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

Alejandro Perdomo-Ortiz,[1, 2, 3, *] Marcello Benedetti,[1, 2, 3] John Realpe-Gómez,[1, 4, 5] and Rupak Biswas[1, 6]

[1]*Quantum Artificial Intelligence Lab., NASA Ames Research Center, Moffett Field, CA 94035, USA*
[2]*USRA Research Institute for Advanced Computer Science (RIACS), Mountain View, CA 94043, USA*
[3]*Department of Computer Science, University College London, WC1E 6BT London, UK*
[4]*SGT Inc., Greenbelt, MD 20770, USA*
[5]*Instituto de Matemáticas Aplicadas, Universidad de Cartagena, Bolívar 130001, Colombia*
[6]*Exploration Technology Directorate, NASA Ames Research Center, Moffett Field, CA 94035, USA*
(Dated: September 1, 2017)

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 machine learning 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 on "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 still struggling, such as generative models in unsupervised or semisupervised learning, instead of the popular and much more tractable ML 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-resolution datasets on 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.

I. INTRODUCTION

With quantum computing technologies nearing the era of commercialization and of quantum supremacy [1], it is a 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 deterministic 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 [8–36]. We restrict the scope of our perspective to this specific angle and refer to it hereafter as quantum-assisted machine learning (QAML).

Although the focus has been on tasks such as classification [14], regression [11, 15, 18], Gaussian models [16], vector quantization [13], and principal component analysis [17] corresponding to the most widely-used strategies by machine learning practitioners nowadays, we do not foresee these would be important for near-term uses of 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 speedup, to reach 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 routine 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, fully generative models, unsupervised and semi-supervised learning as described in Sec. II. Other potential applications have been highlighted in a recent review [34].

(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 Sec. II B we suggest the case of the cognitive sciences, as a potential research domain potentially yielding such datasets.

(iii) Focus on hybrid quantum algorithms that can be easily integrated in the intractable step of the ML algorithmic pipeline, as described in Sec. III and Sec. IV.

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 Refs. [20, 21]). Based on our past experience in implementing QAML algorithms on existing quantum hardware devices, we provide here some insights into the main challenges in developing such opportunities in this steadily growing research field. Along with illustrations from our earlier work and demonstration on quantum annealers, we attempt to provide general insights and challenges applicable to any other quantum computational paradigm such as the widely used gate-based model.

In Sec. II we present examples of domains in ML we believe offer viable opportunities for near-term quantum computers. In Sec. III we present and illustrate the challenges ahead of such implementations and, whenever possible, with demonstrations in real hardware. In Sec. IV we introduce a new and flexible approach, the Quantum-Assisted Hemholtz Machine (QAHM) [35], which has the potential to solve many of the challenges towards a near-term implementation of QAML for real-world industrial scale datasets. In Sec. V we summarize our work.

II. OPPORTUNITIES IN QAML

A. Quantum devices for sampling applications

The majority of data currently 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 machine learning. 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 sparsity and simplicity [37]. When these

techniques are used as preprocessing steps they can improve supervised techniques, such as regressors and classifiers, by substantially reducing the amount of labeled data required to achieve a desired generalization performance [38]. Connections to reinforcement learning and more specific applications are pointed out in Ref. [39] 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, hence, providing a way to generate new data resembling the training set. By following the taxonomy introduced in [39], we distinguish generative models as either *explicit density estimators* 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 [40–44], Monte Carlo sampling approximations [45–48], or a combination of both [49]. Even though there exist tractable explicit models, here we focus on the more interesting task of learning intractable ones. Implicit density estimators achieve the generative task indirectly; for example, by casting the problem into a classification task [50] or finding the Nash equilibrium of a game [51].

Interestingly, generative models with many layers of unobserved stochastic variables (also called hidden units) have the ability to learn multi-modal distributions over high-dimensional data [52]. Each additional layer provides an increasingly abstract representation of the data and improves the generalization capability of the model [46]. These models are usually represented as graphs of stochastic nodes where edges may be directed, undirected, or both. Unfortunately, exact inference, learning, and model selection are intractable in all but the most trivial topologies. For instance, inferring the posterior probabilities of the unobserved variables given an observed data vector is $\sharp P$ in directed graphs [53] such as belief networks. Finding the probability that the model assigns to a configuration of variables requires the computation of exponentially many terms in the partition function in undirected graphs such as Boltzmann machines. Learning the parameters of the generative model is even more demanding. The common approach is to minimize some divergence of the model's density from the empirical distribution of the data (e.g., the information theoretic Kullback-Leibler divergence). But this process requires inference of the unobserved variables for each observed data point and for each learning iteration. The computational bottleneck of all the examples above consists in 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 [54]. It is indeed difficult for the Markov chain to jump from one mode of
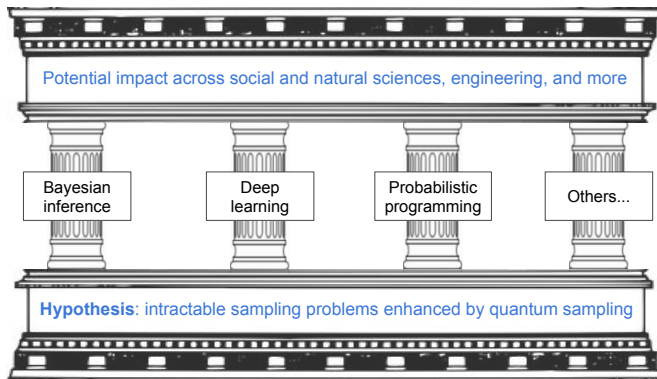
FIG. 1. *Sampling applications in ML as an opportunity for quantum computers:* Quantum devices have the potential to sample efficiently from complex probability distributions. This task is a computationally intractable step in many ML frameworks but with significant potential impact in many application domains.

