History and Future Perspective of the Modern Silicon Bipolar Transistor

Tak H. Ning, *Fellow, IEEE*

Abstract—**A brief historical account of the development of advanced silicon bipolar transistors (SBTs) at IBM Research is described, with a focus on discussing the technical merits of the directions taken. A perspective on the future of silicon bipolar is given, including a discussion on the merits of SiGe-base transistors, and on the scaling limits of both Si-base and SiGe-base transistors. An apples-to-apples comparison of SiGe-base transistors and GaAs HBTs is made, showing that GaAs HBTs are inherently faster and more scaleable than SiGe-base transistors.**

I. Introduction

THE bipolar transistor was invented in 1947, and went through periods of rapid development in the 1950s and 1960s. However, by the mid-1970s, it was generally believed that silicon bipolar technology had become mature. Consequently, research activities on silicon bipolar technology, particularly in the United States, had fallen off very drastically, as evidenced by the lack of publications on the subject. Of the 281 papers published in the IEEE TRANSACTIONS ON ELECTRON DEVICES in the 12 months between September 1976 and August 1977, only 19 were on silicon bipolar, and of the 19, only five were by U.S. authors. Major universities and industrial laboratories in the U.S. either never had research programs on silicon bipolar to begin with, or had phased them out by then. With silicon bipolar being the backbone technology for mainframe computers, IBM had a very strong program on silicon bipolar development. However, in IBM Research, there was no research program on silicon bipolar at that time, even though there had been very strong programs on silicon MOSFET research for many years.

In 1976, upper management at IBM decided that IBM Research should establish a research program on silicon bipolar technology. In January 1977, an exploratory bipolar devices and circuit group[1] was formed at IBM's T. J. Watson Research Center, thus began the exciting years of research in silicon bipolar technology at IBM. The IBM team explored several versions of advanced device structures. However, by far the most successful version has been the double-polysilicon self-aligned bipolar transistor [1], [2], shown schematically in Fig. 1 with trench isolation and pedestal collector. Variations of this transistor have been used widely both inside and outside IBM.

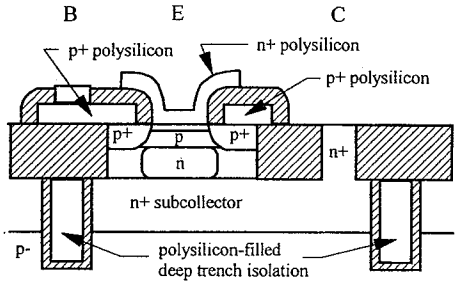


Fig. 1. Schematic cross-sectional view of a trench-isolated double-polysilicon self-aligned bipolar transistor with a pedestal collector (advanced bipolar transistor).

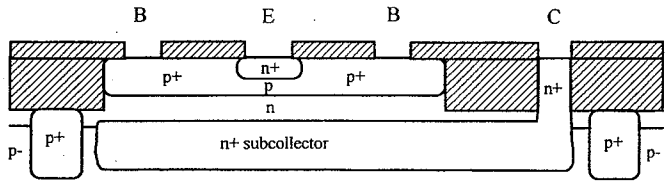


Fig. 2. Schematic cross-sectional view of a state-of-the-art bipolar transistor circa 1977 (conventional bipolar transistor). The base and collector contact windows are about 2.5 μm, and the emitter-stripe width is about 3 μm.

In this paper, a brief historical account of the development of the double-polysilicon self-aligned bipolar transistor is given, focusing on the merits of the key elements of the transistor. The theory that guided the transistor design and scaling effort is also discussed. Finally, a perspective on the future of silicon bipolar is given, including a discussion on its scaling limits and extension to SiGe-base, and a comparison of the SiGe-base transistor with the GaAs heterojunction bipolar transistor (HBT).

Although this paper focuses on the bipolar devices explored at IBM, the technical discussion applies to advanced silicon bipolar devices in general. Perspectives based on other advanced silicon bipolar devices can be found elsewhere in this special issue, and in the literature. For example, see [3].

II. The Conventional Bipolar Transistor Circa 1977

Fig. 2 shows the schematic of a state-of-the-art high-performance silicon bipolar transistor (SBT) circa 1977. Typically it has a 3×7 μm^2 emitter, base contacts on both sides of the emitter to reduce base resistance, recessed field oxide formed by a combination of silicon etching and oxidation, patterned n^+ subcollector for reducing collector resistance, and p^+ pockets

Manuscript received July 27, 1999; revised February 23, 200. The review of this paper was arranged by Editors P. Asbeck and T. Nakamura.

