Weak Charge Quantization as an Instanton of the Interacting σ Model

Alex Kamenev
Department of Physics, Technion, Haifa 32000, Israel
(Received 24 February 2000)

Coulomb blockade in a quantum dot attached to a diffusive conductor is considered in the framework of the nonlinear σ model. It is shown that the weak charge quantization on the dot is associated with instanton configurations of the Q field in the conductor. The instantons have a finite action and are replica nonsymmetric. It is argued that such instantons may play a role in the transition regime to the interacting insulator.

PACS numbers: 73.50.Bk, 72.10.Fk, 72.15.–v, 73.23.Hk

The physics of interacting electronic systems in the presence of disorder has been the subject of an intense study for a few decades already [1]. Various theoretical approaches have been developed for the description of both metallic and insulator phases. The nonlinear σ model (NLσM) in the replica [2,3] or dynamic [4,5] formulation has proven to be the most powerful tool to deal with the weakly disordered (metallic) phase. Despite a lot of effort, the NLσM was not really successful in treating the insulating regime. On the other hand, approaches based on the idea of hopping conductivity [6] are quite efficient in the phenomenological description of the insulating phase. This is not the case in the theory of noninteracting localization, where the NLσM provides a unified treatment [7] of both metallic and insulating regimes (at least in low dimensions, $d \leq 2 + \epsilon$). It is also known to give some insight [8] into the physics of the integer quantum Hall transitions.

To describe the insulating phase one usually assumes that the system may be divided into weakly connected islands whose charge is essentially quantized. The charge may be changed only by an addition or subtraction of an integer amount of e, which costs a finite Coulomb energy. As a result, the important charge configurations are quantized and inhomogeneous. Approaching the metallic phase, the coupling between islands becomes better and charge quantization (Coulomb blockade) effects are less pronounced. Finally, deep in the metallic phase the charge is continuous and practically homogeneous (due to screening). In a *perturbative* treatment of NLσM (which is usually involved in the description of the metallic phase) there is no room for the discreteness of the electron charge. It is doubtful, therefore, that any NLσM perturbative approach can provide a picture of the transition or insulating regimes, where charge quantization is of primary importance. The purpose of this Letter is to show how the weak residual charge quantization may be incorporated in the diffusive NLσM formalism.

To this end I consider a toy problem of a large quantum dot with the capacitive interaction connected to a bulk reservoir by a diffusive noninteracting wire. The dot is open in the sense that the wire conductance is large: $G \gg G_Q \equiv e^2/(2\pi\hbar)$. One may, nevertheless, ask whether any Coulomb blockade effects can be seen on such a dot. In particular, we look for the oscillatory part of the electron number, $\tilde{N}(q)$, on the dot as a function of the continuous background charge, q. (For a closed dot, $G \ll 1$, such oscillations are, of course, extremely pronounced [9].) In the limiting case of a short (see below) wire this problem was recently solved by Nazarov [10]. Employing the model introduced earlier by Matveev [11], he derived an expression for an amplitude, A, of the Coulomb blockade oscillations for a dot connected to a reservoir by an arbitrary (delay free) scattering region:

$$A \sim \prod_i |R_i|^{1/2}, \tag{1}$$

where R_i are eigenvalues of the scattering region reflection matrix. Then, e.g., $\tilde{N}(q) = A\cos(2\pi q + \eta)$, where η is a phase, which depends on the scattering matrix of the contact. For a diffusive wire the eigenvalues R_i are random functions of the disorder realization with the probability density [12] $P(R) = G/[2(1 - R)\sqrt{R}]$, where $G = \int dR(1 - R)P(R)$ is the average conductance of the wire (hereafter conductance is measured in units of G_Q). Employing Eq. (1), one finds for the typical oscillation amplitude

$$e^{\langle \ln A\rangle} = e^{-(\pi^2/8)G}, \tag{2}$$

where the angular brackets stand for disorder averaging. This result holds for a metallic wire with $L \ll \xi$, where L is the wire length and ξ is the localization length, and therefore $G \gg 1$. Moreover, to treat the wire as an instantaneous scattering region one assumes that $L \ll L_T \equiv \sqrt{\hbar D/(2\pi T)}$ [13], where D is the diffusion constant of the wire and T is the temperature. Since, according to Eq. (2), the Coulomb blockade is exponentially suppressed, we refer to it as the weak charge quantization.

