

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/2267350>

Quantum-inspired Neural Networks

Article · June 1995

Source: CiteSeer

CITATIONS

106

READS

2,232

2 authors:



[Tamaryn Menneer](#)

University of Exeter

95 PUBLICATIONS 1,307 CITATIONS

SEE PROFILE



[Ajit Narayanan](#)

Auckland University of Technology

200 PUBLICATIONS 2,850 CITATIONS

SEE PROFILE

Quantum-inspired Neural Networks

Tammy Menneer and Ajit Narayanan

Department of Computer Science

University of Exeter

Exeter EX4 4PT

United Kingdom

{tsimenne,ajit}@dcs.exeter.ac.uk

(Submitted to NIPS 95, Denver, Colorado, December 1995.)

Abstract

Quantum computing (physically-based computation founded on quantum-theoretic concepts) is gaining prominence because of recent claims for its massively increased computational efficiency, its potential for bridging brain and mind, and its increasing relevance as computer technology develops into nanotechnology. Its impact on neural information processing has so far been minimal. This paper introduces some basic concepts inspired by quantum theory for use in neural network training and identifies a method inspired by the ‘many universes’ interpretation of quantum behaviour which promises greater efficiency and perhaps solvability of problems currently not amenable to a neural network approach. Instead of one network being trained on many patterns, many single layer networks are trained on one pattern each. The weights of each network are used to derive a quantum network, the weights of which are calculated as a superposition of individual network weights. Two microfeature tasks are used for network training. Each quantum network is compared with a standard control network. This shows that, on average, the quantum networks learn faster than the controls and give an equal number of errors, or less, in testing.

1 Introduction

Interest in quantum computing (QC) — physically based computation founded on quantum-theoretic concepts — has grown recently in the computer and cognitive sciences as a result of claims by Deutsch (1985, 1989) and Shor (1994) that problems regarded by computer scientists as NP-hard or NP-complete can be ‘solved’ by a quantum computer. Since it is regarded that if a computational solution can be found to one of the problems in the NP-complete class then a solution can be found to all problems in this class, these claims raise deep questions for computer scientists as to the nature of computational and algorithmic processes as well as the relationship between computation and physical process. Also, Penrose’s (1994) recent claims that solving the quantum measurement problem is a prerequisite for understanding the mind, and Hameroff’s (1994) proposal that consciousness emerges as a macroscopic quantum state from a critical level of coherence of quantum-level events in and around cytoskeletal microtubules within neurons, raise important questions for the scientific foundations of cognitive science and the appropriate level for a computational account of mind/brain (Narayanan, 1995). Moreover, as computer technology develops into nanotechnology, there will be a need for appropriate quantum-level descriptions of the behaviour of nanocircuits and quantum logic gates (Deutsch, 1989; Deutsch *et al.*, 1994; Barenco *et al.*, 1994).

At the heart of QC lies the quantum view of atomic and subatomic particles (e.g. atoms, electrons, neutrons, protons, photons, bosons, fermions). An electron circling a nucleus jumps states in discrete quanta (up to six) by absorbing energy or releasing it. The orbits of electrons are described by standing waves: unlike a planet orbiting the sun in a flat ellipse, an electron’s stable orbit takes the form of a wave. An electron at the first quantum level oscillates one full cycle before returning to its start position, at the second level two full cycles, and so on. Energy is not produced or absorbed during an orbit but only on jumping orbits: increased oscillation requires increased energy, reduced oscillation produces energy. QC is founded principally on two fundamental principles of quantum theory. Heisenberg’s uncertainty principle states that *both* the position *and* the momentum of an electron cannot be known (see Haken and Wolf, 1980). This is because in the act of measuring, some energy must be focused on the particle (e.g. a photon from a laser). The energy of the photon can be absorbed by the measured particle, sending it into a different orbit. Hence, if the measurement shows where the particle is (and the more precise the measurement the more focused the measuring photon must be), in this very process the momentum of the particle can be altered. A quantum particle’s *location* can best be described by a *quantum* state vector $|\Psi\rangle$, a weighted sum which in the case of two possible locations equals $w|A\rangle + x|B\rangle$, where w and x are complex number weighting factors of the particle being in locations A and B, respectively, and where $w|A\rangle$ and $x|B\rangle$ are themselves state vectors (Penrose, 1994). For n possible states, there will be n different complex number weighting factors, each describing the weighted probability of the particle being at that location.¹ In this case, the particle is said to be at A and B at the same time: it is in a *linear superposition*. However, in the act of observing a quantum state (or wave function), it collapses to a single state.

