

Learning Interpretable Representations of Entanglement in Quantum Optics Experiments using Deep Generative Models

Daniel Flam-Shepherd,^{1,2,*} Tony Wu,¹ Xuemei Gu,³ Alba Cervera-Lierta,^{1,4} Mario Krenn,^{1,4,2,5,6,†} and Alán Aspuru-Guzik^{1,4,2,7,‡}

¹*Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada*

²*Vector Institute for Artificial Intelligence, Toronto, Ontario M5S 1M1, Canada*

³*Hefei National Laboratory for physical Science at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China,*

⁴*Department of Chemistry, University of Toronto, Toronto, Ontario M5G 1Z8, Canada*

⁵*Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria*

⁶*Max Planck Institute for the Science of Light (MPL), 91058 Erlangen, Germany*

⁷*Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada*

Quantum physics experiments produce interesting phenomena such as interference or entanglement, which is a core property of numerous future quantum technologies. The complex relationship between a quantum experiment’s structure and its entanglement properties is essential to fundamental research in quantum optics but is difficult to intuitively understand. We present the first deep generative model of quantum optics experiments where a variational autoencoder (QOVAE) is trained on a dataset of experimental setups. In a series of computational experiments, we investigate the learned representation of the QOVAE and its internal understanding of the quantum optics world. We demonstrate that the QOVAE learns an interpretable representation of quantum optics experiments and the relationship between experiment structure and entanglement. We show the QOVAE is able to generate novel experiments for highly entangled quantum states with specific distributions that match its training data. Importantly, we are able to fully interpret how the QOVAE structures its latent space, finding curious patterns that we can entirely explain in terms of quantum physics. The results demonstrate how we can successfully use and understand the internal representations of deep generative models in a complex scientific domain. The QOVAE and the insights from our investigations can be immediately applied to other physical systems throughout fundamental scientific research.

I. INTRODUCTION

Quantum mechanics contains a wide range of phenomena that seem counter intuitive from a classical physics perspective. Experimental quantum physics is integral to the investigation of the fundamental questions associated with these phenomena and the quantum mechanical nature of the universe. Quantum entanglement [1–3] is one of those phenomena that is most difficult to reconcile with our picture of reality and also provides the basis for all quantum technologies and applications. Thus, in particular, quantum optics experiments are not only used to test the foundations of quantum physics [4–6], they are also at the heart of manifold quantum technologies many areas including communication [7] and computation [8–10]. The quantum optics experiments we consider here consist of individual optical elements or devices, such as lasers, beam splitters, or non-linear crystals. Complex quantum phenomena such as multi-photon interference effects [11–14], are challenging to understand intuitively. For that reason, in general the connection between experimental structures and its entanglement properties – the so-called structure-property relation – is complicated to

grasp for humans, which leads to undiscovered potential of these technologies.

In order for the continuing advancement of fundamental research and quantum technologies, it is advantageous that researchers develop computational methods that help in the designing of new quantum hardware while providing conceptual understanding of the results [15]. Examples include the Melvin algorithm that learns to expand its own toolbox with useful elements [16], or a graph-based topological optimizer that allows to extract new human-interpretable concepts [17]. Other works show how to optimize setups with genetic algorithms [18–20], reinforcement learning [21] or parametrized optimization [22]. These efforts do not directly generate quantum optics experiments through the use of a learned representation trained on examples of experiments. Such an approach would provide us with the ability to generate with prior knowledge of specific entangled experiments and allow us to directly explore the relationship between experiment structure and entanglement in model’s learned representation. Therefore, in this work, we focus on using deep unsupervised learning [23] and build a generative model of quantum optics experiments.

Deep generative models have had a major impact in the past few years where they have been applied successfully to a variety of data, including images [24], text [25, 26] and audio [27]. In particular, many advances have been made using deep generative models in the chemical sci-

* danielfs@cs.utoronto.edu

† mario.krenn@mpl.mpg.de

‡ alan@utoronto.edu

ences [28]. Specifically, variational autoencoders (VAEs) [29] have been used for molecular design in order to generate drug like compounds [30, 31], polymers [32] and metal-organic frameworks [33]. When trained on a large dataset, these models are effectively able to learn structural distributions of molecules in addition to basic rules of chemistry, entirely by learning the data distribution.

We take inspiration from the field of molecular generative design and study if it is possible for deep generative models to learn insights related to the complex relationship between entanglement and experimental structure by learning a distribution of quantum optics experiments. For our model, the Quantum Optics Variational Auto Encoder (QOVAE), we made use of the variational autoencoder architecture trained on quantum optics experimental setups. We perform a number of investigations in order to understand what the QOVAE can learn and how it learns. We demonstrate that the QOVAE can learn to generate diverse and novel experiments, almost entirely from the space of entangled state producing experiments. Furthermore we see that the model is able to learn specific distributions of entanglement in the experiments it is trained on. We discover, by analyzing its latent space, that the QOVAE learns an interpretable representation of experiments and encodes surprising insights into the relationship between experiment structure and entanglement.

II. QUANTUM OPTICS EXPERIMENTS & ENTANGLEMENT

Experiments. To represent the discrete structure of the experiments, we use a sequence of optical devices as shown in Figure 1. Each sequence uniquely determines the final quantum state and entanglement properties of the system. The quantum system in each experiment is a four photon system with its initial state created by a double spontaneous parametric down-conversion process (SPDC) that experimentally generates two photon pairs. These SPDC processes can produce multipartite entanglement [36, 37], high-dimensional entanglement encoded in the intrinsic orbital angular momentum (OAM) of photons [38–40], and combinations thereof [41, 42]. The experiments are generated using a set of basic elements consisting of beam splitters, mirrors, dove prism, single mode OAM down-converters and holograms [43]. The holograms and the dove prisms have discrete parameters corresponding to the OAM and phase added to the beam, respectively. Every experiment can be represented sequentially, where every element in the sequence represents a device uniquely identified by its location in the graph, specified by the photons propagating through and its order in the sequence. We use a toolbox of 6 devices (Figure 1) operating on 4 photon paths with up to 2 empty paths.

Entanglement. The system we study is a high-dimensional four-photon quantum state. To quantify its

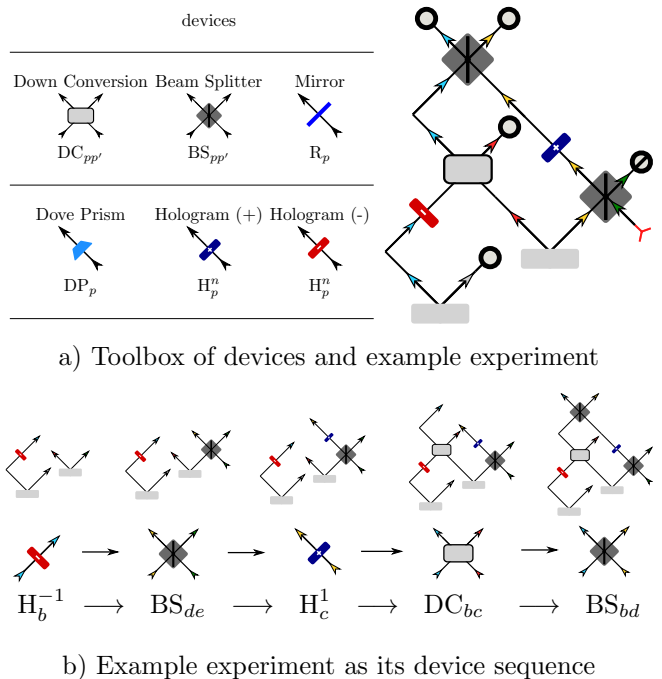


FIG. 1: a) On the (Left), a toolbox of possible devices in any experiment. All elements are used in standard quantum optics laboratories[34], including the technique of entanglement by path identity [35]. For each device included is : its name, visualization and operator sub-scripted by the path(s) it acts on : either a single path p or two paths p, p' . (Right) An example of a quantum optics experiment. Every experiment involves four photon paths beginning with a SPDC crystal to then end with a detector represented by a grey circle with black outline. The paths are designated by arrows and color coded with blue, grey, red and yellow for photons a, b, c, d respectively. In addition, each experiment can have up to two empty paths e, f , color coded with green and purple arrow heads. Empty photon paths start without a crystal and are shown with a red, three pointed star and end with no detector visualized with the detector symbol with a slash. b) The device sequence that defines the experiment in a): first as a graph, then as each device’s visualization from the toolbox in a) and lastly as a sequence of device operators.

entanglement, we derive the entanglement entropy from the discrete Schmidt Rank Vector (SRV) [44]. The SRV is a vector composed by the Schmidt ranks of all bipartitions or subsystems which, in the case of four particles, has a size of seven. For an overall measure of entanglement, we use the sum of all subsystem entanglement entropies, which we denote as S where an experiment with $S > 0$ is entangled and with $S = 0$ is unentangled. More details on the calculation of the quantum state and entanglement can be found in the supplementary.

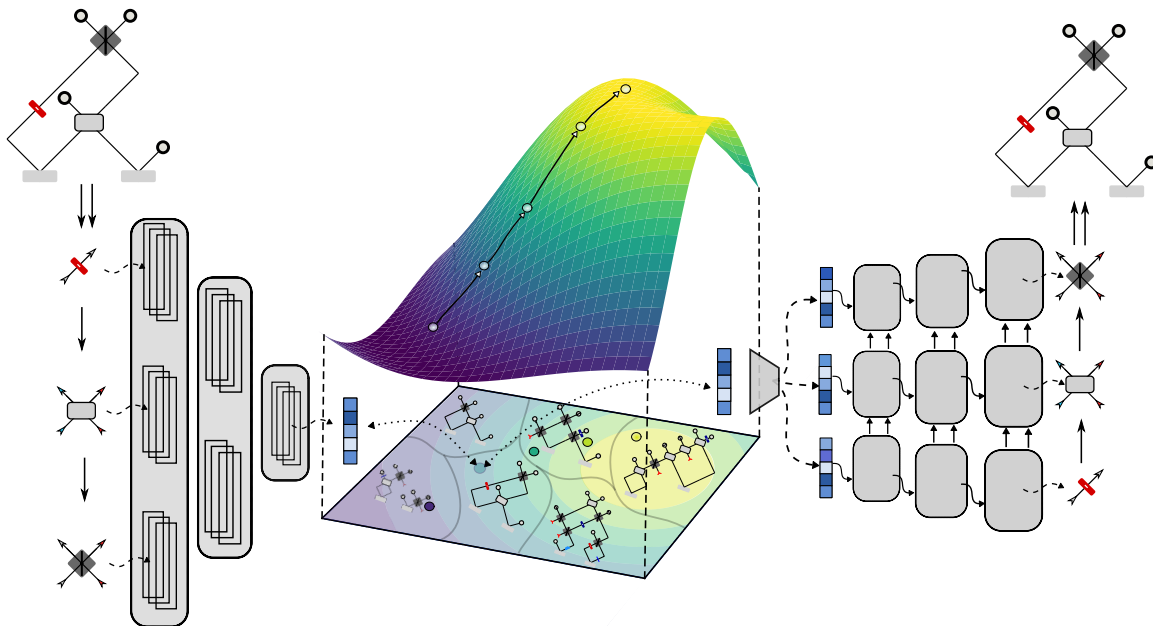


FIG. 2: A visual depiction of the QOVAE. First, an experiment represented by a sequence is encoded into a stochastic latent representation using a convolutional neural network and is reconstructed using another deep recurrent neural network. Over the latent space is the entanglement measure S which is being shown as a function of the latent space \mathbf{z} . On the function arrows are shown moving from a region of low entanglement in violet towards a region of high entanglement in yellow.

