Emergence of biased errors in imperfect photonic circuits

Fulvio Flamini

1 Institut für Theoretische Physik, Universität Innsbruck, Technikerstraße 21a, A-6020 Innsbruck, Austria

We study the impact of experimental imperfections in integrated photonic circuits. We discuss the emergence of a moderate biased error in path encoding, and investigate its correlation with properties of the optical paths. Our analysis connects and deepens previous studies in this direction, revealing potential issues for high-precision tests and optical implementations of machine learning.

I. INTRODUCTION

Photonic integrated technologies are a key resource for classical and quantum applications, also including machine learning and artificial intelligence. Nevertheless, the unavoidable effect of imperfections due to fabrication, calibration and control can severely undermine their performance. With path encoding, for instance, a systematic deviation in the scattering probabilities can lead to a biased error in (un)supervised learning or reinforcement learning. While major achievements have recently been reported for photonic circuits, error correction and model training. The inevitability of such imperfections motivates us to shed more light on the above issue.

In this work, we will consider two canonical unitary decompositions and discuss potential vulnerabilities in realistic settings. In principle, these analyses can be readily extended also to other optical architectures. We reveal and quantify a mild systematic bias in their operation, which originates from a biased error in the matrix elements of the unitary transformation. We also show how this feature connects to previously known results, and consider a graph-based approach to tackle similar analyses on noise spreading in an optical mesh. It is not the aim of this work to propose a unique strategy to mitigate these effects, as done very nicely in related works. We aim to understand these issues and bring them closer to the attention in the field, since practical countermeasures will mostly depend on the actual application and physical constraints.

II. FRAMEWORK

In this section, we will first review related results in the context of circuit optimization and characterization, to highlight the main connections and novel contributions. Then, we will introduce the main quantities and models that will be used in our analyses.

A. Related works

The influence of imperfections in photonic circuits has already been investigated in various works. Here we deepen these studies with a focus on aspects that, to the best of our knowledge, have so far eluded a detailed investigation. Specifically, we study how a biased error can emerge in a noisy circuit, addressing precise, practical questions that concern future applications.

In realistic settings, it is paramount to improve the operation of imperfect components. Pioneering work in this scope was done by Miller, who proposed solutions for automatic alignment, configuration and compensation. Recent works have focused on various other approaches, including numerical techniques, port allocation and compilation, and architecture refinement. Our analysis addresses issues related to error propagation in photonic circuits, so that they can be better addressed by the above strategies.

More closely related to this work, there is a series of studies on the properties of phases and transmissivities in triangular (△) and rectangular (□) meshes. Ref. presents a strategy to dial up Haar-random unitary transformations by sampling transmissivities and phases from independent probability density functions. Besides its use in practical applications, this study clarifies how these parameters are distributed, which may affect their noise resilience in a non-homogeneous way. Similar considerations have been put forward in Ref., showing that, in larger circuits, the distribution of reflectivities is increasingly skewed towards low values. Given that a non-ideal mesh may not achieve these low values, the overall fidelity of these circuits would be limited. More recently, notions of sensitivities and error tolerances of optical components have been expanded in Ref., which provides a rich analytic and numerical exploration of these aspects. Our analysis extends the study of unbalanced error tolerance from a different and fine-grained perspective, establishing connections to previous works. Whenever applicable, we will make these connections explicit.

The above works focus on theoretical properties of linear optical circuits and error mitigation. However, non-ideal transformations have been object of detailed analyses also in the context of Boson Sampling (see Refs. and references within), to investigate whether imperfect optical devices preserve classical computational hardness. In this case, the figures of merit are usually global properties of the output probability distribution (e.g. a distance from the ideal distribution). Conversely, the present analysis abstracts from Boson Sampling and targets individual components of the distributions. We will return to this point in Sec. VIB and VIIA.

In Sec. we start by introducing the graph-based approach to tomography.
imperfections are usually modelled by adding Gaussian noise to the ideal parameters. For beamsplitters, this corresponds to sampling a parameter around the ideal value of the transmission amplitude $(1/\sqrt{2})$, with a width given by the state of the art.

Any $m \times m$ $U$ can be decomposed\cite{22,23} in the product of a diagonal matrix $D$ (with complex unit magnitude elements) and $m(m-1)/2$ unitaries $U_k$

$$U = D \prod_k U_k$$

implementing $m$-mode Givens rotations on the planes spanned by two adjacent optical modes

$$U_k = \begin{bmatrix}
1 & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & u_{11} & u_{12} & 0 & \ldots \\
0 & \ldots & u_{21} & u_{22} & 0 & \ldots \\
\vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & 0 & 1
\end{bmatrix},$$

where the $u_{ij}$ ($i,j \in [1,2]$) are the matrix elements of a MZ (see Eq. 1). In the following sections, for many of our analyses we will also represent optical circuits as graphs (see Fig. 1b,c and the next section).

III. PHOTONIC CIRCUITS AS GRAPHS

A convenient way to investigate the effect of imperfections in photonic circuits is to represent them as directed graphs. For $\triangle$ and $\square$ topologies, such graphs are also acyclic: light enters from a subset of the $m$ input nodes, propagates from left to right and leaves the graph from the $m$ output nodes. This approach, recently discussed in a different context in other works\cite{33,34}, facilitates the study of error propagation. In this section we make this intuition more concrete.

A graph representation for the canonical architectures is shown in Fig. 2a ($\triangle$) and 2b,c ($\square$). This approach inspires a number of questions about symmetry: are there more influential nodes contributing to error spreading? Similar questions have already drawn large attention in areas related to information dissemination, contagious disease transmission and social influence\cite{32}. In these cases, nodes with a high centrality index (see Sec. VIIIB) are known to facilitate spreading, even though this depends on many factors. Motivated by this connection, we propose to leverage ideas from interaction networks to study errors in optical networks.

