

A Quantum Neural Net: with Applications to Materials Science

B. Igel'nik*, M. Tabib-Azar*, Y.-H. Pao*, and S. R. LeClair**

*Electrical Engineering and Computer Science Department
Case Western Reserve University, Cleveland, OH, USA

**Material Directorate, Wright Laboratory, Fairborn, OH, USA

ABSTRACT

In this article a new neural network architecture suitable for learning and generalization is discussed and developed. The architecture is inspired and modeled after quantum electronic devices and circuits where coherent electronic wavefunctions traveling through different parts of the circuit are combined together and result in interferences at detection nodes. These wavefunctions, represented by complex numbers, are implemented as complex weights in our neural net architecture to efficiently and accurately facilitate certain computations. Although similar to the radial basis function (RBF) net, our computational model called quantum net (QN) has demonstrated a considerable gain in performance and efficiency in number of applications compared to RBF net. Its better performance in classification tasks is explained by the cross-product terms in internal representation of its basis functions introduced parsimoniously. These cross-products are the results of interferences naturally occurring in coherent electronic systems. Although we primarily discuss the software implementation of QN on Von Neuman computers, its hardware implementation is also briefly discussed. A number of examples, solved using QN and other networks, are used to illustrate the desirable characteristics of QN.

INTRODUCTION

We explore a new computation method that we call "quantum network" (QN) that uses complex weights in a neural network lattice and can be constructed using quantum wires, dots, and other quantum electronic components. Using coherent electrons or other quantum particles to perform calculations, QN takes advantage of "interferences" that take place between these particles traveling through different paths in the circuit. These interferences are very well known in quantum mechanics and are basis of many physical phenomena such as conductivity of solids, tunnelings in quantum layers, optical properties of clusters, etc. Our QN is the subset of a more general class of computational models and techniques called quantum computers and quantum computing that has received a renewed interest in recent years [1]-[8].

The mathematical model, considered here, is an expansion in basis functions, that is, it connects the multivariate system input x and output y (which without loss of generality can be assumed univariate) by the equation

$$y \doteq f_{N,E_T}(x) = \sum_{n=1}^N a_n g_n(x, b_n),$$

where a_n, b_n are adjustable real parameters, g_n are called basis functions. The types and the number of basis functions determine the architecture of the model, its accuracy and efficiency. The basis functions usually have the same structure for any $n = 1, N$ and constitute a superposition of a simple multivariate function φ of x and b_n (internal function) and a univariate function g (external function), as shown in the equation

$$g_n(x, b_n) = g(\varphi(x, b_n)).$$

In neural networks and RBF networks the internal function φ is a linear sum of univariate functions

$$\varphi(x, b_n) = \sum_{i=1}^d w_n \psi_i(x_i, c_{ni}),$$

where $x = (x_1, \dots, x_d)$, $\psi_i(x_i, c_{ni}) = x_i - c_{ni}$ for neural networks, $\psi_i(x_i, c_{ni}) = (x_i - c_{ni})^2$ for RBF networks, and w_{ni}, c_{ni} are adjustable parameters. The QN architecture constitutes a model of the following form

$$y = a_0 + \sum_{n=1}^N a_n g[(x - d_n)^T \circ c_n \cdot (\bar{c}_n)^T \circ (x - d_n)], \tag{1}$$

where the parameters $a_n, n=0, \dots, N$ are real numbers, while the parameters $c_n, n=1, \dots, N$ are complex vectors, \bar{c}_n means complex conjugate of c_n , and \circ stands for inner product of two vectors. Unlike the neural, or RBF net, the internal representation in a basis function is not a weighted sum of univariate functions but constitutes a quadratic function of d variables with cross-product terms of special form. This form of internal representation, as we show in next two sections, has both "hardware" and "software" implementations.

MOTIVATION

We were intrigued by the ability of QN to solve some interesting problems, such as XOR problem, in a very simple and straightforward manner. While the neural or RBF networks with one nonlinear basis function cannot solve the XOR problem, the "quantum net" can do it with only one node. As will be shown below, this ability is due to the specific construction of internal representation of a basis function in quantum net. This fact suggests that in some applications QN will work better than neural or RBF networks, at least for classification tasks. Several examples confirm this guess. The number of parameters in a basis function of a 'quantum' net is roughly twice that of similar neural or RBF networks. However, savings in the number of basis functions may exceed losses in the number of parameters. Even relatively small decrease in the number of basis functions can compensate increase in the number of parameters since computational time grows as a quadratic function of the number of basis functions. Another rationale for developing the 'quantum net' is its universal approximation capability.

