Heterostructure bipolar transistors: What should we build?

Herbert Kroemer

Citation: Journal of Vacuum Science & Technology B **1**, 126 (1983); doi: 10.1116/1.582513
View online: http://dx.doi.org/10.1116/1.582513
View Table of Contents: http://scitation.aip.org/content/avs/journal/jvstb/1/2?ver=pdfcov
Published by the AVS: Science & Technology of Materials, Interfaces, and Processing

Heterostructure bipolar transistors: What should we build?

Herbert Kroemer

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

(Received 21 October 1982; accepted 11 January 1983)

The paper discusses likely future developments in heterostructure bipolar technology, especially by MBE. This written version concentrates on two new conceptual developments extending earlier concepts. One of these pertains to the problem of emitter/base junction grading. A grading scheme is proposed that extends the grading through the base region and creates a graded-gap base. The other proposes an extension of permeable base transistor technology to bipolar transistors in what is called a gridded-base bipolar transistor. Both promise a further increase in device speed, largely by addressing themselves to the persistent problem of base resistance reduction. Several other topics, already contained in the author's January 1982 review paper and presented orally at the Workshop, have been omitted from this printed version.

PACS numbers: 85.30.Pq, 73.40.Lq

I. INTRODUCTION

This is a sequel to a recent paper[1]—hereafter called HK82—which discussed the mid-1981 state of the development of the heterostructure bipolar transistor (= HBT). That paper stressed the post-1978 development of ideas rather than specific technological achievements implementing those ideas. It preceded the first published reports of HBT's prepared by MBE, but was written in anticipation of extensive MBE developments in this area. The first published reports on MBE-prepared HBT's have by now appeared.[2–4] However, the intent of the present paper is not to review those early MBE achievements: The situation is developing so rapidly that such a review would be outdated by the time it appears in print. The intent is rather to contribute to further advances in the conceptual development of the HBT. Much of this is technology independent, yet is a highly appropriate subject matter for an MBE Workshop: It is hoped that the paper will help the MBE technologist in deciding what to attempt building. As shall be seen, some of this will be quite specific, and perhaps unexpected.

Because of space limitations, the paper will discuss only two selected areas, in which significant conceptual progress has been made since HK82: (a) the question of abrupt versus graded emitter/base heterojunctions, culminating in a proposal for grading not only the emitter/base junction, but to resurrect the old idea[5,6] of grading the entire base region as well (Sec. II), (b) the concept of a gridded-base bipolar transistor, an application of permeable base transistor (PBT) technology[7] to HBT's (Sec. III).

II. THE EMITTER/BASE GRADING PROBLEM

A. The problem

Although the first proposal for a wide-gap emitter, by Shockley,[8] assumed an abrupt heterojunction, until recently most subsequent work, following Kroemer,[9] assumed a graded junction in which both band edges vary monotonically across the emitter/base interface. If all parameters other than the emitter energy gap were kept constant, the effect of the wide-gap emitter would then simply be to increase the *injection ratio* J_n/J_p by the inverse Boltzmann factor involving the energy gap difference

$$(J_n/J_p)_{\text{hetero}} = (J_n/J_p)_{\text{homo}} \times \exp(\Delta\epsilon_g/kT). \quad (1)$$

Here, J_p is the hole current density injected from the base into the electrically neutral part of the emitter body, excluding that portion of the total hole outflow from the base that recombines with electrons in the emitter/base space charge region. Similarly, J_n is the electron current density injected from the emitter into the base. For large values of $\Delta\epsilon_g/kT$, which are readily obtainable, the increase in injection ratio can be many orders of magnitude. As was discussed extensively in HK82, the central idea of HBT design is to trade off this increase in injection ratio for various other design changes that lead to several major improvements in device performance. The first and most important design change made possible is a large increase in base doping, which lowers the base resistance and thereby drastically improves both the high current and the high frequency properties of the device. This base resistance problem continues to play an important role in the present paper.

