The capacitance of microstructures

Kenji Natori

Institute of Applied Physics, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

The down sizing of devices has been so rapidly promoted that the device and the circuit feature size will break into the nanometer range less than 1000 Å at the beginning of the next century. It is of critical importance to investigate the characteristics of microstructure capacitances for analysis of future devices and circuits. We analyze the general features of microstructure capacitance based on the self-consistent field approximation, and clarify that it is decomposed into three components. One is the extension of the capacitance of classical perfect conductors discussed in electrostatics. The quantum-mechanical effect plays an important role in the other two components; one of them is proportional to the electronic density of states at the Fermi potential of the conductor in low temperatures. At room temperature this portion is proportional to the electronic charge in the conductor. The other is due to the delocalization of electronic charge off the surface into the bulk of conductor. These two components have only self-capacitance contributions. Various aspects of the microstructure capacitance, e.g., the order of magnitude, the diagrammatic expression, the charging energy, and the quantum dot charging are discussed. © *1995 American Institute of Physics.*

I. INTRODUCTION

Capacitance is a concept originally introduced in the classical electrostatics; however, it found its most popular application in the field of electric circuits. In microsize electronic circuits used in very large scale integrated systems, the inductance within the chip does not play an important role generally, and the circuit characteristics are dominated by the device conductance and the circuit node capacitance. The importance of the capacitance will be conserved in the future ultralarge scale integrated systems where the device is down sized to the range of nanometer scale. In those systems, it is pointed out[1] that the device conductance will undergo a serious modification due to the change of the carrier transport mechanism. The expression of the circuit capacitance will also be seriously modified due to the dominance of quantum-mechanical effects.

Smith *et al.*[2-5] have pointed out that the microstructure capacitance includes the contribution from the electronic density of states, and they have analyzed their experiments with use of the concept. In 1985 and 1986 they presented the expression of capacitance for the two-dimensional electron gas (2DEG) system, and observed quantum oscillations in the magneto-capacitance measurement of the modulation-doped GaAs–AlGaAs heterostructure.[2,3] They reported the result of the capacitance measurement on a quasione-dimensional system[4] in 1987, and also the zero-dimensional system[5] in 1988. Arnone *et al.*[6] reported in 1993 that the amplitude of the magneto-capacitance oscillation dramatically decreases as the system is reconfigured from the 2DEG to isolated quantum dots.

In the field of semiconductor devices, Luryi[7] reported in 1987 that the 2DEG in a quantum well manifests itself as a capacitor connected in series to the classical capacitance of the system. The capacitance component which he called the quantum capacitance was the contribution from the density of states of the 2DEG.

In nanostructures where the characteristic size of the system is comparable to the electron wavelength, it is well known that the conductance shows a specific quantized behavior.[8] As for the capacitance of the nanostructure Büttiker, Thomas, and Prêtre[9] have shown that the experimentally relevant capacitance is the electrochemical capacitance and it consists of the electrostatic capacitance contribution and the density of states contribution serially connected with each other. Hess, Macucci, and Iafrate[10] have discussed the capacitance of the quantum dot in the density functional theory approach. They gave a quantitative discussion based on the many-body theory and reported the detailed characteristics of the quantum dot of their geometry. They have proposed to discuss it with use of the capacitance averaged over the charge increment of *e* in view of the charge quantization.

The nanostructure capacitance is very important in connection with the recent development of the concept of single electron tunneling.[11] There, the magnitude of capacitance is critical for occurrence of the Coulomb blockade phenomenon. In order to develop a detailed discussion that can be compared with the experiment, it is necessary to evaluate the precise value of capacitance reflecting the electronic structure.

This article discusses the capacitance of the general microstructure within the framework of the one-electron theory based on the Hartree and the Hartree–Fock self-consistent field approximation. In Sec. II the microstructure capacitance is decomposed into three components including the counterpart of the classical capacitance. Section III is devoted to the discussion of various aspects in the microstructure capacitance and an example of the single quantum dot is illustrated.
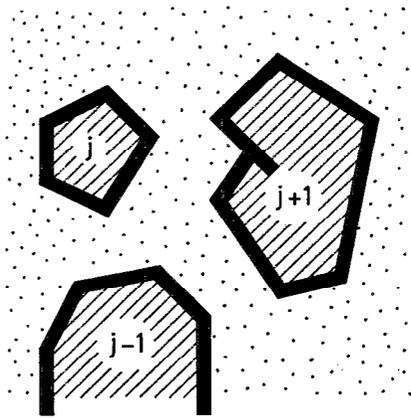
FIG. 1. The investigated system is assumed to consist of a number of microsize bodies of conductors separated from each other by an insulator region.

II. THEORY

We assume a system that consists of a number of microsize bodies of conductors (metals and semiconductors) separated by insulator region from each other, as is exemplified in Fig. 1. When the electric charge and the electric potential of the ith conductor are denoted with Q_i and ϕ_i $(1 \leq i \leq n)$, respectively, the classical electrostatics give that Q_i is related to ϕ_i through the capacitance matrix $\hat{C}$ whose (i,j) component is C_{ij} as

$$Q_i = \sum_j C_{ij} \phi_j. \qquad (1)$$

A counterpart of this expression for microstructures is investigated in this section taking into account the quantum-mechanical effect.

We assume that the system is in static response. Discussion on dynamic response is out of the scope of this article. Classical electrostatics assume that each conductor is a perfect conductor with the constant electric potential throughout the body and the electric charge is localized only on the surface. Every line of electric force from outside terminates on the surface; however, real microconductors consist of neutral atoms, charged ions, and electrons. Ions give the fixed charge and electrons supply the movable charge. The variation of electric charge on the conductor is realized by varying the quantity of electrons filled in the quantum states of the conductor, and it is sufficient to consider only the electronic charge (denoted by q_i) for charging and discharging the capacitance, i.e., $\delta Q_i = \delta q_i$. The charge distribution in the conductor depends on the electronic structure of the filled states. If the electronic potential barrier between the conductor and the insulator is sufficiently large, the electronic wave function nearly vanishes just at the interface and the charge distribution is effectively squeezed into the body of the conductor. Microscopically, the line of electric force penetrates from outside into the conductor. The electric potential in the conductor is not actually constant but varies from surface toward the inside.