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, 55]. In this work, we focus on the use of quantum devices to sample classical [56] or quantum [19, 57] 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 machine learning applications. In this perspective, we argue that even noisy distributions could be used for generative modeling of real-life datasets. This requires to work 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 [58], it is expected that unsupervised learning will become far more important than purely supervised learning in the longer 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 Sec. IV.

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 Ref. [59] 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.

B. 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 [60]. A highly relevant question then is which type of real-life applications could benefit from such quantum supremacy with the small devices that are near-term? 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 [61–63], 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 [62, 63], 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 [64]. 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 data sets elsewhere. Cognitive sciences may offer some potential candidates, as we will describe below.

First, let us observe that one of the reasons for the remarkable success of deep learning is that hierarchical architectures suitably encode the idea that complex objects (e.g., faces) can be built from the composition of other simpler objects (e.g., eyes). In general, the identification of characteristics that are intrinsically quantum, and therefore hard to simulate classically, could be a potential 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 in already standard commercial applications, this may offer the opportunity for quantum technologies to create their own unique niche of commercial applications, thereby becoming indispensable.

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 de-

rived from the maximum entropy principle of information theory [65]. The tools of statistical mechanics have become valuable to study phenomena as complex as human behavior [66, 67], develop record-performance algorithms [68, 69], 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 as well as a set of information-theoretic principles underlying relevant aspects of quantum mechanics [70, 71]. Moreover, there is also increasing evidence suggesting that quantum models could be a valuable mathematical tool to study certain puzzling phenomena in the cognitive sciences (see e.g., Refs. [72–74] and references therein).

Consider, for instance, the following experiment [75]: 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 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.

A general and unexpected prediction concerns order effects in questions [76]. 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 two experimens with 100 people [76].

While it is too early to confirm that this is a correct approach, 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 [76–78]. This may guide the search for commercially relevant datasets that can benefit from quantum supremacy. A report [79] 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 [71] 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. Twenty years may be in sync with the advent of more powerful and more portable quantum computing technologies. While this may not be so, the potential returns may be high.

Although we emphasize the case of cognitive sciences, it would be interesting to explore what other relevant and commercial datasets might exhibit such quantum-like correlations, and where quantum computers can have an advantage even with 50 to 100 of qubits, that is beyond the reach of classical simulation.

III. CHALLENGES IN QAML

As discussed before, 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 Ref. [22], it would be difficult for these algorithms to preserve any of the speedup claimed since they inherit the limitations of the Harrow-Hassidim-Lloyd (HHL) algorithm [80]. Besides the concerns expressed in Ref. [22], we raise the bar even higher in this work since our attention goes to the implementation of algorithms with potential quantum advantage in near-term quantum devices.

As pointed out in Sec. II, 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 circumvents most of the relevant limitations in the HHL algorithm. Be-
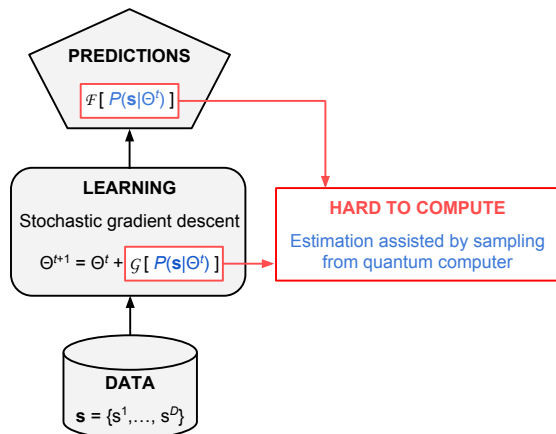
FIG. 2. General scheme for hybrid quantum-classical algorithms as one of the most promising research directions to demonstrate a quantum enhancement in ML tasks. A ML algorithm usually relies on data driving the fine tuning of the model parameters. In the case of training generative models one can use stochastic gradient descent to update the parameters Θ from time t to $t + 1$. The updates often require the estimation of an intractable function $\mathcal{G}$ which could be approximated if sampled from a probability distribution $P(\mathbf{s}|\Theta^t)$; it is in this computationally hard sampling step where calls to the quantum computer can be made to assist the training. In some cases, making predictions out of the trained model is also an intractable task, where sampling from the quantum computer can be of use as well. Examples of such hybrid approaches are illustrated further in specific ML models in Fig. 3 and in Fig. 4.

cause the input size is extremely large (e.g., number of users times number of products), the algorithm promises to significantly speedup 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 few decades, and our focus here is on devices with a few thousand qubits at most. Instead, our attention goes to hybrid quantum-classical algorithms where conventional computers are used in the tractable subroutines of the algorithms, while we make calls to the quantum computer to assist only in the intractable steps. Fig. 2 illustrates an example of this concept for the case of ML tasks.

There are several challenges which will generally impact in general 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 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.

A. 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 models 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 via stochastic gradient descent which 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 the parameters defining both the 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 can indirectly adjust the temperature of a Gibbs distribution by scaling the values of the parameters accordingly; for example, a model with a given temperature would be equivalent to another model with half the temperature if we divide all the parameters specifying the energy of the corresponding Gibbs distribution by two.

In the case of an experimental physical device, such as a quantum annealer, we cannot neglect the temperature since it is not under our control and 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 communication between the 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 Fig. 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, when the model parameters may be below the noise level of the quantum device, and then the quantum computer can be called when desired to continue the process.