We are interested in the wire rather than in the dot. We show, thus, how the weak charge quantization, Eqs. (1) and (2), may be understood (and extended for $L > L_T$) from the point of view of the NLσM of the wire. Since the coupling constant of the NLσM is $1/G$, the expected result, Eq. (2), is the nonanalytic function of the coupling constant. Therefore, it cannot be obtained in any perturbative treatment of NLσM. Indeed, we show that these are the instanton configurations with the finite action, which are responsible for the weak charge quantization. In our

 0031-9007/00/85(19)/4160(4)$15.00

toy model the electron-electron interactions and disorder are spatially separated on the dot and the wire correspondingly. Nevertheless, one has to consider the full interacting NLσM of Finkel'stein [2,3] with the Q field having a structure both in the energy and replica spaces. The finite action instantons, mentioned above, involve rotations between two or more different replicas. Thus each instanton configuration is *replica nonsymmetric*. Eventually the replica symmetry is restored by instantons with all possible replica permutations. The model provides the first example of a nonperturbative and nonreplica diagonal application of the interacting NLσM. We confirm Nazarov's result, Eq. (2), and extend it to the case of a long, $L > L_T$, wire and 2D diffusive conductor.

Let us consider a large [14] quantum dot connected to a reservoir by a noninteracting diffusive wire. The Coulomb interaction on the dot has the form

$$H_{\rm int} = E_c(\hat{N} - q)^2, \tag{3}$$

where $E_c = e^2/(2C)$ is the capacitive charging energy, $\hat{N}$ is the electron number operator on the dot, and q is the background charge. Our aim is to calculate the oscillatory part of the average (over quantum fluctuations) number of electrons on the dot, $\tilde{N}(q)$. The next step is to average the result over realizations of disorder in the diffusive wire. The random phase of oscillations results in $\langle \tilde{N}(q)\rangle = 0$. Therefore, to calculate a typical oscillation amplitude one has to look for the correlation function [15]

$$\langle \tilde{N}(q)\tilde{N}(q')\rangle = \lim_{n\to 0}\frac{1}{n^2}\,\partial_q\partial_{q'}\langle Z^n(q)Z^n(q')\rangle\,, \tag{4}$$

where we introduced the partition function of the dot-wire system, $Z(q)$, along with the replica trick to facilitate the disorder averaging. The partition function of the open dot may be written in the standard way [16] as the imaginary time functional integral over the scalar field, $\Phi(\tau)$. If $T \gg \Delta$ [14], the Φ field configurations may be classified by the integer winding number, W, by imposing the boundary condition $\Phi(\tau + \beta) = \Phi(\tau) + 2\pi W$, where $\beta = 1/T$. In our case both the field and the corresponding winding number are $2n$-component replica vectors, $\mathbf{\Phi} = \{\Phi_1, \ldots, \Phi_{2n}\}$ and $\mathbf{W} = \{W_1, \ldots, W_{2n}\}$. The correlation function takes the form

$$\langle Z^n(q)Z^n(q')\rangle = \sum_{\mathbf{W}=-\infty}^{\infty} \exp\biggl\{2\pi i\biggl(q\sum_{a=1}^{n} W_a + q'\sum_{a=n+1}^{2n} W_a\biggr)\biggr\}\int \mathcal{D}\mathbf{\Phi}\,\exp\biggl\{-\frac{1}{2E_c}\int_0^{\beta}\sum_{a=1}^{2n}\dot{\Phi}_a^2\,d\tau - S_w(\mathbf{\Phi})\biggr\}. \tag{5}$$

Here $S_w(\mathbf{\Phi})$ is the average effective action which is a result of tracing out degrees of freedom of the wire. The latter are connected to the dot's variable, $\mathbf{\Phi}$, through the boundary condition at the dot-wire interface (for $T \gg \Delta$). In terms of the NLσM [2,3] the effective action may be written as

$$e^{-S_w(\mathbf{\Phi})} = \int \mathcal{D}\mathbf{Q}\,\exp\biggl\{\frac{\pi\nu}{4}\int_0^L dx\,\mathrm{Tr}\{-D(\nabla\mathbf{Q})^2 + 4\hat{\boldsymbol{\varepsilon}}\mathbf{Q}\}\biggr\}, \tag{6}$$

