Theory of a Wide-Gap Emitter for Transistors*

HERBERT KROEMER†, MEMBER, IRE

Summary—In order to obtain a high current amplification factor, it is important in transistors that the ratio of the injected minority carrier current over the total emitter current, γ, be close to unity, or that the quantity $1-\gamma$, called the injection deficit, be as small as possible.

It is shown that the injection deficit of an emitter can be decreased by several orders of magnitude if the emitter has a higher band gap than the base region. This effect can be utilized either in addition to the commonly used high emitter doping in order to eliminate the alpha falloff with current, or to decrease the high emitter doping in order to obtain a lower emitter capacitance.

Decreasing the emitter capacitance in high-frequency transistors may be utilized either to extend their frequency range or to increase their power capabilities by increasing the area.

Introduction

An important quantity characterizing the emitter of any transistor is the emitter efficiency, γ, defined as that fraction of the total emitter current that is minority-carrier injection current. Since the current amplification factor, α_{ce}, is proportional to γ, it is desirable that γ be high. Actually, it is very important that γ be close to unity. In the usual grounded-emitter operation the current amplification factor of the transistor is

$$\alpha_{cb} = \frac{\alpha_{ce}}{1-\alpha_{ce}}. \tag{1}$$

From this it follows that a change of γ such that α_{ce} increases from 0.98 to 0.99 (percentagewise a very small change of about 1 per cent) increases α_{cb} from about 49 to about 99, that is, by a factor of two. It is therefore more appropriate to consider the "injection deficit" $1-\gamma \approx (1-\gamma)/\gamma$ rather than γ itself as a measure of the transistor performance. In a good transistor, the injection deficit ought to be small.

In a *p-n-p* transistor,

$$\gamma = \frac{j_p}{j_n + j_p}, \tag{2}$$

$$\frac{1-\gamma}{\gamma} = \frac{j_n}{j_p}, \tag{3}$$

where j_p and j_n are the densities of the hole and electron currents at the emitter junction. In order to obtain a low ratio of electron to hole current, it is necessary in a semiconductor with constant band gap to dope the p side of the junction much more heavily than the n side. There are practical limits, however, as to the magnitude of doping possible, and there are situations where a high doping of the emitter is undesirable for other reasons. In all cases, an improvement could be obtained if the injection deficit could be decreased by other means instead of, or in addition to, the high doping.

The purpose of this paper is to point out that it is possible to lower the injection deficit by using an emitter material with a wider band gap than the base material (Fig. 1).[1] The reason for this is that in such a case the activation energy (or contact potential) qV_n for electrons flowing from the base into the emitter is higher than the activation energy (or contact potential) qV_p for holes entering the base from the emitter (see Fig. 1). The difference in activation energies is the difference in bandwidth, ΔE. Since the activation energy enters exponentially into the current-flow equations, this means a decrease in the injection deficit by a fraction of $\exp(-\Delta E/kT)$, all other things being equal. This is shown quantitatively in the next section.