III. QUANTUM OPTICS VARIATIONAL AUTOENCODER

For our QOVAE model, we use a variational autoencoder to learn distributions of quantum experiments as sequences. The QOVAE model consists of two neural networks: an encoder which maps a quantum optics experiment \mathbf{x} to a continuous latent representation \mathbf{z} and decoder that reconstructs the experiments \mathbf{x} from the latent representation \mathbf{z} . Both the encoder and decoder are parameterized by deep neural networks. Figure 2 displays the main model. For the data, we represent an experiment sequentially as a series of one-hot column vectors \mathbf{x}_t in a matrix where $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T] \in \mathbb{R}^{T \times D}$. Here, T is maximum experiment length (number of devices) and D is the number of devices in the toolbox. This is fed into the encoder of the QOVAE, which learns a representation by using layers of 1D convolutions that are used to generate the mean and log standard deviation of the latent space. The decoder uses the latent representation of the experiment to generate the experimental sequence using a recurrent neural network. Further details about the QOVAE, how it's trained, as well as the encoder/decoder can be found in the methods section.

To train the QOVAE, we use the Melvin computer algorithm, to generate a training dataset of quantum optics experiments. We target experiments that produce entangled states ($S > 0$), thus we split the dataset into entangled setups $p_{\text{data}}(\mathbf{x}_{S>0})$ and unentangled setups

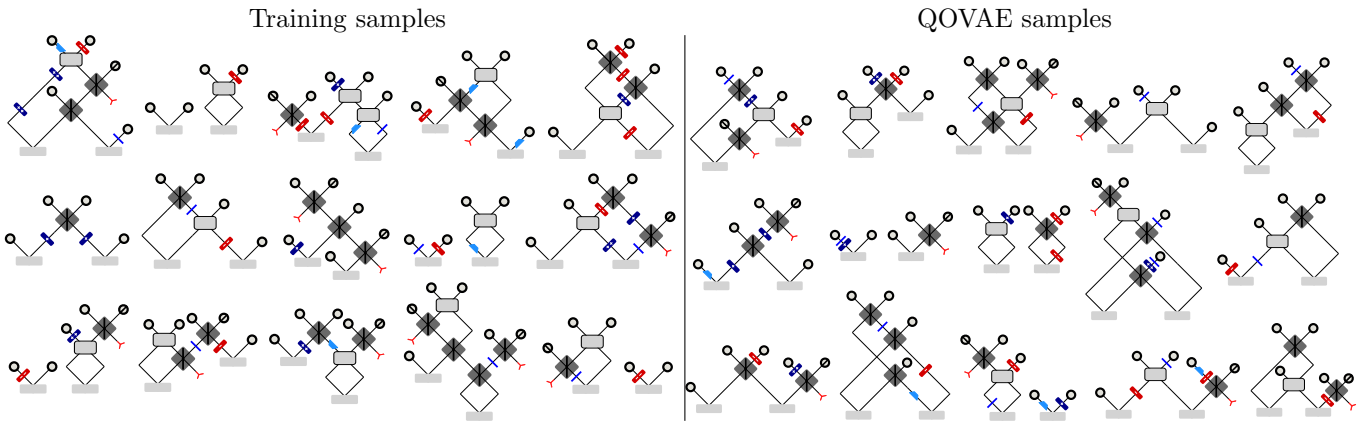
$p_{\text{data}}(\mathbf{x}_{S=0})$. In total, we generated a dataset with approximately 200K (thousand) experiments, half of them producing multipartite entangled states. From the generated experiments, we only use setups that have a maximum of 6 devices that are Beam splitters or Down converters (two path devices). To conduct our investigations, we train two models, the first on a dataset of approximately 80K entangled experiments (after restrictions) with a 6 dimensional latent space (QOVAE-High) and the second on dataset consisted of approximately 35K entangled and unentangled experiments and has a 2 dimensional latent space (QOVAE-Low).

IV. RESULTS AND DISCUSSIONS

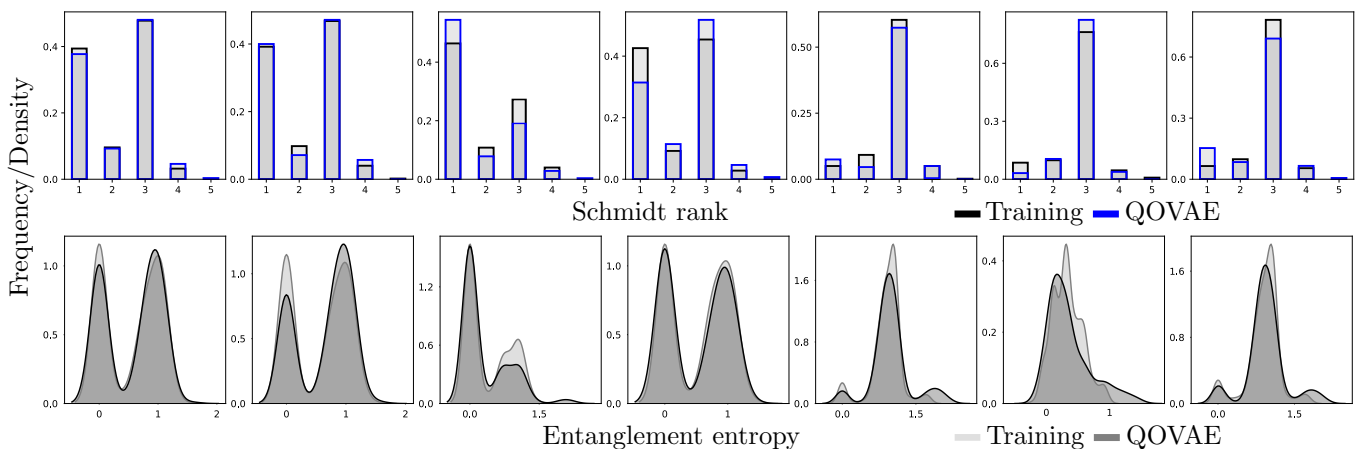
After training both models we conduct a number of experiments to assess the QOVAE and its learned representation. We generate quantum optics experiments using random latent vectors, conduct latent space investigations and assess the entanglement properties of the experiments produced from the model. From the experiments, we extract a series of results that we discuss in the following paragraphs with important conclusions in bold.

The QOVAE learns to generate new, unique experiments almost entirely from the space of entangled quantum optics experiments.

After training QOVAE-High, we randomly sample 10K



a) Comparison of experiments generated by the QOVAE and training experiments.



b) Comparison of entanglement distribution of the QOVAE-High and training experiments.

FIG. 3: (a) Training experiments from the first dataset (left), beside experiments generated by the QOVAE-High. (b) Distribution plots for the entanglement entropy and Schmidt rank of all seven system bi-partitions calculated using experiments generated by the QOVAE-High and training experiments

Measure	QOVAE Experiments	Random experiments
Entangled	0.93	0.47
Uniqueness	0.99	—
Novelty	0.99	—

TABLE I: Entanglement, uniqueness and Novelty ratios for experiments generated from QOVAE-High or randomly

latent vectors and decode them to produce 10K experiments. In this experiment, we study how capable the QOVAE is in generating new data from only the space of quantum optics experimental setups that produce entangled states. We want to observe if QOVAE can ignore the space of unentangled experiments (that produce unentangled states or no state at all), while generating novel and unique entangled experimental setups.

We analyze the 10K sampled experiments using some basic metrics, the results of which are displayed in Table I. First, we calculate the entanglement S of the experiments and determine the ratio between entangled and unentangled experiments (row 1 of Table I). Next, we calculate the uniqueness of the sampled experiments which is the ratio of the number of entangled experiments to the number of entangled experiments after removing the duplicate experiments. Lastly, we calculate the novelty, which is the ratio of unique experiments not appearing in the training data. The QOVAE is able to generate roughly 93% entangled experiments, almost all unique and novel (i.e. not in the training data.)

These results are roughly a double of random sampling. Effectively, the QOVAE can generate from the space of experimental quantum optics setups producing entangled quantum states, that are also not found in the training data and are not all the same setup. These results give a strong indication that the learned internal repre-

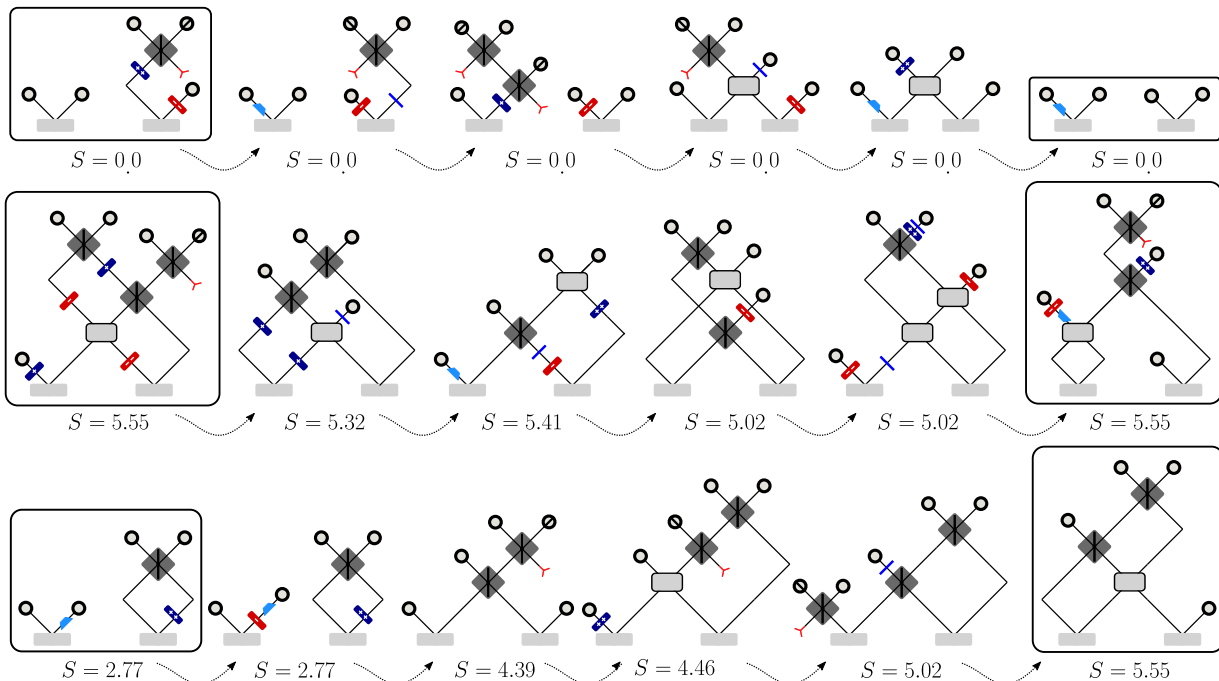


FIG. 4: Three examples of latent space interpolations from the QOVAE. Each interpolation is between experiments with different entanglement measure S . The chosen initial and final experiment along the interpolation path are enclosed in a rectangle and four experiments along the path in between the two are displayed. Below each experiment is the entanglement measure S .

sensation of the QOVAE encodes meaningful structure-entanglement relations.