The ϕ_j in Eq. (1), the electric potential of the jth conductor, is the energy increment per unit charge when an in-

finitesimal charge is added to the conductor. This means that ϕ_j is equivalent to the electronic chemical potential μ_j of the jth conductor divided by the electronic charge $(-e)$, irrespective as to whether the classical or the quantum-mechanical formulation is used. Thus, we put

$$\phi_j = \mu_j / (-e), \qquad (2)$$

and call ϕ_j the electrode potential of the jth conductor so as to distinguish it from the local electric potential. It is natural that the origins of ϕ_j and μ_j coincide with each other and form the system origin of energy; however, we can use local origins for these quantities any time when convenient.

In classical electrostatics the Coulomb interaction between charges plays an essential role in evaluation of the charging energy, and we should consider the Coulomb repulsion between electrons in our calculation of the capacitance. The rigorous discussion based on the many-body theory generally prefers a complicated numerical calculation on some definite system, which approach is not preferable for our generalized discussion. Here we adopt the self-consistent field approximation by the Hartree or the Hartree–Fock method. In the simplified Hartree approximation, the charge distribution provides the electric potential distribution through the Poisson equation, and the electric potential distribution regulates the electronic states through the Schrödinger equation. A self-consistent solution gives the electronic states in the one-electron picture taking into account the electron–electron interaction in a form of averaged field. In the common algorithm of the Hartree or the Hartree–Fock self-consistent field approximation, the electronic state is provided as a solution of the Hartree equation or the Hartree–Fock equation. The one-electron energy $E_l(N)$ associated with the lth particle state φ_l in an N-particle system is expressed as

$$E_l(N) = H_l(N) + \sum_{m=0}^{N-1} W_{lm}(N), \quad 0 \leq l \leq N-1, \qquad (3)$$

where $H_l(N)$ is the mean value of the single-particle portion in the Hamiltonian formed with φ_l, and $W_{lm}(N)$ is the coulomb integral in the Hartree approximation, or the combined contribution of the Coulomb integral and the exchange integral in the Hartree–Fock approximation. We assume $W_{ll}(N) = 0$ for convenience. It is well known that a simple summation of $E_l(N)$ overestimates the total energy of the system due to double counting of the interaction energy. The chemical potential $\mu(N)$, which is the necessary energy to add the Nth electron to the $(N-1)$-electron system, is expressed by

$$\mu(N) = E_{N-1}(N) + \sum_{l=0}^{N-2} [H_l(N) - H_l(N-1)]$$

$$+ \frac{1}{2} \sum_{l=0}^{N-2} \sum_{m=0}^{N-2} [W_{lm}(N) - W_{lm}(N-1)]. \qquad (4)$$

The summation terms on the right-hand side of Eq. (4) give the energy correction due to the relaxation of existing elec-

Kenji Natori

trons, and are small especially when electrons are degenerate. At finite temperatures we assume that electrons are populated on one-electron energy levels according to the Fermi distribution function, and these levels are self-consistently calculated with use of the distribution. Then the chemical potential of the system is well approximated by the Fermi potential of the distribution.

The independent variables in an n-conductor system are either $n-\phi_i$ or $n-Q_i$. If the ϕ_i are externally set by connecting each conductor to the external electrode, for example, Q_i follow. If Q_i are externally given to each independent conductor, the ϕ_i of each conductor is automatically settled.

The electronic charge of the jth conductor is expressed in the one-electron picture as

$$q_j = (-e) \int D_j(E, Q_1, Q_2, ..., Q_j, ..., Q_n) f(\mu_j, E) dE,$$

(5)

where

$$D_j(E, Q_1, Q_2, ..., Q_j, ..., Q_n)$$

is the density of states of the jth conductor calculated with one-electron energy levels in the self-consistent field. This quantity depends on independent variables Q_j and Q_i ($i \neq j$), where Q_i is the total charge of the ith conductor and consists of the ionic charge and the electronic charge. We assume that the ionic charge is definite and fixed (e.g., charged impurity centers). The density of states depends not only on the magnitude of each Q_i but also on its distribution reflecting the electronic states in the conductor; however, the distribution is automatically settled when the magnitude is given. Note that the origin of energy is not set at each conductor but a single origin is chosen for the whole system. $f(\mu, E)$ is the Fermi distribution function,

$$f(\mu, E) = \frac{1}{1 + \exp[(E - \mu)/kT]},$$

(6)

$(-e)$ is the electronic charge, E is the energy, μ_j is the chemical potential of the jth conductor substituted for the Fermi potential, k is the Boltzmann constant, and T is the temperature. If we concentrate on the jth conductor, the dependence of

$$D_j(E, Q_1, Q_2, ..., Q_j, ..., Q_n)$$

on Q_i ($i \neq j$) is through the value of the electric potential on the surface of the jth conductor $V_j(\mathbf{s})$, where $\mathbf{s}$ denotes the site on the surface. The electronic state in the jth conductor is completely specified if Q_j and the boundary condition $V_j(\mathbf{s})$ at each $\mathbf{s}$ point are given. Therefore, we can rewrite Eq. (5) as

$$q_j = (-e) \int D_j[E, Q_j, V_j(\mathbf{s})] f(\mu_j, E) dE.$$

(7)

