

PERSPECTIVE

Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers

To cite this article: Alejandro Perdomo-Ortiz *et al* 2018 *Quantum Sci. Technol.* **3** 030502

View the [article online](#) for updates and enhancements.

You may also like

- [Classical versus quantum models in machine learning: insights from a finance application](#)
Javier Alcazar, Vicente Leyton-Ortega and Alejandro Perdomo-Ortiz
- [Quantum-assisted Helmholtz machines: A quantum–classical deep learning framework for industrial datasets in near-term devices](#)
Marcello Benedetti, John Realpe-Gómez and Alejandro Perdomo-Ortiz
- [Near-term quantum algorithms for linear systems of equations with regression loss functions](#)
Hsin-Yuan Huang, Kishor Bharti and Patrick Rebentrost



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Quantum Science and Technology



PERSPECTIVE

Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers

PUBLISHED
19 June 2018

Alejandro Perdomo-Ortiz^{1,2,3,4,5} , Marcello Benedetti^{1,2,4,5}, John Realpe-Gómez^{1,6,7} and Rupak Biswas^{1,8}

¹ Quantum Artificial Intelligence Lab., NASA Ames Research Center, Moffett Field, CA 94035, United States of America

² USRA Research Institute for Advanced Computer Science, Mountain View, CA 94043, United States of America

³ Qubiterra, LLC., Mountain View, CA 94041, United States of America

⁴ Department of Computer Science, University College London, WC1E 6BT London, United Kingdom

⁵ Cambridge Quantum Computing Limited, CB2 1UB Cambridge, United Kingdom

⁶ SGT Inc., Greenbelt, MD 20770, United States of America

⁷ Instituto de Matemáticas Aplicadas, Universidad de Cartagena, Bolívar 130001, Colombia

⁸ Exploration Technology Directorate, NASA Ames Research Center, Moffett Field, CA 94035, United States of America

E-mail: alejandro.perdomoortiz@nasa.gov

Keywords: quantum-assisted machine learning, quantum machine learning, near-term quantum computers, hybrid algorithms, unsupervised learning, quantum annealing, unsupervised generative models

Abstract

With quantum computing technologies nearing the era of commercialization and quantum supremacy, machine learning (ML) appears as one of the promising ‘killer’ applications. Despite significant effort, there has been a disconnect between most quantum ML proposals, the needs of ML practitioners, and the capabilities of near-term quantum devices to demonstrate quantum enhancement in the near future. In this contribution to the focus collection ‘What would you do with 1000 qubits?’, we provide concrete examples of intractable ML tasks that could be enhanced with near-term devices. We argue that to reach this target, the focus should be on areas where ML researchers are struggling, such as generative models in unsupervised and semi-supervised learning, instead of the popular and more tractable supervised learning techniques. We also highlight the case of classical datasets with potential quantum-like statistical correlations where quantum models could be more suitable. We focus on hybrid quantum–classical approaches and illustrate some of the key challenges we foresee for near-term implementations. Finally, we introduce the quantum-assisted Helmholtz machine (QAHM), an attempt to use near-term quantum devices to tackle high-dimensional datasets of continuous variables. Instead of using quantum computers to assist deep learning, as previous approaches do, the QAHM uses deep learning to extract a low-dimensional binary representation of data, suitable for relatively small quantum processors which can assist the training of an unsupervised generative model. Although we illustrate this concept on a quantum annealer, other quantum platforms could benefit as well from this hybrid quantum–classical framework.

1. Introduction

With quantum computing technologies nearing the era of commercialization and of quantum supremacy [1], it is important to think of potential applications that might benefit from these devices. Machine learning (ML) stands out as a powerful statistical framework to attack problems where exact algorithms are hard to develop. Examples of such problems include image and speech recognition [2, 3], autonomous systems [4], medical applications [5], biology [6], artificial intelligence [7], and many others. The development of quantum algorithms that can assist or entirely replace the classical ML routine is an ongoing effort that has attracted a lot of interest in the scientific quantum information community [8–40]. We restrict the scope of our perspective to this specific angle and refer to it hereafter as quantum-assisted machine learning (QAML).

Research in this field has been focusing on tasks such as classification [14], regression [11, 15, 18], Gaussian models [16], vector quantization [13], principal component analysis [17], and other strategies that are routinely used by ML practitioners nowadays. We do not think these approaches would be of practical use in near-term quantum computers. The same reasons that make these techniques so popular, e.g., their scalability and algorithmic efficiency in tackling huge datasets, make them less appealing to become top candidates as killer applications in QAML with devices in the range of 100–1000 qubits. In other words, regardless of the claims about polynomial and even exponential algorithmic speed-up, reaching interesting industrial-scale applications would require millions or even billions of qubits. Such an advantage is then moot when dealing with real-world datasets and with the quantum devices to become available in the next years in the few thousands-of-qubits regime. As we elaborate in this paper, only a game changer such as the new developments in hybrid classical-quantum algorithms might be able to make a dent in speeding up ML tasks.

In our perspective here, we propose and emphasize three approaches to maximize the possibility of finding killer applications on near-term quantum computers:

- (i) Focus on problems that are currently hard and intractable for the ML community, for example, generative models in unsupervised and semi-supervised learning as described in section 2.
- (ii) Focus on datasets with potentially intrinsic quantum-like correlations, making quantum computers indispensable; these will provide the most compact and efficient model representation, with the potential of a significant quantum advantage even at the level of 50–100 qubit devices. In section 2.2 we suggest the case of the cognitive sciences, as a research domain potentially yielding such datasets.
- (iii) Focus on hybrid algorithms where a quantum routine is executed in the intractable step of the classical ML algorithmic pipeline, as described in section 3 and section 4.

Each one of these tasks has its own challenges and significant work needs to be done towards having experimental implementations on available quantum hardware (see, for example, [20, 21]). Based on our past experience in implementing QAML algorithms on existing quantum hardware, we provide here some insights into the main challenges and opportunities in this steadily growing research field. Along with illustrations from our earlier work and demonstrations on quantum annealers, we attempt to provide general insights and challenges applicable to other quantum computational paradigms such as the gate model.

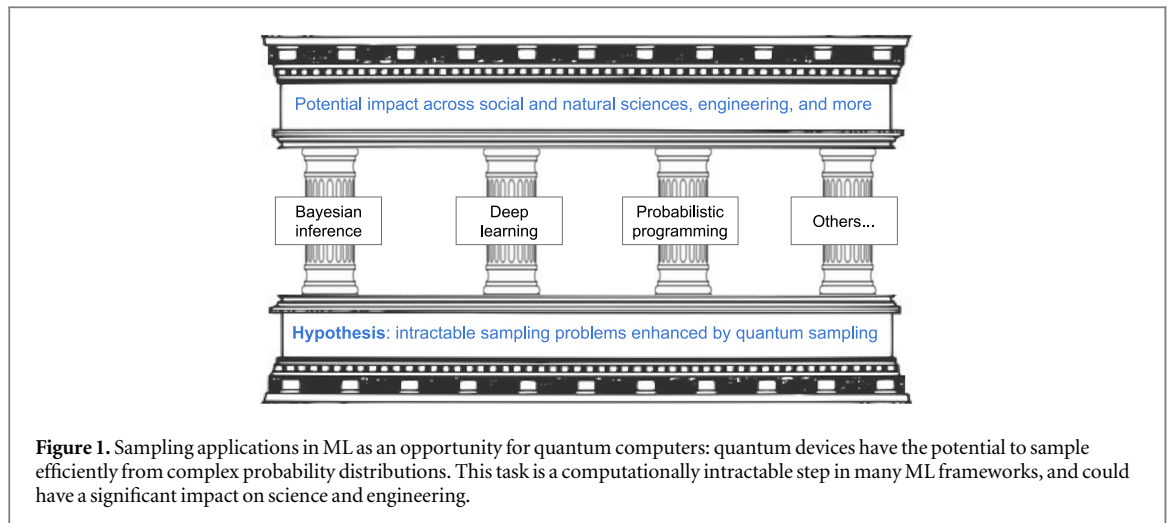
In section 2 we present examples of domains in ML that we believe offer viable opportunities for near-term quantum computers. In section 3 we present and illustrate the challenges ahead of such implementations and, whenever possible, with demonstrations in real hardware. In section 4 we introduce a new and flexible approach, the quantum-assisted Helmholtz machine (QAHM) [38], which has the potential to solve many of the challenges towards a near-term implementation of QAML for real-world industrial-scale datasets. In section 5 we summarize our work.