The author is with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: ning@watson.ibm.com).

Publisher Item Identifier S 0018-9383(01)09058-X.

[1]When the group was first formed in January 1977, it consisted of 6 members: H. N. Yu was acting manager of the group, D. D. Tang focused on device design and modeling; P. M. Solomon focused on circuit design and modeling, T. H. Ning focused on device technology, G. Feth focused on circuit applications, and M. G. Smith focused on system applications. A bit later, additions to the team included S. K. Wiedmann who focused on logic and memory circuits, and R. D. Isaac who focused on process technology development.

for isolation. Although the emitter area A_E is 21 μm^2, the base-collector junction area A_C is about 284 μm^2, giving an A_C/A_E ratio of about 14. The total transistor area, including isolation, is about 945 μm^2, giving a transistor/emitter area ratio of almost 45!

The transistors were typically formed as follows. A patterned n^+ subcollector is first formed on a p^- silicon wafer by diffusion. An n-type epitaxy layer is then grown on top. A p-type pocket for the base region is then formed by diffusion. The n^+ emitter is then formed by diffusion. The p-type region directly underneath the emitter forms the intrinsic base, while the remainder of the p-type pocket forms the extrinsic base. Typical emitter junction depth is about 500 nm, and typical intrinsic-base width is about 250 nm.

By the mid-1970s, the limitations of this conventional bipolar transistor were quite apparent and appeared difficult to overcome. With the advent of ion-implantation technology, formation of shallow base and emitter regions were quite feasible. However, because of shallow-emitter effects [4], reducing the emitter depth to less than 300 nm would increase the base current, and hence reduce the current gain, significantly. To counter the loss of current gain, the intrinsic-base sheet resistance would have to be increased, causing emitter-collector punch-through problems [5]. Thus, the conventional bipolar transistor is inherently not extendable to thin base widths.

The conventional transistor structure has a number of other limitations as well. With the extrinsic- and intrinsic-base regions formed from the same doping step, it is extremely difficult to optimize both regions simultaneously. Also, the large collector area leads to large parasitic capacitance. And, the large overall device area makes the transistor not suitable for dense memory applications.

III. Development of the Advanced Bipolar Transistor

At the time IBM formed a bipolar research group in January 1977, there were reports of several exciting developments in bipolar technology already. First, using doped polysilicon as a diffusion source to form and contact the emitter had been demonstrated [6]–[8]. It opened up a way for forming shallow emitters without worrying about metal penetration of the emitter. Then there were reports of self-aligning the emitter and base contacts employing doped polysilicon to form the emitter [9], [10]. This self-aligned transistor had an A_C/A_E ratio much smaller than that of the conventional transistor. The concept of using deep-trench isolation to replace p-pocket isolation was also evolving at the time [11], [12]. And, the pedestal-collector transistor, where the collector has a higher doping concentration directly underneath the emitter but a lower doping concentration elsewhere to minimize base-collector junction capacitance, had been demonstrated [13], [14].

The IBM team quickly decided that the advanced bipolar transistor should incorporate most or all of these features. The goal was to develop a high-performance manufacturable transistor that was scaleable to small dimensions. The final result was the transistor illustrated schematically in Fig. 1. The major milestones in the research phase of this transistor are discussed below.

A. Self-Aligned Polysilicon-Base Contact

After considering several self-alignment schemes, the IBM team decided to use p^+ polysilicon for forming and contacting the extrinsic base, with the emitter-base separation determined by a vertical sidewall insulator (see Fig. 1). This scheme has several advantages. 1) It minimizes the base-collector junction area by minimizing the emitter-base separation and by allowing the metal-to-base contact to be located over the field oxide. 2) It decouples the extrinsic base formation from the intrinsic base formation, allowing both to be optimized. 3) The emitter opening in the p^+-polysilicon layer provides a natural masking for ion implantation of the intrinsic base, emitter, and pedestal collector.

By 1977, reactive-ion etching (RIE) was already widely used in advanced technology development in IBM [15], [16]. RIE was used to form the vertical sidewall on the base polysilicon layer. The sidewall thickness was typically 0.2–0.3 μm.

