The Capacity of Quantum Neural Networks

Logan G. Wright[1,2*] and Peter L. McMahon[1]**

1. School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA
2. NTT Physics and Informatics Laboratories, NTT Research, Inc., East Palo Alto, CA, USA

* lgw32@cornell.edu, **pmcmahon@cornell.edu, https://mcmahon.aep.cornell.edu

What are Quantum Neural Networks?

Quantum neural networks (QNNs) form an important class of algorithm in quantum machine learning [1,2]. It marries two powerful tools - quantum computing and neural networks - and has potential to improve the performance of learning machines [3-5].

Parameterized Models

We can treat any artificial neural network as a black box with input and output data as well as tunable parameters. Thus, *neural networks* can be defined more broadly as *parameterized models*.

Quantization

If we replace a classical model with a quantum one, for the same input size d, it operates on the vectors $\in \mathbb{C}^{2^d}$, instead of vectors $\in \mathbb{R}^d$.* Thus, the dimension is exponentially increased.

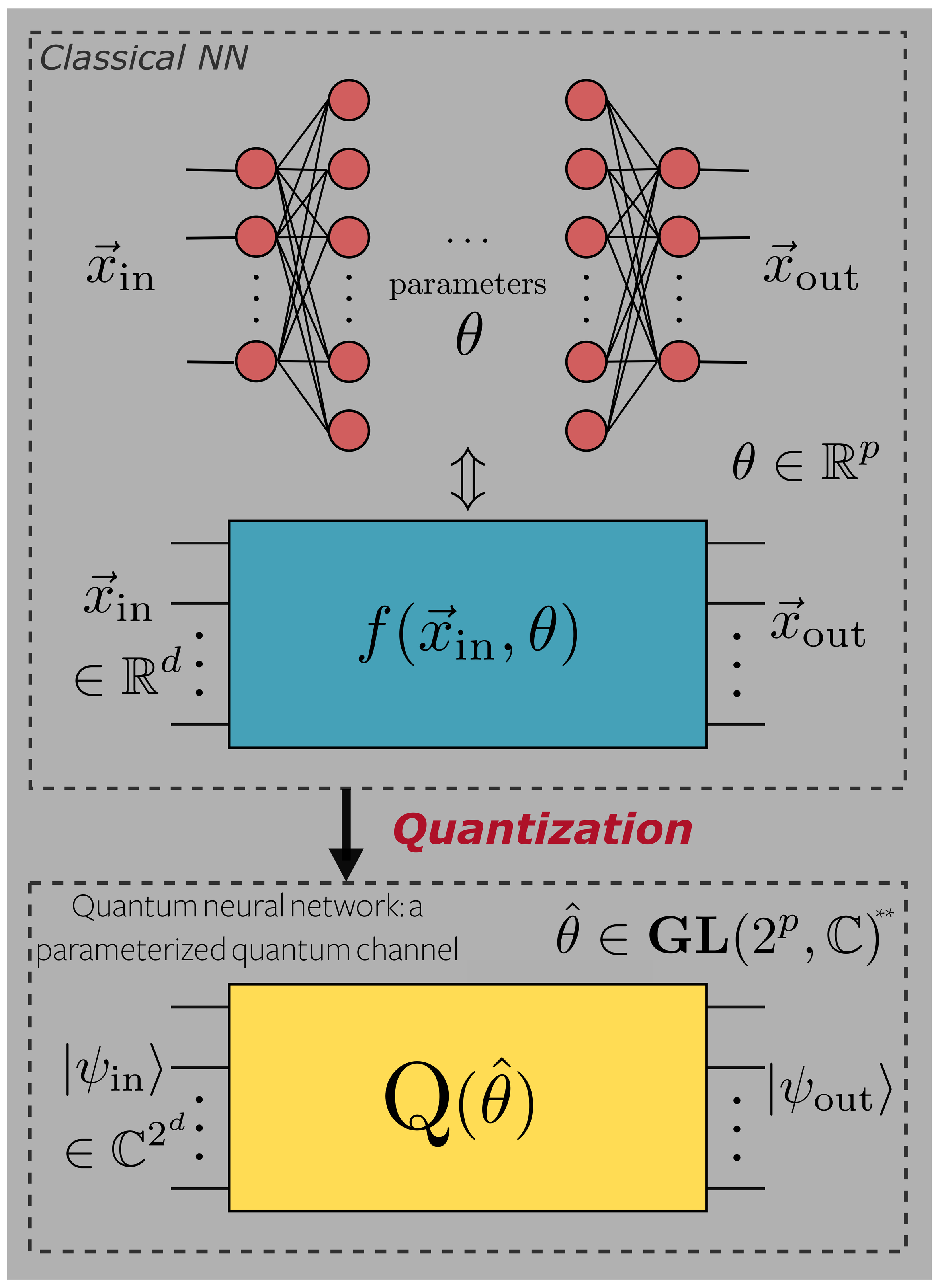
Quantum Training

When training a QNN, we can either use a classical computer, optimizing classical parameters θ, or a quantum one that optimizes parameters stored in a quantum state $\hat{\theta}$.