A notion of sensitivity has been introduced for MZs that depends on their position in the mesh\cite{33}. This quantity is linked to the shape of the distributions induced by the Haar measure over the physical parameters\cite{22,23}. Here we show how consistent considerations can be made from a different perspective, by considering the importance of each MZ (node $n$) in the circuit (graph $G$). We investigate two aspects: centrality measures (discussed in Sec.
and properties related to paths in a graph. The latter approach is convenient to study error propagation in waveguide structures and, assuming a homogeneous level of noise over all optical elements, to estimate the amount accumulated in the circuit. If we focus on a node, it is reasonable to expect that its influence increases the more connected it is to the input and output modes. We can even focus on subsets of input/output modes, or assign different weights with different paths. For instance, even if we apply to circuits with arbitrary topologies. Also, while random unitaries, $\Phi_G(n)$ can in principle be introduced in Ref.\textsuperscript{34} and since, for a given $m$, they both depend only on $I_n$ and $O_n$. The main difference is that, while $\alpha_G(n)$ has a solid theoretical support that is lacking in the present discussion, $\Phi_G(n)$ can in principle be applied to circuits with arbitrary topologies. Also, while $\alpha_G(n)$ is an effective and compact tool to study Haar-random unitaries, $\Phi_G(n)$ is in principle compatible with applications where unitaries are not drawn according to this measure. It would be exciting to design an estimator that generalizes to arbitrary topologies and that is more grounded in the theory of interferometric design.

The above analysis focused on individual nodes $n$ inside the circuit. When $n$ is an output node, the analysis concerns the connection between optical paths and the matrix elements $u_{ij}$ that describe a scattering process. Specifically, we can count the number of paths $\pi_{i,j}$ that connect any pair $(i,j)$ of input and output modes. This information allows to reveal the imbalance in the number of tunable elements that influence all $u_{ij}$, within the same topology and for different topologies (see also Sec. \textsuperscript{[IVB]}). We find that this number is symmetric about the main diagonal $(i = j)$, due to the vertical symmetry of the mesh, while $\square$ is almost symmetric about both. As expected, we also find that $u_{ij}$ in $\triangle$ present a strong imbalance w.r.t. paths, while the $\square$ architecture is significantly more balanced. Our aim is then to quantify this intuitive discrepancy, which is one of the reasons why $\square$ is usually considered superior. However, the relative ease of a brute-force approach (convenient to explore arbitrary, small circuits up to $m \sim 30$) comes at the expense of a computationally intensive evaluation, given the ex-

\textbf{Fig. 2. Photonic circuits as graphs.} $\triangle$ and $\square$ architectures can be represented as planar graphs to study error propagation. Here, white nodes represent $m = 12$ input/output modes, colored nodes represent unit cells (two phase shifters each), while edges represent waveguide connections. Different measures of node centrality can be studied and visualized with this approach (see Sec. \textsuperscript{[VII B]}). a) $\triangle$ topology. Colors represent the value of the flow $\Phi$ for each node (see Eq.\textsuperscript{[1]}), normalized by the maximum. Lighter nodes have a higher value, as in panels (d,e). b) $\square$ topology, same color scheme as panel (a). c) Sensitivity measure proposed in Ref.\textsuperscript{34} (d,e) Data from panels (b,c) projected onto a bar chart (using the same ordering, layer by layer).
ponentially large number of paths. Specific tricks (e.g. leveraging symmetries or propagation in light cones) can only mildly extend this range. Hence, we looked for a more efficient way to estimate the number of paths that contribute to the scattering amplitude $u_{ij}$. We find that an exact, analytical solution can be obtained using a special integer sequence given by the Catalan’s trapezoid. We discuss its derivation in Sec. VIIA with further considerations on the asymptotics. We numerically verify its prediction by comparing it to the graph-based approach, up to the largest size that could be probed on a workstation. Fig. 3 shows an example for $m=50$ (no new pattern emerged for larger $m$).

In general, the formula presented in Sec. VIIA - and estimators analogous to Eq. 3 - could be used to quantify the importance of a node for a transition, as well as the asymmetry between transitions. Conversely, the graph-based approach is flexible, can inspire the design of new estimators (see Fig. 4) and supports arbitrary topologies, for which a closed-form expression is not available. Together, these considerations set the stage to discuss how biased errors emerge in single- and multi-photon dynamics in linear optical circuits.

---

**IV. UNIFORM NOISE AND NONUNIFORM DEVIATIONS IN A SCATTERING PROCESS**

Given the difference between $\triangle$ and $\Box$ meshes shown in Fig. 3 and 4, we ask whether this is reflected in the noise sensitivity of the elements $u_{ij}$ of a unitary implemented with these schemes. This analysis aims to reveal a potential bias that has eluded studies based on coarse-grained measures, such as the total variation distance or the fidelity. Indeed, while these measures provide a simple and useful summary of the overall quality of an implementation, they may not reflect the needs of a practical application. For instance, we do not want a classifier with a different accuracy for different objects (or we would like to compensate for this asymmetry). Also, we expect that high-precision applications, e.g. tests in quantum foundation, would benefit from a fine-grained understanding of the noise resilience of the scattering matrix.

---

**A. Single-photon evolution**

To investigate this aspect, we quantify the average deviation of the noisy element $\tilde{u}_{ij}$ when Gaussian noise is added in the architecture (i.e. not directly to $u_{ij}$). The research question we address here is whether the sensitivity of $\tilde{u}_{ij}$ correlates with the patterns found in Fig. 3.
we find that the sensitivity of $u_{ij}$ reflects patterns in the input-output connectivity (see Fig. 3 and 4), but more rigorous analyses are necessary for a full understanding of this correlation. The overall patterns given by $\zeta_{ij}(U)$ in Eq. 7 depend on how parameters are sampled (top row: $\triangle$ meshes; bottom row: $\square$ meshes). For instance, patterns for Haar-random unitaries are due to the distribution of phases in the circuit. We study this sensitivity by numerically generating $2 \times 10^4$ 50-mode unitaries $U$, for $2 \times 10^3$ random initializations of imperfect components each. Tests are carried out using uniformly sampled phases (a, d), Haar-random unitaries (b, e), and the Fourier transform (c, f). Imperfections are modelled with a mild Gaussian noise on the amplitude transmission of the 50:50 beamsplitters ($\mu = 2^{-1/2}$, $\sigma = 10^{-2}$) and on the phase shifters ($\sigma = 10^{-3}$). In all panels, brighter colors correspond to elements that deviate the most from the ideal value.

