























either a classical discriminator or a quantum discriminator, and the cost function is optimized to minimize the discrimination probability with respect to real samples. The target application is to accelerate classical GANs using quantum computers.

**Quantum neural network architectures.** Several quantum neural network (QNN) architectures have been proposed; for instance, REFS<sup>153,173,174</sup> proposed perceptron-based QNNs. In these architectures, each node in the neural network represents a qubit, and their connections are given by parameterized unitaries of the form in Eq. (4) acting on the input states. In addition, REF<sup>175</sup> introduced quantum convolutional neural networks (QCNNs). QCNNs have been used for error correction<sup>175</sup>, for image recognition<sup>176</sup> and to discriminate quantum states belonging to different topological phases<sup>175</sup>. Moreover, it has been shown that QCNNs and QNNs with tree tensor network architectures do not exhibit BPs<sup>177,178</sup> (which will be discussed later), potentially making them a generically trainable architecture for large-scale implementations.

### New frontiers

In this section, we discuss some exciting, recently proposed applications of VQAs. These applications highlight the fact that VQAs could be used to understand and exploit the mathematical and physical structure of quantum states, and quantum theory in general.

**Quantum foundations.** NISQ computers are likely to play an important part in understanding the foundations of quantum mechanics. In a sense, these devices offer experimental platforms to test foundational ideas ranging from quantum gravity to quantum Darwinism<sup>179</sup>. For example, the emergence of classicality in quantum systems will be soon be a computationally tractable field of study because of the increasing size of NISQ computers. Along these lines, REF<sup>180</sup> proposed the variational consistent histories (VCH) algorithm. Consistent histories is a formal approach to quantum mechanics that has proved useful in studying the quantum-to-classical transition and quantum cosmology. In this formalism, interference between different paths (histories) is quantified in the decoherence functional<sup>181</sup>. The exponential number of terms in this decoherence functional makes the formalism computationally expensive on classical devices. VCH provides a way to prepare a density matrix representation of the entire functional, allowing one to efficiently examine the consistency of a set of histories. The application of standard VQAs to foundational situations can also provide a framework for new insights. For example, REF<sup>182</sup> showed that an FUMC strategy (discussed above) cannot efficiently learn a scrambling unitary. This result provides insight into the black hole information paradox, as one would need to have a representation of a black hole's scrambling unitary to unscramble information from emitted Hawking radiation<sup>183</sup>.

**Quantum information theory.** Another field that is likely to see renewed interest owing to NISQ computers is quantum information theory<sup>184</sup>. For example, in REF<sup>161</sup>

it was remarked that the quantum autoencoder algorithm could potentially be used to learn encodings and achievable rates for quantum channel transmission. Another area of research is using NISQ computers to compute key quantities in quantum information theory, such as the von Neumann entropy or distinguishability measures such as the trace distance. Although it is known that these problems are hard for general quantum states<sup>185</sup>, REF<sup>186</sup> introduced a VQA to estimate the quantum fidelity between an arbitrary state  $\sigma$  and a low-rank state  $\rho$ . Moreover, in REF<sup>85</sup> a VQA was introduced to learn modular Hamiltonians, which provides an upper bound on the von Neumann entropy of a quantum state. Here, one attempts to variationally decorrelate a quantum state by minimizing the relative entropy to a product distribution, and hence this method is suited for states that can be easily decorrelated.

**Entanglement spectroscopy.** Characterizing entanglement is crucial for understanding condensed matter systems, and the entanglement spectrum has proved useful in studying topological order. Several VQAs have been introduced to extract the entanglement spectrum of a quantum state<sup>112,145,187</sup>. As the entanglement spectrum can be viewed as the principal components of a reduced density matrix, algorithms for PCA can be used for this purpose, including the VQAs discussed before. In addition, one can also use the variational algorithm for quantum singular value decomposition introduced in REF<sup>187</sup>. These algorithms could potentially characterize the entanglement (and, for example, topological order) in a ground state that was prepared by VQE, and hence different VQAs can be used together in a complementary manner.

