QUASI-ELECTRIC AND QUASI-MAGNETIC FIELDS IN NONUNIFORM SEMICONDUCTORS*

By

Herbert Kroemer

RCA Laboratories,
Princeton, N. J.

Summary—*In a chemically uniform semiconductor with a constant impurity concentration and without elastic strains, the forbidden band has a constant width, and there are no internal electric fields. A nonhomogeneous impurity distribution introduces electric fields, while nonuniform elastic strains introduce a nonuniform bandwidth. If the semiconductor is an alloy, such as a germanium–silicon alloy, a change in the alloy composition also changes the bandwidth.*

A change in bandwidth means gradients of the band edges which are different for the conduction and the valence bands. These gradients act upon the electron and hole movement as though they were electric fields, but, because the two slopes are different, these "quasi-electric" fields are not the same for electrons and holes, contrary to the case of real electric fields.

In addition, a type of "quasi-magnetic" field is produced when an inhomogeneity produces a shift of the location within the Brillouin zone of the energy minimum of the band. These quasi-magnetic fields not only are different for electrons and holes but they are also different for the electrons (holes) inside the various energy minima of that band.

Inhomogeneous Semiconductors

THIS paper deals with the electron movement in semiconductors in which the crystal potential is not perfectly periodic but where, instead, the *shape* of the atomic potential changes gradually from cell to cell if one proceeds through the crystal, as illustrated in Figure 1. It is this type of inhomogeneity to which we refer in this paper when we speak of "inhomogeneous" semiconductors. An example of an inhomogeneous semiconductor in this sense is a semiconductor under nonuniform elastical strains. Another example is an alloy of several semiconductors such that the composition of the alloy changes gradually throughout the crystal (e.g., a nonuniform germanium–silicon alloy). Actually, the atomic potential in such an alloy does not vary continuously from one cell to the next, but rather in a discontinuous and random fashion. In first-order considerations, however, it is permissible to neglect those statistical fluctuations by treating a

* This paper was presented at the Symposium on "The Role of Solid State Phenomena in Electric Circuits," in New York, April 24, 1957, and is included in the records of this symposium.

"local average" of the potential, which may be formed by averaging the actual potential over a certain number of cells adjoining the cell under consideration. In a nonuniform alloy this local average will then represent a slowly varying atomic potential as in the case of elastical deformation.

It is believed that nonuniform semiconductors in the sense used here are much more common than would be indicated by these two examples. Lattice discontinuities, like surfaces, are believed to lead to potential deformations in their neighborhood. Polarizations of the electronic orbits due to electric fields or due to high densities of injected carriers also belong in this class, and there may be other types. In many cases the inhomogeneity effects may be small compared to other effects so that they are overlooked in experiments not designed to show

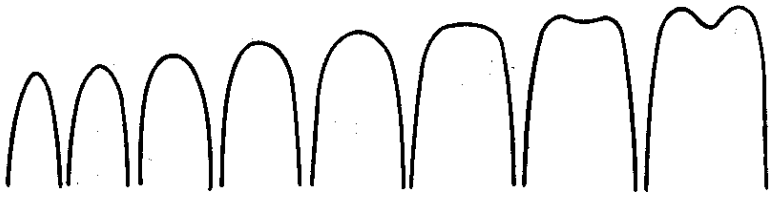

Fig. 1—An example of a nonuniform crystal potential.

them up. The purpose of this paper is to outline the general effects upon the electron movement of inhomogeneous potentials, regardless of their origin, and of the strength of the effects compared with other effects that might be present simultaneously.

QUASI-ELECTRIC FIELDS

In a nonuniform semiconductor one has essentially a different kind of semiconductor in every portion of the crystal. Different semiconductors, in general, have different widths of the forbidden energy band, so the nonuniform semiconductor leads to the concept of a nonuniform band gap. The question immediately arises as to whether this is a legitimate concept, i.e., whether the following three conditions are satisfied:

(1) A local density of states can still be defined, and the distribution of the states over the various energy levels can be described by *sharp* band edges and by effective mass tensors.

(2) The band gap and the mass tensors at each point are the same, as in a homogeneous crystal which has the same

atomic potential throughout as that which the inhomogeneous crystal has at the given point.

(3) The dynamics of the electron are still governed by Newton's Law, using as the mass of the electron the above effective mass and as the force upon the electron the slope of the edge of that band to which the electron belongs.

