

# Variational quantum unsampling on a quantum photonic processor

Jacques Carolan<sup>1\*</sup>, Masoud Mohseni<sup>2</sup>, Jonathan P. Olson<sup>3</sup>, Mihika Prabhu<sup>1</sup>, Changchen Chen<sup>1</sup>, Darius Bunandar<sup>1</sup>, Murphy Yuezhen Niu<sup>2</sup>, Nicholas C. Harris<sup>4</sup>, Franco N. C. Wong<sup>1</sup>, Michael Hochberg<sup>5</sup>, Seth Lloyd<sup>6</sup> and Dirk Englund<sup>1</sup>

**A promising route towards the demonstration of near-term quantum advantage (or supremacy) over classical systems relies on running tailored quantum algorithms on noisy intermediate-scale quantum machines. These algorithms typically involve sampling from probability distributions that—under plausible complexity-theoretic conjectures—cannot be efficiently generated classically. Rather than determining the computational features of output states produced by a given physical system, we investigate what features of the generating system can be efficiently learnt given direct access to an output state. To tackle this question, here we introduce the variational quantum unsampling protocol, a nonlinear quantum neural network approach for verification and inference of near-term quantum circuit outputs. In our approach, one can variationally train a quantum operation to unravel the action of an unknown unitary on a known input state, essentially learning the inverse of the black-box quantum dynamics. While the principle of our approach is platform independent, its implementation will depend on the unique architecture of a specific quantum processor. We experimentally demonstrate the variational quantum unsampling protocol on a quantum photonic processor. Alongside quantum verification, our protocol has broad applications, including optimal quantum measurement and tomography, quantum sensing and imaging, and ansatz validation.**

The construction of a universal error-corrected quantum computer would enable an exponential advantage over the best classical computer in a variety of computational tasks<sup>1,2</sup>. While significant progress has been made in reducing errors on physical qubits beyond the required fault tolerance levels<sup>3,4</sup>, scaling these systems up to a level required for large-scale computing is a major outstanding challenge<sup>5</sup>. Given this difficulty, there has emerged a significant effort towards algorithms for noisy intermediate-scale quantum processors, which can solve problems without the need for full-scale error correction<sup>6,7</sup>. Not only would such a machine reveal a fundamental gap between the computational power of the quantum and classical worlds<sup>8</sup>, it could also potentially advance combinatorial optimization<sup>9</sup>, quantum simulation<sup>10</sup> and neural networks<sup>11–15</sup>.

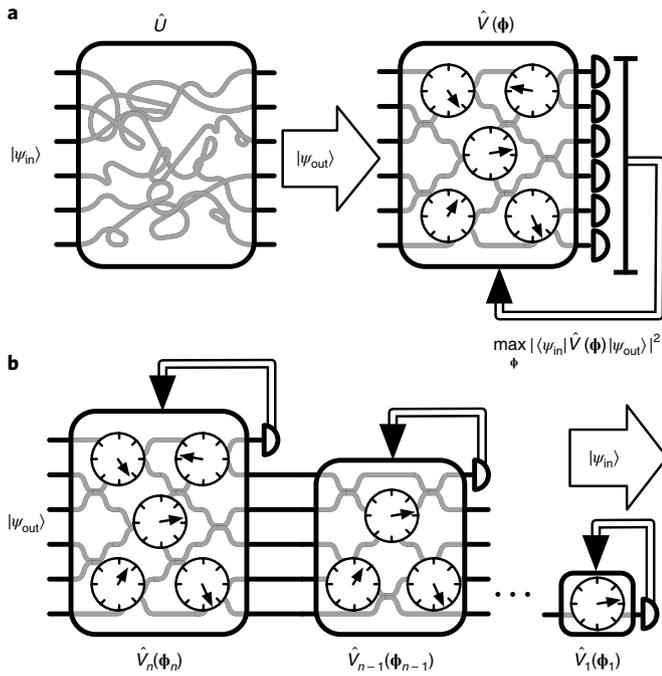
Hardware-specific quantum algorithms have been developed to demonstrate a quantum advantage<sup>16,17</sup>. Moreover, the system requirements (such as noise or qubit number) to show an unambiguous advantage have been analysed<sup>18,19</sup>. Generally, quantum advantage algorithms for noisy intermediate-scale quantum processors follow a similar structure: showing that under reasonable complexity-theoretic conjectures, efficient classical sampling from a distribution  $p_U(x) \equiv |\langle x | \psi_{\text{out}} \rangle|^2$  is intractable. Here  $|\psi_{\text{out}}\rangle = \hat{U} |\psi_{\text{in}}\rangle$  is a quantum state generated by a quantum circuit  $\hat{U}$  acting on an input state  $|\psi_{\text{in}}\rangle$ , and  $\{|x\rangle\}$ , for example, is the set of bit strings in the computational basis. As experiments reach the regime where they can no longer be classically simulated<sup>20–22</sup>, the question of verification becomes paramount. Unlike problems such as factoring that are in the nondeterministic polynomial-time (NP) complexity class and therefore can be efficiently verified<sup>23</sup>, sampling problems typically exist outside this class and efficient verification may not be possible<sup>24</sup>. Machine-level verification techniques have been developed

using information about the physical system to achieve efficient verification<sup>25</sup>, but a hardware-independent approach to verification is outstanding.

Rather than determine properties of an output state given knowledge of the circuit, we ask: given direct access to the state  $|\psi_{\text{out}}\rangle$ , can we efficiently learn the physical/computational operation  $\hat{U}$ , or approximate  $\hat{U}$  such that can we generate  $|\psi_{\text{out}}\rangle$ ? Here we develop the variational quantum unsampling (VQU) protocol that performs optimization on  $|\psi_{\text{out}}\rangle$  using a controllable auxiliary quantum circuit  $\hat{V}(\Phi)$ , which is a functional of control parameters  $\Phi$  (see Fig. 1). Inspired by neural network approaches to machine learning, our approach approximates the effect of an unknown time-reversed quantum operation  $\hat{V}(\Phi) \approx \hat{U}^\dagger$  to learn the quantum circuit that recovers a known input state such that  $\hat{V}(\Phi) |\psi_{\text{out}}\rangle \approx |\psi_{\text{in}}\rangle$ . In general, our variational learning procedure amounts to partial characterization of unknown unitary operations given knowledge of their actions over certain input states. Consequently, our approach can be understood as a variational approach to partial quantum process tomography<sup>26</sup>.