To illustrate the concept of quantum computing, we examine the well-known double-slit interference experiment shown in Figure 1.

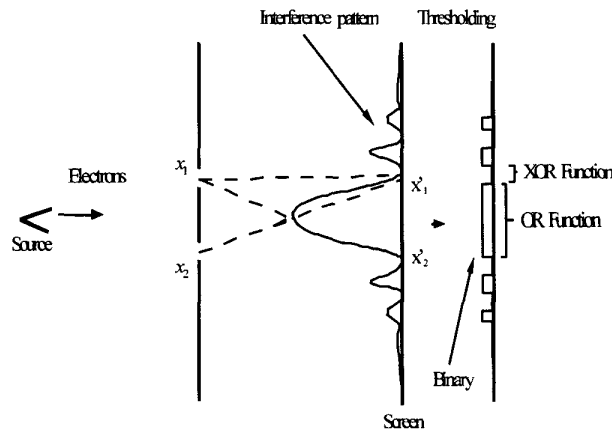


Fig. 1. Schematic of the double-slit interference experiment to illustrate the concept of quantum computing.

We note that the double-slit apparatus can be used to perform both OR and XOR functions depending on the location of the detector on the screen. To perform the OR function the light intensity on the screen is integrated from x_1 to x_2 . Hence, when only one of the slits is open there is enough light at the detector to

exceed the threshold for "1". If both slits are open, the light at the detector almost doubles and after thresholding, a logic one is detected. In the regions where the XOR function is performed, the light intensity is zero when both slits are open (due to destructive interference between lights coming through different slits) or when both of them are closed. The light intensity in this region is high when only one of the slits is open. By choosing an appropriate function for thresholding, this system can be made relatively immune to noise and variations to slit size and other parameters.

Quantum devices in particular and "wave-mechanical" devices in general, can be readily used to perform the following operations: attenuation and amplification, phase change, and detection. Mathematically, attenuation and amplification of an electronic wavefunction (ψ) amounts to multiplication by a real number (A, ψ) while changing its phase ($\psi e^{i\theta}$) is accomplished by multiplication of a complex number. Detection of the electronic wavefunction is accomplished by integration of the square modulus of the wavefunction over a region in space corresponding to the detector's active volume V :

$$\text{Probability of Detection} = \iiint_V \psi(v) * \bar{\psi}(v) dv.$$

When ψ is a slowly-varying function of position and for a unit active volume of the detector, the probability of detection is approximately proportional to the square modulus of the wavefunction:

$$\text{Probability of Detection} = \propto \alpha |\psi|^2.$$

XOR PROBLEM

The XOR problem is to find a curve $f(x, y) = t$ that separates the points $A(0,0), C(1,1)$ from the points $B(1,0)$ and $D(0,1)$ as shown in Figure 2. That means that there exists a real number t and a model $z = f(x, y)$ such that the points A and C are on the one side of the curve $f(x, y) = t$ and the points B and D are on another side of the curve. Obviously we can take $t = 0$. Then we can prove the following propositions [9]

Proposition 1.

Any net of the form

$$f(x, y) = g(\psi_1(x) + \psi_2(y)), \quad 2.$$

where g is a monotonic, fixed univariate function, ψ_1, ψ_2 are arbitrary fixed-shape univariate functions cannot solve XOR problem.

Proposition 2.

There exists a net f of the form (2) with fixed monotonic univariate function g and adaptive-shape differentiable functions ψ_1, ψ_2 , formed from polynomials, which solves XOR problem. Any such net should have at least 8 parameters.

Proposition 3.

There exists a quantum net with one node and 4 parameters which solves XOR problem.