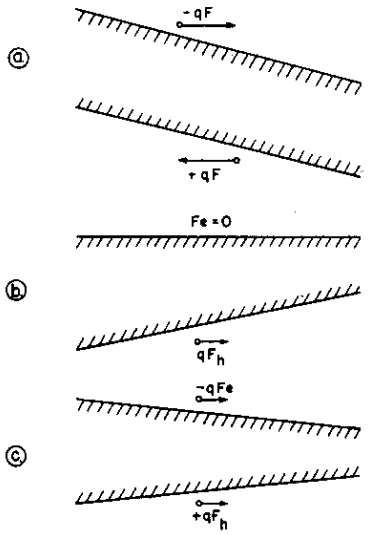

Fig. 2—(a): Effect of a true electric field; (b) and (c): Effects of quasi-electric fields.

The third of these conditions is the most interesting one. In a homogeneous semiconductor the band slope under an external field is the same for all bands and, as a result, the forces upon electrons and holes are equal in magnitude and opposite in direction. This is not the case with a varying band gap. If the concept of a varying band gap is legitimate, the forces would no longer be equal and opposite. It should, for example, be possible to have a force acting only upon one kind of the carriers, or to have a force which acts in the same direction for both (Figure 2). Electrical forces in uniform crystals can never do this. This is why we call these forces "quasi-electric." They present a new degree of freedom for the device designer to enable him to obtain effects with the quasi-electric fields that are basically impossible to obtain with ordinary circuit means involving only "real" electric fields. Two simple examples are given in the next section.

The principal question as to the correctness of the above three conditions and, therefore, the reality of the quasi-electric fields, has been answered in the affirmative; Bardeen and Shockley[1] have shown in their treatment of the electron-lattice scattering via deformation potentials that the concept of a variable band gap is legitimate in the above sense, provided the nonuniformity arises from nonuniform elastical deformation. The present author[2] has extended their proof and has shown that the concept holds true regardless of the shape or the origin of the variation of the atomic potential, provided this variation is a sufficiently gradual one.

Mathematically, this result can be expressed as an extension of the Wannier–Slater[3] Theorem: Assume that $E_{\vec{k}}(\vec{x})$ is the energy of an electron with the wave vector $\vec{k}$, which moves in an exactly periodic potential of such a shape as exists in the nonuniform crystal at the position $\vec{x}$ inside the crystal. Then an operator $E_{-i\nabla}(\vec{x})$ can be derived from $E_{\vec{k}}(\vec{x})$ by replacing $\vec{k}$ with the operator $-i\nabla$. The behavior of the electron can then be described by a Wannier–Slater wave equation:

$$E_{-i\nabla}(\vec{x})\Phi(\vec{x}) = E\Phi(\vec{x}), \tag{1}$$

where $\Phi(\vec{x})$ is the familiar Wannier–Slater amplitude function for the electron.†

In a semiconductor with a single energy minimum at $k=0$,

$$E_{\vec{k}}(\vec{x}) = \frac{\hbar^2 k^2}{2m^*} + E_B(\vec{x}), \tag{2}$$

where $E_B(\vec{x})$ is the position-dependent‡ band edge. This leads to

---

[1] J. Bardeen and W. Shockley, "Deformation Potentials and Mobilities in Non-Polar Crystals," *Phys. Rev.*, Vol. 80, p. 72, October, 1950.

[2] H. Kroemer, "Band Structure of Semiconductor Alloys with Locally Varying Composition," *Bull. Amer. Phys. Soc.*, Vol. 1, p. 143, March, 1957.

[3] J. C. Slater, "Electrons in Perturbed Periodic Lattices," *Phys. Rev.*, Vol. 76, p. 1592, December, 1949.

† Equation (1) is identical with Equation (6) in Slater's paper. The operator $-i\nabla$ in $E_{-i\nabla}$ operates only upon $\Phi(\vec{x})$, not upon $\vec{x}$ inside $E_{-i\nabla}(\vec{x})$.

‡ It is now irrelevant whether this position dependence is due to external electric fields or changes in the band gap. Actually, the two effects cannot be separated from each other. Equation (3), therefore, covers as well the older case of an electric perturbation in a truly periodic potential.

$$\left[-\frac{\hbar^2}{2m^*} \nabla^2 + E_B(\vec{x}) \right] \Phi(\vec{x}) = E\Phi(\vec{x}), \tag{3}$$

the familiar wave equation for an electron in a potential, $E_B(\vec{x})$. Since the position dependence of $E_B(\vec{x})$ will be different, in a nonuniform semiconductor, for the conduction band and for the valence band, Equation (3) represents the mathematical expression for the existence of the quasi-electric fields.

