Proceedings of the 12th Conference on Solid State Devices, Tokyo, 1980;
Japanese Journal of Applied Physics, Volume 20 (1981) Supplement 20-1, pp. 9-13

(Invited) Heterostructures for Everything: Device Principle of the 1980's?

Herbert KROEMER

Department of Electrical and Computer Engineering,
University of California,
Santa Barbara, California 93106, USA

One of the dominant themes of semiconductor device R & D during the 1980's will be the incorporation of heterostructures into most existing kinds of devices, and the emergence of new kinds of devices made possible by heterostructures. In this paper the power of heterostructures as a design tool is illustrated by discussing several ways in which the incorporation of heterostructures can improve the bipolar transistor. The dominant idea is that energy gap variations are a powerful way to control carrier flow; in bipolar structures they permit the control of electrons and holes independently. Several applications of this principle are discussed, going beyond the familiar wide-gap emitter concept, and including several concepts not previously discussed in the literature. The paper closes with a brief discussion of non-bipolar applications and speculative future applications.

§1. Introduction

It has now been ten years since the experimental realization of the double heterostructure laser. Such lasers are today used in actual communications systems, and the heterostructure technology is rapidly spreading to other devices. I believe that one of the dominant themes of semiconductor device R & D during the 1980's will be the incorporation of heterostructures (HS's) into every kind of semiconductor device whose performance can be improved by such an incorporation, and for which the improvement is sufficiently desirable to justify the technology. Such improvements can be made in almost all classes of devices. Finally, new kinds of devices made possible by HS's are rapidly emerging, and will assume an increasing role toward the end of this decade.

Rather than attempting to cover every conceivable application of HS's, I shall try to illustrate the power of HS's as a design principle by concentrating on the variety of ways how their incorporation can drastically improve the familiar bipolar transistor. The underlying idea is that energy gap variations are a powerful way to control carrier flow, in addition to the control exerted by the electrostatic potentials generated by doping and bias. In bipolar structures, energy gap variations can be used to control electron and hole flows *separately*. Various examples of non-bipolar applications will be given in the last part of the paper.

§2. Heterojunction Bipolar Transistors

2.1 *Wide-gap emitters*

The idea that the performance of a bipolar transistor could be improved by increasing the energy gap in the emitter relative to the base is as old as the bipolar transistor itself.[1] Consider an npn transistor with an energy band diagram as shown in Fig. 1. The operating principle of the device is the injection of electrons from the emitter into the base, and their subsequent collection by the collector. Asso-

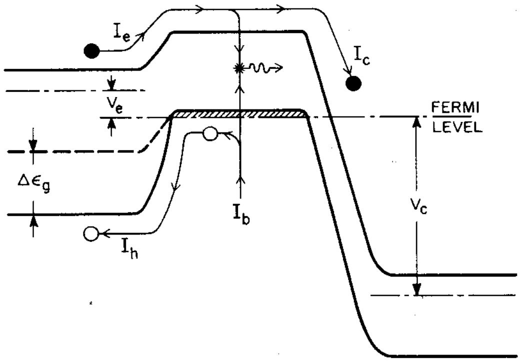

Fig. 1. Energy band diagram and carrier flow in an npn transistor with a wide-gap emitter. The heavy broken line in the emitter shows the valence band edge as it would exist in a homojunction transistor with the same emitter doping.

ciated with the desired emitter-to-base electron injection is an undesirable base-to-emitter hole injection. This hole injection current is part of the base current; at high current levels it is often the dominant part. In order to have a desirable current gain $\beta = I_c/I_b$ of about 100, the hole injection current from base to emitter must be kept below 1% of the emitter-to-base electron injection current. In conventional transistors this is achieved by a high emitter-to-base doping ratio of typically about 100: 1. This is the dominant design constraint in conventional bipolar transistors.