Proof. The equation of separating curve for QN with one node is

$$(x - 0.5)^2 + 2(x - 0.5)(y - 0.5)\cos(\theta_2 - \theta_1) + (y - 0.5)^2 = a^2. \quad 3.$$

Transform the variables x and y to new variables u and v by the turn of coordinate axes on the angle $\pi/4$

$$\begin{aligned} x - 0.5 &= u\sqrt{2}/2 - v\sqrt{2}/2, \\ y - 0.5 &= u\sqrt{2}/2 + v\sqrt{2}/2. \end{aligned} \quad 4.$$

Substituting (4) into (3), one obtains

$$2u^2 \cos^2 \frac{\theta_2 - \theta_1}{2} + 2v^2 \sin^2 \frac{\theta_2 - \theta_1}{2} - a^2 = 0, \quad 5.$$

which is the equation of an ellipse with axes parallel to coordinate axes. Therefore, in coordinates x and y the equation (5) also constitutes an equation of ellipse with the angle between axes of the ellipse and coordinate axes equal $\pi/4$. This is shown in Figure 2.

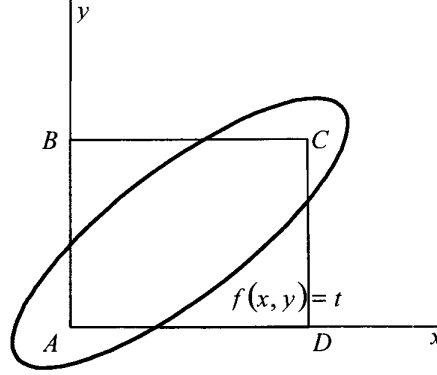


Fig. 2. Geometric illustration of the solution of the XOR problem by quantum net

Substitution of the coordinates of the points A , B , C , and D in the left-hand side of the equation (5) yields

$$\cos^2(\theta_2 - \theta_1) - a^2 > 0, \sin^2(\theta_2 - \theta_1) - a^2 < 0. \quad (6)$$

The simultaneous inequalities (6) are satisfied, for example if

$$a^2 = 0.5, 0 < \theta_2 - \theta_1 < \pi/4.$$

Therefore, there exists a quantum net with only one basis function, which solves the XOR problem. This net requires not more than 4 parameters if we make position of the center of the ellipse adjustable.

SOFTWARE IMPLEMENTATION

For practical consideration the quantum net should be written in the form

$$y = \tilde{f}(x_1, \dots, x_d) = a_0 + \sum_{n=1}^N a_n g \left[\sqrt{w^2 \sum_{j=1}^d e^{i\theta_{nj}} (x_j - c_{nj}) \sum_{j=1}^d e^{-i\theta_{nj}} (x_j - c_{nj})} \right],$$

where the model is in the coordinate form, all the parameters $a_0, a_n, c_{nj}, \theta_{nj}, w$ are real, w is the absolute value and θ_{nj} is the argument (phase) of a complex parameter. We assume that the input variables x_1, \dots, x_d are scaled so that $0 \leq x_j \leq 1$. For certainty we assume that the external function g is the Gaussian

$$g(t) = e^{-t^2/2}.$$

The values of the internal parameters are specified by the following inequalities

$$0 \leq c_{nj} \leq 1, 0 \leq \theta_{nj} \leq \pi - \Delta, \pi + \Delta \leq \theta_{nj} \leq 2\pi,$$

where $\Delta > 0$ is any number small compared to π . We divide all data that are available for learning, into two sets, the training set E_T and the generalization set E_G . The ensemble approach [10], [11] is used for training and testing. It constitutes a further development of ideas of the Functional-Link net [12]. The training set is used to adjust the parameters $a_0, a_n, c_{nj}, \theta_{nj}$ on the criteria of the minimal training error, while the testing set is used to determine the number of basis functions N . The parameter w can be adjusted manually. The learning of the model is made by the ensemble approach. The main features of this approach are as follows. The algorithm of learning is sequential. That means that only one node is learned at a time.

