TOWARDS AN INDUCTIVE BIAS FOR QUANTUM STATISTICS IN GANS

Hugo Wallner & William R. Clements
ORCA Computing
London, United Kingdom
{hugo, wclements}@orcacomputing.com

ABSTRACT

Machine learning models that leverage a latent space with a structure similar to the underlying data distribution have been shown to be highly successful. However, when the data is produced by a quantum process, classical computers are expected to struggle to generate a matching latent space. Here, we show that using a quantum processor to produce the latent space used by a generator in a generative adversarial network (GAN) leads to improved performance on a small-scale quantum dataset. We also demonstrate that this approach is scalable to large-scale data. These results constitute a promising first step towards building real-world generative models with an inductive bias for data with quantum statistics.

For many machine learning tasks, using latent spaces that match the structure of the data has been shown to be beneficial. For instance, latent spaces where time evolution is described by Hamiltonians (Toth et al., 2020; Greydanus et al., 2019) or by physical symmetries (Connor & Rozell, 2020; Quessard et al., 2020) have been used to achieve excellent results in modelling physical data with an analogous underlying structure. In generative adversarial networks (GANs), it has similarly been shown that making the structure of the latent space used by the generator correspond to that of the data can improve performance (Arici & Celikyilmaz, 2016; Karras et al., 2019).

However, a challenge arises when considering data that originates from a quantum process. Several quantum processes cannot be efficiently simulated using classical (i.e. non-quantum) computational methods (Aaronson & Arkhipov, 2011). Therefore, classical computers may struggle to generate a latent space for a GAN that matches the structure of the data. As an alternative, quantum processors are rapidly improving, with several special-purpose processors now in or near the ”quantum advantage” regime that outperforms the best classical supercomputers (Arute et al., 2019; Madsen et al., 2022). These processors can sample from complex distributions that cannot efficiently be sampled from classically (Hangleiter & Eisert, 2022), but due to their small size and the restricted nature of the sampling problem that they solve it is not yet clear how they can be harnessed for practical applications.

In this work, we investigate the use of a “quantum latent space” for generative adversarial networks, where the inputs to the generator are produced by a quantum processor. We study whether this approach can yield an inductive bias for non-classical statistics in a GAN, and whether this approach is scalable despite the small scale of current quantum processors and the unusual structure of this latent space. We show the following two results:

- Using a quantum process to produce inputs for the generator can improve the performance of a GAN on a toy dataset generated by a similar quantum process
- GANs with a quantum latent space can successfully be trained on a large-scale dataset using current quantum processors

These results constitute a promising first step towards leveraging the statistical structures made available by quantum processors for generative modelling tasks. Further research is however required to identify the real-world datasets that are the most likely to benefit from this approach.

1 Background and Related work

Previous work has highlighted the importance of the latent space of the generator in a GAN, and introduced methods for structuring this space in a way that better reflects the data. For instance, Karras et al. (2019) note that uncorrelated latent spaces are not well suited for capturing correlated features in complex data, and propose a model that learns a more structured intermediate latent space. Other types of structured latent spaces have also been proposed for different types of data, such as Gaussian mixture models (Ben-Yosef & Weinshall, 2018) or a combination of discrete and continuous variables (Mukherjee et al., 2019) for highly diverse data. Alternatively, Arici & Celikyilmaz (2016) proposed using a latent space structured by the discriminator. Even without using a priori knowledge about the structure of the data, Brock et al. (2019) observed through large-scale testing that different latent spaces yield significantly different results in GANs.

Prior work has considered using quantum processors as part of hybrid quantum-classical generative models, such as VAEs (Winci et al., 2020) and GANs (Huang et al., 2021; Rudolph et al., 2022). However, the work most closely related to ours (Rudolph et al., 2022) uses the GAN architecture proposed by Arici & Celikyilmaz (2016), which involves a time-consuming training method that does not scale beyond very small quantum processors. Our work is the first to show that hybrid generative models can scale to quantum processors in the quantum advantage regime and to large-scale data.