B. Polysilicon Emitter Contact

A systematic study of the effect of emitter contact on shallow-emitter transistors was carried out. The conventional wisdom at the time was that heavily doped polysilicon behaves like metal in that a heavily-doped-polysilicon contact would behave like an ohmic contact. Consequently, the early experiments treated the heavily doped n^+ polysilicon as a contact material for the shallow emitter, and not as the emitter itself. Perhaps the most exciting finding of these experiments [17], [18] was that the n^+ polysilicon contacts to silicon did not behave like ohmic contacts at all. Minority holes injected from the base into the shallow n-type emitter, instead of recombining at the polysilicon-silicon interface, as expected for an ohmic contact, recombine primarily inside the n^+ polysilicon layer, leading to significant increase in current gain.

Subsequent experiments showed that a polysilicon emitter contact improves the current gain of a transistor without affecting its speed [19], [20]. This result is very significant for it demonstrated for the first time that shallow-emitter transistors (transistors having shallow single-crystalline emitter regions) could be made without the problem of insufficient current gain.

C. Polysilicon Emitter

Nowadays, all advanced SBTs employ polysilicon emitter, where the shallow-emitter doping step is skipped altogether. Instead, a heavily doped n-type polysilicon layer is formed directly on the p-type base layer, followed by just sufficient thermal cycle to obtain low emitter resistance. In IBM, the polysilicon emitter was discovered unexpectedly. A series of experiments, illustrated in Fig. 3, was designed to study the differences among implanted-and-diffused shallow emitters contacted by silicide, aluminum, or arsenic-doped polysilicon. To avoid any ambiguity, all three kinds of emitter contacts were designed to be on the same wafers. The masking procedure was such that one of the devices received no emitter implant at all but had the n^+ polysilicon layer formed directly on the p-type base layer. This device was not supposed to work, since the conventional wisdom at the time suggested that the polysilicon contact would behave like an ohmic contact, and the device would not have meaningful current gain. However, we found

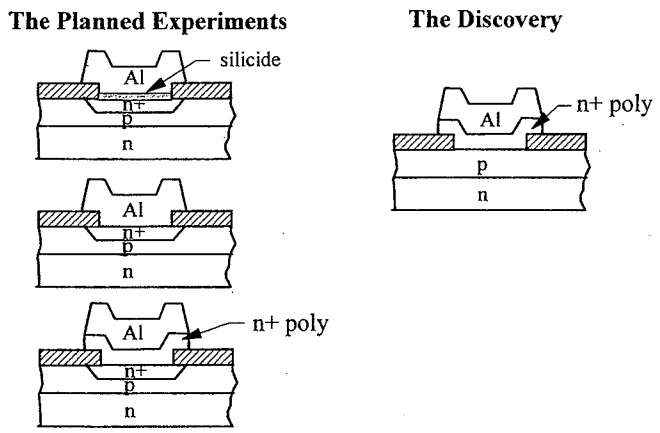


Fig. 3. Schematics illustrating the emitter-contact experiments in which the polysilicon emitter was discovered.

TABLE I
TYPICAL MEASURED CURRENT GAINS OF THE BIPOLAR TRANSISTORS SKETCHED IN FIG. 3, FOR A BASE SHEET RESISTANCE VALUE OF ABOUT 7 KΩ/□

Emitter stack	Al/silicide/n+	Al/n+	Al/n+ poly/n+	Al/n+ poly (polysilicon emitter)
Current gain	40 - 60	65 - 70	145 - 170	~ 400

to our great surprise that the device not only worked properly but also showed the largest current gain of all the transistors! (See Table I.)

The advent of polysilicon emitter marked a major milestone in the evolution of the SBT. With polysilicon emitter, emitter junctions (the n^+ emitter region in the crystalline silicon) can be very shallow, typically only 30 nm or less. Such ultra-shallow emitters enable very thin intrinsic-base layers to be formed. And, with polysilicon emitter, current gain ceased to be a problem, overcoming a fundamental limitation in scaling bipolar transistors. In IBM, we have been using polysilicon emitter since 1981.

The physics and technology of polysilicon emitter have been the subject of intense study in universities and industrial laboratories worldwide. Most of the papers up to 1989 were collected in one publication [21]. It is a good reference and contains an excellent introduction to the theoretical and experimental aspects of polysilicon-emitter bipolar transistors.