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

(a) RBM-ba
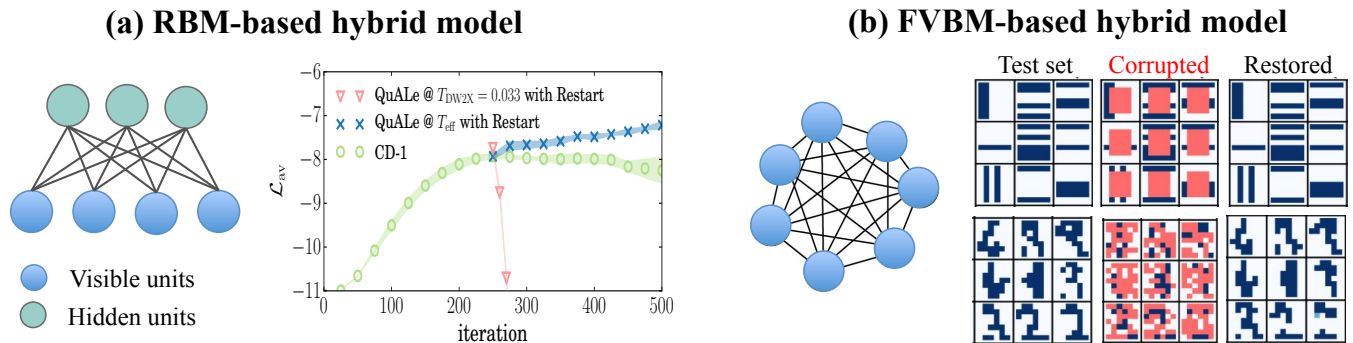

(b) FVBM-based hybrid model

FIG. 3. Examples of QAML implementations of different probabilistic graphical models, illustrating some of the challenges in near-term devices. (a) Restricted Boltzmann machines (RBM) are key components in deep learning approaches such as deep belief networks. Training of these architectures could be significantly improved if one were able to sample from the joint probability (Boltzmann) distribution over the visible and hidden units, a computationally intractable step usually performed with approximate Markov chain Monte Carlo approaches such as Contrastive Divergence (CD) [81]. Although a quantum annealer can be used to generate such samples, the results need to be combined with those obtained from the classical ML pipeline component within the hybrid approach. This *classical-quantum model compatibility* challenge (Sec. III A) is illustrated in the experimental results on the right. $\mathcal{L}_{av}$ is the average log-likelihood and represents the performance metric of the model training; the higher its value, the better the model is expected to represent the training dataset. As demonstrated in Ref. [20], estimating the instance-dependent effective temperature, T_{eff}, is key in this QAML approach. In all three lines, the first 250 iterations are done with the cheapest and widely used version of CD (denoted as CD-1). In the line with open circles, all the 500 iterations in the training is performed with CD-1, and is used as a baseline comparison for our QAML approach. Using our quantum-assisted learning (QuALe) algorithm where T_{eff} is estimated at each learning iteration (crosses), we can restart from the point where the classical CD-1 left off and improve the performance of the model with respect to the baseline. Assuming instead that T_{eff} is the physical temperature of the device, $T_{DW2X} = 0.033$ (triangles), such restart technique fails. (b) Visible-only generative models are often used either because of their tractability [39] or because of the convexity of the associated optimization problem, as in fully-visible Boltzmann machines (FVBM). In the latter, convexity does not mean tractability; exact learning would still require computation of the partition function which is intractable for nontrivial topologies. Even though there exist fast [82] and consistent approximations to the required gradients [83], we here consider quantum annealing as an alternative tool to sample from nontrivial topologies. In Ref. [21], we implemented and trained a hybrid QAML model by introducing a gray-box model (see Sec. III B) expected to be robust to noise in the programmable parameters and to deviations from the desired Boltzmann distribution, for example, due to non-equilibrium effects in the quantum dynamics. On the right we show examples of the reconstruction capabilities of the generative model using samples taken from the quantum device in experiments with up to 940 qubits.

B. 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 also lead to severe non-equilibrium distributions away from Gibbs [56, 84]. For example, this is one of the challenges towards scaling the approach in Refs. [20, 23] where the training of RBMs and DBNs require reliables 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 [84] was to use samples obtained from a quantum device to seed classical Gibbs samplers (e.g., MCMC). This has been explored further in Ref. [56] on a realization of fully-visible Boltzmann machines (FVBM). 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, however, gray-box models such as the FVBMs studied in Refs. [21, 25, 56] and illustrated in Fig. 3(b) may be a near-term solution to these issues. In a gray-box model, although we assume that the samples from the quantum device are described by a Gibbs-like distribution, the samples are compared directly to the training data at the level of the first and second moment statistics, without complete knowledge of the actual parameters that were implemented. This has the potential to increase the resilience against perturbations in the programmable parameters, since the emphasis moves from the model defined by these parameters to the quality of the samples obtained from the quantum device in terms of their closeness to those representing the data. As long as the estimated gradients for stochastic gradient descent

have a positive projection on the direction of the true gradients, the learning algorithm still 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, the gradients could be estimated under the assumption that the samples are from a quantum Gibbs distribution, but a suitable 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 are possible under classical or quantum Boltzmann machines with only visible units. The challenge of model compatibility and the need for temperature estimation previously mentioned arises from the inclusion of hidden units in the model. Since we do not need to know the distribution the quantum device is sampling from, the model is attached to the quantum device used in the training. Anytime we want to perform ML tasks of interest, such as the reconstruction of images or the generation of new images as shown in Fig. 3, we need to run the quantum device that was trained. Gray-box models with hidden variables could exploit the available resources from the quantum device, while coping with its intrinsic noise and parameter misspecification. For example, the gray-box model in Ref. [21] is used here to build a more powerful quantum-assisted Helmholtz machine with hidden variables; this is discussed in Sec. IV and described in detail in Ref. [35].

