J. Barhen, J. Wells, V. Protopopescu, Y. Braiman, "Neuromorphic Pattern Recognition Using Arrays of Quantum Dots," published in the *Proceedings of Fourth World Multiconference on Systemics, Cybernetics and Informatics (SCI 2000)*, Orlando, FL, July 23-26, 2000, pp.257-262, Ed. Nagib C. Callaos.

# *Neuromorphic Pattern Recognition Using Arrays of Quantum Dots*

J. Barhen, J. Wells, V. Protopopescu, Y. Braiman

***Center for Engineering Science Advanced Research***
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6355
barhenj@ornl.gov

**Abstract.** Revolutionary advances in computing, communication, detection, and sensing using nanoscale devices subsume a profound understanding of the complex dynamics of small arrays of quantum structures. Such arrays produce bi-stable and multi-stable robust behavior, which can be harnessed for unconventional, yet powerful computational concepts. In this paper, we propose a novel approach to signal pattern analysis using an array of quantum dots (QD). Our methodology combines an ultrafast neuromorphic learning algorithm with photon-assisted tunneling in the QD array. The latter enables emulation of the plasticity of neural synapses.

## Introduction

Recent developments in nanoscale science and technology have opened exciting opportunities for revolutionary advances in nanoscale computing, communication, detection, and sensing. To fully exploit this emerging potential subsumes a profound understanding of the complex dynamics and properties of small arrays of quantum structures, including quantum dots (QD), ultrasmall Josephson junctions, quantum-dot lasers, and others. Such arrays produce robust bi-stable and multi-stable behavior, which can be harnessed for unconventional, yet powerful computational concepts, and in particular for neuromorphic computing.

Obtaining lattice architectures of sufficient size and regularity to perform actual computations at room temperature has been a formidable challenge, unsurmounted by the scientific community to date. To address this challenge, we have been exploring innovative biophysical assembly techniques to direct appropriately derivatized and passivated gold clusters (i.e., "QD") to preselected locations along stretched strands of engineered DNA sequences. Our goal is to produce an operational device by controlling the placement of the clusters with subnanometer precision. This will enable the electron-transport properties of the array to be engineered precisely through the size and placement of the gold clusters. Our progress, to date, has included the actual synthesis of quantum dots approximately 2 nm in size, the engineering of one-dimensional DNA templates, and elements of device simulation. This progress is discussed in a companion paper [1].

The purpose of this paper is to address the issue of complex information processing by a quantum dot array. In particular, we are interested in demonstrating a capability for pattern classification using neuromorphic algorithms.

## Single-Electron Tunneling in Quantum-Dot Arrays

Consider a one-dimensional array of $N$ tunnel junctions constructed from metallic source and drain electrodes weakly coupled to a linear array of $N$ - 1 metal clusters. We review the results of the "orthodox theory" of single-electron tunneling [1] to describe the charge transport through the array under small, but finite, bias voltage. (We are interested in nanometer-sized metal clusters in which the Coulomb blockade of conductance may be observed at room temperature, but for simplicity here we neglect any discreteness in the density of electronic states in the cluster resulting from their small size.) The vector $\vec{n}$ defines the state of our system, $\vec{n} \equiv (n_1, \cdots, n_i, \cdots, n_{N-1})$, where $n_i$ is the number of excess electrons accommodated by the $i^{th}$ quantum dot. The Gibbs free energy $E(\vec{n}, V)$ describing the electrostatic energy of the array of quantum dots and its interaction with the external voltage is

$$E(\vec{n}, V) = \frac{1}{2}\sum_{i=1}^{N-1} C_{ii} f_i^2 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=0, j\neq i}^{N} C_{ij}\left(f_i - f_j\right)^2 - V_s Q_s - V_d Q_d, \tag{1}$$

where $C_{ij} = C_{ji}$ is the mutual capacitance between conductors $i$ and $j$, and $f_i$ is the electrical potential of cluster $i$ measured with respect to the substrate. The source and drain electrodes are enumerated as $i = 0$ and $i = N$, respectively. The source potential is $f_0 = V_s = V/2$, and the drain potential is $f_N = V_d = -V/2$. $V$ is the transport voltage across the array. The charge on the source electrode is $Q_s = C_{01}(f_0 - f_1) + en_s$, and the charge on the drain electrode is $Q_d = C_{N-1,N}(f_N - f_{N-1}) + en_d$, where $n_s\,(n_d)$ is the number of electrons that has tunneled from the source (drain) electrode through the first (last) junction.

