Quantum capacitance devices

Serge Luryi
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 5 October 1987; accepted for publication 30 November 1987)

Two-dimensional electron gas (2DEG) in a quantum well or inversion layer, unlike an ordinary grounded metallic plane, does not completely screen an applied transverse electric field. Owing to its Fermi degeneracy energy, a 2DEG manifests itself as a capacitor in series, whose capacitance per unit area equals $C_Q = me^2/\pi\hbar^2$, where m is the effective electron mass in the direction transverse to the quantum well. Partial penetration of an external field through a highly conducting 2DEG allows the implementation of several novel high-speed devices, including a three-terminal resonant-tunneling transistor and a gate-controlled thermionic emission transistor.

It is well known that a grounded metal plate completely shields the quasistatic electric fields emanating from charges on one side of the plate from penetrating into the other side. Thus, in a three-plate capacitor, illustrated in Fig. 1(a), application of a voltage to the node 1 changes the electric field only in the space filled with the dielectric ϵ_1. The situation is different if the middle plate Q is made of a *two-dimensional* (2D) metal, like the electron gas (2DEG) in a quantum well (QW) or an inversion layer. In this case, quite generally, the field due to charges on plate 1 partially penetrates through Q and induces charges on plate 2. As will be shown below, the capacitance $C_{\rm tot}$ seen at the node 1 is given by the equivalent circuit of Fig. 1(b), where C_1 and C_2 are the geometric capacitances:

$$C_i = \frac{\epsilon_i}{4\pi d_i}, \quad i = 1,2, \tag{1}$$

and C_Q the "quantum capacitance" per unit area:

$$C_Q = \frac{g_v me^2}{\pi\hbar^2} = g_v \frac{m}{m_0} \times 6.00 \times 10^7 \ {\rm cm}^{-1}. \tag{2}$$

Here m is the effective electron mass in the direction perpendicular to the QW plane, and g_v is the valley degeneracy factor.[1] Thus defined, C_Q coincides with the strong-inversion limit of the so-called inversion-layer capacitance, discussed by Nicollian and Brews[2] and other authors[3] in connection with the carrier-density fluctuations induced by interface charges in a metal-oxide-semiconductor (MOS) system. The quantum capacitance is a consequence of the Pauli principle, which requires an extra energy for filling a QW with electrons. In the classical limit, $\hbar \to 0$ or $m \to \infty$, one has $C_Q \to \infty$, and the capacitances C_2 and C_Q drop out, as they should. For MOS structures on a Si (100) surface, one has $g_v = 2$ and $m = m_l = 0.98 m_0$, so that $C_Q \gg C_1 \equiv C_{\rm oxide}$ at all realistic oxide thicknesses. On the other hand, for a small m one can expect interesting effects when C_Q becomes comparable to the geometric capacitances. Partial penetration of an external field through a highly conducting 2DEG allows the implementation of several novel high-speed devices. Before discussing these devices, it may be worthwhile to go through a derivation[4] of the above equivalent circuit and Eqs. (1) and (2).

Let σ_1, σ_2, and σ_Q be the charge densities, respectively, on electrodes 1 and 2 and in the quantum well. The neutrality condition $\sigma_1 + \sigma_2 + \sigma_Q = 0$ can be written in the form

$$\sigma_2 = -\sigma_1 \sin^2(\phi), \tag{3a}$$

$$\sigma_Q = -\sigma_1 \cos^2(\phi), \tag{3b}$$

where ϕ is a variational parameter to be determined by minimizing the total energy $E_{\rm tot}$ of the system. The latter includes the field energies

$$E_i = \int_0^{d_i} \frac{\epsilon_i F_i^2 \, dx}{8\pi} = \frac{2\pi d_i \sigma_i^2}{\epsilon_i}, \quad i = 1,2, \tag{4}$$

where $F_1 = 4\pi\sigma_1/\epsilon_1$ and $F_2 = -4\pi\sigma_2/\epsilon_2$ are the electric fields in regions 1 and 2, and the Fermi-degeneracy energy

