RESEARCH ARTICLE | MAY 01 1995

Modeling the patterned two-dimensional electron gas: Electrostatics \odot

John H. Davies; Ivan A. Larkin; E. V. Sukhorukov



J. Appl. Phys. 77, 4504–4512 (1995) https://doi.org/10.1063/1.359446





CrossMark

05 November 2023 20:13:03





Modeling the patterned two-dimensional electron gas: Electrostatics

John H. Daviesa) and Ivan A. Larkin

Department of Electronics and Electrical Engineering, Glasgow University, Glasgow G12 8QQ, United Kingdom

E. V. Sukhorukov

Institute of Problems of Microelectronics Technology, Academy of Science of Russia, Chernogolovka, Moscow 142432, Russia

(Received 26 September 1994; accepted for publication 11 January 1995)

We present analytical calculations of the potential in a two-dimensional electron gas (2DEG) generated by patterned polygon gates on the surface of a heterostructure. They give the bare and screened potentials and reveal the effect of different boundary conditions on the surface. The formulas for the bare electrostatic potential from patterned gates are simple enough to be plotted in spreadsheets; they give threshold voltages, estimates of the region occupied by the 2DEG, and the energies of some collective infra-red excitations. We also consider the screened potentials in linear response, where no part of the 2DEG is fully depleted, which can again be found within an electrostatic approximation. The behavior of the exposed surface between the gates affects the potential strongly. Surface states provide perfect pinning of the Fermi energy in the "equipotential" model, the usual assumption, but this requires charge to move to the surface from the 2DEG. The charge on the surface is held fixed in response to a gate voltage in the "frozen" model, which typically gives a lower cutoff voltage but stronger confinement of electrons in a split-gate wire. The difference between the two models is large, emphasizing that accurate modeling needs a thorough understanding of the surface states. © 1995 American Institute of Physics.

I. INTRODUCTION

Recent years have seen rapid development in experiment and theory on "quantum devices," structures whose dimensions are comparable with the wavelength of electrons within them. 1,2 The behavior of these devices depends on the wavelike nature of electrons, and many structures have clear ancestors in microwaves. As in microwaves, a basic need is to guide the waves effectively without significant scattering. Electronic devices are typically based on a two-dimensional electron gas (2DEG) in a III-V heterostructure where scattering is small because of the separation between the donors and electrons. The electrons are then guided by electrostatic fields produced by patterned gates on the surface of the structure to produce long wires^{3,4} (equivalent to electromagnetic waveguides), short wires or quantum point contacts, 5,6 crosses, 7-9 quantum dots, 10 and many other devices. The Fermi wavelength of electrons in a 2DEG is typically 50 nm so the critical dimensions of the patterned gates are around 0.1 µm. Confining the 2DEG on this scale also leads to other physical effects, notably the Coulomb blockade when the electrostatic charging energy of a small puddle of electrons becomes similar to other energy scales of the problem.10

Modeling these structures poses a number of difficult problems.

(1) The geometry is complicated. Typical dimensions of the gates are 0.1μ m and they may have intricate patterns on the surface. The 2DEG is some 0.05μ m below the surface, so modeling should take full account of the three-dimensional geometry.

- (2) Some dimensions of the region occupied by electrons are comparable with their wavelength so that a quantummechanical treatment of the electron may be needed. Accurate modeling of a quantum dot, where motion in all directions is quantized, would need solution of a threedimensional Schrödinger equation for all occupied states self-consistently with the electrostatic potential—a major numerical task.
- (3) Several aspects of the materials are not well understood, particularly the surface and the donors in n-AlGaAs. The small size of the device emphasizes the influence of the surface, and the problems are exacerbated by the high electric fields due to the small size.

Although a number of calculations have been performed with different methods and approximations, ^{11–17} all are fairly complicated and time-consuming. It would be desirable to have something simpler to give a rapid picture of the potential inside a device. We address this goal here by providing results that can be used to calculate the electrostatic potential seen in the 2DEG due to the gates alone. While this neglects the important (Hartree) contribution of the electrons themselves that must be included in a self-consistent treatment, it is useful for several reasons.

- (1) The formulas can readily be evaluated and plotted with a spreadsheet, in contrast to the computational resources required for a self-consistent calculation.
- (2) Some quantities measured experimentally, notably infrared absorption, ¹⁸ may reflect the *bare* confining potential rather than the screened one because of Kohn's theorem. ^{19–23}
- (3) Another experimentally important quantity, the cutoff

a)Electronic mail: jdavies@elec.gla.ac.uk

- voltage, refers to conditions when all electrons have left the 2DEG and can therefore be calculated from the bare potential alone.²⁴
- (4) Important features, such as the edge of the region occupied by electrons, can be estimated from the electrostatic potential within a simple local screening approximation.
- (5) It is also possible to calculate the *linear* response of a 2DEG to the potential from the gate within an electrostatic picture that treats the 2DEG as an equipotential.²⁵

The first aim of this paper is to calculate the electrostatic field due to a number of patterned gates used in practice, ranging from the simple wire to general polygons and to use these to calculate cutoff voltages and the confinement energies of electrons in wires.

Another issue that we consider is the treatment of exposed surfaces, part of the third problem raised above. Submicron devices inevitably have a large surface : volume ratio, which means that surfaces have a greater effect on their properties. Experimental problems may arise through contamination of exposed surfaces; simulations require an accurate model of the surface if they are to give reliable results. While these problems have largely disappeared for Si, the behavior of III-V materials is strongly influenced by surface states. This may not matter in conventional devices, where the active region is covered with a metal gate, but the opposite is true where gates are used to pattern a 2DEG. The basic device is a narrow quantum wire^{3,4} formed under the gap between split gates. A negative voltage on the gate repels electrons from the two-dimensional electron gas (2DEG) underneath, but some electrons remain under the gap to form the wire. The electrons are immediately underneath the exposed surface, and indeed this gap is needed for the device to function. Clearly the operation of this device depends crucially on the behavior of the surface. It will be shown that two different, but physically plausible, models for the surface states give substantially different potentials in the 2DEG. In principle it ought to be easy to discriminate between these experimentally; in practice the differences may be obscured by the non-ideal behavior of other parts of the device, notably the donors in n-AlGaAs. Scanning probe microscopy may be able to resolve this problem.