To determine the free energy, the potentials $\vec{f} = (f_1, \cdots, f_{N-1})$ must be determined from the static charge configuration. Using the charge conservation law, the total charge on island $i$ can be written in terms of the potentials,

$$q_i = C_{ii} f_i + \sum_{i=0, i\neq j}^{N} C_{ij}\left(f_i - f_j\right)^2 = en_i + q_{0,i}, \quad i = 1, \cdots, N-1. \tag{2}$$

The background charges $-e/2 < q_{0,i} < +e/2$ are due to incompletely screened charges in the environment of the islands. Equation (2) can be written in matrix form $\vec{Q} = \underline{C}\vec{f}$, where the generalized capacitance matrix elements are defined

$$\underline{C}_{ii} \equiv \sum_{j=0}^{N} C_{ij}; \quad \underline{C}_{ij} = -C_{ij}, \tag{3}$$

and the augmented charge vector is defined as $Q_i \equiv q_i + C_{i0} f_0 + C_{iN} f_N$. The generalized capacitance matrix can be inverted to obtain the potential distribution given any charge distribution. For convenience, we rewrite the free energy of the array using matrix notation as

$$E(\vec{n}, V) = \frac{1}{2}\vec{Q}^T \underline{C}^{-1}\vec{Q} - V_s Q_s - V_d Q_d. \tag{4}$$

In describing the electron transport through the array, we neglect here the effects of co-tunneling, and consider only single-electron tunneling between nearest neighbors in the array. That is, the final state $\vec{m}$ of the tunneling differs from the initial state $\vec{n}$ by the transfer of a single electron though the $k^{th}$ junction, e.g., $\vec{m} = \vec{n} \pm \Delta\hat{u}_k$, where $\Delta\hat{u}_k \equiv \hat{u}_k - \hat{u}_{k-1}$ and $\hat{u}_k$ is a unit vector for the $k^{th}$ quantum dot. The $\pm$ sign gives the direction of tunneling

through the junction. If the transition rates $\Gamma$ are sufficiently small, one can perform a calculation using Fermi's Golden Rule to obtain [1]

$$\Gamma_k^{\pm}(\vec{n},V) = \Gamma_k^{\pm}\left(\Delta E_k^{\pm}(\vec{n},V)\right) = \frac{-\Delta E_k^{\pm}}{e^2 R_k}\left[1-\exp\left(\frac{\Delta E_k^{\pm}}{k_B T}\right)\right]^{-1} \tag{5}$$

where $\Delta E_k^{\pm}(\vec{n},V) \equiv E(\vec{n}\pm\Delta u_k,V) - E(\vec{n},V)$ is the change in the free energy of the system due to the tunneling, $R_k$ is the effective resistance of the tunnel junction, $e>0$ is the fundamental unit of charge, and the thermal energy is $k_B T$. One can use probability conservation to write the corresponding master equation describing the time evolution of the probability $P(\vec{n},t)$ of finding the circuit in the state $\vec{n}$

$$\frac{dP(\vec{n},t)}{dt} = \sum_k \left[\Gamma_{k-1}^{+}(\vec{n}-\Delta\hat{u}_{k-1})P(\vec{n}-\Delta\hat{u}_{k-1},t) + \Gamma_k^{-}(\vec{n}+\Delta\hat{u}_k)P(\vec{n}+\Delta\hat{u}_k,t)\right] - \sum_{k=1}^{N}\left[\Gamma_k^{+}(\vec{n})+\Gamma_k^{-}(\vec{n})\right]P(\vec{n},t). \tag{6}$$

Practical approaches to solving the master equation are described in Refs. [3-4]. Given the solution, the average tunneling current is given by computing the net flow through any junction $k$ in the array:

$$I(V) = I_k(V) = e\sum_{\vec{n}} P(\vec{n})\left[\Gamma_k^{+}(\vec{n},V) - \Gamma_k^{-}(\vec{n},V)\right]. \tag{7}$$

Since the summation is performed over the charge states, the current is a function of the transport voltage.