A positive answer would suggest an explanation for this phenomenon, which is the first step to counteract it. To this end, we consider the following quantity

$$\zeta_{ij}^{(f)}(U) := |1 - 2 \left( 1 + \left| f(u_{ij}) / f(\bar{u}_{ij}) \right| \right)^{-1}|,$$

which is nothing more than an adaptation of the interferometric visibility, for some suitable complex function $f$. Choosing $f_1(z) = \text{Re}(z)$ and $f_2(z) = \text{Im}(z)$ and considering several random imperfect implementations for each $U$, we can compute the estimator

$$\zeta_{ij} := \gamma_1 \left( \zeta_{ij}^{(\text{Re})}(U) \right)_U + \gamma_2 \left( \zeta_{ij}^{(\text{Im})}(U) \right)_U,$$

where $\gamma_{1,2}$ normalize the two terms by their maximum, for both architectures and different unitaries. We consider two ways to sample $U$: according to the Haar measure (uniform over the unitaries, not in the phase space) and by uniformly sampling the tunable phases in the circuit (not uniform over the unitaries). Results reveal a clear imbalance in the sensitivity to a homogeneous Gaussian noise among the matrix elements of $U$ (see Fig. 5).

In the case of uniformly sampled phases (Fig. 5a,d), we find that the sensitivity of $u_{ij}$ (vaguely) negatively correlates with the number of optical paths: the higher the connectivity is (see Fig. 3 and 4), the lower the deviation tends to be as measured by $\zeta_{ij}$ . However, the pattern that emerges in Fig. 5a suggests that a thorough explanation depends also on other factors. From a practical standpoint, this imbalance prompts us to take it into account when unitaries are not sampled from the Haar measure, e.g. when training machine learning models.

In the case of Haar-random unitaries (Fig. 5b,e), a positive correlation with properties of the optical paths seems to apply (almost opposite to the previous case). We attribute this pattern to the different distribution of phases in the circuit, which was already discussed in Fig. 2. In particular, we see that the $u_{ij}$ that exhibit a large average deviation from the ideal value are correlated with those with a large number of paths in the mesh. However, here too patterns are more complex and call for more refined, formal analyses to retrace the actual cause. Specifically, it would be interesting to understand why a slightly brighter region appears in the bottom-right corner of the image for $\triangle$ (or a darker region in the center for $\square$).

The above discussion, which involves averages over sev-
eral unitaries and random initializations, does not say much about specific transformations. Hence, we applied the same analysis to one illustrative example, the Fourier transform, to see whether this is in any way reflected in a single unitary (Fig. 3d). We observe that, despite its symmetry, the sensitivity has the same pattern as the average Haar-random unitary (Fig. 3b,c).

Overall, this shows that the fidelity is not always a suitable metric for high-precision, path-encoded applications. Interestingly, the above patterns survive if we separately plot the two terms in Eq. 7. This means that, for example, multi-photon dynamics would handle noise in a different way, with a marked discrepancy between the most and the least resilient states. This observation will be the object of the next sections.

### B. Multi-photon evolution

So far, we discussed how noise affects the matrix elements of \( U \), which correspond to single-photon scattering amplitudes. Here we sharpen this analysis by studying how Gaussian noise, applied uniformly on the tunable phases, influences multi-photon transitions. We emphasize that this research question is different from previous analyses in the literature, which looked at global properties of the noisy probability distribution. While this issue may not be too relevant for Boson Sampling, it can beinvolved in that process (\(|\{\xi_{vw}\}^{-1}\)|). We are now ready to outline how the analysis proceeds.

The analysis consists of three stages. (i) [Adding noise] We pick a Haar-random \( U \) and uniformly apply Gaussian noise on all tunable phases (i.e. not on the \( u_{ij} \)). The analysis was repeated for several \( U \) and different sizes \( m \), finding a consistent behaviour. (ii) [From \( u_{ij} \) to dynamics] We numerically generate noisy \( N \)-photon probability distributions (\( N = 2, 3, 4 \)) and compute \( \Delta \xi_k \) using Eq. 8 (see Fig. 4b). (iii) [From dynamics back to \( u_{ij} \)] We investigate a possible connection between statistically significant deviations in the probability distribution, tunable phases and \( u_{ij} \) (Fig. 4c).

An example for this analysis is shown in Fig. 6, for \( N = 2, m = 12 \). For a fixed input state \( \nu \), we find no obvious correlation between the output states and the magnitude of the bias \( \delta_{vw} \sigma_{vw}^{-1} \); sometimes the bias is large when \( p_{vw} \) is large, sometimes the converse is true. However, when we consider all input-output combinations as in Eq. 8 and ask how bias manifests itself in \( U \), we retrieve patterns similar to those in Fig. 3 and 4 - this time from yet another mechanism. To see this, in Fig. 6c we plot the sum of \( \Delta \xi_k \) over all parameters \( \xi_k \) that influence \( u_{ij} \), namely

\[
\Delta \xi_k = \sum_{v,w} \left| \delta_{vw} \right| \frac{\Delta \xi_k}{\sigma_{vw}} \cdot \left\{ \xi_k \in \{\xi_{vw}\} \right\}
\]

where \(|\{\cdot\}\)| is the number of elements in the set, \( k = 1, 2, \ldots, m(m - 1) \) and we used the Iverson bracket

\[
\left\{ \xi_k \in \{\xi_{vw}\} \right\} = \begin{cases} 1, & \xi_k \in \{\xi_{vw}\} \\ 0, & \text{otherwise} \end{cases}
\]

Incidentally, \( \delta_{vw}/\sigma_{vw} \) is the inverse of the coefficient of variation of the single deviations that produce \( \delta_{vw} \). The intuition behind this quantity, and more generally about \( \Delta \xi_k \), is that we are interested in the significance of the deviation \( \delta_{vw} \) from 0 (i.e. how many \( \sigma_{vw} \) is \( \delta_{vw} \) far from 0). If there is a relevant systematic bias in the probability distribution, we should be able to observe a preferred (positive or negative) direction for some \( p_{vw} \) when we consider all combinations. This real-valued measure of significance is then assigned to each parameter \( \xi_k \) that contributed to its magnitude via the Iverson bracket, in proportion to the number of parameters that were involved in that process (\(|\{\xi_{vw}\}^{-1}\)|).