The type of quantum processor considered in our work is a boson sampler. A boson sampler consists of a network of optical components in which identical photons interfere with each other (Gard et al., 2015). Since photons are quantum particles, the output of this network is described by a quantum superposition of all the possible outcomes. When a measurement is performed at the output of the network, a single outcome is realized from this superposition. For example, on the right of figure 1, if 3 photons are sent into a 6-channel boson sampling device, one run of the experiment may yield measurement result [3,0,0,0,0,0], where all three photons were detected in the first channel and none in the others, and the next run may yield result [1,1,1,0,0,0] where the three photons were detected in the first three channels. The underlying probability distribution has a complex structure, such that simulating this sampling task is believed to be intractable classically beyond a few tens of photons (Aaronson & Arkhipov, 2011): further information is provided in the appendix. The rapid development of boson samplers in the last few years (Wang et al., 2019) has resulted in two experiments that claim to have achieved the quantum advantage regime (Zhong et al., 2020; Madsen et al., 2022). For example, Madsen et al. (2022) claim that it would take 9000 years for the world's fastest supercomputer to produce a single sample from the same distribution as their boson sampler.

2 Methods

In this work, we investigate whether the use of a quantum processor to produce inputs to the generator can yield an inductive bias for non-classical statistics, and whether this approach is scalable to large-scale datasets. To do so, we perform two experiments: one on a small-scale toy dataset produced by a quantum process, and one and a large-scale image dataset. For both experiments, as illustrated in figure 1 we compare different types of latent space distributions that produce samples z of length L (see fig. 1):

- The commonly used Gaussian distribution (continuous), $z \in \mathbb{R}^L \sim \mathcal{N}(0, I)$.
- The Bernoulli distribution (discrete), $z \in \{0, 1\}^L$, which yielded good performance in Brock et al. (2019).
- The distribution arising from a boson sampling quantum processor with L channels.

We use the Wasserstein generative adversarial network with gradient penalty (WGAN-GP) (Gulrajani et al., 2017) scheme, which has been shown to be more stable during training than several other GAN proposals. To ensure an apples-to-apples comparison, for both our small-scale and large-scale experiments our models differ only through the nature of the process that produced the samples that are sent to the generator and are identical in every other respect. Model and training details are provided in the appendix.

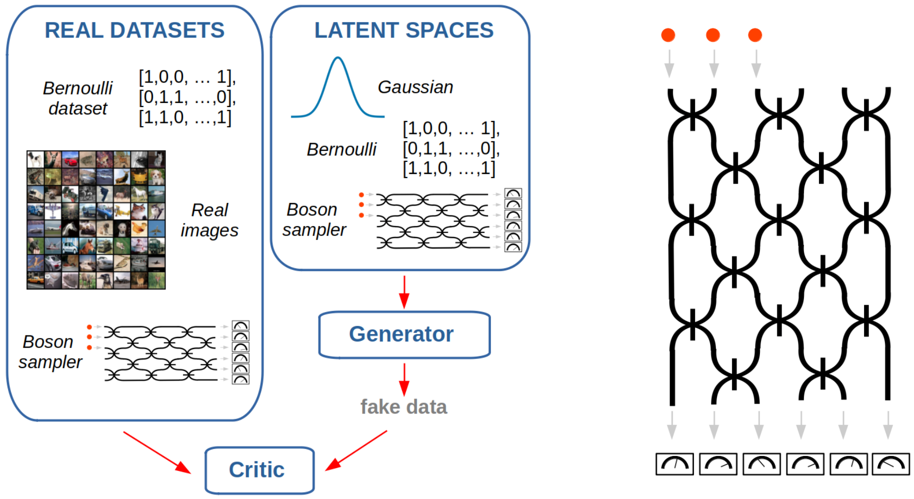

Figure 1: Left: In this work, we compare different types of GANs using different latent spaces and datasets. We focus on quantum latent spaces generated by boson samplers, a type of quantum processor. Right: in a boson sampler, several identical photons enter an optical network and a measurement is performed to determine where the photons left the network. Simulating this experiment beyond a few tens of photons is hard for classical (i.e. non-quantum) computers.

3 Experiments and Results

For our small-scale experiment, we consider two toy datasets: a first classical dataset produced by a 16-dimensional Bernoulli distribution, and a second quantum dataset produced by 8 identical photons in a 16-channel boson sampler. We use a classical algorithm (Clifford & Clifford, 2018) to simulate this small-scale boson sampler. To show that our results are generally applicable, each boson sampler used for each training run uses a different randomly generated optical network, which leads to different outcome probabilities. We also ensure that the boson sampler used to generate the latent space uses a different optical network than that used to generate the data, such that the latent space and the data have different probability densities and the generator cannot simply learn the identity. We also consider an additional latent space produced by a boson sampler with non-identical photons (NIP), which is a system that can be efficiently simulated by a classical process. This classical system is often used as a benchmark against which boson samplers are compared (Walschaers et al., 2016), since even though they produce the same outputs as a boson sampler (in this case, all combinations of 8 photons in 16 channels) they do so with different probabilities.

