

CHAPTER 2.

Nanostructures and Quantum Transport

Driven primarily by the demand for ever faster and smaller computer chips, there has been great experimental developments and theoretical understandings of sub-micron structures known as nanostructures. The most dramatic feature of such structures is that their functions and behaviors are governed directly by quantum mechanical principles, instead of by some classical theory whose empirical parameters can be calculated quantum mechanically. Lower dimensional systems are made by confining electrons in some directions and by frozen them into the ground quantum level in those directions. Electrons can maintain their phase coherence throughout a sample at achievable temperatures, and very interesting behaviors, such as quantized conductance, universal conductance fluctuations and Aharonov-Bohm effects, show up in transport properties through such a sample. Also, in a structure confined in all directions, the addition or subtraction of a single electron can involve a Coulomb charging energy greater than the temperature, blocking the transport through such a structure. Such a Coulomb blockade effect has been used to manipulate single electrons, leading to single electron turnstiles, pumps and transistors.

I. LOWER DIMENSIONAL SYSTEMS

Lower dimensional models are realized in materials of high anisotropy or by quantum confinement. In the former case, one can regard the material as a collection of weakly coupled sheet or chains, but the system is still three dimensional in essence. In the latter case, the particles are spatially confined into a single sheet, line, or dot with very small lateral size. In this section we will concentrate our attention to this type of lower dimensional systems.

A system can be regarded as truly lower dimensional if the motion in the lateral directions is frozen. In quantum mechanics, this means that the temperature is so low that the condition $k_B T \ll \Delta\epsilon$ is satisfied, where $\Delta\epsilon$ is the typical level spacing in the lateral direction. As a rule of thumb estimate, the gap may be taken as $\Delta\epsilon = \frac{\hbar^2}{2mL^2}$, where L is the lateral size of the wave function. Therefore, achieving lower dimensional system requires the combined effort of reducing the size and temperature. At the liquid ^4He temperature $\sim 4\text{K}$, the above condition requires that $L \ll 2500 \text{ \AA}$ in GaAs where $m = 0.067m_e$, and $L \ll 1500 \text{ \AA}$ in Si where $m = 0.2m_e$.

A. Two dimensional electron gases

If an electron gas is confined in a potential well in one direction and is free in the other two directions, the energy spectrum is given by

$$\epsilon = \epsilon_j + \frac{\hbar^2(k_x^2 + k_y^2)}{2m}, \quad (1)$$

where ϵ_j is the energy levels in the confined direction. Thus, the energy spectrum consists of a series of branches of two-dimensional free particle dispersions (Figure). The density of states is given by

$$g(\epsilon) = \frac{m}{\pi\hbar^2} \sum_j \theta(\epsilon - \epsilon_j), \quad (2)$$

where θ is the unit step function, and we have taken into account of spin degeneracy.

A genuine two dimensional electron gas is realized if all but the lowest branch is empty. In this case, the motion in the perpendicular direction is frozen into the ground state. This requires that the temperature be so low that the condition $k_B T \ll \epsilon_1 - \epsilon_0$ is satisfied. Also, the density of electrons should satisfy $n < \frac{m}{\pi\hbar^2}(\epsilon_1 - \epsilon_0)$, otherwise higher branches must also be occupied. The two dimensional electron gas realized this way can be degenerate ($n\lambda_T^2 \gg 1$) or not.

On the surface of liquid ^4He , electrons can be confined by the potential of their image charges...

On the interface of Si and SiO_2 in a MOSFET device, a thin electron layer can be attracted and confined by applying a gate voltage V_g . The confining potential is a linear potential due to the electric field of the gate voltage which is stopped by a high potential barrier of the insulating SiO_2 . The energy levels can be estimated in the WKB approximation as

$$\frac{2\epsilon}{3\alpha} \sqrt{\frac{2m\epsilon}{\hbar^2}} = (j + \frac{3}{4})\pi, \quad (3)$$

where m is the effective mass in silicon which is about 0.2 electron mass in the vacuum, and α is the slope of the linear potential which is on the order of 10^6 eV/m under typical experimental conditions. This gives a size of the ground state wave function on the order of 100 \AA , and a level separation $\epsilon_1 - \epsilon_0$ on the order of 10 meV. Therefore, at temperatures much smaller than 100 K and electron density smaller than 10^{16} m^{-2} , the electron gas behaves as a *bona fide* two dimensional particle system. Because the electronic density can be conveniently changed by varying the gate voltage, the device is one of the most useful for studying two dimensional particle systems.

Another type of widely used devices are GaAs-AlGaAs heterostructures, where the electrons reside either on an interface as in the MOSFET or on a narrow slab of GaAs confined by AlGaAs blocks known as a quantum-well structure. The electron density cannot be easily changed as in a silicon MOSFET, but it offers a number of advantages. The effective mass is smaller, and there is no complication of valley degeneracies as in the silicon device. Also, by the modulation doping technique which makes the impurities separated spatially from the two dimensional electron gas, the electrons can move over great distances (microns) without suffering a single scattering.

