Seventh International Conference on Control, Automation, Robotics And Vision (ICARCV'02), Dec 2002, Singapore

Quantum Dot Neural Network Neurons

R. W. Newcomb[1], C.-H. Yang[1], and A. M. Hodge[1,2]

[1]ECE Dept, UMCP, College Park, MD 20742 USA
[2]U.S. Naval Research Laboratory, Washington D.C. 20375-6800
Email:newcomb@eng.umd.edu

Abstract

Functional descriptions of neurons for neural networks constructed from single electron quantum dot transistors are presented.

Single electron transistors with quantum dots should make it possible to practically construct artificial neural networks with millions of cells on a chip. Here we discuss how quantum dots can serve as weights for neurons made from the transistors. The transistor characteristics are modeled by weight parameterized monotonic functions which can be obtained from practically realized transistors.

1 Introduction

In order for the number of neurons in transistorized realizations of an artificial neural network, ANN, to approach the number in the brain it will be necessary to practically make sub-micron sized neurons. Toward that goal the best option appears to be the single electron transistor, SET [1][2][3]. Here we discuss how the SET with quantum dots can be used to make neurons.

The SET has a number of advantages, these including extremely small size and very low voltage operation. There are also a number of disadvantages, such as noisy operation, and difficulty of characterization. In the following we develop an intuitive model which should allow for the design of artificial neural networks, ANNs, using available theories of ANNs.

By way of an interesting side point relevant to understanding what is being undertaken, the words "single electron" in the name of the SET refers to the transfer of a single electron at a time through a tunnel junction, where this transfer is controlled by various voltages. Even though only a single electron is transferred, this is done continuously so that one can still speak of a measurable current, which really is an average of current pulses arising from the passage of single electrons.

2 Single Electron Transistors

Figure 1a) shows a top view of a fabricated SET with quantum dots of the type in which we are interested with Fig. 1b) giving a diagramatic representation. Figure 1c) shows the circuit symbol we will use.

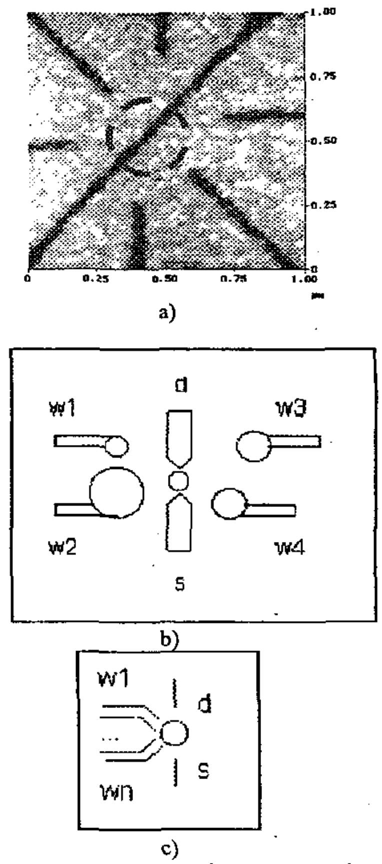


Figure 1. Top view of an SET serving as an artificial neuron

For the SET there is a center quantum dot into and out of which electrons tunnel through the tunnel junctions indicated by the pointed contacts of the source, s, and drain, d, in Fig. 1b) and the diagonal line in Fig. 1a). At a further distance from the center dot there are outer gates made of quantum dots labeled with W's since they serve to set the neuron weights, there being six in Fig. 1a), four in Fig. 1b) and n in Fig. 1c). To interpret the

SET as a neuron we will use the voltages on the outer gates as the synaptic inputs with the (synaptic) weights set by the sizes of the outer dots as well as their distance from the center dot. The output of the neuron is the drain current. To feed synapses of other neurons this current will need to be converted to a voltage which can be accomplished by feeding it into a resistor connected SET.

Although there are many fine points to be considered in the operation of an SET the main ideas can be gleaned by considering the dots as capacitive coupling to the tunnel junctions as shown in Fig. 2.