C. 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-based quantum computer, the overhead comes in the number of swaps required to make distant qubits talk to each other [85]. In any architecture, this compilation requirement needs to be considered in the algorithmic design. In the case of quantum annealers, for instance, achieving the topological connectivity of the desired model is only half the challenge.

Unless there is a way to sidestep the intrinsic locality at the hardware level, 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 hardware-embedded model represents the intended model on logical variables? In the case of quantum annealers, there are no known optimal solutions, although heuristic strategies have been proposed [86–88].

However, there is a way out in the type of ML applications we are considering: since the main goal in the training phase of a ML algorithm is to find the optimal parameters that minimize certain performance metric, ML itself is a parameter setting procedure. This is precisely the demonstration in Ref. [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 minor-embedding problem or to the mapping of the logical model of interest into physical hardware. An equally or even more important challenge, however, 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 refer to here as *the curse of limited connectivity*. In the case of gate-based quantum computation, Ref. [89] presents an illustrative experimental study, highlighting this challenge. It presents a comparison and analysis of the trade-off between connectivity and 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-based quantum computers.

D. 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 [90]. There has been work in quantum ML that follows the latter direction [91, 92]. 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, datasets of images with millions of pixels which can be in gray scale, with 256 intensities per pixel, or in color, represented by 3-dimensional vectors. We refer to this kind 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 vectors are stored into 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 would be able to prepare arbitrary states of this kind on near-term quantum systems. Even reading out all the amplitudes of an output vector might kill or significantly hamper any speedup claims [22].

In this perspective, we argue that near-term QAML algorithms should rather aim at encoding continuous vari-

ables 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; besides the issue of model compatibility described above, it will also have a tendency to overflow the programmable range of the device as the training algorithms naturally push parameters to large values; for instance, in quantum annealing each hidden unit is a qubit and its local field equals the weighted sum of all visible variables. Moreover, sampling the hidden variables $\mathbf{u}$ from the posterior distribution $P(\mathbf{u}|\mathbf{v})$ may be highly problematic because in gate-based computers the preparation of arbitrary states is an open challenge. The problem is also intrinsic in all the proposals that treat quantum annealers as hidden layers of a neural network because quantum annealers are inherently serial devices. If we were to train the model in an on-line setting, i.e., using just one point per parameter update, this may not be a problem. However, the process is very slow in the more common batch training setting; for instance, the standard ML dataset of handwritten digits MNIST is composed of 60 000 training points, hence we would need to program the quantum device 60 000 times for each full learning iteration. As we will see in the next section, deep learning may provide solutions to these challenge. We propose a new hybrid quantum-classical paradigm where the objective is to solve most of the issues discussed here. This approach also allows us to work with a flexible quantum-assisted framework capable of incorporating complex ML datasets as discussed below.

IV. 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, they include the encoding of continuous variables, the limited number of variables, the need to prepare states, and measurement for each training 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 Ref. [35].