The learning starts with a simplest net of the form (node 0) $y = a_0$ and then grows net node by node. Next, consider the best net that can be chosen from the ensemble. This is done by the process of adaptive stochastic optimization. The whole ensemble of K possible choices of the parameters c_{nj}, θ_{nj} is divided in M groups each having L members so that $K = ML$. In first group the parameters c_{nj}, θ_{nj} are generated from the intervals $[0,1]$ and $[0.2\pi] - [\pi - \Delta, \pi + \Delta]$ uniformly. After the first group of the parameters c_{nj}, θ_{nj} have been chosen, the parameters a_0, a_1, \dots, a_n , net output, and the training error have been calculated, the net with the minimal training error is identified. The internal parameters $c_{njopt}, \theta_{njopt}$ of this optimal net are kept in memory and used to correct the distribution of the parameters in the groups 2, 3, ... M . For these groups, instead of the uniform distribution, we use the triangle distribution.

UNIVERSAL APPROXIMATION CAPABILITY OF THE QUANTUM NET

The universal approximation capability of the quantum net was proved [9] in the following theorem

Theorem. (The universal approximation capability of the quantum net).

Suppose the external function g satisfy the conditions

$$0 \neq \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left| g \left(\sum_{j=1}^d t_j^2 \right) \right| dt_1 \dots dt_d < \infty. \quad 7.$$

Define a distance between a function f , defined and continuous on I^d , and a quantum net f_N

$$f_N(x) = a_0 + \sum_{n=1}^N a_n g \left[w^2 \sum_{j=1}^d e^{i\theta_{nj}} (x_j - c_{nj}) \sum_{i=1}^d e^{-i\theta_{ij}} (x_j - c_{ij}) \right] \quad 8.$$

as

$$\rho(f, f_N) = \sqrt{E \int [f(x) - f_N(x)]^2 dx}. \quad 9.$$

Thus for any $\epsilon > 0$ and any function f , defined and continuous on I^d , there exists a quantum net f_N , such that

$$\rho(f, f_N) < \epsilon. \quad 10.$$

EXAMPLES OF APPLICATIONS IN MATERIALS SCIENCE

Example 1.

The comparison between quantum net and RBF net was made also on real data [13]. A body of data constitutes 6431 patterns of ternary systems (systems of 3 chemical elements) with 15 features of the elements in the system, 5 for each element. For each system it is known if it can or cannot form a compound. This information is available through long and expensive experimentation and lengthy calculations. The task is to build a neural net, which can accurately predict possible formation of a compound for a new system, not available in the database. The quantum net and the RBF net were applied to this data, both using the ensemble approach for learning and generalization. With RBF net the results are: 72 (95.5% correct) misclassifications in a testing set of 1607 patterns, using 44 nodes, with quantum net 60 misclassifications (96.2% correct) are obtained with 30 nodes and 49 misclassifications (97.0% correct) with 70 nodes. Our experiments have indicated that the activation function for the quantum net can be chosen from the same set of functions as for RBF net. In particular the minimum of generalization error is achieved with the "thin plate" activation function $f(t) = t^2 \log(t)$. These results are shown in Figure 3.

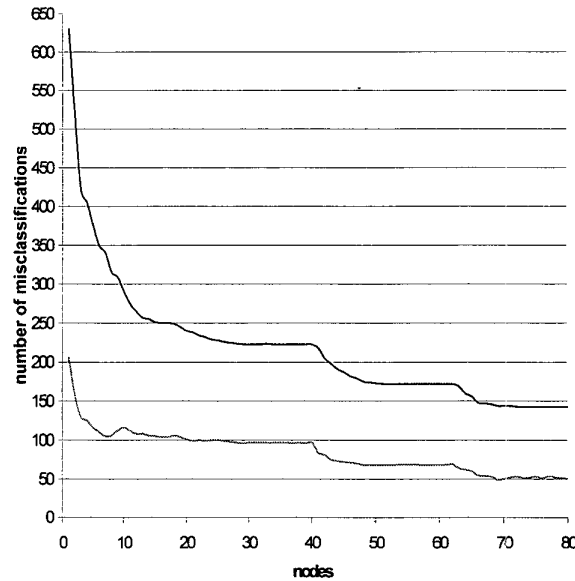


Fig. 3. Training error (upper) and testing error (lower) versus number of nodes for QN -- Villars' data.

Example 2.