Two Examples

In this section the potential usefulness of the quasi-electric fields will be illustrated by two examples of how transistor performance can be improved by the incorporation of quasi-electric fields.[4]

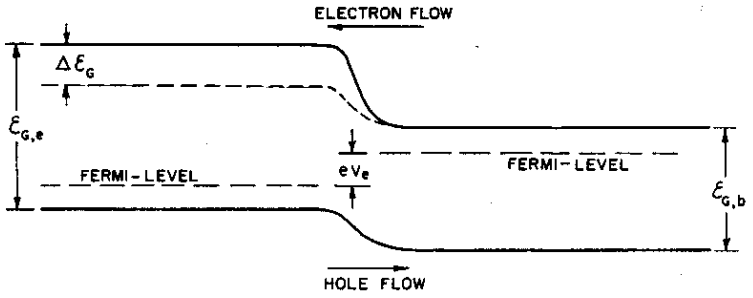

Fig. 3—Wide-gap emitter.

The Wide-Gap Emitter

Figure 3 shows a forward biased p-n junction where the p side has a wider band gap than the n side. In the transition region we then have a quasi-electric field opposing the hole flow to the right, and another quasi-electric field opposing the electron flow to the left. The electron field is stronger. This means that such a junction has a higher ratio of hole-to-electron current than a constant-gap junction with the same impurity distribution and the same mobilities. Used as an emitter in a p-n-p transistor, the transistor has a higher emitter efficiency and a higher current-amplification factor than an otherwise identical constant-gap transistor. This higher emitter efficiency may be utilized in either one of two ways:

[4] H. Kroemer, "The Theory of Diffusion and Drift Transistors, Part III—Dimensional Equations," *Archiv der Elektrischen Übertragung*, Vol. 8, p. 499, November, 1954.

(1) In low-frequency transistors the high emitter efficiency results in a reduction of the well-known and undesirable falloff of the current amplification factor with increasing current.

(2) In high-frequency transistors the intrinsically higher emitter efficiency of a wide-gap emitter may be utilized to decrease the doping of the emitter region without adversely affecting the current amplification factor. Such a decrease in doping results in a decrease of the emitter capacitance, which is one of the most seriously frequency-limiting quantities of modern high-frequency transistors.

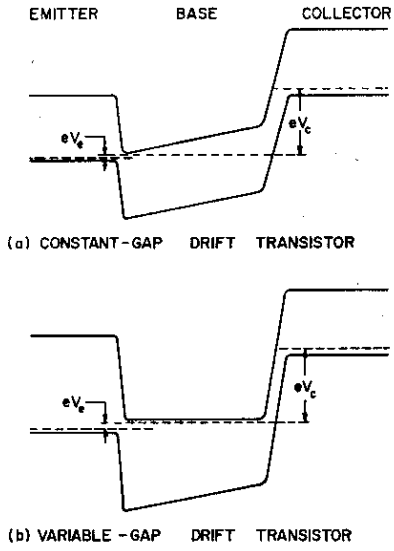

Fig. 4—Two drift-transistor types.

The wide-gap emitter will be described in more detail elsewhere.[5]

Graded-Gap Drift Transistor

In an ordinary drift transistor the drift field is generated by inhomogeneous doping in the base region (Figure 4a). Figure 4b shows a structure in which the quasi-electric field due to a decreasing band gap is used as a drift field. Such a structure has two advantages:[4]

1. Stronger drift fields are obtainable than by inhomogeneous doping.

---

[5] H. Kroemer, "Theory of a Wide-Gap Emitter for Transistors." To be published.

2. In an ordinary drift transistor a neutralizing electron accompanies each hole that travels through the base region. While the hole is aided by the drift field, the electron is opposed by it. It can be shown[4] that this effect wipes out the advantage of the drift field at high current densities. In a graded-base drift transistor, as shown in Figure 4b, there is no quasi-electric field opposing the electron flow, and the drift effect persists to much higher current densities.

Quasi-Magnetic Fields

Equation (3) was derived under the assumption of Equation (2), i.e., of spherical energy surfaces around $k = 0$. In many semiconductors known today this assumption is not fulfilled; instead, there are several nonspherical minima located at symmetrical points in k-space. If one of them is centered around, say, $\vec{k} = \vec{k}_o$ then, in the neighborhood of this minimum one has, instead of Equation (2),

$$E_k(\vec{x}) = \frac{\hbar^2}{2m^*}(\vec{k} - \vec{k}_o)^2 + E_B(\vec{x}), \tag{4a}$$

where $1/m^*$ is now a tensor. If there is a minimum at $\vec{k} = \vec{k}_o$ there will also be an identical minimum at $\vec{k} = -\vec{k}_o$:

$$E_k(\vec{x}) = \frac{\hbar^2}{2m^*}(\vec{k} + \vec{k}_o)^2 + E_B(\vec{x}). \tag{4b}$$

