Nonlocal Transport and Cuspidal Surface Mobility in Semiconductors

R. F. Greene
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received 1 February 1963; revised manuscript received 17 April 1963)

The curious cuspidal behavior of the surface mobility formulas of Greene, Frankl, and Zemel for nearly flat bands has been emphasized by Flietner, who also noted the omission of a size-effect transport term. He gave revised formulas showing higher surface mobility values and no cusp. We show that Grenee, Frankl, and Zemel actually omitted a *pair* of size-effect terms, that these very nearly cancel, and that, therefore, the corresponding thick-sample surface mobility formulas are correct. We also show that the cusp is a characteristic of transport in the nonlocal case (mean free path comparable to Debye screening length) where band bending cannot be treated as a small perturbation.

1. INTRODUCTION

The curious behavior of the surface mobility μ_S in the theory of Greene, Frankl, and Zemel[1] (hereinafter referred to as GFZ), whereby μ_S goes through a strong downward cusp as the surface potential excess $\Delta_S u$ goes through zero, has been emphasized recently by Flietner.[2] He concludes that the cusp is an artifact resulting from the omission of a size-effect term and that μ_S should really vary smoothly near $\Delta_S u=0$ with values much closer to the bulk mobility μ_B.

We shall show that this size-effect term is actually only one of a *pair* of terms, which cancel almost exactly. The omission of both terms by GFZ on intuitive grounds is, therefore, justified, as are the corresponding thick-slab mobility formulas.

Far from being an artifact, the cusp is a characteristic feature of transport in the nonlocal case: mean free path λ comparable to bulk screening distance L_B. It is pointed out that the exact theory provides that $\partial(\mu_S/\mu_B)/\partial\Delta_S u$ suffer a finite positive discontinuity $\frac{1}{2}\pi(\lambda/L_B)^3$ at the cusp. Such singular behavior cannot be obtained by analytic perturbation theory.

By independent arguments we show how the perturbation treatment breaks down over the region of velocity space, which dominates the conductance changes near flat band.

The value of μ_S at the *exact* flat-band point is, to be sure, somewhat unphysical, inasmuch as the accuracy of measurement of μ_S goes to zero there with that of the surface excess of current. Indeed, this happens at finite values of $\Delta_S u$ because of current noise, nonuniformities in $\Delta_S u$ over the surface, etc. Nevertheless, the general cuspidal behavior is important because its effects may be felt even for $|\Delta_S u|\gtrsim 1$.

2. EXACT SURFACE MOBILITY FORMULAS

The surface mobility μ_S for electron transport along the surface of a semiconductor slab of thickness $2d$ is defined[1] in terms of the surface excess of electrons ΔN,

$$\Delta N=\int_0^d dz(n-n_B), \tag{2.1}$$

and of the surface excess of current (I_x-I_{x0})

$$I_x-I_{x0}=\Delta N e\mu_S E_x \tag{2.2}$$

$$I_{x0}=n_B e\mu_B E_x\left\{d-\lambda+\lambda\int_0^\infty d\nu\,\exp\left(-\nu-\frac{d}{\lambda(\pi\nu)^{1/2}}\right)\right\}. \tag{2.3}$$

Here n_B and μ_B are the bulk electron concentration and mobility, respectively, and I_{x0} is the flat-band current, in the half-region $0<z<d$, produced by a uniform electric field E_x parallel to the surface. λ is the mean free path $\tau(kT/2\pi m)^{1/2}$ in a theory with constant bulk relaxation time τ.

Flietner[2] pointed out that the omission of a small term in (I_x-I_{x0}) would cause an error in μ_S which would diverge at the flat-band point where $\Delta N\to 0$. Noting that the size-effect term in I_{x0} [viz., the integral in (2.3)] was omitted in GFZ, he concluded that the GFZ values for μ_S would be quite incorrect near $\Delta_S u=0$.

It is not hard to show that a second size-effect term exists in I_x, and that the difference between the two terms is quite small, vanishing more strongly than ΔN at $\Delta_S u=0$. The omission of both terms by GFZ on intuitive grounds is, therefore, justified. (The cancellation is inexact when $\lambda\gtrsim d$, however.) This can be seen from the exact solution [Eq. (3.4) of GFZ] of the Boltzmann equation for arbitrary d. Treating only the accumulation case, for brevity, we obtain

$$I_x=n_B e\mu_B E_x\left\{d-\lambda+\Delta N/n_B-[\exp(\Delta_d u)-1]\right.$$
$$+\lambda\int_{-\Delta_d u}^\infty d\nu\,\exp[-\nu-2K(v_{zd},v_{zS})]$$
$$\left.+\lambda\int_{-\Delta_S u}^{-\Delta_d u} d\nu\, e^{-\nu}[\exp 2K(v_{zS},0)-1]\right\}, \tag{2.4}$$

Here $u=\ln(n/n_i)$, $\Delta u=u-u_B$, $\Delta_d u=u_d-u_B$, and K and ν are as defined in Eq. (2.2) of GFZ. Since $2K(v_{zd},v_{zS})\to -d/\lambda(\pi\nu)^{1/2}$ as $\Delta_S u\to 0$, the term emphasized by Flietner as being in I_{x0} is clearly also present in I_x, in somewhat more general form.