We will show that performing quantum training can lead to advantages for QNNs. ...

*While we refer to these continuous spaces here, it is essential later to consider that all quantities, inputs, outputs and parameters, are only specifiable to finite-precision.

** Of course, even the most general density matrix has more constraints than a general complex matrix of the same size. Also, it is important to remember that our parameters really have finite precision, regardless of whether they are quantum or classical in nature.

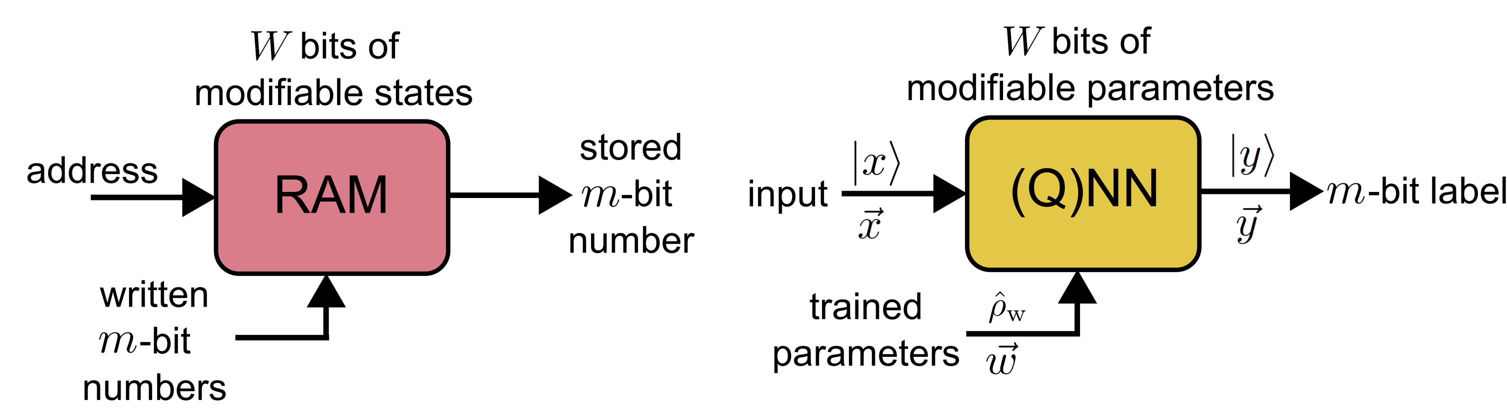


When are QNNs > classical NNs?

Since they may process information within an exponentially-larger space, it seems plausible that QNNs may boast some advantages over purely-classical NNs, in speed of training or inference, in expressive power, or other metrics. Nonetheless, no systematic conditions for when and how QNNs (practical or otherwise) can be better than classical NNs have been developed.

Here, we answer this general question through the lens of information theory. Our results can be viewed as a rigorous means of counting parameters in (Q)NNs, and the basis for an information theory of (Q)NNs (i.e., parameterized machines trained to perform computations, possibly based on quantum physical effects).