Suppose now that the emitter band gap is increased beyond that of the base. If the emitter doping is (initially) kept unchanged, all the increase in energy gap goes into depressing the valence band edge, introducing an additional energy barrier into the path of the hole flow, but not into the path of the electron flow. The result is a reduction of the base-to-emitter hole injection current by a factor $\exp(-\Delta\varepsilon_g/kT)$. This is incredibly effective: Energy gap difference of several tenths of 1 eV are readily available, and their effect is so large that the base-to-emitter hole injection current becomes negligible, regardless of the emitter-to-base doping ratio. The current gain β will be limited only by the recombination currents. It was recognized by Kroemer$^{2)}$ that this could be utilized to improve the transistor by using a much higher base doping and a much lower emitter doping. One result would be a greatly reduced β-falloff with current than in conventional transistors. If the emitter doping were reduced below the value that the base doping has in a conventional transistor ("super-inverted" doping), the emitter capacitance would be reduced, with benefits for the high-frequency performance. It was subsequently recognized$^{3-5)}$ that an even greater improvement in frequency response would result from the reduction in base resistance that could be obtained by a drastic increase in base doping. Maximum oscillation frequencies $f_{\max}$ of 100 GHz and more have been predicted.$^{3,4,6)}$

As a result of rapid progress in the heteroepitaxial growth of III/V compound semiconductors, especially GaAs/(Al, Ga)As, it appears that these predicted improvements are about to become realized in practice. Several authors$^{7-10)}$ have reported β-values over 1000, much larger than the β-values of even the best conventional transistors.* Some of these results were on phototransistors, which are easier to construct than true three-terminal devices.

The high-frequency performance still lags appreciably behind that of Si transistors, and even more behind the theoretical possibilities, largely due to non-optimized technologies. For true three-terminal devices, only three papers$^{10-12)}$ have so far reported performance above 100 MHz, up to $f_t \cong 1$ GHz. The fastest HS transistors so far are phototransistors.$^{8,13,14)}$ In one case,$^{13)}$ response times as short as 1 nsec have been reported. Considering the high gain of these devices ($\gg 100$), this would correspond to sinusoidal operating frequencies up to many gigahertz. Perhaps more significant: This particular device is the first high-performance bipolar transistor reported in the literature that was prepared by MOCVD rather than LPE. With the rapid progress in MOCVD and MBE, one may expect the future progress in transistor frequency performance to be rapid.

The technology of HS transistors is likely to be dominated by III/V compounds, because of the comparative ease with which the new epitaxial technologies permit the preparation of defect-free HS's in lattice-matched III/V compound pairs. Promising lattice-matched III/V systems in addition to (Al, Ga)As-on-GaAs are (Ga, In)P-on-GaAs and InP-on-(Ga, In)As.$^{14)}$

Because of the extremely high state of development of Si-IC technology, there is a strong incentive to develop an HS-IC technology for Si, even if its ultimate performance is less than for an all-III/V compound technology. Very promising results in this direction have been obtained by Matsushita *et al.*$^{15)}$ who used an emitter made from amorphous SiO_x, which has a wider energy gap than Si. It remains to be seen what the high-frequency potential of this combination is; the low-frequency current gain ($\beta \gtrsim 500$) is excellent.

A potentially very promising system is GaP-on-Si. The two semiconductors are fairly well lattice-matched (within 0.4%); however, good

---

*This field lacks a good recent review. References 7–14 give only recent results for which $\beta > 1000$ or $f >$ 100 MHz. For references to most of the earlier pioneering papers see ref. 5.

epitaxy of GaP on Si appears to be hard to achieve. The first GaP-on-Si transistor, prepared by VPE, has been reported by Katoda and Kishi,16) with (so far) very low β-values. In our own laboratory, we are attempting to grow GaP emitters on Si by MBE. A detailed theoretical estimate6) suggests that npn Gap-on-Si transistors with a maximum oscillation frequency $f_{\max}$ up to about 100 GHz should be achievable. It remains to be seen what will come of these GaP-on-Si efforts.