By substituting Equations (4a) and (4b) into Equation (1), one obtains:

$$\left[\frac{\hbar^2}{2m^*}(-i\nabla - \vec{k}_o)^2 + E_B(\vec{x})\right]\Phi = E\Phi \tag{5a}$$

$$\left[\frac{\hbar^2}{2m}(-i\nabla + \vec{k}_o)^2 + E_B(\vec{x})\right]\Phi = E\Phi. \tag{5b}$$

These are exactly the equations for an electron moving under the influence of a vector potential,

$$\vec{A} = \pm\frac{\hbar^2}{e}\vec{k}_o. \tag{6}$$

In a uniform semiconductor, $\vec{k}_o$ is constant and the vector potential has no physical meaning and can be transformed out of the equation by a simple gauge transformation. In a nonuniform semiconductor, however, $\vec{k}_o$ may vary with position. In general, then, curl $\vec{A}$ will no longer vanish. This means that the electron movement is as though a magnetic field

$$\vec{B} = \text{curl}\,\vec{A} = \pm \frac{\hbar c}{e} \cdot \text{curl}\,\vec{k}_o \tag{7}$$

were present. This is the quasi-magnetic field.

Thus far this result is a purely mathematical one. But the existence and the physical origin of the quasi-magnetic field can actually be understood qualitatively without resorting to the above mathematical derivation.

For simplicity we assume E_B to be constant, i.e., no quasi-electric fields are present. We wish to show, then, in a special example that the electron moves in an orbit if curl $\vec{k}_o \neq 0$. We assume that the crystal has a primitive cubic structure and that it is inhomogeneous along one of the [100] directions, which direction we call the x-direction. We further assume that the energy minima lie along the [100] directions of the k-space and that the minima move away from $k = 0$ if one proceeds in the positive-x direction. The k-space, then, is shown by Figures 5a and 5b. In such a case the direction of curl $\vec{A}$ for the six ellipsoids is shown in Figure 5c. Only for the two ellipsoids on the k_x axis is curl $\vec{A} = 0$.† The four other ellipsoids should see finite quasi-magnetic fields of the same magnitude, but of different directions.

An electron near the k_y minimum will be studied in more detail. Figure 6 shows the position of this minimum in k-space for three different positions inside the crystal, $x = 0$ and $x = \pm \Delta x$. Assume now that an electron is located at $x = 0$ with a k-vector corresponding to the point "A" in the diagram. Since the velocity, $\vec{v}$, of the electron is proportional to the gradient of the energy in k-space, the electron in that moment moves exactly in plus-x direction. After the time interval $\Delta x/v$ the electron has arrived at $x = +\Delta x$, still having the same absolute $\vec{k}$ vector. But since the ellipsoid itself has shifted, the position of the electron with respect to $\vec{k} = \vec{k}_o$ has changed. The gradient of the energy now has a component in the minus-k_y direction.

† For these two ellipsoids, div $\vec{A} \neq 0$, contrary to ordinary vector potentials.

Correspondingly, the velocity has a component in the minus-y direction, i.e., the electron is constantly deflected towards the right of its instantaneous direction of motion.

An electron at $x = 0$ and at point "B" in k-space has an initial velocity entirely in the plus-y direction. It therefore does not move into a region of different atomic potential. Nevertheless it, too, will be deflected towards the right. This is due to the fact that the distance of point "B" in k-space from $\vec{k}_o$ will vary for different values of x, and therefore the energy will change with x for fixed $\vec{k} = \vec{k}_B$. As shown in the bottom drawing of Figure 6, the energy decreases with increasing x. This gradient of the energy acts as a force upon the electron

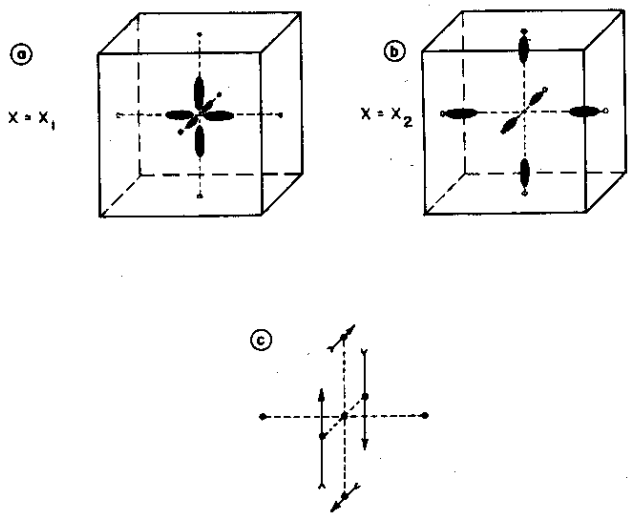

Fig. 5—(a) and (b): Examples of the location of the energy ellipsoids in the Brillouin zone at different positions inside the crystal; (c): Direction of the quasi-magnetic field for the different ellipsoids.

and also deflects it to its right.

