COMMENTS

Comment on "Quantum capacitance of resonant tunneling diodes"

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

(Received 26 February 1991)

In a recent letter[1] Hu and Stapleton (HS) introduced a quantity they call the quantum capacitance of resonance tunneling (RT) diodes. This quantity, which I shall denote by C_{HS}, describes the modulation of charge δQ stored in the quantum well (QW) in response to a variation of external voltage δV applied to the diode, $\delta Q = C_{HS}\delta V$. HS assert that it is C_{HS} rather the geometric barrier capacitances $C_1 = \epsilon S/d_1$ and $C_2 = \epsilon S/d_2$ that determines the time constant of an RT oscillator. Proceeding to calculate C_{HS} in a simple model of a RT diode, HS find that $C_{HS} \ll C_1, C_2$, and in this context they criticize my earlier work.[2] My problem with the HS letter is not that they had calculated their "quantum capacitance" incorrectly but that this quantity is unrelated to the oscillator time constant. Consequently, the use of C_{HS} by HS is based on a misunderstanding and their results are devoid of a definite physical significance.

Consider the general circuit in Fig. 1, where the resistances R_1 and R_2 are arbitrary functions of the voltages V and V_1. For the purpose of calculating either the current-voltage characteristics or the charge distribution inside the device, this circuit is equivalent to the model used by HS. Indeed, it has been rigorously shown by Payne[3] that the following two calculational procedures produce identical results for the RT diode: One method is to first calculate the energy-dependent tunneling probability for a single-step tunneling process between the emitter and the collector, and then calculate the current by integrating over the energies of the emitter electrons. The other method is to use an effective Hamiltonian for calculating the tunneling probabilities for two sequential steps and then calculate the current by a kinetic equation for the probability density. Even though it has become customary to refer to the former method as "coherent" and the latter as "sequential," it should be emphasized that the use of either method has nothing to do with the question of whether the tunneling process itself is coherent or sequential, which is an issue that can be decided upon only by including scattering processes. Within the "sequential" framework, the equivalent circuit of Fig. 1 is reasonable.

In general, the state of the circuit (e.g., the amount of charge Q stored in the central node) is not uniquely determined by the applied voltage V, cf. the well-known intrinsic bistability effect.[4] Therefore, let us choose an operating regime away from any singular switching point and discuss the response of the circuit to a shock excitation $\delta V\,u(t)$, where $u(t)$ is a step function and δV is sufficiently small so as not to force switching into a different regime. Let us further assume that the current-voltage characteristics are sufficiently smooth so that variations in R_1 and R_2 can be neglected within the range δV. Both assumptions are valid in the situation considered by HS.

The time-dependent response δQ is easy to evaluate by the standard methods. For an arbitrary $t>0$ the result is given by

$$\delta Q(t) = \delta Q_\infty (1 - e^{-t/\tau}), \tag{1}$$

where

$$\delta Q_\infty = \delta V \frac{R_1 C_1 - R_2 C_2}{R_1 + R_2} \equiv C_{HS}\delta V, \tag{2}$$

$$\tau = \frac{R_1 R_2 (C_1 + C_2)}{R_1 + R_2}. \tag{3}$$

We see that C_{HS} determines the magnitude of the stored charge in the steady state but not the small-signal dynamics. The "quantum capacitance" can be positive or negative, but the approach to the steady state is characterized by the usual RC time constant, as discussed in my letter.[2]

The actual evaluation of the lumped elements $R_i(V,V_1)$ and $C_i(V,V_1)$ depends on the model used and has been done by various authors with different degrees of refinement. In this context, the HS letter contains other erroneous assertions. For example, it is claimed that the RC model predicts a maximum operating frequency of 4 GHz for diodes with barrier thickness d ranging from 17 to 50 Å. In fact, there is an exponential dependence of R_1 and R_2 on d. The 4 GHz estimate (corresponding to $\tau \approx 40$ ps) has been made specifically for 50 Å $Al_{0.35}Ga_{0.65}As$ barriers and a 50 Å GaAs QW. For $d=30$ Å, simple estimates[5] predict $\tau \approx 2$ ps (depending also on the QW thickness). These estimates, based on the semiclassical approximation, agree reasonably well with more refined calculations.[6] As far as I know, there is no significant disagreement between the RC model and the experiment.

HS state correctly that the sequential model was introduced to account for the terahertz results of Sollner *et al.*[7] in the detection of external infrared signals by an RT diode. However, they seem to attribute to me a notion that the sequential model is somehow faster than the coherent model. In fact, the maximum oscillating frequency, $f_{max} = 1/(2\pi\tau)$, is limited by Eq. (3) in either model. The terahertz data have been explained by Payne[3] who noted that the RT current response to an external $\delta V(t)$ has a slow fall-off at $f \gg f_{max}$, since $d\delta Q/dt$ is constant for $t \ll \tau$. Moreover, Payne showed,[3] that a high-frequency current fall-off would result in a model that assumes a single en-

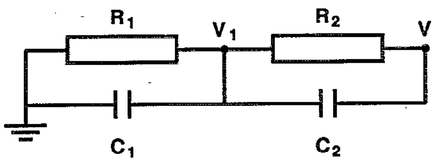

FIG. 1. An equivalent circuit of double-barrier oscillators.

ergy for incident electrons. That model is mathematically *not* equivalent to the two procedures described above and it does not reduce[8] to the equivalent circuit of Fig. 1. Experiments[7] indicate that such a model is invalid. As discussed by a number of authors,[9] in a realistic model the cutoff frequencies for rectification and linear admittance are vastly disparate.

