Queuing-theoretic simulation of single-electronic metal-semiconductor devices and systems

Sharif Babiker, John R. Barker and Asen Asenov

Nanoelectronics Research Centre
Department of Electronics and Electrical Engineering
University of Glasgow, Glasgow G12 8QQ, UK

Abstract
Traffic theoretic or queuing methods are proposed as a natural framework for modelling the correlated tunnelling and transport of single electron solitons in coupled tunnel capacitor structures below the Coulomb blockade threshold. Stationary state and time-dependendent modelling is developed and validated against Monte Carlosimulation. Exact results are obtained for the double tunnel junction or quantum dot. The effects of discrete energy states are evaluated and the extension to multiple junctions illustrated.

I. Introduction

The controlled transport/correlated-tunnelling of single-electronic excitations in coupled tunnelling capacitator structures is now experimentally established in metal-insulator, metal-semiconductor systems and in capacitatively-coupled quantum point contacts in semiconductor 2DEG structures[1-4]. These structures rely on the existence of ultra-small capacitative structures such that the effective charging energies $e^2/2C$ exceed the thermal energy k_BT [5]. By exploiting state-of-the-art nanofabrication it is possible to construct 20nm scale coupled capacitors (metal on semiconductor coupled Schottky dots)[6] which point the way to a future high temperature, high density nanoelectronic systems technology (Fig.1).

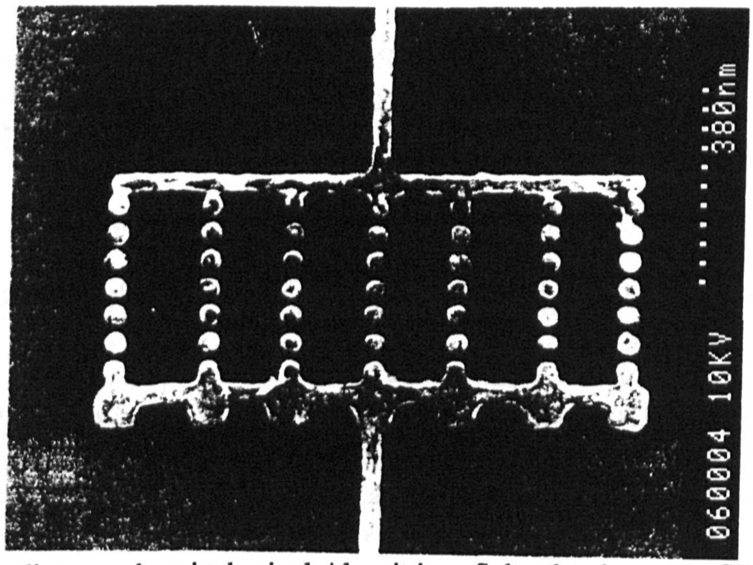
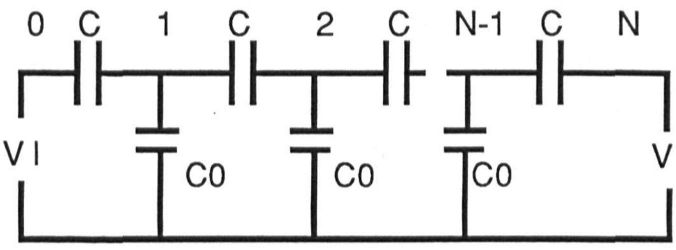


Figure 1. 40 nm diameter hemispherical Aluminium Schottky dots on p-Silicon with 15 nm spacing (Weaver[6]).

Single-electronic, ultra-small capacitor systems provide a new regime of study for semiconductor device modelling wherein the effects of charging become important and the Coulomb interaction provides subtle correlation effects.

II. Monte Carlo method

Simulations of arrays of single-electronic tunnel junctions (see figure 2) and gated arrays (figure 3) have been carried out to date by Monte Carlo methods which have proved to be prohibitive computationally for the extended systems of technological interest.



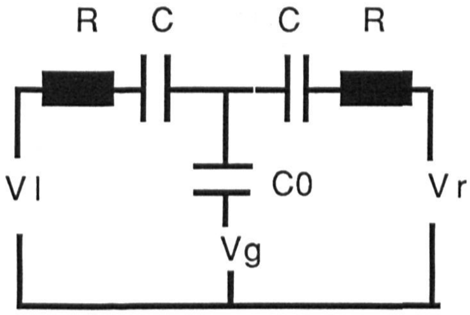
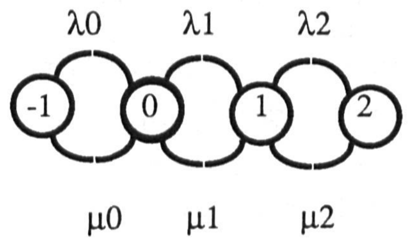


Figure 2. A multi-junction tunnelling array. Figure 3. Gated 2-junction array
The capacitances C are tunnel junctions. C_0 is taken to be non-tunnelling.