2. Opportunities in QAML

2.1. Quantum devices for sampling applications

The majority of data being collected daily is unlabeled. Examples of unlabeled data are photos and videos uploaded to the Internet, medical imaging, tweets, audio recordings, financial time series, and sensor data in general. Labeling is the process of data augmentation with task-specific informative tags. But this task is often expensive as it requires humans experts. It is therefore important to design models and algorithms capable of extracting information and structures from unlabeled data; this is the focus of unsupervised ML. But why is this important at all? The discovery of patterns is one of the central aspects of science; scientists do not always know *a priori* which patterns they should look for and they need unsupervised tools to extract salient spatio-temporal features from unlabeled data. In general, unsupervised techniques can learn useful representations of high-dimensional data, that have desirable properties such as simplicity and sparsity [41]. When used in conjunction with supervised techniques such as regressors and classifiers, unsupervised tools can substantially reduce the amount of labeled data required to achieve a desired generalization performance [42]. Connections to reinforcement learning and more specific applications are pointed out in [43] and references therein.

An unsupervised approach that learns the joint probability of all the variables involved in a problem is often called a *generative model*. The name comes from the possibility of inferring any marginal and conditional distribution, which in turn provide a way to generate new data resembling the training set. By following the taxonomy introduced in [43], we distinguish generative models as either explicit or implicit density estimators. *Explicit density estimators* are a large family of models which include Boltzmann machines, belief networks, and autoencoders. Depending on the characteristics of the model, they can be learned by variational approximations



[44–48], Monte Carlo sampling [49–52], or a combination of both [53]. *Implicit density estimators* achieve the generative task indirectly, for example, by casting the problem into a classification task [54] or finding the Nash equilibrium of a game [55].

Interestingly, generative models with many layers of unobserved stochastic variables (also called hidden variables) have the ability to learn multi-modal distributions over high-dimensional data [56]. Each additional layer provides an increasingly abstract representation of the data and improves the generalization capability of the model [50]. These models are usually represented as graphs of stochastic nodes where the edges may be directed, undirected, or both. Unfortunately, exact inference is intractable in all but the most trivial topologies [57]. In practice, learning the parameters of the model is even more demanding as it requires one to carry out the intractable inference step for each observed data point and for each learning iteration. The computational bottleneck comes from computing expectation values of several quantities under the complex distribution implemented by the model. Classically, this is approached by Markov chain Monte Carlo (MCMC) techniques that, unfortunately, are known to suffer from the slow-mixing problem [58]. It is indeed difficult for the Markov chain to jump from one mode of the distribution to another when these are separated by low-density regions of relevant size.

The capability of quantum computers to efficiently prepare and sample certain probability distributions has attracted the interest of the ML community [10, 59]. In this work, we focus on the use of quantum devices to sample classical [60] or quantum [19, 61] Gibbs distributions as an alternative to MCMC methods. This approach to sampling holds promise for more efficient inference and training in classical and quantum generative models, e.g., classical and quantum Boltzmann machines [25]. Once again though, the small number of qubits and the limitations of currently available hardware may impair the sampling process, making it useless for real ML applications. In this perspective, we argue that even noisy distributions could be used for generative modeling of real-life datasets. This requires working in settings where the operations implemented in hardware are only partially known. We call this scenario a *gray-box*. We also argue that hybrid classical-quantum architectures are suitable for near-term applications where the classical part is used to bypass some of the limitations of the quantum hardware. We call this approach *quantum-assisted*.

As highlighted by ML experts [62], it is expected that unsupervised learning will become far more important than purely supervised learning in the long term. More specifically, most earlier work in generative models in deep learning relied on the computationally costly step of MCMC, making it hard to scale to large datasets. We believe this represents an opportunity for quantum computers, and has been the motivation of our previous work [20, 21] and of the new developments presented in section 4.

We discussed how inference and learning in graphical models can benefit by more efficient sampling techniques. As illustrated in figure 1, sampling is at the core of other leading-edge domains as well, such as probabilistic programming (see [63] for an example of application). If quantum computers can be shown to significantly outperform classical sampling techniques, we expect to see a strong impact across science and engineering.

2.2. Datasets with native quantum-like correlations

Recently, Google demonstrated that quantum computers with as few as 50 qubits can attain quantum supremacy, although in a task with no obvious applications [64]. A highly relevant question then is: which type of real-life applications could benefit from quantum supremacy with near-term small devices? One of the main

motivations underlying the research efforts described here is that quantum computers could speed up ML algorithms. However, they could improve other aspects of ML and artificial intelligence (AI) as well. Recent work shows that quantum mechanics can provide more parsimonious models of stochastic processes than classical models [65–67], as quantified by an entropic measure of complexity. This suggests that quantum models hold the potential to substantially reduce the amount of other type of computational resources, e.g., memory [66, 67], required to model a given dataset.

This invites us to pose the following challenge: identify real-life datasets, unrelated to quantum physics, where quantum models are substantially simpler than classical models, as quantified by standard model comparison techniques such as the Akaike information criterion [68]. In a sense, this is a form of quantum supremacy. While datasets generated in experiments with quantum systems would be an obvious fit, the challenge is to find such datasets elsewhere. Cognitive sciences may offer some potential candidates, as we will describe below.

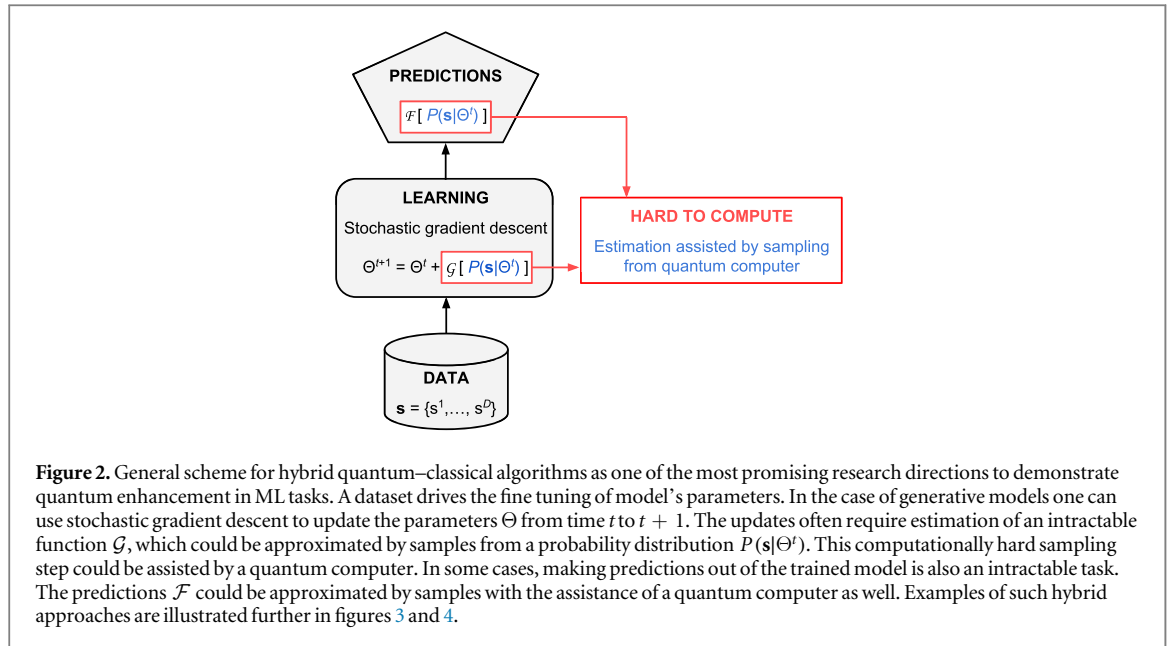
To avoid potential misunderstandings, let us consider first the example of statistical mechanics, which was developed in the 19th century to describe physical systems with many particles. Although statistical mechanics was long thought to be specific to physics only, we know today that certain aspects of it can be derived from very general information-theoretical principles. For instance, the structure of the Boltzmann distribution can be derived from the maximum entropy principle of information theory [69]. The tools of statistical mechanics have become valuable to study phenomena as complex as human behavior [70, 71], develop record-performance algorithms [72, 73], among many other interdisciplinary applications. Indeed, statistical mechanics has played, and continue to play, a relevant role in the development of ML, the Boltzmann machine being an important example.