Note that $D_j[E, Q_j, V_j(\mathbf{s})]$ is the function of E and Q_j, and is the functional of $V_j(\mathbf{s})$ at the same time. The dependence of

$$D_j(E, Q_1, Q_2, ..., Q_j, ..., Q_n)$$

on Q_j is twofold: the one through the fact that $V_j(\mathbf{s})$ depends on Q_j and the direct one, and the direct one is left in Eq. (7). Here we introduce infinitesimal increments δQ_j which are equal to the increments δq_j, respectively. It accordingly produces the increments $\delta \mu_j$ and $\delta V_j(\mathbf{s})$, and

$$\delta Q_j = (-e) \int \frac{\partial D_j[E, Q_j, V_j(\mathbf{s})]}{\partial Q_j} f(\mu_j, E) dE \, \delta Q_j$$

$$+ (-e) \int \int \frac{\delta D_j[E, Q_j, V_j(\mathbf{s})]}{\delta V_j(\mathbf{s})}$$

$$\times f(\mu_j, E) dE \, \delta V_j(\mathbf{s}) d\mathbf{s}$$

$$+ (-e) \int D_j[E, Q_j, V_j(\mathbf{s})] \frac{\partial f(\mu_j, E)}{\partial \mu_j} dE \, \delta \mu_j, \quad (8)$$

where $\delta D_j[E, Q_j, V_j(\mathbf{s})]/\delta V_j(\mathbf{s})$ is the functional derivative of $D_j[E, Q_j, V_j(\mathbf{s})]$. The value of $V_j(\mathbf{s})$ at each $\mathbf{s}$ is dependent on Q_j and Q_i ($i \neq j$), and we have

$$\delta V_j(\mathbf{s}) = \sum_i \frac{\partial V_j(\mathbf{s})}{\partial Q_i} \delta Q_i.$$

Equation (2) gives $\delta \mu_j = (-e) \delta \phi_j$. With use of these relations, Eq. (8) is transformed into a linear relation of δQ_i and $\delta \phi_i$ for ($1 \leq i \leq n$) as

$$\sum_i [(\hat{C}_D^{-1})_{ji} + (\hat{C}_Q^{-1})_{ji} + (\hat{C}_C^{-1})_{ji}] \delta Q_i = \delta \phi_j,$$

(9)

where the coefficients $(\hat{C}_D^{-1})_{ji}$, $(\hat{C}_Q^{-1})_{ji}$, and $(\hat{C}_C^{-1})_{ji}$ are (j, i) components of the inverse matrices of matrices $\hat{C}_D$, $\hat{C}_Q$, and $\hat{C}_C$, respectively. These matrices are defined with

$$(\hat{C}_D)_{ji} \equiv (-e)^2 \int D_j[E, Q_j, V_j(\mathbf{s})] \frac{\partial f(\mu_j, E)}{\partial \mu_j} dE \, \delta_{ji},$$

(10)

$$(\hat{C}_Q)_{ji} \equiv \frac{e \int D_j[E, Q_j, V_j(\mathbf{s})][\partial f(\mu_j, E)/\partial \mu_j] dE}{\int \{\partial D_j[E, Q_j, V_j(\mathbf{s})]/\partial Q_j\} f(\mu_j, E) dE} \, \delta_{ji},$$

(11)

$$(\hat{C}_C^{-1})_{ji} \equiv \frac{\int \int \{\delta D_j[E, Q_j, V_j(\mathbf{s})]/\delta V_j(\mathbf{s})\}[\partial V_j(\mathbf{s})/\partial Q_i] d\mathbf{s} \, f(\mu_j, E) dE}{e \int D_j[E, Q_j, V_j(\mathbf{s})][\partial f(\mu_j, E)/\partial \mu_j] dE},$$

(12)

where $\delta_{ij}=1$ for $i=j$, and 0 otherwise. Note that the derivative of the density of states in the denominator of Eq. (11) should be evaluated keeping the value of $V_j(s)$ fixed. The variation of the density of states through $V_j(s)$ is included in Eq. (12). If we introduce column vectors δQ and $\delta \phi$ whose components are, respectively, $\delta Q_1, \delta Q_2, \ldots, \delta Q_n$ and $\delta \phi_1, \delta \phi_2, \ldots, \delta \phi_n$, δQ is related to $\delta \phi$ through the capacitance matrix of the system $\hat{C}$, as

$$\delta \mathbf{Q} = \hat{C}\, \delta \boldsymbol{\phi}, \tag{13}$$

with

$$\hat{C}^{-1} \equiv \hat{C}_D^{-1} + \hat{C}_Q^{-1} + \hat{C}_C^{-1}. \tag{14}$$

Matrices $\hat{C}_D$ and $\hat{C}_Q$ are diagonal and consist only of the self-capacitance contributions.

III. DISCUSSION

A. Evaluation of capacitance

Here we discuss and evaluate these capacitance components in order to clarify physical images. When electrons are degenerate at sufficiently low temperatures, and when the density of states can be expanded in the vicinity of μ_j, Eq. (10) is approximated by a power series of kT,

$$
\begin{aligned}
(\hat{C}_D)_{ji} = e^2 \Bigg(& D_j[\mu_j, Q_j, V_j(s)] \\
& + \frac{\pi^2}{6}(kT)^2 \frac{\partial^2}{\partial \mu_j^2} D_j[\mu_j, Q_j, V_j(s)] \\
& + \frac{7}{360}\pi^4 (kT)^4 \frac{\partial^4}{\partial \mu_j^4} D_j[\mu_j, Q_j, V_j(s)] + \cdots \Bigg) \delta_{ji} \\
\approx\ & e^2 D_j[\mu_j, Q_j, V_j(s)] \delta_{ji}.
\end{aligned}
\tag{15}
$$

Note that the dominant contribution to $\hat{C}_D$ at low temperatures is the density of states of carriers at the Fermi surface. This contribution is related to the fact that the carrier is the Fermion and obeys the Pauli principle, as Luryi[7] has pointed out. In order to accommodate more carriers in a limited space of the small conductor, higher-energy states must be exploited lifting the chemical potential. At high temperatures where the carrier distribution obeys the Boltzmann distribution, Eq. (10) is reduced to

$$(\hat{C}_D)_{ji} = \frac{e}{kT}|q_j|\delta_{ji}. \tag{16}$$

The capacitance component is proportional to the electronic charge present in the conductor, and increases as the applied bias voltage is increased to increase the charge.