Our results comparing the performance of the trained models are shown in table 1. Since the two target distributions are discrete, to quantify the performance of a trained model we use the average distance between the outputs of the generator and their nearest integers. A smaller distance indicates that the generator has been more successful in learning the discrete nature of the distribution. We find that the model that uses the quantum latent space is best at generating quantum data, whereas a classical model is best at generating the Bernoulli data. In particular, the quantum model outperforms the model using the non-interfering photons. Since both these latent spaces produce the same vectors corresponding to 8 photons in 16 channels but with different probabilities, this shows that the improved performance of the quantum model on the quantum dataset is partly due to the different statistics of the latent space and not only to the different contents of the latent space vectors.

	Gaussian	Bernoulli	Boson sampler	NIP
Quantum dataset	0.061 ± 0.001	0.065 ± 0.001	$\mathbf{0.036 \pm 0.001}$	0.041 ± 0.002
Bernoulli dataset	$\mathbf{0.012 \pm 0.0015}$	0.0195 ± 0.0125	0.0153 ± 0.0015	0.017 ± 0.002

Table 1: L1 distances between the numbers generated by the GANs with different latent spaces (columns) and their closest integers, for different datasets (rows). The error bounds correspond to the uncertainty of the mean estimated over 12 runs. NIP = non-identical photons.

To further investigate the differences between the models trained with difference latent spaces, in figure 2 we plot the cumulative probabilities of the samples generated by the different GANs. We observe that GANs that use a quantum latent space provide a cumulative distribution that is a better fit to the cumulative distribution of the quantum data, whereas the cumulative distribution generated by the GANs with classical latent spaces better match the classical data. Both these differences in the cumulative distributions and the differences in the abilities to produce integer numbers indicate that using a quantum latent space can indeed yield an inductive bias for quantum statistics in GANs.

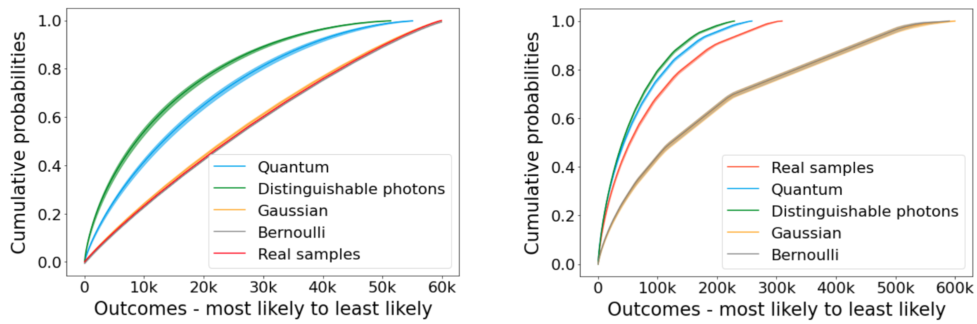

Figure 2: Cumulative probabilities of the real data (red) and the data generated by different GANs trained to approximate Bernoulli (left) and boson sampling distributions (right). For example, in the left figure, the 10,000 most common outcomes samples from the original dataset account for about 20 percent of all outcomes. The curves are averages over 12 runs. We observe that the quantum latent space produces a cumulative distribution closer to the quantum data.

For our large-scale experiment, we use a latent space produced by an actual boson sampling device with an average of 65 photons in 128 channels. To the best of our knowledge, this is the first time a quantum processor of this size has been used for a generative modelling task. Since there is no consensus as to what constitutes a real-world quantum dataset, and this experiment aims firstly to investigate the scalability of this approach, we use the CIFAR-10 image dataset (Krizhevsky et al., 2009). Focusing on this well-studied classical dataset allows us to investigate this approach using pre-existing models and metrics such as the inception score (Salimans et al., 2016).