This connection seems to suggest that a biased error in multi-photon dynamics is correlated with properties of the paths. However, we observe that the pattern in Fig. 6c matches the distribution of the number of tunable parameters that influence each \( u_{ij} \) (Fig. 4b), providing stronger insights into the underlying error propagation. We emphasize that here we are not simply recombining in \( u_{ij} \) noise that was added to each \( \xi_k \) that influences \( u_{ij} \): in that case, the outcome would be trivial. Instead,
we are first mixing the effect of noise in a quantum process (photons do not evolve independently in the circuit), then we are recombining the (inverse of the) coefficient of variation, which is related to a biased error.

V. BIAS IN THE CALIBRATION PROCESS

Understanding how a circuit operates is a necessary step for practical applications. With tunable circuits, it is especially necessary to calibrate the voltage-to-phase relationship for each thermo-optic phase shifter in the device. In this manuscript, we will use \[ \theta \sim \theta_0 + \alpha V^2 \] and \[ \psi \sim \psi_0 + \beta V^2 \] for phase shifters inside and before the MZ, respectively (see Fig. 1). During a calibration, the goal is then to estimate \((\theta_0, \alpha, \psi_0, \beta)\) for each pair of phase shifters in each MZ. While estimating the transmissivity of directional couplers is also important, here we consider them ideal to isolate the contribution related to the phase.

Bayesian estimation and machine learning are promising approaches to calibrate a device without detailed information on the internal structure (see e.g. Refs. 46,47 and references within). In the case of \(\triangle\) and \(\Box\) meshes, instead, calibration can be carried out by directing light to each MZ in a specific order. A strategy to calibrate the \(\Box\) mesh using knowledge about the structure was briefly sketched in Ref. 17–19,48,49. We propose an algorithm in Sec. VIIC, which we use for our discussion on errors (see Fig. 7). In this scheme, parameters are estimated in two main steps: \(i\) all MZs are probed in sequence (Fig. 7a), returning an estimate for \((\theta_0, \alpha)\); \(ii\) using

FIG. 6. Sensitivity and multi-photon transitions. We numerically investigate a systematic deviation from the ideal values of \(N\)-photon transition probabilities. a) Three-stage description of the analysis (see Sec. IVB). Top panel: (example of) ideal two-photon output (collision-free) probability distribution for a 12-mode Haar-random unitary \(U\) from input modes \(v = (1, 2)\). Middle: for each probability \(p_{vw}\) (gray bar), we plot the deviation \(\delta_{vw} = \langle \tilde{p}_{vw} \rangle - p_{vw}\) (black points) with standard deviation \(\sigma_{vw}\), over an orange zero baseline. The take-home message is distilled in the bottom panel. Bottom (mirrored): for each \(p_{vw}\), we plot how many \(\sigma_{vw}\) separate \(\delta_{vw}\) from 0. Some gray and fuchsia peaks seem to correlate but this is not always true. For \(m = 12\), \(\delta_{vw}/\sigma_{vw} = 1.49 \pm 0.18\), averaged over all input-output pairs and 200 random initializations of unitaries with Gaussian noise as in Fig. 5 b) \(\Delta \xi_k\) for all nodes in the \(\triangle\) and \(\Box\) meshes. This pattern is partly due to the fact that cells closer to the output are reached by more input states. c) \(\Delta_{ij}\) normalized to the maximum, for all matrix elements \(u_{ij}\). We recover patterns similar to Fig. 3 and 4 for both \(\triangle\) (left) and \(\Box\) (right) meshes. d) Number of tunable parameters that influence each \(u_{ij}\), normalized to the maximum (see also Fig. 9). This pattern matches the one in panel (c).
FIG. 7. Calibration and noise. The voltage-to-phase dependency for the □ mesh can be inferred in two stages (a,b). a) Light is injected in first input mode and measured in the last output, to characterize the diagonal A. This step is repeated sequentially for the other diagonals. b) Phase shifters outside the MZs are calibrated by constructing larger MZs consisting of two MZs in the 50:50 state (fuchsia) interleaved with and followed by MZs in the bar state (blue). c) The order chosen by the calibration induces a pattern in the noise distribution. In this example, colors describe the mean deviation of the estimated $\psi_0$ (left) and $\beta$ (right) from the actual values (bright: high; dark: low), averaged over $10^3$ calibrations of Haar-random unitaries.

this information, all MZs are adjusted to either the bar or the 50:50 state (Fig. 7c) to estimate $(\psi_0, \beta)$. After this step, additional tricks can be used to improve the estimates. Our main observation, here, is that there is a systematic bias in the uncertainty and accuracy associated with every parameter. This bias is mainly induced by the sequential nature of the calibration, since estimating the latter parameters requires (imperfect) knowledge of several others. In Fig. 7c, for instance, we show how the average error associated to each MZ correlates with its position in the optical mesh, which in turn influences the sequential procedure of the calibration. This issue can become relevant in large-size circuits.

VI. DISCUSSION

Classical and quantum optical technologies provide a promising platform to enable faster and energy-efficient machine learning. Among these technologies, integrated circuits represent a well-established resource with an intuitive operation, good control and advanced miniaturization. In this context, while path-encoding is generally a convenient approach to process information, we argue that it is also necessary to counteract the path-dependent noise hidden in the process. This issue manifests itself in an unbalanced control, both in accuracy and precision, over the map between input and output states, which is potentially relevant in tasks such as classification, reinforcement learning or tests on the foundations of physics.

In this work, we shed light on this aspect by considering single-photon and multi-photon evolutions in triangular and rectangular architectures. We numerically simulate these dynamics in realistic settings, collecting evidence that the impact of noise correlates with properties of the optical paths, as well as with the number of tunable elements they involve. We then present a conjecture for an exact, closed-form expression to compute the number of paths for all input-output combinations and arbitrary size. The validity of this conjecture was verified up to a large circuit size using a graph representation, which we consider of independent interest for several reasons. In fact, we envision the development of techniques to estimate the relevance of nodes in an optical mesh, as done in other fields such as social network analysis. We make a first step also in this direction, reporting on preliminary investigations to identify the most influential regions for error propagation. While some of these tests are inconclusive, we believe that more can be understood by looking carefully with these techniques. This can be especially true for large circuits with arbitrary topology, where special sets of nodes can act as bottlenecks or take on a more influential role for a given application.