Therefore, although the quantum Hall effect was first discovered in the silicon MOSFET, most of the refined research including the discovery of the fractional quantum Hall effect was done on the GaAs devices.

B. Tunneling Structures and Superlattices

Using the molecular epitaxy technology, one can lay down atoms of choice layer by layer with great precision. The most widely used structures in basic research are those made of GaAs and AlGaAs layers. An AlGaAs layer provides an energy barrier (~ 0.2 eV) for the conduction electrons in GaAs, and the barrier height increases with the concentration of Al. Because of the variability of the barrier height and width, one can realize a number of text book examples of quantum mechanics.

Quantum tunneling at an energy of 0.1 eV below the barrier becomes appreciable when the barrier thickness is about 50 \AA . A resonant tunneling structure that we discussed in section 3.32 can be formed with two such barriers separated by a GaAs region which is thick enough to contain at least one quasi-bound state.

A double well structure can be realized with two slabs of GaAs separated by a thin slab of AlGaAs in between and bounded by thick layers of AlGaAs from the outside. With suitable doping, two sheets of two-dimensional electron gases are formed in each well, and they are coupled by tunneling through the thin barrier. Such a structure has been used in recent years for the double layer quantum Hall effect (Chapter 6).

A superlattice is formed with an periodic array of alternating layers of GaAs and AlGaAs, the realization of a textbook example of solid state physics, the Kronig-Penney model. The levels in each quantum well become energy bands (minibands) due to quantum tunneling through the barriers. The Wannier-Stark ladders (see Chapter 5) was first realized in such a structure.

Other interesting structures have been made, including the Fibonacci quasiperiodic lattice to study exotic spectral and wave function properties, and harmonic potential wells in which three dimensional high mobility electron gases are realized.

C. One dimensional structures

With the molecular beam epitaxy, one can realize a number of textbook examples of one dimensional potentials, but one still needs to consider the free motion in the lateral directions.

quantum wire, quantum annulus, ring, disc, cylinder, quantum dash, quantum sphere, tube, stadium.

D. Quantum dots and anti-dots

II. LANDAUER-BUTTIKKER CONDUCTANCE

One important advance in our understanding of transport theory is that conductance cannot be defined locally. This is especially true for nanostructures where the whole sample including its leads form a region of phase coherence. It is not only meaningless to talk about the conductivity at each point of the sample, it also makes no sense to say the conductance of a structure without indicating how it is connected to the leads. The Landauer-Buttiker formulation of conductance takes particular care in this respect, and provides a simple and powerful framework for quantum transport in nanostructures and mesoscopic systems.

A. Two terminal conductance

Consider a structure consisting of a scattering region connected to two particle reservoirs via two ideal leads, and we wish to study the electric conduction between the reservoirs. In the scattering region, no inelastic scatterings are allowed, but the structure and scattering potentials are arbitrary otherwise. The leads are ideal in the sense that there is no scattering either elastically or inelastically. The reservoirs are maintained at equilibrium within each, whose temperatures and chemical potentials can be separately controlled. The particles emitted from each reservoir are distributed according to the Fermi-Dirac distribution with the reservoir's temperature and chemical potential. The particles going into the reservoirs are perfectly absorbed and become thermalized there.

Therefore, this is a structure where momentum loss and energy dissipation of the particles occur in different places. Such a structure can be easily modeled theoretically as a scattering region connected to scattering free regions extending to positive and negative infinities, which serve as perfect particle reservoirs. Such models have become relevant to reality only in the last decade when clean and well controlled nanostructures are realized.

The electric current that is originated from the left and going to the right can be written as

$$I_L = -e \int_0^\infty \frac{dk}{2\pi} v_k f_L(\epsilon) T(\epsilon) = -\frac{e}{h} \int d\epsilon f_L(\epsilon) T(\epsilon), \quad (4)$$

where $T(\epsilon)$ is the transmission coefficient at energy ϵ , and we have used the velocity formula $v_k = \frac{\partial \epsilon}{\hbar \partial k}$. Subtracting the current originated from the right and going to the left, and using the fact that the transmission coefficient is independent of direction of incidence, we obtain the net current going from the left to the right as

$$I = -\frac{e}{h} \int d\epsilon [f_L(\epsilon) - f_R(\epsilon)] T(\epsilon). \quad (5)$$

This formula is very general, and can be used to analyze nonlinear conductivity and thermoelectric effects. In the following, we will limit ourself to zero temperature and linear conductivity.

