Surface and soliton charge in insulating systems

Qian Niu*
Physics Department, University of Washington, FM-15, Seattle, Washington 98195
(Received 3 September 1985)

By employing the locality property of the Green's function, it is shown that the charge of a filled band can only be changed in integral units under a local perturbation. This result is used to show that the surface charge depends on the surface potential only in a discrete way. The half-integral quantization of the surface charge in a case of an inversion-symmetric and periodic bulk potential has been rederived. It is also shown that the fractional part of a soliton charge associated with a topological defect in the periodic potential is simply determined by the period misfit of the potential on the two sides of the defect.

Consider a one-dimensional noninteracting electron system in a periodic potential. The energy spectrum is known to consist of Bloch bands separated by energy gaps. If the periodic potential is altered in a finite region, the energy spectrum is not changed except for a number of discrete levels appearing between the bands or below the lowest band. Nonlocal perturbation of the periodic potential can also be considered. One example is that the potential at the right of some point be uniformly shifted to the right by less than a period, with the resulting void region being filled with an arbitrary potential. We call the resulting imperfection a topological defect (or shift-generated topological defect, to be specific). Another kind of topological defect can be obtained by taking off the semi-infinite part of the periodic potential at the right of some point, and replacing it with the reflection of the left-side potential about this point. This gives rise to a reflection-generated topological defect, which is distinguished from the above case if the periodic potential does not have a center of reflection. It is clear that the energy spectrum in such a system is still the same as in the ideal case, with perhaps a number of discrete levels off the bands. We can also consider an "edge defect" resulting from replacing a semi-infinite range of the potential by a potential barrier; the energy spectrum below the barrier is also the same as before, discounting the discrete levels.

Because of the defects in the potential, the electron charge density of a filled band is distorted from the ideal case. By now, there already exists a large number of works concerning the charge accumulation or depletion due to such distortions. It has been shown by Heine[1] that in case of a reflection-symmetric and periodic bulk potential, the surface charge is quantized in half-integral units. Extensive discussion and various generalizations of this result has been given by a number of people.[2] In the presence of a topological defect, fractional charge deficiency has been found by Su, Schrieffer, and Heeger.[3] This problem has also been extensively studied on various models using both conventional solid-state theory[4-7] and field-theoretical methods.[8]

Here we present an elementary derivation of the aforementioned results, attempting to achieve a certain unity and abstractness of the approach. Our basic tool is the Green's function, supplemented by a "δ-function-potential cutting" trick. We develop this idea by the statement and proof of a series of theorems.

Theorem 1. The total electron charge of the states enclosed in a contour in the complex energy plane is unchanged by the perturbation of the potential in a finite region, if the contour crosses the real energy axis only through the gaps of the unperturbed spectrum, and there are no energy levels that can pass through the contour during the continuous turning-on of the perturbation.

Here the unperturbed system may already have a finite number of defects, topologically trivial or nontrivial. Because the energy spectrum is real, the contour mentioned in the theorem is free to change so long as it encloses the same energy range. Another point to note is that although the concept of "total charge" is meaningless beause of its divergence when one or more energy bands are enclosed in the contour, the change Δ of the total charge can actually be defined in a precise way as the integral of the local electron-density distortion due to the perturbation:

$$\Delta=\int_{-\infty}^{\infty} dx\,[\rho(x)-\rho_0(x)]\ ; \tag{1}$$

here (and hereafter) the electron charge is set to be unity. If $G(x,x';z)$ is the perturbed Green's function and $G_0(x,x';z)$ the unperturbed, then the density distortion due to the perturbation $V(x)$ can be written as

$$\rho(x)-\rho_0(x)=\oint_C \frac{dz}{2\pi i}[G(x,x;z)-G_0(x,x;z)]\ . \tag{2}$$

Using the Dyson relation between the perturbed and unperturbed Green functions, the above equation can be turned into the following form:

$$\rho(x)-\rho_0(x)=\oint \frac{dz}{2\pi i}\int_{-\infty}^{\infty} dx' G_0(x,x';z) \times V(x')G(x',x;z)\ . \tag{3}$$

Because our contour of integration is away from the spectrum of G_0 (and therefore of G, assuming no discrete level on the contour), the Green's functions are exponentially bounded for large distances, $|x-x'|$. As a result, the

density distortion diminishes exponentially as x is far from the region of the perturbation. The change of the total charge given by Eq. (1) is thus well defined.