II. PHYSICAL MODEL OF HETEROSTRUCTURES

Patterned 2DEGs are almost always formed in III-V heterostructures (silicon inversion layers are occasionally used but will not be considered here). These are layered structures, typically comprising a patterned gate covering only part of the surface of a cap layer of GaAs, an *n*-doped layer of AlGaAs, a spacer layer of undoped AlGaAs and an undoped channel of GaAs in which the 2DEG resides. Unfortunately there are gaps in our understanding both of the surface and the donors.

As we have just discussed, the behavior of the exposed surface is crucial in modeling the response of a 2DEG to a patterned gate, and recent numerical work has emphasized this point. ¹⁷ It is usually assumed, particularly for devices at room temperature, that the Fermi energy E_F is pinned at a fixed energy below the conduction band on free surfaces of

GaAs. 26 This is unambiguous in equilibrium, where E_F is constant throughout the structure. However, it presents problems as soon as a negative bias V_g is applied to the gate, even though almost no current flows, as it is no longer clear whether the surface is pinned to E_F in the 2DEG or to that in the gate. Clearly the device would not produce a wire if the surface was pinned to the gate, as no difference between the gap and gate would be seen from inside the material. Thus it must be assumed that the surface remains pinned to E_F in the 2DEG, which we shall take as the zero of energy and potential. This means that the exposed surface behaves as an equipotential at zero, a convenient boundary condition that gives simple electrostatics. This is a "pinned" surface.

The difficulty with the pinned surface is that charge must move from the 2DEG to the surface to keep its potential constant in response to a change in V_g . While this may pose no problem at room temperature, it seems unlikely at the low temperatures (around 1 K) at which experiments on quantum devices are usually performed, and the charge on the surface may well be unable to change under these conditions. Although the equipotential model can be used to calculate the density of electrons in the 2DEG at equilibrium at room temperature, the surface should instead be treated as a simple dielectric boundary, with a fixed charge density, in response to a change in V_g applied at low temperature. This is a "frozen" surface. The high dielectric constant of semiconductors means that the boundary condition for this model is approximately $\partial \phi / \partial n = 0$.

We assume that the layer of donors simply behaves as a fixed charge density which does not respond to the gate voltage. This implies that the donors must be fully ionized or that any electrons around the donors are firmly trapped in deep states (DX centers²⁷). Experiments using a large gate²⁸ shows that this is a reasonably good picture below some 150 K but not at higher temperature when there are usually free electrons around the donors. There may be a further problem with a patterned gate because this produces strong horizontal components to the electric field which could alter the occupation of donors by inducing electrons to tunnel between them. Again, we ignore this possibility.

For calculating the "bare" electrostatic potential seen by the electrons due to the gate alone, the only property of the 2DEG that we need is its depth, d. Although the main contribution to d is the depth of the GaAs-AlGaAs interface at which the electrons are trapped, the peak density of the 2DEG is some 5-10 nm below the interface and this should also be included.

An important ingredient that we shall need is the threshold voltage of uniform layers, V_t , which is the (negative) voltage that must be applied to a large gate to reduce the density of electrons just to zero. Let the density of the 2DEG be $n_{\rm 2D}$ when $V_g=0$. There are several components 24,28,29 to V_t , the major one being the "parallel-plate capacitor" voltage needed to remove the charge density, $-en_{\rm 2D}d/e\epsilon_0$. There is also the energy of the bound state in the potential well at the heterojunction, and the kinetic (Fermi) energy of the 2DEG. The former is included roughly if we measure d to the peak density of the 2DEG rather than to the interface, as suggested above. The kinetic energy effectively adds a

further $\frac{1}{4} a_0$ to d where a_0 is the Bohr radius scaled by the effective mass and dielectric constant of the semiconductor, about 10 nm in GaAs. Typically d>25 nm so the electrostatic energy is bigger by a factor of at least 10 and $V_t \approx -en_{2D}d/\epsilon\epsilon_0$ is a good approximation. We shall use the inequality $d \gg \frac{1}{4} a_0$ again in Sec. VI to argue that a 2DEG may be treated as an equipotential for linear response.

We shall now use this model to calculate the potential generated by gates of various patterns both for an empty 2DEG and for a filled 2DEG where the response is linear.

III. PATTERNED GATES ON A PINNED SURFACE

The aim of this section is to calculate the electrostatic potential in a plane d below the surface of the semiconductor partially covered by a patterned gate. The surface occupies the x-y plane and z is measured into the material. The exposed surface is taken to be pinned so we can set the potential $\phi(\mathbf{r},0)=0$ there, with $\phi(\mathbf{r},0)=V_g$ on the gate. We use lower-case letters like $\mathbf{r}=(x,y)$ to denote two-dimensional vectors with the corresponding upper-case letters for three-dimensional vectors like $\mathbf{R}=(x,y,z)=(\mathbf{r},z)$. Thus the problem is to find a solution $\phi(\mathbf{R})$ to Laplace's equation $\nabla^2\phi=0$, given the value $\phi(\mathbf{r},0)$ on the plane z=0, and subject to the further boundary condition $\partial\phi/\partial z=0$ as $z\to\infty$.

One route is to start by making a two-dimensional Fourier transform from $\phi(\mathbf{r},0)$ to $\tilde{\phi}(\mathbf{q},0)$. The dependence on z is a decaying exponential to satisfy Laplace's equation and the boundary condition at $z=\infty$, so $\tilde{\phi}(\mathbf{q},z)=\tilde{\phi}(\mathbf{q},0)\exp(-|qz|)$. This multiplication of the Fourier transform is equivalent to a convolution in real space. Taking the two-dimensional inverse Fourier transform of $\exp(-|qz|)$ leads to the general result

$$\phi(\mathbf{r},z) = \int \frac{|z|}{2\pi(z^2 + |\mathbf{r} - \mathbf{r}'|^2)^{3/2}} \phi(\mathbf{r}',0) d\mathbf{r}'.$$
(3.1)

We shall now evaluate the potential in the plane of the 2DEG, z=d, for a number of important shapes. The strategy here is to solve for a triangle and build other polygons from this, but we give a more general if cumbersome form in Sec. IV.