$$E_Q = \pi\hbar^2\sigma_Q^2/2g_v me^2. \tag{5}$$

(Corrections due to electron interaction[1] have been neglected.) Varying $\delta E_{\rm tot}(\phi) = 0$, we find

$$\tan^2(\phi) = \hbar^2\epsilon_2/4mg_v d_2 e^2 \equiv C_2/C_Q, \tag{6}$$

which proves the equivalent circuit of Fig. 1. In particular, the charge induced on the ground metal plate equals

$$\sigma_2 = -\sigma_1 [C_2/(C_2 + C_Q)]. \tag{7}$$

For possible applications an important consideration is by how much does the electrostatic potential Φ_Q of the QW vary in response to a variation of the voltage on electrode 1, at fixed voltages on the QW and electrode 2. This can be described by an "ideality factor,"

$$n \equiv \left(\frac{\partial \Phi_Q}{\partial V_1}\right)^{-1}_{V_Q - {\rm const}} = 1 + \frac{C_Q + C_2}{C_1}. \tag{8}$$

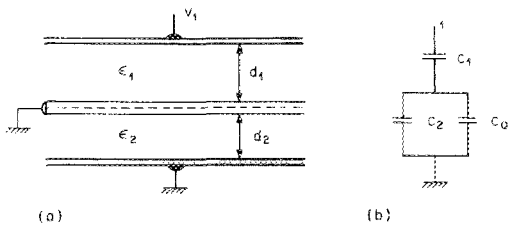

FIG. 1. (a) Schematic illustration of a three-plate capacitor in which the middle plate represents a two-dimensional metal. The space between the plates is assumed filled with dielectrics of permittivity ϵ_1 and ϵ_2. (b) Equivalent circuit for the capacitance seen at node 1.

 0003-6951/88/060501-03$01.00

It should be clearly understood that the quantum capacitance has a different origin from the inversion-layer capacitance in the weak-inversion (subthreshold) regime of a field-effect transistor.[2] The latter capacitance, associated with partial occupation of the inversion-layer states at a finite temperature, vanishes at $T=0$. Motion of the depletion boundary in a subthreshold MOS system corresponds to a trivial penetration of the electric field through an "insulator," whereas the quantum-capacitance effect is a property of a 2D Fermi system.[5] It corresponds to the field penetration through a highly conducting 2D metal.

Let us now discuss possible device applications of this property. Consider a three-terminal device illustrated in Fig. 2, assuming a GaAs/AlGaAs heterostructure implementation. In this structure an undoped GaAs QW is separated by an undoped AlAs barrier of thickness d_1 from a heavily doped n-GaAs gate layer and by a graded undoped $Al_xGa_{1-x}As$ layer of thickness d_2 from an n-doped GaAs emitter layer. (Alternatively, the emitter can be organized as an undoped AlAs layer of thickness d_2 on top of a heavily doped n-AlAs layer.) Electrical contacts are provided to all three layers. For concreteness, let us consider a common-collector configuration. Application of a positive bias V_G to the gate induces 2DEG in the QW, which will serve as a collector. The Fermi level in the QW is fixed at the commom ground level, $E_F/e \equiv V_Q = V_C = 0$. Application of a negative emitter bias V_E gives rise to a thermionic current into the QW over the emitter triangular barrier. Because of the above-discussed penetration of the field through the QW, the thermionic current is controlled by V_G. Like in all "potential effect" transistors,[6] the mutual conductance

$$g_m \equiv \left(\frac{\partial J_C}{\partial V_G}\right)_{V_C=\text{const}} = \frac{\partial J_C}{\partial \Phi}\left(\frac{\partial \Phi}{\partial V_G}\right)_{V_C=\text{const}} \approx \frac{eJ_C}{nkT} \qquad (9)$$

is proportional to the current (J_C is the collector current density) so long as the $J_C(V_G)$ characteristic is exponential. Neglecting variation of the potential across the narrow QW, we can identify the potential barrier height Φ with the QW potential Φ_Q and use (8) for the evaluation of n. The small-signal transistor delay τ is then limited by