where ν is the density of states (DOS) in the wire and $\hat{\boldsymbol{\varepsilon}} = \delta^{ab}\varepsilon_m\delta_{mk}$ is a matrix diagonal in replica and Matsubara spaces, $\epsilon_m = 2\pi T(m + \frac{1}{2})$. The $\mathbf{Q}$ field is also a matrix, $Q^{ab}_{\varepsilon_m\varepsilon_k}(x)$, both in the replica space, $a, b = 1, \ldots, 2n$, and the Matsubara space, $m, k = 0, \pm 1, \pm 2, \ldots$.

On the wire-reservoir and the wire-dot interfaces the $\mathbf{Q}$ field obeys the following boundary conditions:

$$\mathbf{Q}(x = L) = \mathbf{\Lambda} \equiv \delta^{ab}\,\mathrm{sgn}(\varepsilon_m)\delta_{mk}\,; \tag{7a}$$

$$\mathbf{Q}(x = 0) = e^{-i\mathbf{\Phi}(\tau)}\mathbf{\Lambda}_{\tau-\tau'}e^{i\mathbf{\Phi}(\tau')}, \tag{7b}$$

correspondingly. In the last equation we employed imaginary time representation of the $\mathbf{Q}$ field instead of the frequency representation. In addition, the normalization condition $\mathbf{Q}^2 = 1$ must be obeyed at every point in space.

For a good metal, $G \equiv h\nu D/L \gg 1$, the $\mathbf{Q}$ integral may be calculated in the saddle point approximation. The extremal $\mathbf{Q}$ configurations are given by the Usadel equation,

$$\nabla(D\mathbf{Q}\nabla\mathbf{Q}) - [\hat{\boldsymbol{\varepsilon}}, \mathbf{Q}] = 0\,, \tag{8}$$

where $\mathbf{Q}$ satisfies $\mathbf{Q}^2 = 1$ and the boundary conditions (7). Our strategy, thus, is to find the solution of Eq. (8) for a fixed $\mathbf{\Phi}$ and identify $S_w(\mathbf{\Phi})$ with the action on the extremal $\mathbf{Q}$ configuration.

The oscillatory (in $q - q'$) component of the correlation function originates from nonzero winding numbers; cf. Eq. (5). Before proceeding in this direction, we make a few remarks about the $\mathbf{W} = 0$ sector. In this case the functional integral on the right-hand side of Eq. (6) may be minimized by the *replica diagonal* anzatz [4]

$$\mathbf{Q}_{\tau\tau'}(x) = e^{-i\mathbf{K}(\tau,x)}\mathbf{\Lambda}_{\tau-\tau'}e^{i\mathbf{K}(\tau',x)}, \tag{9}$$

where $\mathbf{K} = K_a(\tau, x)$ is replica and imaginary time diagonal matrix with $\mathbf{K}(\tau, L) = 0$ and $\mathbf{K}(\tau, 0) = \mathbf{\Phi}(\tau)$. Substituting the anzatz (9) into Eq. (6) and minimizing the action with respect to $\mathbf{K}$ [17], one finds

$$S_w(\mathbf{\Phi}) = \frac{G}{4\pi\beta}\sum_{\omega_m}\frac{|\omega_m|^{3/2}}{E_T^{1/2}}\coth\sqrt{\frac{|\omega_m|}{E_T}}\,|\mathbf{\Phi}(\omega_m)|^2, \tag{10}$$

where $E_T \equiv \hbar D/L^2$ is the Thouless energy of the wire. In the short wire limit, $|\omega_m| \ll E_T$, one recovers the phenomenological action of an Ohmic environment [16,18] $S_w = G/(4\pi\beta)\sum_m |\omega_m|\,|\mathbf{\Phi}_m|^2$. Although configurations with $\mathbf{W} = 0$ do not affect the thermodynamics of the dot [cf. Eq. (5)], they change its tunneling DOS $\sim \langle \exp\{i[\mathbf{\Phi}(\tau) - \mathbf{\Phi}(\tau')]\}\rangle_{\mathbf{\Phi}}$. In the small temperature (or short wire) limit, $T \ll E_T$, one finds the well-known [19] power law DOS $\sim T^{2/G}$. In the other limiting case, $T \gg E_T$, the DOS is given by $\exp\{-\frac{2}{G}\sqrt{\frac{E_T}{T}}\}$. This is nothing but the 1D Altshuler-Aronov zero-bias anomaly [4,20,21] for the case of the interacting dot and noninteracting wire. One can easily include long-range interactions inside the wire, which lead to the dynamical screening term in the K action [4,17], to obtain the standard 1D result [20]. Therefore, both the phenomenological theory of an Ohmic bath and the zero-bias anomaly are consequences of the $\mathbf{W} = 0$, replica diagonal sector of the NLσM.