2.2 *Wide-gap collectors*

In most bipolar logic families (ECL is the dominant exception) the collector is forward-biased during part of the logic cycle. If the base is more heavily doped than the collector, this causes major base-to-collector injection of holes, which increases dissipation and slows down the switching speed. Using a wider energy gap on the collector as well as on the emitter side, suppresses this highly undersirable effect.6) In I^2L, this is not a problem, but I^2L has problems of its own, which also suggest HS's as a solution; see below.

2.3 *Utilizing turn-on voltage differences*

2.3.1 *Double base layers*

We may view the role of HS's as providing barriers to control the flow of electrons and holes independently of each other. The idea of the wide-gap emitter and collector was to block the flow of holes (in an npn transistor, electrons in a pnp). The idea can be extended to control the flow of the carriers with opposite polarity as well. Ladd and Feucht3) pointed out that it is desirable, in an HS transistor just as in a conventional transistor, to have a thick base region outside the emitter, and that the realization of the full promise of the HS transistor might well hinge on achieving a suitable device geometry. It was pointed out by this writer6) that a natural solution to this problem lies in a thick planar design in which the outer base consists of two layers, the upper one of which has the same wide energy gap as the emitter and is very heavily doped (Fig. 2). In a homostructure transistor, such a design would lead to a disastrous loss of β, emitter-base tunneling effects and a high emitter capacitance. In an HS transistor, because the vertical part of the emitter junction has a wide energy gap on both sides, this part will have its current density reduced by the same Boltzmann factor by which the base-to-emitter hole injection is depressed. Using a different terminology: the vertical part of the junction is biased below its turn-on voltage, which is higher than that of the horizontal part by approximately $\Delta\varepsilon_g/q$. Because the emitter would be weakly doped, the effect of a deep sidewall on the emitter capacitance is small, and emitter-base tunneling cannot occur at all. Yet the high conductivity of the upper base layer in effect brings the base contact as close to the emitter-base junction as is physically possible. In fact, this design has been used in the HS transistor reported in ref. 11.

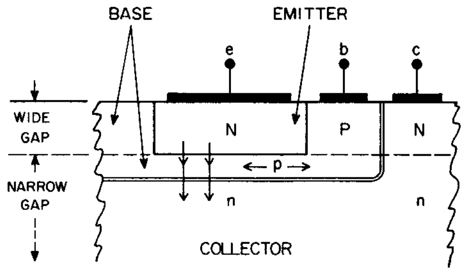

Fig. 2. Two-base layer design in a heterojunction transistor. The external base layer has, on top of a thin narrow-gap layer, a thick very heavily doped second base layer extending right to the emitter junction, minimizing the external base resistance. Because of the wider energy gap, negligible currents will flow through this part of the emitter junction.

2.3.2 *Wide-gap sidewalls and injection barriers*

The idea to use energy gap variations to suppress carrier injection into portions of the base region where no injection is desired, is an important new concept, the power of which does not appear to have been widely recognized. The lateral pnp transistor in I^2L is an example of a device that could be greatly improved by incorporating this idea. In homostructure designs, this transistor is always a poor transistor. It could be greatly improved by an HS geometry as, for example, in Fig. 3. The actual transistor is the all-narrow-gap p^+np structure shown embedded between the two wide-gap layers. The two wide-gap p^+np transistors above and below it are biased below turn-on; their n-type base regions simply act as walls to confine the injected holes in the true n-base. Because of the small hole diffusion lengths in III/V compounds, the implementation of such a structure will require submicron technologies.

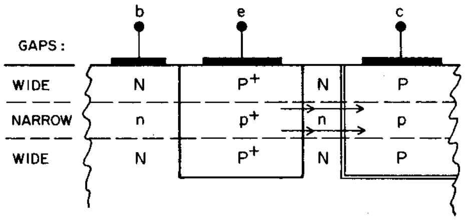

Fig. 3. Injection current in a lateral pnp transistor, as in I^2L. Only the narrow-gap portion (n) of the base carries current.

