Quantum neuromorphic hardware for quantum artificial intelligence

Enrico Prati
Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: enrico.prati@cnr.it

Abstract. The development of machine learning methods based on deep learning boosted the field of artificial intelligence towards unprecedented achievements and application in several fields. Such prominent results were made in parallel with the first successful demonstrations of fault tolerant hardware for quantum information processing. To which extent deep learning can take advantage of the existence of a hardware based on qubits behaving as a universal quantum computer is an open question under investigation. Here I review the convergence between the two fields towards implementation of advanced quantum algorithms, including quantum deep learning.

1. Learning from the working principles of the brain
The invention of the modern hardware architecture of computers, the so called Von Neumann architecture, is often attributed to John Von Neumann. It was only in 1998 that Konrad Zuse was officially recognized by the Conference on the History of Computing in Padernborn as the inventor of such architecture as he developed and patented the Z1 model, based on floating point and binary digits in 1938 and the Z3 with the full modern architecture in 1941 [1], before Von Neumann [2]. In order to acknowledge both the two fathers of such an architecture, we should call it Von Neumann-Zuse architecture. Though, there is an additional reason for which Von Neumann should be acknowledged: when inventing the computer, he started with the aim to create an artificial brain, but in order to ensure control of the computation with the tools available at that time, he opted to impose two major simplifications, i.e. the system, contrarily to the unreliable and asynchronous building blocks of brain namely the real neurons, was based on deterministic and synchronous design. [2]. This enabled him to create the computer with telegraph relays and vacuum tubes, but he left open the alternative of pursuing the opposite strategy. As he was aware that brain processes information in parallel, and considered the brain a parallel and asynchronous computer, he implicitly also defined the characteristics of a kind of Non-Von Neumann (NVN) architecture. Indeed, he pointed out two prominent aspects of biological neurons: they process information in a parallel fashion and they are very unreliable (a synapse works in average once every 10 solicitations) and affected by several unavoidable sources of noise, ranging from atomic fluctuations across their membrane to potential fluctuation due to release of molecules by vesicles. Despite such sources of inaccuracy and indeterminism, neurons and synapses successfully carry information processing by redundancy and parallelism. As discussed later, there are several hardware approaches enabling device functionalities towards
integrate and fire response of neurons and synaptic plasticity, which generally operate in low disturbance conditions. For researchers working in quantum information processing, where noise is an unavoidable source of loss of coherence and therefore of information, the fact that the brain perfectly works despite the noise may open a very tempting path. Furthermore, its working principles become surprising if one observes that the brain works not only despite the noise, but more precisely thanks to noise. There is a long list of useful and necessary effects of noise in order to achieve and improve brain functionalities in several cortical areas[4, 5, 6]. The deep learning algorithms [7, 8] inspired by the layered structure of visual cortex has pushed to the right direction the exploitation of such a parallelism and it has provided a new paradigm for the pattern recognition in complex data structures. In some cases, such as for Boltzmann machines, the statistical behavior reproduces the intrinsic variability of neuronal activity. Deep learning neural networks can be applied not only to image processing, but to any kind of data set, ranging from DNA research to economics to identify patterns and regularities often hidden to human intuition. In the next section the role of quantum mechanics in brain is discussed.

2. Quantum effects and information processing in the brain
There is a number of researches claiming that quantum mechanics plays a direct role in the information processing in the brain. By direct role, here we intend that the brain has been supposed to exploit quantum superposition of states, for instance in relatively large molecules such as chains in microtubules. Even if the topic is debated, there is a relatively large leap between the existence of quantum superposition effects in the brain, and their exploitation in the information processing. Quantum effects are normally present at the atomic scale and they are involved from instance at the scale of ion channel processes, where single Na, K and Ca ions move across the membrane of the neurons, to molecular dynamics and matching in the complex creation of proteins from genes as an expression of DNA. Several noise processes present in the brain are ultimately related to random quantum fluctuations.[9, 10, 11, 12] Despite a number of
interesting claims about room temperature observable effects in biology of quantum mechanics [13], including some structures present in the brain [14], there is currently no evidence that such a superposition is used to perform the signaling process between neurons. Therefore, if one is interested to create a neuromorphic hardware exploiting quantum mechanics, one has two options. One may look either to include quantum fluctuations in the classical hardware, in order to natively implement sources of fluctuations which proved effective for the information processing of several cortical areas, or alternatively to export the organization structure of some cortical areas such as visual cortex, and to extend the deep learning architectures towards a quantum domain. Currently, there is no evidence that superposition of states is a key ingredient per se to support neuronal information processing. Among the neuromorphic classical hardware, we may include aluminum atomic bridges, CuS and AgS synapses, silicon CMOS platform, HfO₂ RRAMs, InGaZnO synapses, GST synapses, hybrid organic/silicon transistors [10]. Among them, only silicon devices and HfO₂ devices proved effective to cover both plasticity and spontaneous fluctuations [10, 15, 16, 17, 18]. Artificial neurons are currently built only by silicon devices. A prominent example of deep learning network designed on CMOS silicon hardware at circuit level has been provided in [19]. The opposite strategy, namely the exploitation of quantum hardware to extend the deep learning by quantum algorithms has been investigated. In the last section of this mini-review, the latter topic is outlined.