A. Triangular gate

4506

Consider an infinite triangle with one edge along the x-axis whose vertex at the origin includes an angle 2A. Expressed in two-dimensional polar coordinates $\mathbf{r} = (r, \theta)$, the general result (3.1) for the potential in the plane z = d becomes

$$\frac{\phi(\mathbf{r},d)}{V_g} = \frac{d}{2\pi} \int_0^{2A} d\theta' \int_0^{\infty} dr' \times \frac{r'}{[d^2 + r^2 + r'^2 - 2rr'\cos(\theta - \theta')]^{3/2}}.$$
(3.2)

The integral over r' is standard [Ref. 31, Eq. (2.264.5)], leaving

$$\frac{\phi(\mathbf{r},d)}{V_g} = \frac{d}{2\pi} \int_0^{2A} \frac{d\theta'}{R - r\cos(\theta - \theta')},$$
(3.3)

where $R = \sqrt{r^2 + d^2}$ is the three-dimensional distance from the vertex of the gate to the point of observation, and r is its projection onto the x-y plane. This integral [Ref. 31, Eq. (2.553.3)] gives

$$\frac{\phi(\mathbf{r},d)}{V_g} = \frac{1}{\pi} \arctan \frac{d\sin A}{(R-x)\cos A - y\sin A}.$$
 (3.4)

A problem now arises in choosing the branch of the arctangent. It should be interpreted in the sense of the numerical function "atan2"(y,x)" that is available in most programming systems. In other words, the quadrant of the arctangent depends on the *individual* signs of the numerator and denominator within it, not just on their ratio. This interpretation is essential to get the correct branch and is needed, for example, to ensure that $\phi = V_g$ everywhere in the limit of a gate covering the whole surface, $A = \pi$.

B. Split-gate wire

A long wire formed by a pair of split gates is the building block of devices based on quantum wires. The potential of a single gate occupying the upper half-plane can be found by putting $A = \pi/2$ in Eq. (3.4),

$$\frac{\phi(\mathbf{r},d)}{V_c} = \frac{1}{\pi} \arctan \frac{d}{-y} = \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \frac{y}{d} \right). \tag{3.5}$$

Thus the potential due to a stripe gate occupying |x| < a is

$$\frac{\phi(x,d)}{V_{\sigma}} = \frac{1}{\pi} \left(\arctan \frac{a+x}{d} + \arctan \frac{a-x}{d} \right). \tag{3.6}$$

This can be extended to a periodic structure, which might be used to form an array of wires or a superlattice. If the gaps have width 2b, giving a period of 2(a+b), then

$$\frac{\phi(x,d)}{V_g} = \frac{1}{\pi} \left(\arctan \frac{\tan\left[\frac{1}{2} \pi(a+x)/(a+b)\right]}{\tanh\left[\frac{1}{2} \pi d/(a+b)\right]} + \arctan \frac{\tan\left[\frac{1}{2} \pi(a-x)/(a+b)\right]}{\tanh\left[\frac{1}{2} \pi d/(a+b)\right]} \right). \tag{3.7}$$

Similarly, the important case of a split-gate wire, where the gate covers the whole surface except for a slit |x| < a, gives

$$\frac{\phi(x,d)}{V_g} = 1 - \frac{1}{\pi} \left(\arctan \frac{a+x}{d} + \arctan \frac{a-x}{d} \right). \tag{3.8}$$

C. Infinite rectangle

Setting $A = \pi/4$ in Eq. (3.4) gives

$$\frac{\phi(\mathbf{r},d)}{V_o} = \frac{1}{\pi} \arctan \frac{d}{R - x - y} \equiv f_1(x,y). \tag{3.9}$$

This is the potential generated by an infinite rectangular gate occupying the first quadrant with its vertex at the origin. Although this expression is simple, it will be convenient later to have an expression with definite symmetries in x and y. Splitting Eq. (3.9) into even and odd parts gives, after lengthy manipulation, the alternative form

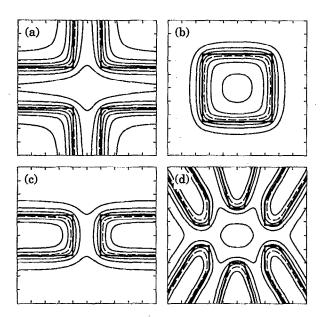


FIG. 1. Examples of the electrostatic potential in the 2DEG generated by patterned gates (thick lines) on the surface. The regions shown are 1 μ m square, the depth of the 2DEG is d=0.05 μ m, there is unit potential on all gates and the contours (thin lines) are separated by 0.1 with a broken line at 0.5: (a) Crossed long wires, (b) quantum dot defined by a square gate, (c) short wire defined by rectangular gates, (d) quantum dot defined by six polygonal gates.

$$\frac{\phi(\mathbf{r},d)}{V_g} = \frac{1}{2\pi} \left(\frac{\pi}{2} + \arctan \frac{x}{d} + \arctan \frac{y}{d} + \arctan \frac{xy}{dR} \right)$$
$$\equiv f_2(x,y). \tag{3.10}$$

These results can be used with superposition to calculate the potential from many practical patterned gates. For example, the Hall effect may be measured in the cross-shaped geometry shown in Fig. 1(a), defined by the gaps between four infinite rectangles. Equation (3.9) was used to calculate the potential contours, which show the flared shape of the junction; this is important for the dynamics of ballistic electrons and may lead to a reversal of the sign of the Hall effect. ^{8,9} It is also possible to apply a voltage that depletes the areas under the gates and the gaps between them but leaving a puddle of electrons under the junction to form a quantum dot.

D. Finite rectangle

Consider a rectangle defined by L < x < R, B < y < T. Its potential can be constructed from Eqs. (3.9) or (3.10) by superposing four infinite rectangles with their vertices on the four corners; those on (L,B) and (R,T) carry $+V_g$ while those on (L,T) and (R,B) carry $-V_g$. Many terms cancel if the symmetrized form $f_2(x,y)$ is used and we find

$$\frac{\phi(\mathbf{r},d)}{V_g} = g(x - L, y - B) + g(x - L, T - y) + g(R - x, y - B) + g(R - x, T - y)$$

$$+ g(R - x, y - B) + g(R - x, T - y)$$
(3.11)

where

$$g(u,v) = \frac{1}{2\pi} \arctan \frac{uv}{dR}$$
; $R = \sqrt{u^2 + v^2 + d^2}$. (3.12)

This result can be used to construct more complicated shapes as in the next example. By itself it can also be used to calculate the potential under a quantum dot and an example with a square gate, is shown in Fig. 1(b). The potential near the center of the dot, where the electrons reside, is nearly circular, which has also been seen in numerical work. ¹⁵