In the Monte Carlo approach[7,8] the state of the tunnel junction array is described by the number of solitons (electron plus polarisation field) at each node. The total charge on each junction/tunnel junction/ground capacitance is the sum of the charges due to the voltage sources and the charges induced by the soliton structure.These charges are linearly related to the voltages and number of solitons at each node. The evolution of the array is then determined by the stochastic process of single electron tunnelling which alters the soliton occupancy vector **n**.

III. Queuing theory: stationary case

In the present paper we introduce a new approach to modelling single-electronic systems that captures the Poisson stochastic nature of tunnel events and provides a fast, physically transparent and efficient method of calculating the steady-state characteristics of multi-junction configurations. This new method involves a re-formulation of the transport equations in terms of of *queuing theory* (traffic theory[9]) and centres on determining the distribution P_i of quasi-electrostatic soliton excitations that are formed during the transport/tunnelling process. For an N junction array the state of the system is described by $S_i = (k_1, k_2, \ldots k_N)$ where k_i are the number of excess electrons(solitons) at node k_i(figure 4 shows states for a 2-junction system).



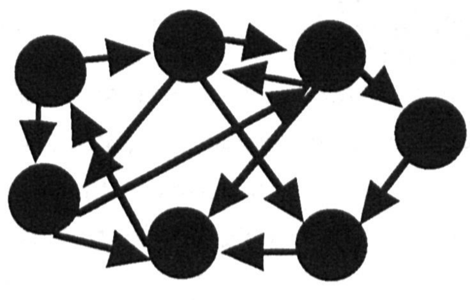


Figure 4. Soliton states in a 2-junction. Figure 5. Transitions of soliton states

An electron may tunnel from a node to a neighbouring node provided a tunnelling conection exists (connection matrix $X_{ij} = 1$). Electrons form a queue at each arrival node where they wait for "service" (unlike conventional traffic theory the queues are always full and the "service rate"

which is equal to μ_i the total departure rate from each node i varies as a function of the state of the system). The current passing between two neighbouring nodes (l,k) which are connected by a tunnel junction with capacitance C_{lk} is calculated from the P_i by :$I_{kl}=e.\sum_i P_i(\Gamma_i(k,l) - \Gamma_i(l,k))$ where P_i is the probability of finding the system in state S_i, and $\Gamma_i(k,l)=$ tunnelling rate from node k to neighbouring node l when the system is at state S_i.The average voltage across the capacitor C_{kl} can also be calculated as $v_{kl} = \sum_i P_i . q_{kl}(i)/C_{kl}$.
To determine the distribution of the soliton states P_i, consider a single tunnel event from node k to node l at state S_i to form state S_j. Define the departure rate $\mu_{ij} = \Gamma_{ni}(\kappa,\lambda)$ for $X_{kl} = 1$,from state S_i to Sj for the system already occupying S_i . The Poisson nature of the tunnelling events allows us to find the total departure rate from state S_i as:$\mu_i= \sum_j \mu_{ij}$ and the transition probability from state S_i to S_j is thus,$r_{ij} = \mu_{ij} / \mu_i$ defing the routing matrix **R** .The the average residence time (per visit) at state S_i is $1/\mu_i$. The input traffic to node i is given by $\lambda_l = \Sigma \lambda_l r_{ij}$. or in matrix form, the traffic erquations are: $\lambda.(\mathbf{I}-\mathbf{R}) = 0$ where λ is the row vector $(\lambda_1,\lambda_2,..\lambda_n)$. Unfortunately, the matrix (**I-R**) has a zero determinant and therefore, the traffic equations, have an infinite number of solutions. To proceed let us define the occupancy vector $\underline{m}=(m_1,m_2,...,m_n)$ where $m_i=1$ if the system is found at state S_i and $m_i=0$ otherwise. It is clear that the elements of m always satisfy $\sum_i m_i = 1$. The model described is similar to a closed network of servers in which there is exactly one job travelling (trapped) between the service centres. Let λ^* be some none zero solution of the traffic equations. The probability distribution of **m** is given as [12]: $p(\mathbf{m}) = \alpha_1(m_1).\alpha_2(m_2)...\alpha_n(m_n)/G$, where $\alpha_i(0)= 1$ & $\alpha_i(1)=\lambda_i^*/\mu_i$. G is a normalisation constant evaluated as :$G = \sum \prod_i \alpha_i(m_i)$, where the summation is realised over all possible vectors**m**. G reduces to $G = \sum_i \lambda_i^*/\mu_i$. Finally, the distribution of soliton structures is found as,

$P_i = G^{-1}.\lambda_i^*/\mu_i$ which is the key result of this paper.