A quantum net and an RBF net are compared in the task of building a cellular automata (CA) based model of thin film growth [14]. The atoms of types A and B are sent to the substrate by two heated sources. Those atoms which bond between each other or/and with the substrate, form a surface film. The geometric features of the surface, such as average roughness, are of great importance to the quality of the film. Depending on the current state of the surface and substrate, an incoming atom can form different types of bonding with the surface or remain as a vapor. For the current state of the model, 6 possible states of an atom are assumed. These are AA bond, AB bond, adsorbed, wall-adsorbed, cliff-adsorbed, and vapor. In the CA model it is supposed that the actual state of an atom depends not on the entire substrate and surface but only on the states of atoms in the neighborhood of the incoming atom. The neighborhood constitutes 26 cells that together with the incoming atom form a cube with the incoming atom at the center. Surrounding cells are filled by atoms of type A or B, or they are empty. The state of an incoming atom can be predicted given the state of the neighborhood, temperature and some probabilities calculated by using laws of statistical physics. It is impossible for such a model to operate in a reasonable time given that calculations must be carried out for millions of atoms. That is why a neural net is used. After training on a number of known examples, the net can predict the current state of the incoming atom.

In the current model, we used two discrete variables to characterize each of the neighborhood, the temperature, and 3 probabilities -- altogether 6 variables -- as inputs to the net, and one discrete output taking 6 possible values. The number of training patterns was 3208, and the number of testing patterns was 1069. Comparison of the RBF and quantum net is made in terms of the number of misclassifications of the output and the time required to predict the state of an incoming atom. Results are shown in Table 1.

Table 1: Comparison between RBF and quantum nets in terms of time and accuracy.

Type of net	# training patterns	# testing patterns	# misclassified patterns.	% correct test patterns	Time/ test pattern
RBF	3208	1069	70	93.4	14 μ sec
QN	3208	1069	60	94.4	12 μ sec

Example 3.

The data set consisted of 676 points describing dependency of the optical thickness of a thin film (output) on its spectral pattern (input) [15]. The input constituted a 33-dimensional vector. The output values were

uniformly distributed in the range $[0.5, 5.5]$ with the average value equal to 3. Thus, 1% of error corresponds to 0.03 or 0.0009 MSE. Three quarters of the data (507 patterns) were used for training and one quarter (169 patterns) was used for testing. This is an example of learning a continuous function with a large number of variables. The results of training and testing for a quantum net are shown in Figure 4.

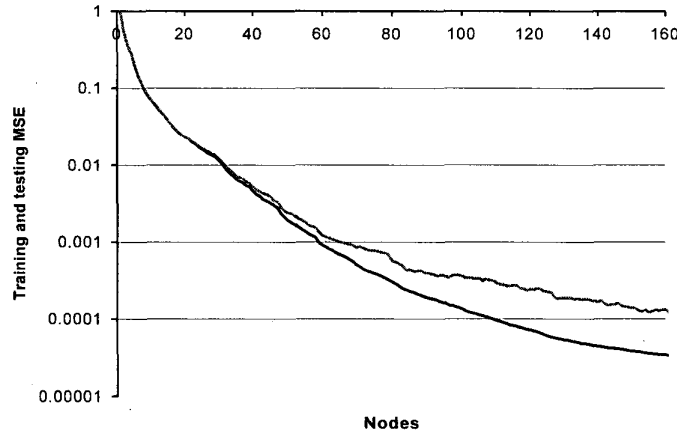


Fig. 4. Training MSE (lower) and testing MSE (upper) versus number of nodes for quantum net.

The level of 1% of testing error was achieved with a net of 68 nodes (0.000895 MSE), while the training error was 0.8% (0.000597 MSE). The best results were obtained with a net of 170 nodes: testing error 0.37% (0.000121 MSE), training error 0.16% (0.000031 MSE). The corresponding results for RBF net of the same size were : testing error 0.5% (0.000225 MSE), training error 0.27% (0.000063 MSE). These examples confirm that the quantum net has a visible advantage in accuracy and efficiency of learning and generalization compared with the RBF net. These advantages will become even more when the quantum computers will make calculations with complex numbers much faster than now.

HARDWARE IMPLEMENTATION

To demonstrate the practicality of the hardware implementation of QNs, we discuss a specific example below involving the XOR function. A QN capable of performing XOR is shown in Figure 5 where an Aharonov-Bohm ring is used with two FET switches. The magnetic field perpendicular to the ring is chosen so that when both the FETs are on (conducting) the current through the ring is zero (i.e., destructive interference case) and when either of the FETs are closed, the current is non-zero. It should be noted that this structure is very similar to the regular microwave waveguide structure (with angstrom dimensions) and that the reflection paths should be properly terminated /matched to prevent unwanted interferences.