[Code] for all the analyses presented in this manuscript is publicly available online. We conclude with extra remarks on the applicability of this work (see Table I). On the one hand, the positive side is that the imbalance revealed by these analyses seems to be of small magnitude. This means that the above concerns are practically not relevant for most applications today. We leave an absolute estimate of its impact to more specialized studies, which focus on individual technologies and implementations. On the other hand, as the circuits size and the desired accuracy increase, it is reasonable to expect that these concerns will become more relevant in the future. In the highly plausible scenario where optical technologies become a go-to platform for machine learning, similar analyses would also be useful to build trust with the end consumers and funding agencies. We hope this research will help make progress also in this sense.
We thank Sebastian Gstir and Robert Keil for their very careful and constructive feedback, which helped improve the manuscript, and Caterina Taballione, Jehner J. Renema and Jörn P. Epping for fruitful discussions. We also acknowledge support and stimulating discussions with the Quantum Information & Computation group and the Photonics group at University of Innsbruck.

The project leading to this application has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 885567 and No 801110, and from the Austrian Federal Ministry of Education, Science and Research (BMBWF). It reflects only the author’s view, the EU Agency is not responsible for any use that may be made of the information it contains. The project has received partial support from the Austrian Science Fund (FWF) through the project SFB BeyondC F71.

VII. APPENDIX

In the following sections, we study three aspects that were only briefly mentioned in the main text. Sec. VII A presents an exact and closed-form expression to compute the number of optical paths inside the canonical meshes. This information should facilitate the development of measures to counteract the influence of biased errors. Sec. VII B presents considerations on a strategy to estimate the influence of a node on error propagation. Finally, Sec. VII C completes the discussion started in Sec. IV presenting a pseudocode to calibrate the □ mesh. This can be useful in practical applications, as well as to understand how a bias originates during the process.

A. Counting optical paths

Here we present analytical expressions to compute the exact number of paths that light can take from mode $i$ to $j$ in an $m$-mode $\triangle$ and $\square$ optical meshes. We denote this set of paths as $\{\pi_{ij}^m\}$ and its cardinality as $|\{\pi_{ij}^m\}|$. This information is relevant to estimate the susceptibility of each input-output transition to circuit imperfections. We also present analytical expressions for the exact number of paths that connect one input MZ (i.e. a MZ in the first vertical layer) to any MZ inside the $\square$ architecture. This quantity can be used in conjunction with the flow of a MZ (see Eq. 4) to assess its criticality in a large mesh. Further work is necessary to connect this result to applications with specific requirements.

1. Paths from input to output

For a fixed mesh, we expect $|\{\pi_{ij}^m\}|$ to depend on the horizontal layers to which $i$ and $j$ belong (nodes in the same input/output MZ have the same number of paths) and on the parity of $m$, since $\square$ has horizontal symmetry when $m$ is odd. Below we present an analytic expression that indeed fulfills these observations.

Given its formulation, we approach the problem using combinatorics. We find by inspection [13] that a solution can be expressed using Catalan’s trapezoid [13] $C_s(a, b)$, a generalization of the Catalan’s triangle (in turn generalizing the Catalan numbers, defined using binomial coefficients). A closed-form expression for $C_s(a, b)$ is

$$ C_s(a, b) = \begin{cases} \binom{a + b}{b}, & 0 \leq b < s \\ \binom{a + b}{b} - \binom{a + b}{b - s}, & s \leq b \leq s + a - 1 \\ 0, & b > s + a - 1 \end{cases} $$

(11)

where $s = 1, 2, 3, \ldots$, $a = 0, 1, 2, \ldots$ and $b = 0, 1, 2, \ldots$

We find that, by introducing the compact notation

$$ \sigma_{ij}^m := \frac{1}{2} \text{sgn} \left( \frac{1}{2} i + j - m \right) $$

$$ T_{ij}^m := -i + j + \frac{1}{2} (-1)^i + \frac{1}{2} (-1)^{m+j} $$

(12)

the combination $(S_{ij}^m, A_{ij}^m, B_{ij}^m)\square$, with

$$ S_{ij}^m := \frac{m + 2}{2} + \sigma_{ij}^m (m - 2i + (-1)^i + 1) $$

$$ A_{ij}^m := \frac{m - 1}{2} - \sigma_{ij}^m T_{ij}^m $$

$$ B_{ij}^m := \frac{m - 1}{2} + \sigma_{ij}^m T_{ij}^m $$

(13)

indeed provides the answer to our counting problem

$$ |\{\pi_{ij}^m\}|^\square = C_{S_{ij}^m}(A_{ij}^m, B_{ij}^m) $$

(14)

Recalling that Eq. 11 is a piecewise function over three regions, we observe that the last condition is always satisfied for all values given by Eq. 13 due to the constraints of the problem ($1 \leq i, j \leq m$). From a physical perspective this makes sense, since $|\{\pi_{ij}^m\}|^\square$ cannot be 0.

We do not know why Eq. 14 is true, but we numerically verified its correctness for all $i, j \in [1, m]$ and $m \in [1, 27]$. We leave the pleasure of the proof to the curious and more
skilled reader. For circuits with larger size ($m \sim 30$) it becomes computationally challenging to list all paths on a workstation with 36 cores and 128 GB of RAM.

By a similar counting argument, choosing

$$\begin{align*}
(S_i, A_i^m, B_j^m)_{\triangle} &= (i, m - i, m - j)
\end{align*}$$

we also find the expression for the $\triangle$ mesh

$$
\left\{\pi_{ij}^m\right\}_{\triangle \triangle} = \left(\begin{array}{c}
2m - i - j \\
m - j
\end{array}\right) - \left(\begin{array}{c}
2m - i - j \\
m
\end{array}\right)
$$

Notice that, due to the symmetry rule for binomial coefficients, Eq. 16 is symmetric in $i$ and $j$, as expected.

We emphasize that $\left\{\pi_{ij}^m\right\}_{\square}$ and $\left\{\pi_{ij}^m\right\}_{\triangle}$ provide the exact number of paths between modes ($i$, $j$). Using these quantities, we can quantify the asymmetry between any single-photon transitions. We can also combine this information to study multi-photon processes with distinguishable photons (since they evolve independently), while correlations may be found in dynamics that involve quantum interference. Indeed, we recall that each $u_{ij}$ depends only on a subset of tunable parameters, related to the waveguide connectivity in the circuit (see Fig. 9).