3. From machine learning to quantum artificial intelligence
In the years, three main methods have been developed to train both feedforward and recurrent neural networks: supervised learning (the training is based on correct input-output pairs), reinforcement learning (the correct input-output pairs are never presented), and unsupervised learning (contrarily to previous, the input data are unlabelled so there is no estimate of the accuracy of the output) and hybrid combinations. The implementation can be based on either feedforward networks, whose topology is such that input signal can only travel in one direction paths from the input neurons to the output neurons, or recurrent networks where loops in the topology are present. Deep learning approach consists in the separation of the nodes into groups providing multiple layers used as nonlinear processing units, which create a hierarchy from low-level to high-level features representation layers (Figure 1). Deep learning has been employed for deep supervised learning such as supervised deep feedforward multilayer perceptrons invented by Ivakhnenko and Lapa in 1965 [20] and the more recent long short-term memory recurrent neural networks [21] respectively, deep reinforcement learning [22, 23] applied to both recurrent (the backgammon player of [24]) and later feedforward (more specifically of convolutional kind, used to beat Atari2000 arcades and in AlphaGo [25]) neural networks respectively, and deep unsupervised learning in both versions based on feedforward (autoencoders hierarchies [26] since 1987) and recurrent (both deep belief networks based on restricted Boltzmann machines [27] and neural history compressor [28]) neural networks. It is worth mentioning at this point that restricted Boltzmann machines are a subclass of recurrent neural networks. They are a special case of a Hopfield network which is a neural network where all the nodes are symmetrically connected with the other nodes so that its weights matrix is symmetric. In a Boltzmann machine the internal connections of a Hopfield network define a bipartite network split between a visible layer and a hidden layer, and the units are stochastic and subject to thermalization according to the Boltzmann statistics. A restricted Boltzmann machine (RBM) consists in a subclass of Boltzmann machines for which no connections internally exist between nodes of the visible and of the hidden layers respectively. A deep RBM is composed by many RBM layers (see Figure 2). The prospect of employing spiking neuron in artificial neural networks has been summarized in the review of Schmidhuber [29], even if currently spiking neurons cannot yet compete with the best traditional deep NNs in practical applications. On the other hand, the advancement of scalable hardware qubits such as superconductive [30] and silicon
Figure 2. Hierarchy of different kind of Recurrent Neural Networks (RNN) including quantum domain for unsupervised learning. Starting from Hopfield networks (its energy cost function is based on a classical Ising Hamiltonian, symmetric and real valued, corresponding to a symmetric operator acting on usual classical states), one can both restrict the topology to more sophisticated cases, such as bipartite (between visible $v$ and hidden $h$ layer) and stochastic, which corresponds to so called Boltzmann machines, and by further restricting the connection so that there are neither $v$-$v$ nodes connections, nor $h$-$h$ nodes connections – the so called restricted Boltzmann machine (RBM) – which in turn can be used for deep belief networks by using a stack of them. On the other side, the Hopfield network can be implemented in a more general domain such that provided by quantum hardware for quantum annealing with Ising Edwards-Anderson Hamiltonian or a general Ising Hamiltonian acting on quantum states. Quantum deep restricted Boltzmann machines, or more briefly quantum deep learning, corresponds to a Deep Belief Network exploited in a more extended class of Hamiltonians spanning a quantum space of vectors.