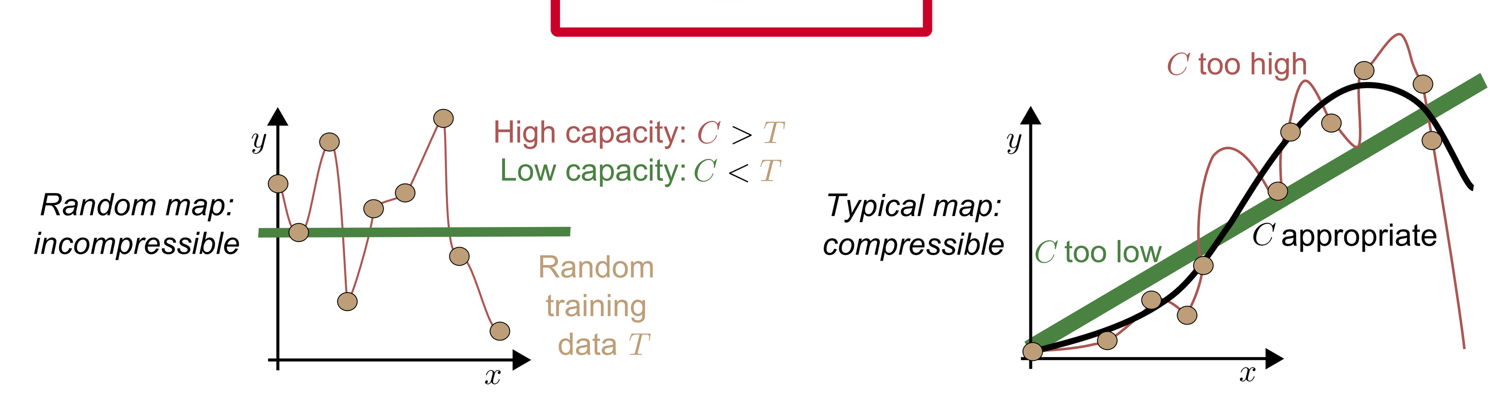
Memory Capacity of (Q)NNs



The capacity of a (Q)NN can be defined analogously to the memory capacity of a RAM: how much information can be `written' to it and then later reliably retrieved? In the limit of a random task (mapping random inputs to random output labels), learning becomes equivalent to memorization, and we can apply well-known results and concepts from Shannon theory. A (Q)NN can be thought of most generally as a parameterized quantum channel, whose inputs, outputs, and parameters may be classical and/or quantum in nature.

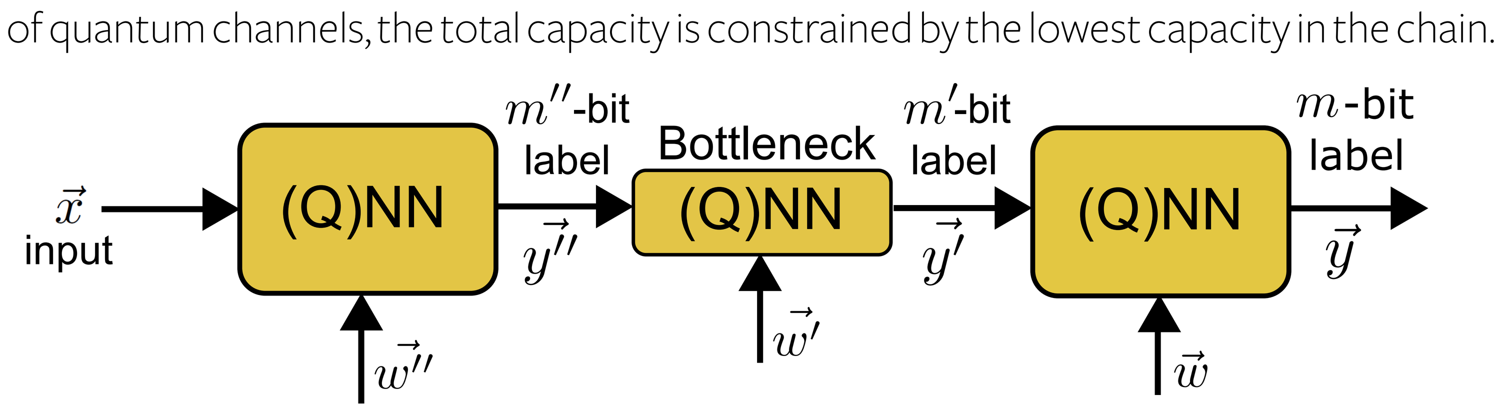
A simple but key result applying to all (Q)NNs follows from the pigeonhole principle: the capacity of a (Q)NN is limited by the information which can be stored in its parameters (W) via the training process:

$$\boxed{C \leq W.}$$



The capacity of a (Q)NN determines how complex the models it can learn are. A crucial innovation of our approach is that it defines capacity in information units (bits and/or qubits). C is therefore the maximum amount of information required to specify a model learned by the (Q)NN, as well as the amount of information that must be provided in training data in order to ensure that the (Q)NN generalizes (i.e., that it is not able to simply memorize the training data).

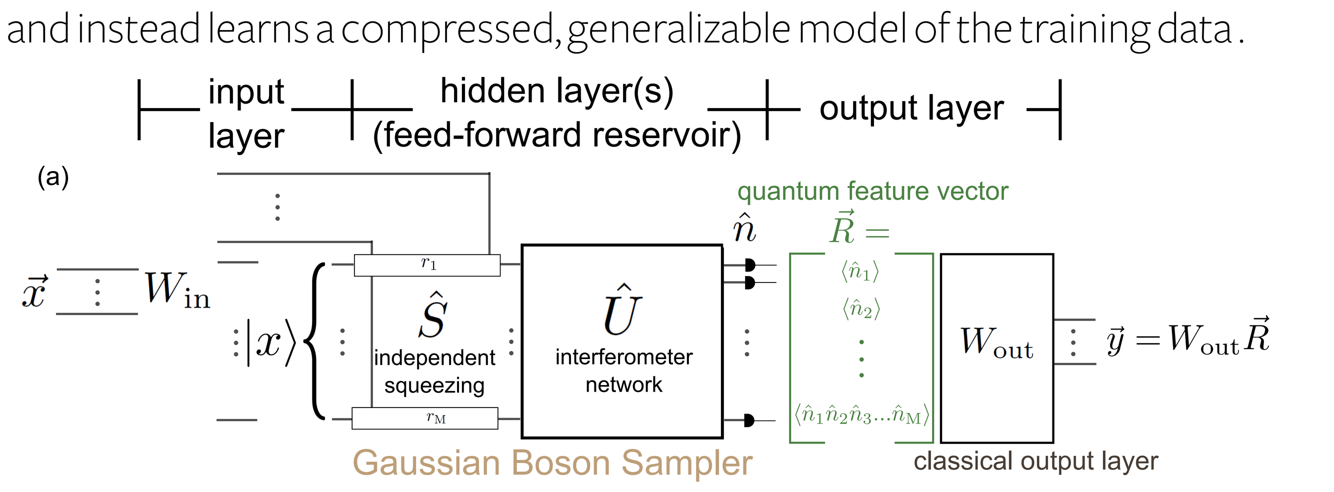
An important concept in the information theory of (Q)NNs is the capacity bottleneck. As a direct result of the information processing inequality, if we view any (Q)NN as a chain of quantum channels, the total capacity is constrained by the lowest capacity in the chain.