We can also use Eq. 14 and 16 to quantify the asymmetry between architectures w.r.t. the number of paths that connect two modes. This information can be useful since $\left\{\pi_{ij}^m\right\}_{\square}$ correlates with the resilience to noise, which can be very different in the two cases. For instance, for $m$ large and $i = j = 1$, using the Stirling approximation for the binomial coefficient we obtain the scaling

$$
\begin{align*}
\frac{\left\{\pi_{11}^m\right\}_{\triangle}}{\left\{\pi_{11}^m\right\}_{\square}} &\sim \frac{\sqrt{\frac{2m - \frac{3}{2}}{m}}}{\Gamma(\frac{3}{2})} \\
\end{align*}
$$

which shows an exponential separation between the two architectures (see also Fig. 3). Similarly, when $i = 1$ and $j = \left\lfloor \frac{m}{2} \right\rfloor$ we can use

$$
\begin{align*}
\begin{pmatrix} a + b \\ b \end{pmatrix} \sim \sqrt{\frac{a}{\pi(a - b)}} b^a (a - b)^{a - \frac{1}{2}},
\end{align*}
$$

which works when both $a$ and $b$ are large, and $\Gamma(x) \sim \sqrt{2\pi x^{x^{\frac{1}{2}}}} e^{-x}$, to get

$$
\begin{align*}
\frac{\left\{\pi_{11}^m\right\}_{\triangle}}{\left\{\pi_{11}^m\right\}_{\square}} &\sim e^{w m - 1},
\end{align*}
$$

with $w = \log 3/8 + 5/4 \log 3 \sim 0.39$. We tested the approximations in Eq. 17 and 19 in the regime $30 < m < 300$, finding a very good agreement with the exact value given by Eq. 14 and 16. Together, they provide quantitative understanding of the asymmetry between the two architectures. This knowledge can guide the choice of the most appropriate platform for a given application that needs to fulfill specific physical constraints, as well as to correct for this inherent imbalance.

2. Paths from a Mach-Zehnder to input and output

Here we discuss how to compute the number of paths that connect any MZ in the $\square$ architecture (located in layer $l$, $q$-th position from top) with the input and the output layers of the circuit. In the representation based on a directed graph, we would count in how many ways we can reach a node from the input $|\{\pi_{ij}^l,q\}_{IN}|$, as well as in how many ways $n$ can reach the output $|\{\pi_{ij}^l,q\}_{OUT}|$. Intuitively, this corresponds to the number of paths that lay in the light cone of a MZ (summarized by $I_{in}$ and $O_{in}$ in Sec. 11). This information can help evaluate to what extent a MZ affects scattershot-like applications, where light is injected and measured in all input and output modes. We will return to this point in Sec. VII B.

We tackle this task by considering the number of paths $|\{\pi_{ij,l,q}^1\}|$ that connect a MZ in the first layer ($q$-th from top) with any MZ in the circuit ($l \geq 1$, $q$-th from top).

FIG. 8. $\square$ and number of paths. Logarithm of the number of paths $|\{\pi_{ij}^m\}|$ that connect mode $i$ to mode $j$ in the $\square$ mesh using Eq. 14. Three surfaces are shown for a different size of the mesh ($m = 20, 40, 60$). The same plot for $\triangle$ would exhibit peaks on one of the corners, as in Fig. 9.
Notice that each MZ in the first layer can be reached by two input modes, so a factor of 2 may be introduced to describe the problem in terms of modes. In both cases, we consider again the Catalan’s trapezoid in Eq. 11 and look for a combination \((s_{l,m}^{Aq_q}, A_{q_q}^{l,m}, B_{q_q}^{l,m})\) that fits the problem. Here, the superscript \(G\) serves to remember that we are describing the problem in terms of nodes in a graph. After some experimentation, we find by inspection that the following expressions

\[
S_{q_q}^{l,m} := \frac{m + 2}{2} + \sigma_m^{q_q} \left( m - 2q' + (-1)^{q'} - (-1)^l \right)
\]

\[
A_{q_q}^{l,m} := \frac{l - 1}{2} - \sigma_m^{q_q} \left( T_m^{q_q} - (l - 1) \mod 2 \right)
\]

\[
B_{q_q}^{l,m} := \frac{l - 1}{2} + \sigma_m^{q_q} \left( T_m^{q_q} - (l - 1) \mod 2 \right)
\]

solve the problem for circuits with odd size \(m\), i.e.

\[
|\{\pi_{l,m}^{q_q}\}| = G_{S_{q_q}^{l,m}}(A_{q_q}^{l,m}, B_{q_q}^{l,m}). \tag{21}
\]

Similar considerations lead to an analogous of Eq. 21 for even \(m\) and for the \(\triangle\) architecture.

Indeed, by using Eq. 11 it is not hard to see that for \(l = 1\) we have \(|\{\pi_{l,m}^{q_q}\}|_N = 0\), since \(A_{q_q}^{1,m} = -B_{q_q}^{1,m} = \sigma_m^{q_q} T_m^{q_q}\) (no path connects two MZs in the same layer). In addition, when \(l = m = 2r + 1\) \((r = 0, 1, 2, \ldots)\) we retrieve the same expression in Eq. 11 as desired. As mentioned, the only difference is that Eq. 11 refers to input-output modes while here we refer to MZs. To match Eq. 11 and Eq. 20 it is sufficient to observe that \(q = \left\lfloor \frac{r+1}{2} \right\rfloor = 1, 2, \ldots, \frac{m}{2}\) and \(q' = \left\lfloor \frac{r+1}{2} \right\rfloor = 1, 2, \ldots, \frac{m}{2}-1\).

By using Eq. 21 (odd size) and the symmetry of the problem, we can now evaluate

\[
|\{\pi_{l,m}^{q_q}\}|_N = \sum_{q'=1}^{m-1} C_{S_{q_q}^{l,m}}(A_{q_q}^{l,m}, B_{q_q}^{l,m}), \tag{22}
\]

\[
|\{\pi_{l,m}^{q_q}\}|_{OUT} = \sum_{q'=1}^{m-1} C_{S_{q_q}^{l,m}}(A_{q_q}^{l,m}, B_{q_q}^{l,m}),
\]

which provides a baseline to study the centrality of a node. This task will be the focus of the next section.