D. Bipolar Transistor Design, Scaling, and the Pedestal Collector

Base-widening effect on f_T was demonstrated in 1962 [22]. However, to demonstrate the detrimental effect of base widening on a digital circuit requires the transistors to be operated at sufficiently large current densities where diffusion capacitance dominates the circuit delay. This usually implies operating the transistors at a collector current density of $J_C > 0.3qv_{\text{sat}}N_C$ [4], where v_{sat} is the saturated velocity of electrons and N_C is the collector doping concentration. For a transistor circa 1977, with $A_E = 20\ \mu\text{m}^2$ and $N_C = 2 \times 10^{16}\ \text{cm}^{-3}$, base widening is not significant until the operating current is larger than 2 mA. Even

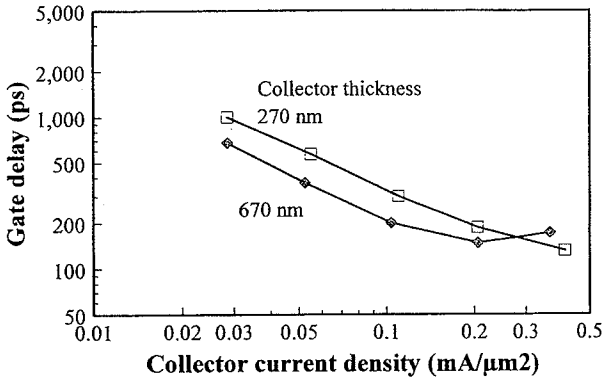


Fig. 4. Typical switching delay as a function of collector current density for advanced bipolar transistors, with collector thickness as a parameter (after [23]).

at these currents, the circuit delays are often still dominated by the load capacitance instead of diffusion capacitance.

With the double-polysilicon self-aligned bipolar transistor, the load-capacitance delay component has been reduced sufficiently that base widening can easily dominate circuit delays. This is demonstrated in Fig. 4, which shows the circuit delay versus collector current for two transistors of different collector thickness (thickness between the intrinsic base and the subcollector). The transistor with the thicker collector has a smaller base-collector junction capacitance but larger minority-carrier-storage volume in the collector region. This transistor slows down as the collector current increases once the current density is large enough to cause significant base widening [23].

The results shown in Fig. 4 imply that a device design optimized for one operating point may be far from optimum for another operating point. This important point is central to a new procedure for designing the optimal bipolar transistor [24]. The procedure "synthesizes" the optimal transistor instead of analyzes a given design. It optimizes the collector doping concentration for the intended operating current density. A theory for scaling this optimally designed transistor to smaller dimensions was also developed [25]. Several constraints and requirements are imposed in scaling the transistor. These constraints and requirements and the resulting scaling rules for ECL circuits are shown in Tables II and III [25]. Here V is the power supply voltage, ΔV is the logic swing, C_{DE} is the emitter diffusion capacitance, and C_{dBC} is the base-collector junction depletion-layer capacitance. The power supply voltage remains constant in bipolar scaling because the turn-on voltage of a p-n diode is relatively independent of its area. Reducing the diode area by 10× increases its turn-on voltage by only 60 mV.

The scaling theory serves as a valuable guide for understanding the potential and limitation of small-dimension bipolar transistors. It indicates that circuit speed can be improved in proportion to the scaling factor κ if both collector current density and doping concentration are increased in proportion to κ^2. To increase collector doping concentration and yet minimize base-collector junction capacitance, the pedestal collector design [13] was used. As discussed earlier, the self-aligned polysilicon-base transistor allows the pedestal collector to be formed readily by ion implantation. In the literature, the implanted pedestal collector is often referred to as self-aligned implanted collector (SIC).

TABLE II
CONSTRAINTS AND REQUIREMENTS IN ECL SCALING (AFTER [25])

Parameter	Constraint or requirement
Voltage	V, ΔV = constant
Capacitance	C_{DE}/C_{dBC} = constant
Base doping concentration	$N_B \propto W_B^{-2}$
Collector doping concentration	$N_C \propto J_C$

TABLE III
SCALING RULES FOR ECL CIRCUITS (AFTER [25])

Parameter	Scaling rule (Scaling factor $\kappa > 1$)
Feature size or emitter-stripe width	$1/\kappa$
Base width W_B	$1/\kappa^{0.8}$
Collector current density J_C	κ^2
Circuit delay	$1/\kappa$

E. Deep Trench Isolation

Even with recessed field oxide, the p^+ isolation pockets in the conventional transistor take up a lot of area because doping impurities diffuse laterally as well as vertically. A 2-μm-deep p^+ isolation formed by diffusion through a 1-μm wide window typically has a lateral dimension of about 4 μm. In addition, the p^+ pockets must not be too close to the n^+ subcollector in order to minimize collector-substrate junction capacitance.