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 Sec. II A for a brief introduction). Here $\mathbf{v}$ are the visible units that represent the data and $\mathbf{u}$ are unobservable or hidden units that serve to capture non-trivial correlations. It is common to use binary valued stochastic hidden units as they can express the presence or absence of features in the data. The set of hidden units 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 *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 parametrized density matrix, $P_{QC}(\mathbf{u}) = \langle \mathbf{u}|\hat{\rho}|\mathbf{u}\rangle$. As an example, the density could be parametrized 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. On the other hand, the conditional distribution $P(\mathbf{v}|\mathbf{u})$ is a classical component 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 solving one of the aforementioned problems. Because the quantum device works on a lower dimensional binary representation of the data, we are 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 Sec. II A, 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 [93, 94] 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 [40–44, 46–49]. 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 recogni-
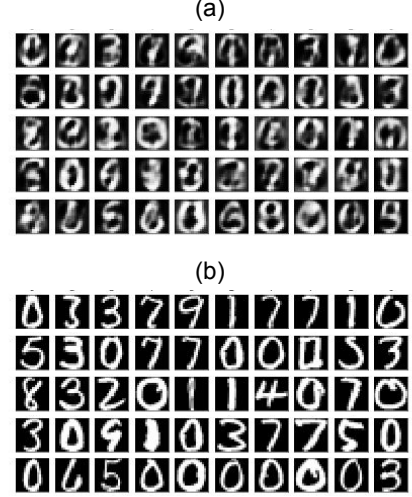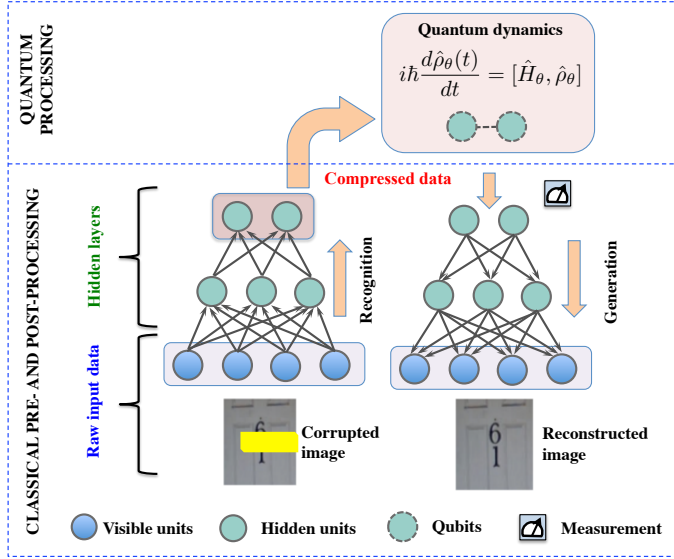
FIG. 4. *Generation of handwritten digits with a type of Quantum-Assisted Helmholtz Machine (QAHM).* (Left) The QAHM is the framework we propose to work with intractable generative models and complex ML datasets in near-term devices. By complex, we refer to datasets where the number of variables is much larger than the number of qubits available in the quantum device; also, as in many real-world applications, the data may be continuous rather than discrete. The framework is very flexible as it employs a quantum computer to model the deepest hidden layers, containing the most abstract representation of the data. This low-dimensional compact representation is where we believe the quantum device can capture non-trivial correlations and where quantum distributions might have a significant effect. The number of hidden variables in the deepest layers of a network is much smaller than the number of visible variables, making it ideal for near-term implementation on early quantum technologies either on quantum annealers, or universal gate-model quantum computers. θ indicates the parameters of the quantum computer to be learned and that are used to control the samples obtained from it. Although we illustrate in panel (a) and (b) a realization on a quantum annealer from Ref. [35], extension to gate-based architectures are currently in progress. (a) Artificial data generated by a QAHM implemented on the D-Wave 2000Q quantum annealer, trained on a sub-sampled version of the MNIST dataset with 16×16 continuous valued pixels and 10 binary variables indicating the class in $\{0, \ldots, 9\}$. Both recognition and generator networks have 266 visible variables and two layers of 120 and 60 hidden variables, respectively. The samples are generated after the quantum-assisted training phase by first sampling the deepest layer with the D-Wave 2000Q and then transforming those samples via a down pass through the classical part of the generator network. The experiments are run using 1644 qubits of the D-Wave 2000Q quantum annealer. Some of the samples resemble blurry variations of digits written by humans; this problem affects other approaches as well—only the development of generative adversarial networks (GANs) [51] led to much sharper artificial images. (b) Samples from the MNIST dataset that are closest in Euclidean distance to those generated by our model. The model does not simply memorize the training set, but rather reproduce its statistics. The QAHM can be fine-tuned to provide sharper results and can be extended to other quantum hardware architectures as well.

tion 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 [47, 93, 95].

The general architecture of a type of QAHM is illustrated in Fig. 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 units 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 Sec. III B), 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.

We have 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. While each gray-scale pixel is characterized by an integer value in $\{0, \ldots, 255\}$, we normalize it and interpret it as a continuous variable in the range $[-1, +1]$. Furthermore, there are 10 binary variables indicating the class in $\{0, \ldots, 9\}$. The 266 visible variables needed to encode this dataset could, in principle, be embedded directly on the D-Wave 2000Q using only-visible models and a much poorer representation of the data via a naive binarization. However, even in this approach, we would have to restrict to only-visible models with relatively sparse topologies because only about 60 variables with effective all-to-all connectivity can be embedded in the D-Wave 2000Q. The

sparsity and the absence of hidden variables alone can severely limit the capacity to model the dataset. So, our approach not only improves on the data representation by using continuous variables and on the modelling capacity by including hidden variables, but it also enables the handling of larger datasets than would be possible in state-of-the-art quantum annealers, even if restricting to less powerful approaches.

We used a classical recognition network and a quantum-assisted generator network, both with 266 visible units and two hidden layers of 120 and 60 units, respectively. The deepest layer of 60 units was mapped to 1644 qubits in D-Wave 2000Q using the approach in Ref. [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 Fig. 4(a). Although these are only preliminary results, and we cannot compete with state-of-the-art ML, the artificial data often resemble digits written by humans. Figure 4(b) shows the image in the training set that is closest in Euclidean distance to each of the samples at the corresponding location. We can see that the artificial data generated by the network is not merely a copy of the training set, but they do present variations and novelty in some cases; this reflects *generalization* capabilities which is a desired feature in a unsupervised generative model. While the artificial data may also looks blurry, this is a problem that affects other approaches as well—only the development of GANs [51] led to much sharper artificial images.

V.   SUMMARY

Machine learning (ML) has been presented as one of the application 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. More specifically, while most of the significant progress demonstrating potential quantum-enhancement for ML has its merits in showing that quantum computers have the potential of being very efficient at doing linear algebra (e.g., [11, 13–18, 80]), as discussed in Ref. [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 speedup 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 Ref. [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 such highly desirable intractable applications, in Sec. II we presented the case of sampling from complex probability distributions with quantum devices. There is potential here for boosting the training of fully 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 seem to be the case of some datasets from the field of cognitive sciences (see e.g., Refs. [71, 75, 76, 96]; also [72–74] and references therein). Other hard ML problems have been mentioned elsewhere [34]. 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, limited qubit connectivity, which will result in an overhead of resources in terms of qubits or the number of gate operations in the case of the adiabatic or the gate model, respectively. It is also important to take into consideration the model complexity 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 modelling 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 Sec. III 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* [35]. This new approach aims to solve some of the most relevant pressing issues towards implementing 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 but discrete dataset with a 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 the reconstruction of images, or the generation of images resembling those in the dataset, but that are generated from samples from the quantum device.