The $\hat{C}_C$ component in Eq. (12) is evaluated as follows. Instead of $V_j(s)$, we here introduce another set of variables: the mean value of $V_j(s)$, i.e., $V_{j0} \equiv \int V_j(s)ds/\int ds$, and the deviation from the mean value denoted by $u_j(s) \equiv V_j(s) - V_{j0}$. Accordingly, the variables that regulate the value of the density of states shift from E, Q_j, and $V_j(s)$ to E, Q_j, V_{j0}, and $u_j(s)$. [Note that the relation $\int u_j(s)ds = 0$ decreases the degree of freedom by unity and the total number of independent variables is conserved.] $V_j(s)$ is the potential

value on the closed surface of the jth conductor and the uniform shift of this parameter causes the energy shift of all particles within it. This means the variation of V_{j0} causes the shift of $D_j[E, Q_j, V_{j0}, u_j(s)]$ along the E axis, and the V_{j0} dependence of $D_j[E, Q_j, V_{j0}, u_j(s)]$ is effectively expressed as

$$D_j[E + eV_{j0} + \text{const}, Q_j, u_j(s)].$$

Let us examine the effect of variations of δV_{j0} and $\delta u_j(s)$ on the value of

$$\int D_j[E, Q_j, V_{j0}, u_j(s)]f(\mu_j, E)dE.$$

δV_{j0} directly shifts $D_j[E, Q_j, V_{j0}, u_j(s)]$ along the E axis, and changes the effective interval of integration yielding the first-order contribution. However, the contribution of $\delta u_j(s)$ is expressed as

$$\int \int \frac{\delta D_j[E, Q_j, V_{j0}, u_j(s)]}{\delta u_j(s)} \delta u_j(s)ds f(\mu_j, E)dE,$$

and if the functional derivative portion does not strongly depend upon s (which we can expect in many cases) the first-order contribution has the tendency to be cancelled due to the relation $\int \delta u_j(s)ds = \delta \int u_j(s)ds = 0$. Consequently, the dominant contribution to $\hat{C}_C^{-1}$ is transformed to

$$
\begin{aligned}
(\hat{C}_C^{-1})_{ji} = & \frac{\displaystyle \int \frac{\partial D_j[E, Q_j, V_{j0}, u_j(s)]}{\partial V_{j0}} f(\mu_j, E)dE \frac{\partial V_{j0}}{\partial Q_i}}{\displaystyle e\int D_j[E, Q_j, V_{j0}, u_j(s)] \frac{\partial f(\mu_j, E)}{\partial \mu_j} dE} \\
& + \frac{\displaystyle \int\int \frac{\delta D_j[E, Q_j, V_{j0}, u_j(s)]}{\delta u_j(s)} \frac{\partial u_j(s)}{\partial Q_i} ds f(\mu_j, E)dE}{\displaystyle e\int D_j[E, Q_j, V_{j0}, u_j(s)] \frac{\partial f(\mu_j, E)}{\partial \mu_j} dE} \\
\approx & \frac{\displaystyle \int \frac{\partial D_j[E, Q_j, V_{j0}, u_j(s)]}{\partial V_{j0}} f(\mu_j, E)dE \frac{\partial V_{j0}}{\partial Q_i}}{\displaystyle e\int D_j[E, Q_j, V_{j0}, u_j(s)] \frac{\partial f(\mu_j, E)}{\partial \mu_j} dE} \\
= & \frac{\partial V_{j0}}{\partial Q_i}.
\end{aligned}
\tag{17}
$$

In a classical perfect conductor system where the electric potential is constant over each conductor and the charge is localized on its surface, V_{j0} naturally coincides with the electric potential of the conductor and it is trivial that $\hat{C}_C$ gives the conventional capacitance matrix. In the realistic system with the same geometry, there is no guarantee that V_{j0} coincides with the classical electric potential of the conductor and $\hat{C}_C$ does not generally coincide with the classical capacitance value. However, $\hat{C}_C$ gives the capacitance contribution from the insulator regions between those conductors and we can regard it as the extended concept of the classical capacitance. In the case of 2DEG layers such as the inversion layer of the metal–oxide–semiconductor (MOS) junction or the electron gas layer at the interface of the

AlGaAs–GaAs junction, the system is uniform along the layer interface. The electric potential at the surface is expressed by V_{j0} alone and we do not need to introduce $u_j(\mathbf{s})$. $\hat{C}_C$ reduces to a two-body capacitance value C_C and coincides with the classical capacitance.

The expression of the $\hat{C}_Q$ component is also simplified. If we remember that the density of states is expressed as

$$D_j[E,Q_j,V_j(\mathbf{s})]=\sum_l \delta\{E-E_{lj}[Q_j,V_j(\mathbf{s})]\},$$

where $\delta(E-\cdots)$ denotes the δ function and $E_{lj}[Q_j,V_j(\mathbf{s})]$ is the lth electronic energy level, the denominator of Eq. (11) is transformed to

$$\int \frac{\partial D_j[E,Q_j,V_j(\mathbf{s})]}{\partial Q_j} f(\mu_j,E)dE$$

$$=\int \frac{\partial \Sigma_l \delta\{E-E_{lj}[Q_j,V_j(\mathbf{s})]\}}{\partial Q_j} f(\mu_j,E)dE$$

$$=-\int \sum_l \delta\{E-E_{lj}[Q_j,V_j(\mathbf{s})]\}$$

$$\times \frac{\partial E_{lj}[Q_j,V_j(\mathbf{s})]}{\partial Q_j}\frac{\partial f(\mu_j,E)}{\partial \mu_j} dE. \qquad (18)$$

When electrons are degenerate at low temperatures, $\partial f(\mu_j,E)/\partial\mu_j$ has a sharp peak at $E=\mu_j$. If $\partial E_{lj}[Q_j,V_j(\mathbf{s})]/\partial Q_j$ is well behaved for variation of l, its value at $E_{lj}[Q_j,V_j(\mathbf{s})]\approx\mu_j$ is substituted for this factor and we obtain

$$(\hat{C}_Q)_{ji}\approx \left(-\frac{\partial E_{lj}[Q_j,V_j(\mathbf{s})]}{e\partial Q_j}\right)^{-1}_{E_{lj}[Q_j,V_j(\mathbf{s})]\approx\mu_j}\delta_{ji}. \qquad (19)$$