For points intermediate between "A" and "B" on the energy ellipsoid, both of the two described effects are present and deflect the electron to the right. The net result is in all cases that the electron circles in k-space about $\vec{k} = \vec{k}_o$ along a path of constant energy, as in the presence of a true magnetic field.

The magnitude of the quasi-magnetic field is given by Equation (7). We assume a favorable case, namely that $\vec{k}_o$ wanders from the center of the Brillouin zone to its edge, and that the transition is rather steep, namely only about 2×10^4 atomic distances. For a lattice constant of 5×10^{-8} centimeter, then, $B = 625$ gausses.

This is not a negligible field. Of course, in practical cases the change of $\vec{k}_o$ may be smaller, but it is conceivable that the transition, on the other hand, can be made steeper. It appears reasonable, therefore, to expect magnetic fields of several hundred gausses.

The consequences of the quasi-magnetic fields are limited. As shown already in Figure 5, the fields for the two ellipsoids of each pair are in opposite directions. In thermal equilibrium, the two ellipsoids are equally populated. As a result, all those magnetic effects cancel which are linear in the magnetic field. There is, therefore, no self-Hall effect, no change in the low-field–high-temperature diamagnetic properties,

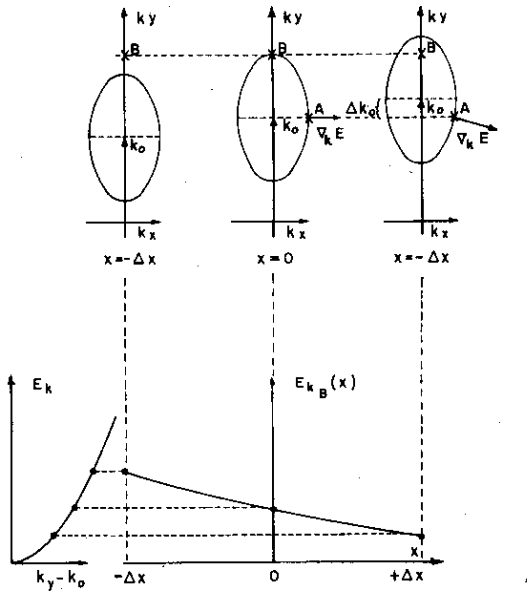

Fig. 6—Explanation of the origin of the quasi-magnetic field.

etc. There should be, however, an indication of effects that are nonlinear in the magnetic field, such as the magnetoresistance. For example, the mobility in a graded semiconductor alloy with noncentral energy bands will always be lower than in a uniform semiconductor alloy, due to the magnetoresistance effect caused by the quasi-magnetic field. It is possible, however, that this decrease of the mobility will not be observable, because already without the quasi-magnetic field the mobilities in a semiconductor alloy are lowered due to disorder scattering, and, the lower the original mobilities are the less they will be changed by a magnetic field. The situation might be different, how-

ever, if the nonuniform band structure were due to elastic deformation rather than to nonuniform composition, because then there is no disorder scattering. This should lead to a mobility reduction with nonuniform strains. One such type of strain is the thermoelastic lattice waves which are responsible for the lattice scattering of the electrons. Our theory, then, predicts that, in addition to the already-known scattering mechanism, these waves also reduce the lattice mobility due to the quasi-magnetic field they produce in multi-valley semiconductors. It is believed that this effect has not yet been studied.

Another effect that should be altered by quasi-magnetic fields is the cyclotron resonance. There the external magnetic field is added to the quasi field producing different net fields for different ellipsoids. If the line widths are small, the result should be a splitting of each cyclotron resonance line into a *multiplett*. The line width condition here means in particular, that no large change of the effective mass is associated with the change in the position of the ellipsoids.

For example, assume the external magnetic field to be parallel to the y-direction in Figure 5. The resonance corresponding to the four ellipsoids on the k_x and the k_z axes would then split up into a triplet, the center line corresponding to the two k_x ellipsoids, the two (symmetrically lying) side lines corresponding to the two k_z ellipsoids.

Other effects in which the quasi-magnetic field would show up could be expected if the equal population of the symmetrical minima could be disturbed.