¹The weighting factors are used in ratios to calculate the probability of the particle being at A compared with that of it being at B. They are not the probabilities themselves.

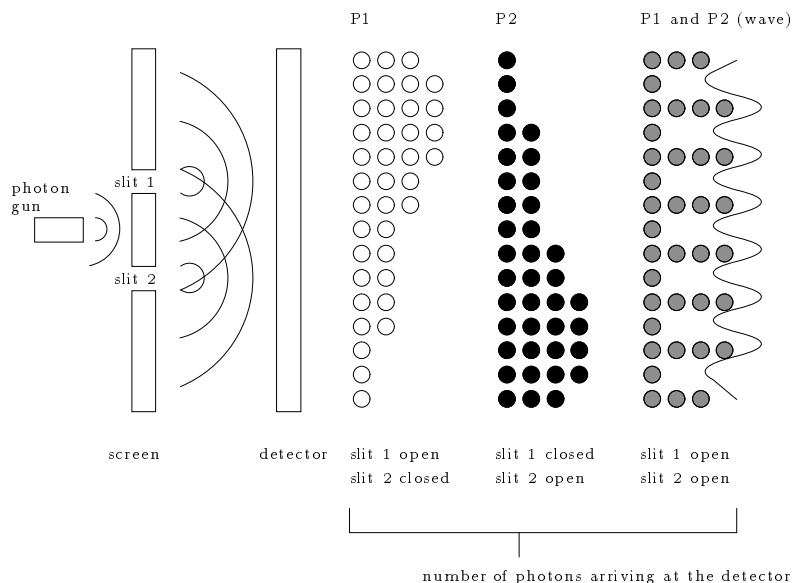


Figure 1: A diagram of Young's double-slit experiment (adapted from Hey and Walters, 1987). Photons are fired individually from a 'photon gun', through a screen with two slits, onto a detector. The pattern of photons detected when only one slit is open follows a probability distribution expected of discrete particles (white and black dots for slit 1 being open and slit 2 being open, respectively). However, when both slits are open, the pattern obtained follows a typical interference pattern obtained with waves and is not the sum of P1 and P2. Since it cannot be determined which slit a photon goes through in this case, the photons are shaded grey.

The second fundamental principle concerns a particular interpretation of Thomas Young's famous 'double-slit' experiment in 1801. Imagine that a photon is fired at a screen which has two parallel slits and behind which there is a detector. If only one slit is open, the pattern of photons arriving at the detector is as expected if photons are discrete particles. If both slits are open, however, the pattern of arriving photons takes a wave pattern (Figure 1), resulting in alternating bright and dark regions. The interpretation of the wave pattern is that one part of the wave pattern of the photon went through one slit and the other part through the other slit, and the location of the photon on the detector is determined by how and where the two waves interfere with each other when they reach the detector. The path differences between the slits and the detector result in constructive and destructive interference. In this case, the photon as wave has gone through both slits. Problems of interpretation arise when objects at the classical level (e.g. snooker balls, tables, cats) are subsumed under the laws of quantum superposition. A snooker ball does not appear to us in a superposed state but has a unique location and, when hit, momentum, both of which can be precisely measured. The 'many universes' principle is that, just as a photon can pass through two slits at the same time, a classical object exists in many different worlds at the same time, where each world represents one of the