For a more general case, if we define an average of a quantity g over states in the neighborhood of Fermi potential as

$$\bar{g}\equiv \frac{\int D_j[E,Q_j,V_j(\mathbf{s})]g[\partial f(\mu_j,E)/\partial\mu_j]dE}{\int D_j[E,Q_j,V_j(\mathbf{s})][\partial f(\mu_j,E)/\partial\mu_j]dE}, \qquad (20)$$

then Eq. (11) is rewritten as

$$(\hat{C}_Q)_{ji}=\left(-\overline{\frac{\partial E_{lj}[Q_j,V_j(\mathbf{s})]}{e\partial Q_j}}\right)^{-1}\delta_{ji}. \qquad (21)$$

The meaning of this component is clarified with a familiar example. Suppose a metal–oxide–p-type-semiconductor junction is biased with the positive bias on the metal electrode. The semiconductor part includes negative electronic charges, which are trapped at each acceptor atom site forming the localized charge of uniform charge density $(-eN_A)$ where N_A is the acceptor concentration, and give rise to the depletion layer as is illustrated in Fig. 2. The applied bias V_g is assumed to be not so large as to cause the inversion layer. The $\hat{C}_Q$ component for the semiconductor region of unit interface area, which we denote with C_Q simply, is easily evaluated. With use of the depletion layer approximation in evaluating the potential distribution, the energy E of an electron at x is approximated as
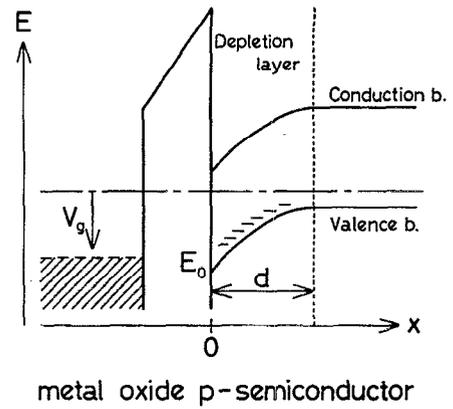

metal oxide p–semiconductor

FIG. 2. Energy band diagram of the metal–oxide–p-semiconductor junction. The gate bias voltage V_g is applied and the charged acceptors near the interface form the depletion layer.

$$E=\frac{e^2}{2\epsilon_s}N_A[d^2-(d-x)^2]+E_0+E_b, \qquad (22)$$

where ϵ_s is the permittivity of the semiconductor, d is the width of the depletion layer, E_b is the binding energy of the acceptor level with respect to the valence-band maximum, and E_0 is the energy of the valence-band maximum at the semiconductor interface. The energy is not the result of a self-consistent field calculation, but it reflects the classical electrostatic potential distribution and the result of Sec. II is applicable. The electronic charge per unit area denoted by q, which is equal to Q, the total charge per unit area here, is approximated by $(-eN_Ad)$, and Eq. (22) leads us to the expression of the density of states per unit area designated by $D(E,Q)$,

$$D(E,Q)=\frac{1}{\sqrt{(2e^2/\epsilon_s N_A)[(Q^2/2\epsilon_s N_A)+E_0+E_b-E]}},$$

$$E_0+E_b\leqslant E\leqslant \frac{Q^2}{2\epsilon_s N_A}+E_0+E_b. \qquad (23)$$

Variation of Q induces the shift of $D(E,Q)$ along the E axis, and with use of

$$\frac{\partial D(E,Q)}{\partial Q}=\frac{\partial D(E,Q)}{\partial E}\frac{(-Q)}{\epsilon_s N_A}=\frac{\partial D(E,Q)}{\partial E}\frac{ed}{\epsilon_s},$$

Eq. (11) yields $C_Q=(\epsilon_s/d)$, the depletion layer capacitance. Note that the contribution of this term to the total capacitance Eq. (14) vanishes as $(d\to0)$, even if $(Q\neq0)$; (i.e., $N_A\to\infty$). The term $\hat{C}_Q$ gives a nonzero contribution when the charge distribution is extended inside the conductor region without localizing on the surface.

In the present case this is fulfilled by the electron trapping at the spatially distributed acceptor site. There are also some other cases that cause nonzero contribution of this term. When the potential barrier at the interface between the conductor and the insulator is high enough, the wave function at the interface vanishes and the electronic charge distribution is effectively pushed inside the conductor region. The electronic interaction through the Coulomb repulsion

and the exclusion effect due to the Pauli principle will also promote the delocalization of the electronic charge. The spatial distribution of the charge gives rise to the spatial distribution of the potential energy, and the charge increment brings about the variation of potential energy distribution pushing up and down the electronic energy levels causing the variation of the density of states.

The $\hat{C}_Q$ component of the MOS inversion charge layer is estimated based on a simplified model. Suppose that the inversion charge is trapped in a triangular potential well at the MOS interface. The energy level E_l of the trapped electron is expressed as[12]

$$E_l = \alpha(l)|\mathbf{F}|^{2/3} = \tfrac{3}{2}e|\mathbf{F}|Z_l, \qquad (24)$$

where $\alpha(l)$ is a constant including the quantum number l, $|\mathbf{F}|$ is the magnitude of electric field constituting the triangular well, and Z_l is the mean separation of the lth energy level charge from the MOS interface. The field distribution near the MOS interface is controlled by the charge distribution there. If we adopt the weighted average of the field value nearby as the effective triangular field, we obtain[1]

$$|\mathbf{F}| = |\mathbf{F}_0| + \frac{|Q|}{2\epsilon_s}, \qquad (25)$$

where $|\mathbf{F}_0|$ is the value of $|\mathbf{F}|$ without the inversion charge and ϵ_s is the permittivity of the semiconductor. Noting that an increment of Q reduces the number of inversion electrons and causes the reduction of $|\mathbf{F}|$, we obtain

$$C_Q \approx \left(-\frac{\partial E_l}{e\,\partial Q} \right)^{-1} = \frac{2\epsilon_s}{Z_l}. \qquad (26)$$

The factor 2 in the numerator of the right-hand side of Eq. (26) is due to the fact that the electric field $|\mathbf{F}|$ that confines the inversion charge is originated from the charge itself.