In a similar vein, decades of research in quantum foundations and quantum information have allowed the identification of certain features, like interference, entanglement, contextuality, among others, that are encoded naturally by quantum systems [74, 75]. Moreover, there is increasing evidence suggesting that quantum models could be a valuable mathematical tool to study certain puzzling phenomena in the cognitive sciences (e.g., see [76–78] and references therein).

Consider, for instance, the following experiment [79]: a participant is asked to play a game where she can either win \$200 or lose \$100 with equal probability, and afterwards she is given the choice of whether or not to play the same gamble again. The experimentalist decides whether or not to tell the participant the result of the first gamble, i.e., whether she won or lost. Results showed that although participants typically preferred to play the second gamble when they knew the outcome of the first gamble, regardless of whether they won or lost, they typically preferred not to play if they did not have such information. More specifically, the participants choose to play the second gamble (G) with probabilities $P(G|W) = 0.69$ or $P(G|L) = 0.59$, respectively, if they knew that they won (W) or lost (L) the first gamble, and with probability $P(G) = 0.39$ if they did not have such information. These results violate the law of total probability $P(G) = P(G|W)P(W) + P(G|L)P(L)$, regardless of the values of the marginal probabilities $P(W)$ and $P(L)$. By interpreting this as an interference phenomenon, the authors have managed to fit the experimental results using a quantum model substantially more parsimonious than alternative classical models, as quantified by standard Bayesian model comparison techniques.

Another general and unexpected prediction is concerned with the effects of the order of questions [80]. Asking a ‘yes–no’ question to a human subject can create a context that affects the answer to a second ‘yes–no’ question. So the probability of, say, answering ‘yes’ to both questions depends on which order the questions are asked. For instance, in a 1997 poll in USA, 501 respondents were first asked the question: ‘do you generally think Bill Clinton is honest and trustworthy?’ Afterwards, they were asked the same question about Al Gore. Other 501 respondents were asked the same two questions, but in reverse order. The number of respondents answering ‘yes’ to both questions significantly increased when Al Gore was judged first. A quantum model for this phenomenon is based on the assumption that the respondent’s initial belief regarding the idea of ‘honest and trustworthy’ can be encoded using a quantum state ρ . The two possible answers, ‘yes’ (Y) or ‘no’ (N), to the question about Clinton or Gore are represented by basis $\{|Y\rangle_C, |N\rangle_C\}$ or $\{|Y\rangle_G, |N\rangle_G\}$, respectively, with respect to which measurements are performed. If the projectors associated to Clinton and Gore do not commute, we have order effects. A general parameter-free equality can be derived from these quantum models for which experimental support has been found in about 70 surveys of about 1000 people each, and 2 experiments with 100 people [80]. Quantum computers may allow for a complementary experimental exploration of these ideas at larger scales, now that experiments with hundreds of people are becoming more common [80–82].

A report [83] from the White House in 2016 reads ‘it is unlikely that machines will exhibit broadly-applicable intelligence comparable to or exceeding that of humans in the next 20 years.’ Some of the most relevant reasons are technical. Recent work [75] suggests why quantum models may be useful in the cognitive sciences and, if so, it may offer new approaches to tackle some of the technical hurdles in AI. The 20-year span predicted above may



be in sync with the advent of more powerful and more portable quantum computing technologies. The potential returns may be high.

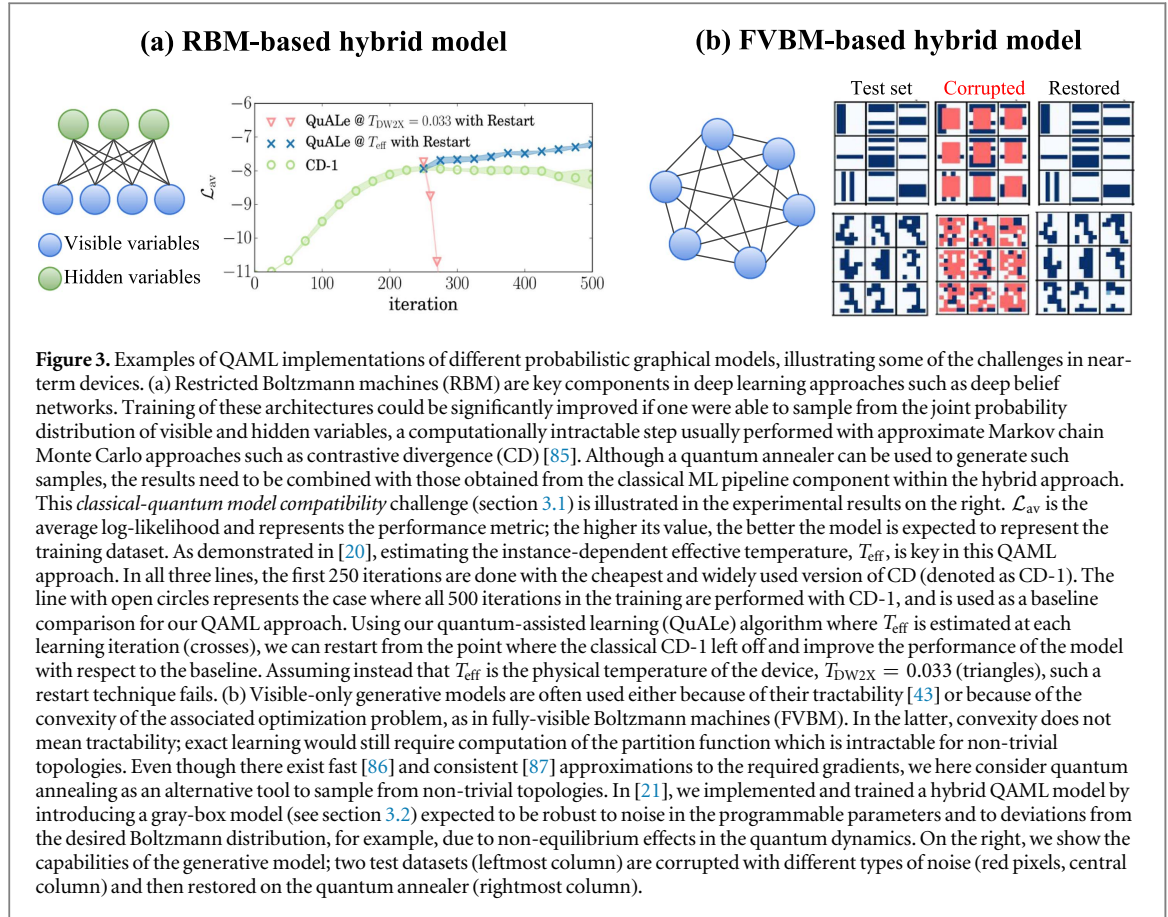
Although we emphasized the case of cognitive sciences, it would be interesting to explore what other relevant and commercial datasets exhibit quantum-like correlations, and where quantum computers can have an advantage even at the level of 50–100 qubits. In general, the identification of characteristics that are intrinsically quantum, and therefore hard to simulate classically, could be a game changer in the landscape of applications for near-term quantum technologies. Rather than trying to catch up with mature classical technologies in a competition for supremacy, this may offer the opportunity for quantum technologies to create their own unique niche of commercial applications, thereby becoming indispensable.