Heterojunctions grown by MBE tend to be quite abrupt, unless specific measures are taken to grade the transition. As was discussed in HK82, an abrupt emitter/base junction introduces a potential barrier $\Delta\epsilon_B = \Delta\epsilon_c - \Delta\epsilon_N$ into the path of the electron flow (Fig. 1), the height of which tends to be very close to the conduction band offset $\Delta\epsilon_c$ that is characteristic of the semiconductor pair employed. As a result, the improvement in injection ratio is much less than for a graded junction. To the first order, for a sufficiently thin and sufficiently heavily doped base region, the energy gap difference $\Delta\epsilon_g$ in Eq. (1) is simply replaced by $\Delta\epsilon_g - \Delta\epsilon_B = \Delta\epsilon_g - \Delta\epsilon_c + \Delta\epsilon_N \approx \Delta\epsilon_g - \Delta\epsilon_c$, that is, by the valence band offset $\Delta\epsilon_v$, which tends to be much smaller than $\Delta\epsilon_g$, leading to a large reduction in the magnitude of the exponential factor in Eq. (1), and hence in the injection ratio. The simple first-order replacement $\Delta\epsilon_g \rightarrow \Delta\epsilon_v$ is *quantitatively* valid only if the base is sufficiently thin that the speed of diffusive electron flow across the base is supply limited, as is the flow across the spike barrier. This is a reasonable approximation for transistors of greatest interest, with base thicknesses of the order 10^{-5} cm.

The reduction of injection ratio by the conduction band spike barrier is particularly severe for $Al_xGa_{1-x}As/GaAs$ heterojunctions, because of their small valence band discontinuity, only about $\Delta\epsilon_v \cong 0.15\Delta\epsilon_g$ (this equals $\cong 0.176\Delta\epsilon_c$, or about 1.87 meV per percent of Al in the Al:Ga ratio). Now, in order to avoid injection of electrons into the low mobility upper valleys of the GaAs band structure, it is necessary to keep the height of this spike below 0.3 eV. In an *abrupt* $Al_xGa_{1-x}As/GaAs$ emitter with $\Delta\epsilon_c = 0.3$ eV, corresponding to $x \cong 0.28$, this implies a valence band offset $\Delta\epsilon_v$ of only $\cong 0.053$ eV ($\cong 2kT$), which is too small to permit the full range of device design adjustments (especially the high base doping) that are the essence of HBT design.

One possible way out of this dilemma has been to deliberately grade the emitter/base junction, which reduces the height of the electron barrier. In fact, the first successful MBE-grown HBT's, reported by Asbeck *et al.*,[2] incorporate just such grading. We will return to this point shortly.

On the other hand, it has been argued[1,10] that the injection of electrons into the base from a "ballistic launching ramp" would actually be desirable, so long as this does not lead to transfer into higher valleys. The high speed of such ballistic electrons would lead to higher speed transport through the base. Such ballistic effects would make up for some (but not all) of the disadvantages of the greatly reduced injection ratio. However, there is actually no need to sacrifice any of these desirable properties to obtain the others: They can be obtained simultaneously, and it is evidently desirable to do so.

To this end it is necessary either to choose a different semiconductor pair with a higher $\Delta\epsilon_v:\Delta\epsilon_c$ ratio, or to modify the energy gap grading procedure.

B. (Ga, In)P/GaAs as an alternate

The undesirably small valence band offset in the (Al,Ga) As/GaAs system is not an accident: It is a direct consequence of the two semiconductors having the same anion, namely arsenic. Replacing the emitter by a semiconductor that contains a different anion with a higher electronegativity, namely phosphorus, would automatically increase the valence band offset.[11] The ternary alloy $Ga_{0.5}In_{0.5}P$ is lattice

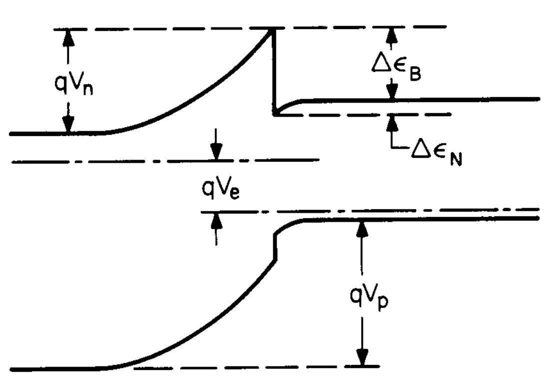

FIG. 1. Band structure of an abrupt wide-gap emitter, showing the electron spike barrier. No interface charge is assumed. (From Ref. 1.)

matched to GaAs. Using the Harrison theory[12] of heterojunction band lineups, we estimate a valence band offset of 0.29 eV ($\cong 11kT$), which should yield an injection ratio equal to that of a graded $Al_xGa_{1-x}As$ emitter with $x \cong 0.27$. But in contrast to the latter, a (Ga, In)P abrupt emitter would still retain a conduction band offset estimated at $\cong 0.16$ eV, available as a significant ballistic launching ramp.