E. Short wire

The contours of potential under a finite wire, behaving as a short constriction between two large areas of 2DEG, is shown in Fig. 1(c). A very short wire with ballistic propagation can also be viewed as a quantum point contact. The length along y, the direction of propagation, is 2b, the width in x is 2a and the device is centered on the origin. The potential can be built from the previous results in a number of ways. The simplest is to regard this as a stripe gate with |y| < b and to remove from it a rectangular region with |x| < a, |y| < b to form the wire. This gives

$$\frac{\phi(\mathbf{r},d)}{V_g} = \frac{1}{\pi} \left(\arctan \frac{b+y}{d} + \arctan \frac{b-y}{d} \right)$$

$$-g(a+x,b+y) - g(a+x,b-y)$$

$$-g(a-x,b+y) - g(a-x,b-y). \tag{3.13}$$

Another route is to start from a split gate with gap |x| < a and to remove four infinite rectangles like x > a, y > b. This can be used to treat tapered gates by using the result for a triangle, Eq. (3.4) instead of that for the rectangle.

Clearly this process can be extended to deal with more complicated gates, but we now give a different approach based on the potential produced by a general polygon.

IV. POLYGON GATE

Although it is possible to construct the potential from any polygonal gate using the simple formula developed above for the triangle, we shall now develop a general formula for an arbitrary polygon, following earlier work that developed a general formula for the normal electric field throughout the plane of the gates. ³⁰ The problem is to find the potential $\phi(\mathbf{R})$ with the boundary conditions

$$\phi(\mathbf{r},z=+0) = \begin{cases} V_i & \text{if } \mathbf{r} \in S_i; \\ 0 & \text{otherwise}, \end{cases}$$
 (4.1)

where S_i is the area of gate i, and $\partial \phi / \partial z = 0$ as $z \to +\infty$.

To calculate $\phi(\mathbf{R})$ we introduce the Green's function $G(\mathbf{R}, \mathbf{R}')$ of Laplace's equation in the half-space z > 0. This obeys

$$\nabla^2 G(\mathbf{R}, \mathbf{R}') = -\delta(\mathbf{R} - \mathbf{R}'); \quad G(\mathbf{R}, \mathbf{R}')|_{z=0, R=\infty} = 0$$
(4.2)

and takes the form

$$G(\mathbf{R}, \mathbf{R}') = \frac{1}{4\pi\sqrt{|\mathbf{r} - \mathbf{r}'|^2 + (z - z')^2}}$$
$$-\frac{1}{4\pi\sqrt{|\mathbf{r} - \mathbf{r}'|^2 + (z + z')^2}},$$
(4.3)

where the second term is an image to satisfy the boundary condition on z = 0. Green's theorem then leads to³⁰

$$\phi(\mathbf{R}) = \int \int ds' \left[\phi(\mathbf{R}') \frac{\partial}{\partial z'} G(\mathbf{R}, \mathbf{R}') \right]_{z'=0}$$

$$= \frac{1}{2\pi} \sum_{i} V_{i} I_{i}(\mathbf{R}), \qquad (4.4)$$

where

$$I_i(\mathbf{R}) = \iint_{\mathcal{S}_i} \frac{zds'}{(|\mathbf{r} - \mathbf{r}'|^2 + z^2)^{3/2}}$$
(4.5)

and the integral is over the area of each gate. This is a more elegant derivation of our earlier result (3.1).

We shall now use the two-dimensional divergence theorem to reduce $I_i(\mathbf{R})$ to an integral along the boundary ∂S_i . This requires a function whose divergence is $z(r^2+z^2)^{-3/2}$. Using the relation

$$\operatorname{div} \frac{\mathbf{r}}{r^2} = 2 \pi \delta(\mathbf{r}) \tag{4.6}$$

in two dimensions, we obtain $\operatorname{div}(\mathbf{r}M(r)/r^2) = M'(r)/r + 2\pi M(0)\delta(\mathbf{r})$, where M(r) is an arbitrary function of $r = |\mathbf{r}|$ and M'(r) = dM(r)/dr. Thus the divergence theorem gives

$$\iint_{S_{i}} \frac{M'(r)ds}{r} = -\oint_{\partial S_{i}} \frac{M(r)\mathbf{r} \cdot d\mathbf{l}}{r^{2}} \\
-\begin{cases} 2\pi M(0) & \text{if } (\mathbf{r}=0) \in S_{i}; \\ 0 & \text{otherwise}. \end{cases}$$
(4.7)

The element of length $d\mathbf{l}$ (which would be an element of area in the usual three-dimensional case) is directed inside the area S_i , perpendicular to the boundary ∂S_i . Now let $M'(r)/r = z(r^2 + z^2)^{-3/2}$, whence $M(r) = -z(r^2 + z^2)^{-1/2}$; Eq. (4.6) allows us to ignore the constant of integration. Substituting M(r), and $(\mathbf{r} - \mathbf{r}')$ instead of \mathbf{r} , in Eq. (4.7) gives

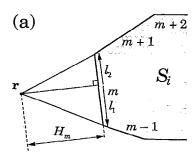
$$I_{i}(\mathbf{R}) = \oint_{\partial S_{i}} \frac{z(\mathbf{r}' - \mathbf{r}) \cdot d\mathbf{l}'}{|\mathbf{r}' - \mathbf{r}|^{2} \sqrt{|\mathbf{r}' - \mathbf{r}|^{2} + z^{2}}} + 2\pi C_{i}(\mathbf{r}), \quad (4.8)$$

where $C_i(\mathbf{r}) = 1$ if $\mathbf{r} \in S_i$ and 0 otherwise.