The QOVAE can learn distributions of entangled states that it is trained on.

Figure 3 (a) displays 15 random samples of experiments from QOVAE-High and its training data. From the figure, we can assess how well the model has learned to generate experiments that are similar to the training experiments by comparing the graph structure of the experiments in both samples. We can see that both sets of samples have similar devices: 1-5 single path and 1-4 double path devices, as well as empty paths where there are 11 in the training experiments and 10 in the QOVAE's. Hence, we conclude that the QOVAE learns a similar distribution as the training set.

Next, we assess the distribution of entanglement learned by the QOVAE-High. Suppose the QOVAE-High can learn to create experiments with a distribution of entangled states. In that case, the distribution of entanglement of every bi-partition or subsystem should be similar between experiments from the training data and the model. To test this, we sample 10K experiments from the model and training data and calculate the entanglement entropy and Schmidt ranks for all seven bi-partitions for every experiment sampled. Now we want to compare the distribution of values between the training data and model. We do this visually with a distribution plot for each subsystem. To plot the seven distributions for each

partition, we use kernel density estimators [45] to estimate the entanglement entropy densities and histograms for the Schmidt ranks. The resulting distribution plots are displayed in figure 3 b), for both the QOVAE and training data.

We can see from the distribution plots in Figure 3 b) that in the first four subsystems (first four columns) the QOVAE-High successfully learns that the training entanglement distribution has two modes (or peaks). For the last three subsystem plots (last three columns), the QOVAE-High is able to learn that there is a single mode of entanglement in the training distribution. Therefore we conclude that the QOVAE has learned to match the training distribution of entanglement for every partition of the system.

In addition to the previous experiments, we test if the QOVAE-Low learns to distinguish the two distributions within its training data: the distribution of entangled experiments $p_{\text{data}}(\mathbf{x}_{S>0})$ and experiments that are not entangled $p_{\text{data}}(\mathbf{x}_{S=0})$. We test if the QOVAE-Low can produce the same ratio of entangled to unentangled experiments present in the training data. As before, we generate 10K experiments from the QOVAE by sampling random vectors $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then decode those vectors to produce 10K quantum optics experiments. We calculate the entanglement measure for each experiment S and see that 54% of those setups have $S > 0$ and, therefore, are entangled. This means that the QOVAE reproduces the same bi-modal distribution of entangled

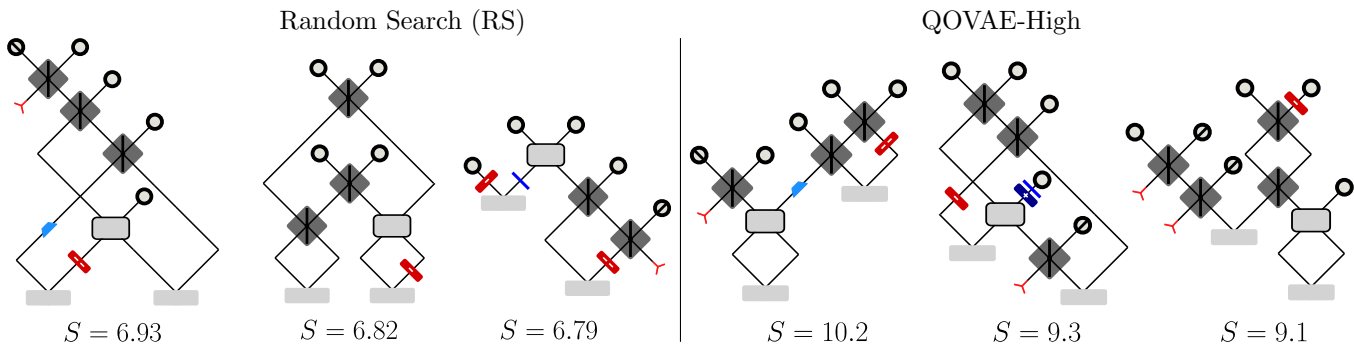


FIG. 5: (Left) Quantum optics experiments with highest entanglement S found using Random Search. (Right) The top three experiments found by searching the latent space of the QOVAE.

vs not entangled experiments in the training data.

The QOVAE learns a quasi-continuous embedding in terms of entanglement.

We analyse the smoothness of the latent space in terms of the entanglement measure S . Specifically, we test if experiments that are close in the latent space have similar entanglement properties. This can be done by interpolating in the latent space from one latent representation \mathbf{z}_1 of an experiment to another \mathbf{z}_2 , and decoding experiments on the path from $\mathbf{z}_1 \rightarrow \mathbf{z}_2$. We use spherical linear interpolation [46] for the path and decode at 5 equally spaced steps along the path.

In total, we show three interpolation from the latent space of the QOVAEs, as shown in Figure 4. In order to see if the model has learned a notion of similarity between experimental setups corresponding to their entanglement, we perform interpolations from different entanglement measures S . The first one shown in Figure 4 interpolate between experiments that do not produce entangled states ($S = 0$) and from QOVAE-Low. The last two interpolations from QOVAE-High between experiments that produce high-dimensional entangled states ($S > 0$), the second between experiments with $S = 5.55$ and the last between $S = 2.77$ and $S = 5.55$.

For the first interpolation, the experiments decoded along the interpolation path preserve the entanglement of the initial/final experiments, experiments remain unentangled along the path. Similarly, in the second interpolation, decoded experiments maintain entanglement at $S \approx 5.0$ of the initial and final experiments'. In the last interpolation, the entanglement S of the experimental setups decoded along the path increase linearly from $S = 2.77$ towards the final experiment's entanglement of $S = 5.55$.

This experiment demonstrates that the QOVAE learns a rather smooth representation of quantum optics experiments that encodes a measure of similarity between the experimental structures and entanglement properties.

The QOVAE can be used to efficiently search

	Sequence
QOVAE	$DC_{ab} \rightarrow D_b \rightarrow BS_{bd} \rightarrow BS_{ae} \rightarrow H_c^{-2} \rightarrow BS_{cd}$
	$DC_{cd} \rightarrow BS_{ae} \rightarrow BS_{af} \rightarrow BS_{bc} \rightarrow BS_{bc} \rightarrow H_c^{-1}$
	$BS_{cf} \rightarrow DC_{ac} \rightarrow R_a \rightarrow H_b^{-1} \rightarrow H_a^1 \rightarrow BS_{bd} \rightarrow BS_{bc}$
RS	$H_a^{-2} \rightarrow DC_{ad} \rightarrow D_b \rightarrow BS_{bc} \rightarrow BS_{cd} \rightarrow BS_{cf}$
	$BS_{ab} \rightarrow H_c^{-1} \rightarrow DC_{cd} \rightarrow BS_{ac} \rightarrow BS_{bd}$
	$R_a \rightarrow H_c^{-2} \rightarrow H_b^{-1} \rightarrow BS_{ce} \rightarrow BS_{cd} \rightarrow DC_{ac} \rightarrow D_a$

TABLE II: Experiments from Fig 5 in ascending order.

for new highly entangled quantum setups in the latent space.

We use the QOVAE to search for new experiments that produce high dimensional entangled states. We use random search in the latent space of QOVAE-High (that is trained only on entangled experiments). We compare the random search in the latent space with a random search directly on the experimental device sequence. We want to find experiments that produce systems with the largest possible entanglement measured with S . We are effectively comparing the following optimizations

$$\max_{\mathbf{x}_\pi \in \mathcal{X}} S(\mathbf{x}_\pi) \text{ versus } \max_{\mathbf{z} \in \mathcal{Z}} S(f(\mathbf{z})),$$

where $\mathbf{x}_\pi \in \mathcal{X}$ is an experiment from the space of randomly sampled quantum optics experiments and \mathcal{Z} is the learned latent space of the QOVAE. The learned latent is used as a proxy optimization space by randomly sampling a latent vector $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then computing $S(f(\mathbf{z}))$, where f is the QOVAE decoder.

We performed a single run of random search with the same number of iterations (10K) for both searches and display the top 3 experimental setups that we found by random sampling on device sequences and the QOVAE in figure 5. Table II displays the same experiments in their sequential representation. The search in the QOVAE's latent space finds experiments with, in total, higher entanglement S . This demonstrates the utility

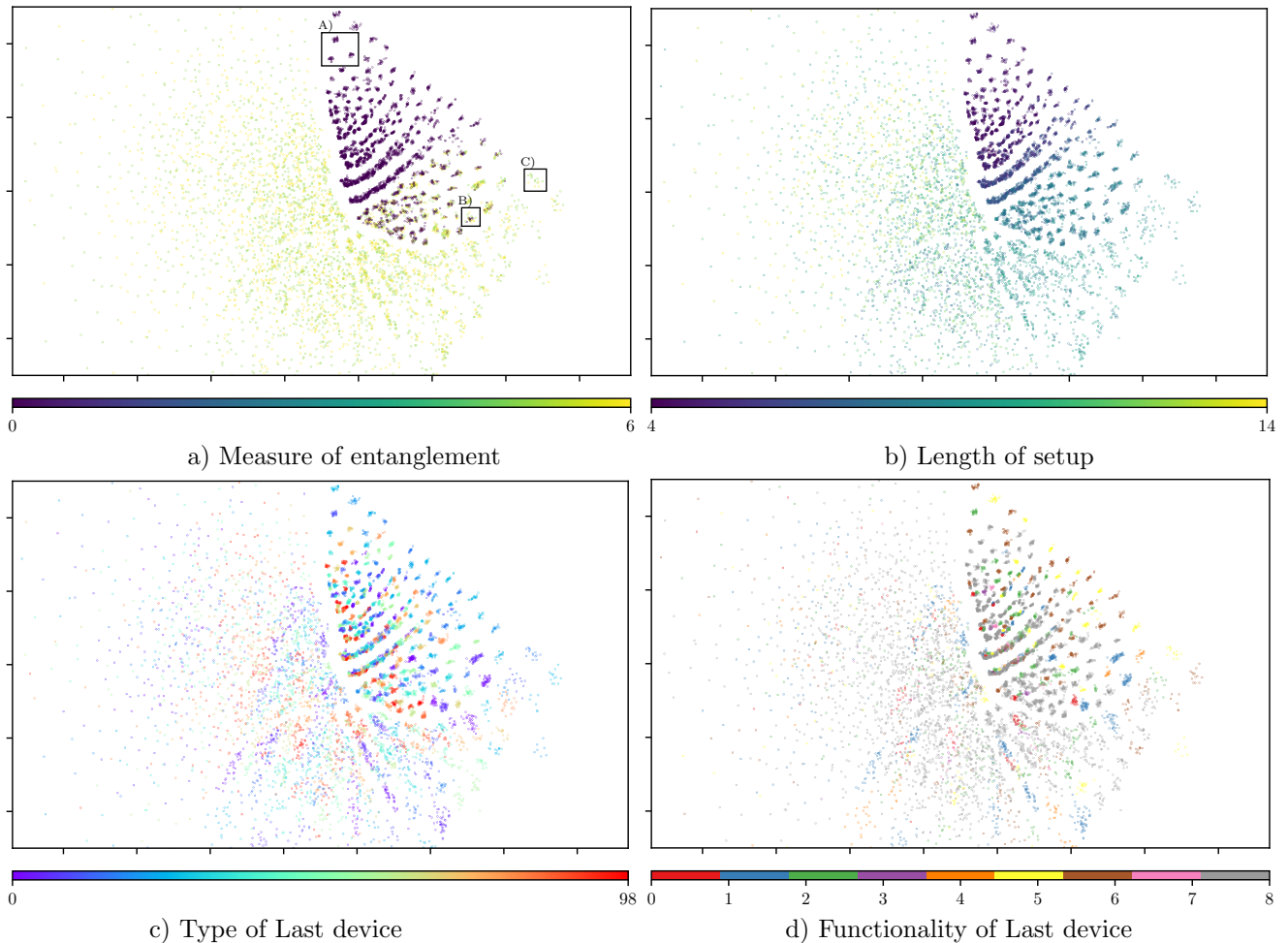


FIG. 6: An exploration of the two dimensional latent space of the QOVAE-Low colored *a)* by entanglement measure S , *b)* by length, *c)* by last devices, and *d)* Functionality of Last device.

of using a learned representation of quantum optics experiments when searching for new highly entangled experiments.