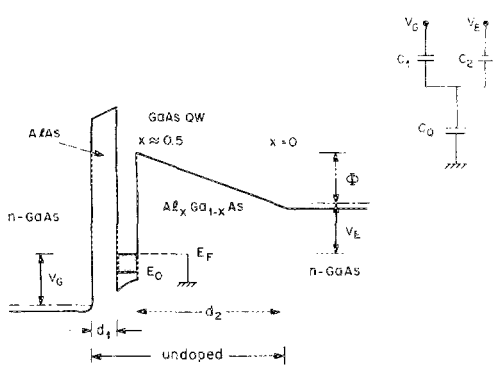

FIG. 2. Cross section of a thermionic quantum-capacitance transistor. Insert shows an equivalent circuit for determining the stored charges in a three-terminal arrangement (neglecting the current flow). The "voltage" drop across the quantum capacitor C_Q equals E_F/e.

$$\tau_{\min} \equiv \frac{C_{\text{tot}}}{g_m^{\max}} = \frac{(C_Q + C_2)kT}{eJ_{\max}} = \frac{d_2}{v_T}\left(\frac{C_Q}{C_2} + 1\right), \qquad (10)$$

where $v_T \equiv (kT/2\pi m)^{1/2}$, and I used the expression[6]

$$J \ll (\epsilon kT/4\pi e d_2)(v_T/d_2) \equiv J_{\max} \qquad (11)$$

for the current at which the $J_C(V_G)$ and $J_C(V_C)$ characteristics lose their exponential nature because of the slowing down of the diffusion velocity on the uphill slope.

The expected performance of the described device should be virtually identical to a conceptually similar device, recently proposed by Kastalsky and Grinberg[7] and called the QW emission transistor or QWET. The structure of QWET is similar to that shown in Fig. 2, but the QW serves as an emitter and the electrode 2 as a collector of thermionic current controlled by the gate via its raising or lowering the QW Fermi level relative to the barrier top. Although the authors of Ref. 7 chose not to discuss their device in terms of the field penetration, it is clear that QWET should work precisely because of this effect. Accordingly, the figure of merit for QWET is similar to that described by Eq. (10), the only difference being that v_T must be replaced by the scattering-limited saturated velocity v_S on the—now downhill—slope d_2. At low temperatures, when $v_T < v_S \approx 10^7$ cm/s, the QWET configuration may be preferable. The most optimistic estimate for GaAs ($C_Q = 4\times 10^6$ cm^{-1}) is obtained taking $d_2 \sim 300$ Å, which gives $\tau_{\min} \sim 4$ ps. Further improvement in speed of the thermionic quantum-capacitance devices is possible if materials with lower effective mass are used for the QW.

Next, consider a three-terminal resonant tunneling (RT) device illustrated in Fig. 3. Again for concreteness, I assume an implementation with the AlGaAs/GaAs heterostructure technology. The structure is similar to my interpretation (Ref. 8, p. 555) of the so-called Stark-effect transistor (SET) proposed by Bonnefoi *et al.*,[9] but the conceptual principle of operation is quite different. The difference is in the effect used for the gate field to control the RT of electrons from the doped emitter layer into the QW collector. In the SET the control is effected via the shift by the gate field of the position of 2D subbands, which are sensitive to the shape of the QW (the Stark effect). For a narrow well, that effect is relatively small compared to the field-penetration effect that I presently propose to utilize.