	Gaussian	Bernoulli	Quantum
Inception score	7.57 ± 0.02	$\mathbf{7.66 \pm 0.039}$	7.54 ± 0.011

Table 2: Inception scores (IS) obtained on CIFAR 10 using different latent spaces. The uncertainty of the mean is estimated over 5 runs.

Table 2 shows the inception scores (IS) Salimans et al. (2016) achieved by the trained models. We find that the performance of the model that uses the quantum latent space is comparable to that of the models that use classical latent spaces, which indicates that the model has trained successfully. This result demonstrates that a quantum processor can indeed be used within a large-scale GAN without a significant loss in performance. We also find that small but statistically significant differences exist between the different latent spaces, which reflects the observation by Brock et al. (2019) that different latent spaces generally lead to different performance. Further research is however required to identify the real-world datasets on which the use of a quantum latent space can be beneficial.

4 Conclusion

In this work, we have shown that using a quantum processor to produce inputs to the generator in a GAN can lead to an inductive bias for non-classical statistics on a toy dataset, and that this approach is scalable to large-scale datasets. These results are a promising first step towards using the statistical structures made available by current quantum processors for real-world generative modelling tasks. We note however that further research is required to identify the real-world datasets that are most likely to benefit from this approach. We also anticipate that this method can be further improved by developing scalable methods for jointly training the quantum processor and the generator, and investigating the use of different types of quantum processors.

REFERENCES

Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pp. 333–342, 2011.

Tarik Arici and Asli Celikyilmaz. Associative adversarial networks. *NeurIPS 2016 Workshop on Adversarial Training*, 2016.

Juan Miguel Arrazola and Thomas R Bromley. Using gaussian boson sampling to find dense subgraphs. *Physical review letters*, 121(3):030503, 2018.

Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

Leonardo Banchi, Nicolás Quesada, and Juan Miguel Arrazola. Training gaussian boson sampling distributions. *Physical Review A*, 102(1):012417, 2020.

Matan Ben-Yosef and Daphna Weinshall. Gaussian mixture generative adversarial networks for diverse datasets, and the unsupervised clustering of images. *arXiv preprint arXiv:1808.10356*, 2018.

Marco Bentivegna, Nicolò Spagnolo, Chiara Vitelli, Fulvio Flamini, Niko Viggianiello, Ludovico Latmiral, Paolo Mataloni, Daniel J Brod, Ernesto F Galvão, Andrea Crespi, et al. Experimental scattershot boson sampling. *Science advances*, 1(3):e1400255, 2015.

Kamil Bradler and Hugo Wallner. Certain properties and applications of shallow bosonic circuits. *arXiv preprint arXiv:2112.09766*, 2021.

Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

Peter Clifford and Raphaël Clifford. The classical complexity of boson sampling. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 146–155. SIAM, 2018.

Marissa Connor and Christopher Rozell. Representing closed transformation paths in encoded network latent space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3666–3675, 2020.

Bryan T Gard, Keith R Motes, Jonathan P Olson, Peter P Rohde, and Jonathan P Dowling. An introduction to boson-sampling. In *From atomic to mesoscale: The role of quantum coherence in systems of various complexities*, pp. 167–192. World Scientific, 2015.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in neural information processing systems*, 32, 2019.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. *Advances in neural information processing systems*, 30, 2017.

Craig S Hamilton, Regina Kruse, Linda Sansoni, Sonja Barkhofen, Christine Silberhorn, and Igor Jex. Gaussian boson sampling. *Physical review letters*, 119(17):170501, 2017.

Dominik Hangleiter and Jens Eisert. Computational advantage of quantum random sampling. *arXiv preprint arXiv:2206.04079*, 2022.

He-Liang Huang, Yuxuan Du, Ming Gong, Youwei Zhao, Yulin Wu, Chaoyue Wang, Shaowei Li, Futian Liang, Jin Lin, Yu Xu, et al. Experimental quantum generative adversarial networks for image generation. *Physical Review Applied*, 16(2):024051, 2021.

Joonsuk Huh, Gian Giacomo Guerreschi, Borja Peropadre, Jarrod R McClean, and Alán Aspuru-Guzik. Boson sampling for molecular vibronic spectra. *Nature Photonics*, 9(9):615–620, 2015.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Lars S Madsen, Fabian Laudenbach, Mohsen Falamarzi Askarani, Fabien Rortais, Trevor Vincent, Jacob FF Bulmer, Filippo M Miatto, Leonhard Neuhaus, Lukas G Helt, Matthew J Collins, et al. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912): 75–81, 2022.