3. Challenges in QAML

Near-term implementation of QAML algorithms that can compete with state-of-the-art classical ML algorithms most likely will not come from the quantum version of popular ML algorithms (see, e.g., [11, 13–18]). As mentioned in [22], it would be difficult for these algorithms to preserve the speed-up claimed since they inherit the limitations of the Harrow-Hassidim-Lloyd (HHL) algorithm [84]. Here, we raise the bar even higher as our attention goes to the implementation of algorithms with potential quantum advantage in near-term quantum devices.

As pointed out in section 2, problem selection is key. For example, consider the recent work in quantum recommendation systems [27]. The authors developed a custom data structure along with a quantum algorithm for matrix sampling that has polylogarithmic complexity in the matrix dimensions. The result is a quantum recommendation system, and a proposal to circumvent most of the relevant limitations in the HHL algorithm. Because the input size is extremely large (e.g., number of users times number of products), the algorithm promises to significantly speed-up the task compared to currently employed ML approaches that require polynomial time. For the very same reasons, however, millions of qubits would be needed to handle datasets where state-of-the-art classical ML starts to struggle. We do not expect such devices to appear in the next decade. Instead, our attention goes to hybrid quantum–classical algorithms where conventional computers are used in the tractable subroutines of the algorithms and quantum computers assist only in the intractable steps. Figure 2 illustrates an example of this concept for the case of ML tasks.

There are several challenges which will generally impact any QAML algorithm, such as the limited qubit connectivity, the finite dynamic range of the parameters dictated by the intrinsic energy scale of the interactions in the device, and intrinsic noise in the device leading to decoherence in the qubits and uncertainty in the programmable parameters. We now emphasize some of these practical challenges, with particular attention to execution and implementation of hybrid QAML algorithms in near-term devices.



3.1. Issue of classical and quantum model compatibility

Essential to a hybrid approach is the need to have a flow of information between the classical preprocessing and the quantum experiments. The possibility of sharing information back and forth between the different architectures might pose a significant challenge, arising from the need to match samples from the classical and the quantum model. That is the case, for example, when assisting the training of restricted Boltzmann machines (RBMs) or deep belief networks (DBNs) with quantum devices [20, 23]. There, updates of the parameters are performed by a stochastic gradient descent algorithm that requires two key components. Due to the bipartite structure of RBMs, the first component, also known as the ‘positive phase,’ can be estimated very efficiently with conventional sampling techniques, while the second component, also known as the ‘negative phase,’ is, in general, intractable and can be assisted with quantum sampling. Since the two terms are subtracted in the same equation and originate from the same model, it is important to have control and to match all of the parameters defining both classical and quantum probability distributions.

A challenge here is related to the temperature of the Gibbs distribution of the model we are sampling from. For simulations in conventional computers, it is irrelevant to explicitly specify the temperature. Since it is a multiplicative factor, we could set it to 1 and ignore it altogether. In the case of an experimental physical device, such as a quantum annealer, we cannot neglect the temperature because (i) it is not under our control, and (ii) it is determined by many factors. These include not only the operational temperature of the device, but also the details of the quantum dynamics. The lack of knowledge of this parameter implies that the communication between classical and quantum components of our hybrid algorithm is broken. In previous work [20], we showed the significant difference in performance that can arise from tackling this challenge. As shown in figure 3(a), a proper estimation of the temperature also allows us to use restart techniques. For instance, the learning process can initially be carried out on a classical computer because the model parameters may be below the noise level and precision of the quantum device. Then, when desired, the quantum computer can be called to continue the process.

The challenge can also be addressed at hardware level by designing devices that can prepare a class of quantum Gibbs states at will. However, this strategy might open up other difficulties, some of these are detailed below.

3.2. Robustness to noise in programmable parameters of the quantum device

Preparing quantum Gibbs states with a quantum annealer or with a universal gate-model quantum computer is not a straightforward task, given the intrinsic noise in the programmable parameters. In the case of quantum annealers, freezing on quantum distributions or dynamical effects can lead to non-equilibrium distributions away from Gibbs [60, 88]. This is one of the challenges towards scaling the approaches in [20, 23] where the training of RBMs and DBNs requires reliable samples from a classical Boltzmann distribution. Furthermore, any quantum computing architecture, with intrinsic noise in the programmable parameters, can always lead to deviations from the desired target state. A proposed solution [60, 88] was to use samples obtained from a quantum device to seed classical Gibbs samplers (e.g., MCMC). While this is a promising approach, one of its drawbacks is that it forces the model to have a specific form, even if the quantum device has a genuinely different structure. If the quantum distribution prepared by the device is far from the one assumed, most of the burden would lie on the classical post-processing. It is not clear how much of the method's efficiency is given by the quality of the samples from the quantum devices and how much is achieved by the post-processing steps.

Interestingly, gray-box models such as the fully-visible Boltzmann machine (FVBM) studied in [21, 25, 60] and illustrated in figure 3 (b) may be a near-term solution to these issues. In a gray-box model, although we assume that samples come from a Gibbs-like distribution, we work directly at the level of the first and second moment statistics, without complete knowledge of the actual parameters that were implemented. Therefore, the emphasis is on the quality of the samples obtained from the quantum device, and their closeness to the data. This has the potential to increase the resilience against perturbations in the programmable parameters. In fact, as long as the estimated gradients for stochastic gradient descent have a positive projection on the direction of the true gradients, the model moves towards the optimum. In cases where this is not so, it might still be possible to design algorithms that mix model-based and model-free information in a suitable way. For instance, a proxy could check whether the estimated gradients actually project on the correct direction, if not, then the system can move in the opposite direction.

Gray-box models with hidden variables could exploit all of the available resources from the quantum device, while coping with its intrinsic noise and parameter misspecification. As an example, we used a gray-box approach to implement a quantum-assisted Helmholtz machine where the hidden variables are sampled from a D-Wave device; the framework is discussed in section 4 and described in detail in [38]. A caveat of the gray-box approach is that the final model is inevitably tailored for the quantum device used during the training. That is, any time we want to perform ML tasks of interest, such as reconstruction or the generation of new images as shown in figure 3, we will need to use the same quantum device.

3.3. The curse of limited connectivity

The basic principle behind this challenge is that physical interactions are local in nature. Although engineering advances can push the degree of qubit connectivity in quantum annealers or gate-model quantum computers, required qubit-to-qubit interactions not available in the device will cost an overhead in the computational resources. In the case of quantum annealers, a standard solution is to produce an embedding of the logical problem of interest into the physical layout, therefore increasing the number of required qubits. In the case of a gate-model quantum computer, the overhead comes in the number of swaps required to make distant qubits talk to each other [89]. In any architecture, this compilation requirement needs to be considered in the algorithmic design.

In the case of quantum annealers, achieving the topological connectivity of the desired model is only half of the challenge. Another significant challenge is the problem of parameter setting associated with the additional interactions present in the embedded model. In other words, how does one set the new parameters such that the embedded model accurately represents the intended model? There are no known optimal solutions, although heuristic strategies have been proposed [90–92]. However, in the type of ML applications we are considering, there is a way out. The main goal in the training phase of an ML algorithm is to find the optimal parameters that minimize a certain performance metric, suggesting that ML itself is a parameter setting procedure. This is precisely the demonstration in [21], where we show how to train models with arbitrary pairwise connectivity. In this case, the difficult task is not the embedding, which can be readily obtained by known heuristics, but rather the training of the whole device, implicitly solving the parameter setting problem.

To summarize, emphasis has been given to the embedding problem and to the mapping of the logical model of interest into physical hardware. An equally, or even more important challenge, is to determine how to set the parameters, including those associated to the embedding, such that the device samples from the desired distribution. This combined problem is what we call *the curse of limited connectivity*. In the case of gate-model quantum computer, [93] presents an illustrative experimental study, highlighting this challenge. It presents a comparison and analysis of the trade-off between connectivity and the quality of computation due to the

aforementioned overhead in computational resources. A similar trade-off would need to be taken into consideration in implementations of QAML algorithms in near-term gate-model quantum computers.