Partial tomography of general quantum states requires a number of measurements that scales with the dimensionality of the system<sup>27</sup>. However, in the special case where quantum states are well described by an efficient representation such as matrix product states, efficient tomography schemes have been proposed<sup>28</sup>. In contrast, variational approaches may be able to adaptively learn the underlying structure of the quantum channel to perform efficient partial tomography in special cases<sup>29</sup>. While variational quantum algorithms have been developed for computational tasks such as classification<sup>12,13,30</sup> and simulation<sup>10,31</sup>, we instead focus on the verification of circuit outputs.

<sup>1</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Google Quantum AI Laboratory, Venice, CA, USA. <sup>3</sup>Zapata Computing Inc, Cambridge, MA, USA. <sup>4</sup>Lightmatter, Boston, MA, USA. <sup>5</sup>Elenion Technologies, New York, NY, USA. <sup>6</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. \*e-mail: [jacques.carolan@nbi.ku.dk](mailto:jacques.carolan@nbi.ku.dk)



**Fig. 1 | VQU.** **a**, Given a state  $|\psi_{out}\rangle = \hat{U} |\psi_{in}\rangle$ , the task is to find the circuit that returns  $|\psi_{in}\rangle$ , thus determining some features of  $\hat{U}$ . The VQU protocol feeds this state into a controllable quantum circuit and optimizes the parameters  $\phi$  to find the time-reversed condition that  $\hat{V}(\phi)|\psi_{out}\rangle \approx \hat{U}^\dagger |\psi_{out}\rangle$  over the known input state. **b**, Directly optimizing for this condition is inefficient in the qubit number; therefore, the layer-wise approach breaks the problem up such that at each stage only a polynomially sized subset of the entire Hilbert space is optimized for.

**Variational learning**

Formally, the unsampling problem asks: given direct access to a polynomial number of copies of  $|\psi_{out}\rangle$ , find a circuit that returns the known input state  $|\psi_{in}\rangle$ , thus determining some elements of  $\hat{U}$ . Prima facie, one can imagine taking  $|\psi_{out}\rangle$  and coherently passing it through an appropriately parametrized circuit  $\hat{V}(\phi)$  (see Fig. 1a). In the language of machine learning, we can define a loss function

$$L(\phi) = 1 - |\langle \psi_{in} | \hat{V}(\phi) | \psi_{out} \rangle|^2 \tag{1}$$

that quantifies the distance between the output state and the input state, and is bounded  $L(\phi) \in [0, 1]$ . Searching for the condition that

$$\min_{\phi} L(\phi) = 0 \tag{2}$$

leads to  $\hat{V}(\phi) \approx \hat{U}^\dagger$  over a given input state. That is, the circuit that generates  $|\psi_{in}\rangle$  is found, corresponding to a single column of  $\hat{U}$ . Note, however, that without a well-chosen ansatz, we could have  $|\langle \psi_{in} | \hat{V}(\phi) | \psi_{out} \rangle|^2 \approx 1/D$ , where  $D$  is the dimension of the system, which typically scales exponentially in the particle number. The probability for an individual event is therefore exponentially unlikely and estimating the associated probability takes exponential time. Moreover, it has recently been shown that gradient-based quantum circuit learning becomes exponentially inefficient due to a very flat loss landscape if one starts with a generic random initial state over the entire Hilbert space<sup>32</sup>; however, initialization strategies have been proposed to overcome this obstacle<sup>33</sup>. Here we opt for a divide-and-conquer approach that selects efficiently accessible

subspaces of the entire Hilbert space for the stochastic optimization, within a layer-wise model of learning.

Towards this end, we use multiple unitaries  $\hat{V}_k$ , or ‘layers’, and a layer-by-layer training approach, which at each stage optimizes over only a polynomially sized subset of the full Hilbert space (see Fig. 1b). To illustrate this procedure, consider an  $n$ -qubit system with the known pure initialization state, with tensor product structure (such as a mean-field state),  $|\psi_{in}\rangle = |\alpha_1, \alpha_2, \dots, \alpha_n\rangle$ , where  $|\alpha_i\rangle$  is the state of the  $i$ th qubit. The first training stage feeds  $|\psi_{out}\rangle$  into a circuit  $\hat{V}_n(\phi_n)$  acting on all  $n$  qubits. Letting  $\rho_1 = \hat{V}_n(\phi_n) |\psi_{out}\rangle \langle \psi_{out}| \hat{V}_n^\dagger(\phi_n)$ , the optimization then varies circuit parameters  $\phi_n$  to minimize