The QOVAE learns an interpretable representation.

Since QOVAE-Low only has two latent dimensions, we can directly plot the latent space and interpret the latent space in terms of structure-property relations. These plots help us to explain the results we have obtained in the previous experiments.

We perform this interpretation by encoding every training example into the latent space, and observe remarkable structures and distinct clusters of setups (see Fig.6). We can now color-code every point in the latent space with the entanglement measure of the corresponding training example. We find a very clear region in the latent space that is unentangled (colored in violet), and another structure being entangled (see Fig.6a). This is an interesting and exciting result, as the model was trained

entirely without the corresponding entanglement information, just by using the structure of the setup. This result indicates that the model implicitly distinguishes between entangled and unentangled experiments. To understand how the QOVAE does this we analyse the training examples that correspond to three different regions, one unentangled region A), one that shares entangled and unentangled states B) and one that contains only entangled states C).

In region A), we find three different clusters, none of which contain entangled states. We find that all three clusters contain experiments with only four devices, while in region B) and C), all setups have a device length of nine. So we understand - to first order - that the QOVAE encodes the training examples by the number of the optical devices. We confirm this by color-coding the latent space by the length of the setups, as shown in Fig.6b.

However, the size of the setup cannot alone be responsible for the very peculiar structure of the latent space and its distribution of entanglement. After all, region B) and C) have the same number of elements, however B)

contains both entangled and unentangled setups, while C) contains only entangled ones. We analyse those two regions further and find, that the last element in every setup of region B) is always H_b^{-2} while in C) it is always DC_{ac} . Therefore we understand that the last element of the setup is the second property the QOVAE uses to separate setups, which we can confirm by color-coding all points by their final element (see Fig.6c).

With this new insight, we can finally explain why the two regions B) and C) have different entanglement properties: the initial two elements in every setup (by construction) consist of nonlinear crystal in path a, b and one in c, d , which creates a correlated state in all four paths. In order to get a second correlated state in a superposition (thus entanglement), the setup needs to create another combination of correlations in the four paths. The only two other ways are by creating correlations in paths ac and bd or in paths ad and bc , which can be understood in a graph-theoretical way [47, 48]. In region C), the last element is a crystal in path ac , which already creates the necessary correlation. In this setup, the previous eight elements only need to create another correlation in path bd to create entanglement, while in region B), the fully correlated state needs to be created, which is much more unlikely to happen. We can further strengthen our interpretation by clustering all possible last elements into eight functional groups for instance, all holograms have the same effect on the entanglement property, so we can combine them to a single functional group.

From this result, we have been able to fully interpret the structure-property relation that has been learned by the QOVAE in a quantum physical sense. With this new understanding, we can also explain properties that we have observed in previous experiments. For example, the encoding is smooth in terms of entanglement because the number of optical elements significantly determines the possible correlations build from inside the setup.

V. CONCLUSION

We presented the QOVAE which is the first deep generative model for the design of quantum optics hardware. Deep generative models are widely used but there has never been any investigation or understanding developed of their internal representation in a complex scientific domain. In a series of complex computational experi-

ments, we investigated the QOVAE's *internal picture of the quantum world*. The QOVAE was able to generate novel entangled experiments, learn distributions of entanglement and was shown to interpolate smoothly in its latent space— which can also be used to search effectively for new highly entangled experiments. When plotting the QOVAE's latent space we find complex internal structure, that surprisingly the QOVAE implicitly discovered properties of entangling quantum experiments in an unsupervised way. Our results go beyond designing new quantum optics— they tackle the question of interpretability and explainability of black-box models [49, 50] in a scientific domain. This is particularly promising since understanding what these model learn could lead to new computer-inspired scientific insights and discoveries.

The QOVAE could, in principal, be directly be applied to other physical science domains, such as in the design of new quantum circuits for quantum computing. Currently, Noisy Intermediate-Scale Quantum (NISQ) computing algorithms [51–53] are promising candidates to surpasses the classical computational capabilities for numerous applications. Most of these approaches require good priors to explore efficiently and represent the space of parameters and solutions. The exponentially large Hilbert space formed by all possible quantum circuits makes this task computationally intractable when the structure-properties relation of these circuits is still not fully-understood. QOVAE's ability to learn meaningful representations as understood by domain experts could provide insights about how the Hilbert space is organized within these parameterized quantum circuits. QOVAE learns an intepretable representation of entanglement in quantum optics experiments. Our work with the QOVAE is an example in the physical sciences of opening the black-box of deep generative models to develop promising scientific insights.

ACKNOWLEDGEMENTS

A.A.-G. acknowledge support from the Canada 150 Research Chairs Program, the Canada Industrial Research Chair Program, and from Google, Inc. in the form of a Google Focused Award. M.K. acknowledges support from the FWF (Austrian Science Fund) via the Erwin Schrödinger fellowship No. J4309. The authors declare that there are no competing interests.

-
- [1] E. Schrödinger, “Discussion of probability relations between separated systems,” *Mathematical Proceedings of the Cambridge Philosophical Society* **31**, 555 (1935).
 - [2] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?” *Physical review* **47**, 777 (1935).
 - [3] J. S. Bell, “On the einstein podolsky rosen paradox,” *Physics Physique Fizika* **1**, 195–200 (1964).
 - [4] M. Giustina, M. A. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, *et al.*, “Significant-loophole-free test of bell’s theorem with entangled photons,” *Physical review letters* **115**, 250401 (2015).
 - [5] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, *et al.*, “Strong loophole-free

- test of local realism,” *Physical review letters* **115**, 250402 (2015).
- [6] K.-W. Bong, A. Utreras-Alarcón, F. Ghafari, Y.-C. Liang, N. Tischler, E. G. Cavalcanti, G. J. Pryde, and H. M. Wiseman, “A strong no-go theorem on the wigner’s friend paradox,” *Nature Physics* **16**, 1199 (2020).
- [7] J. Yin, Y. Cao, Y.-H. Li, J.-G. Ren, S.-K. Liao, L. Zhang, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, *et al.*, “Satellite-to-ground entanglement-based quantum key distribution,” *Physical review letters* **119**, 200501 (2017).
- [8] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nature communications* **5**, 1 (2014).
- [9] S. Paesani, Y. Ding, R. Santagati, L. Chakhmakchyan, C. Vigliar, K. Rottwitt, L. K. Oxenløwe, J. Wang, M. G. Thompson, and A. Laing, “Generation and sampling of quantum states of light in a silicon chip,” *Nature Physics* **15**, 925 (2019).
- [10] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, “Quantum computational advantage using photons,” *Science* **370**, 1460 (2020).
- [11] L. Wang, X. Zou, and L. Mandel, “Induced coherence without induced emission,” *Physical Review A* **44**, 4614 (1991).
- [12] T. Herzog, J. Rarity, H. Weinfurter, and A. Zeilinger, “Frustrated two-photon creation via interference,” *Physical review letters* **72**, 629 (1994).
- [13] A. J. Menssen, A. E. Jones, B. J. Metcalf, M. C. Tichy, S. Barz, W. S. Kolthammer, and I. A. Walmsley, “Distinguishability and many-particle interference,” *Physical review letters* **118**, 153603 (2017).
- [14] L.-T. Feng, M. Zhang, D. Liu, Y.-J. Cheng, G.-P. Guo, D.-X. Dai, G.-C. Guo, M. Krenn, and X.-F. Ren, “Observation of nonlocal quantum interference between the origins of a four-photon state in a silicon chip,” arXiv:2103.14277 (2021).
- [15] M. Krenn, M. Erhard, and A. Zeilinger, “Computer-inspired quantum experiments,” *Nature Reviews Physics* **2**, 649 (2020).
- [16] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, “Automated search for new quantum experiments,” *Physical Review Letters* **116** (2016), [10.1103/physrevlett.116.090405](https://doi.org/10.1103/physrevlett.116.090405).
- [17] M. Krenn, J. Kottmann, N. Tischler, and A. Aspuru-Guzik, “Conceptual understanding through efficient automated design of quantum optical experiments,” *Physical Review X* **11**, 031044 (2021).
- [18] P. Knott, “A search algorithm for quantum state engineering and metrology,” *New Journal of Physics* **18**, 073033 (2016).
- [19] R. Nichols, L. Mineh, J. Rubio, J. C. Matthews, and P. A. Knott, “Designing quantum experiments with a genetic algorithm,” *Quantum Science and Technology* **4**, 045012 (2019).
- [20] L. O’Driscoll, R. Nichols, and P. A. Knott, “A hybrid machine learning algorithm for designing quantum experiments,” *Quantum Machine Intelligence* **1**, 5 (2019).
- [21] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, “Active learning machine learns to create new quantum experiments,” *Proceedings of the National Academy of Sciences* **115**, 1221–1226 (2018).
- [22] J. M. Arrazola, T. R. Bromley, J. Izaac, C. R. Myers, K. Brádler, and N. Killoran, “Machine learning method for state preparation and gate synthesis on photonic quantum computers,” *Quantum Science and Technology* **4**, 024004 (2019).
- [23] R. Salakhutdinov, “Learning deep generative models,” *Annual Review of Statistics and Its Application* **2**, 361 (2015).
- [24] A. Razavi, A. v. d. Oord, and O. Vinyals, “Generating diverse high-fidelity images with vq-vae-2,” arXiv preprint arXiv:1906.00446 (2019).
- [25] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, “Generating sentences from a continuous space,” arXiv preprint arXiv:1511.06349 (2015).
- [26] S. Semeniuta, A. Severyn, and E. Barth, “A hybrid convolutional variational autoencoder for text generation,” arXiv preprint arXiv:1702.02390 (2017).
- [27] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, in *International Conference on Machine Learning* (PMLR, 2018) pp. 4364–4373.
- [28] B. Sanchez-Lengeling and A. Aspuru-Guzik, “Inverse molecular design using machine learning: Generative models for matter engineering,” *Science* **361**, 360 (2018).
- [29] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114 (2013).
- [30] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, “Automatic chemical design using a data-driven continuous representation of molecules,” *ACS central science* **4**, 268 (2018).
- [31] W. Jin, R. Barzilay, and T. Jaakkola, “Junction tree variational autoencoder for molecular graph generation,” arXiv preprint arXiv:1802.04364 (2018).
- [32] W. Jin, R. Barzilay, and T. Jaakkola, “Hierarchical generation of molecular graphs using structural motifs,” *International Conference on Machine Learning* , 4839 (2020).
- [33] Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr, *et al.*, “Inverse design of nanoporous crystalline reticular materials with deep generative models,” *Nature Machine Intelligence* **3**, 76 (2021).
- [34] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, “Multiphoton entanglement and interferometry,” *Reviews of Modern Physics* **84**, 777 (2012).
- [35] M. Krenn, A. Hochrainer, M. Lahiri, and A. Zeilinger, “Entanglement by path identity,” *Physical Review Letters* **118**, 080401 (2017).
- [36] D. Bouwmeester, J.-W. Pan, M. Daniell, H. Weinfurter, and A. Zeilinger, “Observation of three-photon greenberger-horne-zeilinger entanglement,” *Physical Review Letters* **82**, 1345–1349 (1999).
- [37] X.-C. Yao, T.-X. Wang, P. Xu, H. Lu, G.-S. Pan, X.-H. Bao, C.-Z. Peng, C.-Y. Lu, Y.-A. Chen, and J.-W. Pan, “Observation of eight-photon entanglement,” *Nature photonics* **6**, 225 (2012).
- [38] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, “Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes,” *Physical Review A* **45**, 8185–8189 (1992).