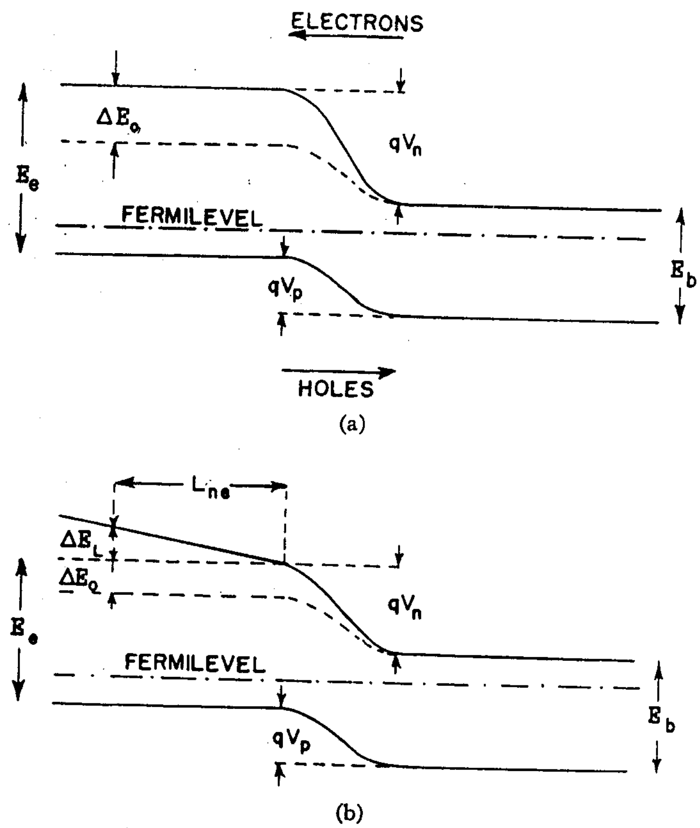

Fig. 1—Band structure of a wide-gap emitter junction. (a) With constant gap outside the depletion region and (b) with linear gap variation inside the emitter.

* Original manuscript received by the IRE, April 12, 1957; revised manuscript received, July 29, 1957.

† RCA Labs., Princeton, N. J.

[1] It has been pointed out to the author that this principle was first suggested by W. Shockley in U. S. Patent No. 2,569,347, issued September 25, 1951.

The Injection Deficit in a Wide-Narrow Junction

We introduce the following notation:

j_n, j_p = electron and hole current densities at the junction.
D_{ne}, D_{pb} = diffusion constants for electrons in the emitter and holes in the base, and
L_{ne}, L_{pb} = corresponding diffusion lengths.
n_{oe}, p_{ob} = equilibrium minority carrier densities in the emitter (electrons) and the base (holes) adjoining to the junction.
V = applied bias.
n_{ie}, n_{ib} = intrinsic carrier densities of the emitter and base semiconductors.
P_e, N_b = net acceptor density in the emitter and net donor density in the base.
m_{ne}^*, m_{pe}^*, m_{nb}^*, m_{pb}^* = effective masses of the electrons and the holes in the emitter region and the base region.
E_e, E_b = emitter band gap and base band gap. $\Delta E_0 = E_e - E_b$.

With this notation for a simple p-n junction,[2]

$$i_n = \frac{qD_{ne}n_{oe}}{L_{ne}}\left(e^{\frac{qV}{kT}} - 1\right) \tag{4a}$$

$$i_p = \frac{qD_{pb}p_{ob}}{L_{pb}}\left(e^{\frac{qV}{kT}} - 1\right) \tag{4b}$$

$$\frac{j_n}{i_p} = \frac{D_{pe}}{D_{pb}}\frac{L_{pb}}{L_{ne}}\frac{n_{oe}}{p_{ob}}. \tag{5}$$

Now,

$$n_{oe} = \frac{n_{ie}^2}{P_e}, \qquad p_{ob} = \frac{n_{ib}^2}{N_b}. \tag{6a, b}$$

Furthermore,

$$\frac{n_{ie}^2}{n_{ib}^2} = \left(\frac{m_{ne}^* m_{pe}^*}{m_{nb}^* m_{pb}^*}\right)^{3/2} \exp = (\Delta E_0/kT) \tag{7}$$

Therefore,

$$\frac{j_n}{i_p} = \frac{D_{ne}}{D_{pb}}\frac{L_{pb}}{L_{ne}}\frac{N_b}{P_e}\left(\frac{m_{ne}^* m_{pe}^*}{m_{nb}^* m_{pb}^*}\right)^{3/2} \exp\left(-\Delta E_0/kT\right). \tag{8}$$

This expression differs from that for a junction between semiconductors of equal band gaps by the factor

$$\left(\frac{m_{ne}^* m_{pe}^*}{m_{nb}^* m_{pb}^*}\right)^{3/2} \exp\left(-\Delta E_0/kT\right) \tag{9}$$

of which the important part is the exponential factor. If, for example, ΔE is 0.2 *ev*, then at room temperature $kT = 0.025$ *ev* and $\Delta E/kT = 8$. Assuming the effective masses to be identical, the injection deficit is then decreased by a factor of $e^{-8} = 1:3000$.