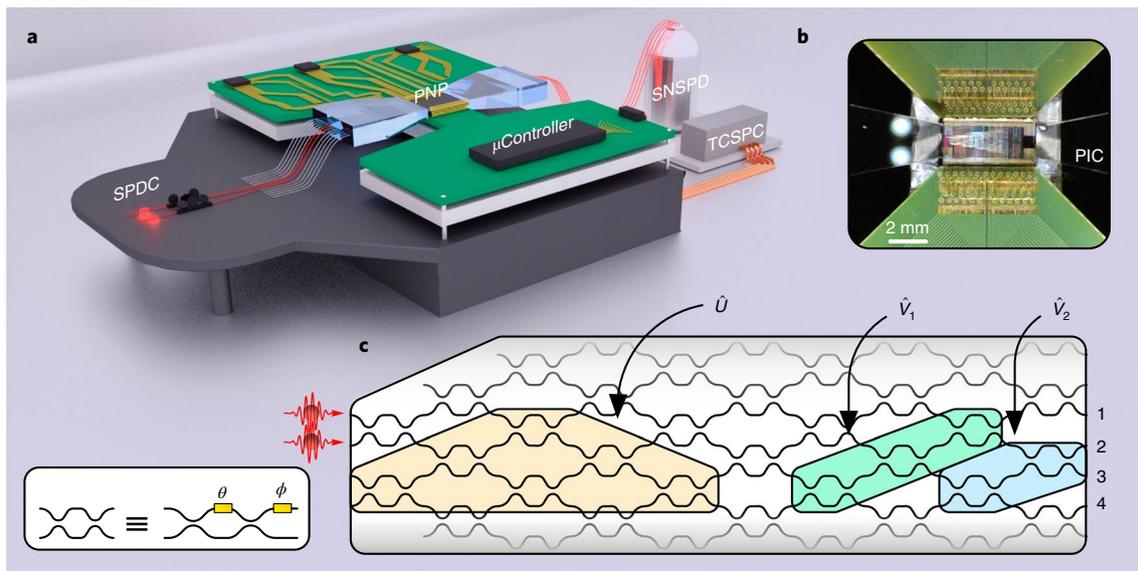
$$L_1(\phi_n) = 1 - \langle \alpha_1 | \text{tr}_{2\dots n}(\rho_1) | \alpha_1 \rangle \tag{3}$$

If  $L_1(\phi_n) = 0$ , then the first qubit is successfully found in the state  $|\alpha_1\rangle$  and the remainder of the qubits  $\rho_1' = \text{tr}_1(\rho_1)$  are in a pure state. The state  $\rho_1'$  is then fed into a circuit  $\hat{V}_{n-1}(\phi_{n-1})$  acting on the remaining  $n-1$  qubits and  $L_2(\phi_{n-1})$  is minimized to maximize the overlap between the second qubit and  $|\alpha_2\rangle$ . This process is repeated for  $n$  stages and if successful each subsequent stage will disentangle a qubit until the total output is  $|\alpha_1, \alpha_2, \dots, \alpha_n\rangle$ . Critically, the probability estimated at each stage is now exponentially boosted, with  $L_i(\phi)$  scaling as  $\mathcal{O}(1)$  (independently of  $n$ ). While boosting the cost function does not guarantee an efficient gradient, layer-wise approaches can minimize the circuit depth, which may ameliorate vanishing gradients. Moreover, the error in a single stage of unsampling  $L(\phi) \approx \epsilon$  should scale as  $\epsilon \ll 1/n$ , such that as  $n$  becomes large the overall unsampling fidelity does not vanish.

Our protocol enables a twofold approach to verification. First, the solution unitary is given by  $\hat{V}_{sol} = \prod_{i=1}^n \hat{I}_{n-i} \otimes \hat{V}_i(\phi_i)$  (where  $\hat{I}_j$  is the identity operation acting on the first  $j$  qubits), which enables direct verification of the sampling circuit. Repeating this process for multiple input basis states yields additional information about  $\hat{U}$ . Second, deviation from  $|\psi_{in}\rangle$  signals decoherent error in the sampling protocol, which can be further inspected by tomography on a reduced subset of qubits. A layer-wise training approach with conditional feedforward was recently used for quantum state discrimination<sup>14</sup> and recognizing quantum states of matter<sup>29</sup>.

We numerically tested this protocol for up to five qubits. These simulations, which are provided in Supplementary Section I, converged to numerical precision in all instances. We conjecture that this efficacy is due to an over-parameterization effect: each layer effectively decouples a qubit from the remainder of the state, and there may be many such circuit settings  $\phi$  that achieve this condition, whereby  $L_i(\phi) = 0$ . For certain classes of classical deep neural networks, this over-parameterization has been shown to both increase expressiveness<sup>34</sup> and accelerate training<sup>35</sup>. Similar layer-wise training approaches have found success in training particular classes of classical neural networks without the need for backpropagation<sup>36,37</sup>.

While the existence of a circuit  $\hat{V}_i(\phi_i)$  is guaranteed provided that  $\rho_{i-1}$  is pure, the above example does not specify how one can physically build the circuit that implements the unitary, and in general constructing an arbitrary unitary requires a circuit depth that grows exponentially in the number of qubits<sup>38</sup>. Each VQU protocol can leverage unique structure in the specific sampling problem to construct a reduced family of unitaries that can be efficiently implemented. If the system parameterization (otherwise known as the ansatz)  $\mathcal{A}_U$  is known, then VQU can be used to characterize the sampling circuit. However, if the system parameterization is unknown or not known to be optimal, VQU can be used to assess whether a test ansatz  $\mathcal{A}_V$  can represent the state, or set of states given by  $\mathcal{A}_U$ . We refer to this procedure as ‘ansatz validation’, which is related to quantum circuit compiling<sup>39,40</sup>.



**Fig. 2 | Optical VQU in a quantum photonic processor.** **a**, Pairs of photons are first generated via SPDC and delivered to a PNP that is mounted on top of a Peltier cooling system to maintain thermal stability. After propagating through the circuit, photons are out-coupled and delivered to an array of superconducting nanowire single-photon detectors (SNSPDs). Coincidence events are recorded by a time-correlated single-photon counting system (TCSPC), and are output to a classical computer that controls a micro-controller unit ( $\mu$ Controller) that drives all 176 on-chip thermo-optic phase-shifters. **b**, An optical micrograph of the 26-mode PNP showing all 240 wire bonds (including grounds) and in/out coupling via two custom-built photonic integrated circuits (PIC). The total footprint of the device is  $4.9 \times 2.2$  mm. **c**, A schematic of the PNP with separate regions marked for the unsampling protocol. Each MZI comprises an internal  $\theta$  and external  $\phi$  phase-shifter (inset). The orange circuit implements the sampling operation, and the green and blue circuits implement the first and second layer of the unsampling protocol, respectively. The full protocol requires active control of 46 thermo-optic phase-shifters.

