the entire system and the green line is the error when each component is averaged across. All three graphs were created with the variance of the error in a single phase shifter being  $0.005 \text{ rad}^2$  and for a)  $N = 2$ , in b)  $N = 4$  and in c)  $N = 16$ .

be seen in Figure 9 (e) where the first order approximation fails almost instantly and the second order approximation soon after while in (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 however is that so too does the variation in the statistical simulation suggesting greater amount of averaging reduces the variability in all aspects of the applied phase.

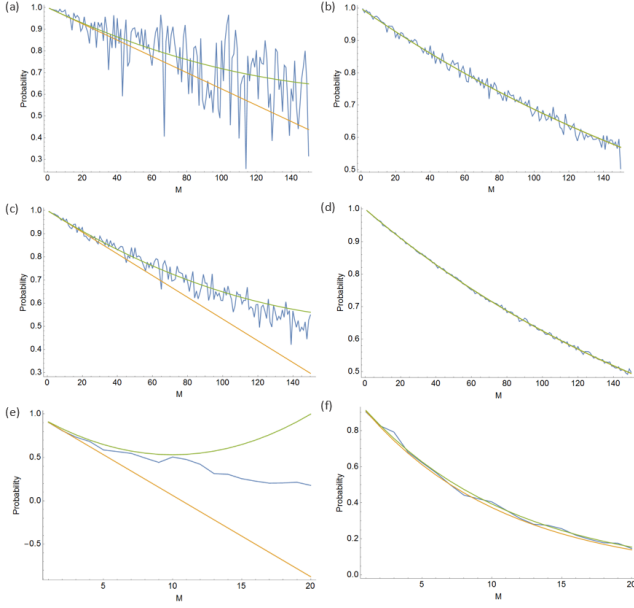


FIG. 9. Probability of success as a function of the number of phase components  $M$  for averaging across the system (left) and averaging across each component (right). Blue line is the statistical model, orange is the first order approximation and green is the second order approximation of the analytical value. a) and b) individual phase shifter variance is  $0.005 \text{ rad}^2$  with  $N = 4$ , c) and d) individual phase shifter variance is  $0.005 \text{ rad}^2$  with  $N = 16$  and e) and f) individual phase shifter variance is  $0.1 \text{ rad}^2$  with  $N = 16$ .

To gain a better insight into how the error behaves and what the difference between the two averaging methods actually is, the actual phase being applied with each system was modelled, as discussed in the following subsection.

#### F. 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 was repeated 5000 times and the results are shown in Figure 10 and Figure 11. This again shows a difference between averaging across the entire system and averaging at each step. In particular the range of applied phases is noticeably smaller when each phase shifter is corrected individually, a clear indication that averaging each step is more effective. By comparing Figure 10 with Figure 8 at  $M = 15$  we can infer that the difference between the two error correction methods seen in Figure 8 is not entirely due to the quality of the approximations used in each case.

To further investigate this behaviour the variance of the applied phase was calculated based on the statistical simulation of the total applied phase. This was then

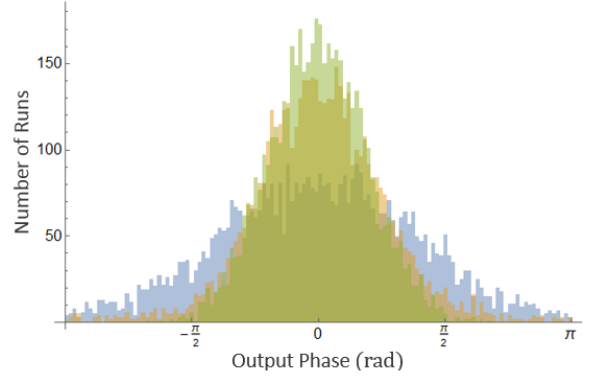


FIG. 10. Top: Total applied phase over 5000 runs for no averaging (Blue), averaging across the entire system (Orange) and averaging each phase shifter individually (Green). Bottom: Histogram of the total applied phases. Each individual phase shifter has a variance of  $0.1 \text{ rad}^2$  and each system has 15 phase shifters in series. The two error averaged circuits are each averaged 4 times.