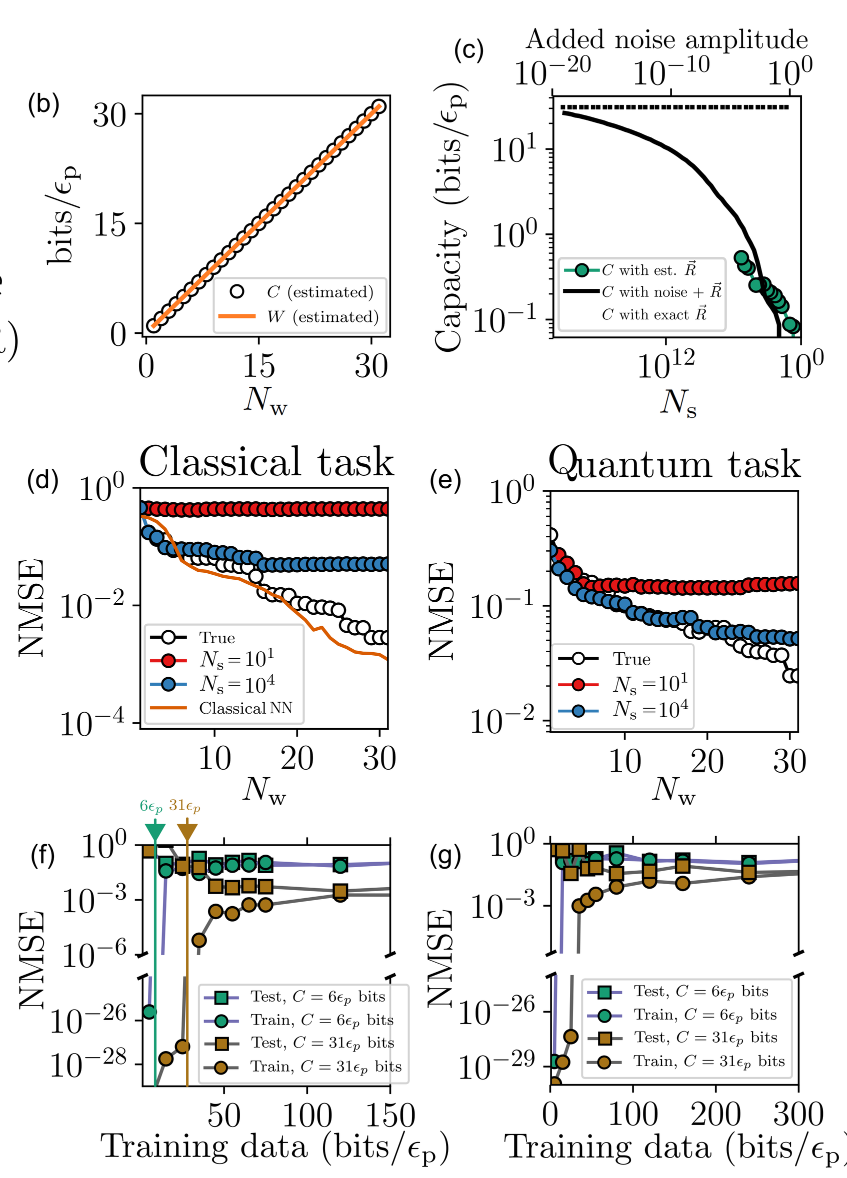


Example: a Boson Sampler QNN

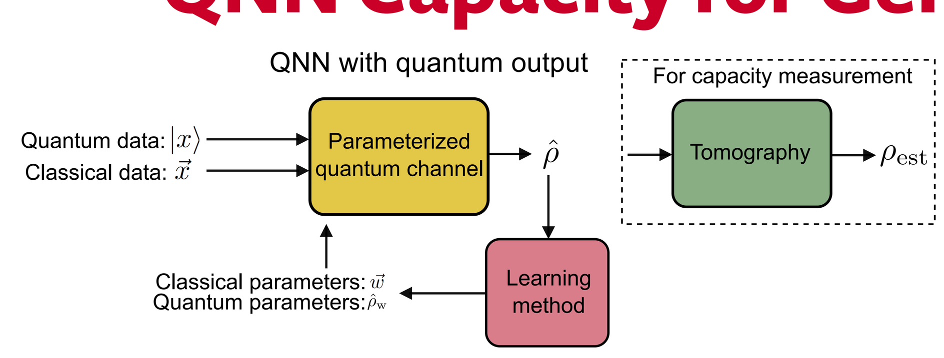
As an example, we simulated a QNN based on a Gaussian Boson Sampler (a), a quantum version of a reservoir computer [6]. We find that the capacity grows linearly with the number of parameters, N_{w}, (b) and logarithmically with the number of measurements made on each run of the device, N_{s} (c). These can be understood analogously to the Shannon-Hartley theorem: $C \sim N_{\text{w}} \log_2(\text{SNR}) \sim N_{\text{w}} \log_2(N_{\text{s}})$

This latter observation demonstrates a read-out-based capacity bottleneck. The QNN can be used to perform tasks on both quantum and classical data (d-e). The capacity of the QNN predicts how much data is required to ensure generalization (f-g): as the amount of training data exceeds the capacity, training and test error converge, evidenticing that the QNN fails to memorize, and instead learns a compressed, generalizable model of the training data.