states as described by a state vector for that object. This interpretation is used by Shor in his QC algorithm for extracting prime factors of very large integers, where a memory register is placed in a superposition of all possible integers it can contain, followed by a different calculation being performed in each ‘universe’ (path after slit). The computation halts when the different universes ‘interfere’ with each other because repeating sequences of integers (standing waves) are found in each universe. Although there is no guarantee that the results are correct, a subsequent check can be made at this point to identify whether the numbers returned are indeed prime factors of the given large integer. According to Shor (1994), a QC solution to the prime factors problem for large integers can be found in a few seconds, whereas earlier in 1994 1600 computers linked via Internet took 8 months to identify the factors of the 129 digit number *RSA-129*.

2 QUINNs

The application of quantum theory to connectionism is quite recent (e.g. Kak, 1995; Pylkko, 1995) and has been at the theoretical level. This paper lays the foundations of a practical approach by initially identifying the practical benefits of adopting quantum concepts in neural network design, development and implementation. Training a standard neural network entails repeated exposure to a set of training patterns until the network achieves an appropriate output for each pattern. A neural network approach *strongly* inspired by quantum concepts regards each pattern in the training set as a particle which is processed by a number of distinct neural networks (probably homogeneous but not necessarily so) in different universes. This is analogous to an electron or photon passing through many slits simultaneously. A *weakly* inspired approach regards each pattern in the training set as a particle which is processed in its own universe and no other. This paper describes a weakly inspired approach, where a number of homogeneous networks will each be trained on just one training pattern, with there being as many networks as training patterns. Each network and its associated training pattern exist in a separate universe. Once each network has been successfully trained in its own universe, a quantum superposition of these networks is calculated, resulting in a quantum-inspired neural network (QUINN) which has generalised across all input patterns. The resulting superpositional weight vector is the quantum-inspired wave-function (QUIWF) for the QUINN which ‘collapses’ upon actual input (testing). The QUIWF collapses in different ways depending on the pattern and collapse method.

To test the feasibility of QUINNs and QUIWFs, a set of experiments were carried out, with traditional neural networks used as controls. Consider, for example, a set of 3 networks each having 4 input units, 1 output unit and 4 weights (one from each input unit to the output unit), labelled a1–d1, a2–d2 and a3–d3 for networks 1, 2 and 3, respectively. Table 1 shows the input patterns the networks have been trained on and the labeled weights.

After training, the weights of individual networks are combined to give the superpositional weights (QUIWF) for the resulting QUINN. For example, the QUINN link from input unit 1 to the output unit (link 1) has a quantum weight derived from a ‘superpositional composition’ of a1 and a2 and a3. A simple probabilistic metric is used here to derive the QUIWF. Each input bit is a subparticle of the

	Weight values				Training input			
Net	1	2	3	4	1	2	3	4
1	a1	b1	c1	d1	1	1	1	1
2	a2	b2	c2	d2	1	0	1	0
3	a3	b3	c3	d2	0	0	0	1

Table 1: Weight values a1–d1, a2–d2 and a3–d3 for each of three networks, and the input pattern that each network is trained on. (There is only one output unit.)

overall pattern. The QUINN links are objects. The individual weights (e.g. a2) are states. When a subparticle (1/0) interacts with (hits) an object (QUINN link) it enters a superposition of states which depend on the object (QUINN link) and the subparticle (1/0). In this way when a ‘1’ interacts with QUINN link 1 it enters superpositional states a1 and a2. This is because the individual weight values a1 and a2 have evolved to process ‘1’, while a3 has not (see Table 1, columns 2 and 6). Therefore, when a ‘1’ is input to QUINN unit 1, the probabilistic metric associated with the QUIWF for QUINN link 1 returns positive for a1 and a2 and negative for a3, where positive (P) refers to acceptance of the particle into this state, and negative (N) refers to its rejection. For simplicity P is 1 and N is 0 in these preliminary experiments.