If the band gap in the emitter region is not constant but increases linearly with increasing distance from the junctions [Fig. 1(b)], (8) still does not give a full account of the change in injection deficit. If the above ΔE_0 is the gap difference across the depletion layer and if ΔE_L is the gap variation along a diffusion length, L_{ne}, in the emitter, then it can be shown that the factor

$$f(\Delta E_L) = \sqrt{\left(\frac{\Delta E_L}{2kT}\right)^2 + 1} - \frac{\Delta E_L}{2kT} \tag{10a}$$

has to be added to (4a), (8), and (9). For $\Delta E_L \gg 2kT$

$$f(\Delta E_L) \rightarrow \frac{kT}{\Delta E_L} \ll 1. \tag{10b}$$

Therefore, a band-gap variation outside the depletion layer also reduces j_n/j_p. However, this is to a smaller degree, namely only linearly rather than exponentially.

Eq. (8) does not hold for arbitrarily large ΔE's, however. This is because (4a) holds only so long as the density of electrons injected into the p-type region remains small compared to the electron density in the source, that is in the n-type region. The analogous statement holds for holes. Mathematically this means

$$V_n - V \gg kT \tag{11a}$$

$$V_p - V \gg kT. \tag{11b}$$

In a p-type wide-gap emitter $V_n > V_p$ and (11a) is fulfilled automatically if (11b) is. Consequently, (8) holds only for voltages that satisfy (11b). But, (11b) also implies that for a workable wide-gap emitter V_p must not

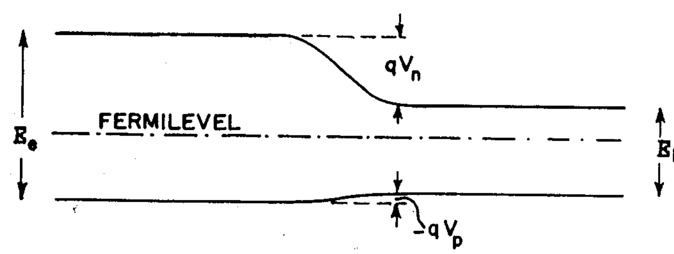

Fig. 2—Wide-narrow junction with negative V_p, due to a low doping ratio. Not suited as wide-gap emitter.

be negative. Therefore, a structure like in Fig. 2 would not be a good wide-gap emitter because the minority carrier density in the base is larger than the majority carrier density in the emitter.

[2] We treat the case of a simple p-n junction rather than a transistor to obtain more symmetrical notation. In a transistor with a base width $w_b \ll L_{pb}$, one has to replace L_{pb} by w and to omit the "minus one" behind the exponential factor. For the derivation of (4) see W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

Alpha and Alpha Falloff

In order to estimate the influence of the decreased injection deficit upon α_{cb} in a transistor, one has to take the recombination losses into account. If ρ is the fraction of injected carriers that recombine on the way to the collector, then for $\rho \ll 1$, from (1) and (2)

$$\frac{1}{\alpha_{cb}} \approx \rho + (1 - \gamma) \approx \rho + \frac{j_n}{j_p} \cdot \qquad (12)$$

If j_n/j_p is small compared to ρ, α_{cb} is determined solely by the recombination losses. This is the case for many transistors at low injection currents. At high injection currents the injected hole density in the base becomes comparable to the donor density. To maintain electrical neutrality, the electrons in the base increase by the same number. This means that the electron current into the emitter is bigger than the value given by (4a) by the factor by which the electron density has increased. For increasing current, therefore, j_n/j_p is not constant but increases (linearly) with current. Eventually j_n/j_p becomes comparable with and larger than ρ, resulting in the well-known alpha falloff.[3]