Now we are ready to prove theorem 1. First we consider the perturbation $\lambda V(x)$, with $0<\lambda\leq 1$. By substituting Eq. (3) into Eq. (1) and expanding G_λ in terms of G_0 and $\lambda V(x)$, we have

$$\Delta_\lambda=\oint\frac{dz}{2\pi i}[\mathrm{Tr}(G_0\lambda VG_0)+\mathrm{Tr}(G_0\lambda VG_0\lambda VG_0)+\cdots], \tag{4}$$

where the traces in coordinate representation is just the integration in coordinates. The series in the square brackets converges if

$$\int_{-\infty}^{\infty}dx\,|V(x)| \tag{5}$$

is bounded and λ sufficiently small. In fact, this convergence is uniform for z on the contour, so that we can interchange the order of the summation and the integration in z. By using the cyclic invariant property of the trace operation and the following relation,

$$(G_0)^2=-\frac{\partial}{\partial z}G_0, \tag{6}$$

the nth-order term can be written as

$$-\frac{\lambda^n}{n}\frac{\partial}{\partial z}\mathrm{Tr}(G_0VG_0V\cdots), \tag{7}$$

which gives a zero result after the integration in z because of the single-valuedness of G_0 on the contour. Thus Δ_λ vanishes for a sufficiently small but finite value (say, λ_1) of λ.

To raise λ further, we treat the system with $\lambda_1V(x)$ contained in the potential as the unperturbed system, and $(\lambda-\lambda_1)V(x)$ as the perturbation. The same argument goes through for sufficiently small but positive $\lambda-\lambda_1$, yielding

$$\int_{-\infty}^{\infty}dx\,[\rho_\lambda(x)-\rho_{\lambda_1}(x)]=0. \tag{8}$$

Thus

$$\begin{aligned}\Delta_\lambda&\equiv\int_{-\infty}^{\infty}dx[\rho_\lambda(x)-\rho_0(x)]\\&=\int_{-\infty}^{\infty}dx\{[\rho_\lambda(x)-\rho_{\lambda_1}]+[\rho_{\lambda_1}(x)-\rho_0(x)]\}\\&=\int_{-\infty}^{\infty}dx[\rho_\lambda(x)-\rho_{\lambda_1}(x)]+\int_{-\infty}^{\infty}dx\,[\rho_{\lambda_1}(x)-\rho_0(x)]\\&=0.\end{aligned} \tag{9}$$

In this way, we can raise λ step by step until $\lambda=1$, while keeping the change of the total charge being zero.

The condition in theorem 1 may well be violated for a general local perturbation; for example, there might be some levels passing through the contour during the continuous turning on of the perturbation. In this case we have the following theorem.

Theorem 2. The total electron charge of the states with their energies enclosed in a given contour in the complex energy plane is changed only in integral units, if the perturbation is in a finite region, and if the contour crosses the real energy axis only through the gaps of both the unperturbed and the perturbed spectrum.

Again, we first consider the perturbation $\lambda V(x)$. As λ is increased from 0, the spectrum of the perturbed system varies continuously from the unperturbed. Suppose at $\lambda=\lambda_1$, a level crosses the contour C for the first time. We distort our contour a little bit from C to C' to give way to the level. This allows λ to increase to $\lambda_1+\epsilon$ (with ϵ small and positive) with C' never being crossed. According to theorem 1, the change of the total charge of the states enclosed in C' is zero, i.e.,

$$\oint_{C'}\frac{dz}{2\pi i}\int_{-\infty}^{\infty}dx\,[G_{\lambda_1+\epsilon}(x,x;z)-G_0(x,x;z)]=0. \tag{10}$$