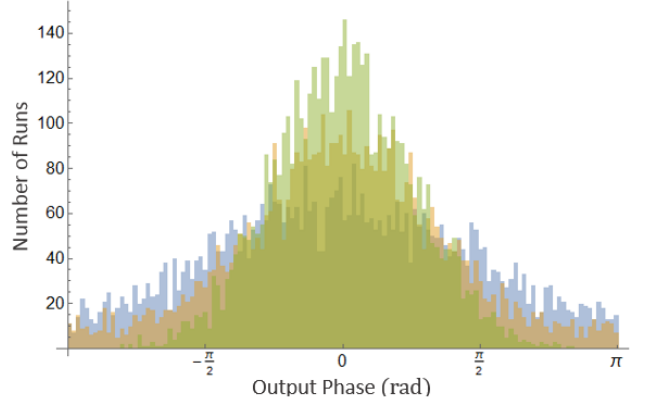


FIG. 11. Top: Total applied phase over 5000 runs for no averaging (Blue), averaging across the entire system (Orange) and averaging each phase shifter individually (Green). Bottom: Histogram of the total applied phase. Each individual phase shifter has a variance of  $0.3 \text{ rad}^2$  and each system has 8 phase shifters in series. The two error averaged circuits are each averaged 4 times.

plotted as a function of  $M$ , the number of phase shifters in a series as shown in Figure 12. Given the individual applied phase are Gaussian random variables they are expected to simply add, such that the variance without any error averaging is expected to simply be

$$\text{Total Variance} = vM \quad (73)$$

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

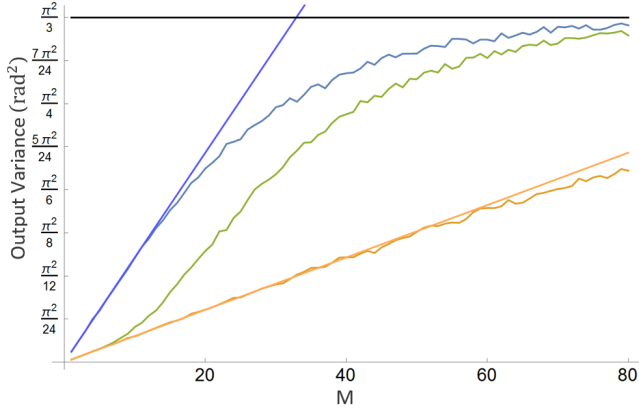


FIG. 12. Variance in the total applied phase without error averaging (Blue), when averaging across the entire system (Green) and when averaging each component individually (Orange), all plotted as a function of the number of phase shifters in series ( $M$ ). The variance of a single phase shifter is  $0.1 \text{ rad}^2$  and the two error averaged systems averaged 4 times. The predicted, linear variance without any averaging (lighter blue) and with averaging (lighter orange) 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. 75, which is also shown in black.

$$\text{Total Variance} = \frac{vM}{N} \quad (74)$$

where  $N$  is the number of times the system is averaged, again  $N = 1$  implies no averaging. This is expected as previously the effect of error averaging was to reduce the error by a factor of  $\frac{1}{N}$  and so a similar behaviour in the variance is likely. As the phase is an angle this behaviour cannot hold for sufficiently large  $vM$ . A completely random phase  $\theta$  is still limited by the possibly range of values, in our case chosen to be  $-\pi < \theta \leq \pi$ . If the value of  $\theta$  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

$$\begin{aligned} \text{Maximum variance} &= \int_{-\pi}^{\pi} \theta^2 P(\theta) d\theta \\ &= \frac{\pi^2}{3} \end{aligned} \quad (75)$$

Figure 12 shows that the two methods of error correction do indeed initially have the same effect. However the averaging across the system method falls out of this linear regime from approximately  $M = 6$  after which it follows the general form of applying no correction. Averaging each step however does not appear to fall out of the linear regime. The down turn at higher total output variance can be attributed to the variance approaching the maximum possible variance and so any behaviour in