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. We hope researchers interested in QAML can find useful the insights we provide here not only considering algorithmic quantum speedups, but also looking into intractable problems and under more meaningful performance metrics for ML practitioners such as statistical efficiency. The intersection of finding small-in-size intractable problems and the development of new hybrid QAML algorithms which factor in the challenges of working with real-world devices, is what we find the ideal scenario; regardless of the devices being quantum annealers or gate model quantum computers. We hope the community makes a leap in this direction now that more powerful and larger quantum annealers and gate model quantum devices are beccoming available to the scientific community.

[1] Masoud Mohseni, Peter Read, Hartmut Neven, Sergio Boixo, Vasil Denchev, Ryan Babbush, Austin Fowler, Vadim Smelyanskiy, and John Martinis, "Commercialize quantum technologies in five years," Nature **543**, 171–174 (2017).

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems* (2012) pp. 1097–1105.

[3] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine **29**, 82–97 (2012).

[4] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *Intelligent Vehicles Symposium (IV), 2011 IEEE* (IEEE, 2011) pp. 163–168.

[5] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature **542**, 115–118 (2017).

[6] Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Wei Xie, Gail L. Rosen, Benjamin J. Lengerich, Johnny Israeli, Jack Lanchantin, Stephen Woloszynek, Anne E. Carpenter, Avanti Shrikumar, Jinbo Xu, Evan M. Cofer, David J. Harris, Dave DeCaprio, Yanjun Qi, Anshul Kundaje, Yifan Peng, Laura K. Wiley, Marwin H. S. Segler, Anthony Gitter, and Casey S. Greene, "Opportunities and obstacles for deep learning in biology and medicine," bioRxiv (2017), 10.1101/142760.

[7] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," Nature **518**, 529–533 (2015).

[8] Harmut Neven, Vasil S Denchev, Marshall Drew-Brook, Jiayong Zhang, William G Macready, and Geordie Rose, "Binary classification using hardware implementation of quantum annealing," in *Demonstrations at NIPS-09, 24th Annual Conference on Neural Information Processing Systems* (2009) pp. 1–17.

[9] Zhengbing Bian, Fabian Chudak, William G Macready, and Geordie Rose, *The Ising model: teaching an old problem new tricks*, Tech. Rep. (D-Wave Systems, 2010).

[10] Misha Denil and Nando De Freitas, "Toward the implementation of a quantum RBM," NIPS Deep Learning and Unsupervised Feature Learning Workshop (2011).

[11] Nathan Wiebe, Daniel Braun, and Seth Lloyd, "Quantum algorithm for data fitting," Physical review letters **109**, 050505 (2012).

[12] KristenL. Pudenz and DanielA. Lidar, "Quantum adiabatic machine learning," Quantum Information Processing **12**, 2027–2070 (2013).

[13] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," arXiv:1307.0411 (2013).

[14] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd, "Quantum support vector machine for big data classification," Phys. Rev. Lett. **113**, 130503 (2014).

[15] G. Wang, "Quantum Algorithm for Linear Regression," ArXiv e-prints (2014), arXiv:1402.0660 [quant-ph].

[16] Z. Zhao, J. K. Fitzsimons, and J. F. Fitzsimons, "Quantum assisted Gaussian process regression," ArXiv e-prints (2015), arXiv:1512.03929 [quant-ph].

[17] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, "Quantum principal component analysis," Nature Physics **10**, 631–633 (2014).

[18] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "Prediction by linear regression on a quantum computer," Physical Review A **94**, 022342 (2016).

[19] Krysta M. Svore Nathan Wiebe, Ashish Kapoor, "Quantum deep learning," arXiv:1412.3489 (2015).

[20] Marcello Benedetti, John Realpe-Gómez, Rupak Biswas, and Alejandro Perdomo-Ortiz, "Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning," Phys. Rev. A **94**, 022308 (2016).

[21] Marcello Benedetti, John Realpe-Gómez, Rupak Biswas, and Alejandro Perdomo-Ortiz, "Quantum-assisted learning of graphical models with arbitrary pairwise connectivity," arXiv:1609.02542 (2016).

[22] Scott Aaronson, "Read the fine print," Nature Physics **11**, 291–293 (2015), commentary.

[23] Steven H. Adachi and Maxwell P. Henderson, "Application of quantum annealing to training of deep neural networks," arXiv:1510.06356 (2015).

[24] Nicholas Chancellor, Szilard Szoke, Walter Vinci, Gabriel Aeppli, and Paul A Warburton, "Maximum-entropy inference with a programmable annealer," Scientific reports **6** (2016).

[25] Mohammad H. Amin and Evgeny Andriyash and Jason Rolfe and Bohdan Kulchytskyy and Roger Melko, "Quantum Boltzmann Machine," arXiv:1601.02036 (2016).

[26] Maria Kieferova and Nathan Wiebe, "Tomography and generative data modeling via quantum boltzmann training," arXiv preprint arXiv:1612.05204 (2016).

[27] Iordanis Kerenidis and Anupam Prakash, "Quantum recommendation systems," arXiv preprint arXiv:1603.08675 (2016).

[28] Peter Wittek and Christian Gogolin, "Quantum enhanced inference in markov logic networks," Scientific Reports **7** (2017).

[29] Thomas E. Potok, Catherine Schuman, Steven R. Young, Robert M. Patton, Federico Spedalieri, Jeremy Liu, Ke-Thia Yao, Garrett Rose, and Gangotree Chakma, "A study of complex deep learning networks on high performance, neuromorphic, and quantum computers," arXiv:1703.05364 (2017).

[30] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione, "An introduction to quantum machine learning," Contemporary Physics **56**, 172–185 (2015).