The isolation area can be reduced significantly with deep-trench isolation. The advanced transistor structure shown in Fig. 1 has a device area that is typically about half that of the conventional transistor shown in Fig. 2. This area reduction is particularly important in memory applications.

There are other benefits of using trench isolation as well. For example, with trench isolation, there is no need to pattern the subcollector. A blanket subcollector layer is patterned automatically by the isolation trenches. Also, the collector-substrate junction capacitance of a trench-isolated transistor is reduced to a relatively insignificant level.

Deep-trench widths are determined by the lithography rules used. The deep trenches are formed by first etching the silicon trenches, then filling the trenches with oxide or with a combination of oxide and polysilicon, followed by planarization using chemical-mechanical polishing. In IBM, the first successful integration of all the advanced elements of the bipolar transistor shown schematically in Fig. 1 was carried out using 1.25-μm lithography [1].

IV. SCALING LIMITS OF SBTS

Bipolar scaling theory [25] suggests that SBTs can be scaled down in size and achieve performance improvement in proportion to the scaling factor κ. However, there are several limitations to bipolar scaling. These include limits due to collector current density, base-collector junction avalanche, chip power density, and emitter series resistance [4]. For device performance, it is the base-collector junction avalanche effect [26] that limits the scaling of silicon bipolar.

The limitation due to avalanche breakdown can be understood as follows. In bipolar scaling, voltages remain constant while collector current density and doping concentration increase as κ^2 in order to achieve performance improvement in proportion to κ. This leads to increased base-collector junction avalanche. Once the base-collector junction avalanche limit is reached, collector doping concentration, and hence collector current density, cannot be increased further, leading to limited performance improvement in further scaling.

Often there are reports of silicon transistors running at "record" speeds. Unfortunately, many of these reports do not state clearly if the transistors also have acceptable breakdown voltages. Without such information, it is impossible to judge the significance of the record speeds claimed.

V. BIPOLAR TRANSISTOR BEYOND SILICON

A. SiGe-Base Bipolar Transistor

In recent years, the most exciting development beyond the double-polysilicon self-aligned bipolar transistor is perhaps the SiGe-base bipolar transistor. The first successful SiGe-base transistor was made using an MBE process to form the SiGe layer [27]. With the advent of UHV-CVD for forming SiGe layers [28], progress in the development of SiGe-base transistors accelerated [29]–[31]. It should be noted that all advanced SiGe-base transistors have been built upon the foundations of advanced Si-base transistors. That is, they employ polysilicon base contact to reduce area and capacitance, and polysilicon emitter to ensure adequate current gain and to achieve thin intrinsic base [29]–[31]. The most advanced SiGe-base transistors use the same self-alignment and deep-trench isolation schemes as Si-base transistors [31]–[33].

A linearly graded Ge profile is often used in SiGe-base bipolar transistors [29]–[31]. The Ge profile and energy-band diagram of a typical SiGe-base transistor are illustrated in Fig. 5. Also shown for comparison are the same schematics for a Si-base transistor. The base bandgaps of the two transistors are about the same near the base-emitter junction. The base bandgap of the SiGe-base transistor narrows gradually toward the base-collector junction. Both transistors employ the same polysilicon emitter. Since the base current of a bipolar transistor is a function of the emitter parameters only, and is independent of the base parameters [4], the Si-base and SiGe-base transistors have the same base current. Only the collector currents, which depend on the base parameters, are different for the two transistors. In other words, the commonly practiced SiGe-base transistor is NOT a true heterojunction bipolar transistor (HBT) since it does not make use of the heterojunction to provide a wide-bandgap emitter. A wide-bandgap emitter