**Quantum metrology.** Quantum metrology is a field in which one seeks the optimal set-up for probing a parameter of interest (such as a magnetic field) with minimal shot noise. In the absence of noise during the probing process, the analytical solution for the optimal probe state can be derived. However, when general physical noises are present, an analytical solution is hard to find. Variational-state quantum metrology variationally searches for the optimal probe state<sup>188–191</sup>. For state preparation, variational quantum circuits are used in REFS<sup>188,190,191</sup> whereas optical tweezer arrays are considered in REF<sup>189</sup>. More concretely, one prepares a probe state with variational parameters, probes the magnetic field with physical noises, measures quantum Fisher information (QFI) as a cost function and updates the parameters to maximize it. Note that since QFI cannot be efficiently computed, an approximation of QFI can be heuristically found by optimizing the measurement basis, or by computing upper and lower bounds on the QFI<sup>191</sup>.

### Challenges and potential solutions

Despite the tremendous developments in the field of VQAs, there are still many challenges that need to be addressed to maintain the hope of achieving quantum speed-ups when scaling up these near-term architectures. Understanding the limitations of VQAs is crucial to developing strategies that can be used to construct

#### Quantum Fisher information

The quantum Fisher information quantifies the sensitivity of a quantum state to a parameter or a set of parameters.



and chooses the values of the remaining parameters so that the circuit is a sequence of shallow unitaries that evaluates to the identity. The main idea behind this method is to reduce the randomness and depth of the circuit to break the assumption that the circuit approximates a 2-design, a condition necessary for BPs to arise in deep ansatzes. Similar to the previous method, other schemes have been introduced to prevent BP by restricting the randomization of the ansatz. For instance, the proposal in REF.<sup>204</sup> showed that correlating the parameters in the ansatz effectively reduces the dimension of the hyperparameter space and can lead to large cost-function gradients. In addition, REF.<sup>205</sup> introduced a method whereby one uses layer-by-layer training: one initially trains shallow circuits and progressively adds components to the circuit. Whereas the latter guarantees that the number of parameters and randomness remain small for the first steps of the training, it has been shown<sup>206</sup> that this method can lead to an abrupt transition in the ability of quantum circuits to be trained. Finally, a method was introduced in REF.<sup>207</sup> where one pre-trains the parameters in the quantum circuits by using classical neural networks.

- **Ansatz strategies.** Another strategy for preventing BPs is to use structured ansatzes that are problem-inspired. The goal here is to restrict the space explored by the ansatz during the optimization. As discussed in the section on ansatzes, the UCC ansatzes for VQE of the quantum alternating operator ansatz<sup>36,37</sup> for optimization are problem-inspired ansatzes that are usually trainable even when randomly initialized. Other ansatz strategies include the proposals in REF.<sup>207</sup> for learning a mixed state, where one leverages knowledge of the target Hamiltonian to create a Hamiltonian variational ansatz. In addition, REFS<sup>73,74</sup> presented an approach in which the ansatz for the solution is  $|\psi(\{c_\mu\})\rangle = \sum_\mu c_\mu |\psi_\mu\rangle$ , for a fixed set of states  $\{|\psi_\mu\rangle\}$  determined by the problem at hand. Here, the optimization over the coefficients  $\{c_\mu\}$  can be solved using a quadratically constrained quadratic programme.

Finally, we remark that along with ansatz strategies there are other ways of potentially addressing BPs. These include optimizers tailored to mitigate the effect of BPs<sup>208</sup>, local cost functions<sup>112</sup> or architectures such as the QCNN, which has been shown to avoid BPs<sup>177</sup>.

### Efficiency

Another requirement that must be met for VQAs to provide a quantum advantage is having an efficient way to estimate expectation values (and more general cost functions). The existence of BPs can exponentially increase the precision requirements needed for the optimization portion of VQAs, as discussed in the subsection ‘Barren plateaus’, but even in the absence of such plateaus, these expectation value estimations are not guaranteed to be efficient. Indeed, early estimations of resource requirements suggested that the number of measurements that would be required for interesting quantum chemistry VQE problems would be astronomical, hence,

addressing this issue is essential for realizing quantum advantage<sup>41</sup>. More reasonable resource estimates can be reached for restricted problems such as the Hubbard model<sup>209,210</sup>. Although in principle one could always take projective measurements onto the eigenbasis of the operator in question, in general both the computational complexity of finding the required unitary and the depth required to implement that transformation may be intractable. However, given that arbitrary Pauli operators are diagonalizable with one layer of single-qubit rotations, it is common for the operators of interest (such as quantum chemistry Hamiltonians) to be expressed by their decomposition into such Pauli operators. That is,  $H = \sum_i c_i \sigma_i$ , where  $\{c_i\}$  are real coefficients and  $\{\sigma_i\}$  are Pauli operators. The drawback of this approach is that, for many interesting Hamiltonians, this decomposition contains many terms. For example, for chemical Hamiltonians, the number of distinct Pauli strings scales as  $n^4$  where  $n$  is the number of orbitals (and thus qubits) for large molecules. In what follows, we discuss several methods whose goal is to obtain measurement frugality in estimating the cost function.