[31] Jonathan Romero, Jonathan P. Olson, and Alan Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," arXiv:1612.02806 (2017).

[32] Jeremy Adcock, Euan Allen, Matthew Day, Stefan Frick, Janna Hinchliff, Mack Johnson, Sam Morley-Short, Sam Pallister, Alasdair Price, and Stasja Stanisic, "Advances in quantum machine learning," arXiv preprint arXiv:1512.02900 (2015).

[33] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," arXiv preprint arXiv:1611.09347 (2016).

[34] C. Ciliberto, M. Herbster, A. Davide Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, "Quantum machine learning: a classical perspective," ArXiv e-prints (2017), arXiv:1707.08561 [quant-ph].

[35] Marcello Benedetti, John Realpe-Gómez, and Alejandro Perdomo-Ortiz, "Quantum-assisted helmholtz machines: A quantum-classical deep learning framework for industrial datasets in near-term devices," Preprint available at arXiv.org (2017).

[36] John Realpe-Gómez, Marcello Benedetti, and Alejandro Perdomo-Ortiz, "Quantum models for quantum-like datasets in near-term devices," In preparation (2017).

[37] Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence 35, 1798–1828 (2013).

[38] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio, "Why does unsupervised pre-training help deep learning?" Journal of Machine Learning Research 11, 625–660 (2010).

[39] Ian Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," arXiv preprint arXiv:1701.00160 (2016).

[40] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114 (2013).

[41] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," arXiv preprint arXiv:1401.4082 (2014).

[42] Andriy Mnih and Karol Gregor, "Neural variational inference and learning in belief networks," arXiv preprint arXiv:1402.0030 (2014).

[43] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther, "Ladder variational autoencoders," in Advances in Neural Information Processing Systems (2016) pp. 3738–3746.

[44] Jason Tyler Rolfe, "Discrete variational autoencoders," arXiv preprint arXiv:1609.02200 (2016).

[45] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski, "A learning algorithm for boltzmann machines," Cognitive science 9, 147–169 (1985).

[46] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh, "A fast learning algorithm for deep belief nets," Neural computation 18, 1527–1554 (2006).

[47] Jörg Bornschein and Yoshua Bengio, "Reweighted wake-sleep," arXiv preprint arXiv:1406.2751 (2014).

[48] Jorg Bornschein, Samira Shabanian, Asja Fischer, and Yoshua Bengio, "Bidirectional helmholtz machines," in International Conference on Machine Learning (2016) pp. 2511–2519.

[49] Ruslan Salakhutdinov and Geoffrey Hinton, "Deep boltzmann machines," in Artificial Intelligence and Statistics (2009) pp. 448–455.

[50] Yoshua Bengio, Eric Laufer, Guillaume Alain, and Jason Yosinski, "Deep generative stochastic networks trainable by backprop," in International Conference on Machine Learning (2014) pp. 226–234.

[51] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in Advances in neural information processing systems (2014) pp. 2672–2680.

[52] Yoshua Bengio et al., "Learning deep architectures for ai," Foundations and trends® in Machine Learning 2, 1–127 (2009).

[53] Dan Roth, "On the hardness of approximate reasoning," Artificial Intelligence 82, 273–302 (1996).

[54] Yoshua Bengio, Grégoire Mesnil, Yann Dauphin, and Salah Rifai, "Better mixing via deep representations," in Proceedings of the 30th International Conference on Machine Learning (ICML-13) (2013) pp. 552–560.

[55] V. Dumoulin, I. J. Goodfellow, A. C. Courville, and Y. Bengio, "On the challenges of physical implementations of RBMs," in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada. (2014) pp. 1199–1205.

[56] Dmytro Korenkevych, Yanbo Xue, Zhengbing Bian, Fabian Chudak, William G Macready, Jason Rolfe, and Evgeny Andriyash, "Benchmarking quantum hardware for training of fully visible boltzmann machines," arXiv preprint arXiv:1611.04528 (2016).

[57] Anirban Narayan Chowdhury and Rolando D Somma, "Quantum algorithms for gibbs sampling and hitting-time estimation," arXiv preprint arXiv:1603.02940 (2016).

[58] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," Nature 521, 436 – 444 (2015).

[59] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum, "Human-level concept learning through probabilistic program induction," Science 350, 1332–1338 (2015).

[60] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, John M Martinis, and Hartmut Neven, "Characterizing quantum supremacy in near-term devices," arXiv preprint arXiv:1608.00263 (2016).

[61] Mile Gu, Karoline Wiesner, Elisabeth Rieper, and Vlatko Vedral, "Quantum mechanics can reduce the complexity of classical models," Nature Communications 3, 762 (2012).

[62] Matthew S Palsson, Mile Gu, Joseph Ho, Howard M Wiseman, and Geoff J Pryde, "Experimentally modeling stochastic processes with less memory by the use of a quantum processor," Science Advances 3, e1601302 (2017).

[63] Thomas J Elliott and Mile Gu, "Occam's vorpal quantum razor: Memory reduction when simulating continuous-time stochastic processes with quantum devices," arXiv preprint arXiv:1704.04231 (2017).

[64] Kenneth P Burnham and David R Anderson, Model selection and multimodel inference: a practical information-theoretic approach (Springer Science & Business Media, 2003).

[65] E. T. Jaynes, "Information theory and statistical mechanics," Phys. Rev. 106, 620–630 (1957).

[66] Matjaž Perc, Jillian J Jordan, David G Rand, Zhen Wang, Stefano Boccaletti, and Attila Szolnoki, "Statistical physics of human cooperation," Physics Reports (2017).