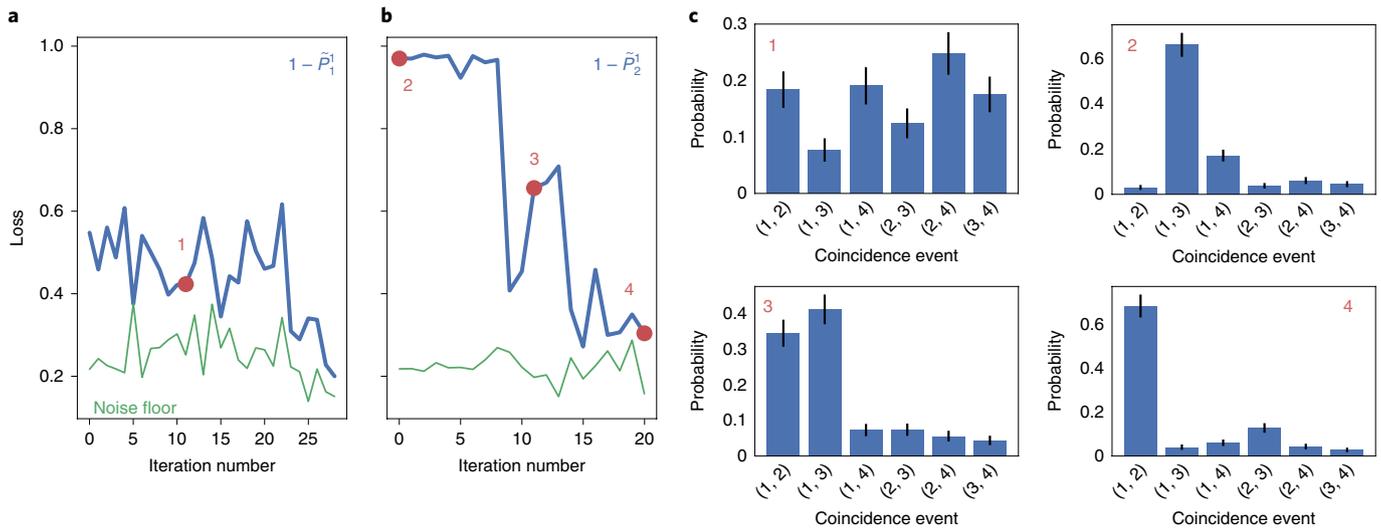
Certain classes of state (for example, Haar random) may require an exponential number of circuit elements to exactly generate; however, this provides a unique opportunity: VQU may compile an approximate form of the state with respect to a given ansatz that, given the noise inherent in deep quantum circuits, may generate a more accurate representation of the target state than an ansatz that is theoretically able to generate the exact state in a noiseless environment. If the form of the ansatz is poorly compatible with the chosen state, it is possible that the optimization of certain layers could be ignorant of the form of the global optimization and lead to a local minimum. This is common when choosing an ansatz independent of the form of the target state and can potentially be mitigated if certain properties or symmetries of the target state are known a priori<sup>41</sup>, additional steps are added to the training<sup>15</sup> or additional gates are adaptively added to the ansatz<sup>42</sup>. While it remains an open question exactly which ansatz are amenable to VQU, in Supplementary Section II we give an example of one such ansatz that is related to the fractional quantum Hall effect.

### Optical VQU

Boson sampling is a mathematical proof (based on conjectures) that shows that ensembles of indistinguishable photons, when acted on by linear optical circuits (arrays of beamsplitters and phase-shifters), generate samples from a probability distribution that cannot be efficiently generated classically<sup>16</sup>. Formally, given an  $n$ -photon initialization state of one photon per mode  $|\psi_{\text{in}}\rangle = |1_1 1_2 \dots 1_n\rangle$  (where  $|i_j\rangle$  represents  $i$  photons in the  $j$ th optical mode), each amplitude of the output state  $|\psi_{\text{out}}\rangle = \varphi(\hat{U}_m)|\psi_{\text{in}}\rangle$  is given by the permanent of a unique  $n \times n$  submatrix of the  $m$ -dimensional unitary  $\hat{U}_m$  (ref. 43), where  $\varphi(\hat{U}_m)$  represents the multi-photon unitary. The output distribution  $p_U(x) = |\langle x|\psi_{\text{out}}\rangle|^2$  is therefore also related to permanents, a notoriously difficult function to calculate<sup>44</sup>, with  $\{|x\rangle\} = \{|i_1 i_2 \dots i_m\rangle\}$  being the set of collision-free computational basis states such that  $\sum_j i_j = n$  with  $i_j \leq 1$ .

In optics, an arbitrary  $m$ -dimensional unitary operator  $\hat{U}_m$  across  $m$  optical modes can always be constructed out of  $m(m-1)/2$  reconfigurable beamsplitters and phase-shifters<sup>45</sup>. This theorem therefore provides an efficient circuit ansatz for the optical VQU protocol. Efficient characterization schemes exist to estimate  $\hat{U}_m$  using either classical states of light<sup>46</sup> or small-scale quantum states<sup>47</sup>. However, these approaches do not capture the full multi-photon process  $\varphi(\hat{U}_m)$ , which may involve non-trivial entanglement between photons or imperfections in the photonic input state. Coarse-grain multi-photon verification techniques have also been developed that rule out efficiently simulable alternative distributions, rather than verify the quantum state directly<sup>25</sup>. Limiting our discussion to the regime of  $n$  photons in  $m = n^2$  optical modes, the optical VQU protocol first feeds  $|\psi_{\text{out}}\rangle$  into a  $n^2$ -dimensional circuit  $\hat{V}_1(\Phi_n)$  and minimizes the loss function (3), which maximizes  $\hat{P}_1$ , the probability of one and only photon in the first mode. Note, in the optical case  $|\psi_{\text{in}}\rangle = |1_1\rangle$  and the trace operation occurs over the optical mode basis. Critically, the probability of exactly one photon in the  $i$ th optical mode  $\hat{P}_i$  scales as  $\mathcal{O}(1/n)$  and is therefore an efficiently accessible measurement (see Supplementary Section III-A for a proof of this). Given  $L_1(\Phi_n) \approx 0$ , the second layer  $\hat{V}_2(\Phi_{n-1})$  is an  $n^2 - 1$  mode circuit acting on  $n - 1$  photons that maximizes the probability of one photon in the second optical mode. The optical VQU protocol proceeds for a total of  $n$  layers until the initialization state  $|1_1 1_2 \dots 1_n\rangle$  is recovered. In Supplementary Section III-B we show that the full protocol requires  $\mathcal{O}(n^3)$  parameters.