3.4. Representation of complex ML datasets into near-term devices

Quantum information does not have to be encoded into binary observables (qubits), it could also be encoded into continuous observables [94]. There has been work in quantum ML that follows the latter direction [95, 96]. However, most available quantum computers do work with qubits, nicely resembling the world of classical computation. Datasets commonly found in industrial applications have a large number of variables that are not binary. For instance, images may have millions of pixels, where each pixel is a 3D vector and each entry of the vector is a number specifying the intensity of color. We refer to these kinds of datasets as complex ML datasets. A naive binarization of the data will quickly consume the qubits of any device with 100–1000 qubits. Many QAML algorithms [11, 14, 18] rely on amplitude encoding instead, a technique where continuous data is stored in the amplitudes of a quantum state. This provides an exponentially efficient representation upon which one could perform linear algebra operations. Unfortunately, it is not clear how one could prepare arbitrary states of this kind on near-term quantum devices. Even reading out all of the amplitudes of an output vector might kill or significantly hamper any speed-up [22].

In this perspective, we argue that near-term QAML algorithms should rather aim at encoding continuous variables stochastically into abstract binary representations, a strategy we refer to as *semantic binarization*. In this approach, we use quantum states that can be prepared by near-term devices for sampling from unique quantum probability distributions that may capture correlations hard to model with conventional classical ML models. In the context of quantum annealers, such design may allow sampling from non-trivial graph topologies that are usually avoided in favor of restricted ones, for example, bipartite graphs are favored in classical neural networks for convenience.

One way to obtain such an abstract representation is to use hybrid approaches where visible variables \mathbf{v} are logically implemented by classical hardware, and hidden variables \mathbf{u} are physically implemented by quantum hardware. However, this idea comes with further challenges. First, the issue of model compatibility described above would apply. Second, sampling hidden variables \mathbf{u} from the posterior distribution $P(\mathbf{u}|\mathbf{v})$ may be highly problematic because the preparation of arbitrary quantum states is an open challenge. Finally, we might have to sample a binarization for each data point and that would be impractical for near-term quantum computers. For instance, the standard ML dataset of handwritten digits by the Modified National Institute of Standards and Technology (MNIST) is composed of 60 000 training points, hence we would need to program the quantum device at least 60 000 times.

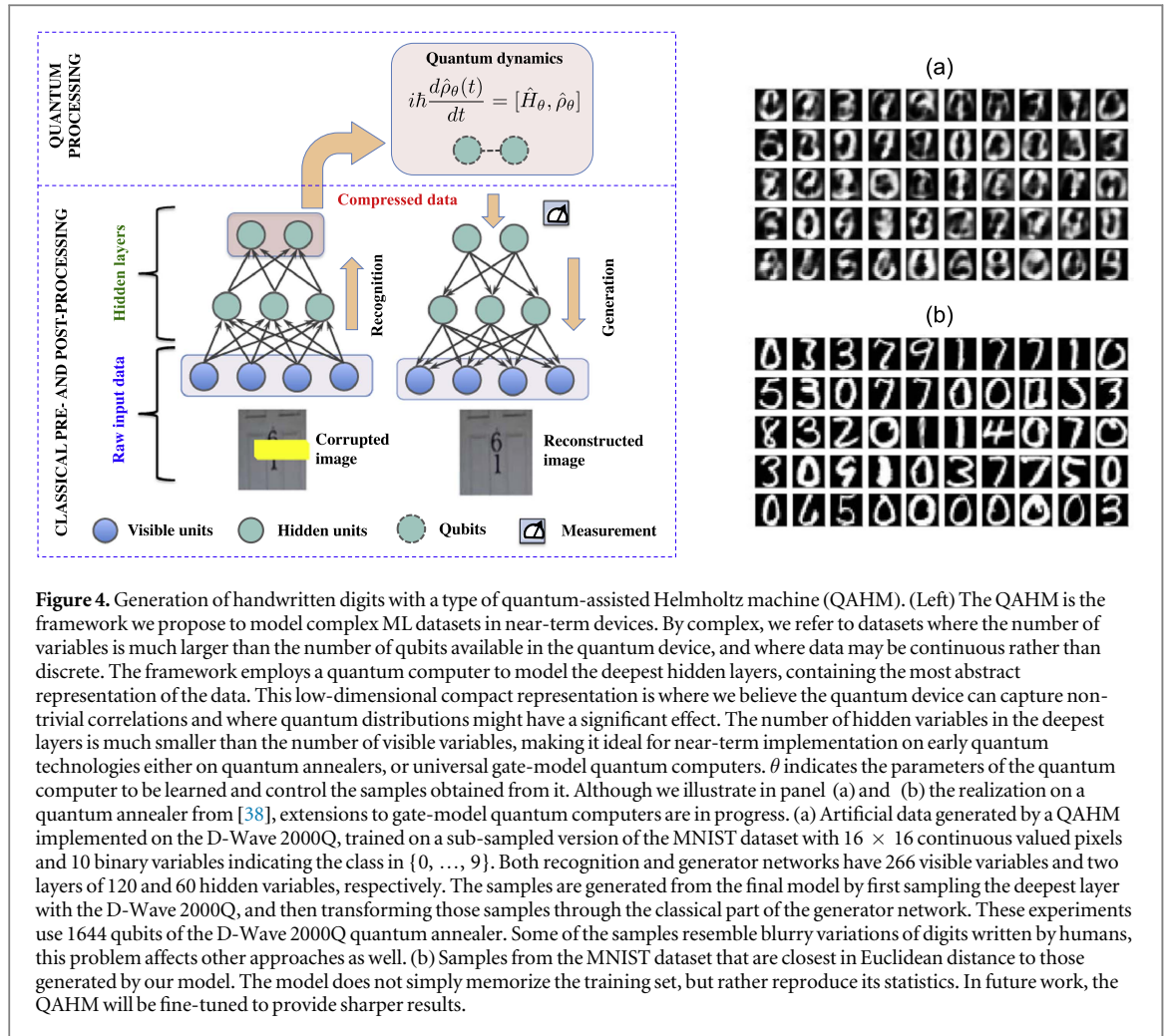
As we will see in the next section, deep learning may provide solutions to these challenges. We propose a new hybrid quantum–classical paradigm where the objective is to tackle most of the issues discussed here and, at the same time, to deal with complex ML datasets.

4. The quantum-assisted Helmholtz machine

The quantum-assisted Helmholtz machine (QAHM) is a framework for hybrid quantum–classical ML with the potential of coping with real-world datasets. We already pointed out some of the challenges in developing near-term QAML capable of competing with conventional ML pipelines, most importantly, the encoding of continuous variables, the limited number of variables, the need to prepare and measure quantum states for each data point. Here we show how some early ideas from the deep learning community can help avoid some of these difficulties. Details about the formalism and preliminary implementation of QAHM can be found in [38].

Consider a dataset and the task of modeling its empirical distribution with a generative model $P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{v}|\mathbf{u})P_{QC}(\mathbf{u})$ (see section 2.1 for a brief introduction). Here \mathbf{v} are the visible variables that represent the data and \mathbf{u} are unobservable or hidden variables that serve to capture non-trivial correlations. It is common to use binary valued stochastic hidden variables as they can express the presence or absence of features in the data. The set of hidden variables can be partitioned into a sequence of layers that encode increasingly abstract features. In other words, $P(\mathbf{v})$ is a deep neural network, called the *generator network*.

Here we suggest using a quantum device to model the most abstract representation of the data, that is, the deepest layers of the generator network. The samples obtained from a quantum device are described by the diagonal elements of a parameterized density matrix, $P_{QC}(\mathbf{u}) = \langle \mathbf{u} | \hat{\rho} | \mathbf{u} \rangle$. As an example, the density could be parameterized by a quantum Gibbs distribution $\hat{\rho} = e^{-\beta H} / \mathcal{Z}$, where H is the Hamiltonian implemented in quantum hardware and \mathcal{Z} is the corresponding partition function. The conditional distribution $P(\mathbf{v}|\mathbf{u})$ is then a classical neural network that transforms samples from the quantum device into samples with the same structure of those in the dataset. Hence, visible variables \mathbf{v} could be continuous variables, discrete variables, or other objects, effectively tackling the challenge of representing complex data (see section 3.4). Because the quantum



device works on a lower dimensional binary representation of the data, this model is also able to handle datasets whose dimensionality is much larger than the number of qubits available in hardware.