From the experimental point of view, the magnitude of these quantities is of critical interest and we present the order of magnitude. With use of the density of states of a two-dimensional subband, the magnitude of the $\hat{C}_D$ component for (100) silicon MOS junction at low temperatures is estimated to be $\sim 1.6 \times 10^{14}$ (e/V·cm^2). The magnitude of the $\hat{C}_C$ component $\sim (e_i/t_i)$, where e_i and t_i are the permittivity and the effective thickness of the insulator region, respectively, is $\sim 4.3 \times 10^{12}$ (e/V·cm^2) if we assume that the insulator is SiO$_2$ and $t_i \sim 5$ nm. The $\hat{C}_Q$ component is estimated to be $\sim 5.2 \times 10^{13}$ (e/V·cm^2) with use of the value $Z_l \sim 2.5$ nm for the Si inversion layer.[12] C_D and C_Q are, respectively, about 37 times and 12 times as large as C_C. For a one-dimensional electron gas system exemplified by the quantum wire, electronic states consist of one-dimensional subbands. For simplicity, we assume that the cross section of the wire is circular and an external electrode facing the wire is set across an insulator layer in a concentric geometry. The $\hat{C}_D$ component per unit length for a $\langle 100 \rangle$-oriented Si wire with the radius $r_1 = 5$ nm is estimated to be $\sim 6.9 \times 10^7$ (e/V·cm), when the Fermi potential is at the bottom of the second subband. For other capacitances,

$$C_C \sim 2\pi(r_1 + 0.5t_i)(\epsilon_i/t_i) = 2.0 \times 10^7 \quad \text{(e/V·cm)},$$

and

$$C_Q \sim 2\pi(r_1 - 0.5Z_l)(2\epsilon_s/Z_l) = 1.2 \times 10^8 \quad \text{(e/V·cm)},$$

if we use the same parameters as in the MOS junction. Note that these three contributions have come closer to the same order of magnitude. As an example of the zero-dimensional electron system, we assume a Si spherical quantum dot of the radius $r_2 = 5$ nm covered with both the insulator layer of thickness t_i and the opposing electrode set in a concentric geometry. Electronic states in the dot consist of discrete energy levels and the $\hat{C}_D$ component is approximated by $C_D \sim 2g_v e^2/\Delta E$, where the factor 2 and g_v give the spin and the valley degeneracy, respectively, and ΔE gives the average level spacing. Taking account of the level degeneracy, we obtain $C_D \sim 390$ (e/V). Other capacitance components are

$$C_C \sim 4\pi(r_2 + 0.5t_i)^2(\epsilon_i/t_i) = 30 \quad \text{(e/V)}$$

and

$$C_Q \sim 4\pi(r_2 - 0.5Z_l)^2(2\epsilon_s/Z_l) = 91 \quad \text{(e/V)}.$$

The ratio of C_D to C_C seems to have increased from the quantum wire to the quantum dot but this is because the estimation has not been carried out on the same basis. Anyway, these estimated values suggest that the contribution of $\hat{C}_D$ and $\hat{C}_Q$ will be by no means negligible in future ultrasmall microstructures. So far, we have discussed the case where the conductor is of the semiconductor Si. In case of metallic conductors where the carrier density is very large, the contribution of $\hat{C}_D$ and $\hat{C}_Q$ in Eq. (14) tends to be far smaller due to the large density of states at the Fermi surface as well as the short screening length.

B. Application aspects

When the present result is applied to actual circuit networks, we should be careful that the direct application of Eq. (14) may cause apparent difficulty. If we constitute the matrix $\hat{C}_C$ as an assembly of the two-body capacitance $C_C(i,j)$ and want to evaluate the total capacitance of the system given by Eq. (14), we need to express the matrix element $(\hat{C}_C)_{ij}$ in terms of $C_C(i,j)$. The most simple rearrangement $(\hat{C}_C)_{ii} = \Sigma_{j \neq i} C_C(i,j)$ and $(\hat{C}_C)_{ij} = -C_C(i,j)$ $(i \neq j)$ will lead us to $\det(\hat{C}_C) = 0$ and to the absence of the inverse matrix. This is because the network of the two-body capacitance demands that $\Sigma_i Q_i = 0$ and this has reduced the degree of freedom of the system. To avoid such a situation, we rearrange $\hat{C}_C$ explicitly specifying the potential level of one circuit node (e.g., the ground level) so as to reduce the effective number of the node, but then the equivalence of each circuit node is lost. A simple solution to hold the equivalence consists in the diagrammatic method. Suppose that the matrix $\hat{C}_C$ is constituted with the two-body capacitance and is expressed around the jth node as the left-hand-side part of Fig. 3. The total capacitance is easily evaluated by directly replacing the jth node with the one in the right-hand-side part of the figure. C_{Dj} and C_{Qj} are those components of the jth node defined by Eqs. (10) and (11), respectively. In evaluation of the circuit response, the condition that the hypothetical circuit nodes j' and j'' accommodate no charge, i.e., $Q_{j'} = Q_{j''} = 0$ $(1 \leq j \leq n)$ should be considered.
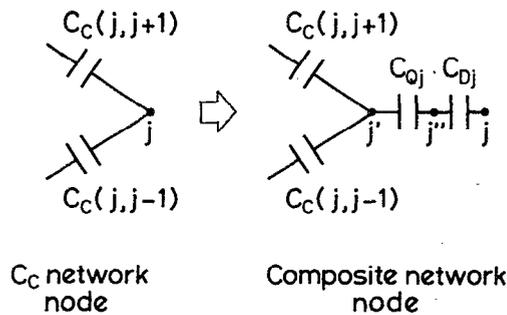
C_C network node

Composite network node

FIG. 3. The capacitance network in an electrical circuit is usually organized with two-body capacitances. When the $\hat{C}_C$ component of capacitance around the jth node is expressed as the left-hand-side part of the figure, the diagrammatic solution consists in replacing it by the right-hand-side part of the figure and computing the capacitance.