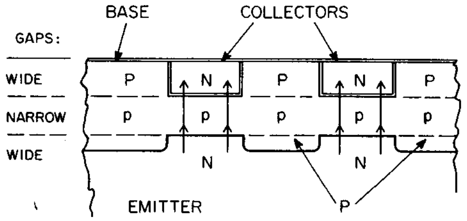

Fig. 4. Supression of electron injection into selected portions of the base region in I^2L, by means of wide-gap injection barriers, achieved by pulling the emitter-base junction into the wide-gap region of the structure.

Another example of suppression of undesired injection is the following. In I^2L, a significant fraction of the base area is in contact with the emitter, but not with one of the collectors. Electron injection into those portions of the base creates stored charge that wastes power and slows down the switching speed. In an HS-I^2L design, this could be avoided by simply pulling the emitter-base pn junction below the hetero-interface, into the wide-gap region, as shown in Fig. 4.

§3. Beyond Bipolar Transistors

The detailed discussion of the bipolar transistor was intended as an example. There probably does not exist a kind of device that cannot be similarly improved by the incorporation of HS's. Space does not permit me to say more than a few words about these other possibilities. Perhaps the most powerful general design principle applicable to many non-bipolar devices involves the use of thin ($\ll 1000$ Å) alternating layers of different semiconductors. At sufficiently abrupt HS's, there usually occurs a sharp discontinuity in the conduction band edge, similar to the discontinuity at Schottky barriers, giving the alternating layers an energy band structure as in Fig. 5. Because of the band edge steps, electrons from the high-ε_c layers will drain into the potential wells inside the low-ε_c layers, provided the layers are thin enough to minimize the built-in voltages created by the resulting space charges. If the low-ε_c layers are left undoped, with all donors placed into the high-ε_c layers, one obtains a high net electron concentration, without any impurity scattering. Such structures have an enhanced electron mobility compared to a bulk semiconductor of the same net electron concentration, especially at low temperatures.

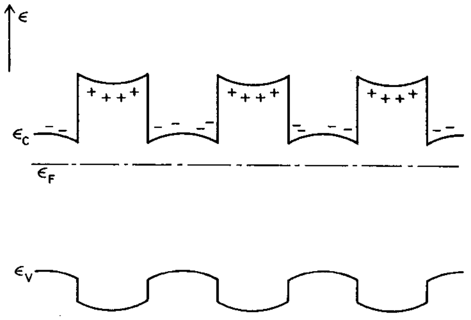

Fig. 5. Energy bands in a modulation-doped multilayer structure. All donors are contained in the wide-gap layers, all electrons in the narrow-gap layers.

This is the concept of **modulation doping**17) (MD), one of the most important new ideas to have emerged from the new hetero-epitaxial technologies. Its first utilization in an FET has been reported,18,19) leading to a major improvement in low-T performance. The use of MD is a very general concept applicable to many devices besides FET's, and it may even give rise to new kinds of devices.

One of the most promising aspects of MD is that it should permit the utilization, in the low-ε_c layers, of narrow-gap semiconductors such as InAs. These have much higher mobilities to begin with. Because of the absence of impurity scattering, MD structures would retain these mobilities; they would also minimize several other drawbacks of low-gap semiconductors that have prevented their device use in bulk structures. I believe that their use at low temperatures will lead to a re-appraisal of the use of semiconductors compared to Josephson devices at low temperatures.

Of particular interest in MD structures will be the non-linear high-field transport properties along the potential wells, including especially negative-differential-conductance(NDC) effects. They should be more pronounced than in bulk semiconductors. Scattering of hot electrons out of the potential wells should lead to a new and strong NDC mechanism,(20) similar to that of the Gunn effect, and of great promist for millimeter wave applications.