Generalisation is defined here as the acceptance of all (or most) of the input bits of a pattern in a single universe. In other words, the input is processed correctly by the universe of states in which all or most of the bits can exist. For example, say the test input to the QUINN is 1010:

- 1 is input to unit 1, giving the superposition Pa1 and Pa2 and Na3;
- 0 is input to unit 2, giving Nb1 and Pb2 and Pb3;
- 1 is input to unit 3, giving Pc1 and Pc2 and Nc3;
- 0 is input to unit 4, giving Nd1 and Pd2 and Nd3.

The only universe (original network) that has accepted all the input bits is the one with all P probabilities. This is universe (net) 2 of states {a2, b2, c2, d2}. The correct output can only occur in this universe. Therefore a collapse occurs from input bits in different universes to each bit in one state of the same universe. If the test input is 1001:

- 1 is input to unit 1, giving the superposition Pa1 and Pa2 and Na3;
- 0 is input to unit 2, giving Nb1 and Pb2 and Pb3;
- 0 is input to unit 3, giving Nc1 and Nc2 and Pc3;
- 1 is input to unit 4, giving Pd1 and Nd2 and Pd3.

Since universe (net) 3 accepts more of the inputs than the other two, this is the net chosen for processing the test pattern.

The superpositional composition method described above is inspired by a *fermion* view of particles and universes: their wave functions cannot overlap and when col-

lapse occurs a single existing universe (trained network) is the result. A *boson* view of particles inspires collapse methods where the wave functions of individual particles can overlap and their corresponding universes merge, resulting in a collapsed net which may not have existed independently during training. A boson-collapse method may need to use sophisticated probabilistic *ratios* in its description of a QUINN to identify which are the most suitable links from each of the universes to use in the merged, collapsed quantum network on presentation of a pattern.

3 Experimental method

The networks simulated were single layer feed forward networks. The learning algorithm used was the Delta rule (Rumelhart *et al*, 1986, p .53). Two tasks were used. Both required the input to be classified as living or non-living. For the first task (A) the input consisted of four microfeatures: *has_hair*, *has_legs*, *makes_noise*, and *moves* (set to ‘1’ for *yes* and ‘0’ for *no*). The output was ‘1’ for *living* and ‘0’ for *non-living*. Of the 16 potential patterns in this domain, 10 described real world information. The second task (B) was an extension of task A. The above microfeatures were used for input, with two additional features: *has_energy_source*, and *soft*. The output was ‘10’ for *living* and ‘01’ for *non-living* (two output units). Of the 64 potential patterns, 16 described real-world information. Only patterns describing real-world information were used in these experiments.

Each QUINN network was derived from individual networks each trained on one pattern in the training set. The control networks (similar architecture) were trained for all patterns in a training set, as normal. In both cases, subsets of the real-world patterns were selected as training sets and the test set consisted of all real-world patterns. Typically, the number of training patterns varied from 30% to 100% of the real-world patterns. Also, different subsets of the real-world patterns were generated for training, resulting in 14 training sets for task A and 20 for task B. Three different weight seeds (0.001, 2, 45 for task A and 10, 20, 30 for task B) were used for training. For each type of network (QUINN and control) there were therefore 42 networks for task A and 60 for task B.

Every network was tested on the corresponding full input pattern set (10 and 16, respectively). The number of patterns for which the network gave the wrong output was counted. The number of weight changes for a QUINN is the summed number of changes for each individually trained network of the QUINN.

	Task A		Task B	
	Control	QUINN	Control	QUINN
No. of weight changes	59	27	211	122
No. of wrong outputs	1.5	1.5	7.0	6.9

Table 2: Mean number of weight changes and errors for each network type.

The QUINNs performed well in both tasks. On average they required fewer weight changes than the control in training. In testing of task A the mean number of errors in testing was equal to the control. For task B it was slightly less (see Table 2).