We turn now to our main subject—the nontrivial winding numbers, $\mathbf{W} \neq 0$. One may try to solve Eq. (8) by the replica diagonal anzatz, Eq. (9). Then the phase $K_a(\tau, x)$ acquires W_a vortices in the (x, τ) plane. As a result, the corresponding action is logarithmically large, $S_w \sim G \ln L/\lambda$, where λ is a microscopic cutoff length. (Note that this logarithmic divergence continues well in the ballistic regime; therefore λ is likely to be of the order of the Fermi wavelength.) On the other hand, one may find solutions of Eq. (8) with much smaller action, $S_w \sim G$. For this one *must* allow rotations between different replicas and impose the following condition on the possible $\mathbf{W}$ configurations:

$$\sum_{a=1}^{2n} W_a = 0. \tag{11}$$

Equations (4), (5), and (11) immediately lead to the following conclusions: (i) $\langle \tilde{N}(q) \rangle = 0$; (ii) the correlation function $\langle \tilde{N}(q) \tilde{N}(q') \rangle$ is a function of $q - q'$ only.

To demonstrate how such solutions may be constructed, let us consider the optimal realization of the $\mathbf{\Phi}$ field, $\mathbf{\Phi}(\tau) = 2\pi \mathbf{W} \tau/\beta$. Employing Eq. (7b), one finds

$$\mathbf{Q}(x = 0) = \delta^{ab} \operatorname{sgn}(\varepsilon_{m-W_a}) \delta_{mk}. \tag{12}$$

This is equivalent to the local, replica dependent shift of the chemical potential on W_a Matsubara units. Consider, e.g., the simplest nontrivial winding number configuration consistent with Eq. (11) having only two nonzero components of $\mathbf{W}$:

$$W_a = 1;\ a \in [1, n]; \qquad W_b = -1;\ b \in [n + 1, 2n], \tag{13}$$

and $W_c = 0$ for $c \neq a, b$. In this case one may satisfy the boundary conditions Eqs. (7a) and (12) along with $\mathbf{Q}^2 = 1$ by choosing

$$\begin{pmatrix} Q^{aa}_{\frac{\pi}{\beta}\frac{\pi}{\beta}} & Q^{ab}_{\frac{\pi}{\beta}-\frac{\pi}{\beta}} \\ Q^{ba}_{-\frac{\pi}{\beta}\frac{\pi}{\beta}} & Q^{bb}_{-\frac{\pi}{\beta}-\frac{\pi}{\beta}} \end{pmatrix} = \begin{pmatrix} \cos\theta(x) & \sin\theta(x) \\ \sin\theta(x) & -\cos\theta(x) \end{pmatrix}, \tag{14}$$

where $\theta(0) = \pi$, $\theta(L) = 0$ and all other elements of $\mathbf{Q}$ are equal to those of $\mathbf{\Lambda}$. The Usadel equation takes the form of the anharmonic pendulum equation

$$L_T^2 \nabla^2 \theta - \sin\theta = 0. \tag{15}$$

Its solution may be written in terms of the elliptic integrals. The corresponding action is equal to

$$S_w = G \frac{L}{2L_T} \left[\frac{4}{\sqrt{\kappa}} E(\kappa) - \frac{1 - \kappa}{\kappa} \frac{2L}{L_T} \right] + O(n), \tag{16}$$

where $\sqrt{\kappa}\, K(\kappa) = L/L_T$ and $K(\kappa)$ and $E(\kappa)$ are the complete elliptic integrals of the first and second kind. In the two limiting cases one obtains (in the $n \to 0$ limit)

$$S_w = \begin{cases} \frac{\pi^2}{4} G(1 + \frac{4T}{\pi E_T}) & T \ll E_T; \\ 2G(L_T), & T \gg E_T, \end{cases} \tag{17}$$