There are two primary noise sources in linear optical systems: coherent (for example, over/under-rotation in the MZIs) and incoherent (for example, photon loss and distinguishability). In Supplementary Section V, we show that VQU is inherently robust to coherent error, which can be understood from the well-bounded spectrum norm of linear optical Hamiltonians<sup>48</sup>. Distinguishability between photons may be mitigated through filtering, which due to post-selection in photonic experiments only reduces the overall efficiency and does not affect the cost function  $L_k$  at each iteration of optimization.



**Fig. 3 | Experimental results.** Variation in the loss function during the optical VQU protocol. **a**, The first layer (blue) minimizes  $L_1 = 1 - \tilde{P}_1^i$  to find a photon in the first optical mode. **b**, The second layer (blue) minimizes  $L_2 = 1 - \tilde{P}_2^i$  to find a photon in the second optical mode. The green line plots the experimental noise floor (see the text). The red points correspond to probability distribution time slices as shown in **c**. **c**, The probabilities for all six twofold coincidence events are plotted with error bars assuming Poissonian counting statistics. The first plot (top left) shows support across all coincidence events while at the end of the VQU protocol (bottom right)  $P = 0.695 \pm 0.053$  is found in the  $(1, 2)$  coincidence event, corresponding to the initialization state  $|1, 1\rangle_2$ .

### Experimental unsampling

We implement a proof-of-concept demonstration of the optical VQU procedure on a state-of-the-art quantum photonic processor comprising three stages: photon generation via spontaneous parametric down-conversion (SPDC), reconfigurable quantum circuitry on a programmable nanophotonic processor (PNP), and single-photon detection, all within an actively configured feedback loop for optimization (see Fig. 2a). See Methods for a description of the full experimental set-up. The PNP comprises 88 reconfigurable Mach-Zehnder interferometers (MZIs) arranged in a mesh, which enables different regions of the device to be used for separate quantum operations. In Fig. 2c the sampling circuit is shown in orange, and the unsampling layers are shown in green and blue.

The sampling circuit (Fig. 2c, orange) directly dials six MZIs (12 phases) to generate a four-mode random unitary according to the Haar measure<sup>49</sup>. Two photons pass through the sampling circuit and the output state is fed into the first unsampling layer (Fig. 2c, green); a four-mode circuit acting on modes  $\{1, 2, 3, 4\}$ . The classical optimizer is programmed to find a single photon in optical mode 1 by minimizing  $L_1(\Phi_1) = 1 - \tilde{P}_1^i(\Phi_1)$ . Each iteration of the optimization collects approximately 100 twofold coincidence events. As shown in Fig. 3a, the phases  $\Phi_1$  are first randomly initialized and  $L_1 = 0.55$  and after 28 iterations of the optimization  $L_1 = 0.20$ , which is at the noise floor of our experiment. The output state is then fed to the second unsampling layer; a three-mode circuit acting on modes  $\{2, 3, 4\}$  (Fig. 2c, blue). The optimizer is set to find a single photon in mode 2 by minimizing  $L_2(\Phi_2) = 1 - \tilde{P}_2^i(\Phi_2)$ . Crucially, by leaving mode 1 untouched  $L_1$  cannot increase. The phases  $\Phi_2$  are randomly initialized and  $L_2 = 0.97$ . As shown in Fig. 3b, after just 20 stages of optimization  $L_2 = 0.31$ , which is  $1.3\sigma$  from the noise floor of our experiment.

The final fidelity of the VQU protocol, defined as the overlap between the initialization state and the output state  $\mathcal{F} = |\langle \psi_{\text{in}} | \psi_{\text{out}} \rangle|^2$ , was found to be  $\mathcal{F} = 0.695 \pm 0.053$ , which is  $1.2\sigma$  from the maximal achievable fidelity given the noise floor of our experiment (Fig. 3a,b, green). This noise floor is primarily due to a low signal-to-noise ratio, caused by photon emission from our thermo-optic phase-shifters and high fibre-to-chip coupling loss ( $-8$  dB facet to facet).

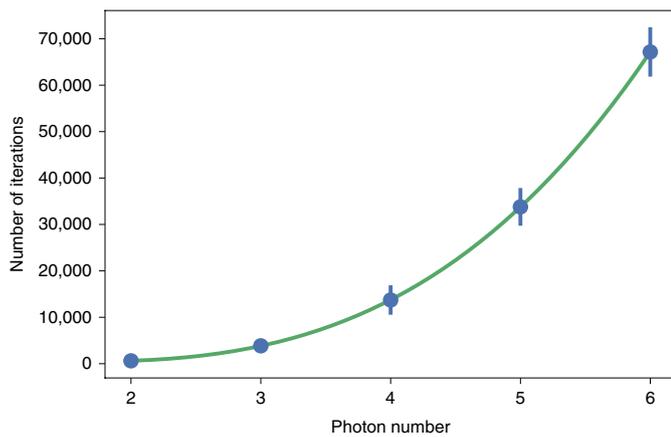
The deviation from the maximum possible fidelity is probably due to the performance of our optimizer in the presence of finite counts. Notwithstanding, we estimate the probability observed samples are due to random circuit settings to be  $p = 0.004$  (see Methods for further details). Future implementations will use either low-loss in/out couplers<sup>50</sup> or on-chip single-photon sources<sup>51,52</sup> and detectors<sup>53</sup> to increase the signal-to-noise ratio and boost fidelity.