Sudipto Mukherjee, Himanshu Asnani, Eugene Lin, and Sreeram Kannan. Clustergan: Latent space clustering in generative adversarial networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4610–4617, 2019.

Robin Quessard, Thomas Barrett, and William Clements. Learning disentangled representations and group structure of dynamical environments. *Advances in Neural Information Processing Systems*, 33:19727–19737, 2020.

Manuel S Rudolph, Ntwali Bashige Toussaint, Amara Katabarwa, Sonika Johri, Borja Peropadre, and Alejandro Perdomo-Ortiz. Generation of high-resolution handwritten digits with an ion-trap quantum computer. *Physical Review X*, 12(3):031010, 2022.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. *Advances in neural information processing systems*, 29, 2016.

Justin B Spring, Benjamin J Metcalf, Peter C Humphreys, W Steven Kolthammer, Xian-Min Jin, Marco Barbieri, Animesh Datta, Nicholas Thomas-Peter, Nathan K Langford, Dmytro Kundys, et al. Boson sampling on a photonic chip. *Science*, 339(6121):798–801, 2013.

Max Tillmann, Borivoje Dakić, René Heilmann, Stefan Nolte, Alexander Szameit, and Philip Walther. Experimental boson sampling. *Nature photonics*, 7(7):540–544, 2013.

Peter Toth, Danilo Jimenez Rezende, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, and Irina Higgins. Hamiltonian generative networks. In *International Conference on Learning Representations*, 2020.

Mattia Walschaers, Jack Kuipers, Juan-Diego Urbina, Klaus Mayer, Malte Christopher Tichy, Klaus Richter, and Andreas Buchleitner. Statistical benchmark for BosonSampling. *New Journal of Physics*, 18(3):032001, 2016.

Hui Wang, Jian Qin, Xing Ding, Ming-Cheng Chen, Si Chen, Xiang You, Yu-Ming He, Xiao Jiang, L You, Z Wang, et al. Boson sampling with 20 input photons and a 60-mode interferometer in a 10^{14}-dimensional hilbert space. *Physical review letters*, 123(25):250503, 2019.

Walter Winci, Lorenzo Buffoni, Hossein Sadeghi, Amir Khoshaman, Evgeny Andriyash, and Mohammad H Amin. A path towards quantum advantage in training deep generative models with quantum annealers. *Machine Learning: Science and Technology*, 1(4):045028, 2020.

Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.

A Appendix

A.1 Model and training details

For our small-scale experiments, we use fully connected networks with 2 hidden layers of 512 neurons for both the generator and the critic. The generator output is of dimension 16. Both models use a Relu activation function in all their hidden layers. All the latent spaces that we consider also have a dimension of 16. The training is done using batches of 500 samples over 40k iterations. As in Gulrajani et al. (2017), we update the critic 5 times for each generator update. We use a RMSProp optimizer with a learning rate of 5×10^{-4}. The quantum latent space is obtained by sampling from a simulated boson sampler with 8 photons in an interference network with 16 channels. The interference network is mathematically described by a 16×16 unitary matrix, which we draw randomly from the Haar measure for each experiment. For each of the 12 experimental runs, we independently drew 3 random unitary matrices: one for the quantum latent space, one for the non-interfering photons, and one for the data the GAN is trained on.

To obtain the cumulative distributions shown in figure 2, we generate 10M samples from the generator. We compute the probability p_s of each sampled output s, and order these probabilities in decreasing order. The cumulative distribution is then defined by:

$$F_i = \sum_{s=0}^{i} p_s \tag{1}$$

For a uniform distribution, the cumulative distribution is linear in the limit of infinite sampling.

To demonstrate the scalability of our approach, we used the same architecture as was used in Gulrajani et al. (2017) for the CIFAR dataset. All the latent spaces considered are of dimension 128. For the generator, we use a fully connected layer followed by three residual blocks. The colored images that we generate are 32 pixels by 32 pixels. The critic uses a similar architecture with the exception that only the generator uses batch normalisation. We use the Adam optimiser with an initial learning rate of 10^{-4} and a linear decay. The training is done on a single GPU for 100k iterations. The quantum latent space is obtained using 10M samples from a boson sampler with an average of 65 photons in 128 channels.