The first few published reports[13–16] of MBE growth of $Ga_{0.5}In_{0.5}P$ on GaAs have appeared. Although the papers reveal the usual startup problems one expects for a new materials system, these difficulties appear, if anything, less severe than those exhibited by (Al, Ga)As at the same stage of its development, and they contain nothing suggesting any really serious problems. I propose that we take (Ga, In)P *very* seriously for HBT's.

C. Graded emitter/base junctions

If the conduction band spike barrier is lowered, through compositional grading, by an amount $\Delta\epsilon_l$, the effect on the injection ratio is the same as if $\Delta\epsilon_l$ were added to the valence band offset, leading to the energy $\Delta\epsilon_v + \Delta\epsilon_l$ to appear instead of either $\Delta\epsilon_g$ or $\Delta\epsilon_v$ alone in the exponential in Eq. (1). The HBT's reported by the Rockwell group[2] have employed this principle with results that appear promising. However, such barrier lowering also reduces the benefits of ballistic carrier injection. One may obtain both a large injection ratio *and* strong ballistic effects by combining the grading with an increase of the Al fraction, to somewhere near 40%. In the absence of grading, this would yield conduction and valence band offsets of about 0.42 and 0.07 eV. Reducing the conduction band barrier by grading, to between 0.20 and 0.25 eV, would increase the effective energy in the injection ratio enhancement factor to between 0.24 and 0.29 eV, a more than adequate value.

D. Graded-base long-gradient HBT

Given the desirability of grading the energy gap, what should be the ϵ_g-vs-position profile? We argue here in favor of a design that grades not only the emitter/base junction proper, but that extends grading through the base, to the edge of the base/collector depletion layer, as shown in Fig. 2. The idea of base grading is to introduce a strong *quasielectric field*[6] into the base to aid the minority carrier transport. Such a design was first proposed by this writer in 1954[5] and 1957.[6] It was recently taken up again by Levine *et al.*,[17] and by Asbeck *et al.*[18] For a given base thickness, such a quasielectric field can greatly reduce the base transit time τ_b. Now, in a well-designed bipolar transistor, the base transit time is only a relatively small fraction of the total signal propagation delay. Hence, the improvement obtainable by incorporating a drift field into a base with *fixed thickness* is quite limited. But the high electron drift velocity can be traded off to retain a *fixed transit time* for a much thicker base region, which would have a much lower base resistance, which in turn increases the speed of the transistor.

If one keeps the total potential energy drop $\Delta\epsilon_b$ within the base below the energy at which electrons can transfer into higher low mobility valleys, the electric field in the base may be allowed to exceed the threshold field above which such

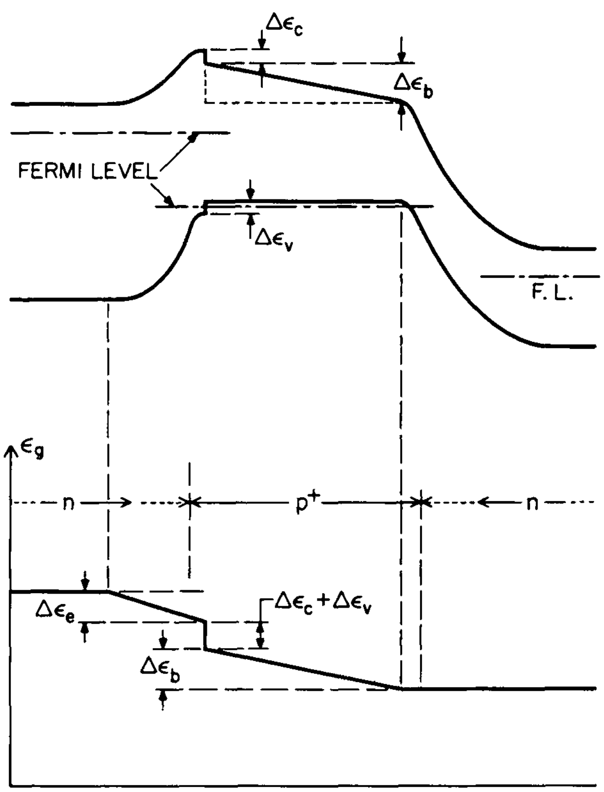

FIG. 2. Proposed heterostructure bipolar transistor structure with graded-gap base region. (a) Band diagram. (b) ϵ_g-vs-position profile.