Alongside the proof-of-concept experimental demonstration, extensive numerical simulations were performed for up to six photons. In Fig. 4, we plot the number of iterations required to converge to a fidelity of  $\mathcal{F} = 1 - 10^{-5}$ , alongside an expected cubic fit for  $n = 100$  runs (see Methods for further details). The efficiency of these numerical experiments suggests that the presence of local optima is limited and unlikely to prevent convergence for optical unsampling experiments.

### Concluding remarks

We have introduced the VQU protocol: a nonlinear quantum neural network approach for verification and inference of near-term quantum processors. Our protocol leverages a divide-and-conquer approach that selects efficiently accessible subspaces of the entire Hilbert space for optimization. Within a layer-wise learning model, we simulate the effect of an unknown time-reversed quantum operation to recover a known input state. We demonstrated this protocol optically on a quantum photonic processor.

Our approach can be directly applied to the verification and certification of circuit outputs, and for the comparison and training of circuit ansatz. Moreover, VQU could also lend itself to the characterization of other physical processes that can be probed by quantum signals such as molecular excitations<sup>54</sup>. Applied to optical systems, VQU may find application as a subroutine in quantum cryptographic protocols<sup>55</sup> or for optimal receivers for optical communications<sup>56</sup>. Here, VQU is akin to adaptive optics for quantum optical systems, correcting quantum data transmission through turbulent free-space links or mode-mixing fibres. More generally, partial tomography schemes have direct application in a wide range of quantum information protocols where properties such as the topology or symmetry of a state are required. For example, in



**Fig. 4 | Monte Carlo numerics.** Full boson unsampling Monte Carlo numerics for up to  $n=6$  photons in  $m=36$  modes. The total number of steps required to converge to  $\mathcal{F}=1-10^{-5}$  for each photon number is plotted in blue, alongside an expected third-order polynomial fit in green with  $R^2=1-10^{-6}$ . The error bars represent one standard deviation from  $N=100$  runs.

quantum error correction partial information about a quantum state can be used to estimate an error syndrome for bosonic error correction codes<sup>57</sup> or diagnose systematic errors in quantum circuits<sup>58</sup>. Applying variational approaches such as VQU to error correction problems is a promising research direction.

As quantum processors push the limits of what is classically simulable and coherent control of quantum phenomena advances, the problem of quantum state and circuit verification represents a formidable challenge. We therefore anticipate VQU in particular, and other layer-wise learning models more generally, serving as a vital tool in the arsenal of the quantum engineer.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-019-0747-6>.

Received: 3 June 2019; Accepted: 11 November 2019;  
Published online: 13 January 2020

### References

- Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2010).
- Montanaro, A. Quantum algorithms: an overview. *npj Quantum Inf.* **2**, 15023 (2016).
- Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
- Gaebler, J. P. et al. High-fidelity universal gate set for  ${}^9\text{Be}^+$  ion qubits. *Phys. Rev. Lett.* **117**, 060505 (2016).
- Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
- Mohseni, M. et al. Commercialize quantum technologies in five years. *Nature* **543**, 171–174 (2017).
- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Harrow, A. W. & Montanaro, A. Quantum computational supremacy. *Nature* **549**, 203–209 (2017).
- Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. Preprint at <http://arXiv.org/abs/1411.4028v1> (2014).
- McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 023023 (2016).
- Romero, J., Olson, J. P. & Aspuru-Guzik, A. Quantum autoencoders for efficient compression of quantum data. *Quantum Sci. Technol.* **2**, 045001 (2017).
- Farhi, E. & Neven, H. Classification with quantum neural networks on near term processors. Preprint at <http://arXiv.org/abs/1802.06002v1> (2018).
- Schuld, M., Bocharov, A., Svore, K. & Wiebe, N. Circuit-centric quantum classifiers. Preprint at <http://arXiv.org/abs/1804.00633> (2018).
- Chen, H., Wossnig, L., Severini, S., Neven, H. & Mohseni, M. Universal discriminative quantum neural networks. Preprint at <http://arXiv.org/abs/1805.08654v1> (2018).
- Steinbrecher, G. R., Olson, J. P., Englund, D. & Carolan, J. Quantum optical neural networks. *npj Quantum Inf.* **5**, 60 (2019).
- Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. In *Proc. 43rd Annual ACM Symposium on Theory of Computing* 333–342 (ACM, 2011).
- Boulard, A., Fefferman, B., Nirkhe, C. & Vazirani, U. On the complexity and verification of quantum random circuit sampling. *Nat. Phys.* **15**, 159–163 (2019).
- Boixo, S. et al. Characterizing quantum supremacy in near-term devices. *Nat. Phys.* **14**, 595–600 (2018).
- Neville, A. et al. Classical boson sampling algorithms with superior performance to near-term experiments. *Nat. Phys.* **13**, 1153–1157 (2017).
- Zhang, J. et al. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature* **551**, 601–604 (2017).
- Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring. In *Proc. 35th Annual Symposium on Foundations of Computer Science* 124–134 (Society for Industrial and Applied Mathematics, 1994).
- Hangleiter, D., Kliesch, M., Eisert, J. & Gogolin, C. Sample complexity of device-independently certified ‘quantum supremacy’. *Phys. Rev. Lett.* **122**, 210502 (2019).
- Carolan, J. et al. On the experimental verification of quantum complexity in linear optics. *Nat. Photon.* **8**, 621–626 (2014).
- Mohseni, M., Rezaekhani, A. T. & Lidar, D. A. Quantum-process tomography: resource analysis of different strategies. *Phys. Rev. A* **77**, 032322 (2008).
- O’Donnell, R. & Wright, J. Efficient quantum tomography. In *Proc. 48th Annual ACM Symposium on Theory of Computing* 899–912 (ACM, 2016).
- Cramer, M. et al. Efficient quantum state tomography. *Nat. Commun.* **1**, 149 (2010).
- Cong, I., Choi, S. & Lukin, M. D. Quantum convolutional neural networks. *Nat. Phys.* <https://doi.org/10.1038/s41567-019-0648-8> (2019).
- Grant, E. et al. Hierarchical quantum classifiers. *npj Quantum Inf.* **4**, 65 (2018).
- Kokail, C. et al. Self-verifying variational quantum simulation of lattice models. *Nature* **569**, 355–660 (2019).
- McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812 (2018).
- Grant, E., Wossnig, L., Ostaszewski, M. & Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. Preprint at <http://arXiv.org/abs/1903.05076> (2019).
- Eldan, R. & Shamir, O. The power of depth for feedforward neural networks. In *Proc. Conference on Learning Theory* (eds Feldman, V., Rakhlin, A. & Shamir, O.) 907–940 (JMLR, 2016).
- Arora, S., Cohen, N. & Hazan, E. E. On the optimization of deep networks: implicit acceleration by overparameterization. In *Proc. 35th International Conference on Machine Learning, ICML 2018* (eds Dy, J. & Krause, A.) 372–389 (International Machine Learning Society, 2018).
- Bengio, Y., Lamblin, P., Popovic, D. & Larochelle, H. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing 19* (eds Schölkopf, B., Platt, J. C. & Hoffman, T.) 153–160 (MIT Press, 2007).
- Hettinger, C. et al. Forward thinking: building and training neural networks one layer at a time. Preprint at <http://arXiv.org/abs/1706.02480v1> (2017).
- Barenco, A. et al. Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).
- Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. *Phys. Rev. A* **98**, 032309 (2018).
- Khatris, S. et al. Quantum-assisted quantum compiling. *Quantum* **3**, 140 (2019).
- Romero, J. et al. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Sci. Technol.* **4**, 014008 (2018).
- Grimsley, H. R., Economou, S. E., Barnes, E. & Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **10**, 3007 (2019).
- Scheel, S. Permanents in linear optical networks. Preprint at <http://arXiv.org/abs/quant-ph/0406127> (2004).