At zero temperature, only the states between the two chemical potentials contribute to the net current. For small chemical potential difference, the current satisfies the ohm's law, $I = \sigma V$, where $-eV = \mu_L - \mu_R$, and the conductance is given by

$$\sigma = \frac{e^2}{h} T(\epsilon_f). \quad (6)$$

This is the celebrated Landauer formula, which says that conductivity is transmission. This formula and its generalizations play a central role in the physics of nanostructures and mesoscopic systems.

In a real system, the leads also have different states in the transverse (including spin) directions which serve as different channels of conduction. If the system is separable between the longitudinal direction and the transverse direction, the above derivation for the Landauer formula can be repeated for each lateral state, and the conductance for the total current is given by

$$\sigma = \frac{e^2}{h} \sum_j T(\epsilon_f - \epsilon_j), \quad (7)$$

where the sum is over the lateral states. The sum is really limited to those channels with $\epsilon_j < \epsilon_f$, because otherwise there is no longitudinal states for transmission.

In a heterostructure where the transverse directions form an extended plane, the summation over the lateral states may be replaced by an energy integral weighted by the two dimensional free particle density of states, and the conductance per unit area of the cross section (and per spin direction) is given by

$$\frac{e^2}{h} \frac{m}{2\pi\hbar^2} \int_{-\infty}^{\epsilon_f} d\epsilon T(\epsilon_f - \epsilon). \quad (8)$$

In the more general situation, there is cross transmission between the different channels, $T_{jj'}$, and one may not be able to make a sensible one to one correspondence between the channels on the left lead and those on the right lead. The number of available channels on the two leads may even be different. The transmission coefficient $T_{jj'}$ may be determined, by definition, in the following way. Consider a quantum state at energy ϵ_f , with the boundary condition that there is only a single in-coming wave which is on the left lead and in the j th transverse mode. If we normalize the wave function to a unit in-coming flux, then $T_{jj'}$ is the out-going flux on the right lead and in the j' th transverse mode. The conductance is then given in the general case by

$$\sigma = \frac{e^2}{h} \sum_{jj'} T_{jj'}. \quad (9)$$

In a more elegant way, the sum is sometimes written as the trace of $t^\dagger t$, where t is the matrix transmission amplitudes.

B. Wave guide transport

Conductance quantization. For an ideal wave guide structure in which each conduction channel has transmission coefficient equal to unity, the Landauer formula predicts that the conductance is

equal to $\frac{e^2}{h}$ multiplied by the number of channels whose transverse (and spin) energy is below the Fermi level. In other words, the conductance is quantized. To observe this phenomenon in a real system, the wave guide needs to be smooth and sufficiently short to avoid impurities. Partial reflections by any impurity and sharp feature in the wave guide can deteriorate the quantization. An adiabatic argument shows that a wave guide gradually and smoothly widen up at its ends to particle reservoirs can be used to observe the quantization of conductance. [1]

The observation was first made on a split gate device, where two regions of two dimensional electrons are connected at a very small region known as a point contact. [2] As the Fermi-energy is changed by varying the gate voltage, the conductance shows a series of plateaus whose heights are integer multiples of e^2/h accurate on the level of 1%. The result was somewhat surprising, because the point contact really does not resemble anything like a usual wave guide with an elongated structure. Nevertheless, numerical and analytical calculations showed that clear plateaus of the conductance should indeed appear for a sharp contact between particle reservoirs.[3] Nature is more readily showing her beauty than we can anticipate! Once the phenomenon is established, quantization of conductance is seen in different systems including those occurring at room temperature between tips of scanning tunneling microscopes and sample surfaces.

Interference transistor. The transmission coefficient can depend sensitively on the geometry of the structure in the presence of interference. For example, in the ‘T’ shaped wave guide, the transmission coefficient along the horizontal line oscillates between zero and unity as a function of the length of the side arm. Theoretical calculations show that the period of oscillation is half of the wave length, suggesting that this is due to interference between the amplitude for going directly through the intersection and that making a detour in and out of the side arm. [4] A gate voltage can be applied near the end of the side arm to change its effective length, thereby executing the transistor action of controlling the conductance of the horizontal line. A variant of this idea has been demonstrated experimentally. [5]

Nonlocal conductance. According to the conventional circuit theory, the presence of dead ends (like the side arm in the above example) should not affect the conduction of the circuit. The result on the interference transistor shows clearly that conduction in the regime of quantum coherence is inherently nonlocal. A more vivid and earlier demonstration of this principle was carried out along a narrow wire with a side arm wrapped into a ring. The conductance was found to vary periodically with the magnetic flux going through the ring with a period of h/e . [6]