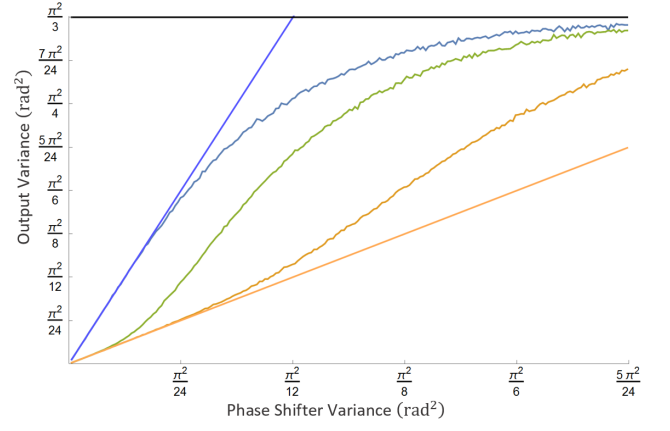


FIG. 13. Variance in the total applied phase without error averaging (Blue), when averaging across the entire system (Green) and when averaging each component individually (Orange), all plotted as a function of the variance in each individual phase shifter. Each system applied 4 phase shifters in series, that is  $M = 4$ , and the two error averaged systems averaged 4 times, that is  $N = 4$ . The predicted variance without any averaging (lighter blue) and with averaging (lighter orange) is also shown along with the maximum allowable variance.

this region can be considered insignificant as the output results generated from a system in this regime is likely to be quite random. To determine if and when the averaging each step method of error correction fails this process was repeated, now 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 13 demonstrated that again the two error correction methods initially are equivalent. Once more averaging across the system falls out of the linear regime and now we can clearly see so too does the averaging each step.

This is highly 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  $0.5 \text{ rad}^2$  when averaging across each element and  $\frac{0.5}{m} \text{ rad}^2$  when averaging across a system of  $m$  phase shifters. If each individual beam splitter has a variance of  $0.1 \text{ rad}^2$ , 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 12. 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 threshold.

#### IV. FEASIBILITY OF ERROR AVERAGING

##### A. A Short Note on Boson Sampling

Having achieved some semi-general results in terms of what benefit can be obtained from error averaging, the next question which arises is; when is it useful? To this end, the behaviour for a Boson Sampling system will be discussed. It is known that as a Boson Sampling system scales, the error terms for each beam splitter combine such that they grow exponentially [?] and as such it provides a good basis system in need of correcting. What is more, Boson Sampling requires the sort of simple linear optical components previously discussed for its implementation and has been shown to demonstrate the benefit of quantum computers over their classical counterparts [?]. And so error averaging could provide a useful tool in the quest for the experimental observation of quantum supremacy. To explore this, a general  $m$  mode system is considered. It would be constructed such that each input mode is connected to each output mode and every path will have the same number of beam splitters and phase shifters. The first requirement was for the system to be considered in some sense general. Note that this system while being general in that 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 it was decided to require  $m$  to be even and 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. This then means that the system is made up of  $c$  columns where  $c = \lceil \log_2(m) \rceil$ .

This system as it currently stands is not suitable for implementing an arbitrary unitary transformation. This is known because for a system to be suitable to simulate  $\mathcal{U}(m)$ , it must have on the order of  $O(m^2)$  beam splitters and phase shifters [?]. This goes back to the number of generators of the  $\mathcal{U}(m)$  group. This system however only has

$$b \times c = \frac{m}{2} \log_2(m) \quad (76)$$

beam splitters and so, if each beam splitter has a single phase shifter than the number of controllable components will be  $m \log_2(m)$  rather than the necessary  $O(m^2)$ . Using such a system we can either average across the system or average each component individually to protect

against errors. This then suggests error averaging, in its current state, is particularly useful for any system in which the result cannot simply be checked to verify the results. The question still remains of how best to implement error averaging. In particular, does the method of averaging across the system indeed always have the same effect as averaging each step? If this is the case it is clearly the more efficient method of implementing error averaging due to the drastic reduction in encoding resources necessary.