## Neuromorphic Algorithms for Complex Information Processing

Quantum dot nanoelectronic devices represent a promising hardware technology that offers both conceptual opportunities and engineering challenges for complex information processing applications. One such application, namely pattern recognition, is of considerable interest to the development of modern intelligent systems and will be considered here. In recent years, the quest for innovative approaches to machine intelligence has received considerable attention. The proven ability of neuromorphic algorithms to deal with uncertain information and to interact with dynamic environments is therefore providing a strong incentive to explore the feasibility of their implementation on arrays of quantum dots. However, in contrast to conventional hardware approaches, we must develop here computational paradigms that exploit from the onset not only the concept of massive parallelism but also, and most importantly, the physics of the underlying device.

Artificial neural networks are adaptive systems that process information by means of their response to discrete or continuous input [5]. Neural networks can provide practical solutions to a variety of artificial intelligence problems, including pattern recognition [6], autonomous knowledge acquisition from observations of correlated activities [7], real-time control of complex systems [8], and fast adaptive optimization [9]. At the heart of such advances lies the development of efficient computational methodologies for "learning" [10]. The development of neural learning algorithms has generally been based upon the minimization of an energy-like neuromorphic error function or functional [11]. Gradient-based techniques have typically provided the main computational mechanism for carrying out the minimization process, often resulting in excessive training times for the large-scale networks needed to address real-life applications. Consequently, to date, considerable efforts have been devoted to: (1) speeding up the rate of convergence [12-14] and (2) designing more efficient methodologies for deriving the gradients of these functions or functionals with respect to the parameters of the network [15,16]. The primary focus of such efforts has been on recurrent architectures. However, the use of

gradient methods presents challenges even for the less demanding multi-layer feed-forward architectures, which naturally occur in QD arrays. For instance, entrapment in local minima has remained one of the fundamental limitations of most currently available learning paradigms. The recent development of the innovative global optimization algorithm TRUST [17] has been suggested [18] as a promising new avenue for addressing such difficulties.

Roychowdhury and his collaborators were the first to propose the implementation of neural networks in terms of QD arrays [19]. In their *Gedankenexperiment* a generic array of nanometer-sized metallic islands would be deposited on a resonant tunneling diode. Furthermore, all islands would have a direct conductive/capacitive link to their nearest neighbors, established, for example via organic molecular wires. They considered both continuous and discrete charge networks. The latter are of particular interest here. The Roychowdhury team showed that the evolution of an initial charge distribution toward a stable final equilibrium distribution can be given a neuromorphic interpretation and that this property emerges purely as a result of the discreteness of the electronic charge [20]. There are several shortcomings in their proposal. First, they assumed that all inter-island capacitance could be modified arbitrarily, but offered no mechanism to achieve this essential property. Moreover, their architecture involved capacitive coupling between all islands, a "*floating*" plate, and a grounded plate. Tunneling was assumed to occur only between the islands and the floating plate, but not between islands. Thus, even though their paradigm would allow some elementary form of combinatorial optimization, it could not be used for neural learning needed in pattern recognition.

In the previous section we illustrated the underlying physical concepts of single-electron transport in arrays of quantum dots. As pointed out by Roychowdhury and coworkers [19-20], there is a profound similarity between the dynamics of neural networks and that of QD arrays. In the latter, the free energy of an array characterized by a charge distribution can be lowered in terms of tunneling events. For neural networks, on the other hand, Hopfield has shown that the stable states of the network are the local minima of a bounded Lyapunov function of the net's output parameterized by the synaptic interconnection weights. A careful analysis, however, reveals that this formal similarity is not adequate for implementing learning algorithms for pattern recognition. By comparing the leading terms of the free energy in Eq. (4), i.e., $\frac{1}{2}\vec{Q}^T \underline{C}^{-1}\vec{Q}$ and the Lyapunov function in a Hopfield network, i.e., $L(x,\underline{W}) = -\frac{1}{2} x^T \underline{W}\, x$, we see that the inverse of the augmented capacitance matrix would have to play the role of the synaptic matrix. However, the elements of $\underline{C}_{ij}$ are *fixed*, and cannot be modified. An alternative approach for controlling the dynamics of the system has to be found. In principle, one could manipulate the free energy of the array via capacitive gating of each of the quantum dots. However, for an array of quantum dots 1-to-2 nm in size, which is necessary for room temperature operation, we are not aware of technology capable of implementing such gating on a nanometer scale.