44. Valiant, L. G. The complexity of computing the permanent. *Theor. Comput. Sci.* **8**, 189–201 (1979).
45. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, H. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994).
46. Rahimi-Keshari, S. et al. Direct characterization of linear-optical networks. *Opt. Exp.* **21**, 13450–13459 (2013).
47. Laing, A. & O'Brien, J. L. Super-stable tomography of any linear optical device. Preprint at <http://arXiv.org/abs/1208.2868> (2012).
48. Rubin, N. C., Babbush, R. & McClean, J. Application of fermionic marginal constraints to hybrid quantum algorithms. *New J. Phys.* **20**, 053020 (2018).
49. Russell, N. J., Chakhmakhchyan, L., O'Brien, J. L. & Laing, A. Direct dialling of Haar random unitary matrices. *New J. Phys.* **19**, 033007 (2017).
50. Notaros, J. et al. Ultra-efficient CMOS fiber-to-chip grating couplers. In *Proc. Optical Fiber Communication Conference M21–M25* (2016).
51. Silverstone, J. W. et al. On-chip quantum interference between silicon photon-pair sources. *Nat. Photon.* **8**, 104–108 (2014).
52. Carolan, J. et al. Scalable feedback control of single photon sources for photonic quantum technologies. *Optica* **6**, 335–340 (2019).
53. Najafi, F. et al. On-chip detection of non-classical light by scalable integration of single-photon detectors. *Nat. Commun.* **6**, 5873 (2015).
54. Brinks, D. et al. Visualizing and controlling vibrational wave packets of single molecules. *Nature* **465**, 905–908 (2010).
55. Guha, S. et al. Quantum enigma machines and the locking capacity of a quantum channel. *Phys. Rev. X* **4**, 011016 (2014).
56. Guha, S. Structured optical receivers to attain superadditive capacity and the holevo limit. *Phys. Rev. Lett.* **106**, 240502 (2011).
57. Niu, M. Y., Chuang, I. L. & Shapiro, J. H. Hardware-efficient bosonic quantum error-correcting codes based on symmetry operators. *Phys. Rev. A* **97**, 032323 (2018).
58. Moroder, T. et al. Certifying systematic errors in quantum experiments. *Phys. Rev. Lett.* **110**, 180401 (2013).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

**Numerical methods.** While the protocols we present in the body of the manuscript are agnostic to the particular optimization algorithm, in practice we necessitate two conditions: the optimization algorithm should be (1) local, so as to converge efficiently, and (2) gradient free, as in general the gradients will not be a priori known. We have determined through extensive numerical studies that the BOBYQA algorithm<sup>59</sup> performs well in terms of speed and accuracy, satisfies (1) and (2) and is readily implemented in the NLOPT library<sup>60</sup>. Consequently, all numerical experiments presented in this manuscript use this algorithm.

**Experimental set-up.** Pairs of degenerate photons at 1,582 nm are generated via SPDC from a custom-fabricated periodically poled KTiOPO<sub>4</sub> crystal under extended phase-matching functions<sup>61</sup>. Photon pairs are then collected into optical fibres and delivered to a PNP<sup>62–64</sup> via a custom-built optical interposer that reduces the mode field diameter of the input fibres to better match that of silicon waveguides (Fig. 2b). The PNP consists of 176 individually tuneable phase-shifters across 26 optical modes, fabricated in a complementary metal–oxide–semiconductor (CMOS) compatible silicon photonics process. On-chip MZIs are controlled via two thermo-optic phase-shifters, with an internal phase shift  $\theta$  for splitting ratio configuration and an external phase shift  $\phi$  for phase configuration.