Finally, there will be **quantum wells**, structures as in Fig. 5, but with dimensions to small that discrete quantum levels are formed within the wells. Many future semiconductor lasers will go beyond present HS designs, towards quantum well designs.(21) Even more exotic possibilities exist for the use of large numbers of quantum wells in periodic arrays, as artificial **superlattices.**(22)

The list could easily be extended, especially if one were to include light-sensing devices other than phototransistors, and various ultra-specialized applications. In fact, many future applications of HS's have probably not even been perceived yet. But considering the major examples I have given, and doing so against the background of the proven role of HS's in lasers, I have probably made my point:

Heterostructures for Everything? Well . . .!?

References

1) W. Shockley: U.S. Patent 2569347, filed 26 June 1948, issued 26 Sept. 1951.
2) H. Kroemer: Proc. IRE **45** (1957) 1535.
3) G. O. Ladd and D. L. Feucht: IEEE Trans. Electron Devices **ED-17** (1970) 413.
4) W. P. Dumke, J. M. Woodall and V. L. Rideout: Solid-State Electronics **15** (1972) 1339.
5) For an excellent review of the various potential benefits of a heterojunction transistor (to the extent they had been recognized at the time) see Milnes and D. L. Feucht: *Heterojunctions and Metal-Semiconductor Junctions* (Academic Press, New York, 1972) Chap. 3.
6) H. Kroemer: Device Research Conference, 1978, Santa Barbara; see IEEE Trans. Electron Devicse **ED-25** (1978) 1339; Bull. Am. Phys. Soc. **24** (1979) 230.
7) B. W. Clark, H. G. B. Hicks, I. G. A. Davis and J. S. Heeks: *Gallium Arsenide and Related Compounds, 1974* (Inst. Phys. Conf. Series, Bristol, 1975) Vol. 24, pp. 373 ff.
8) H. Beneking, P. Mischel and G. Schul: Electron. Lett. **12** (1976) 375.
9) M. Konagai, K. Katsukawa and K. Takahashi: J. Appl. Phys. **48** (1977) 4389.
10) P. W. Ross, H. G. B. Hicks, J. Froom, J. G. Davies, F. J. Probert and J. E. Carroll: Electron. Engin. **49** (1977) 35.
11) D. Ankri and A. Scavennec: Electron. Lett. **16** (1980) 41.
12) J. P. Bailbe, A. Marty, P. H. Hiep and G. E. Rey: IEEE Trans. Electron. Devices **ED-27** (1980) 1160.
13) R. A. Milano, T. H. Windhorn, E. R. Anderson, G. E. Stillman, R. D. Dupuis and P. D. Dapkus: Appl. Phys. Lett. **34** (1979) 562.
14) M. Tobe, Y. Amemiya, S. Sakai and M. Umemo: Appl. Phys. Lett. **37** (1980) 73.
15) T. Matsushita, N. Oh-uchi, H. Hayashi and H. Yamato: Appl. Phys. Lett. **35** (1979) 549.
16) T. Katoda and M. Kishi: J. Electron. Matls. **9** (1980) 783.
17) R. Dingle, H. L. Störmer, A. C. Gossard and W. Wiegmann: Appl. Phys. Lett. **33** (1978) 665.
18) T. Mimura, S. Hiyamizu, T. Fugi and K. Nanbu: Jpn. J. Appl. Phys. **19** (1980) L225.
19) For a review of other HS benefits in FET's, see D. Boccon-Gibod, J.-P. André, P. Baudet and J. P. Hallais: IEEE Trans. Electron. Devices **ED-27** (1980) 1141.
20) K. Hess, H. Morkoc, H. Schichijo and B. G. Streetman: Appl. Phys. Lett. **35** (1979) 469.
21) J. P. van der Ziel, R. Dingle, R. C. Miller, W. Wiegmann and W. A. Nordland, Jr.: Appl. Phys. Lett. **26** (1975) 463. For a recent review of the developments since this first paper, see N. Holonyak, R. M. Kolbas, R. D. Dupuis and D. D. Dapkus: IEEE J. Quantum Electron. **16** (1980) 170.
22) For a recent compact review of the very extensive literature see L. L. Chang and L. Esaki: Prog. Crystal Growth Charact. **2** (1979) 3.