where $G(L_T) \equiv e^2 \nu D/L_T = G\sqrt{2\pi T/E_T}$ is the dimensionless conductance of the length L_T of the wire. In the zero temperature case the instanton spreads uniformly over the wire, $\theta(x) = \pi(L - x)/L$. In the opposite limit, $L \gg L_T$, it is confined to the region of the size L_T near the dot: $\theta(x) = 2\pi - 4 \arctan[\exp\{x/L_T\}]$. The low temperature (short wire) limit may be directly compared to Nazarov's result, Eq. (2) (we calculate $\langle \tilde{N}^2 \rangle$ rather than $\langle |\tilde{N}| \rangle$, thus the factor of 2 difference). In the general case Eq. (16) provides the entire temperature and wire-length dependence of the oscillation amplitude (with the exponential accuracy). Winding number configurations other than Eq. (13) are exponentially less probable. The analytical continuation, $n \to 0$, is straightforward if one notices that there are precisely n^2 different configurations like Eq. (13) with the same action and $q - q'$ dependence. Thus the n^2 factor in Eq. (4) is canceled and the remaining expression is n independent. It is important to notice that the rotational symmetry between the replicas is broken down to the permutation one by the choice of the winding number configuration. As a result the degenerate saddle point manifolds, playing the central role in the case of noninteracting level statistics [22], are absent in the present case.

One can define the current operator on the instanton trajectory as $\mathbf{J} \equiv i\mathbf{Q}\nabla\mathbf{Q} = \sigma_y \nabla\theta(x)$, where σ_y is the Pauli matrix in the space defined in Eq. (14). In the limit of the long wire, $L \gg L_T$, the current is nonuniform along the wire since, $\nabla\theta \neq \text{const}$. By virtue of the continuity equation, this seemingly leads to the charge accumulation in the region $x \sim L_T$. If it were the case, our model of the noninteracting wire would loose its validity and finite temperature results should be modified. This is not the case, since the current, $\mathbf{J}$, is pure replica off diagonal (because of the σ_y matrix) and therefore completely decouples from the Coulomb interactions. Interpretation of this counterintuitive phenomena is that the replica structure of the instanton takes care of the mesoscopic fluctuation effect. It describes sample specific energy (and thus temperature) dependence of the transmission coefficients, introducing retardation neglected in the Nazarov approach. This retardation is not associated with the accumulation of any real charge in some portion of the wire.

Finally, we address the 2D setup with the dot of the radius d placed at the center of a 2D disordered disk of the radius L. In this case the elementary instanton is given by Eqs. (13) and (14) with $\theta = \theta(r)$ which is a solution of the following equation:

$$L_T^2 \partial_r (r \partial_r \theta) - r \sin\theta = 0 \tag{18}$$

with $\theta(d) = \pi$ and $\theta(L) = 0$; r is the polar radius. For $L \ll L_T$ the equation is solved by [23]

$$\theta(r) = \pi \left[1 - \frac{\ln r/d}{\ln L/d} \right], \tag{19}$$

with the corresponding action

$$S_w = \frac{\pi^2}{4} \frac{2\pi e^2 \nu D}{\ln L/d}. \quad (20)$$

At higher temperature, $d \ll L_T < L$, with the logarithmic accuracy one must substitute L_T instead of L in Eq. (20). By analogy with Eq. (17) the final result may be written as $S_w = \pi^2 G(L_T)/4$, where $G(L_T) = 2\pi e^2 \nu D / \ln L_T/d$ is the conductance of the 2D disk with the outer radius L_T and the inner one d. Note that the temperature dependence of the 2D action is $[\ln D/(d^2 T)]^{-1}$ much slower than $\sqrt{T}$ in one dimension.

In conclusion, we have shown that the residual charge quantization in the diffusive interacting systems may be described by the instantons of the interacting NLσM. These instantons are replica nonsymmetric finite action configurations of the **Q** field, which extremize the action with the "twisted" boundary conditions. As an illustration, we have calculated the (exponentially small) amplitude of the Coulomb blockade oscillation in the open quantum dot connected to the 1D (2D) diffusive conductor of arbitrary length (radius). Although the instantons are exponentially rare in the diffusive regime, they are responsible for the charge quantization and inhomogeneity—the features which dominate the physics of the insulating phase. One may expect them to be increasingly important as the insulator regime is approached.