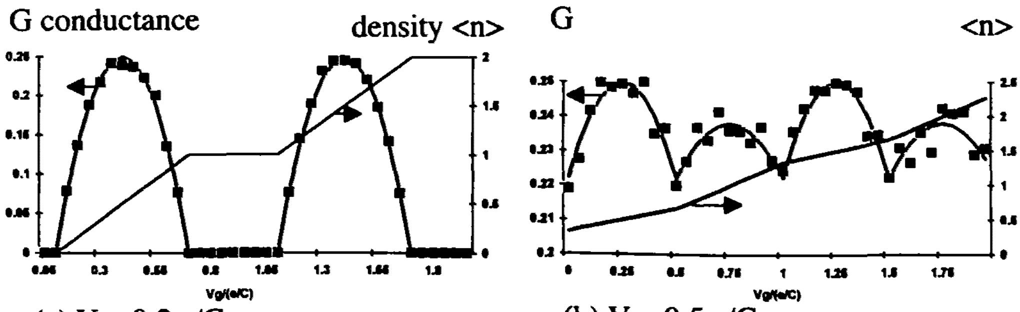
IV. Exactly solvable example: the double junction

Figure 6 shows an application to the 2-junction (or single quantum dot) compared with results obtained by the Monte Carlo method(750 events).Above a threshold voltage V_{th} electrons can tunnel into and away from the dot, one electron at a time. For N excess electrons the total electrostatic energy due to this charge is $(Ne)^2/2C_T$ where C_T is the total capacitance seen by the charge, $C_T=C_1+C_2+C_0$ and C_1,C_2 are the capacitances between the dot and the metallic electrodes and C_0 is the capacitance to the ground electrode. In this case, on leaving state S_i the system can only make a transition to states S_{i+1} or S_{i-1}(see figure 4). This is a Markovian birth-death process which can be solved exactly using the transition rates λ_n (S_{n-1} -> S_n) and μ_n (S_n -> S_{n-1}) given in terms of the tunnelling rates by: $\lambda_n = \Gamma_{n-1}(l,c) + \Gamma_{n-1}(r,c)$; $\mu_n = \Gamma_n(c.l) + \Gamma_n(c,r)$.The probability of finding the system at state S_n is found as:

$$P_n = Z^{-1}. \prod_{i=k_1+1}^{n} (\lambda_i/\mu_i) \quad (n > k_1) \text{ or } = Z^{-1} \quad (n = k_1)$$

$$Z = 1 + \sum_{i=k_1}^{k2} \prod_{n=k_1+1}^{i} (\lambda_n/\mu_n)$$

and where k_1 and k_2 are the minimum and maximum possible number of excess electrons that can be accommodated on the dot. The states contributing to the process can be discovered according to the condition: if $\lambda_i > 0$ and $\mu_i > 0$ then both S_i and S_{i-1} are legal states.
The oscillations in the dc conductance with gate voltage at low bias voltage correspond to transport through the same number of soliton states. In Fig6(a) transport in the different segments corresponds to states (0,1),(1,2), (2,3) and so on. At higher bias more states are generated leading to splitting of the segments. In Fig 6(b) the sequences are: (0,1),(0,1,2),((1,2) and (1,2,3). The maximum of the conductance peaks increases with applied bias whereas the minima decrease in amplitude. For high bias V>>Vth , the conductance approaches the constant valueG -> $G_t/2$. The speed-up over Monte Carlo simulation (750 departures) is about 300 x in CPU time.



(a) V = 0.2 e/C (b) V = 0.5 e/C

Figure 6. Conductances G/G_0 and soliton density <n> as a function of voltage V_g for 2 tunnel junctions at bias voltages V=0.2, 0.5 e/C. Squares:Monte Carlo .lines: queuing theory .

V. Time evolution

Suppose the system is found in state $S = S_k$ at $t = 0$ with initial densities $P_i(0) = \delta_{ik}$. At a later time t the system will be in a mixture of all possible legal states with densities $P_i(t)$ given by the rate equation: $\partial P_i/\partial t = \Sigma \{P_j \mu_{ji}\} - P_i \mu_i$. As an application, consider a single quantum dot described by the 2 soliton state $\{S\} = \{ n, n+1\}$, with birth and death coefficients λ and μ respectively. If the system is known to be in state S_n at t=0, the rate equation may be solved to give: $P_n(t) = \{\mu+\lambda \exp(-(\lambda+\mu)t \} /(\lambda+\mu)$; $P_{n+1}(t) = \lambda \{1 - \exp(-(\lambda + \mu)t \} /(\lambda + \mu)$
which in the limit t -> ∞ recovers our earlier result $P_n(\infty) = \mu / (\lambda+\mu)$; $P_{n+1}(\infty) = \lambda / (\lambda+\mu)$.