The task B control network failed to learn for 5 of the 20 training sets. This was because they contained patterns that were not linearly separable. The results of the corresponding QUINNs have therefore been excluded from the averages given above.

4 Conclusions and further work

The results of these experiments are promising. The QUINNs for both tasks required fewer weight changes (about 50%) than the control networks. Both showed generalisation equal to that of the control. Therefore, the many universes view from quantum theory has provided a basis for quicker training in some cases with no generalisation loss, for single layer networks with the Delta learning rule.

This method has the potential for eliminating catastrophic forgetting. This is because one network is trained for each pattern and therefore no interference occurs between patterns during learning. It also provides single layer network solutions of linearly inseparable problems. This was shown by the fact that the control did not learn certain training sets, while the quantum network did.

Cognitively, QUINNs (weakly and strongly inspired) may help overcome one of the main objections to neural and connectionist networks being able to provide adequate accounts of mind: that the repeated presentation of patterns (in some cases several hundred epochs) has no basis in human learning. The idea of multiple universes each tailored for its own pattern which combine to form a superpositional representation of the domain may be cognitively attractive.

This paper presents promising preliminary research. There is therefore a need to continue and expand the proposal. It is proposed to extend the microfeature domain, since this was the most successful. Both the size of the pattern set and the length of the input will be increased. Catastrophic forgetting will also be examined by the addition of new and contradictory patterns after training. Also, the theory behind this research requires formalising. At present the process is weakly quantum inspired. It is intended to incorporate further motivation from quantum theory in areas such as the superpositional construction of the quantum network and interference. In this way a novel input will be processed by a network that uses merged and overlapping weights from different individual networks.

References

- Barenco, A., Deutsch, D., Ekert, A. and Jozsa, R. (1994). Logic gates for quantum circuits. Research Report, Clarendon Laboratory, Department of Physics, University of Oxford, OX1 3PU, UK.
- Brown, J. (1994). A quantum revolution for computing. *New Scientist*, Sept. 24th, 21–24.
- Deutsch, D. (1985). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London*, A400, 97–117.
- Deutsch, D. (1989). Quantum computational networks. *Proceedings of the Royal Society of London*, A425, 73–90.

Deutsch, D., Barenco, A. and Ekert, A. (1994) Universality in quantum computation. Research Report, Clarendon Laboratory, Department of Physics, University of Oxford, OX1 3PU, UK.

Deutsch, D. and Jozsa, R. (1992). Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London*, A439, 553–558.

Haken, H. and Wolf, H. C. (1980) *The Physics of Atoms and Quanta: Introduction to Experiments and Theory*. Third Edition. Translated by W. D. Brewer. Berlin: Springer-Verlag.

Hameroff, S. (1994) Quantum coherence in microtubules: A neural basis for emergent consciousness? *Journal of Consciousness Studies*, 1(1), 91–118.

Hey, T. and Walters, P. (1987) *The Quantum Universe*. Cambridge, UK: Cambridge University Press.

Kak, S. C. (1995). Quantum neural computation. Research Report, Louisiana State University Department of Electrical and Computer Engineering, Baton Rouge, LA 70803-5901, USA.

Narayanan, A. (1995). Biomolecular cognitive science. *Proceedings of the Foundations of Cognitive Science Workshop*, AISB95, Sheffield, UK. Available through ftp: atlas.ex.ac.uk, Research Report 325.

Penrose, R. (1994). *Shadows of The Mind: A Search for the Missing Science of Consciousness*. New York: Oxford University Press.

Pylkko, P. (1995). Indeterminacy and experience. In *Current Trends in Connectionism*, L. F. Niklasson and M. B. Bodén (Eds.), Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rumelhart, D. E., Hinton, G. E. and McClelland, J. L. (1986a). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1*. Cambridge, Massachusetts: The MIT Press, A Bradford Book.

Shor, P. W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. To appear in *The Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*.