C. Universal conductance fluctuations

The transmission coefficient depends sensitively on the impurity configurations, which implies that the conductance in the regime of quantum coherence is sample specific. It was shown that even the displacement of a single impurity can cause the conductance to change on the order of e^2/h . The telegraph noise seen on relatively short and narrow wires was explained this way. Somewhere near the wire an impurity can move between two positions, causing the conductance of the wire

fluctuate between two values. The relative duration of the two values scales exponentially with the temperature as $\exp(-\frac{\Delta}{k_B T})$, where Δ presumably corresponds to the energy difference of the impurity at the two positions.[7] In larger samples, $1/f$ noise was predicted and observed because of the movement of many impurities between their metastable positions with distributed values of Δ . [8]

We now estimate the size of the conductance fluctuation due to the displacement of impurities, following an argument due to P. A. Lee [9]. The conductance may be written as

$$\sigma = \frac{e^2}{h} (N - \sum_{ii'} R_{ii'}), \quad (10)$$

where N is the number of channels on the input lead, and $R_{ii'}$ is the reflection coefficient from the i th to the i' th channel. The above expression is a result of current conservation, $\sum_j T_{ij} + \sum R_{ii'} = 1$. For a disordered structure, it is equally probable for an incident particle to reflect to any channels, so that $R_{ii'}$ is on the order of $1/N$. Also, it is evident from numerical simulations that, when the impurities are changed from one configuration to another, the reflection coefficients fluctuate independently for different pairs of incident and reflection channels. Therefore, the root mean square of the conductance is given by

$$\delta\sigma = \frac{e^2}{h} \sqrt{\sum_{ii'} [(\langle R_{ii'} R_{ii'} \rangle - \langle R_{ii'} \rangle \langle R_{ii'} \rangle) + (\langle R_{ii'} R_{i'i} \rangle - \langle R_{ii'} \rangle \langle R_{i'i} \rangle)],} \quad (11)$$

where the second term in the summation is for correlation between the coefficients with mutually reversed incident and reflection channels and should be removed when $i' = i$. Since $R_{ij} \sim 1/N$, each term in the summand is on the order of $1/N^2$, which implies that the size of the conductance fluctuation is always on the order of $\frac{e^2}{h}$ independent of the number of channels. This is known as the universal conductance fluctuation[10].

In practice, it is more convenient to apply a weak magnetic field to achieve the same effect of changing the impurity configuration. The field should be so weak that the Lorentz force on the electrons is negligible, but the Aharonov-Bohm phase between different scattering paths can cause the reflection and transmission coefficients to fluctuate as a function of the field.

D. Coherent back reflection.

Consider the reflection coefficient $\langle |r_{ii'}|^2 \rangle$ averaged over the impurity configurations. For a given impurity configuration, we write the reflection amplitude as a sum of contributions from different scattering paths

$$r_{ii'} = \sum_l A(l). \quad (12)$$

Here we have implicitly assumed a semiclassical picture, which makes sense if the wave length is much shorter than the elastic mean free path, the average distance that the particle travels before two successive events of scattering by the impurities. Because of the disorder in the impurity

positions, the amplitudes for different scattering paths take random phases with respect to one another. We therefore have

$$\langle |r_{ii'}|^2 \rangle = \sum_l \langle |A(l)|^2 \rangle, \quad (13)$$

where the interference terms between different scattering paths are averaged out.

The above arguments have neglected the following fact: the amplitudes for a pair of mutually time reversed paths have exactly the same phase in the absence of a magnetic field. This fact is particularly important for the case of $i' = i$, where for each multiple scattering path l , there is a time reversed path \bar{l} with $A(\bar{l}) = A(l)$. We therefore have

$$\langle |r_{ii}|^2 \rangle = \sum_l [\langle A(l)A^*(l) \rangle + \langle A(l)A^*(\bar{l}) \rangle] = 2 \sum_l \langle |A(l)|^2 \rangle, \quad (14)$$

which says that the back reflection coefficient is enhanced by a factor of two due to constructive interference between time reversed scattering paths.

Such an effect was originally observed in light scattering off a random dielectric medium [11]. There is a sharp peak of the scattering intensity in the direction of back scattering, with the peak height twice as big as the background and with the peak width about $\frac{\lambda}{\ell}$ in radians, where λ is the wavelength of the light and ℓ is the elastic mean free path. Beyond this angular width, the phase difference between time reversed paths becomes comparable to π , and constructive interference is lost on average.

For the electronic transport problem discussed above, the coherent enhancement of back reflection is retained in $\langle |r_{ii'}|^2 \rangle$ even if $i' \neq i$, as long as the difference in the transverse wave lengths for the i th and i' th channels is less than the mean free path, in analogy to the situation of light scattering. This implies that for a lead with transverse size $W < \ell$, coherent enhancement of back scattering is present between all participating channels. For wider leads, there are channels whose transverse wavelengths differ more than the mean free path, between which coherent enhancement of reflection is absent.