Since there is a level of the perturbed system contained between C and C' at $\lambda=\lambda_1+\epsilon$, we must have

$$\begin{aligned}&\oint_C\frac{dz}{2\pi i}\int_{-\infty}^{\infty}dx[G_{\lambda_1+\epsilon}(x,x;z)-G_0(x,x;z)]\\&\quad=\oint_{C'}\frac{dz}{2\pi i}\int_{-\infty}^{\infty}dx[G_{\lambda_1+\epsilon}(x,x;z)-G_0(x,x;z)]\\&\qquad+\oint_{C-C'}\frac{dz}{2\pi i}\int_{-\infty}^{\infty}dx\,G_{\lambda_1+\epsilon}(x,x;z)\\&\quad=\pm 1,\end{aligned} \tag{11}$$

where the plus (minus) sign corresponds to the case that the level crossed the contour C from outside (inside). We can continue this argument until $\lambda=1$, yielding

$$\int_{-\infty}^{\infty}dx\,[\rho(x)-\rho_0(x)]=n,$$

where n is the net number of levels that went inside of the contour C during the course of increasing λ from 0 to 1.

A direct consequence of theorem 2 is that the increase of the total charge of a band of states is equal to an integer, which is the net number of discrete levels merged into the band during the continuous turning-on of the perturbation.

The following theorem has a special value for the discussion of the "edge charge" of a semi-infinite system, and the "soliton charge" due to a topological defect.

Theorem 3. The insertion of a δ-function impurity in the potential changes the charge of a band by one or nothing. With a δ-function potential at $x=a$, the perturbed Green's function satisfies

$$G(x,x';z)=G_0(x,x';z)+VG_0(x,a;z)G(a,x';z), \tag{12}$$

where V is the amplitude of the δ-function. We first set $x=a$ in the above equation to solve for $G(a,x';z)$, which is then substituted back into the equation to give an expression for the perturbed Green's function as

$$G(x,x';z)=G_0(x,x';z)+\frac{G_0(x,a;z)G_0(a,x';z)V}{1-VG_0(a,a;z)}. \tag{13}$$

From Eq. (12) it is clear that the poles of G are not at those of G_0; thus the poles of G must come from the solution of the following equation:

$$VG_0(a,a;z)=1. \tag{14}$$

This has solutions only when z is real and away from the bands. From the spectral representation of $G_0(a,a;z)$, it

is a strictly decreasing function of z when z is in an energy gap. Just above a pole $G_0(a,a;z)$ is positive infinite; just below a pole it is negative infinite. It may or may not be infinite at a band edge, depending on whether the wave function $\psi(x=a;z)$ at the band edge is nonzero or zero. It is now clear that in the course of increasing V from zero to positive (negative) infinity, all the discrete levels are shifted upwards (downwards), and the number of levels split from or merged into an energy band is at most one. The present theorem immediately follows from theorem 2.

To discuss the charge associated with an edge, we have to choose a reference charge distribution, which is to be subtracted from the actual charge density. A natural choice of this is the charge distribution in the corresponding infinitely extended potential, with the part outside the boundary being set to zero. According to theorems 2 and 3, the fractional part of the edge charge must be independent of the details of the edge potential once the boundary point is chosen. For different choice of the boundary points d and d' in the reference charge distribution, $\rho_0(x)$, the fractional part of the edge charge changes by the amount

$$\int_d^{d'} dx\,\rho_0(x) \pmod 1 . \tag{15}$$

If the bulk periodic potential has a center of reflection, a general theorem claims that the edge charge is an integer or half-integer with the boundary point (which determine the reference charge distribution) at a center of reflection of the periodic potential. This result was first obtained by Heine[1] using a quite different method, and was then discussed by a number of people.[2] We now give a simple proof of this theorem. Suppose the bulk periodic potential is infinitely extended (removing the edge), we insert a δ-function potential at a reflection center, and let the amplitude go to positive infinity. Theorem 3 says that this will change the total charge of a band by one or nothing. Because the charge density before and after the insertion of the δ potential is symmetric about the position of the δ function, the charge deficiency must be shared equally by the two sides. Each side is equivalent to a semi-infinite system with a hard wall at the edge, so we conclude that the edge charge is $\pm\frac{1}{2}$ or zero. Since the modification of the surface potential can only change the surface charge by an integer, the surface charge in the general case must be an integer or a half-integer.