- [39] J. Romero, D. Giovannini, S. Franke-Arnold, S. M. Barnett, and M. J. Padgett, “Increasing the dimension in high-dimensional two-photon orbital angular momentum entanglement,” *Physical Review A* **86** (2012), [10.1103/physreva.86.012334](https://doi.org/10.1103/physreva.86.012334).
- [40] M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, “Generation and confirmation of a (100 x 100)-dimensional entangled quantum system,” *Proceedings of the National Academy of Sciences* **111**, 6243–6247 (2014).
- [41] M. Erhard, M. Malik, M. Krenn, and A. Zeilinger, “Experimental greenberger–horne–zeilinger entanglement beyond qubits,” *Nature Photonics* **12**, 759 (2018).
- [42] Y.-H. Luo, H.-S. Zhong, M. Erhard, X.-L. Wang, L.-C. Peng, M. Krenn, X. Jiang, L. Li, N.-L. Liu, C.-Y. Lu, A. Zeilinger, and J.-W. Pan, “Quantum teleportation in high dimensions,” *Physical review letters* **123**, 070505 (2019).
- [43] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial, “Measuring the orbital angular momentum of a single photon,” *Physical Review Letters* **88** (2002), [10.1103/physrevlett.88.257901](https://doi.org/10.1103/physrevlett.88.257901).
- [44] M. Huber and J. I. De Vicente, “Structure of multidimensional entanglement in multipartite systems,” *Physical review letters* **110**, 030501 (2013).
- [45] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization* (John Wiley & Sons, 2015).
- [46] K. Shoemake, in *Proceedings of the 12th annual conference on Computer graphics and interactive techniques* (1985) pp. 245–254.
- [47] M. Krenn, X. Gu, and A. Zeilinger, “Quantum experiments and graphs: Multipartite states as coherent superpositions of perfect matchings,” *Physical review letters* **119**, 240403 (2017).
- [48] X. Gu, M. Erhard, A. Zeilinger, and M. Krenn, “Quantum experiments and graphs ii: Quantum interference, computation, and state generation,” *Proceedings of the National Academy of Sciences* **116**, 4147 (2019).
- [49] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence* **1**, 206 (2019).
- [50] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (IEEE, 2018) pp. 80–89.
- [51] J. Preskill, “Quantum computing in the nisq era and beyond,” *Quantum* **2**, 79 (2018).
- [52] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational quantum algorithms,” *Nature Reviews Physics* , 1 (2021).
- [53] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, “Noisy intermediate-scale quantum (nisq) algorithms,” *arXiv preprint arXiv:2101.08448* (2021).
- [54] M. Mirhosseini, O. S. Magaña-Loaiza, M. N. O’Sullivan, B. Rodenburg, M. Malik, M. P. J. Lavery, M. J. Padgett, D. J. Gauthier, and R. W. Boyd, “High-dimensional quantum cryptography with twisted light,” *New Journal of Physics* **17**, 033033 (2015).
- [55] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, “Experimental high-dimensional two-photon entanglement and violations of generalized bell inequalities,” *Nature Physics* **7**, 677–680 (2011).

SUPPLEMENTARY MATERIALS

1. Methods

a. Latent variable models

Consider some dataset $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ of i.i.d. samples of discrete or continuous data. We assume the data is generated via some unobserved continuous random variable \mathbf{z} such that the joint distribution can be written as

$$p_{\theta}(\mathbf{x}, \mathbf{z}) = p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z}).$$

The true posterior can be found by the Bayes rule as

$$p_{\theta}(\mathbf{z}|\mathbf{x}) = p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})/p_{\theta}(\mathbf{x}),$$

but is intractable because of the marginal likelihood $p_{\theta}(\mathbf{x})$, which can be computed as

$$p_{\theta}(\mathbf{x}) = \int p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})d\mathbf{z},$$

and is also intractable because the likelihood is a complicated nonlinear function of the latent variables.

The marginal likelihood of the data is composed of a sum over the marginal likelihoods of individual datapoints:

$$\log p_{\theta}(\mathbf{X}) = \log p_{\theta}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i=1}^N \log p_{\theta}(\mathbf{x}_i).$$

b. Variational Bayes

By defining an approximation to true posterior $q_{\phi}(\mathbf{z}|\mathbf{x})$, we can rewrite a single marginal likelihood as

$$\log p_{\theta}(\mathbf{x}) = \mathbb{D}_{\text{KL}}[q_{\phi}(\mathbf{z}|\mathbf{x})|p_{\theta}(\mathbf{z}|\mathbf{x})] + \mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x}),$$

where the first term on the right is a KL divergence of the approximate posterior to the true posterior. Since the KL is positive,

$$\log p_{\theta}(\mathbf{x}) \geq \mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x}),$$

hence $\mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x})$ is a lower bound on the marginal likelihood, known as the ELBO, and can be written as

$$\begin{aligned} \mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x}) &= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[\log \frac{p_{\theta}(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] \\ &= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}, \mathbf{z})] - \mathbb{H}[q_{\phi}] \\ &= \mathbb{E}_{q_{\phi}} [\log p_{\theta}(\mathbf{x}|\mathbf{z})] - \mathbb{D}_{\text{KL}}[q_{\phi}|p_{\theta}(\mathbf{z})]. \end{aligned}$$

We want to optimize the variational lower bound of $\mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x})$ with respect to the variational and generative parameters ϕ, θ .

c. stochastic gradient variational Bayes (SGVB)

SGVB introduces a simple practical estimator of the ELBO and its derivatives. Under specific chosen approximate posteriors $q_{\phi}(\mathbf{z}|\mathbf{x})$, we can re-parameterize the random variable \mathbf{z} using a differentiable transformation from some auxiliary noise ϵ [29] as

$$\mathbf{z} = \mathbf{t}_{\phi}(\epsilon, \mathbf{x}) \text{ where } \epsilon \sim p(\epsilon).$$

We can now differentiate through sampling and form Monte Carlo estimates of the elbo as follows:

$$\mathcal{L}(q_{\phi}, p_{\theta}, \mathbf{x}) = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\ell(\mathbf{z})] = \mathbb{E}_{p(\epsilon)} [\ell(\mathbf{t}_{\phi}(\epsilon, \mathbf{x}))],$$

and take gradients simply by

$$\frac{\partial}{\partial \phi} \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\ell(\mathbf{z})] = \mathbb{E}_{p(\epsilon)} \left[\frac{\partial \ell}{\partial \mathbf{t}} \frac{\partial \mathbf{t}}{\partial \phi} + \frac{\partial \ell}{\partial \phi} \right].$$

d. Mean field assumption

In mean-field variational inference [?], we specify an approximate posterior that factorizes, for example, the diagonal multivariate Gaussian:

$$q_{\phi}(\mathbf{z}|\mathbf{x}) = \prod_{i=1}^d q_{\phi}(\mathbf{z}_i|\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2).$$

We can reparameterize $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$ and evaluate the KL divergence and entropy,

$$\begin{aligned} \mathbb{D}_{\text{KL}}[q_{\phi}|p_{\theta}(\mathbf{z})] &= \frac{1}{2} \sum_{j=1}^d (1 + 2 \log \sigma_j - \mu_j^2 - \sigma_j^2) \\ \mathbb{H}[q_{\phi}] &= \frac{d}{2} \log(2\pi e) + \sum_{j=1}^d \log \sigma_j, \end{aligned}$$

with prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbb{I})$, where d is dimension of the latent variables. In a variational autoencoder (VAE), we parameterize both the encoder and decoder using neural networks as described in the next sections.

e. Encoder

Our encoder is a diagonal Gaussian whose parameters are a mapping from the data manifold to the latent space. Our data, the quantum optics experiments $\mathbf{x} \in \mathbb{R}^{d \times T}$, are represented as a sequence with T elements, each from a toolbox of d possible devices,

$$q_{\phi}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z} | \boldsymbol{\mu}_{\phi}(\mathbf{x}), \boldsymbol{\sigma}_{\phi}^2(\mathbf{x})),$$

where $\boldsymbol{\mu}, \log \boldsymbol{\sigma} = g_\phi(\mathbf{x})$. First, g consists of a convolutional neural network with 3 layers,

$$\mathbf{h} = \text{Conv1d}_3(\text{Conv1d}_2(\text{Conv1d}_1(\mathbf{x}))).$$

A single layer takes on the form

$$\mathbf{x}' = \text{Conv1d}(\mathbf{x}) = \text{ReLU}(\mathbf{w} \odot \mathbf{x} + \mathbf{b}),$$

where $\mathbf{w} \in \mathbb{R}^{n_f \times \ell \times d}$ is the convolution filter tensor which consists of n_f filters each with length ℓ and d features. Also the layer output is $\mathbf{x}' \in \mathbb{R}^{n_f \times T - \ell + 1}$.

For the input layer, this is just $d = D$ the number of devices in the toolbox used to create any experiment. $\text{ReLU}(\cdot)$ is the element-wise rectified linear unit function and \odot is the convolution operator which outputs a tensor with the has following elements

$$(\mathbf{w} \odot \mathbf{x})_{ft} = \sum_{l=1}^{\ell} \mathbf{w}_l^f \cdot \mathbf{x}_{t+l-1},$$

where $\mathbf{w}_\ell^f \in \mathbb{R}^d$ is the f^{th} convolutional filter ℓ^{th} weight vector operating on the t^{th} element (device) in the sequence (experiment).