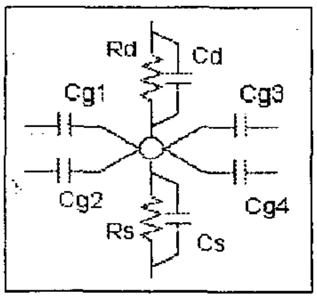


Figure 2. Equivalent Capacitance View of SET

To set up a simplified model we present the following arguments while noting that what follows is rather intuitive and still needs proper experimental verification.

For an electron to tunnel into the center dot it needs to acquire an energy to make a quantum jump, q^2/Cs, where q is the (absolute value) of the charge of an electron, q=160E(-21) Coulomb. This energy of an electron is obtained from the voltage on the center dot, which in turn comes from the voltages on the weight dots and the voltage at the drain, Vds (we will measure all voltages with respect to the source, the terminal with the lower potential). Starting at zero voltage there is no current until that voltage supplies enough energy to overcome the thermal energy, KbT, and thus, there is a Coulomb block of $q^2/(2Cs) >> KbT$, where Kb is Boltzman's constant, Kb=13.8E(-24), and T is the absolute temperature. The one half coming from the quantum jump passing through zero voltage being from negative to positive voltage. Since the voltage on the center dot is influenced by the outer dot gate capacitor voltages, there is control of the change of charge on the center dot by these gates. The actual current is determined by the voltage giving rise to a transfer of tunneling charge through the source capacitor, Cs, divided by the tunneling resistance, Rs. Similarly for a charge leaving via the drain capacitor (by the symmetry of Fig. 1a) we will assume Rd=Rs, Cd=Cs). As the drain voltage increases from zero initially there will be no current until the Coulomb block is overcome, then the current is constant with one electron at a time constantly moving in and out with the voltage jump being twice the Coulomb block voltage. After the voltage increases eventually two electrons move in and out, with the process continuing until the flow of electrons looks more on a macro scale than a micro scale. Figure 3 shows a typical I-V curve. The data for this curve was obtained from a fabricated quantum dot device as in Fig. 1 [4] simulated in MathCad using the model formulated in the following equations.

Figure 3 gives the drain current for an SET with three outer gates, the voltage of one of which is varied to give the three different curves, for Vgs1=0, 0.15v, 0.1v (for the right to left curves respectively). It is noted that the

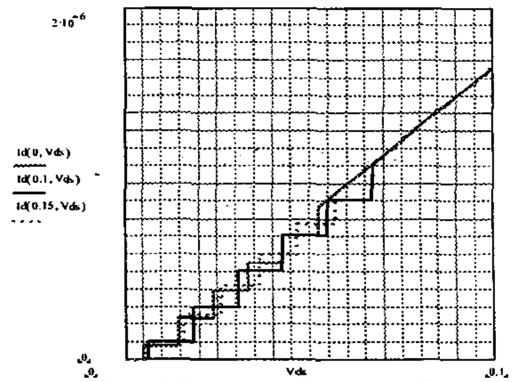


Figure 3. Typical SET curves for three different outer gate voltages

drain voltages and currents are very small, the outer limits being 0.1 volt, while the upper current is 1.7 micro-Amp, so that these devices should take very small power. It is also seen that the Coulomb block is dependent upon the outer gate voltages since they affect the main dot charge. This dependence is quite interesting in its influence by the charge coming from the outer gates since if this charge goes up by one electron that electron tunnels out. Further, that charge is a non-integer multiple of an electron's charge; it increases up to one-half an electron and then decreases symmetrically up to one electron with periodic repetition [2, p. 341]. This periodicity appears to be a rather complicating factor as far as neuron usage is concerned. But if the gate (synapse) voltages are also kept low the behavior can be kept single-valued.