Studies of the dynamics of arrays of QD in the presence of the time-dependent excitation (e.g., RF signal [21,22]) reveal a rich structure of dynamical behaviors that offers a tremendous potential for performing the computation we need. In particular, a team led by Oosterkamp has recently made available an extensive survey of experiments and methods for photon-assisted tunneling in QD [23]. In the absence of a time-dependent field, current flows through a quantum dot via tunneling when an unoccupied internal energy state is aligned to the Fermi energy of the leads. However, as pointed out by Oosterkamp *et al* following seminal work by Likharev *et al.*, if a time-varying AC voltage $A_0 \cos(2pwt)$ is applied, inelastic tunnel events are induced when electrons exchange photons of energy $\hbar n$ with the oscillating field. Tien and Gordon first described theoretically this phenomenon of multiphoton-assisted tunneling [24]. A direct inclusion of this phenomena in a master equation that takes into account Coulomb blockade can be made by writing the

tunneling rate $\bar{\Gamma}$ through each barrier in the presence of an electromagnetic excitation in terms of the rates without the external AC field,

$$\bar{\Gamma}_k^{\pm}(\vec{n},V,A_0,n) = \sum_{a=-\infty}^{\infty} J_a\left(\frac{eA_0}{\hbar n}\right)\Gamma_k^{\pm}(\Delta E_k^{\pm} + a\hbar n) \tag{8}$$

where $J_a$ denotes the Bessel function of the first kind and $a$ denoting the number of photons exchanged. This generalized master equation is obtained by substituting the rates in Eq. (8) into Eq. (6). The current through this device becomes a function of the transport voltage $V$, and the amplitude $A_0$ and frequency ν of the AC field,

$$I(V,A_0,n) = e\sum_{\vec{n}} P(\vec{n})\left[\bar{\Gamma}_k^{+}(\vec{n},V,A_0,n) - \bar{\Gamma}_k^{-}(\vec{n},V,A_0,n)\right]. \tag{9}$$

We consider the transport voltage $V$ as the input variable and the current $I$ as the output function in designing the neuromorphic computation. For a two-dimensional QD array with $K$ input and $K$ output nodes, we can readily generalize the description given here to consider $K$ input voltages $V_k$ and $K$ output currents $I_k(V_k,A_0,n)$ (see Fig. 1). This vector-valued function $\mathbf{I}_K$ is controllable through the parameters of the external, alternating field, $A_0$ and ν, by minimizing the error function $E$, defined over the number of $L$ training patterns as the squared difference between the $l$-th observed current, ${}^l\mathbf{I}_K$ and the target currents, ${}^l\mathbf{I}^*_K$ (see below),

$$E \equiv \frac{1}{L}\sum_{l=1}^{L}\left[{}^l\mathbf{I}_K(\mathbf{V}_K,A_0,n) - {}^l\mathbf{I}_K^*\right]^2. \tag{10}$$

For convenience, matrix and vector dimensions are explicitly indicated as subscripts. If a larger number of controls is necessary, then a polychromatic AC field may be considered for the global control, rather than a monochromatic field.