The second component of the encoder g is a MLP with three layers that maps the flattened output \mathbf{h} from the Convolutional neural net. to the parameters of latent distribution

$$\boldsymbol{\mu}, \log \boldsymbol{\sigma} = \text{MLP}_g(\text{Flatten}(\mathbf{h})).$$

f. Decoder

For the observation model, every data point is a sequence of devices from the d element toolbox of possible device elements, thus we can model the data as independent categoricals whose mean vector is mapped from a latent samples using a neural network,

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{z}) = \prod_{t=1}^T \text{Categorical}(\mathbf{p}_{\mathbf{x}_t}),$$

where $\mathbf{p}_{\mathbf{x}_t} \in [0, 1]^d$ is the probability vector of each device in the toolbox. Our encoder outputs these probabilities as

$$\mathbf{p}_{\mathbf{x}_1}, \dots, \mathbf{p}_{\mathbf{x}_t}, \dots, \mathbf{p}_{\mathbf{x}_T} = f_\theta(\mathbf{z}).$$

First, f consists of a three layer Recurrent neural network each with with Gated recurrent units (GRUs),

$$\mathbf{h} = \text{GRU}_3(\text{GRU}_2(\text{GRU}_1(\text{MLP}_f(\mathbf{z})))),$$

where MLP_f is single layer MLP with a $\text{ReLU}(\cdot)$ activation. A single layer $\mathbf{h} = \text{GRU}(x)$ takes on the form

$$\begin{aligned} \mathbf{z}'_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \\ \mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \\ \mathbf{h}'_t &= \text{Tanh}(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h), \\ \mathbf{h}_t &= (1 - \mathbf{z}'_t) \odot \mathbf{h}_{t-1} + \mathbf{z}'_t \odot \mathbf{h}'_t, \end{aligned}$$

where $\mathbf{W}, \mathbf{U}, \mathbf{b}$ are parameters of the GRU layer. The \mathbf{z}'_t and \mathbf{r}_t are update and reset gate vectors, respectively, and σ is the sigmoid function. The input to the first GRU layer is the T length sequence of vectors $[\text{MLP}_f(\mathbf{z}), \dots, \text{MLP}_f(\mathbf{z})]$. The output of the RNN is mapped to the device probabilities using a softmax layer

$$\mathbf{p}_{\mathbf{x}_t} = \text{Softmax}(\mathbf{W} \mathbf{h}_t + \mathbf{b}) \quad t = 1, \dots, T.$$

g. Training and data details

We manually performed an initial hyperparameter search for both QOVAEs. We found an initial set of training hyperparameters using Grid search. We found training on around 1600 epochs to produce better validation accuracy and ELBO values. We trained the QOVAE using SGD and the Adam optimizer [29] with a low learning rate $\sim 10^{-4}$. Training was done using the KERAS machine learning package from tensorflow. All models were trained on a V100 node on the Beluga supercomputer. The minibatch size grid was $\{32, 64, 128, 256\}$ with an optimal batch size of 64. The model architecture is the same for all experiments with both models. We use grid search to search over the encoder and decoder architecture. For the convolutional layers we search over 6, 12, 18, 36 filter numbers with lengths 3, 4, 5 For the MLPs in both the decoder and encoder, we consider 32, 64, 128, 256 hidden units in each layer. For the GRU layer, we consider hidden states of size 128, and a 64 unit layer that maps to the GRU layers. To generate the data using random search/ Melvin, in order to randomly sample an experiment $\mathbf{x}_\pi = (x_{\pi_1}, \dots, x_{\pi_\ell})$, we first sample an experiment length $\ell \sim \text{uniform}\{1, \dots, T\}$, and then sample ℓ devices x_{π_i} from the toolbox with replacement $\pi_i \sim \text{uniform}\{1, \dots, D\}$. Next, we calculate the total entanglement measure S of the experiment and classify those with $\mathbf{x}_{S>0}$ or $\mathbf{x}_{S=0}$.

Device	Token	Visual	Operation	Operator
Down Conversion	DownConv(Ψ, p, p')		$ \Psi\rangle + \sum_{\ell} \ell\rangle_p -\ell\rangle_{p'}$	DC $_{pp'}$
Beam Splitter	BS(Ψ, p, p')		$ \ell\rangle_p \rightarrow \frac{ \ell\rangle_{p'} + i -\ell\rangle_p}{\sqrt{2}}$	BS $_{pp'}$
Mirror	Ref(Ψ, p)		$ \ell\rangle_p \rightarrow i -\ell\rangle_p$	R $_p$
Dove Prism	DP(Ψ, p)		$ \ell\rangle_p \rightarrow ie^{i\pi\ell} -\ell\rangle_p$	DP $_p$
Hologram	OAMHo1o(Ψ, p, n)		$ \ell\rangle_p \rightarrow \ell + n\rangle_p$	H $_p^n$
Hologram	OAMHo1o($\Psi, p, -n$)		$ \ell\rangle_p \rightarrow \ell - n\rangle_p$	H $_p^{-n}$

TABLE III: The main table of devices divided into columns displaying 1) the standard device name 2) The token used in the sequence representation of the experimental setups and used to convert any setup to a sequence of one hot vectors 3) The visual of the device used in the graph representation of the experimental setups 4) the operator that the device represents that acts on the systems state 5) the action the operator performs on any specific ket in the state

2. The quantum state

a. The quantum system

We are investigating multipartite entanglement in a four photon system with two pairs of OAM-entangled photons. The state is represented by the OAM of the photon which is its orbital angular momentum. Each maximally entangled quantum optics experimental setup will produce some state which lives in the hilbert space defined by the tensor product of the individual subsystems defined by photons a, b, c, d given by $\mathcal{H} = \mathcal{H}_a \otimes \mathcal{H}_b \otimes \mathcal{H}_c \otimes \mathcal{H}_d$

$$|\Psi\rangle = \sum_{ijkl \in \ell_{\text{OAM}}} \alpha_{ijkl} |\psi_{ijkl}\rangle \quad (1)$$

where we can define a general state in the system as a superposition of the basis kets as defined

$$|\psi_{ijkl}\rangle = |i\rangle_a \otimes |j\rangle_b \otimes |k\rangle_c \otimes |l\rangle_d \quad (2)$$

$$= |i\rangle_a |j\rangle_b |k\rangle_c |l\rangle_d = |i, j, k, l\rangle \quad (3)$$

where i, j, k, l are the OAM quantum number of the photon, OAM states provide a suitable physical realisation of multilevel qudit systems which have been shown to improve the robustness of quantum key distribution schemes. In general, photon OAM states take on discrete integer values $m \in \mathbb{Z}$ with OAM $m\hbar$. A proof-of-principle experiment with 7 OAM modes from $-3 \rightarrow 3$ has been demonstrated [54].

$$\ell_{\text{OAM}} = \{-m_{-\ell}, -m_{-\ell} + 1, \dots, -1, 0, 1, \dots, m_{\ell} - 1, m_{\ell}\}$$

where m_{ℓ} is the maximal OAM number that can be reached by the setup, and $-m_{-\ell}$ is the smallest that can be reached. This means each OAM quantum number can

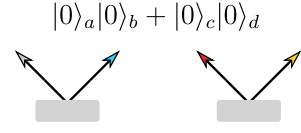


FIG. 7: Two SPDC crystals create the initial state

take on a possible $m_{-\ell} + m_{\ell} + 1$ discrete values and since the system consists of four photons ($i \in \{a, b, c, d\}$) the dimension of the Hilbert space is $\prod_i (m_{-\ell_i} + m_{\ell_i} + 1)$ and each basis ket lives in $\mathbb{C}\Pi_i^{(m_{-\ell_i} + m_{\ell_i} + 1)}$

b. Initial State and SPDC

Here, the initial state is created by a double spontaneous parametric down-conversion (SPDC) process. SPDC is a widespread source for the experimental generation of photon pairs. Multiple SPDC processes can be used to produce multipartite entanglement, as it is well known for the case of two-dimensional polarization entanglement [36, 37]. However, instead of polarization, we are using the OAM of photons [38–40, 55], which is a discrete high-dimensional degree of freedom based on the spatial structure of the photonic wave function

The input state of example 1 is a double-emission from SPDC, which leads to the initial state is of the form of a product state $|0\rangle_a|0\rangle_b \otimes |0\rangle_c|0\rangle_d$ with general form and arbitrary order

$$|\psi\rangle = N \sum_{\ell=-d_c}^{d_c} |\ell\rangle_a |-\ell\rangle_b + |\ell\rangle_c |-\ell\rangle_d \quad (4)$$

with d_c being the highest order of SPDC considered, with photon pairs a,b and c,d. N is a normalization constant. During creation of the dataset we only consider $d_c = 0$ with initial state

$$|\Psi\rangle = |0\rangle_a|0\rangle_b + |0\rangle_c|0\rangle_d \quad (5)$$

This is the unnormalized state produced from the two initial SPDC devices, when we visualize the experiments we depict this as two grey rectangles to depict the crystal as in Figure 7.

Algorithm 1: State Calculation

- 1: **Input** $\mathbf{x} \sim p_{\text{data}}(\mathbf{x})$
 - 2: **Initialize** $|\Psi_0\rangle = |0\rangle_a|0\rangle_b + |0\rangle_c|0\rangle_d$
 - 3: $\mathcal{O}_1, \dots, \mathcal{O}_\ell \leftarrow \mathbf{x}$
 - 4: $|\Psi\rangle \leftarrow \mathcal{O}_\ell \cdots \mathcal{O}_1 |\Psi_0\rangle$
 - 5: $|\Psi\rangle \leftarrow |\Psi\rangle \otimes |\Psi\rangle$
 - 6: $|\Psi\rangle \leftarrow \frac{|\Psi\rangle}{\langle\Psi|\Psi\rangle}$
 - 7: $\rho \leftarrow \text{PartialTrace}(|\Psi\rangle)$
 - 8: $\mathbf{s} \leftarrow -\text{Tr}(\rho \log \rho)$
 - 9: $S = \sum_k s_k = \sum_j S(\rho_{aj}) + \sum_i S(\rho_i)$
 - 10: **return** S
-

3. State and entanglement calculations

To calculate the state of some experimental setup \mathbf{x} that has an initial state from the SPDC process $|\Psi_0\rangle = |0\rangle_a|0\rangle_b + |0\rangle_c|0\rangle_d$ we define the set of operations the state will undergo by extracting the operators defined from the sequence of optical devices that define the experiment. For each $\mathbf{x}_t \in \mathbf{x}$ we have a corresponding operator \mathcal{O}_t that changes the state by acting on it according to Table 2. Each device changes the state in the order defined by the sequence of the experiment and leads to a state $|\Psi\rangle$ which we then square collect four dimensionally entangled terms then normalize: $|\Psi\rangle \otimes |\Psi\rangle / \langle\Psi|\Psi\rangle$. If there are no four particle or the state consists of one of the basis kets then the experiment is unentangled.