Coherent enhancement of reflection reduces conductance on average, and is the fundamental mechanism for a whole range of weak-localization effects (see Chapter 6). Since the phase between a pair of time reversed paths can be modulated by the magnetic field (or flux) due to the Aharonov-Bohm phase, this effect can be observed through the field (or flux) modulation of the conductance (see the next section).

E. multiple-terminal structures

4 terminal multiple terminal Onsager relations

III. AHARONOV-BOHM EFFECTS

A. h/e and $h/2e$ Oscillations

Consider the structure of a ring connected to two ideal leads. We wish to study the two terminal transport through the ring when a magnetic field is applied. For simplicity, we assume that the ring is very thin compared to its diameter, so that the effect of the magnetic field enters only through the Aharonov-Bohm phase. Also, our discussions will be in the language appropriate for the single channel situation, but the results can be generalized to the multiple channel case.

The transmission amplitude t_{12} has two major contributions (to leading order of the transmission coefficients), that via point a and that via point b :

$$t_{12} = t_{1a}t_{a2} \exp[i\frac{e}{\hbar} \int_{C_a} \mathbf{A} \cdot d\mathbf{r}] + t_{1b}t_{b1} \exp[i\frac{e}{\hbar} \int_{C_b} \mathbf{A} \cdot d\mathbf{r}], \quad (15)$$

where C_a and C_b are paths via a and b , respectively. The conductance is then given by $\frac{e^2}{h}$ times

$$|t_{12}|^2 = T_a + T_b + 2\sqrt{T_a T_b} \cos(\frac{e}{\hbar}\phi + \theta), \quad (16)$$

where $\phi = \int_{C_a-C_b} \mathbf{A} \cdot d\mathbf{r}$ is the magnetic flux through the ring, θ is the phase difference between the amplitudes $t_{1a}t_{a2}$ and $t_{1b}t_{b2}$ in the absence of the magnetic field, and $T_a = |t_{1a}t_{a2}|^2$ and similarly for T_b . It is seen that the conductance oscillates with the flux ϕ with the flux quantum h/e being the period. Like the universal conductance fluctuation, the size of the oscillation component is on the order of $\frac{e^2}{h}$, independent of the number of quantum channels. Such oscillations have indeed been observed on the conductance through individual rings of metal or semiconductor nanostructure [12].

The usage of a single ring in such an experiment is important, because the oscillations would be averaged out in an ensemble of rings through the random phase θ . Webb et al have studied arrays of rings, and found that the relative strength of the Fourier component corresponding to the fundamental period of oscillation decays like the inverse root of the number of rings as expected [13]. Such a Fourier component is also absent in the conductance through a thin and long metal cylinder, presumably due to a similar reason [14].

However, the second harmonics survives the ensemble average as was originally predicted by Altshuler, Aronov and Spivak [15]. We need to go beyond the approximation in the last paragraph in order to see the $h/2e$ -period oscillation. The easiest approach is to consider the reflection amplitude in the following expansion

$$r_{11} + t_{1a}t_{ab}t_{b1}e^{-i\frac{e}{\hbar}\phi} + t_{1b}t_{ba}t_{a1}e^{+i\frac{e}{\hbar}\phi} + \dots, \quad (17)$$

where r_{11} is the reflection amplitude without going through the ring, $t_{1a}t_{ab}t_{b1}$ is the amplitude for going around the ring once clockwise. We have separated out the Aharonov-Bohm phase explicitly. When we consider the reflection coefficient, interference terms arise which depend on the flux periodically. The interference between the first and the second (or third) terms gives an oscillation

of h/e period, equivalent to those considered before. Such terms do not survive an ensemble average because of the random phase in $r_{11}^* t_{1a} t_{ab} t_{b1}$ for instance. The interference between the second and third terms gives an contribution equal to twice of the real part of

$$t_{1a} t_{ab} t_{b1} (t_{1b} t_{ba} t_{a1})^* e^{-i2\frac{e}{\hbar}\phi}. \quad (18)$$

Because of the time reversal symmetry in the absence of the field, the coefficient of the exponential in the above expression is equal to $|t_{1a} t_{ab} t_{b1}|^2$, which does not have a random phase and can thus survive the ensemble average. Such interference terms would normally double the back scattering coefficient, but is now modulated by the magnetic field as $\cos(2\frac{e}{\hbar}\phi)$.

B. Scattering by Magnetic Fluxes

In recent years, one can perform transport measurements on two dimensional electrons in the presence of a dilute set of magnetic fluxes of If one deposit a superconductor on the top of a heterostructure

In the classic paper of Aharonov and Bohm [16], the scattering of electrons by an infinitely thin magnetic flux line was calculated. The differential cross section is found to be

$$\sigma(\theta) = \frac{\sin^2 e\phi/2\hbar}{2\pi k} \frac{1}{\sin^2(\theta/2)}, \quad (19)$$

where θ is the angle relative to the forward direction, and k is the wave number of the incident electrons.