[31, 32] qubits make closer the implementation of quantum algorithms. Indeed, fault tolerant quantum information processing requires handling of the fragile states of physical qubit. A number of quantum error correction codes such as Steane code [36] and surface code [37] have been developed to assess logical error-free qubits. Such logical qubits are managed by the application layer for algorithm implementation with no care about the underlying hardware. Most significant architecture are classified among circuital (digital) quantum computing, based on universal sets of quantum logic ports, and quantum annealing (and adiabatic quantum computing). A major example of hybrid architecture has also been demonstrated [30], enabling the control of 5 logical qubits distilled by a virtual qubit layer constituted by 500 physical qubits. The quantum annealing corresponds to the minimization of a complex functional and is suitable for optimization problems by exploiting minimization search via quantum tunneling in a complex potential landscape mapping the problem to be solved. It appears that the analogy with minimization problem solved by deep learning algorithm in the regression operation is more than just qualitative. For instance, optimization problems like the four-queens on a chessboard has been mapped onto an adiabatic quantum computing framework by exploiting the analogy between Hopfield network and its quantum counterpart [33]. In the remaining of this section,
two examples of quantum algorithm for quantum artificial intelligence are reviewed. The first examples consists in the set of quantum deep learning algorithms provided by Microsoft in 2014 [34]. Such algorithms is based on the quantum mechanical extension of a deep unsupervised recurrent neural network where the recurrent neural network is organized as a deep restricted Boltmann machine. In its classical version, the latter is mathematically represented by an Ising model which is in thermal equilibrium. Let’s clarify where the quantumness sets in. Usually the training of the network is carried by employing gradient descent methods. The energy of the system is expressed in terms of parameters called weights and biases respectively, which enable to define a gradient descent to find such weights and biases in order to maximize a like-hood function during training by appropriately updating such values. The difficulty of computing the gradients for updating the parameters is exponentially hard by raising the number of both visible and hidden nodes, so two quantum algorithms have been proposed to speed up such process: the Gradient Estimation via Quantum Sampling (GEQS) and the Gradient Estimation via Quantum Amplitude Estimation (GEQAE) respectively. In general, the starting point when training the network is far from the Gibbs state which would maximize the probability of a given Boltzmann machine, so the quantum speedup is provided by its natural capability to rapidly converge thanks to qubits to a mean field state (a non-uniform prior distribution for the probabilities of each configuration) which is closer to the Gibbs state and therefore suitable for accelerating the search. GEQS consists in the preparation of a mean field state by exploiting the polynomial time required by qubits in order to achieve it. GEQAE algorithm is a new form of training, possible in cases where the training data is provided via a quantum oracle (a black box whose implementation is not specified in the algorithm), which allows access to the training data in superposition rather than sequentially. This method is preferable for large training sets of data. Finally, one may observe that advantages by employing even small scale of noisy physical qubits is sufficient to efficiently compute solutions even without applying quantum error correction codes. A five superconductive qubits called transmons chip has been fabricated and programmed to use it as a quantum oracle and to speed up a learning task, known as binary classification (to identify an unknown mapping between a set of bits onto 0 or 1).[35]

4. Conclusion
To conclude, the development of artificial intelligence based on the emulation of the functional organization of the visual cortex enabled efficient machine learning called deep learning. Such algorithm may be extended to its quantum version to be implemented on a quantum computer. Quantum deep learning by a quantum deep belief network is the first example of a class of quantum algorithms which has been proved to outperform classical algorithms. Such algorithm creates abstraction layers based on quantum superpositions which belong to a representations space not accessible to human brain understanding.

References
[1] Zuse H 2016 Konrad Zuse’s Computer Z3 http://www.zuse.de
[2] Von Neumann J 1993 First Draft of a Report on the EDVAC IEEE Annals of the History of Computing 15 4 27-75
[3] Von Neumann J 2000 The Computer and the Brain (New Haven: Yale University Press, 2nd Edition, Mrs. Hepsa Ely Sillman Memorial Lectures) Applied Physics Letters 102 123701
[4] Burton R M and Mpitsos G J 1992 Event-dependent control of noise enhances learning in neural networks Neural Networks 5 4 627–637
[5] Brascamp J W et al. 2006 The time course of binocular rivalry reveals a fundamental role of noise Journal of vision 6 11 8
[6] Orlandi J G et al. 2013 Noise focusing and the emergence of coherent activity in neuronal cultures Nature Physics 9 9 582–590
[7] LeCun Y et al 1998 Gradient-Based Learning Applied to Document Recognition Proc of the IEEE 86 11 2276–2324
[8] Hinton G E and Salakhutdinov R R 2006 Reducing the Dimensionality of Data with Neural Networks *Science* **313** 504