[1] Y. Hu and S. Stapleton, Appl. Phys. Lett. **58**, 167 (1991).
[2] S. Luryi, Appl. Phys. Lett. **47**, 490 (1985).
[3] M. C. Payne, J. Phys. C **19**, 1145 (1986).
[4] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. **58**, 1256 (1987); *ibid.* **59**, 1623 (1987); Solid State Electron. **31**, 731 (1988).
[5] S. Luryi, in *Heterojunction band discontinuities*, edited by F. Capasso and G. Margaritondo (Elsevier Science, Amsterdam, 1987), Chap. 12, p. 550.
[6] Y. Nomura, S. Nara, S. Maruno, M. Gotoda, Y. Morishita, and H. Ogata, Superlatt. Microstruct. **6**, 73 (1989).
[7] T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983).
[8] The absence of a high-frequency response requires that $fI(f)\to 0$ as $f\to\infty$. In the equivalent circuit of Fig. 1, such as effect can be simulated by adding an inductance in the R_1 branch.
[9] V. Kislov and A. Kamenev, Appl. Phys. Lett. **59**, 1500 (1991); and references therein.

Response: The capacitance associated with a device is defined as the derivative of the charges stored in the device with respect to the applied voltage across the device. Such definition gives a constant capacitance $C=\epsilon S/d$ for a parallel plate capacitor, or a voltage dependent capacitance $C=\epsilon S/d(V)$ for a semiconductor *p-n* junction. This, however, does not imply that any plane structured devices can be partitioned into subdivisions so that the total capacitance is the series of the individual capacitor representing each subdivision. This partition may be realized if and only if it can be proved from the definition. For this reason, we criticize the method to use parallel plate capacitor formulas to estimate the capacitance associated with a resonant tunneling diode (RTD), since the accumulation of charges in the quantum well depends on the quantum coherent effects.[1,2] We calculated the capacitance contributed from charges in the well based on the above definition and a damped resonant tunneling model. However, Luryi claims that our results "are devoid of a definite physical significance" and "the use of $C_{\rm HS}$ by HS is based on a misunderstanding."

Luryi's criticism is based on an equivalent circuit model of the RTD (see Fig. 1 in Comment).[3] The construction of this model derives from dividing the RTD into two subdivisions with each subdivision to be represented by a resistor and a capacitor in parallel. However, such dividing has not be previously justified, nor is it obvious from the point of view of coherent tunneling. This model may be justified only when the coherence of tunneling electrons is completely destroyed, which may be from the following factors such as; large width of the well, too many defects, or high temperature etc. In this case, the RTD can be truly regarded as the simple combination of two single barriers, but at the same time the negative differential resistance also vanishes, which is not the case of our interests. Since the model itself lacks foundation, any results derived from such a model will be inconclusive. We also fail to see how Eq. (2) in Comment[3] will yield our results of $C_{\rm HS}$ if R_1 and R_2 literally represent the effective resistance of the two barriers.

Luryi correctly states that $C_{\rm HS}$ describes the response of charges in the well to the change of the external voltage, but claims that $C_{\rm HS}$ has no effect on the time constant of the RTD. In order to answer this question, we need to analyze the charge configuration in the RTD: With an applied voltage V across the diode, there are negative charges Q_a in the accumulation region at the cathode, negative charges Q_w in the well, and position charges Q_d in the depletion region at the anode. Charge neutrality guarantees that the total negative charges $Q_a + Q_w$ is equal to the total positive charge Q_d. Then the total capacitance is given by

$$C=\frac{dQ_a}{dV}+\frac{dQ_w}{dV}=\frac{dQ_a}{dV}+C_{\rm HS}\frac{dV_1}{dV}, \qquad (1)$$

where $C_{\rm HS}=dQ_w/dV_1$ and V_1 is the voltage drop across the double barrier structure (excluding the accumulation and depletion regions). The total capacitance C and the effective resistance R of the RTD determine the time constant $\tau=(RC)^{-1}$. Equation (1) clearly indicates that $C_{\rm HS}$ directly contributes to the time constant of the RTD, though the relative contribution of $C_{\rm HS}$ depends on the ratio of $C_{\rm HS}$ and dQ_a/dV, the latter may be approximated by a parallel plate capacitor with d measuring the total width of the RTD from the accumulation region to the depletion region.

We owe Luryi an apology for our improper wording on the 4 GHz estimate frequency limit and we should have not written the sentence "As an alternative, the sequential tunneling... ." in Ref. 2. More careful wording will be used in our future publications. However, we would also like to point out an incorrect assertion made by Luryi that we claim $C_{\rm HS}$ "determines the time constant of an RT oscillator." It is the total capacitance C, which is under the current investigation, that determines the time constant.

Yuming Hu and Shawn Stapleton
School of Engineering Science, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada

(Received 26 March 1991)

[1] S. Luryi, Appl. Phys. Lett. **47**, 490 (1985).
[2] Y. Hu and S. Stapleton, Appl. Phys. Lett. **58**, 167 (1991).
[3] S. Luryi, Appl. Phys. Lett. **59**, 2335 (1991).