transfer would take place in a long sample. It is then justified to use the below-threshold mobility μ_b to estimate the drift velocity. For given values of τ_b, $\Delta\epsilon_b$, and μ_b one easily finds a base width

$$w_b = (\mu_b \Delta\epsilon_b \tau_b / q)^{1/2}. \quad (2)$$

It is not obvious which mobility value to use. Because of the high base doping, it would be unrealistic to assume low field mobilities as high as those in high purity GaAs. At the same time, because the electrons in such a structure tend to be hot electrons less subject to impurity scattering then lattice-temperature electrons, it would also be incorrect to assume the low field mobilities of heavily doped GaAs. Assuming the *adhoc* value $\mu_b = 2500\ \text{cm}^2\text{V}^{-1}\text{s}^{-1}$, together with $\Delta\epsilon_b = 0.25$ eV and $\tau_b = 1$ ps, one finds $w_b = 2.5\times10^{-5}$ cm, far above current design practices for high speed bipolar transistors. The associated quasielectric field is $E_b = \Delta\epsilon_b / qw_b = 10^4$V/cm, far above the intervalley transfer field in a long GaAs specimen. The drift velocity would be $v_b = \mu_b E_b \cong 2.5\times10^7$ cm/s.

Because of the high drift field inside the base, any ballistic launching ramp at the entrance into the base should be kept low. In effect the electrons are traveling under near-ballistic conditions in the high field anyway, and a high launching ramp would only introduce the danger of transfering electrons to the higher low mobility valleys. What *is* probably desirable is a shallow ($\cong 1kT$) ballistic "kicker," to bring the electrons up to the desired high base drift velocity instantaneously. If the energy gap gradient in the base is achieved by controlled temperature ramping of the MBE beam sources during growth, such a shallow kicker may be easily obtained by simply shuttering the column-III sources for an appropriate time while the temperature ramping proceeds.

The remainder of the energy gap variation shown in Fig. 2 takes place inside the emitter/base space charge layer, possibly extending into the neutral emitter body itself. It is only this energy gap variation to the left of the kicker step that contributes to the desirable high injection ratio enhancement factor in Eq. (1), and it should therefore be kept large compared to kT, preferably no less that $9kT$ ($e^9 \sim 8000$). In Fig. 2 we have shown linear gap variations with position (with different slopes on the two sides of the ballistic kicker step). The width of the transition on the emitter side was chosen to coincide with the width of the emitter/base space charge layer; the optimal grading width is expected to be close to this value.

The structure discussed here differs from the ballistic HBT recently discussed by Ankri and Eastman.[10] In their structure the initial electron launching energy is just below the intervalley transfer energy, but the base region is of uniform gap. The initial injection velocity in the Ankri/Eastman device is much higher, but electrons of such a high energy can lose energy very rapidly. Following the earlier proposal in HK82, Ankri and Eastman argue that, because of the forward-directed nature of polar scattering, the electron retains a high forward velocity for a large distance. This would presumably make a fairly thick base possible. But it is not clear to what extent this remains valid in a base region as heavily doped as would be desirable in an HBT, in which other scattering mechanisms may be very important. In our present long-gradient design, a natural "sloping floor" limiting the energy loss is placed underneath the electron. It is believed that in this way a larger distance can be traversed in a given transit time, and at higher doping level, two factors that should combine to yield a most desirable lower base resistance.

III. THE GRIDDED-BASE BIPOLAR TRANSISTOR

A. Bipolars vs FET's

No discussion on the merits of HBT's can ignore their principal competitors, field effect transistors (FET's). Hence, HK82 contained an extensive discussion on this comparison. Two new developments have entered and changed the picture since then: The sudden emergence of the new High Electron Mobility Transistor[19] (HEMT), and the striking technological progress that has been made in implementing the permeable base transistor (PBT) idea.[7]

It is assumed that the reader is sufficiently familiar with the HEMT to be aware that it is basically a FET with higher electron mobilities and hence higher speed than conventional FET's, and that the mobilities and the device speeds increase drastically with decreasing temperature. Therefore, if highest raw speed at any cost is desired, and if "at any cost" includes a willingness to go to cryogenic operation, the HEMT is unquestionably preferable to any form of bipolar transistor, and probably even superior to Josephson devices. However, for operation under more common conditions, at or above room temperature, the comparisons between HBT's and FET's made in HK82 largely carry over to the

HEMT version of FET's, with only some quantitative shifts in favor of the HEMT in borderline cases.