C. Persistent Current

Now we consider the Aharonov-Bohm effect in isolated systems, where the energy spectrum is discrete and the energy levels depends periodically and symmetrically on the magnetic flux through each hole of such a system. Since the partial derivative of the Hamiltonian with respect to the vector potential gives e times the velocity operator, the electric current averaged over the length of a ring may be written as

$$I = - \sum_j f_j \langle \psi_j | \frac{\partial H}{\partial \phi} | \psi_j \rangle \quad (20)$$

where f_j is the occupation number of the eigenstate $|\psi_j \rangle$, and ϕ is the magnetic flux through the hole of the ring. Using the fact that $H|\psi_j \rangle = \epsilon_j|\psi_j \rangle$, the expectation value of $\frac{\partial H}{\partial \phi}$ may be written as $\frac{\partial \epsilon_j}{\partial \phi}$, which is known as the Feynman-Herman theorem in the general context of ϕ being any parameter of the Hamiltonian. The current can therefore be written as

$$I = - \sum_j f_j \frac{\partial \epsilon_j}{\partial \phi} = - \frac{\partial F}{\partial \phi}, \quad (21)$$

where F is the free energy of the system.

1. *Ideal gases*

The persistent current for a perfect single channel ring can be calculated analytically. From the energy dispersion $\epsilon_j = \frac{\hbar^2}{2mL^2}(j + \frac{e\phi}{\hbar})^2$ calculated in Chapter 1, the current in the state is found as $-\frac{\hbar}{mL^2}(j + \frac{e\phi}{\hbar})$. The total current at temperature T is given by

$$I(\phi) = \sum_j \frac{eh}{mL^2} (j + \frac{e\phi}{\hbar}) f(\frac{\hbar^2}{2mL^2} (j + \frac{e\phi}{\hbar})^2), \quad (22)$$

where f is the distribution function. To show the periodic dependence of the current on the flux, we can use the Poisson summation formula

$$\sum_j g(j) = \sum_n \int dx e^{i2\pi nx} g(x), \quad (23)$$

which is valid for any function $g(x)$. After some manipulations, the persistent current may be written as

$$\sum_n \sin(ne\phi/\hbar) \frac{2e\hbar}{m} \rho'(nL), \quad (24)$$

where $\rho(nL)$ is the density matrix at the off-diagonal distance $x - x' = nL$ for the ideal gas of infinite size.

At zero temperature, the density matrix for electrons decays algebraically as $\frac{\sin(k_f x)}{k_f x}$, we find that the current is given by

$$I(\phi) = \sum_j \sin(ne\phi/\hbar) \frac{2e\hbar}{m} \frac{\cos(nk_f L)}{nL}, \quad (25)$$

where we have assumed that $k_f L \gg 1$ and thrown away a term of order $(k_f L)^{-1}$ smaller. This is a saw-tooth periodic function of ϕ of amplitude $\frac{I_0 = eh}{mL^2}$. For a ring of $L = 1 \mu\text{m}$, the current amplitude is on the order of 10^{-10} A.

The density matrix in two dimensions decays at large off-diagonal distances as

$$\rho(x) = \frac{k_f^2}{\pi^3} \left(\frac{\pi}{k_f x}\right)^{\frac{3}{2}} (\sin k_f x - \cos k_f x). \quad (26)$$

This leads to a persistent current density around a cylinder as

$$J = J_0 \sum_n \sin(ne\phi/\hbar) \frac{(\sin nk_f L + \cos nk_f L)}{n^{\frac{3}{2}}}, \quad (27)$$

where $J_0 = \frac{2e\hbar}{m} \left(\frac{k_f}{\pi L}\right)^{\frac{3}{2}}$.

At finite temperatures, the density matrix decays exponentially at large off-diagonal distances, and the leading term of the persistent current is given by

$$I(\phi) = \frac{4I_0 T}{\pi T^*} e^{-T/T^*} \cos(k_f L) \sin(e\phi/\hbar), \quad (28)$$

where $T^* = \frac{\epsilon_f}{\pi k_B k_f L}$, assuming that $k_B T \ll \epsilon_f$.

If the electrons were bosons, they would all stay in the ground state at zero temperature, giving rise to a persistent current of $-N \frac{e^2 \phi}{mL^2}$ for $|\phi| < \frac{h}{2e}$ with periodic extensions in other intervals of the flux, where N is the total number of particles. The amplitude of the persistent current is N times larger than that for a single ring of corresponding fermions. At finite temperature, the persistent current is still macroscopic below the critical temperature of Bose-Einstein condensation, and decays exponentially with the length of the ring above the critical temperature. According to the discussions in Section 1.4.5, the exponent of decay is given by $\sqrt{\frac{-2m\mu}{\hbar^2}}$, where μ is the chemical potential which is negative below the critical temperature.