Discussions with L. Levitov and A. Shytov have initiated this project. Valuable comments of I. Aleiner, A. Andreev, M. Feigelman, Y. Gefen, and I. Gruzberg are highly acknowledged. This research was supported in part by BSF Grant No. 9800338.

[1] *Electron-Electron Interaction in Disordered Systems,* edited by A.L. Efros and M. Pollak (Elsevier Science Publishers B.V., North-Holland, Amsterdam, Oxford, New York, Tokyo, 1985).

[2] A.M. Finkel'stein, in *Electron Liquid in Disordered Conductors*, Soviet Scientific Reviews Vol. 14, edited by I.M. Khalatnikov (Harwood Academic Publishers GmbH, London, 1990).

[3] D. Belitz and T.R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).

[4] A. Kamenev and A. Andreev, Phys. Rev. B **60**, 2218 (1999).

[5] C. Chamon, A.W.W. Ludwig, and C. Nayak, Phys. Rev. B **60**, 2239 (1999).

[6] A.L. Efros and B.I. Shklovskii, in Ref. [1], pp. 409–482.

[7] K. Efetov, *Supersymmetry in Disorder and Chaos* (Cambridge University Press, Cambridge, England, 1997).

[8] H. Levine, S.B. Libby, and A.M.M. Pruisken, Nucl. Phys. **B240**, 30 (1984); **B240**, 49 (1984); **B240**, 71 (1984); H.A. Weidenmuller and M.R. Zirnbauer, Nucl. Phys. **B305**, 339 (1988).

[9] D.V. Averin and K.K. Likharev, in *Mesoscopic Phenomena in Solid,* edited by B.L. Altshuler, P.A. Lee, and R.A. Webb (Elsevier Science Publishers B.V., North-Holland, Amsterdam, Oxford, New York, Tokyo, 1991).

[10] Yu.V. Nazarov, Phys. Rev. Lett. **82**, 1245 (1999).

[11] K.A. Matveev, Phys. Rev. B **51**, 1743 (1995).

[12] Yu.V. Nazarov, Phys. Rev. Lett. **73**, 134 (1994).

[13] To disregard the thermal smearing of the Coulomb blockade, the inequality $T \ll E_c$ [see Eq. (3)] is assumed hereafter.

[14] A large dot means $\Delta \ll T$, where Δ is the mean level spacing of the dot. If $T \sim \Delta$ the mesoscopic effects play an important role; see I. Aleiner and L.I. Glazman, Phys. Rev. B **57**, 9608 (1998).

[15] The better quantity to average is $\Re \ln \tilde{N}(q)$ [10]. It may be calculated in the same way with a slightly more complicated algebra.

[16] G. Schön and A.D. Zaikin, Phys. Rep. **198**, 237 (1990).

[17] The action in terms of the **K** field has the form [4], $S_w = \nu/2 \int dx \sum_m [\hbar D|\omega_m||\nabla \mathbf{K}_m|^2 + \omega_m^2 |\mathbf{K}_m|^2]$, where $\omega_m = 2\pi T m$. The last term in the action originates from the static compressibility of the electron gas.

[18] A.O. Caldeira and A.J. Leggett, Ann. Phys. (N.Y.) **149**, 374 (1983).

[19] Yu.V. Nazarov, Sov. Phys. JETP **68**, 561 (1989); M.H. Devoret *et al.,* Phys. Rev. Lett. **64**, 1824 (1990); S.M. Girvin *et al.,* ibid. **64**, 3183 (1990); A. Kamenev and Y. Gefen, Phys. Rev. B **54**, 5428 (1996).

[20] B.L. Altshuler and A.G. Aronov, in Ref. [1], pp. 1–154.

[21] L.S. Levitov and A.V. Shytov, JETP Lett. **66**, 214 (1997).

[22] A. Kamenev and M. Mézard, J. Phys. A **32**, 4373 (1999); Phys. Rev. B **60**, 3944 (1999); I.V. Yurkevich and I.V. Lerner, Phys. Rev. B **60**, 3955 (1999).

[23] I am indebted to M. Feigelman for pointing out the mistake in the previous version.