B. The PBT: The best of both worlds?

From a bipolar perspective, the progress in PBT technology[7] is perhaps more important. The PBT is basically a vertical FET, in which the current flows vertically through a thin epitaxial layer, and in which the controlling gate electrodes have been embedded into the semiconductor in the form of narrow and very closely spaced metal figures. The vertical flow geometry permits a much closer source-to-drain spacing and hence a higher speed than in horizontal FET's, including HEMT's. In some ways the PBT resembles a bipolar device, but without the base resistance that is the nemesis of the latter. However, this speed improvement comes at a very high technological cost: The finger spacing must not only be very close, requiring x-ray or electron beam lithography but—worse—it must be very uniform within a device, and in IC's also from device to device. Even small fractional variations in this spacing cause large variations in turn-on voltage. Variations within a single device smear out the turn-on characteristics and reduce the transconductance of the device. Variations from device to device reduce their integratability. Quite possibly, there does not, at this time, exist a device that is more demanding on horizontal lithography.

C. Applying PBT technology to bipolars

If one contemplates the reaons for this difficulty, one realizes that it is not so much related to the PBT technology itself as to the fact that the PBT is a vertical FET rather than some other kind of device less sensitive to lithography tolerances. But this connection need not exist! Divorced from its connection to a specific kind of device, PBT technology may be viewed as being simply a technology to embed a conductive metal grid into a single-crystalline semiconductor body without compromising the device quality of the semiconductor. The remarkable fact that the PBT works as well as it does means nothing less than that such embedded-metal structures must be considered as very serious contenders for all kinds of future device structures. This includes specifically their use to create improved internal electrical access to the base region of a bipolar transistor, significantly improving what has always been the main bottleneck of bipolar transistor design, the base resistance. Figure 3 shows the envisaged configuration, which I would like to call a *gridded bipolar transistor*. It differs from the PBT principally by the insertion of a very heavily p-type doped base region in low resistance electrical contact with the metal grid. And of course the emitter region has been changed to a wide-gap emitter, making the structure a heterostructure bipolar device. It may (or may not) be desirable to surround the grid metal by a very thin heavily p-type doped wide-gap "sheath," as shown. This would suppress parasitic current flow across the forward-biased grid-to-emitter Schottky barrier, as well as loss of injected electrons from the base into the grid. It should be readily possible to create such a sheath—if in fact necessary—by outdiffusion of a suitable dopant (Be?) from the metal during subsequent semiconductor growth.

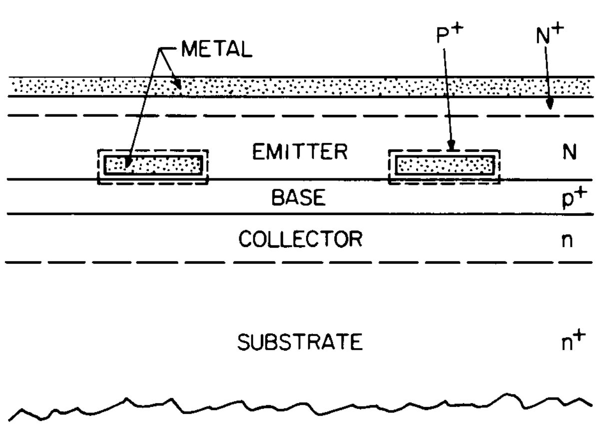

FIG. 3. Proposed gridded-base bipolar transistor.

Except for possible unforseen difficulties with this p-type sheath—which might not be needed anyway—the structure should not be much more difficult to build than a conventional PBT, and it might have significantly better properties: (a) Being a true bipolar, it should have the higher transconductance of a bipolar compared to an FET. (b) The turn-on voltage should no longer be highly sensitive to the exact value of the grid finger spacing, but should depend principally on the well-defined energy gap of the base region semiconductor.