2. Real samples

A persistent current generates a magnetic dipole moment, which can be detected by a highly sensitive SQUID. Since the persistent current is typically very small, and decreases with the ring size, it is preferable to use small rings, which also makes it easier to maintain phase coherence over the size of the sample at an achievable low temperature.

A ring made from a high quality two dimensional electron gas in a GaAs-AlGaAs heterostructure has been used to detect the persistent current [17]. The mean free path of the two dimensional electron gas (about $5 \mu\text{m}$) is comparable to the circumference of the ring ($8 \mu\text{m}$), and the measured signal is compatible with a persistent current I_0 predicted for the ideal model.

An experiment has also been performed on a disordered metal ring, in which the mean free path l_e is about a percent of the circumference L [18]. Theoretical calculations predicted [19] that the persistent current should be reduced by a factor of $\frac{l_e}{L}$ from I_0 , but the measured result shows no such reduction. Electron-electron interaction has been invoked to explain the discrepancy, but controversy still remains.

Experiments have also been carried out on arrays of small metal rings, and a persistent current about $\frac{l_e}{3L} I_0$ per ring has been observed [20]. A very interesting point on statistical mechanics has been raised by this result. In the absence of electron-electron interaction, the average persistent current calculated in the grand canonical ensemble (fixed chemical potential) is exponentially small in the system size e^{-L/l_e} [21]. On the other hand, the average in the canonical ensemble (fixed number of particles) is non-exponential, is $h/2e$ -periodic in the flux and paramagnetic near zero field [22]. The samples are certainly better described by the canonical ensemble, but the theoretical prediction on the size of the persistent current is still an order of magnitude smaller than the experimental results.

With electron-electron interaction, a non-exponential result is obtained even in the grand canonical ensemble [23]. It is known that electron-electron interaction suppresses grand-canonical fluctuations in the particle number, which is intimately related to the Coulomb blockade effect to be described in the next section.

IV. SINGLE ELECTRONICS

A. Charge quantization

Consider a small island weakly coupled to an electrode at chemical potential μ . We wish to know how the average number of electrons and its fluctuations in the island depend on the chemical potential and temperature. In the limit of weak coupling, we may formulate the system as a grand canonical ensemble. The probability for the island to have N electrons is proportional to

$$\frac{\exp[-(F(N, T) - \mu N)]}{k_B T}, \quad (29)$$

where $F(N, T)$ is the free energy of the N electron system.

In the limit of zero temperature, the free energy is just the ground state energy $E_0(N)$ for N electrons. The system overwhelmingly favors N electrons relative to $N - 1$ or $N + 1$ electrons in the island, if the chemical potential lies in the interval $E_0(N + 1) - E_0(N) > \mu > E_0(N) - E_0(N - 1)$. The boundaries of these intervals are the set of ionization levels at which the particle number can fluctuate between two consecutive integers. Therefore, the average number of electrons is a step function of the chemical potential, taking integer values and jumps by one unit when μ passes one of these ionization levels. We call such a phenomena as charge quantization.

At finite temperatures, we can formally define the ionization levels in terms of the free energy as $\{\mu_N = F(N, T) - F(N - 1, T), N = \text{integers}\}$. The physical meaning of these levels is that they are the boundaries of intervals of the chemical potential, in which it is most probable for the system to have a particular number of particles. However, the probability distribution of the particle number is broad if one or more such intervals around the chemical potential can be fit within the thermal energy $k_B T$. Therefore, a necessary condition of charge quantization is that $k_B T$ be much smaller than the spacing between the ionization levels. Under this condition, the system will have N particles with certainty, if $\mu_{N+1} \gg \mu \gg \mu_N$ on the scale of $k_B T$. When μ is far from all but one ionization level μ_{N+1} , the average particle number is given by

$$\langle N \rangle = N + \frac{1}{1 + e^{\frac{\mu_{N+1} - \mu}{k_B T}}}. \quad (30)$$

We now calculate these ionization levels in a few simple models.

Non-interacting electrons – At zero temperature, the ionization levels are the same as the single particle levels, $\mu_N = \epsilon_N$. Charge quantization is possible only when the temperature is much smaller than the level spacing.

Capacitance model – The total energy is given by $E_c N^2 + NV + U_0$, where $E_c = \frac{e^2}{2C}$ with C being the capacitance of the island with respect to the outside world, V is the potential energy due to charges outside the island and background (fixed) charges inside the island, and U_0 is a constant. The ionization levels are given by $\mu_N = (2N - 1)E_c + V$, whose spacing is $2E_c = \frac{e^2}{C}$. Charge quantization is possible when $k_B T \ll \frac{e^2}{C}$.