If the emitter has a wide band gap, the ratio j_n/j_p increases with current by the same *factor* as for an ordinary emitter. Since on an *absolute* scale, however, j_n/j_p is lower by the factor (9), the alpha-falloff effect sets in at much higher current densities. Since j_n/j_p increases linearly with the total current, alpha falloff sets in at currents which are bigger by the reciprocal of (9), compared to an otherwise identical constant-gap transistor. In our numerical example of $\Delta E = 0.2$ *ev*, the alpha falloff sets in at 3000 times the current. This means the alpha falloff is practically nonexistent.

Capacitance

Another consequence of the exponential factor in (9) is the following:[4] In many transistors for small-signal operation, it is of not primary importance to minimize the falloff effect to the point of vanishing. In these cases the exponential factor may be used to decrease the doping in the emitter by this same factor (9) and still have an unchanged falloff characteristic. It would then be possible to have a usable emitter efficiency with an emitter that has a considerably lower impurity density than the base region. This, however, would imply a reduced emitter transition capacitance.

In the case of audio-frequency large-signal transistors, the emitter transition capacitance is of no great importance while alpha falloff is a very serious effect. In this case one therefore should maintain the high doping in the emitter. The situation is completely reversed, however, for very-high-frequency transistors, like the drift transistor or the *p-n-i-p* transistor. In these transistors the current amplification factor is, under usual operating conditions, not limited by the injection deficit but rather by transit time effects. The low-frequency alpha falloff therefore is not an important quantity in this case. However, since high-frequency transistors have a rather high impurity density in the base region, the emitter capacitance is rather high. As a result, the emitter capacitance often becomes the limiting factor for the *over-all* frequency behavior of the transistor. In such a case, a wide-gap emitter with a lower doping might improve the over-all frequency limit considerably.

Quantitatively, the capacitance of an abrupt junction is[5] (per unit area)

$$C = \sqrt{\frac{q\epsilon}{8\pi} \frac{NP}{N+P} \frac{1}{V+V_c}} \qquad (13a)$$

where V_c is the contact potential. If the two sides have a different dielectric constant,

$$C = \sqrt{\frac{q\epsilon_n N \cdot \epsilon_p P}{8\pi(\epsilon_n N + \epsilon_p P)} \frac{1}{V+V_p}} \cdot \qquad (13b)$$

If $P \gg N$ this simplifies to

$$C = \sqrt{\frac{q\epsilon_n N}{8\pi(V+V_c)}} \qquad (14a)$$

while for $P \ll N$

$$C = \sqrt{\frac{q\epsilon_p P}{8\pi(V+V_c)}} \cdot \qquad (14b)$$

If, in a constant-gap transistor, a doping ratio $P:N = 30$ is assumed as an example, the introduction of a 0.2 *ev* wider emitter band gap allows a reduction of this ratio by 1/3000, namely to $P:N = 1:100$ without a change in γ. The capacitance, then, would be decreased to one tenth of the original value assuming identical dielectric constants.

A reduction of the emitter capacitance of this order could be utilized either to increase the frequency limit of the transistor or to increase the emitter (and collector) area. In the latter case, one would obtain a higher power capability for the same frequency response.

The Wide-Gap Collector

The use of a wide-gap semiconductor in the collector region would have an advantage only if the collector region at the same time had a lower impurity concentration than the base region.[4] Then one would obtain the lower collector capacitance of a high-resistivity collector region without the increased saturation current that is associated with a higher resistivity collector region in the constant-gap transistor.

[3] W. M. Webster, "On the variation of junction transistor current-amplification with emitter current," Proc. IRE, vol. 42, pp. 914–920; June, 1954.

[4] H. Kroemer, "Zur theorie des diffusions und des driftransistors, part III," *Archiv der Elektrischen Ubertragung*, vol. 8, pp. 499–504 November, 1954.

[5] Shockley, *loc. cit.*