FIG. 4. A quantum dot coupled with the classical electrode biased to the ground level.

The charging energy of the capacitor system is briefly discussed. Let us assume that the capacitor is charged up at a constant temperature in equilibrium with the external reservoir. Since the free-energy increment is expressed as $dF = -S\,dT + \Sigma_i \mu_i\,dN_i$, where S is the entropy, the free-energy variation in the isothermal charging is

$$dF = \sum_i \phi_i\,dQ_i. \tag{27}$$

In case the charging process is performed without exchanging the heat, we should apply the idea of the adiabatic charging. The internal energy increment is expressed as $dU = T\,dS + \Sigma_i \mu_i\,dN_i$ and the condition $dS = 0$ leads us to

$$dU = \sum_i \phi_i\,dQ_i. \tag{28}$$

Note that the adiabatic charging generally brings about the variation of the temperature. The charging energy of the system is easily computed with use of these expressions combined with Eq. (13), if the capacitance is given as a function of the electrode potential. Another charging process is possible where the capacitor is charged up with high-energy electrons and the excess energy is dissipated inside the capacitor; but, such a nonequilibrium charging process is out of scope of this article.

Now we examine a specific example of the quantum dot coupled with the classical electrode as is shown in Fig. 4. The capacitance is charged up or discharged by varying the number of electrons in the quantum dot in some way or other. Such a configuration is frequently used in single electron transistor devices.[11] It is also important for ultrasmall circuit application of the present theory. This example includes the δ-function-type density of states, whose aspect is absent in the one-dimensional or the two-dimensional electron gas system and we need a special consideration for it. $T = 0$ is assumed for simplicity. The capacitance is defined here as the differential coefficient of charge as we see in Eq. (13). We first assume that the circuit node is continuously charged up neglecting the fact that the charge is quantized with e. Suppose that the quantum dot has the δ-function-type density of states $D(E)$ with its peaks located at E_0, $E_1, E_2, \ldots$. The local origin of energy is set at the level of the
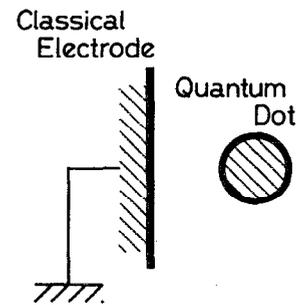
conduction-band bottom averaged over the surface of the dot. The electric potential of this level is denoted with V_0. The system origin of energy is set at the classical electrode. The dot charge (denoted by Q) dependence of the density of states is neglected for simplicity and the contribution from the $\hat{C}_Q$ component is absent. If we assume that the $\hat{C}_C$ component which depends on the geometry is a constant C_C, we obtain $Q = C_C V_0$. Q consists of the fixed ionic charge Q_{ion} and the variable electronic charge q. The value of q is varied by e as each energy level is filled up with an electron. If the chemical potential of the quantum dot μ shares the origin with $D(E)$, the electrode potential ϕ of the dot measured from the system origin is expressed as $\phi = V_0 - \mu/e$. The total capacitance $C = (C_C^{-1} + C_D^{-1})^{-1}$, where $C_D = q^2 D(\mu)$, is plotted as a function of μ in Fig. 5(a). Note that the figure is modified as is shown in Fig. 5(b), if we replot it as a function of ϕ. This is because ϕ increases continuously while a single energy level is filled up with μ fixed at the same level. In actuality, the value of q is varied stepwise by e, and therefore ϕ is varied discretely. For example, if ϕ is decreased from $(Q_{ion} - e)/C_C - E_0/e$ (the level E_0 is just filled) to $(Q_{ion} - 2e)/C_C - E_1/e$ (the level E_1 is just filled) by $|\Delta\phi| = (E_1 - E_0)/e + e/C_C$ and Q is decreased by $|\Delta Q| = e$ the capacitance value observed is not the curve of Fig. 5(b) but the averaged value,

$$C = \frac{|\Delta Q|}{|\Delta\phi|} = \frac{e^2}{(E_1 - E_0) + (e^2/C_C)}. \tag{29}$$

The charging free energy is calculated with Eq. (27). Noting that $dQ = C\,d\phi$ where C is a function of ϕ shown in Fig. 5(b), we can integrate Eq. (27) to give the free energy as a function of ϕ in Fig. 6(a). The value of the free energy at $\phi = Q_{ion}/C_C$ is expressed in a form of $f_0 + (Q_{ion})^2/(2C_C)$ with some constant f_0 for convenience. Figure 6(a) is rearranged as a function of the discrete variable q as is shown in Fig. 6(b).

Let us examine how this discussion will be modified when the Q dependence of the electronic states in the quantum dot is considered. This effect is due to the electron–electron interaction inside the dot and causes the contribution of the $\hat{C}_Q$ component. In the framework of the many-body theory approach also, the concept of the chemical potential is useful and we can express the capacitance with use of it. The

J. Appl. Phys., Vol. 78, No. 7, 1 October 1995

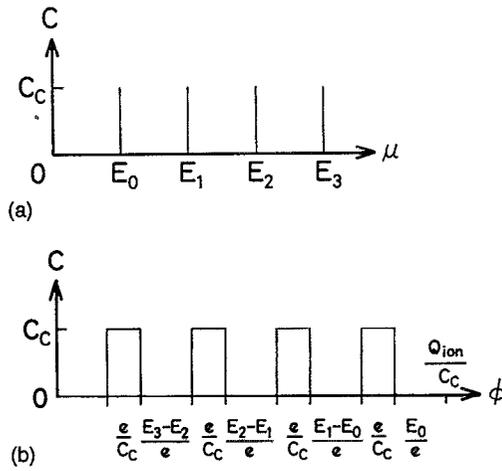Kenji Natori   4549

(a)

(b)