The pattern recognition scheme can now readily be implemented using the following method. We will assume that two sets of $L$ vectors are used for training. They are stored as rows of two matrices $\mathbf{\Omega}_{LI}$ and $\mathbf{R}_{LO}$ respectively, which represent the input signal patterns and the target outputs. We denote the number of columns of each matrix as $I$ for input and $O$ for output, without confusion. Since typically $L >> I$, two preprocessing steps are used [25]. First, appropriate clustering transforms $\mathbf{\Omega}_{LI}$ and $\mathbf{R}_{LO}$ into smaller size matrices $\mathbf{\Omega}^*_{KI}$ and $\mathbf{R}^*_{KO}$. Then, two successive *nonlinear* transformations map $\mathbf{\Omega}^*_{KI}$ into $\mathbf{H}_{KK}$, a nonsingular $K$ x $K$ presynaptic matrix, which constitutes the actual input into the quantum-dot array. We also decouple the nonlinearity of the transfer function, φ, at the output layer of the neural net from the linear interlayer pattern propagation mediated by the synaptic weights $\mathbf{W}_{KO}$ . This transformation nominally computes the postsynaptic input to the output layer of the neural net as a $K$ x $O$ rectangular matrix. Since the latter is connected via a bijective sigmoid mapping to the output training examples, the synaptic interconnection matrices $\mathbf{W}_{KO}$ can be determined by solving the linear system $\mathbf{H}_{KK}\,\mathbf{W}_{KO} = \varphi^{-1}(\mathbf{R}^*_{KO})$ using gradient iteration. In simulation on a conventional computer, this can be accomplished by exactly solving the system of linear equations using singular-value decomposition techniques. On nanoelectronic hardware, the same result canl be achieved by directly optimizing the error function in Eq. (10), where the sum is now running from 1 to $K$ clusters, in terms of the parameters of the external field, $A_0$ and ν. If the dimension $O$

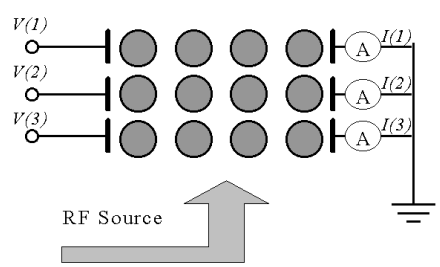


**Figure 1.** Two-dimensional QD array showing voltages as input channels and currents as outputs.

of the output pattern is smaller than the number $K$ of output nodes of the quantum-dot array, the output error is calculated using the Euclidean distance in $O$ dimensions.

Future efforts in this area will focus on directly accounting for uncertainties when attempting to obtain best estimates for the device parameters and responses of interest. For example, nominal values for the elements of the capacitance matrix will be computed from "first-principles" simulations of the metal clusters and substrate via density-functional-theory-based molecular dynamics [1], and the current through the device will be computed via numerical solutions to the master equation. To obtain best estimates for critical parameters (e.g., $A_0$ and $\nu$), we must consistently combine computational results and experimental measurements. We will achieve this by optimizing a generalized Bayesian loss function that simultaneously minimizes the differences between the best estimate responses and the measured responses on one hand, and the best estimate and calculated parameters on the other hand. Our optimization process will use the inverse of a generalized total covariance matrix as the natural metric of the calculational manifold in conjunction with response sensitivities to all parameters [26].