In contrast to halide–VPE, MBE has so far exhibited difficulties in growing high quality crystals on top of a metal. Progress in this direction has recently been made,[20] but it is not all clear whether, say, semi-insulating overgrowth between the emitter contact and the base grid fingers would even be a drawback.

ACKNOWLEDGMENTS

I have greatly benefitted from numerous discussions with more individuals than can be listed here, but none more than Dr. Peter Asbeck and Dr. David Miller at Rockwell. This work was supported by the U.S. Army Research Office and by the Office of Naval Research.

[1] H. Kroemer, Proc. IEEE **70**, 13 (1982). For a shorter version, containing, however, some additional material, see also: H. Kroemer, Jpn. J. Appl. Phys. Suppl. **20-1**, 9 (1981).
[2] P. M. Asbeck, D. L. Miller, R. A. Milano, J. S. Harris, Jr., G. R. Kaelin, and R. Zucca, *IEEE Internat. Electron Devices Meeting*, (IEEE, New York, 1981), p. 629; J. S. Harris, Jr., D. L. Miller, and P. M. Asbeck, *14th Conf. on Solid State Devices, Tokyo 1982*, (The Japan Society of Applied Physics, Tokyo, 1982) Digest of Papers, p. 199 (1982). D. L. Miller, P. M. Asbeck, and W. Petersen, *2nd Internat. Symposium on Molecular Beam Epitaxy and Related Clean Surface Techniques, Tokyo 1982*, (The Japan Society of Applied Physics, Tokyo, *1982*) Collected Papers, p. 121 (1982).
[3] W. V. McLevige, H. T. Yuan, W. M. Duncan, W. R. Frensley, F. H. Doerbeck, H. Morkoc, and T. J. Drummond, IEEE Electron Device Lett. **EDL-3**, 43 (1982).
[4] D. Ankri, W. Schaff, C. E. C. Wood, and L. F. Eastman, *GaAs and Related Compounds* (to be published).
[5] H. Kroemer, Arch. Elect. Übertraguug **8**, 499 (1954).
[6] H. Kroemer, RCA Rev. **18**, 332 (1957).
[7] For a review, see C. O. Bozler and G. D. Alley, Proc. IEEE **70**, 46 (1982).
[8] W. Shockley, U.S. Patent No. 2 569 347 (25 Sept. 1951).
[9] H. Kroemer, Proc. IRE **45**, 1535 (1957).

[10]D. Ankri and L. F. Eastman, Electron. Lett. **18**, 751 (1982). See also Ref. 4.
[11]J. O. McCaldin, T. C. McGill, and C. A. Mead, Phys. Rev. Lett. **36**, 56 (1976).
[12]W. A. Harrison, J. Vac. Sci. Technol. **14**, 1016 (1977).
[13]G. B. Scott and J. S. Roberts, Inst. Phys. Conf. Ser. **45**, 181 (1979); G. B. Scott, J. S. Roberts, and R. F. Lee, Appl. Phys. Lett. **37**, 30 (1980); J. S. Roberts, G. B. Scott, and J. P. Gowers, J. Appl. Phys. **52**, 4018 (1981); G. B. Scott, G. Duggan, and J. S. Roberts, J. Appl. Phys. **52**, 6312 (1981).
[14]Y. Kawamura, H. Asahi, and H. Nagari, Jpn. J. Appl. Phys. **20**, L807 (1981).
[15]P. Blood, J. S. Roberts, and J. P. Stagg, J. Appl. Phys. **53**, 3142 (1982).
[16]H. Asahi, Y. Kawamura, and H. Nagai, J. Appl. Phys. **53**, 4928 (1982).
[17]B. F. Levine, W. T. Tsang, C. G. Bethea, and F. Capasso, Appl. Phys. Lett. **41**, 470 (1982). This paper gives apparently complete references to earlier work on this topic.
[18]P. M. Asbeck, D. L. Miller, R. Asatourian, and C. G. Kirkpatrick, IEEE Electron. Device Lett. **EDL-3**, 403 (1982).
[19]See, for example: S. Hiyamizu and T. Mimura, J. Cryst. Growth **56**, 2 (1982); T. Mimura, Surf. Sci. **113**, 454 (1982); D. Delagebeaudeuf and N. T. Linh, IEEE Trans. Electron. Devices **ED-29**, 955 (1982).
[20]A. R. Calawa, B. A. Vojak, G. M. Metze, and M. J. Manfra (these proceedings).