Typically, learning algorithms for generative models attempt to maximize the average likelihood of the data. As pointed out in section 2.1, this is not feasible in models with multiple layers of discrete hidden variables as we would need to sample from the intractable posterior distribution $P(\mathbf{u}|\mathbf{v})$. A Helmholtz machine [97, 98] consists of a generator network along with a *recognition network* $Q(\mathbf{u}|\mathbf{v})$ that learns to approximate $P(\mathbf{u}|\mathbf{v})$. This is a key approach behind many variational learning and importance sampling algorithms employed nowadays [44–48, 50–53]. In principle, the recognition network can also be implemented as a deep neural network whose hidden layers are modeled by a quantum device. However, this design requires quantum state preparation and measurement of the hidden variables for each data point in the training set and for each learning iteration, a daunting process in near-term implementations. To avoid doing this, we will implement the recognition network as a classical deep neural network. Restricting the recognition network to be classical is not a feature of the QAHM framework, instead it is an option to speed up learning of large datasets assisted by serial quantum devices (e.g., quantum annealers). To force the recognition network $Q(\mathbf{u}|\mathbf{v})$ to be close to the true posterior $P(\mathbf{u}|\mathbf{v})$ a notion of distance between them is minimized at each learning iteration. Using the Kullback-Leibler divergence, for instance, leads to the so-called wake-sleep algorithm [51, 97, 99].

The general architecture of a type of QAHM is illustrated in figure 4. The recognition network (left) infers hidden variables via a bottom-up pass starting from the raw data. The most abstract representation is obtained either from a classical layer (near-term) or from a quantum device (future implementations). The generator network generates samples of the visible variables via a top-down pass starting from samples obtained from a quantum device. The final model is an implicit density model when a gray-box approach is used to characterize the quantum hardware (see section 3.2), but we can turn it into an explicit density model if further processing is employed (e.g., quantum annealing to seed Gibbs samplers). Tasks such as reconstruction, generation, and classification can also be implemented in the QAHM framework.

In [38], we tested these ideas using a D-Wave 2000Q quantum annealer for the generation of artificial images. For this task we used a sub-sampled 16×16 pixels version of the standard handwritten digit dataset

MNIST. Each gray-scale pixel is characterized by an integer value in $\{0, \dots, 255\}$; we rescale this value in the range $[-1, +1]$ and interpret it as a continuous variable. There are also 10 binary variables indicating membership to one of the classes. The 266 visible variables needed to encode this data could, in principle, be embedded directly on the D-Wave 2000Q using a FVBM and a much poorer representation of the data via a naive binarization. Yet we would have to choose a relatively sparse model topology as we cannot embed an all-to-all connectivity in the D-Wave 2000Q for the 266 variables. The sparse connectivity and the absence of hidden variables can severely limit the ability to model the dataset. Our approach tackles both challenges and also enables the handling of larger datasets than would be possible in state-of-the-art quantum annealers.

We used a classical recognition network and a quantum-assisted generator network, both with 266 visible variables and two hidden layers of 120 and 60 variables, respectively. The deepest layer of 60 variables was mapped to 1644 qubits in D-Wave 2000Q using the approach in [21]. We run the wake-sleep algorithm for 1000 iterations and generated samples with the quantum-assisted generator network. The images generated are shown in figure 4(a). Although these preliminary results cannot compete with state-of-the-art ML, the artificial data often resemble digits written by humans. Figure 4(b) shows the images in the training set that are closest in Euclidean distance to the generated samples. We can see that the artificial images generated by the network are not merely copies of the training set, instead, they present variations and novelty in some cases, reflecting the generalization capability of the model. While the artificial data may also look blurry, this problem affects other approaches as well. Only the recent development of GANs [55] led to much sharper artificial images.

5. Summary

Machine learning has been presented as one of the applications with commercial value for near-term technologies. However, there seems to be a disconnect between the quantum algorithms proposed in much of the literature and the needs of the ML community. While most of the quantum algorithms for ML show that quantum computers have the potential of being very efficient at doing linear algebra (e.g., [11, 13–18, 84]), as discussed in [22], these proposals do not address the issues related to any near-term implementation. More importantly, to date there are no concrete benchmarks indicating that such work can be close to outperforming their conventional classical ML counterparts. In this perspective we stress this disconnect and provide our views on key aspects to consider towards building a robust readiness roadmap of QAML in near-term devices.

If a demonstration of quantum advantage on industrial applications is a first milestone to be pursued, we emphasize the need to move away from the popular and tractable ML implementations. We should rather look for applications that are highly desirable, but not-so-popular because of their intractability. It is in this domain where we believe quantum computers can have a significant impact in ML. In that sense, quantum speed-up in itself is not enough; if the chosen ML applications are tractable to a great accuracy with classical ML methods (as it is the case of [27]), then the number of qubits required to tackle industrial-scale applications may be far larger than those available in near-term devices.

As an example of intractable applications, in section 2 we presented the case of sampling from complex probability distributions with quantum devices. There is potential here for boosting the training of generative models in unsupervised or semi-supervised learning. Another approach we suggested is to explore applications where quantum distributions naturally fit the model describing the data correlations. That seems to be the case for some datasets from the field of cognitive sciences (see, e.g., [75, 79, 80, 100], also [76–78] and references therein). Other hard ML problems have been mentioned elsewhere [37]. We think working in any of these currently intractable applications will yield a higher payoff towards demonstrating that quantum models implemented in near-term devices might surpass models trained with classical resources.

Here we focused on some of the most pressing challenges we foresee in near-term implementations. For instance, the limited qubit connectivity will result in an overhead of qubits in the adiabatic model, and an overhead of gate operations in the gate-model. It is also important to take into consideration the model complexity that each physical hardware might present, which could have significant consequences on the ML task. For instance, applications to cognitive sciences might require a universal quantum computer capable of preparing and fine-tuning tailored quantum distributions, while applications to generative modeling might only need sampling from a quantum Gibbs distribution. For the latter, there already exist proposals with both quantum annealers and gate-model quantum computer architectures.

Coping with the challenges presented in section 3 is certainly an ongoing research activity. One key strategy proposed here towards the near-term demonstration of quantum advantage is the development of hybrid quantum–classical algorithms capable of exploiting the best of both worlds. In this perspective, we also put forward a new framework for such hybrid QAML algorithms, referred here as the *quantum-assisted Helmholtz machine* [38]. This new approach aims to solve some of the most pressing issues towards handling industrial-scale datasets with a large number of continuous variables. It is motivated by the idea that a quantum computer

should be used only to tackle the more abstract representation of the data, after trimming the information that can be handled classically. Here we use a deep neural network to transform the large continuous data into a new abstract discrete dataset with reduced dimensionality. It is to this abstract representation that we apply the quantum computer. This approach can be used to solve practical ML tasks, including reconstruction, classification, and generation of images.

Certainly, more work is needed to address the question of identifying the first killer ML application that can be implemented in near-term quantum computers with the order of a few thousand qubits. Finding intractable problems and developing new hybrid QAML algorithms that tackle the challenges of working with real-world devices is what we find the ideal scenario. We hope the community makes a leap in this direction now that more powerful and larger quantum annealers and gate-model quantum computers are becoming available to the scientific community.

Acknowledgments

The work of AP-O, J-R-G, and MB was supported in part by the AFRL Information Directorate under grant F4HBKC4162G001, the Office of the Director of National Intelligence (ODNI), the Intelligence Advanced Research Projects Activity (IARPA) via IAA 145483, and the US Army TARDEC under the ‘Quantum-assisted Machine Learning for Mobility Studies’ project. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, US Army TARDEC, or the US Government. The US Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon. MB was partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and by Cambridge Quantum Computing Limited (CQCL). The authors would like to thank Leonard Wossnig, Jonathan Romero, and Max Wilson for useful feedback on an early version of this manuscript.