To simplify this further we assume that S_i is a polygon. Then Eq. (4.8) reduces to a sum of integrals over the edges \mathcal{R}_m of the polygon as shown in Fig. 2(a),

$$I_i(\mathbf{R}) = \sum_{m} J_{im}(\mathbf{R}) + 2\pi C_i(\mathbf{r}); \qquad (4.9)$$

$$J_{im}(\mathbf{R}) = \int_{\mathcal{B}_m} \frac{z H_m dl}{(H_m^2 + l^2) \sqrt{H_m^2 + z^2 + l^2}},$$
 (4.10)



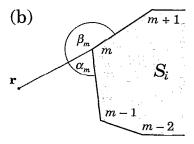


FIG. 2. Solution of the electrostatic potential from a polygon gate S_i . This shows the plane z=0 so \mathbf{r} is the projection of the point of observation $\mathbf{R}=(\mathbf{r},z)$. (a) Sum over edges m; note that $l_{1,2}$ are both measured away from the foot of the perpendicular H_m . (b) Sum over vertices m; the third angle γ_m involves the altitude z and does not appear on this projection.

where H_m is the distance between the edge \mathcal{R}_m and \mathbf{r} , the projection of the observation point \mathbf{R} onto the plane z=0. This is a standard integral [Ref. 31, (Sec. 2.28)] and we obtain, omitting the indexes i and m,

$$J(\mathbf{R}) = \arctan \frac{zl_1}{HR_1} + \arctan \frac{zl_2}{HR_2}$$
; $R_{1,2} = \sqrt{z^2 + H^2 + l_{1,2}^2}$. (4.11)

The distances $l_{1,2}$ are both measured outwards from the foot of the perpendicular H, and both $l_{1,2}$ and H may be negative [Fig. 2(a)]. In this case the arctangents are defined conventionally and *not* as atan2(y,x).

We must now verify that this solution obeys the boundary conditions. At $z \to +\infty$, $z/R_{1,2}=1$ and $J(\mathbf{R}) = \arctan l_1/H + \arctan l_2/H = \theta$, where θ is the angle (with sign) subtended by the edge at the projected observation point \mathbf{r} . Therefore, if the gate is of finite size, and the projection of the observation point lies outside the gate, we obtain $\phi = 0$ at infinity by summing over all edges. If the projection of the observation point lies inside the gate we get -2π after summation over all edges; this is cancelled by $2\pi C_i$ so $\phi = 0$ everywhere at $z \to +\infty$.

Now let the observation point **R** lie in some plane with finite z, and consider moving the projected point **r** along a line of varying H while keeping $l_{1,2}$ constant. Clearly both arctangents in Eq. (4.11) go to zero as $H \to \pm \infty$ but we must look carefully at the singularities when H=0 and each arctangent jumps by $\pm \pi$. If this line does not intersect the edge of the polygon, then l_1 and l_2 have opposite signs and the jumps in the two arctangents cancel one another when H=0. Thus the potential is continuous, and goes to zero everywhere along this line when $z\to 0$. Both $l_{1,2}$ are positive

if the line intersects the edge. The jumps in the arctangents now add as H goes from +0 (outside the gate) to -0 (inside), and this contribution to the potential jumps by -2π . The discontinuity is exactly cancelled by $2\pi C_i(\mathbf{r})$ so the potential is continuous everywhere. Only the C_i term survives as $z{\to}0$ which shows that $I_i(\mathbf{R})$ obeys the desired boundary conditions on the surface. We have therefore shown that ϕ is continuous everywhere and obeys all the boundary conditions.

The integral $I(\mathbf{R})$ may instead be expressed in terms of angles at each vertex, defined in Fig. 2(b). Defining $l_1/H = \cot \beta$, $l_2/H = \cot \alpha$ and $z/R = \sin \gamma$, we obtain

$$I_i(\mathbf{R}) = \sum_{m} \left[\arctan(\sin \gamma_m \cot \alpha_m) + \arctan(\sin \gamma_m \cot \beta_m) \right]$$

$$+2\pi C_i(\mathbf{r}),\tag{4.12}$$

where m labels the vertex of the polygon including those at infinity.

As an example of the sort of structure to which these formulas can be applied, Fig. 1(d) shows the contours of bare potential in a quantum dot defined by a complicated pattern of 6 polygonal gates with 28 vertices.

V. THRESHOLD VOLTAGES

The threshold voltage V_t of a 2DEG is the bias which must be applied to a large gate to remove the electrons underneath. There are typically several "thresholds" in a device with a patterned gate. For example, the constriction in Fig. 1(c) has a "definition voltage" close to V_t when the regions underneath the gates become depleted and current becomes restricted to the constriction. Later there will be a further "cutoff voltage" where the constriction itself is depleted and the current ceases. It is possible to use the electrostatic results that we have calculated to make a good estimate of the cutoff voltage in structures which become fully depleted, and to make a rough estimate in other cases.

The simplest case is the long wire confined by split gates which has been considered before. Let there be a gap of 2a between the gates; then the last electrons will remain in the middle of the wire where Eq. (3.8) shows that $\phi(0,d) = V_g[1-(2/\pi)\arctan(a/d)]$. The potential from the gate must reach V_t in the 2DEG to drive all the electrons away, so the required gate voltage is

$$V_g = V_t \left[1 - \frac{2}{\pi} \arctan \frac{a}{d} \right]^{-1}. \tag{5.1}$$

This neglects the binding energy of the lowest quantum-mechanical transverse state in the confining potential along x from the gate, but is otherwise exact (subject of course to the limitations of the model).

The separation of the energy levels at cutoff can also be estimated from the curvature of the potential, and an obvious use of this result is to find the width of the gap that gives the strongest confinement of electrons at a fixed depth d. Simple algebra gives $a \approx 0.8d$ and $V_g \approx 1.7V_t$. The separation of energy levels for this optimum wire is $\Delta \varepsilon \approx 17\sqrt{E_F/d}$ where the energies are measured in meV, d is in nm, E_F is the

Fermi energy of the uniform 2DEG, and the prefactor depends only on the dielectric constant of the host semiconductor. Clearly a shallow 2DEG with a high Fermi energy is needed to maximize the separation between the energy levels in a wire. The two parameters are linked; roughly $E_F \propto 1/d$ for a deep structure so $\Delta \varepsilon \propto 1/d$, but the benefits become less as d is reduced and E_F saturates. A typical shallow structure with $n_{\rm 2D} = 5 \times 10^{15} \, {\rm m}^{-2}$, $E_F = 18 \, {\rm meV}$ and $d = 35 \, {\rm nm}$ gives $\Delta \varepsilon \approx 12 \, {\rm meV}$, which would allow subbands to be resolved at temperatures up to 40 K. Unfortunately a gap of only 55 nm must be drawn, and self-consistent calculations show that $\Delta \varepsilon$ falls rapidly as electrons occupy the wire.