[1] R. F. Greene, D. R. Frankl, and J. N. Zemel, Phys. Rev. 118, 967 (1960).
[2] H. Flietner, Physica Status Solidi 1, 484 (1961).

From (2.2), (2.3), and (2.4) exact expressions for μ_S, valid for any value of d, may be obtained. The exact expression for μ_S then differs from the GFZ thick-slab expression [(2.10) of GFZ] by the following additional *pair* of terms:

$$\delta\left(\frac{\mu_S}{\mu_B}\right)=\frac{\lambda n_B}{\Delta N}\left\{\int_{-\Delta_d u}^{\infty} d\nu \exp[-\nu-2K(v_{zd},v_{zS})]\right.,$$
$$\left.-\int_0^{\infty} d\nu \exp\left(-\nu-\frac{d}{\lambda(\pi\nu)^{1/2}}\right)\right\}. \quad (2.5)$$

To show that this pair of terms nearly cancels, it is only necessary to examine the region $|\Delta_S u|\ll 1$ in which

$$\Delta u=\Delta_S u \cosh\left(\frac{d-z}{L_B}\right)\Big/\cosh\frac{d}{L_B}, \quad (2.6a)$$

$$\Delta N=n_B L_B \Delta_S u \tanh(d/L_B), \quad (2.6b)$$

where

$$L_B=L_D/(\cosh u_B)^{1/2}=(\kappa kT/8\pi n_i e^2)^{1/2}(\cosh u_B)^{1/2}. \quad (2.6c)$$

In this region, also, Eqs. (2.2a) and (2.2f) of GFZ give

$$2K(v_{zd},v_{zS})$$
$$=\frac{1}{\lambda[\pi(\nu+\Delta_d u)]^{1/2}}\left\{d-\frac{L_B\Delta_d u}{2(\nu+\Delta_d u)}\left[\sinh\left(\frac{d}{L_B}\right)-1\right]\right\},$$
$$\nu>-\Delta_d u. \quad (2.7)$$

Then, when $d/L_B\gtrsim 3$, (2.5) gives, for the flat-band point

$$\delta\left(\frac{\mu_S}{\mu_B}\right)=\left(\frac{\lambda}{d}\right)\int_0^{\infty} d\nu \exp\left[-\nu-\frac{d}{\lambda(\pi\nu)^{1/2}}\right]$$
$$\ll\frac{\lambda}{d}, \quad \text{when } d/\lambda\lesssim 3. \quad (2.8)$$

Since $\delta(\mu_S/\mu_B)$ is greatest near the flat-band point, (2.8) shows that the thickness-dependent corrections to the mobility formula (2.10) of GFZ are, indeed, negligible.

3. FAILURE OF PERTURBATION THEORY FOR NONLOCAL TRANSPORT

The cusp is a property of the *exact* solution of the Boltzmann equation boundary value problem. One can show, in fact, from Eq. (2.10) of Ref. 1 that the discontinuity in slope of μ_S/μ_B has the finite positive value

$$\left.\frac{\partial\mu_S/\mu_B}{\partial\Delta_S u}\right|_{\Delta_S u=0^-}^{\Delta_S u=0+}=\lim_{\Delta_S u\to 0}\frac{\lambda}{L_B\Delta_S u^2}$$
$$\int_{-\Delta_S u}^{-\Delta_d u} d\nu \exp[-\nu-2K(v_{zS},0)]$$
$$=\frac{\pi}{2}\left(\frac{\lambda}{L_B}\right)^3. \quad (3.1)$$

Thus, the cusp is closely associated with the long mean free path or nonlocal aspect of the transport.

In a perturbation treatment of the flat-band region one would assume that f_1 could be expanded as a power series in $\Delta_S u$. But this would mean that μ_S could also be so expanded, because μ_S is a linear functional of f_1. Since a cusp *does* exist, we must conclude that no analytic perturbation treatment of the flat-band region when $\lambda\sim L_B$ will work.

Independent of the above arguments, one can also see in detail how a perturbation treatment fails by examining the change in f_1 produced, in the extreme case $\lambda\gg L_B$, by a small attractive potential. Consider electrons approaching a scattering surface with small *normal* velocity v_z. In the flat-band case, these electrons do not yet 'know' of the surface, and so for them f_1 has the bulk value f_{1B}. But when there is a weak accumulation layer, electrons with this normal velocity will have had their last scattering on the physical surface and will, therefore, have an f_1 close to zero. [These properties of f_1 are easily seen in the exact solutions (2.7) of Ref. 1.] Thus, f_1 is strongly altered in a certain region of velocity space by a small change in $\Delta_S u$.

Finally, it is important to realize that the velocity space region of strongly altering f_1 dominates the conductance change near flat-band: f_1 changes strongly for all electrons having *normal* velocity $|v_z|<(2kT\Delta_S u/m)^{1/2}$, when $\lambda\gg L_B$, and there are $\sim 4n_B L_B(\Delta_S u/\pi)^{1/2}$ such electrons. This number is greater even than the total carrier excess $\Delta N\sim 2n_B L_B\Delta_S u$. This means, of course, that surface scattering reduces the current contribution not only for the excess carriers, but also for much of the flat-band population in the surface region.

ACKNOWLEDGMENTS

The author would like to thank J. N. Zemel and R. Bartlett for several stimulating discussions.