A.2 Overview of boson sampling

A.2.1 Boson sampling theory

Boson sampling is a non-universal model of quantum computation proposed by Aaronson & Arkhipov (2011), in which identical photons are sent into an interference circuit, and a measurement is performed to determine where the photons left the circuit. An interference circuit with N channels is described by an $N \times N$ unitary matrix U, which can be physically implemented by a wide range of physical devices such as integrated chips (Spring et al., 2013) or optical fiber loops (Madsen et al., 2022). If k identical photons are sent into the first k channels of an interference device, then the probability of measuring an output configuration where t_i photons were found in output channel i is

$$p(t_1, ..., t_N) = \frac{|\text{Perm}(U_{k,T})|^2}{t_1!...t_N!}$$

where Perm indicates the matrix permanent and $U_{k,T}$ is the submatrix of U obtained by taking the first k columns of U and by repeating each row i t_i times (if $t_i = 0$ then row i is not in $U_{k,T}$). The permanent of an n by n matrix M with elements $x_{i,j}$ is defined as

$$\text{Perm}(M) = \sum_{\sigma \in S_n} \prod_{i=i}^{n} x_{i,\sigma_i}$$

where S_n is the set of all permutations of the numbers 1 to n. Although this expression is similar to that of a matrix determinant, unlike a determinant calculating the permanent of a matrix is in general a computationally hard problem. Indeed, if M is an arbitrary real or complex matrix then this problem is #P-hard (Aaronson & Arkhipov, 2011). Since calculating a permanent is hard, Aaronson & Arkhipov (2011) show that the problem of sampling from the output distribution of a boson sampler, even approximately, is also hard. The classical complexity of simulating a boson sampler increases exponentially with the number of photons, such that current classical computers cannot simulate more than a few tens of photons (Clifford & Clifford, 2018).

A.2.2 Technological state of the art

Early proof of principle boson samplers used only a small handful of photons (Spring et al., 2013; Tillmann et al., 2013). These early demonstrations were limited by the fact that the single photon generation process is probabilistic, so the probability of generating N single photons decreases exponentially with N. However, alternative boson sampling schemes that can accommodate the probabilistic nature of single photon generation while still being hard to simulate were soon developed. For example, scattershot boson sampling (Bentivegna et al., 2015) requires only a random subset of photon sources to fire, while Gaussian boson samplers (Hamilton et al., 2017) use photon number superpositions that can be produced deterministically as an input to the boson sampler.

Combined with these improved schemes for boson sampling, significant progress over the last few years in photon source, circuit and detector technology led to two boson sampling experiments in the quantum advantage regime in which a classical simulation is no longer feasible (Zhong et al., 2020; Madsen et al., 2022). Zhong et al. (2020) performed a boson sampling experiment in which up to 76 photons were measured in 100 channels using a fixed interference circuit. Later, Madsen et al. (2022) performed a boson sampling experiment in an optical fiber loop system involving up to 219 photons in 216 channels. The latter experiment contains programmable parameters that determine the interference network and the number of photons and channels, and was made available to the public via a cloud API.

Despite these impressive results, boson sampling was proposed initially solely as a relatively simple method of obtaining a demonstrable quantum advantage versus existing classical computers, and it is not yet clear how they can be harnessed for practical applications. Some promising suggestions have included using them to solve graph problems (Arrazola & Bromley, 2018), spectroscopy problems (Huh et al., 2015), or combinatorial optimisation problems (Banchi et al., 2020; Bradler & Wallner, 2021). However, these proposals have yet to lead to a practical real-world advantage versus other more traditional techniques.

A.2.3 Boson samplers vs qubit-based processors

Many other currently available quantum processors are based on the qubit model of quantum computation, however we felt that for this study boson samplers were more appropriate. Unlike boson samplers, qubit-based processors are universal for quantum computing. However, current publicly available qubit-based quantum processors are either too small to provide inputs to a GAN generator for large-scale data, or their qubits are too noisy to reach the quantum advantage regime. Moreover, the type of random qubit-based circuits that have been used to demonstrate quantum advantage produce measurement results that are very hard to distinguish from a classical Bernoulli distribution (Arute et al., 2019). Using a boson sampler allows us to overcome both challenges and use a large quantum processor in the quantum advantage regime with a latent space that is very different to common alternative classical latent spaces.