The voltage needed to remove electrons from the middle of a wire of finite length 2b can be calculated in the same way using Eq. (3.13), giving

$$V_g = V_t \left[\frac{2}{\pi} \arctan \frac{b}{d} - \frac{2}{\pi} \arctan \frac{ab}{d\sqrt{a^2 + b^2 + d^2}} \right]^{-1}.$$
(5.2)

In this case the result is only a rough guide because the 2DEG is not fully depleted when the wire is cut off: large areas of electrons remain to the left and right of the short wire. In using this approximation we are making the assumption that the response of the electrons is entirely local, so the the density of electrons at a point depends only on the external potential (from the gate, donors and surface) at that point. This is not correct for the contribution to the potential generated by other electrons in the 2DEG, which has a range set roughly by the depth of the 2DEG, d. The approximation is therefore more accurate for shallow 2DEGs and has been used to calculate the shape of the saddle point.³⁰ Its accuracy is difficult to check except by numerical methods for all but the simplest geometries. However it is appealingly simple, and can be extended to estimate the edge of the region occupied by electrons in the general case: one simply finds the contour where $\phi(\mathbf{r},d) = V_t$. The curvature of the bare potential along x can again be used to estimate the separation of energy levels in the middle of the wire; it hardly differs from the result for an infinitely long wire at the same gate voltage, even when b=a.

Other geometries can be treated with the same technique; for example one can find the voltage needed to drive electrons out of the quantum dot formed under the middle of a set of four infinite rectangular gates (Sec. III C).

VI. LINEAR RESPONSE

Another limit that can be treated simply with electrostatics is that of linear response, where the voltage on the gate is small enough that electrons remain throughout the 2DEG. The usual dielectric function³² modified to account for the boundary conditions on the surface³³ can be used but requires numerical solution in almost all cases. Instead we can exploit the approximation used in Sec. II that the kinetic energy of the 2DEG is much smaller that the electrostatic potential between the 2DEG and the surface.²⁵ Typically this ratio $d/\frac{1}{4}a_0 > 10$ so it is a reasonable approximation to neglect the change in kinetic energy induced by the potential on

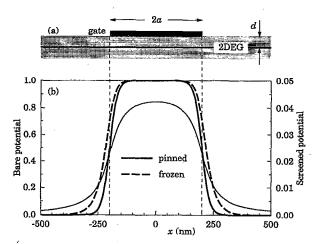


FIG. 3. (a) Stripe gate carrying unit potential at |x| < a on the surface of a semiconductor with a = 200 nm and the 2DEG at a depth d = 50 nm. (b) Bare potential in 2DEG from gate with pinned surface (thin line, left scale), and screened potential (right scale) with a pinned surface (full line) and frozen surface (broken line).

the gate. The 2DEG can therefore be treated as an equipotential surface for the electrostatic problem. The change in energy of the 2DEG is then estimated from the induced charge in the 2DEG, given by the product of the normal derivative $E_z = -\partial \phi/\partial z$ at z = d and $\epsilon \epsilon_0$. Dividing the charge by the density of states, $m/(\pi \hbar^2)$, then gives the change in kinetic energy to lowest order.²⁵ Thus the screened potential in the 2DEG is approximately

$$\phi_{\rm scr}(\mathbf{r},d) = \frac{\pi\hbar^2}{m} \frac{\epsilon \epsilon_0}{e^2} E_z(\mathbf{r},d) = -\frac{a_0 \partial \phi}{4 \partial z} \bigg|_{z=d}.$$
 (6.1)

Consider the stripe gate at |x| < a shown in Fig. 3(a). We must solve Laplace's equation subject to the boundary conditions $\phi = 0$ for z = d in the 2DEG, with $\phi = V_g$ for |x| < a on the gate, and $\phi = 0$ for |x| > a on the free surface where the potential is pinned by the surface states. A conformal map is straightforward and gives

$$\frac{\phi(x,z)}{V_g} = V \left\{ 1 - \frac{2}{\pi} \operatorname{Im} \operatorname{arctanh} \left[\frac{\tanh(\frac{1}{2} \pi w/d)}{\tanh(\frac{1}{2} \pi a/d)} \right] \right\}, \quad (6.2)$$

where w=x+iz. The screened potential then follows from Eq. (6.1),

$$\frac{\phi_{\rm scr}(x,d)}{V_g} = \frac{a_0}{4d} \frac{\tanh(\frac{1}{2} \pi a/d)}{1 + {\rm sech}^2(\frac{1}{2} \pi a/d) \sinh^2(\frac{1}{2} \pi x/d)}.$$
(6.3)

This is plotted with the potential from the gate alone (Eq. 3.6) in Fig. 3(b). Note that the screened potential falls off exponentially with x in contrast to the slow x^{-2} decay of the gate potential. This is perhaps surprising as one might not expect exponential screening in a two-dimensional system, but it must be noted that the length scale is set by the depth of the 2DEG, not the Bohr radius, and is therefore due to electrostatics rather than dielectric response.

We can also use this result to calculate the "definition voltage" at which the 2DEG under the middle of the gate is just depleted. This will occur when $\phi_{scr}(x=0,d)$

 $=(\frac{1}{4} a_0/d)V_t$, whence $V_g=V_t \coth(\frac{1}{2} \pi a/d)$. This is increased over the value for a large gate but only by a factor of 1.09 for a=d. Narrow gates are often used in devices like quantum point contacts and it is reassuring that they do not need severely increased voltages to deplete the areas underneath them.

VII. FROZEN SURFACE

A frozen surface is often more difficult to treat analytically, although this approximation has been used before in the theory of narrow wires. 25,34 First, it gives a mixed boundary condition for the potential with the derivative specified on free surfaces but the value on the gate. Second, superposition is no longer so straightforward and it is not possible to treat the gate and 2DEG separately; for example, a "solution" for the gate potential alone would be $\phi = V_g$ everywhere, which obeys all the boundary conditions on the surface.

At equilibrium there is a uniform 2DEG of density $n_{\rm 2D}$ and no applied gate voltage, so the common potential of both 2DEG and gate is taken as zero. Changes in response to a voltage applied to the gate can then be superposed on this state, with the boundary condition that the charge on exposed surfaces remains fixed. We only consider the simplest geometries in this section for comparison with a pinned surface: the screened potential and definition voltage of a stripe gate, and the cutoff voltage of a split-gate wire.