In order to see that the outer gates can act as weights we normalize all the capacitors to the capacitance of the inner dot, Cdot=Cs+Cd=2Cs, taking the kth outer dot capacitance to be Cgk=Wk.Cdot with Wk the kth weight. That is, the kth outer dot (gate) input (synapse) voltage, Vgsk, has influence on the inner dot charge via

$$Qk = [(q*Cdot)Wk]Vgsk \qquad (1)$$

However, the amount of this charge which affects the drain current is a bit different, since it is a fixed amount for which only its excess electron over that coming in from the source contributes to charge motion [2]. As an electron entering from the source only goes out when an electron leaves via the drain, the placement of one electron from the source is on the average equivalent to one-half an electron as far as current flow is concerned. This leads to the effective charge from n control weight gates to be given by the following equation [2, Fig. 3] {here 1(.) is the unit step function and mod(1) takes the fractional part of the expression preceding it}

$$\frac{Q_{Weffec}}{q} = m_{Weffec}(Vgs) =$$

$$\sum_{k=1}^{n} \{[CdotWkVgs\,k]mod(1).\,1(\frac{1}{2} - [CdotWkVgs\,k]mod(1)) + \quad (2)$$

$$(1 - [CdotWkVgs\,k]mod(1)).\,1([CdotWkVgs\,k]mod(1) - \frac{1}{2})\}$$

Noting that to overcome the energy of the Coulomb blockage, $q^2/(2C)=qV$, requires a voltage q/2C. For the remaining quantum jumps a voltage of q/C is needed. Thus, the voltage for the dot itself, needed for one electron to tunnel in and then out continuously is given by

$$Vqb(Vgs) = q\frac{1+\frac{1}{2}+m_{Weffec}(Vgs)}{2C_T} \qquad (3a)$$

where Vgs is the vector of synapse voltages. In this the total capacitance, C_T, is given by

$$C_T = Cs + Cd + \sum_{k=1}^{n} Cgk =$$

$$Cdot + \sum_{k=1}^{n} Cdot \cdot Wk = Cdot(1 + \sum_{k=1}^{n} Wk) \qquad (3b)$$

and the one-half is as explained above due to averaging a single electron onto the center dot which awaits an electron transfer out the drain. The current at the Coulomb block voltage is

$$Iqb(Vgs) = qb(Vgs)/2 \qquad (3c)$$

For m electrons to tunnel in and out we have each jump in voltage to be twice the Coulomb block, 2Vqb(Vgs). This continues for a maximum of K electron transfers, after which the current becomes linear with slope 1/(2Rs). Thus we get

$$Id(Vgs, Vds) = Iqb(Vgs)1(Vds - Vqb(Vgs)) +$$

$$2 * Iqb(Vgs)[\sum_{k=1}^{K} 1(Vds - (2k+1)Vqb(Vgs))] +$$

$$\frac{(Vds - (2K+1)Vqb(Vgs))1(Vds - (2K+1)Vqb(Vgs))}{2Rs} \qquad (4)$$

The resulting current is as seen in Fig. 3 which is for a center quantum dot of diameter 50 nanometers in Silicon with Cdot=0.023 femto-Farads, three synaptic dots with weight vector $W=[2, 1, 2]^T$, variable Vgs1 along with fixed Vgs2=Vgs3=0 Volts. What is notable in Fig. 3 is the very small operational voltages and currents, under 150 milli-Volts and 0.9 micro-Amps. Consequently, besides the small size these devices take very little biasing power. However, we do need to satisfy the constraint of $q^2/(2C)>>kBT$ which means that temperature needs to be low or the capacitance very small; to operate at room temperature nanometer scale capacitors become a necessity, something which is still in the future for all but highly experimental devices.

Figure 4 shows a 3-D plot of this drain current versus Vds, on the front axis, varying from 0 to 0.1v, and Vgs on the right axis varying from 0 to 0.3v. The periodicity in Vgs is seen and shows that these curves are globally non-monotonic in Vgs though locally they are monotonic.