Now we must quantify and calculate the entanglement in our system using the systems final state calculated in the previous paragraph and in Algorithm 1. Since our state lives in the hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_D$ as

$$|\Psi\rangle = \sum_{ijkl} \alpha_{ijkl} |i\rangle_a |j\rangle_b |k\rangle_c |l\rangle_d \quad (6)$$

the density matrix of the system is given by

$$\rho = |\Psi\rangle\langle\Psi| = \sum_{ijkl} \sum_{i'j'k'l'} \alpha_{ijkl} \alpha_{i'j'k'l'} |ijkl\rangle\langle i'j'k'l'| \quad (7)$$

We need to keep track of the four reduced density matrices for each subsystem or photon $\rho_a, \rho_b, \rho_c, \rho_d$. For example in the case of a we can calculate $\rho_a = \text{Tr}_{bcd}(\rho)$ by tracing out the other subsystems. Explicitly :

$$\rho_a = \sum_{jkl} \langle l|_d \langle k|_c \langle j|_b \cdot |\Psi\rangle\langle\Psi| \cdot |j\rangle_b |k\rangle_c |l\rangle_d \quad (8)$$

The other three can be calculated in similar fashion. We also need to keep track of the three reduced density matrices for each photon pair $\rho_{ab}, \rho_{ac}, \rho_{ad}$. Similarly in the case of the photon pair ab we can calculate $\rho_{ab} = \text{Tr}_{cd}(\rho)$ by tracing out the other subsystems. Explicitly :The Melvin algorithm for generating our Dataset– 1. generate a random quantum optics setup 2. calculate its state and entanglement measure 3. add

to corresponding dataset depending on entanglement

$$\rho_{ab} = \sum_{kl} \langle k|_c \otimes \langle l|_d \cdot |\Psi\rangle\langle\Psi| \cdot |k\rangle_c \otimes |l\rangle_d \quad (9)$$

The other two can be calculated in similar fashion. we are interested in two main vectors quantifying the entanglement of the system ; 1) the von neumann entropy vector $\mathbf{s} \in \mathbb{R}^7$ and 2) the Schmidt rank vector $\mathbf{r} \in \mathbb{Z}^7$ where

$$\rho = \begin{bmatrix} \rho_a \\ \rho_b \\ \rho_c \\ \rho_d \\ \rho_{ab} \\ \rho_{ac} \\ \rho_{ad} \end{bmatrix} \quad \mathbf{s} = \begin{bmatrix} S(\rho_a) \\ S(\rho_b) \\ S(\rho_c) \\ S(\rho_d) \\ S(\rho_{ab}) \\ S(\rho_{ac}) \\ S(\rho_{ad}) \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} \text{rank}(\rho_a) \\ \text{rank}(\rho_b) \\ \text{rank}(\rho_c) \\ \text{rank}(\rho_d) \\ \text{rank}(\rho_{ab}) \\ \text{rank}(\rho_{ac}) \\ \text{rank}(\rho_{ad}) \end{bmatrix} \quad (10)$$

To explain the SRV, consider the simpler case of 3 particles and the state $|\Psi\rangle$ which has a SRV of (4,2,2)

$$|\Psi\rangle = \frac{1}{2}(|000\rangle + |101\rangle + |210\rangle + |311\rangle) \quad (11)$$

Here, the first particle is four-dimensionally entangled with the other two parties, whereas particle two and three are both only two dimensionally entangled with the rest. Also, $S(\cdot)$ is the von neumann entropy given by

$$S(\rho_a) = -\text{Tr}(\rho_a \log \rho_a) = -\sum_s p_s \log p_s \quad (12)$$

where p_s are the eigenvalues of the quantum system ρ_a .

Entanglement entropy is a measure of the degree of quantum entanglement between two subsystems constituting a two-part composite quantum system. Given a pure bipartite quantum state of the composite system, the reduced density matrix describes the state of a subsystem. The entropy of entanglement is the Von Neumann entropy of the reduced density matrix for any of the subsystems. If it is non-zero, the subsystem is in a mixed state and the two subsystems are entangled.

We define our entanglement measure S as the sum of all the entanglement entropies of all bipartitions of the system

$$S = \sum_{j \neq a} S(\rho_{aj}) + \sum_i S(\rho_i) \quad (13)$$

A bipartition of the system is a partition which divide the system into two parts a and b , containing n_1 and n_2 particles respectively with $n_1 + n_2 = n$ supposing the quantum system consist of n particles. Bipartite entanglement entropy is defined with respect to this bipartition.

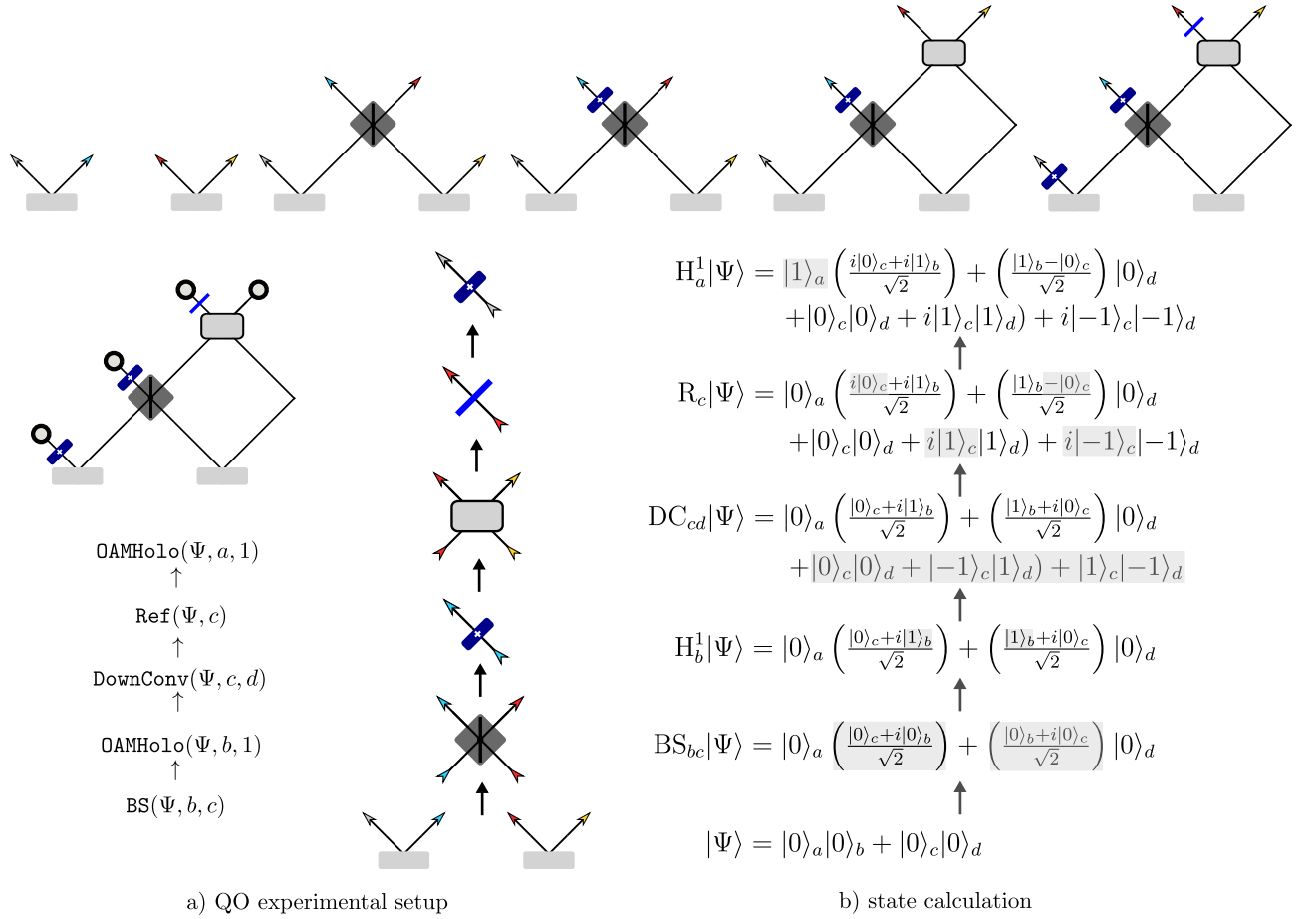


FIG. 8: the graph of QO experimental setup whose state is calculated in the following section

4. Example state calculation

The state calculation is automated using symbolic algebra from the sympy python package cite. For demonstration purposes let's consider an example state calculation of a simple quantum optics experiment defined in Fig 8 the sequence of devices in the experiment operate on the state through the sequence of operators

$$BS_{bc} \rightarrow H_b^1 \rightarrow DC_{cd} \rightarrow R_c \rightarrow H_a^1 \quad (14)$$

starting with the initial state $|\Psi_0\rangle = |0\rangle_a|0\rangle_b + |0\rangle_c|0\rangle_d$ We apply each operator in order to find the final state given by

$$|\Psi\rangle = H_a^1 \cdot R_c \cdot DC_{cd} \cdot H_b^1 \cdot BS_{bc} \cdot |\Psi_0\rangle$$

The first device in the setup is a beamsplitter on photon path b and c with operator BS_{bc} acting on the initial state replacing b and c kets with their superposition as

$$|0\rangle_b \rightarrow \left(\frac{|0\rangle_c + i|0\rangle_b}{\sqrt{2}} \right) \text{ and } |0\rangle_c \rightarrow \left(\frac{|0\rangle_b + i|0\rangle_c}{\sqrt{2}} \right)$$

next the device H_b^1 will add 1 OAM to all b kets, so that the two zero OAM for photon b become $|0\rangle_b \rightarrow |1\rangle_b$ Then applying device $DC_{c,d}$

$$|\Psi\rangle = |\Psi\rangle + |0\rangle_c|0\rangle_d + |-1\rangle_c|1\rangle_d + |1\rangle_c|-1\rangle_d$$

Then applying device R_c and device H_a^1 we flip c and add a i prefactor as well as increase the OAM of kets a. Then we square of the state and normalize:

$$|\Psi\rangle \rightarrow |\Psi\rangle \otimes |\Psi\rangle \text{ and } |\Psi\rangle \rightarrow \frac{|\Psi\rangle}{\langle\Psi|\Psi\rangle}$$

we are left with the following terms that contribute to 4 dimensional entanglement :

$$|\Psi\rangle = |1, 1, -1, -1\rangle + |1, 1, 0, 0\rangle + |1, 1, 1, 1\rangle$$

5. Further experiments

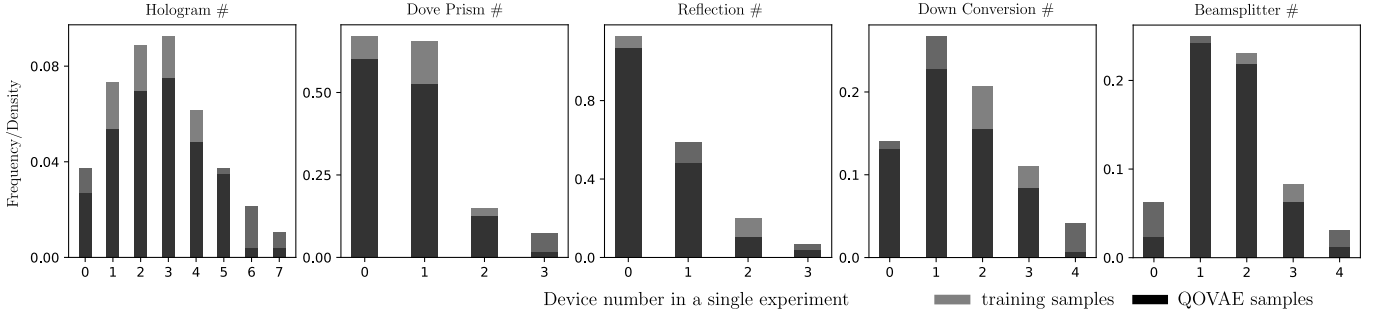


FIG. 9: **Learning structural distributions** We check explicitly if the QOVAE-High has learned global properties of the training distribution. We look at the distribution of devices in generated experiments to tell us what the model has learned about the explicit structure of training experiments. By plotting histograms capturing the distribution of the number of basis element devices per experiment found from random samples of experiments from the training data and the QOVAE we can see how well the QOVAE captured the structural distribution of the training data. To do so we sample 10k experiments from the QOVAE-High and its training data then plot histograms of experiment device number. We focus on the main devices including number of holograms, dove prisms and reflection devices as well as the number of double path devices like Down Converters and Beam-splitters. We can see that the model exactly learns how many DPs and R's exist in the experiments on average. Similar for the double path devices, the histograms are similar and the QOVAE leans them both reasonably well but underestimates slightly. For the hologram we also get a decent but not exact match. Overall, it is safe to say the QOVAE has learned the global structure in the training experiments.