The half-integral quantization of the surface charge has been employed by Kallin and Halperin[9] to prove the vanishing of the piezoelectric constant of a crystal, which has a reflectional symmetry but may not be terminated symmetrically. We can follow exactly the same kind of reasoning to prove that the piezoelectric effect is a bulk effect in the general case, using the fact that the surface charge depends on the surface potential only in a discrete way.

Now we consider the problem with a topological defect in the periodic potential,

$$V(x)=\begin{cases} V_0(x) & \text{if } x\le 0 , \\ V_0(x-a) & \text{if } x>0 , \end{cases} \tag{16}$$

where $V_0(x+1)=V_0(x)$, and $0<a<1$. The case with $a=l/2$ has been studied in a remarkable work of Su, Schrieffer, and Heeger,[3] and the charge deficiency due to such a defect has been found to be one-half. For a general shift of the potential, the charge deficiency has been found to be a/l by Prange[5] and Thouless.[6] Here we give a new derivation of these results.

The soliton charge associated with the defect is defined as

$$\Delta_d=\frac{1}{2L}\int_{-L}^{L} dx_0 \left[\int_{-\infty}^{x_0} dx[\rho(x)-\rho_0(x)] + \int_{x_0}^{\infty} dx[\rho(x)-\rho_0(x-a)]\right] \quad (L\gg l) , \tag{17}$$

where $\rho(x)$ is the electron density of a band in the potential $V(x)$, and $\rho_0(x)$ is that of the same band in the potential $V_0(x)$. It can be easily shown that the above definition of the soliton charge is equivalent to that given by Kivelson and Schrieffer.[4] We take the present form only for convenience. Suppose two δ potentials with positive infinite amplitudes are inserted at $x=x_0$ and $x=x_0+a$ in the potential $V(x)$, resulting in a electron density of $\rho_\delta(x)$. According to theorem 3, we have

$$\int_{-\infty}^{\infty} dx[\rho_\delta(x)-\rho(x)]=\text{integer} . \tag{18}$$

On the other hand,

$$\int_{-\infty}^{x_0} dx[\rho_\delta(x)-\rho_0(x)]+\int_{x_0+a}^{\infty} dx[\rho_\delta(x)-\rho_0(x-a)] = \int_{-\infty}^{x_0} dx[\rho_\delta(x)-\rho_0(x)] + \int_{x_0}^{\infty} dx[\rho_\delta(x+a)-\rho_0(x)]=\text{integer} \tag{19}$$

because the potential giving rise to the density $[\rho_\delta(x), x<x_0; \rho_\delta(x+a), x>x_0]$ can be regarded as

$$V_1(x)=A\delta(x-x_0)+\begin{cases} V(x), & x<x_0 \\ V(x+a), & x>x_0 \end{cases} \qquad A\to+\infty \tag{20}$$

which differs from $V_0(x)$ only in a finite region. Collecting Eqs. (17)–(19), we have

$$\Delta_d=\text{integer}+\frac{1}{2L}\int_{-L}^{L} dx_0 \int_{x_0}^{x_0+a} dx[\rho_\delta(x)-\rho_0(x-a)] . \tag{21}$$

Since $\rho_\delta(x)$ $(x_0<x<x_0+a)$ is the charge density of the discrete levels (with their energies within the band under consideration) bounded by the two δ potentials of positive infinite amplitudes, we must have

$$\Delta_d=\text{integer}-\frac{1}{2L}\int_{-L}^{L} dx_0 \int_{x_0}^{x_0+a} dx\,\rho_0(x-a) = \text{integer}-a/l . \tag{22}$$

According to theorem 2, the above result should be still valid if the potential (16) is modified in a finite region.