VI. General application to multiple junction systems and quantum dots

Very complex behaviour is possible with soliton propagation in multiple junctions. Figure 7 shows an example from the traffic theory where we have enhanced the magnitude of the conductance oscillations in a 3-junction array by choosing one element to have a large RC value. Here Vr = 0 and Vl = -V. In region (a) a maximum of one excess electron can stay in the system; the possible states are (0.0), (0.1), (1,0) and the system stays mainly in state (0,1). In region (b) the dominany state is (1,1) and up to 2 electrons excess occur. In region (c) 11 states contribute to the conduction with betwen 0 and 3 excess electrons. Region (d) has 12 states with between -1 and 3 excess charges. The conductance evidently peaks each time the system can accommodate one more electron.

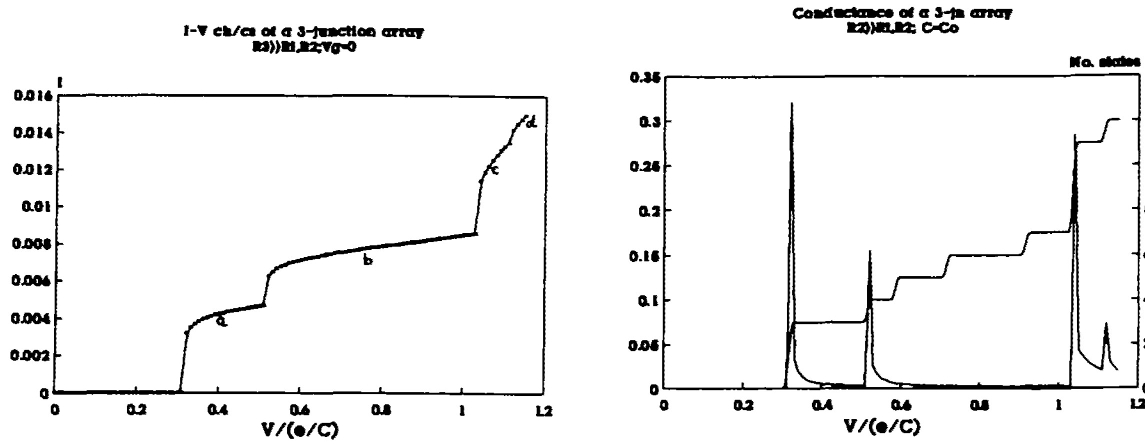


Figure 7. I-V characteristics, conductance and number of soliton states for a 3-junction having $R_3 = 50\ R_1$, $R_1=R_2$; $C_1=C_2=C_3=C$; $C=C_0$.

The methodology has been extended to (a) quantum dot systems where the isolated charging islands have discrete energy levels[13]; (b) macroscopic quantum tunnelling processes [10] The theory is being applied to the simulation of ultra-small coupled Schottky dot structures[6] for which the capacitance matrx requirs a full 3D Poisson solver.

References

[1] T A Fulton and G J Dolan, Phys Rev Letters, **89** 109 (1987).

[2] L S Kuzmin, P Delsing, T Claeson, K K Likhaerev, Phys Rev Letters **60** 309 (1989), **62** 2539 (1989)

[3] L J Geerligs, V F Anderagg, P A M Holweg, J E Mooij, H Pothier, D Esteves, C Urbina, M H Devoret, Phys Rev. Letters **64** 2691 (1990).

[4] U Meirev, M A Kastner and S J Wind Phys RevB**40** 5871 (1989).

[5] D V Averin and K K Likhaerev, J. Low Temp Phys **62** 345 (1986).

[6] J R Barker, J M R Weaver, S Babiker and S Roy, *Theory, modelling and construction of single-electronic systems*, Proceedings Second International Symposium on New Phenomena in Mesoscopic Structures, Hawaii, (1992).

[7] N S Bakhavalov, G S Kazacha, K K Likhaerev, S I Serdyukova, Sov. Phys. JETP **68** 581 (1989).

[8] J R Barker, S Roy and S Babiker, *Trajectory representations, fluctuations and stability of granular electronic devices,* Chapter 22, pp213-232, Science and Technology of Mesoscopic Structures, edited by S. Namba, C. Hamaguchi, T. Ando. Springer-Verlag: London, Tokyo, New York. (1992)

[9] L.Kleinrock,"Queuing Systems",New-York, London,(1974,1976).

[10] D.Averin & A.Odintsov, Phys. Lett.(A),**140**, 251 (1989) .

[11] M Schwartz, "Computer communication Network Analysis", Prentice-Hall Inc (1977).

[12] J Jackson, "Networking of waiting lines", Operations Research, 518, (1959).

[13] C Beenakker, Phys. Rev B44, 1646 (1991).