The structure contains an undoped QW collector, separated by a thin ($d_2 \sim 30$ Å) undoped $Al_xGa_{1-x}As$ tunneling barrier from an n-doped GaAs emitter, and by a slightly

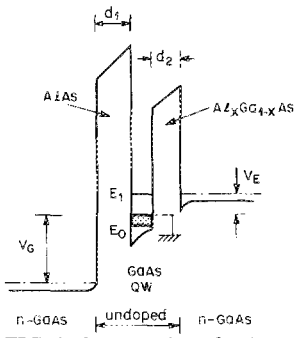

FIG. 3. Cross section of a three-terminal resonant-tunneling device based on the incomplete screening of an external field by a two-dimensional metal.

thicker ($d_1 \sim 60$ Å) undoped AlAs *nontunneling* barrier from an n-doped GaAs gate layer. First, an application of a positive gate bias V_G induces a 2DEG in the QW; subsequently, application of an emitter bias $V_E < 0$ gives rise to characteristics typical of a QW diode.[10] As above, the gate field controls the tunneling characteristics because of the field-penetration effect through the QW. The effectiveness of the gate control depends on an ideality factor n of the structure [Eq. (8)]. The expected transconductance characteristic is well approximated by the current-voltage characteristic of a *symmetric* double-barrier RT diode, with barriers of thickness d_2 but the voltage axis must be scaled by a factor $\xi \approx n/2$. With the assumed geometrical parameters, $\xi \approx 2.6$; again the performance can be improved by using materials with smaller effective mass for the QW. Ideally, as $m \to 0$, one has $\xi \to (d_1 + d_2)/2d_2$.

From these considerations both the collector and transconductance characteristics can be constructed in a straightforward manner. It is clear that the device will exhibit regions of both positive and negative transconductance. As discussed before,[11,8] such a property can have important applications. Indeed, the celebrated usefulness of complementary MOS (CMOS) logic circuits results mainly from the fact that transconductances of p- and n-channel transistors are of opposite sign, which allows high-speed switching combined with low-power dissipation in the steady state. Similar circuits can be obtained from the described RT transistor. For example, a pair of such transistors can perform the logic functions of a CMOS inverter pair.

I wish to express my gratitude to J. R. Brews for useful references, and to A. A. Grinberg and A. Kastalsky for helpful discussions and a preprint of their work.[7]

[1]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[2]E. H. Nicollian and J. R. Brews, *MOS Physics and Technology* (Wiley, New York, 1982).
[3]F. Stern, J. Vac. Sci. Technol. **11**, 962 (1974); J. R. Brews, J. Appl. Phys. **46**, 2181 (1975); E. Arnold, Surf. Sci. **58**, 60 (1976).
[4]This elementary derivation is presented here in the belief that it might elucidate the field-penetration aspect of the problem, important for the applications discussed below.
[5]It is instructive to consider the transition to a 3D case. If more than one 2D subband are occupied, then the value of C_Q must be multiplied by the number of filled subbands (weighted by the Fermi function). The quantum capacitance increases, and in the 3D limit it drops out. In an ordinary metal sheet, however thin, the quantum capacitance associated with one metal surface is so large that the field never penetrates to the other surface. Another way of saying the same thing is that the screening length in a 3D metal is usually much shorter than any realistic film thickness.
[6]S. Luryi and S. M. Sze, in *Silicon Molecular Beam Epitaxy*, edited by E. Kasper and J. C. Bean (CRC Uniscience, Boca Raton, FL, 1987), Vol. 1, Chap. 8, and references cited therein.
[7]A. Kastalsky and A. A. Grinberg (unpublished).
[8]S. Luryi, in *Heterojunction Band Discontinuities: Physics and Device Applications*, edited by F. Capasso and G. Margaritondo (Elsevier, Amsterdam, 1987), Chap. 12.
[9]A. R. Bonnefoi, D. H. Chow, and T. C. McGill, Appl. Phys. Lett. **47**, 888 (1985).
[10]The fact that a *single-barrier* QW structure does give rise to a negative differential resistance characteristic was predicted by S. Luryi, in Technical Digest of the IEEE International Electron Devices Meeting, 2–4 Dec., 1985 Washington, DC, p. 666, and confirmed experimentally by H. Morkoç J. Chen, U. K. Reddy, T. Henderson, and S. Luryi, Appl. Phys. Lett. **49**, 70 (1986).
[11]S. Luryi and F. Capasso, Appl. Phys. Lett. **47**, 1347 (1985); **48**, 1693 (E) (1986).