[67] John Realpe-Gómez, Giulia Andrighetto, Gustavo Nardin, and Javier Antonio Montoya, "Balancing selfishness and norm comformity can explain human behaviour in large-scale prisoners dilemma games," arXiv preprint arXiv:1608.01291v2 (2017).

[68] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina, "Analytic and algorithmic solution of random satisfiability problems," Science **297**, 812–815 (2002).

[69] Marc Mezard and Andrea Montanari, *Information, Physics, and Computation* (Oxford University Press, Inc., New York, NY, USA, 2009).

[70] Giacomo Mauro D'Ariano, Giulio Chiribella, and Paolo Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, 2017).

[71] John Realpe-Gómez, "Quantum theory as a self-referential theory of inference," arXiv preprint arXiv:1705.04307 (2017).

[72] Peter D Bruza, Zheng Wang, and Jerome R Busemeyer, "Quantum cognition: a new theoretical approach to psychology," Trends in cognitive sciences **19**, 383–393 (2015).

[73] Jerome R Busemeyer and Peter D Bruza, *Quantum models of cognition and decision* (Cambridge University Press, 2012).

[74] Diederik Aerts, Jan Broekaert, Liane Gabora, and Sandro Sozzo, "Quantum structures in cognitive and social science," Frontiers in psychology **7** (2016).

[75] Jerome R Busemeyer, Zheng Wang, and Richard M Shiffrin, "Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency." Decision **2**, 1 (2015).

[76] Zheng Wang, Tyler Solloway, Richard M Shiffrin, and Jerome R Busemeyer, "Context effects produced by question orders reveal quantum nature of human judgments," Proceedings of the National Academy of Sciences **111**, 9431–9436 (2014).

[77] Carlos Gracia-Lázaro, Alfredo Ferrer, Gonzalo Ruiz, Alfonso Tarancón, José A Cuesta, Angel Sánchez, and Yamir Moreno, "Heterogeneous networks do not promote cooperation when humans play a prisoners dilemma," Proceedings of the National Academy of Sciences **109**, 12922–12926 (2012).

[78] Mario Gutiérrez-Roig, Carlos Gracia-Lázaro, Josep Perelló, Yamir Moreno, and Angel Sánchez, "Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence," Nature communications **5** (2014).

[79] White House, "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 (2016).

[80] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd, "Quantum algorithm for linear systems of equations," Physical review letters **103**, 150502 (2009).

[81] Geoffrey E Hinton, "Training products of experts by minimizing contrastive divergence," Neural Computation **14**, 1771–1800 (2002).

[82] Miguel A Carreira-Perpinan and Geoffrey E Hinton, "On contrastive divergence learning." in *Aistats*, Vol. 10 (2005) pp. 33–40.

[83] Aapo Hyvärinen, "Consistency of pseudolikelihood estimation of fully visible boltzmann machines," Neural Computation **18**, 2283–2292 (2006).

[84] Jack Raymond, Sheir Yarkoni, and Evgeny Andriyash, "Global warming: Temperature estimation in annealers," arXiv:1606.00919 (2016).

[85] Robert Beals, Stephen Brierley, Oliver Gray, Aram W Harrow, Samuel Kutin, Noah Linden, Dan Shepherd, and Mark Stather, "Efficient distributed quantum computing," in *Proc. R. Soc. A*, Vol. 469 (The Royal Society, 2013) p. 20120686.

[86] Vicky Choi, "Minor-embedding in adiabatic quantum computation: Ii. minor-universal graph design," Quantum Information Processing **10**, 343–353 (2011).

[87] Alejandro Perdomo-Ortiz, Joseph Fluegemann, Rupak Biswas, and Vadim N Smelyanskiy, "A performance estimator for quantum annealers: gauge selection and parameter setting," arXiv:1503.01083 (2015).

[88] Kristen L. Pudenz, "Parameter setting for quantum annealers," arXiv:1611.07552 (2016).

[89] Norbert M Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A Landsman, Kenneth Wright, and Christopher Monroe, "Experimental comparison of two quantum computing architectures," Proceedings of the National Academy of Sciences , 201618020 (2017).

[90] Seth Lloyd and Samuel L Braunstein, "Quantum computation over continuous variables," Physical Review Letters **82**, 1784 (1999).

[91] Hoi-Kwan Lau, Raphael Pooser, George Siopsis, and Christian Weedbrook, "Quantum machine learning over infinite dimensions," Physical Review Letters **118**, 080501 (2017).

[92] S. Das, G. Siopsis, and C. Weedbrook, "Continuous-variable quantum Gaussian process regression and quantum singular value decomposition of non-sparse low rank matrices," ArXiv e-prints (2017), arXiv:1707.00360 [quant-ph].

[93] Geoffrey E. Hinton, Peter Dayan, Brendan J. Frey, and Radford M. Neal, "The wake-sleep algorithm for unsupervised neural networks," Science **268**, 1158 (1995).

[94] Peter Dayan, Geoffrey E Hinton, Radford M Neal, and Richard S Zemel, "The helmholtz machine," Neural computation **7**, 889–904 (1995).

[95] Peter Dayan and Geoffrey E Hinton, "Varieties of helmholtz machine," Neural Networks **9**, 1385–1403 (1996).

[96] Peter D Kvam, Timothy J Pleskac, Shuli Yu, and Jerome R Busemeyer, "Interference effects of choice on confidence: Quantum characteristics of evidence accumulation," Proceedings of the National Academy of Sciences **112**, 10645–10650 (2015).