##### B. Feasibility

To determine the feasibility and usefulness of error averaging it is necessary to determine both what problems will benefit from error averaging as well as under what circumstances the assumptions which have been made are valid. Namely the assumption that perfect, fixed 50 : 50 beam splitters can be used to build the averaging circuit. 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 [?] however this is outside the scope the current paper. At this point it is worth noting that in the paper *Universal Linear Optics* a single, controllable, beam splitter was implemented using the same Mach-Zehnder interferometer set-up discussed previously with fidelity  $\mathcal{F} = 0.992 \pm 0.008$  [?]. Where fidelity is defined as

$$\mathcal{F} = \sum_i \sqrt{p_i^{exp} \times p_i^{th}} \quad (77)$$

This is useful to consider as it suggests that currently we can have quite high quality, low error 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.

It is known [?] that a system requires  $O(m^2)$  controllable components to be sufficient to implement  $\mathcal{U}(m)$ . 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 [?] 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

$$\begin{aligned} 2xm(N-1) &\ll \frac{mn}{2} \log_2(m) \\ 4x(N-1) &\ll n \log_2(m) \end{aligned} \quad (78)$$

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  $x \rightarrow n \log_2(m)$  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 many 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 [?]. It has also been shown that it is necessary for the error to scale as  $\frac{1}{n \log_2(m)}$  [?] 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. 78 then gives

$$4x(n^2 \log_2(m) - 1) \ll n \log_2(m) \quad (79)$$

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

$$\begin{aligned} 2xm &\ll \frac{mn}{2} \log_2(m) \\ \Rightarrow 4x &\ll n \log_2(m) \end{aligned} \quad (80)$$

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 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

$$4x \ll m \log_2(m) \quad (81)$$

where now  $x \rightarrow m \log_2(m)$  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.

## V. CONCLUSION

This has shown two methods of error averaging that, when combined with post selection, forms a rudimentary error correction protocol. In particular error averaging appears to filter the error free states from the error states which can then be post selected away. The errors 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. This has been seen by analytically determining the photon number expectation values in numerous two mode systems for both one and two photon inputs, by



simulating the output expectation values in four mode systems for both one and two photon inputs and by simulating the variance in the output of a series of beam splitters. From this it can be concluded that this error correcting behaviour can be observed for phase applying systems and Mach-Zehnder interferometer beam splitters. This is also generalised to a four mode system for both single and two photon inputs. Significantly, it also generalises to an arbitrary unitary transformation so can correct any linear optical system although more work is needed to fully characterise this behaviour.

Two similar methods of error averaging have been presented which appear to have similar effects under certain conditions. In particular averaging across the entire system has the same behaviour as averaging each step provided the error within the system is small enough. It has been shown that some hybrid scheme is desirable where the entire  $m$  mode system is decomposed into  $x$  subsystems which are then averaged across independently. Work has begun to identify the thresholds necessary for error averaging to remain in the linear regime which allows an appropriate value for  $x$  to be determined. This has also been related back to one of the main underlying assumptions to gain a loose bound on the amount of

correction that can be applied and the number of subsystems,  $x$ , necessary for error averaging to be effectively implemented. This then leads to the conclusion that error averaging could be very useful given it is an easily implementable method of error correction for small scale system likely to be realised in the laboratory in the near future.

Despite the success of the project so far there are still numerous unsolved problems still to be considered. The first is to fully characterise the effect of averaging an arbitrary unitary transformation. Further work is also necessary to fully understand the effects of the underlying assumptions, namely the quality of the components used, detector inefficiencies and dark counts as well as photon loss throughout the system. This is necessary before any conclusions can be made about the fault tolerance of error averaging. Another potential direction for further research is to determine if traditional methods for protecting against loss [? ? ] could be employed such that post selection is unnecessary and error averaging can be considered a true error correction technique. Finally given the simplicity of the code and the existence of small scale linear optical devices it would be interesting to implement error averaging in the laboratory to better understand its performance under realistic conditions.