Only the magnetic field is varied as the adjustable parameter to train this QN. The output is monitored while the magnetic field is varied to ensure zero output when both FETs are on. A small feedback circuitry can be devised to automate the training (very much like the Hopfield net). This hardware implementation is shown schematically in Figure 5.

CONCLUSION AND FUTURE WORK

The new architecture of neural network, suggested in this paper, has a solid motivation and has proved a considerable and measurable advantage over the RBF net in performance and efficiency in a number of applications. Our future work will be concentrated both on applications of this architecture, in particularly in the area of smart sensors, and on theoretical development of a new, completely adaptive architecture.

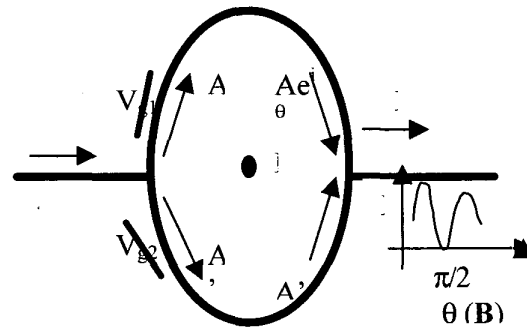


Fig. 5. An example of hardware implementation of QN using an Aharonov-Bohm ring. The ring, schematically shown here, is composed of a high-quality gold or AlGaAs/GaAs 2-D gas layer. The gates are used to raise or lower the potential barrier and control the current flow through a given arm. The magnetic field, \mathbf{B} , is perpendicular to the ring's plane and it causes the phase difference of 2θ between the electronic wavefunction flowing through the two arms. The value of θ depends on the strength of the magnetic field.

REFERENCES

1. A. Garcia and M. Tabib-Azar, 1995. Sensing Means and Sensor Shells: A New Method of Comparative Study of Piezoelectric, Piezoresistive, Electrostatic, Magnetic, and Optical Sensors. *Sensors and Actuators A. Physical*, 48(2), 87-100.
2. M. Tabib-Azar, 1998. *Microactuators; Electrical, Magnetic, Thermal, Optical, Mechanical, Chemical and Smart Structures*. Boston, MA: Kluwer Academic Publishers.
3. P.W. Shor, 1994. Algorithm for Quantum Computation: Discrete Logarithms and Factoring. In *Proceedings of 35th Annual IEEE Symposium on Foundations of Computer Science*, 124-137.
4. D. P. DiVincenzo, 1995. Two-bit gates are universal for quantum computation. *Phys. Rev. A*, 51(2), 1015-1022.
5. C. H. Bennett, 1973. The Fundamental Physical Limits of Computation. *IBM J. Res.Dev.*, 17, 525.
6. D. Deutsch and R. Jozsa, 1992. Rapid Solution of Problems by Quantum Computation. *Proc. Royal Soc., A* 439, 553-558.
7. D. Deutsch, 1985. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. Royal Soc., A* 400, 97-117.
8. S. Lloyd, 1993. A Potentially Realizable Quantum Computer. *Science*, 261(17) 1569-1571.
9. B. Igelnik, M. Tabib-Azar, S. R. LeClair,, 1999. Quantum net: a net with complex coefficients. Submitted to *IEEE Transactions in Neural Networks*.
10. B. Igelnik and Y.-H. Pao, 1995. Stochastic choice of basis functions and adaptive function approximation. *IEEE Transactions on Neural Networks*, 6(6), 1320-1329.
11. B. Igelnik, Y.-H. Pao, S. R. LeClair, and C.-Y. Shen, 1999. The ensemble approach to neural network learning and generalization. *IEEE Transactions on Neural Networks*, 10(1), 19-30.
12. Pao, Y-H, 1989. *Adaptive pattern recognition and neural networks*. Reading, MA: Addison-Wesley.
13. P. Villars. 1998. Private communication.
14. A. Jackson, M. Benedict, 1997. Private Communication.
15. S. Fairchild, 1998. Private Communication.