In order to form combinations of these neurons we will need the inverse function. As seen by turning over Fig. 3, this is a set valued function which can be represented for most purposes also in terms of step functions. Using the initial current jump of (3c) and the current at the Kth jump as

$$IK(Vgs) = Iqb(Vgs)(1 + 2\sum_{k=0}^{K-1} 1) \qquad (5a)$$

we get

$$Id^{-1}(Vgs, Id) = Vd(Vgs, Id) = 2Rs[Iqb(Vgs) +$$

$$2Iqb(Vgs)(\sum_{k=0}^{K-1} 1(Id - (2k+1)Iqb(Vgs))$$

$$+ (Id - IK(Vgs))1(Id - IK(Vgs)))] \qquad (5b)$$

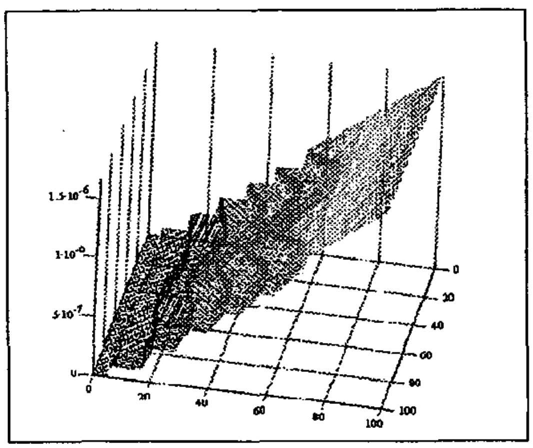
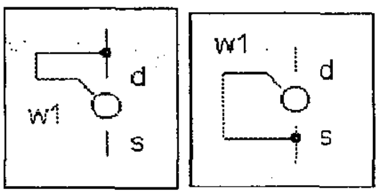


Figure 4. Id(Vgs,Vds) curves

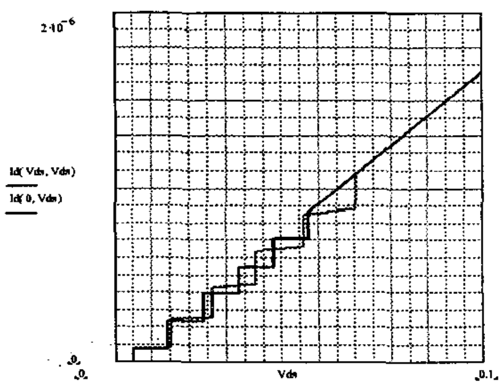
The layout of these devices is symmetrical in the drain and source so will work with Vds<0 as well, except that, since Vgs is measured with respect to the negative terminal of Vds, the reference change means that Vgd should be the gate control voltage when Vds<0. In such a case the curves are symmetrical to those so far discussed.

3 Single Electron Neurons

Having the SET in hand the means to make a neuron becomes clear. The external, synaptic, inputs are voltages into the outer (gate) quantum dots. The synaptic weights are set by the geometry of these gate dots; the larger the dot and the closer to the center dot the larger the weight since the capacitance is proportional to these two dimensions. Since it is the ratio of the outer dot capacitance to the center dot capacitance that determines the weight, it is relative dimensions which matter, something of advantage to a designer since absolute dimensions do not need to be held.



a)



b)

Figure 5. a)Two diode connected SETs for neuron loads b) I-V diode curves for SET of Fig. 3

The output of the SET itself is a current whereas in order to feed other neurons or take the signal to the outside we normally desire voltages. Such can be obtained by using another single outer gate SET as a load, with its gate connected to its drain as illustrated in Fig. 5a). As shown in Fig. 5b), the characteristics are seen to be resistor like so that when used as a load on the SET neuron, the output voltage will look like the (inverse of the) SET characteristics (as seen through (5b)) but saturating at the lower (ground) and upper bias voltages.

Figure 6 shows a typical SET neuron with n synaptic inputs on the left and one neuron output on the right. The activation function for this neuron peaks at about one-half of Vbias, as expected by voltage divider action, and as shown in Fig. 7. The jumps in the SET characteristic actually cancel to a large degree due to their appearance in the weight and the load transistor. The describing equation is given in (6). A solution of this nonlinear equation, found via point by point