B. Centrality measures in an optical mesh

In this section, we turn our attention to the impact of individual MZs in realistic implementations. The objective is to identify the most important nodes within the optical meshes, according to a suitable real-valued measure of importance. To this end, motivated by the analogy between light propagation in a noisy circuit (where each MZ can introduce noise into the evolution) and information spreading, we will consider tools from graph theory and network analysis, where this notion of importance is generally called centrality. The research question we ask is whether network analysis can help understand what noisy MZs are more likely to influence the performance of a circuit, without any information on the underlying quantum process. In other words, is it possible to infer the influence of a MZ by just looking at the network topology?

Centrality measures rank nodes according to their flow (for different types of flow) or to their involvement in the cohesiveness of the network. They are only good within the application they were designed for, and they are likely to disagree otherwise. These rankings provide the order of the most important nodes (although not the relative importance), while little can be said for the majority of the others. This limitation has led to the development of measures to assess the influence of all nodes in a network, e.g. the accessibility (how accessible the network is from a node using random walks) and the expected force of infection. This metric is more related to the flow discussed in Sec. VIIA and in Eq. 1. Another aspect of centrality measures is that they operate from a node-centric perspective, i.e. they don’t consider the synergy between neighbor nodes (which plays a major role in disease and information spreading). In this case, game-theoretic approaches are usually more effective. Centrality and node influence can be correlated but they are in general distinct.

Given these considerations, we computed various measures of centrality for each MZ\(^{29}\): closeness centrality, betweenness –, degree –, eigenvector –, Katz –, PageRank –, HITS –, radiality –, status –, edge betweenness –, and LinkRank –. Results for some of them - those with the most interesting behaviour - are shown in Fig. 10. Contrary to our expectation, we did not find a strong correlation with the reported results.

We also carried out additional numerical simulations to single out the most influential MZs, by adding realistic levels of Gaussian noise on each phase shifter one by one (assuming all the other elements are ideal or nearly ideal). In all these cases, a Monte Carlo simulation using the Fidelity as figure of merit did not reveal any consistent pattern of resilience: the reduction in the Fidelity was too small, comparable to the average fluctuation. We also simulated an imperfect implementation where the error on each phase was proportional to the magnitude: in this case, we indeed recovered the expected pattern due to the phase distribution for Haar-random unitarities.\(^{23}\) We believe that more informative analyses should be tailored to specific architectures, to circumvent the bias induced by the Haar measure.

C. Calibration

In this section we outline an algorithm to calibrate \(\square\) meshes. While the idea was already sketched in previous works\(^{17,19,25,29}\), to the best of our knowledge no detailed procedure was proposed that is amenable to immediate reuse and improvement. Understanding this operation is
FIG. 10. **Centrality of a Mach-Zehnder.** Centrality measures of each node of the optical mesh are displayed with a color scheme, from low (green) to high (yellow), for the △ (a) and □ (b) meshes. Absolute values of the centrality are not relevant.

also useful to evaluate the results in Fig. and gauge how different approaches may induce different errors.

In principle, calibration could also be seen as an optimization problem, where one looks for the set of parameters that best reproduces all measurements, according to a suitable figure of merit. To get a rough understanding of the feasibility of this approach, we attempted a calibration with a simple, unpretentious least-squares optimization (LSO). Specifically, we (i) considered a □ circuit of increasing size $m$, (ii) simulated single-photon ($P_1$) and two-photon ($P_2$) output probability distributions for a set of random parameters $\{\xi\}$, and (iii) compared $P_1$, $P_2$ with some $P'_1$ and $P'_2$ obtained by adding Gaussian noise on both transmissivities and phases. Before this preliminary test, our guess was that, for sufficiently small $m$ and low noise, a LSO would return a good estimate of the noiseless parameters. Contrary to our expectation, we found that, even though LSO does a very good job in minimizing the residuals, it returns parameters that can be quite different from the original ones. In addition, both the quality of the result and the required evaluation time become rapidly unfavorable for circuit sizes as low as $m \sim 7$ (the number of parameters scales quadratically in $m$). While we emphasize that one cannot draw any conclusive consideration from this simple test, we gain evidence that (a) specialized strategies are likely to perform significantly better, and (b) simple LSO-like calibrations can fail due to the large number of parameter combinations that approximate the desired output. Importantly, (c) using the output of a LSO to study specific input-output scattering processes may result in large errors, since $\{\xi\}$ was estimated from a global optimization problem.

The proposed approach consists of two parts (see Algorithm 1). (i) The first two loops serve to characterize the phase shifters in each MZ, by sequentially focusing on the elements below and above the main diagonal of the circuit, respectively. Specifically, each iteration selects a path in the circuit and calls one subroutine (Algorithm 2) that characterizes the phase shifters inside the MZs by LSO. This step reveals what voltages correspond to the bar ($V^-; \text{MZ}_-$), cross ($V^\times; \text{MZ}_\times$) and balanced ($V^{50}$) states, which allow to guide light to specific regions of the circuit. Some of these steps can commute and can be parallelized for a faster execution on large circuits. (ii) The last two loops in Algorithm 1 provide information on the phase shifter before each MZ. First, the linear dependency on the applied voltage is characterized (yielding the $\beta$ coefficients). Once these are known, we can remove the voltage and focus on the offsets $\psi_0$, which completes the calibration.
To test the algorithm, we numerically simulated a full calibration for (several random instances of) noisy settings and circuits with size up to \( m = 15 \) and realistic levels of Gaussian noise. For the sake of convenience, we make three assumptions: (a) we assume \( 0 \leq \psi_0^{(k)} \leq \frac{\pi}{2} \) (to remove some ambiguity in the determination of the offset); (b) we assume that it is possible to inject light from pairs of input modes with a known phase relationship (which we set to zero w.l.o.g.); (c) directional couplers are ideal (an imperfect fabrication can be included in Algorithm 1). Condition (b) is illustrated in Fig. 7 by means of an extra beamsplitter before the circuit, which splits the input light into two balanced paths that recombine in one of the MZs, thus forming a larger MZ.