FIG. 5. The capacitance of a quantum dot electrically coupled with the classical electrode: (a) The capacitance as a function of the chemical potential μ measured from the averaged dot surface potential; (b) the capacitance as a function of the dot potential ϕ measured from the ground level.
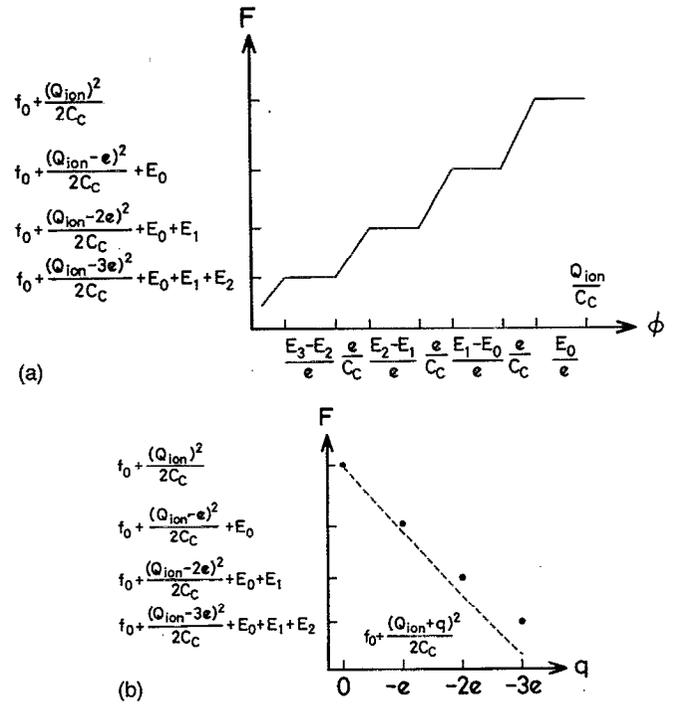

(a)

(b)

FIG. 6. The charging free energy of a quantum dot: (a) The free energy as a function of the dot potential ϕ measured from the ground level; (b) the free energy as a function of the electronic charge on the dot q.

origin of energy is the same as before. The average capacitance during charging from the N-particle state to the $(N+1)$-particle state is

$$C = \frac{|\Delta Q|}{|\Delta \phi|} = \frac{e}{(e/C_C) + \{[\mu(N+1) - \mu(N)]/e\}}. \tag{30}$$

Equation (30) is further transformed in the one-electron picture. We denote the energy level of each one-electron state as $E_0(N), E_1(N), E_2(N),...$ where N particles are included in

the quantum dot. The above $\mu(N)$ in the many-body approximation corresponds to $E_{N-1}(N)$ in the one-electron approximation, and Eq. (30) is rewritten as

$$C = \frac{1}{\dfrac{1}{C_C} + \left(\dfrac{e^2}{E_N(N) - E_{N-1}(N)}\right)^{-1} + \left(\dfrac{e^2}{E_N(N+1) - E_N(N)}\right)^{-1}}. \tag{31}$$

Note that this expression is basically equivalent to Eq. (14). The first term in the denominator of the right-hand side of Eq. (31) shows the contribution of the C_C component. The second and the third terms there, respectively, give the contribution of the $\hat{C}_D$ and the $\hat{C}_Q$ component, as is easily verified by comparing these terms with Eqs. (15) and (19).

IV. SUMMARY

The capacitance of the microstructure is discussed taking account of the electron–electron interaction in the self-consistent field approximation. It was clarified that it consists of three components serially connected. One is the analog or the extension of the classical capacitance, which is basically due to the separation of conductors across the insulator region ($\hat{C}_C$). The other two are due to the quantum-mechanical effects. One of them ($\hat{C}_D$) is proportional to the electronic density of states at the Fermi potential of the conductor in low temperatures. At the room temperature this portion is

proportional to the electronic charge in the conductor. The other one ($\hat{C}_Q$) is the contribution from the spatial distribution of charge in the conductor and is due to the delocalization of charge off the surface into the bulk of the conductor. These two components have only the self-capacitance contributions. In the classical electrostatics, the contributions of $\hat{C}_D$ and $\hat{C}_Q$ vanish leaving $\hat{C}_C$ alone because they are infinitely large [Eq. (14)]. The order of magnitude is estimated and it was shown that the contribution of those quantum-mechanical capacitances cannot be neglected in ultrasmall microstructures made of the semiconductor. The charging energy is briefly discussed and an example of charging a quantum dot is examined.

Kenji Natori

[1] K. Natori, J. Appl. Phys. **76**, 4879 (1994).

[2] T. P. Smith III, B. B. Goldberg, P. J. Stiles, and M. Heiblum, Phys. Rev. B **32**, 2696 (1985).

[3] T. P. Smith III, W. I. Wang, and P. J. Stiles, Phys. Rev. B **34**, 2995 (1986).

[4] T. P. Smith III, H. Arnot, J. M. Hong, C. M. Knoedler, S. E. Laux, and H. Schmid, Phys. Rev. Lett. **59**, 2802 (1987).

[5] T. P. Smith III, K. Y. Lee, C. M. Knoedler, J. M. Hong, and D. P. Kern, Phys. Rev. B **38**, 2172 (1988).

[6] D. D. Arnone, T. J. B. M. Janssen, N. K. Patel, M. Pepper, J. A. A. J. Perenboom, D. A. Ritchie, J. E. Erost, and G. A. C. Jones, J. Phys. Condens. Matter **5**, L1 (1993).

[7] S. Luryi, Appl. Phys. Lett. **52**, 501 (1988).

[8] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhouven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).

[9] M. Büttiker, H. Thomas, and A. Prêtre, Phys. Lett. A **180**, 364 (1993).

[10] M. Macucci, K. Hess, and G. J. Iafrate, Phys. Rev. B **48**, 17 354 (1993).

[11] K. K. Likharev, IBM J. Res. Dev. **32**, 144 (1988).

[12] F. Stern, Phys. Rev. B **5**, 4891 (1972).

J. Appl. Phys., Vol. 78, No. 7, 1 October 1995

Kenji Natori   4551