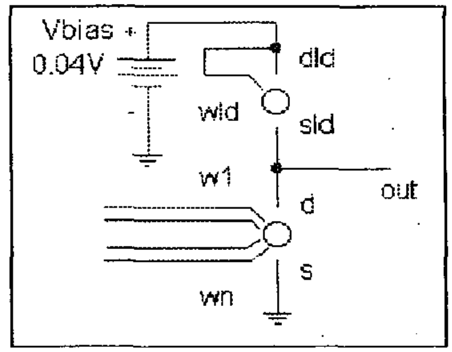


Figure 6. An SET neuron with n synaptic quantum dot weights

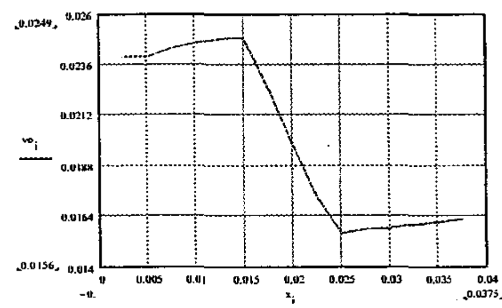


Figure 7. Vout vs Vgs, Vbias=0.04V

calculation in MathCad, is shown in Fig. 7.

$$v_{out} = Vbias - Id_{ld}^{-1}(Vbias - v_{out}, Id(Vgs, v_{out})) \qquad (6)$$

For this type of curve, design methods still need to be carried out. This will mean that standard neural networks with monotonic activation functions will need to be somewhat modified for design of ANNs using SET neurons. By using different ranges of synaptic voltages Figure 6 does show that "decreasing" or "increasing" activation functions can be obtained.

4 Discussion

In the current output form, shown in Fig. 5, these SETs are already neurons and open up a large number of open problems for research. Primary among these is determination of practical means of conversion to voltage; one method is presented in Figs. 7 and 8, which shows this can be done by using another SET as a load. Another difficulty is presented by the periodicity in the synaptic input voltages, as indicated by the use of the modulus function, mod(.), in Equation (1) for the central quantum dot voltage needed to transfer m electrons. Along similar lines is the fact that the describing equations need a better derivation method. Although the equations seem to be physically meaningful and in some agreement with experiments, they are somewhat cursory and need a more rigorous derivation. The simplified equations we have given do seem to be the most useful for analytic work. However real transistors exhibit somewhat smoother continuous curves, something which can be accommodated by replacement of the unit step functions in the equations. Also the number of jumps, K in the equations, prior to domination of the quantum effects, needs a theoretical determination. It should be noted that discontinuous curves with the quantum jumps should be useful for fuzzy neural networks as discussed in [5].

We have assumed that signals can be readily applied to and taken out of an SET neuron. However, the interconnects do pose a problem especially since they can upset the single electron transfer of an SET. Consequently, new means of interconnection must also be devised in which case each cell will probably need to interconnect to nearest neighbors; most likely there will be a maximum of four if cellular neural network configurations are used.

The beauty of the SET is that it improves as the scales get smaller, so it appears to be a very practical device for the future. As sizes decrease the temperature can come up to room temperature in which case the curves we have given will smooth out. Physically the jumps are sharp at zero degrees so the curves we have presented are meaningful in the limit as temperature decrease to zero.

The curves presented above are from MathCad simulations. However, the numbers used appear to reflect those obtained in fabricated devices made by the second author and his students.

References

[1]. T. J. Thornton, "Mesoscopic Devices - What are They?" in Quantum Transport in Ultrasmall Devices Edited by D. K. Ferry, H. L. Grubin, C. Jacoboni, and A.-P. Jaubo, Plenum Press, New York, 1995, pp. 141 - 169.

[2]. M. Amman and K. Mullen, " The Charge-effect Transistor," Journal of Applied Physics, Vol. 65, No. 1, January 1, 1989, pp. 338 - 346.

[3]. H. Grabert, "Single Charge Tunneling: a Brief Introduction," Zeitschrift fur Physik B, Condensed Matter, Vol. 85, No. 3, January 1991, pp. 319 - 325.

[4]. T. H. Chang, C. H. Yang, and M. J. Yang, "STM Observation of Single Electron Charging in 70 nm InAs Quantum Dots," Phys. Stat. Sol. (b) Vol. 224, No. 3, 2001, pp. 693 - 696.

[5]. K. W. Tang and S. Kak, "Fast Classification Networks for Signal Processing," Circuits, Systems, Signal Processing, Vol. 21, No. 2, 2002, pp. 207 - 224.