**Commuting sets of operators.** In the interest of reducing the number of measurements required to estimate an operator expectation value, various methods have been proposed for partitioning sets of Pauli strings into commuting (simultaneously measurable) subsets. The choice of the subsets is also, of course, non-unique and has been mapped onto the combinatorial problems of graph colouring<sup>211,212</sup>, finding the minimum clique cover<sup>213–216</sup> or finding the maximal flow in network flow graphs<sup>217</sup>, which makes it possible to import the heuristics and formal results from those problems.

Perhaps the simplest approach to such a partitioning is to look for subsets that are qubit-wise commuting (QWC), which is to say that the Pauli operators on each qubit commute. Indeed, this was the initial method introduced<sup>218</sup>. However, whereas the QWC methods help to reduce the number of operators, they do not change the asymptotic scaling for quantum chemistry applications, motivating more general commutative groupings to be considered. To this end, it has been shown that by considering general commutations (and increasing the number of gates of the circuit quadratically with  $n$ ), the scaling of the number of measurements can be reduced to  $n^3$  (REFS<sup>211,212,214–217</sup>).

For using VQE on fermionic systems, this scaling can actually be brought down to either quadratic or, for simpler cases, even linear<sup>219</sup> in  $n$ . This improvement is found by considering factorizations of the two-electron integral tensors, rather than working at the operator level. The success of this approach suggests that using background information on the problem may greatly improve the measurement efficiency of estimating an expectation value.

**Optimized sampling.** In addition to reducing the number of individual operators that need to be measured, measurement efficiency can also be improved by carefully allocating the number of shots among the Pauli operators. Since operators with smaller coefficients will tend to

contribute less to the overall variance, assigning the same number of shots to each operator is usually inefficient. Instead, the optimal approach<sup>220</sup> is to give each Pauli operator a number of shots proportional to  $|c_i| \sqrt{\text{Var}(\sigma_i)}$ , where  $c_i$  is the coefficient of the  $i$ th Pauli operator  $\sigma_i$  and  $\text{Var}(\sigma_i)$  is the variance of  $\langle \sigma_i \rangle$ . During an optimization in which low-precision steps may be allowed early on, this allocation can instead be performed randomly with probabilities proportional to  $|c_i| \sqrt{\text{Var}(\sigma_i)}$ . Making the allocation randomly in this way allows for unbiased estimates with as little as one shot, potentially markedly increasing the efficiency of the optimization<sup>221</sup>. Optimizing the sampling of the metric tensor has also been explored, with the conclusion that these costs need not be dominant in metric-aware VQAs<sup>222</sup>.

**Classical shadows.** Another promising approach to efficient measurements is the construction of classical shadows<sup>223</sup>, also known as shadow tomography. In this approach, an approximate classical representation of the state (the classical shadow) is constructed by summing over the collection of states onto which a sequence of different measurements projects. These measurements are taken on the basis of randomly chosen strings so that a partial tomography of the state is completed. By combining the measurements in this way, each shot contributes to the estimation of each Pauli operator expectation value, resulting in a number of measurements that scale logarithmically with  $n$ . As with direct measurement approaches discussed above, this approach can be further optimized by tuning the probability distribution for the Pauli operators that define the measurements to match the properties of the operator and state<sup>224</sup>.

**Neural network tomography.** A different approach using partial tomography is to train an approximate RBM representation of the desired quantum state<sup>225</sup>. This RBM is fitted using measurements of the Pauli operators that are needed to directly estimate a given operator's expectation value, and so does not inherently reduce the number of operators to measure. However, by computing the expectation value on an approximate RBM instead of directly from measurements, the sampling variance for a given number of shots is substantially reduced at the cost of introducing a small, positive bias<sup>225</sup>.

#### Accuracy

One of the main goals for VQAs is to enable a practical use for NISQ devices. For this goal, VQAs provide a strategy to deal with hardware noise, as they can potentially minimize quantum circuit depth. Moreover, error mitigation methods<sup>24,99,226–243</sup> can be combined with VQAs to further improve accuracy (see Supplementary information). However, one can still ask what the impact of hardware noise will be on the accuracy of a VQA.