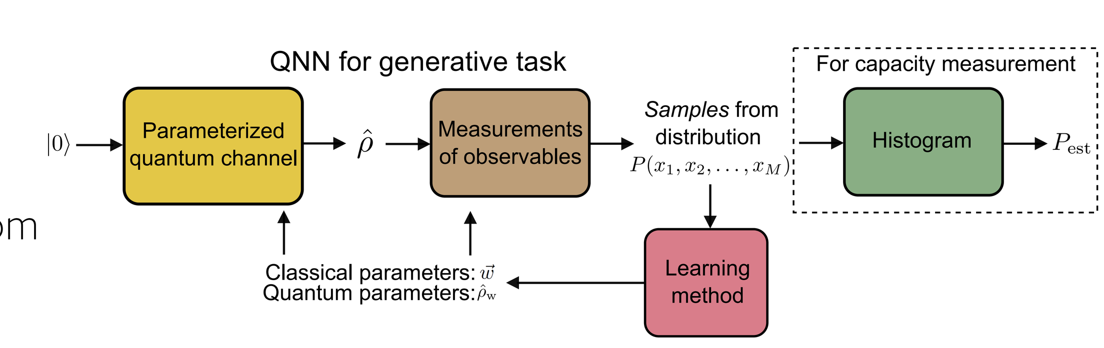


QNN Capacity for Generative and Quantum Tasks



For generative tasks (below), a (Q)NN is used to produce samples from an intended distribution. Here, we can quantify the generative capacity as the amount of information which can be stored and then retrieved from the histograms of samples the QNN can be trained to produce.

We can apply the capacity concept to QNNs that are tasked with performing operations that produce quantum outputs simply by considering the amount of information that can be stored, and then retrieved, through the degrees of freedom controllable within the density matrix describing the states the QNN can learn to produce (above).



Conclusions

1. By defining a measure of expressive power in information units -- the (Q)NN's **capacity** -- we can understand (Q)NNs using information theory.
2. The capacity is the amount of information needed to describe models learnable by the (Q)NN. It is physically bound by the information trainable into the parameters, W. A (Q)NN will generalize when training data provided exceeds C.
3. (Q)NNs with classical parameters do not have a capacity advantage over classical NNs. They may still have other quantum advantages if carefully designed.
4. The capacity of QNNs is often limited (bottlenecked) by measurement noise*.
5. QNNs with quantum parameters may have an exponential capacity advantage over classical NNs, but will require exponentially more data to train.

*notable exceptions apply to generative QNNs and QNNs that act on and/or produce quantum data

Acknowledgements

We thank Tatsuhiro Onodera, Yihui Quek, Keisuke Fujii, Dmitri Pavlichin, Gerald Frieldand, Mario Krell, Brian Coyle and Ashley Montanaro for helpful comments and discussions, and to Shiyuan Ma for assistance in the design of this poster. This work was partially supported by the Impulsing Paradigm Change through Disruptive Technologies (ImPACT) Program of the Council of Science, Technology and Inovation (Cabinet Office, Government of Japan). LGW acknowledges further support from Cornell Neurotech.

References

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. *Nature* **549**, 195–202 (2017).
[2] M. Schuld and N. Killoran, *Phys. Rev. Lett.* **122**, 040504 (2019).
[3] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. *Phys. Rev. A* **98**(3), 032309 (2018).
[4] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd. Preprint at arXiv:1806.06871 (2018).
[5] **L. G. Wright and P. L. McMahon. Preprint at arXiv:1908.01364 (2019).** **[our work]**
[6] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. H. Liew, *npj Quantum Information* **5**, 35 (2019).