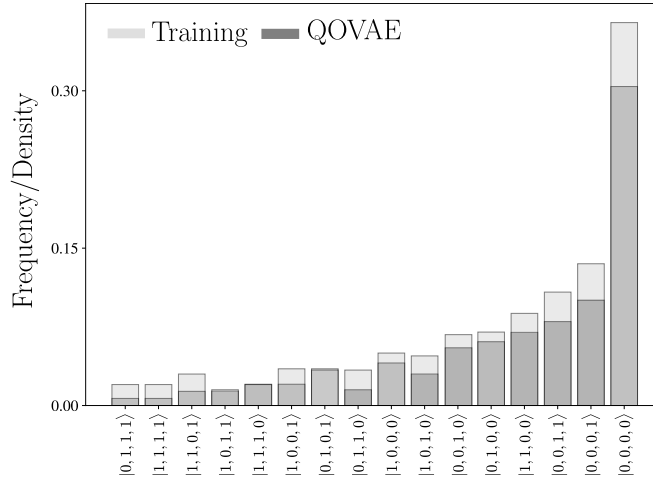


FIG. 10: **Distribution of States** To assess if the QOVAE-high learns a distribution of quantum states in the training data, we directly compare the frequency of basis kets $|i\rangle_a \otimes |j\rangle_b \otimes |k\rangle_c \otimes |l\rangle_d$ arising in the states calculated from the 10K sampled experimental setups from the QOVAE and from the training data. To do this, here we plot histograms of basis kets with photon OAM mode either zero or one from both samples. It is clear that the basis kets with 0 or 1 OAM occur with the same frequency in quantum states produced by the QOVAE or states from the training experiments.

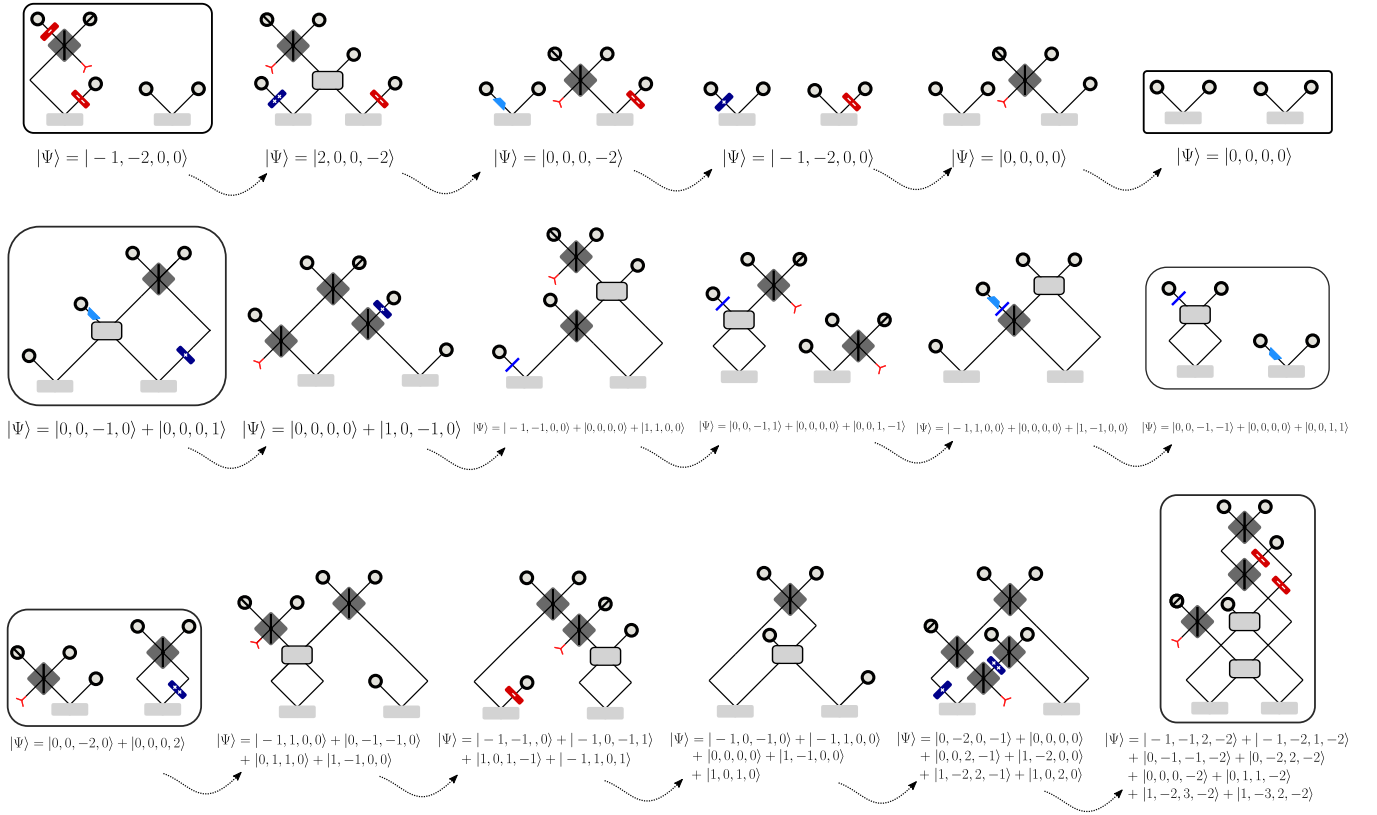


FIG. 11: Interpolations between different quantum states To demonstrate the QOVAE learns a latent representation that encodes a measure of similarity of any experiment's quantum state we provide three interpolations between different states. The first using QOVAE-Low interpolates between two single ket's state. The next two from QOVAE-High : interpolate between a two ket state and three ket state as well as between a two ket state and eight ket state. For the first, we see that along the interpolation path the model decodes single ket states as well. For the second, we see that the interpolated kets have two then three kets. For the third, we see that the states gradually increase their ket number from two to eight.

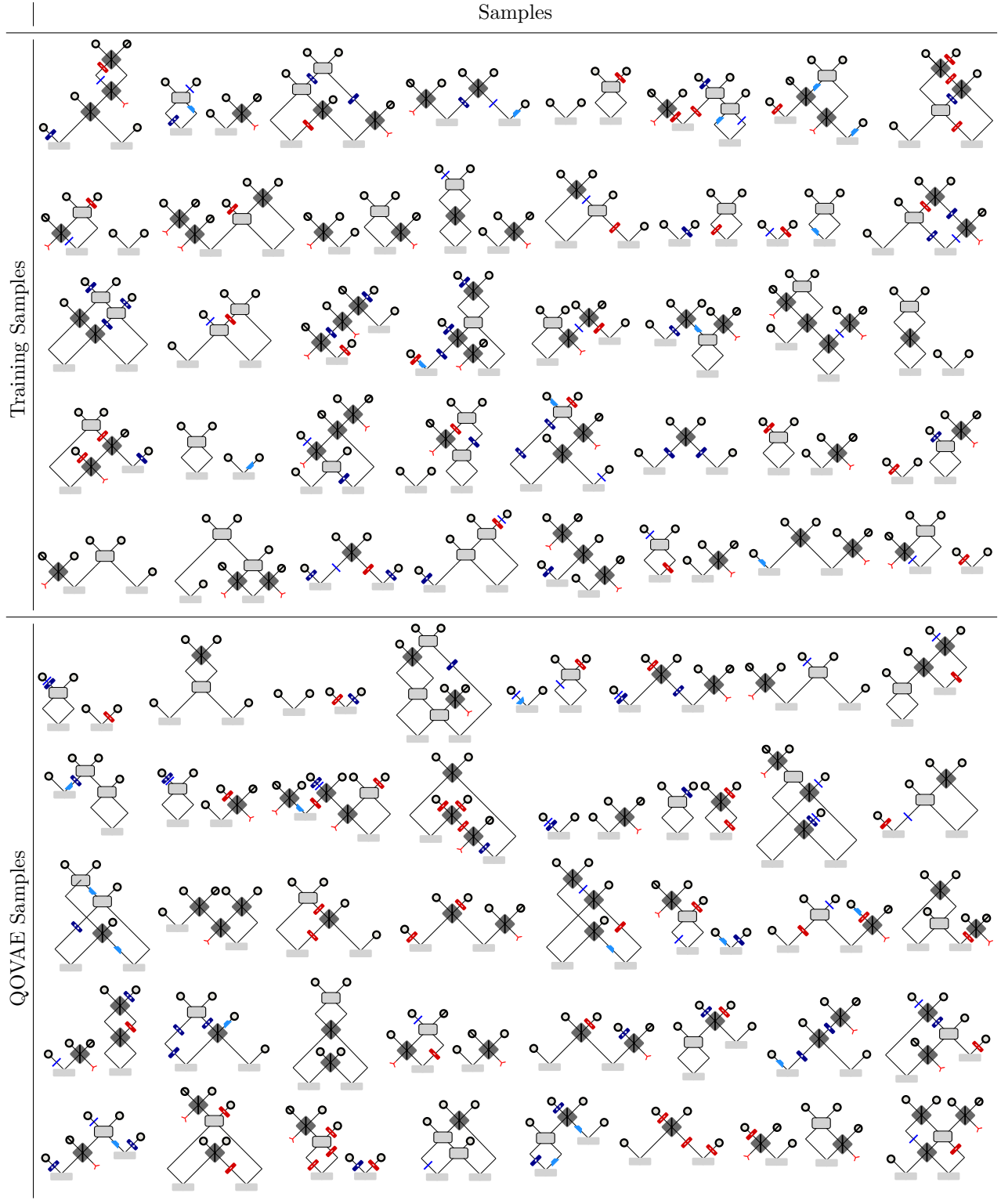


FIG. 12: **More Samples.** We random sample 40 experiments from the training data and from the QOVAE by sampling from the prior $\mathbf{z}^{(s)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and passing that through the decoder $\{f(\mathbf{z}^{(s)})\}$. We display their graph representations. It is clear that both samples display a similar placement of devices and connectivity structures across the four possible photon paths.