A. Linear response to a stripe gate

We can again make a purely electrostatic approximation to solve this, as in Sec. VI; the only difference is the boundary condition $\partial \phi / \partial z = 0$ rather than $\phi = 0$ for |x| > a. A conformal transformation can again be used and requires an elliptic function. The screened potential is found to be

$$\frac{\phi_{\text{scr}}(x,d)}{V_g} = \frac{a_0}{4d} \frac{\pi}{2K(k')} [1 + \text{sech}^2(\frac{1}{2} \pi a/d) \times \sinh^2(\frac{1}{2} \pi x/d)]^{-1/2}, \tag{7.1}$$

where K is the complete elliptic integral of the first kind and $k' = \operatorname{sech}(\frac{1}{2} \pi a/d)$. This is also plotted in Fig. 3(b) and falls off more slowly than the result for a pinned surface because of the square root; it is still exponential but with twice the decay length. Thus a gate on a frozen surface is more effective at depleting a 2DEG than one on a pinned surface. This is because some electric field lines from the gate can terminate on the surface if it is pinned and charge may be induced, but all must lead to the 2DEG and drive out electrons if the surface is frozen. The definition voltage, when the 2DEG under the middle of the stripe is just depleted, is given by $V_g = [(2/\pi)K(k')]V_t$. This reduces to the correct result for a large gate $(a \to \infty)$, as $k' \to 0$ and $K(0) = \pi/2$. Again the effect of the narrow gate is small with $V_g = 1.04V_t$ for a = d.

B. Split gate at cutoff

The electronic structure of a 2DEG under a split gate with a frozen surface was first calculated numerically by

Laux, Frank and Stern¹¹; here we calculate only the electrostatic potential under a split gate at cutoff, when all electrons have just been driven out of the 2DEG. This can be found by supposing three potentials.

Recall that we start with the 2DEG and the gate both at zero potential. First, apply a voltage V_t to the whole surface, both gates and gap. This just depletes the 2DEG, leaving its potential at zero, and sets up a uniform electric field of $-en_{2D}/\epsilon\epsilon_0$ between the surface and 2DEG because we have effectively superposed a charge density of $+en_{2D}$ on the 2DEG. Unfortunately this violates the requirement of frozen charge on the gap between the gates which means that no normal field is permitted there. Another term ϕ_2 must be added to correct this, whose boundary conditions on the surface are $\partial \phi_2/\partial z = -en_{2D}/\epsilon\epsilon_0$ in the gap (|x| < a), and it is convenient to take $\phi_2 = 0$ under the gate (|x| > a). This is a standard example in conformal mapping whose solution is

$$\phi_2(x,z) = \frac{en_{2D}}{\epsilon \epsilon_0} (\text{Re}\sqrt{a^2 - w^2} - z), \tag{7.2}$$

where w=x+iz again. In the mid-point of the channel $\phi_2(0,d)=(en_{2D}/\epsilon\epsilon_0)(\sqrt{a^2+d^2}-d)$. This violates the requirement that the potential be zero at this point, because the last electron has only just left and this point should therefore be at the Fermi energy which we have taken to be zero. Thus a third term is needed and a constant potential of $\phi_3=-\phi_2(0,d)$ is sufficient. This adds to the voltage on the gate, whose final value is

$$V_g = V_t + \phi_3 = \left(V_t + \frac{e n_{2D} d}{\epsilon \epsilon_0}\right) - \frac{e n_{2D}}{\epsilon \epsilon_0} \sqrt{a^2 + d^2}. \tag{7.3}$$

The term in parentheses vanishes if we use the electrostatic approximation $V_t = -en_{2D}d/\epsilon\epsilon_0$, leaving $V_g = V_t\sqrt{1+(a/d)^2}$. This can be compared with the result for a pinned surface, Eq. (5.1). The cutoff voltage is always larger if the surface is pinned, by a factor of $\pi/2$ for wide gaps; this factor should be larger for short wires, which might therefore be better for an experimental test. One might hope that this factor would be large enough for measurements of the cutoff voltage to discriminate between a pinned and frozen surface, but the position remains unclear.

Another important quantity is the curvature $\partial^2 \phi / \partial x^2$ because this sets the separation between the energy levels in a wire which is plotted as a function of a in Fig. 4(a). Confinement is stronger under a frozen surface when the gap is large, despite the smaller cutoff voltage. This is confirmed by the plots of $-e\phi(y,d)$ in Fig. 4(b) for a=200 nm. The potentials for both pinned and frozen surfaces are shown, biassed at their respective cutoff voltages of -1.28 V and -0.83 V.

For comparison, the structure modeled by Laux, Frank and Stern¹¹ using a frozen surface had a=200 nm, $n_{2D}=4.4\times10^{15}$ m⁻² and d=80 nm, allowing 10 nm for the thickness of the 2DEG. The "capacitor" estimate is $V_r=-0.49$ V for the uniform layers. The above formulas then predict the cutoff voltage of the wire to be -2.0 V for a pinned surface and -1.3 V for the frozen case, giving excellent agreement with the numerical result of -1.3 V. It is also interesting to compare our results with recent numerical

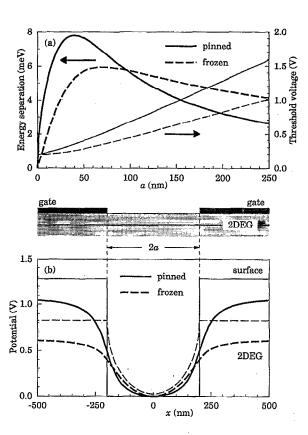


FIG. 4. Conditions at cutoff under a split gate with a gap |x| < a on layers with depth d=50 nm and threshold voltage $V_1=-0.2$ V, comparing results for pinned (solid lines) and frozen surfaces (broken lines). (a) Separation of energy levels at cutoff (thick lines, left scale) and cutoff voltage (thin lines, right scale) as a function of the half-gap a. (b) Confining potential in the 2DEG at cutoff (thick lines) with a=200 nm for pinned and frozen surfaces, with the corresponding potentials on the surface (thin lines). The gates are biased at the cutoff voltages of -1.28 V (pinned) and -0.83 V (frozen surface).

work (Chen et al., ¹⁷ Fig. 5 in particular). These authors considered both pinned (Dirichlet) and frozen (Neumann) surfaces as well as a more realistic model. The potentials on the surface are very similar in form to ours as are those in the 2DEG, where Chen et al. found that it was much easier to remove electrons from a wire if the surface was frozen.

It might be possible to distinguish between the two models for the surface using scanning probe microscopy. The potential across the gap is of course fixed at zero if the surface is pinned, but varies if the charge is frozen. This is also shown in Fig. 4(b). Note that the potential is close to zero in the middle of the gap even for the frozen surface, blurring the distinction between the two cases. A wide gap would be more suitable for this experiment, but the spatial resolution is likely to remain a problem.