After passing through the PNP, photons are out-coupled and delivered to four tungsten silicide superconducting nanowire single-photon detectors with ~65% quantum efficiency for photon counting. Correlations across each channel are recorded by a time-correlated single-photon counting system, and are then fed to a classical computer for processing. On the basis of recorded coincidence events across all  $\binom{4}{2} = 6$  coincidence channels  $\{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$

(where  $(i, j)$  represents a coincidence event between optical modes  $i$  and  $j$ ), a classical optimizer running the local derivative-free BOBYQA algorithm<sup>59</sup> varies the PNP layer phases to minimize a user-defined loss function.

**Monte Carlo data analysis.** To estimate the probability that our photon statistics are due to random circuit settings, we perform a Monte Carlo analysis. To do this, we numerically model our experiment and randomly sample  $N=1,000$  phase-shifter values uniformly in the range  $\phi \in [0, 2\pi)$ . On the basis of these random circuit settings, we calculate  $N=1,000$  unique multi-photon probability distributions assuming a Hong–Ou–Mandel visibility of  $V=0.9$  (independently verified). For each distribution, we calculate the fidelity with the desired (that is, identity) distribution, finding a mean fidelity over random circuit settings to be  $\bar{\mathcal{F}} = 0.247$ . By analysing the subsequent probability density function of fidelities, we estimate the probability of finding a fidelity greater or equal to our measured fidelity  $\mathcal{F} = 0.695$  to be  $p = 0.004$ . We also determine the average fidelity when unitaries are sampled uniformly (according to the Haar measure) to be  $\mathcal{F} = 0.167$  and estimate the probability that our results are due to random unitaries as  $p = 0.002$ .

**Optical VQU numerical experiments.** In the following, we describe the optical VQU numerical experiments, for up to  $n=6$  photons, presented in Fig. 4. While optimizing  $\tilde{P}_j^1$  for  $j \in [1, n]$  is sufficient to perform the VQU protocol, in many cases this performs poorly due to the number of parameters involved in the optimization and the absence of an analytic expression for the gradient. We determined that reducing the parameter set by first bringing all photons into the first  $n$  modes was superior in terms of speed and accuracy of the unsampling protocol.

To compress  $n$  photons into the first  $n$  modes, we perform the following protocol.

1. Generate an  $n^2$ -dimensional sampling unitary via the Haar measure.
2. Pass the output state into an  $n^2$ -dimensional unsampling circuit with all phases  $(\alpha, \phi)_{jk} = (0, 0)$ .
3. For  $j \in [1, n]$  and  $k \in [n^2, 1]$ , optimize  $(\alpha, \phi)_{jk}$  to minimize the photon flux in the  $k+1$  optical mode.

A single iteration of this protocol is sufficient to yield a  $>0.99$  probability of all photons in the first  $n$  modes. To achieve numerical accuracy, we repeat this three times.

Next, to unsample  $n$  photons in  $n$  modes, we perform the following protocol.

1. Pass the output state into an  $n$ -dimensional unsampling circuit with all phases  $(\alpha, \phi)_{jk} = (0, 0)$ .

**Table 1 | Error in fit parameters for optical VQU scaling**

Model	$1 - R^2$ error
$a + bx$	0.12
$a + bx + cx^2$	$1.5 \times 10^{-3}$
$a + bx + cx^2 + dx^3$	$9.6 \times 10^{-7}$
$a + be^{cx+d}$	$1.1 \times 10^{-3}$

2. Maximize the probability of any event,  $\tilde{P}_1^n$ , in the  $j=1$  optical mode over all parameters  $(\alpha, \phi)_{jk}$ .
3. Append an  $(n-1)$ -dimensional unsampling circuit acting on modes  $[2, n]$  and maximize  $\tilde{P}_2^n$ .
4. Repeat for a  $(n-j)$ -dimensional circuit and  $\tilde{P}_j^n$  for  $j \in [3, n-1]$ .

In case numerical accuracy is not achieved, we allow for random restarts, which is included in the total number of iterations plotted in Fig. 4. To estimate the scaling, we fit the number of iterations required to reach numerical accuracy, against the photon number, for a range of hypothesis models. In Table 1 we show the  $1 - R^2$  error for each model.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

59. Powell, M. J. The bobyqa algorithm for bound constrained optimization without derivatives (2009).
60. Johnson, S. G. The NLOpt nonlinear-optimization package (2011); <http://ab-initio.mit.edu/nlopt>
61. Chen, C. et al. Efficient generation and characterization of spectrally factorable biphotons. *Opt. Exp.* **25**, 7300–7313 (2017).
62. Carolan, J. et al. Universal linear optics. *Science* **349**, 711–716 (2015).
63. Harris, N. C. et al. Quantum transport simulations in a programmable nanophotonic processor. *Nat. Photon.* **11**, 447–452 (2017).
64. Harris, N. C. et al. Linear programmable nanophotonic processors. *Optica* **5**, 1623–1631 (2018).

## Acknowledgements

We thank E. Farhi, E. Grant, D. Hangleiter, I. Marvian, J. McClean, M. Pant, M. Schuld, P. Shadbolt, S. Sim and G. Steinbrecher for insightful discussions. This work was supported by the AFOSR MURI for Optimal Measurements for Scalable Quantum Technologies (FA9550-14-1-0052), the MITRE Quantum Moonshot Program and by the AFOSR programme FA9550-16-1-0391, supervised by G. Pomrenke. J.C. is supported by EU H2020 Marie Skłodowska-Curie grant number 751016.

## Author contributions

J.C., M.M., S.L. and D.E. conceived the project. J.C., M.M., J.P.O., M.Y.N., S.L. and D.E. developed the theory. J.C., M.P., C.C., D.B., N.C.H., F.N.C.W., M.H. and D.E. contributed to the experimental set-up. J.C., M.P., C.C. and D.B. performed the experiment and analysed data. J.C. performed numerical experiments. All authors contributed to the discussion of the results and the writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41567-019-0747-6>.

**Correspondence and requests for materials** should be addressed to J.C.

**Peer review information** *Nature Physics* thanks Ashley Montanaro and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).