**Algorithm 1** Calibrations

**Input:** Graph \( G \) for a rectangular architecture of size \( m \).

**Output:** An estimate of the voltage-to-phase relationship for all tunable phase shifters.

1: \( N_{MZ} \leftarrow \frac{m(2m-1)}{2} \)
2: MZpath, MZ\(_{-}\) \( \leftarrow \) Empty lists
3: Models\(_{\psi} \leftarrow \frac{1}{2} + s \cos(2\psi_0 + 2\beta V_\psi) + t \sin(\psi_0 + 2\beta V_\psi)\)
4: for \( d \leftarrow 1 \) to \( \left\lfloor \frac{m-1}{2} \right\rfloor \) do
5: \( MZ_- \leftarrow MZ_- \cup MZpath \)
6: MZpath \( \leftarrow \) Shortest path in \( G \) from \( 2d - 1 \) to \( N_{MZ} \)
7: FitPath(\( 2d - 1, m \))
8: end for
9: for \( d \leftarrow 1 \) to \( \left\lfloor \frac{m-1}{2} \right\rfloor \) do
10: \( MZ_- \leftarrow MZ_- \cup MZpath \)
11: MZpath \( \leftarrow \) Shortest path in \( G \) from \( 1 \) to \( N_{MZ} - 2d + 1 \)
12: FitPath(\( 1, m - 2d + 1 \))
13: end for
14: for vertical layer \( l_v \) in \( G \) do
15: for all \( k \in \{1, N_{MZ}\} \) do
16: \( V_k^{(\theta)} \leftarrow \begin{cases} V_{k,0}^{\theta}, \ k \in l_v \\ V_k^{\theta}, \ \text{otherwise} \end{cases} \)
17: \( V_k^{(\psi)} \leftarrow \begin{cases} V_{\psi}, \ k \in l_v \\ 0, \ \text{otherwise} \end{cases} \)
18: end for
19: for horizontal layer \( l_h \) in \( G \) do
20: Inject light in \( l_h \) and measure output distribution.
21: \( \beta_{V_k}^{(\theta)} \leftarrow \) Scan \( V_{\psi} \) and fit data using Model\(_{\psi} \).
22: end for
23: end for
24: for vertical layer \( l_v \) in \( G \) do
25: for all \( k \in \{1, N_{MZ}\} \) do
26: \( V_k^{(\theta)} \leftarrow \begin{cases} V_{k,0}^{\theta}, \ k \in l_v \\ V_k^{\theta}, \ \text{otherwise} \end{cases} \)
27: \( V_k^{(\psi)} \leftarrow 0 \)
28: end for
29: for horizontal layer \( l_h \) in \( G \) do
30: Inject light in \( l_h \) and measure output distribution.
31: \( \psi_{k}^{(\theta)} \leftarrow \) Approximate distribution using Model\(_{\psi} \).
32: end for
33: end for

**Algorithm 2** FitPath — Pseudocode for the subroutine called by Algorithm 1 to calibrate the diagonals.

**Input:** Input mode \( I \); Output mode \( O \)

**Output:** \((\theta_0^{(k)}, \alpha_0^{(k)}, V_0^x, V_0^y)\) for each MZ \( k \).

1: \( MZ^\times \leftarrow \) Empty list
2: Models\(_{\theta} \leftarrow \frac{1}{2} (1 + \cos(2\theta_0 + 2\alpha V_\theta)) \)
3: FitMZ(b) \( \leftarrow 1 \). Inject light in \( I \); 2. Measure light from \( O \); 3. Scan voltage \( V_\theta \) in MZ \( b \); 4. Fit data using Model\(_{\psi} \).
4: for \( b \in \text{Reverse}(MZpath) \) do
5: \( MZ^\times \leftarrow MZ^\times \cup \{b\} \)
6: for all \( k \in \{1, N_{MZ}\} \) do
7: \( V_k^{(\theta)} \leftarrow \begin{cases} 0, \ k \notin MZ^- \cup MZ^\times \\ V_k^-, \ k \in MZ^- \land k \neq b \\ V_k^x, \ k \in MZ^\times \land k \neq b \end{cases} \)
8: \( V_k^{(\psi)} \leftarrow 0 \)
9: end for
10: FitMZ(b)
11: Evaluate \( V_k^-, V_k^x, V_k^b \) by inverting Model\(_{\psi} \).
12: Append values \( V_k^-, V_k^x, V_k^b \) to lists \( V^-, V^x, V^b \).
13: end for

In our numerical tests, Algorithm 1 and 2 retrieved very good estimates of the actual (unknown) voltage-to-phase relationship, for different \( m \) and levels of noise, with a fully automated procedure. We also observe how the quality of the characterized unitary (with no voltage) decreases for larger \( m \), especially since errors propagate through the sequential calibration. These errors can be mitigated in various ways: (\( i \)) by repeating the procedure multiple times, choosing a different order where possible, and averaging the estimates; (\( ii \)) using optimization algorithms, e.g. particle swarm optimization, starting from the parameters given by the calibration (the Fidelity can be used a simple loss function); (\( iii \)) exchanging the input and the output and averaging the estimates.

**REFERENCES**

1. J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, “Integrated photonic quantum technologies,” Nat. Photonics 14, 273-284 (2020).
2. W. Bogaerts, D. Pérez, and J. Capmany, “Programmable photonic circuits,” Nature 586, 207-216 (2020).
3. G. Wetzstein, A. Ozcân, S. Gigan, et al. “Inference in artificial intelligence with deep optics and photonics,” Nature 588, 39-47 (2020).
4. B. J. Shastry, A. N. Tait, T. Ferreira de Lima, et al. “Photonics for artificial intelligence and neuromorphic computing,” Nat. Photonics 15, 102-114 (2021).
5. P. Hollis, J. S. Harper, and J. Paulos, “The effects of precision constraints in a backpropagation learning network,” Neural Computation 2, 363-373 (1990).
6. M. Y.-S. Fang, S. Manipatruni, C. Wierzych, et al. “Design of optical neural networks with component imprecisions,” Opt. Express 27, 14009-14029 (2019).
7. Y. Shen, N. Harris, S. Skirlo, et al. “Deep learning with coherent nanophotonic circuits,” Nat. Photonics 11, 441–446 (2017).