VIII. CONCLUSIONS

We have derived formulas for calculating the electrostatic potential due to the most common patterned gates on the surface of a heterostructure, with a general formula for a polygonal gate on a pinned surface. Unfortunately we have not been able to find simple formulas for circular gates. Although these results only provide a rough idea of the occupation of the 2DEG in most cases, they do so for very little effort. The electrostatic potential is also adequate for calculating cutoff voltages in some devices and for finding the energies of collective excitations measured by far infra-red absorption. We shall describe a simple numerical method for finding the self-consistent potential in the 2DEG in a subsequent paper.³⁵ Although our calculations have been restricted to gated structures, the same ideas can also be applied to devices where etching is used to pattern the 2DEG.²

We have also shown how the potential in the 2DEG depends on the behavior of the exposed surface using two simple models. In wires confined by split gates, both the cutoff voltage and the energies of the bound states depend strongly on the surface. Measurements of the cutoff voltage do not appear to be able to resolve this problem, perhaps because the donors are affected by the strong electric fields, but scanning probe microscopy may be successful given adequate spatial resolution. Smaller periodic potentials can be measured from commensurability oscillations in the magnetoresistance, but detailed attempts to fit the observed potential showed that strain was the most likely cause³³ and could not distinguish the behavior of the surface states. The "pinned" and "frozen" models are of course highly simplified cases, and scanning tunneling microscopy³⁶ has shown that a more detailed model may be needed. Our calculations show that experimental input is urgently needed for the accurate modeling of ultra-small structures at low temperature, not just for quantum devices but also for conventional ones like field-effect transistors.

ACKNOWLEDGMENTS

We would like to thank Frank Stern and A. R. Long for many fruitful discussions. Part of this work was performed at the Ohio State University with the support of the U.S. Office of Naval Research, and at the Institute for Theoretical Physics, Santa Barbara, with the support of NSF grant PHY89-04035. The work at Glasgow was supported by UK SERC grants GR/H44714 and GR/J98349, and EVS thanks ISF for partial support.

- ¹C. W. J. Beenakker and H. van Houten, in Solid State Physics, edited by H. Ehrenreich and D. Turnbull (Academic, Boston, 1991), p. 1.
- ² J. H. Davies and A. R. Long, *Physics of Nanostructures* (Institute of Physics, Bristol, 1992).
- ³T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews, and G. J. Davies, Phys. Rev. Lett. 56, 1198 (1986).
- ⁴H. Z. Zheng, H. P. Wei, D. C. Tsui, and G. Weimann, Phys. Rev. B 34,
- ⁵D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C 21, L209 (1988).
- ⁶B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. 60, 848 (1988).
- ⁷M. L. Roukes, A. Scherer, S. J. Allen, H. G. Craighead, R. M. Ruthen, E. D. Beebe, and J. P. Harbison, Phys. Rev. Lett. 59, 3011 (1987).
- ⁸C. J. B. Ford, S. Washburn, M. Büttiker, C. M. Knoedler, and J. M. Hong, Phys. Rev. Lett. 62, 2724 (1989).
- ⁹ A. M. Chang, T. Y. Chang, and H. U. Baranger, Phys. Rev. Lett. 63, 996
- ¹⁰ U. Meirav, M. A. Kastner, and S. J. Wind, Phys. Rev. Lett. 65, 771 (1990).
- ¹¹S. E. Laux, D. J. Frank, and F. Stern, Surf. Sci. 196, 101 (1988).
- ¹²G. L. Snider, I.-H. Tan, and E. L. Hu, J. Appl. Phys. 68, 2849 (1990). ¹³J. A. Nixon and J. H. Davies, Phys. Rev. B 41, 7929 (1990).
- ¹⁴J. A. Nixon, J. H. Davies, and H. U. Baranger, Phys. Rev. B 43, 12638
- ¹⁵A. Kumar, S. E. Laux, and F. Stern, Phys. Rev. B 42, 5166 (1990).
- ¹⁶ K. Lier and R. R. Gerhardts, Phys. Rev. B 48, 14416 (1993).
- ¹⁷M. Chen, W. Porod, and D. J. Kirkner, J. Appl. Phys. 75, 2545 (1994).
- ¹⁸T. Demel, D. Heitmann, P. Grambow, and K. Ploog, Phys. Rev. B 38, 12732 (1988).
- ¹⁹ W. Kohn, Phys. Rev. 123, 1242 (1961).
- ²⁰P. Ruden and G. H. Döhler, Phys. Rev. B 27, 3547 (1983).
- ²¹L. Brey, N. Johnson, and P. Halperin, Phys. Rev. B 40, 10647 (1989).
- ²²P. Maksym and T. Chakraborty, Phys. Rev. Lett. 65, 108 (1990).
- ²³ V. Shikin, T. Demel, and D. Heitmann, Surf. Sci. **229**, 276 (1990).
- ²⁴ J. H. Davies, Semicond. Sci. Technol. 3, 995 (1988).
- ²⁵L. I. Glazman and I. A. Larkin, Semicond. Sci. Technol. 6, 32 (1991).
- ²⁶R. B. Darling, Phys. Rev. B 43, 4071 (1991).
- ²⁷P. M. Mooney, Semicond. Sci. Technol. 6, B1 (1991).
- ²⁸ A. R. Long, J. H. Davies, M. Kinsler, S. Vallis, and M. C. Holland, Semicond. Sci. Technol. 8, 1581 (1993).
- ²⁹G. George and J. R. Hauser, IEEE Trans. Electron Devices 37, 1193
- ³⁰I. A. Larkin and E. V. Sukhorukov, Phys. Rev. B 49, 5498 (1994).
- ³¹I. S. Gradshteyn and I. M. Ryhzik, Table of Integrals, Series and Products (Academic, New York, 1980).
- 32 F. Stern, Phys. Rev. Lett. 18, 546 (1967).
- ³³ J. H. Davies and I. A. Larkin, Phys. Rev. B 49, 4800 (1994).
- ³⁴I. A. Larkin and V. B. Shikin, Phys. Lett. 151, 335 (1990).
- ³⁵ J. H. Davies and J. A. Nixon (unpublished).
- ³⁶M. D. Pashley and K. W. Haberern, Phys. Rev. Lett. **67**, 2697 (1991).