**Impact of hardware noise.** There are multiple aspects of the impact of hardware noise: it could potentially slow the training process, it could bias the landscape so that the noisy global optimum no longer corresponds to the noise-free global optimum, and it could affect the final value of the optimal cost.

- **Effect of noise on training.** The question of whether noise can help with the training process was posed in REF.<sup>244</sup>. In practice, it is typical to observe that noise slows down the training. For example, it was heuristically observed that the noise-free cost achieves lower values with noise-free training than with noisy training<sup>11,221,245</sup>. As discussed in the section on BPs, the intuition behind this slowing down is that the cost landscape is flattened, and hence gradient magnitudes are reduced, by the presence of incoherent noise<sup>200,246,247</sup>. Moreover, gradients decay exponentially with the algorithm's depth, meaning that the deeper the circuit, the more it will be affected. This can be further understood from the fact that cost functions are typically extremized by pure states, and since incoherent noise reduces state purity, one expects this noise to erode the extremal points of the landscape<sup>201</sup>. The presence of noise-induced BPs and their effect on the trainability is one of the leading challenges for VQAs, with potential solutions being to develop better quantum hardware or shorter-depth algorithms. It is worth remarking that the results discussed here do not account for the use of error mitigation techniques (see Supplementary information), and the degree to which these could help is still an open question.

- **Effect of noise on cost evaluation.** In REFS<sup>200,201</sup> it was also shown that in the presence of local Pauli noise, the cost landscape concentrated exponentially with the depth of the ansatz around the value of the cost associated with the maximally mixed state. Whereas the proof of this exponential concentration of the cost was for general VQAs, some previous work had also observed this effect for the special case of the QAOA<sup>246,247</sup>. The exponential concentration of the cost is, of course, important beyond the issue of trainability. Even if one is able to train, the final cost value will be corrupted by noise. There are certain VQAs for which this is not an important issue (for example, in QAOA where one can classically compute the cost after sampling). However, for VQE problems, this is important, since one is ultimately interested in an accurate estimation of the energy. This emphasizes the importance of understanding to what degree error mitigation methods (see Supplementary information) can correct for this issue.

**Noise resilience.** One reason for the interest in VQAs is their ability to naturally overcome certain types of noise in hardware, especially in near-term implementations. This noise resilience is a crucial, non-trivial feature of VQAs.

- **Inherent resilience to coherent noise.** By construction, VQAs are insensitive to the specific parameter values, ultimately only sampling physical observables from the resulting state. More specifically, if the physical implementation of a unitary results in a coherent error within the parameter space, or  $U(\theta)$  actually results in  $U(\theta + \delta)$ , then under mild assumptions the optimizer can calibrate this block unitary on the fly to improve the physical state produced. This effect was first conjectured theoretically<sup>218</sup> and later seen









supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) grant nos. JPMXS0120319794, JPMXS0118068682 and JST ERATO grant no. JPMJER1601. K.F. was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grant no. 16H02211, JST ERATO JPMJER1601 and JST CREST JPMJCR1673. K.M. was supported by JST PRESTO grant no. JPMJPR2019 and JSPS KAKENHI grant no. 20K22330. K.M. and K.F. were also supported by MEXT QLEAP grant no. JPMXS0118067394 and JPMXS0120319794. X.Y. acknowledges support from the Simons Foundation. L.C. was initially supported by the LDRD programme of LANL under project no. 20190065DR, and later supported by the US DOE, Office of Science, Office of Advanced Scientific Computing Research under the Quantum Computing Application Teams

(QCAT) programme. P.J.C. was initially supported by the LANL ASC Beyond Moore's Law project, and later supported by the US DOE, Office of Science, Office of Advanced Scientific Computing Research, under the Accelerated Research in Quantum Computing (ARQC) programme. Most recently, M.C., L.C. and P.J.C. were supported by the Quantum Science Center (QSC), a National Quantum Information Science Research Center of the US DOE.

#### Author contributions

All authors have read, discussed and contributed to the writing of the manuscript.

#### Competing interests

The authors declare no competing interests.

#### Peer review information

*Nature Reviews Physics* thanks Lei Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1038/s42254-021-00348-9>.

© Springer Nature Limited 2021