These models emphasize different aspects of the electron system: quantization of energy levels when the electronic wave function is confined in a small region, and electrostatic interaction between

the discrete electron charges. If L is the typical size of the system, the single particle level spacing goes like $1/L^2$ while the charging energy goes like $1/L$. Because of these different scaling, charge quantization was first seen experimentally in a regime where the Coulomb charging effect dominates over the effect of discrete single particle levels. However, for sufficiently small system sizes the discreteness of the single particle levels can become more important than the charging effect.

Constant-interacting electrons – The total energy is given by the sum of the energies of the occupied single particle levels plus the energy given in the capacitance model. At zero temperature, the ionization levels are given by $\mu_N = \epsilon_N + (2N - 1)E_c + V$, whose spacing is the sum of the single particle level spacing and the charging energy $\frac{e^2}{C}$. Under present experimental conditions, one can achieve small enough metal islands for which the charging energy becomes larger than $k_B T$ below $1K$, but the islands are still sufficiently large that the single particle levels can be treated as a continuum, which is the basis of the orthodox theory. In semiconductor devices, because the combined effect of large dielectric constant and small effective mass, the single particle level spacing is a sizable fraction of the charging energy under present experimental conditions.

B. Capacitance spectroscopy

C. Conductance Spikes

D. Electron pumps

Coulomb Blockade Charge quantization -conductance spikes Electron turnstiles Electron Pumps
Quantum pumps

V. NANO-REFRIGERATORS

Quantum Dots BEER Superconductor

-
- [1] L. I. Glazman et al, JETP Lett. 48, 238 (1988).
 - [2] B. J. van Wees, et al, Phys. rev. Lett., 60, 848 (1988).
 - [3] A. Szafer and A. D. Stone, Phys. Rev. Lett. 62, 300 (1989).
 - [4] F. Sols, M. Macucci, U. Ravaioli, and Karl Hess, Appl. Phys. Lett 54, 350 (1989).
 - [5] D. C. Miller, et al, Nanostructure Physics and Fabrication, edited by M. A. Reed and W. P. Kirk, (Academic Press, 1989) p165.
 - [6] See the next section on Aharonov-Bohm effects.
 - [7] N. Zimmerman, B. Golding, and W. Haemmerle, Phys. Rev. Lett. 67, 1322 (1991).
 - [8] S. Feng, P. Lee, and A. Stone, Phys. Rev. Lett. 56 1960 (1986). N. Birge, B. Golding, and W. Haemmerle, Phys. Rev. Lett. 62, 195 (1989). T. Meisenheimer and Giodano, Phys. Rev. B 39, 9929 (1989).
 - [9] P. A. Lee, Physica, 140A, 169 (1986)
 - [10] B. L. Al'tshuler, JETP Lett. 41, 648 (1985); P. A. Lee, and A. D. Stone, Phys. Rev. Lett. 55, 1622 (1985).
 - [11] E. Akkermans, P. E. Wolf, and R. Maynard, Phys. Rev. Lett. 56, 1471 (1986)

- [12] R. A. Webb, et al, Phys. Rev. Lett. 54, 2696 (1985); S. Washburn et al, Phys. Rev. B 32, 4789 (1985); V. Chandrasekhar et al, Phys. Rev. Lett. 55, 1610 (1985); S. Datta et al, Phys. Rev. Lett. 55, 2344 (1985).
- [13] S. Washburn and R. A. Webb, Adv. Phys. 35 375 (1986).
- [14] D. Y. Sharvin and Y. V. Sharvin, JETP Lett. 34, 272 (1981)
- [15] B. L. Altshuler, A. Aronov and B. Z. Spivak, JETP Lett. 33 101 (1981).
- [16] Y. Aharonov and D. Bohm, Phys. Rev. 115, 485 (1959)
- [17] D. Mailly, C. Chapelier, and A. Benoit, Phys. Rev. Lett. 70, 2020 (1993).
- [18] V. Chandrasekhar et al., Phys. Rev. Lett. 67, 3578 (1991).
- [19] H. F. Cheung, E. K. Riedel, and Y. Gefen, Phys. Rev. Lett. 62, 587 (1989).
- [20] L. P. Lévy, G. Dolan, J. Dunsmuir, and H. Bouchiat, Phys. Rev. Lett. 64, 2074 (1990).
- [21] Y. Gefen and O. Entin-Wohlman, Ann. Phys. 206, 68 (1991).
- [22] H. Bouchiat and G. Montambaux, J. de Physique 50, 2696 (1989); B. L. Altshuler, Y. Gefen, and Y. Imry, Phys. Rev. Lett. 66,88 (1991); A. Schmid, Phys. Rev. Lett. 66, 80 (1991); F. von Oppen and E. K. Riedel, Phys. Rev. Lett. 66, 85 (1991); E. Akkermans, Europhys. Lett. 15 709 (1991).
- [23] V. Ambegaokar and U. Eckern, Phys. Rev. Lett. 67, 3192 (1991).