[9] Richardson MJE and Gerstner W 2005 Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance *Neural computation* **17** 4 923–947

[10] Prati E 2016 Atomic scale nanoelectronics for quantum neuromorphic devices: comparing different materials *Int. J. of Nanotech.* **13** 7 509–523

[11] Prati E 2016 Noise-assisted transmission of spikes in Maeda–Makino artificial neuron arrays *Int. J. of Paral., Emerg. and Distr. Syst.* doi:10.1080/17445760.2016.1189914

[12] Hasani R et al 2017 Control of the Correlation of Spontaneous Neuron Activity in Biological and Noise–activated CMOS Artificial Neural Microcircuits arxiv:1702.07426v1

[13] Panitchayangkoon G et al 2010 Long-lived quantum coherence in photosynthetic complexes at physiological temperature *PNAS* **107** 29 12766–12770

[14] Sahni S et al. 2013 Multi-level memory-switching properties of a single brain microtubule

[15] Prati E, Fanciulli M, Calderoni A, Ferrari G and Sampietro M 2007 Microwave irradiation effects on random telegraph signal in a MOSFET *Phys. Lett. A* **370** 491–493

[16] Prati E, Fanciulli M 2008 Manipulation of localized charge states in n-MOSFETs with microwave irradiation *Phys. Lett. A* **372** 3102–3104

[17] Prati E et al 2010 Measuring the Temperature of a Mesoscopic Electron System by means of Single Electron Statistics *Applied Physics Letters* **96** 113109

[18] Prati E et al 2016 Band transport across a chain of dopant sites in silicon over micron distances and high temperatures *Scientific Reports* **6** 19704

[19] Lu J et al 2015 A 1 TOPS/W analog Deep machine-learning engine with floating-gate storage in 0.13 μm CMOS *IEEE J. of Solid-State Circ.* **50** 1 1-12

[20] Ivakhnenko A G and Lapa V G 1965 *Cybernetic Predicting Devices* CCM Information Corporation

[21] Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory *Neural Computation* **9** 8 1735–1780

[22] Sun R and Sessions C 2000 Self-segmentation of sequences: Automatic formation of hierarchies of sequential behaviors *IEEE Transactions on Systems, Man, and Cybernetics: Part B Cybernetics* **30** 3 403–418

[23] Volodymyr M et al 2015 Human-level control through deep reinforcement learning *Nature* **518** 529533

[24] Tesauro G 1994 TD-gammon, a self-teaching backgammon program, achieves master-level play *Neural Computation* **6** 2 215–219

[25] Mnih V et al. Playing Atari with deep reinforcement learning 2013 *Deepmind Technologies Technical Report* arXiv:1312.5602

[26] Ballard D H 1987 Modular learning in neural networks *Proc. AAAI* 270-284

[27] Hinton G and Salakhutdinov R 2006 Reducing the dimensionality of data with neural networks *Science* **313** 504-507

[28] Schmidhuber J 1992 Learning complex, extended sequences using the principle of history compression *Neural Computation* **4** 2 234-242

[29] Schmidhuber J 1995 Deep Learning in Neural Networks: An Overview *Neural Networks* **61** 85–117

[30] Barends R et al. 2016 Digitized adiabatic quantum computing with a superconducting circuit *Nature* **534** 222–226

[31] Rotta D et al. 2016 Maximum density of quantum information in a scalable CMOS implementation of the hybrid qubit architecture *Quantum Information Processing* **15** 6 2253–2274

[32] Rotta D, Sebastiano D, Charbon E and Prati E 2017 *npj Quantum Information* arXiv:1704.06365

[33] Kinjo M, Sato S, Nakamiya Y and Nakajima K (2005) Neuromorphic quantum computation with energy dissipation *Physical Review A* **72** 5 052328

[34] Wiebe N, Kapoor A and Svore K M 2014 Quantum deep learning arXiv:1412.3489

[35] Risté D et al. 2017 Demonstration of quantum advantage in machine learning *npj Quantum Information* **3** 16

[36] Steane A M 1996 Error correcting codes in quantum theory *Phys. Rev. Lett.* **77** 793

[37] Fowler A G et al. 2012 Surface codes: towards large scale quantum computation *Phys. Rev. A* **86** 032324