The formula (22) has been derived by Prange[5] and Thouless[6] based on quite different theories. This result can be intuitively understood as follows.[10] During the formation of the defect, electrons far away at the right of the defect are carried to the right by a distance a. This must leave a charge deficiency of a/l in the vicinity of the defect, assuming one electron per unit period. Further distortion of the potential around the defect can only introduce a charge difference equal to an integer.

Another way to derive this result[11] is to make a scale transformation in the region in the topological defect, such that the periods of the potential fit in both sides. The Hamiltonian in the new coordinates differs from the ideal one by a local perturbation of the potential and the kinetic energy. According to theorem 2, this ideal Hamiltonian can be used to calculate the fractional part of the soliton charge. The soliton charge is given by Eq. (22), because of an "ideal" distribution of the electron density in the new coordinates corresponds exactly to a deficiency of a/l in the original coordinates.

In the case of a reflection generated topological defect, we can insert a δ potential at the reflection center, and let the amplitude of the δ potential go to positive infinity. It is immediately clear that the electron charge associated with this defect is, apart from an additive integer, twice the surface charge of the semi-infinite system at one side of the δ potential.

So far we have been talking about one-dimensional electron systems. Our theorems 1 and 2 remain valid in higher dimensions. The results about surface charge and soliton (domain wall) charge can also be carried over to higher dimensions if the system is periodic in the newly added dimensions.

The effect of many-body interaction can be treated by a self-consistent approximation.[9] The effective potential differs from the bare potential of the ions in two ways. First, the periodic part is modified while the periodicity remains the same. This, however, will introduce no difficulty, since our theory does not refer to any detailed structure of the periodic potential. Second, a local perturbation of the bare potential may lead to a long-range perturbation of the effective potential, because the electron-density distortion is not necessarily confined in a local region. But one expects that the long-range tail should be exponentially small, due to the insulating property of the Bloch bands. From Eq. (3), further distortion in the charge density due to such a small tail in the effective potential remains exponentially small. By a perturbation theory similar to the proofs of our theorems, one expects that the charge of a filled band cannot be changed by such an exponential tail of the effective potential.

Now we summarize our basic results. By employing the locality property of the Green's function when its energy parameter is away from the spectrum, we have shown that the charge of a band can only be changed in integral units by a local perturbation. According to this result, the fractional part of the surface charge only depends on the bulk periodic potential and the position of the boundary point in the definition of the reference charge-density distribution. The half-integral quantization of the surface charge in case of an inversion-symmetric bulk potential has been rederived. It is also shown that the fractional part of a soliton charge associated with a topological defect in the periodic potential is simply determined by the asymptotic period misfit of the potential on the two sides of the defect.

ACKNOWLEDGMENTS

The author is grateful to Professor David Thouless and Professor John Rehr for many valuable commments on this work. He also thanks Mr. Dan Lu and Yan-Jun Ma for many useful discussions. This work was supported by the National Science Foundation under Grants No. DMR 83-19301 and No. DMR 84-15063.

*Present address: The University of Illinois, Department of Physics, 1110 West Green Street, Urbana, IL 61801.

[1]V. Heine, Phys. Rev. **145**, 593 (1966).

[2]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B **10**, 4973 (1974); F. Claro, *ibid.* **17**, 699 (1977); L. Kleinman, *ibid.* **11**, 858 (1975).

[3]W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. B **22**, 2009 (1980).

[4]S. Kivelson and J. R. Schrieffer, Phys. Rev. B **10**, 6447 (1980); see also R. Rajaraman and J. S. Bell, Phys. Lett. **116B**, 151 (1982).

[5]R. E. Prange, Phys. Rev. B **26**, 991 (1982).

[6]D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[7]S. Kivelson, Phys. Rev. B **26**, 4269 (1982); A. J. Niemi and G. W. Semenoff, Phys. Rev. Lett. **51**, 2077 (1983).

[8]See a recent article by M. Stone, Phys. Rev. B **31**, 6112 (1985), and references therein.

[9]C. Kallin and B. I. Halperin, Phys. Rev. B **29**, 2175 (1984).

[10]J. R. Schrieffer, Mol. Cryst. Liq. Cryst. **77**, 209 (1981).

[11]D. J. Thouless (private communication).