ORCID iDs

Alejandro Perdomo-Ortiz  <https://orcid.org/0000-0001-7176-4719>

References

- [1] Mohseni M, Read P, Neven H, Boixo S, Denchev V, Babbush R, Fowler A, Smelyanskiy V and Martinis J 2017 Commercialize quantum technologies in five years *Nature* **543** 171–4
- [2] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *NIPS’12 Proc. of the 25th Int. Conf. on Neural Information Processing Systems* vol. 1 (Lake Tahoe, NV, Dec, 2012) (Curran Associates Inc.) 1097–105
- [3] Hinton G, Deng L, Yu D, Dahl G E, Mohamed A-R, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T N et al 2012 Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups *IEEE Signal Process. Mag.* **29** 82–97
- [4] Levinson J et al 2011 Towards fully autonomous driving: systems and algorithms *Intelligent Vehicles Symposium (IV) 2011 IEEE* (Piscataway, NJ: IEEE) 163–8
- [5] Esteva A, Kuprel B, Novoa R A, Ko J, Swetter S M, Blau H M and Thrun S 2017 Dermatologist-level classification of skin cancer with deep neural networks *Nature* **542** 115–8
- [6] Ching T et al 2017 Opportunities and obstacles for deep learning in biology and medicine *bioRxiv* (New York: Wiley)
- [7] Mnih V et al 2015 Human-level control through deep reinforcement learning *Nature* **518** 529–33
- [8] Neven H, Denchev V S, Drew-Brook M, Zhang J, Macready W G and Rose G 2009 Binary classification using hardware implementation of quantum annealing *Demonstrations at NIPS-09, 24th Annual Conference on Neural Information Processing Systems* 1–17
- [9] Bian Z, Chudak F, Macready W G and Rose G 2010 The Ising model: teaching an old problem new tricks *D-Wave Systems*
- [10] Denil M and Freitas N D 2011 Toward the implementation of a quantum RBM *NIPS Deep Learning and Unsupervised Feature Learning Workshop*
- [11] Wiebe N, Braun D and Lloyd S 2012 Quantum algorithm for data fitting *Phys. Rev. Lett.* **109** 050505
- [12] Pudenz K L and Lidar D A 2013 Quantum adiabatic machine learning *Quantum Information Processing* **12** 2027–70
- [13] Lloyd S, Mohseni M and Rebentrost P 2013 Quantum algorithms for supervised and unsupervised machine learning arXiv:1307.0411
- [14] Rebentrost P, Mohseni M and Lloyd S 2014 Quantum support vector machine for big data classification *Phys. Rev. Lett.* **113** 130503
- [15] Wang G 2017 Quantum algorithm for linear regression *Phys. Rev. A* **96** 012335
- [16] Zhao Z, Fitzsimons J K and Fitzsimons J F 2015 Quantum assisted Gaussian process regression, ArXiv e-prints arXiv:1512.03929 [quant-ph]
- [17] Lloyd S, Mohseni M and Rebentrost P 2014 Quantum principal component analysis *Nat. Phys.* **10** 631–3
- [18] Schuld M, Sinayskiy I and Petruccione F 2016 Prediction by linear regression on a quantum computer *Phys. Rev. A* **94** 022342
- [19] Svore K M, Wiebe N and Kapoor A 2015 *Quantum deep learning* arXiv:1412.3489
- [20] Benedetti M, Realpe-Gómez J, Biswas A and Perdomo-Ortiz A 2016 Estimation of effective temperatures in quantum annealers for sampling applications: a case study with possible applications in deep learning *Phys. Rev. A* **94** 022308
- [21] Benedetti M, Realpe-Gómez J, Biswas R and Perdomo-Ortiz A 2017 Quantum-assisted learning of hardware-embedded probabilistic graphical models *Phys. Rev. X* **7** 041052

- [22] Aaronson S 2015 Read the fine print *Nat. Phys.* **11** 291–3 commentary
- [23] Adachi S H and Henderson M P 2015 *Application of quantum annealing to training of deep neural networks* arXiv:1510.06356
- [24] Chancellor N, Szoke S, Vinci W, Aeppli G and Warburton P A 2016 Maximum-entropy inference with a programmable annealer *Sci. Rep.* **6** 22318
- [25] Amin M H, Andriyash E, Rolfe J, Kulchitsky B and Melko R 2016 Quantum Boltzmann Machine arXiv:1601.02036
- [26] Kieferová M and Wiebe N 2017 Tomography and generative training with quantum boltzmann machines *Phys. Rev. A* **96** 062327
- [27] Kerenidis I and Prakash A 2016 *arXiv preprint* arXiv:1603.08675
- [28] Wittek P and Gogolin C 2017 Quantum enhanced inference in markov logic networks *Sci. Rep.* **7** 45672
- [29] Potok T E, Schuman C, Young S R, Patton R M, Spedalieri F, Liu J, Yao K-T, Rose G and Chakma G 2017 A study of complex deep learning networks on high performance, neuromorphic, and quantum computers *Proc. IEEE* arXiv:1703.05364 47–55
- [30] Schuld M, Sinayskiy I and Petruccione F 2015 An introduction to quantum machine learning *Contemp. Phys.* **56** 172–85
- [31] Romero J, Olson J P and Aspuru-Guzik A 2017 Quantum autoencoders for efficient compression of quantum data *Quantum Sci. Technol.* **2** 045001
- [32] Adcock J, Allen E, Day M, Frick S, Hinchliff J, Johnson M, Morley-Short S, Pallister S, Price A and Stanisic S 2015 Advances in quantum machine learning arXiv:1512.02900
- [33] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2016 *Quantum machine learning* arXiv:1611.09347
- [34] Alvarez-Rodriguez U, Lamata L, Escandell-Montero P, Martín-Guerrero J D and Solano E 2016 Supervised quantum learning without measurements *Sci. Rep.* **7** 13645
- [35] Lucas L 2017 Basic protocols in quantum reinforcement learning with superconducting circuits *Sci. Rep.* **7** 1609
- [36] Schuld M, Fingerhuth M and Petruccione F 2017 Implementing a distance-based classifier with a quantum interference circuit *Europhys. Lett.* **119** 60002
- [37] Ciliberto C, Herbster M, Davide Ialongo A, Pontil M, Rocchetto A, Severini S and Wossnig I 2017 Quantum machine learning: a classical perspective *Proc. R. Soc. A* **474** 2209
- [38] Benedetti M, Realpe-Gómez J and Perdomo-Ortiz A 2017b Quantum-assisted helmholtz machines: a quantum–classical deep learning framework for industrial datasets in near-term devices arXiv:1708.09784
- [39] Benedetti M, Garcia-Pintos D, Nam Y and Perdomo-Ortiz A 2018 *A generative modeling approach for benchmarking and training shallow quantum circuits* arXiv:1801.07686
- [40] Farhi E and Neven H 2018 *Classification with quantum neural networks on near term processors* arXiv:1802.06002
- [41] Bengio Y, Courville A and Vincent P 2013 Representation learning: a review and new perspectives *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1798–828
- [42] Erhan D, Bengio Y, Courville A, Manzagol P-A, Vincent P and Bengio S 2010 Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* **11** 625–60
- [43] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial networks *Advances in Neural Information Processing Systems 27 (NIPS 2014)* arXiv:1701.00160
- [44] Kingma D P and Welling M 2013 Auto-encoding variational bayes arXiv preprint arXiv:1312.6114
- [45] Rezende D J, Mohamed S and Wierstra D 2014 Stochastic backpropagation and approximate inference in deep generative models *Int. Conf. on Machine Learning* 1278–86 arXiv:1401.4082
- [46] Mnih A and Gregor K 2014 Neural variational inference and learning in belief networks *Int. Conf. on Machine Learning* arXiv:1402.0030
- [47] Sønderby C K, Raiko T, Maaløe L, Sønderby S K and Winther O 2016 Ladder variational autoencoders *Advances in Neural Information Processing Systems* 3738–46
- [48] Rolfe J T 2016 *Discrete variational autoencoders* arXiv preprint arXiv:1609.02200
- [49] Ackley D H, Hinton G E and Sejnowski T J 1985 A learning algorithm for boltzmann machines *Cogn. Sci.* **9** 147–69
- [50] Hinton G E, Osindero S and Teh Y W 2006 A fast learning algorithm for deep belief nets *Neural Comput.* **18** 1527–54
- [51] Bornschein J and Bengio Y 2014 *Reweighted wake-sleep* arXiv preprint arXiv:1406.2751
- [52] Bornschein J, Shabani S, Fischer A and Bengio Y 2016 Bidirectional helmholtz machines *International Conference on Machine Learning* 2511–9
- [53] Salakhutdinov R and Hinton G 2009 Deep boltzmann machines *Artificial Intelligence and Statistics* 448–55
- [54] Bengio Y, Laufer E, Alain G and Yosinski J 2014 Deep generative stochastic networks trainable by backprop *International Conference on Machine Learning* 226–34
- [55] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* 2672–80
- [56] Bengio Y et al 2009 Learning deep architectures for ai *Foundations and Trend in Machine Learning* **2** 1–127
- [57] Roth D et al 1996 On the hardness of approximate reasoning *Artif. Intell.* **82** 273–302
- [58] Bengio Y, Mesnil G, Dauphin Y and Rifai S 2013 Better mixing via deep representations *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* 552–60
- [59] Dumoulin V, Goodfellow I J, Courville A C and Bengio Y 2014 On the challenges of physical implementations of RBMs *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27-31, 2014, Québec City, Québec, Canada.* 1199–205
- [60] Korenkevych D, Xue Y, Bian Z, Chudak F, Macready W G, Rolfe J and Andriyash E 2016 Benchmarking quantum hardware for training of fully visible boltzmann machines arXiv:1611.04528
- [61] Chowdhury A N and Somma R D 2016 *Quantum algorithms for gibbs sampling and hitting-time estimation* arXiv preprint arXiv:1603.02940
- [62] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- [63] Lake B M, Salakhutdinov R and Tenenbaum J B 2015 Human-level concept learning through probabilistic program induction *Science* **350** 1332–8
- [64] Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Martinis J M and Neven H 2018 Characterizing quantum supremacy in near-term devices *Nat. Phys.* arXiv preprint arXiv:1608.00263
- [65] Gu M, Wiesner K, Rieper E and Vedral V 2012 Quantum mechanics can reduce the complexity of classical models *Nat. Commun.* **3** 762
- [66] Palsson M S, Gu M, Ho J, Wiseman H M and Pryde G J 2017 Experimentally modeling stochastic processes with less memory by the use of a quantum processor *Science Advances* **3** e1601302
- [67] Elliott T J and Gu M 2017 Occam’s vorpal quantum razor: memory reduction when simulating continuous-time stochastic processes with quantum devices arXiv preprint arXiv:1704.04231