## References


1. J. WELLS, L. MAYA, K. STEVENSON, M. GOVINDARAJAN, T. THUNDAT, J. BARHEN, D. DEAN, and M. STRAYER, "DNA Templates for Directed Assembly of Computing Nanostructures", *Procs., SCI-2000* (in press, 2000).
2. K.K. LIKHAREV, "Single Electron Devices and their Applications", *Procs. IEEE*, **87**(4), 606-632 (1999).
3. L.R.C. FONSECA, A.N. KOROTKOV, K.K. LIKHAREV, and A.A. ODINTSOV, "A Numerical Study of the Dynamics and Statistics of Single-Electron Systems", *J. Appl. Phys.* **78**, 3238 (1995).
4. A.S. CORDAN and A. GOLTZENE., "Optimization of the Master Equation Set for a Multidot Tunnel Structure", *Eur. Phys. J AP* **7**, 137 (1999).
5. M. HASSOUN, *Fundamentals of Artificial Neural Networks*, MIT Press (1995).
6. M. BISHOP, *Neural Networks for Pattern Recognition*, Oxford University Press (1997).
7. M. BECKERMAN, *Adaptive Cooperative Systems*, Wiley Interscience (1997).
8. D. WHITE and D. SOFGE, *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, Van Nostrand Reinhold (1992).
9. A. CICHOCKI and R. UNBENHAUEN, *Neural Networks for Optimization and Signal Processing*, Wiley (1993).
10. P. MARS, J. CHEN, and R. NAMBIAR, *Learning Algorithms*, CRC Press (1996).
11. Y. CHAUVIN and D. RUMELHART, *Backpropagation: Theory, Architectures, and Applications*, Lawrence Erlbaum (1995).
12. J. BARHEN, S. GULATI, and M..ZAK, "Neural Learning of Constrained Nonlinear Transformations", *IEEE Computer*, **22**(6), 67-76 (1989).
13. N. TOOMARIAN and J. BARHEN, "Fast Temporal Neural Learning Using Teacher Forcing", *U.S. Patent No. 5,428,710* (March 28, 1995).
14. J. BARHEN, R. COGSWELL, and V. PROTOPOPESCU, "Single Iteration Training Algorithm for Multi-layer Feed-forward Neural Networks," *Neural Proc. Lett*, **11**(2) (in press, 2000).
15. J. BARHEN, N. TOOMARIAN, and S. GULATI, "Applications of Adjoint Operators to Neural Networks", *Appl. Math. Lett*., **3**(3), 13-18 (1990).
16. N. TOOMARIAN and J. BARHEN, "Neural Network Training by Integration of Adjoint Systems of Equations Forward in Time", *U.S. Patent No.* 5,930,781 (July 27, 1999).
17. J. BARHEN, V. PROTOPOPESCU, and D. REISTER, "TRUST: A Deterministic Algorithm for Global Optimization", *Science*, **276**, 1094-1097 (1997).
18. A. SHEPHERD, *Second-Order Methods for Neural Networks*, Springer (1997).
19. V. ROYCHOWDHURY, D. JANES, S. BANDYOPADHYAY, and X. WANG, "Collective Computational Activity in Self-Assembled Arrays of Quantum Dots: Novel Neuromorphic Architectures for Nanoelectronics", *IEEE Trans. Electron. Devices* **42**(10), 1688-1698 (1996).
20. S. BANDYOPADHYAY and V. ROYCHOUDHURY, "Computational Paradigms in Nanoelectronics: Quantum Coupled Single Electron Logic and Neuromorphic Networks", *Jap. Jour. Applied Phys*., **35**(6), 3350-3362 (1996).
21. L. P. KOUWENHOVEN A. JOHNSON, N. VAN DER VAART, C. HARMANS, and C. FOXON, "Quantized Current in a Quantum-Dot Turnstile Using Oscillating Tunnel Barriers", *Phys. Rev. Lett*. **67**, 1626-1629 (1991).
22. M. N. WYBOURNE, L. CLARKE, M. YAN, S. CAI, L. BROWN, J. HUTCHISON, and J. KEANA, "Coulomb-Blockade Dominated Transport in Patterned Gold-Cluster Structures", *Jap. Jour. Appl. Phys*., **36**, 7796 (1997).
23. T. OOSTERKAMP, W. VAN DER WIEL, S. DeFRANCESCHI, C. HARMANS, and L. KOUWENHOVEN, "Photon Assisted Tunneling in Quantum Dots", *arXiv:cond-mat/9904359*, LANL (26 April 1999).
24. P.K. TIEN and J.P. GORDON, "Multiphoton Process Observed in the Interaction of Microwave Fields with the Tunneling between Superconductor Films," *Phys. Rev*. **129**, 647 (1963).
25. P. BALLONE and W. ANDREONI, "Density Functional Theory and Car-Parrinello Molecule Dynamics for Metal Clusters," in *Metal Clusters*, Ed. W. Ekardt, (Wiley, 1999).

26. J. BARHEN and V. PROTOPOPESCU, “Ultrafast Neural Network Training for Robot Learning from Uncertain Data”, in *Distributed Autonomous Robotic Systems*, L. Parker ed., Springer (in press, 2000).