- [68] Burnham K P and Anderson D R 2003 Model selection and multimodel inference: a practical information-theoretic approach *Springer Science & Business Media*
- [69] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620–30
- [70] Perc M, Jordan J J, Rand D G, Wang Z, Boccaletti S and Szolnoki A 2017 Statistical physics of human cooperation *Phys. Rep.* **687** 1–51
- [71] Realpe-Gómez J, Andrighetto G, Nardin G and Montoya J A 2018 *Phys. Rev. E* **97** 042321
- [72] Mézard M, Parisi G and Zecchina R 2002 Analytic and algorithmic solution of random satisfiability problems *Science* **297** 812–5
- [73] Mezard M and Montanari A 2009 *Information, Physics, and Computation* (New York, NY, USA: Oxford University Press, Inc.,)
- [74] D'Ariano G M, Chiribella G and Perinotti P 2017 *Quantum Theory from First Principles: An Informational Approach* (Cambridge: Cambridge University Press)
- [75] Realpe-Gómez J 2017 *Modeling observers as physical systems representing the world from within: quantum theory as a physical and self-referential theory of inference* arXiv preprint arXiv:1705.04307
- [76] Bruza P D, Wang Z and Busemeyer J R 2015 Quantum cognition: a new theoretical approach to psychology *Trends in cognitive sciences* **19** 383–93
- [77] Busemeyer J R and Bruza P D 2012 *Quantum models of cognition and decision* (Cambridge: Cambridge University Press)
- [78] Aerts D, Broekaert J, Gabora L and Sozzo S 2016 Quantum structures in cognitive and social science *Frontiers in Psychology* **7**
- [79] Busemeyer J R, Wang Z and Shiffrin R M 2015 Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency *Decision* **2** 1
- [80] Wang Z, Solloway T, Shiffrin R M and Busemeyer J R 2014 Context effects produced by question orders reveal quantum nature of human judgments *Proc. Natl Acad. Sci.* **111** 9431–6
- [81] Gracia-Lázaro C, Ferrer A, Ruiz G, Tarancón A, Cuesta J A, Sánchez A and Moreno Y 2012 Heterogeneous networks do not promote cooperation when humans play a prisoner dilemma *Proc. Natl Acad. Sci.* **109** 12922–6
- [82] Gutiérrez-Roig M, Gracia-Lázaro C, Perelló J, Moreno Y and Sánchez A 2014 Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence *Nat. Commun.* **5** 4362
- [83] White House 2016 Artificial intelligence, automation, and the economy *Executive Office of the President* <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>
- [84] Harrow A W, Hassidim A and Lloyd S 2009 Quantum algorithm for linear systems of equations *Phys. Rev. Lett.* **103** 150502
- [85] Hinton G E 2002 Training products of experts by minimizing contrastive divergence *Neural Comput.* **14** 1771–800
- [86] Carreira-Perpinan M A and Hinton G E 2005 On contrastive divergence learning *Aistats* **10** 33–40
- [87] Hyvärinen Aapo 2006 Consistency of pseudolikelihood estimation of fully visible boltzmann machines *Neural Comput.* **18** 2283–92
- [88] Raymond J, Yarkoni S and Andriyash E 2016 Global warming: Temperature estimation in annealers *Frontiers in ICT* **3** 23
- [89] Beals R, Brierley S, Gray O, Harrow A W, Kutin S, Linden N, Shepherd D and Stather M 2013 Efficient distributed quantum computing *Proc. R. Soc. A, Vol. 469* (The Royal Society) 20120686
- [90] Choi V 2011 Minor-embedding in adiabatic quantum computation: Ii. minor-universal graph design *Quantum Information Processing* **10** 343–53
- [91] Perdomo-Ortiz A, Fluegemann J, Biswas R and Smelyanskiy V N 2015 *A performance estimator for quantum annealers: gauge selection and parameter setting* arXiv:1503.01083
- [92] Pudenz K L 2016 Parameter setting for quantum annealers *Proc. IEEE* arXiv:1611.07552
- [93] Linke N M, Maslov D, Roetteler M, Debnath S, Figgatt C, Landsman K A, Wright K and Monroe C 2017 Experimental comparison of two quantum computing architectures *Proceedings of the National Academy of Sciences* 201618020
- [94] Lloyd S and Braunstein S L 1999 Quantum computation over continuous variables *Phys. Rev. Lett.* **82** 1784
- [95] Lau H-K, Pooser R, Siopsis G and Weedbrook C 2017 Quantum machine learning over infinite dimensions *Phys. Rev. Lett.* **118** 080501
- [96] Das S, Siopsis G and Weedbrook C 2017 Continuous-variable quantum Gaussian process regression and quantum singular value decomposition of non-sparse low rank matrices *Phys. Rev. A* **97** 022315
- [97] Hinton G E, Dayan P, Frey B J and Neal R M 1995 The wake-sleep algorithm for unsupervised neural networks *Science* **268** 1158
- [98] Dayan P, Hinton G E, Neal R M and Zemel R S 1995 The helmholtz machine *Neural Comput.* **7** 889–904
- [99] Dayan P and Hinton G E 1996 Varieties of helmholtz machine *Neural Netw.* **9** 1385–403
- [100] Kvam P D, Pleskac T J, Yu S and Busemeyer J R 2015 Interference effects of choice on confidence: quantum characteristics of evidence accumulation *Proc. Natl Acad. Sci.* **112** 10645–50