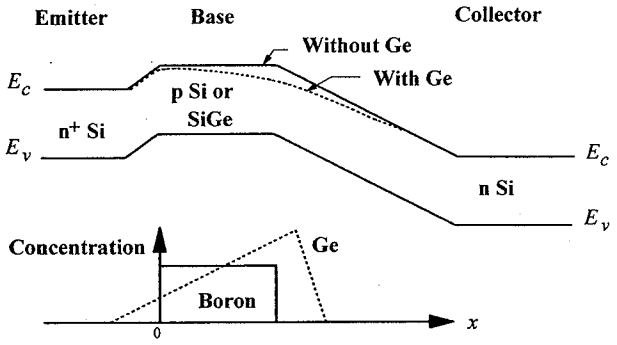


Fig. 5. Schematics showing the energy bands and Ge profile of a typical SiGe-base transistor (dashed). Also shown for comparison are the energy bands of Si-base transistor (solid) (after [30]).

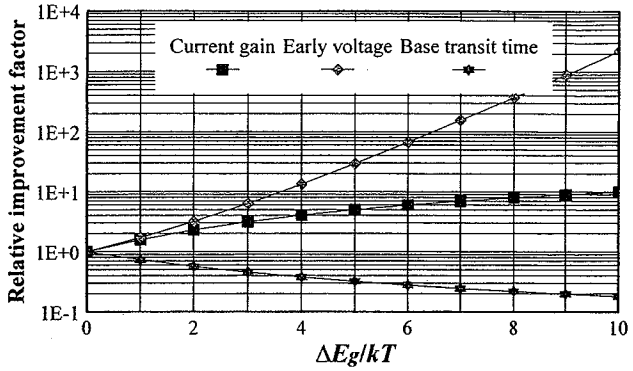


Fig. 6. Relative improvements in current gain, Early voltage, and base transit time due to SiGe base as a function of maximum base bandgap grading across the intrinsic-base layer.

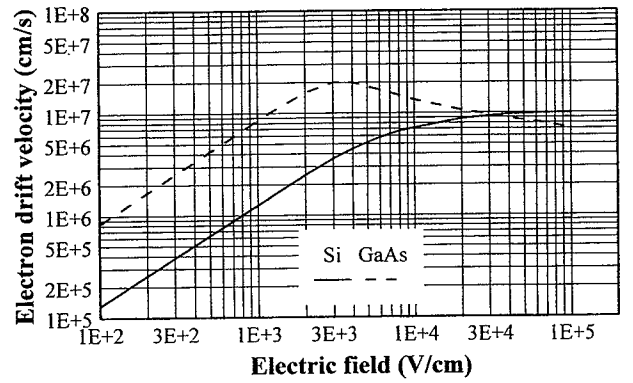


Fig. 7. Electron drift velocity versus electric field for silicon and GaAs (after [37]).

would have resulted in greatly reduced base current. Instead, it is a graded-base-bandgap bipolar transistor. This distinction is important when we compare SiGe-base transistors with true HBTs made of compound semiconductors. (See next subsection.)

Compared to Si-base transistors, SiGe-base transistors offer greatly improved device characteristics for analog and high-frequency applications. For a linearly graded Ge profile, the improvement factors in current gain β, Early voltage V_A, and base transit time t_B, are given by (1)–(3), respectively, where $\Delta E_{g,\mathrm{SiGe}}$ is the maximum bandgap grading across the SiGe-base region [4]. Fig. 6 is a plot of these improvement factors as a function of $\Delta E_{g,\mathrm{SiGe}}/kT$. For digital logic applications, current gain and Early voltage have relatively little effect on circuit speed [20], although base transit time does have some effect [34]. SiGe-base transistors have the same scaling limitations as Si-base transistors. Both are limited by the conflicting requirements of suppressing base-widening and maintaining acceptable base-collector junction avalanche:

$$\frac{\beta(\mathrm{SiGe})}{\beta(\mathrm{Si})} = \frac{\Delta E_{g,\mathrm{SiGe}}/kT}{1-\exp(-\Delta E_{g,\mathrm{SiGe}}/kT)} \tag{1}$$

$$\frac{V_A(\mathrm{SiGe})}{V_A(\mathrm{Si})} = \frac{kT}{\Delta E_{g,\mathrm{SiGe}}}[\exp(\Delta E_{g,\mathrm{SiGe}}/kT)-1] \tag{2}$$

$$\frac{t_B(\mathrm{SiGe})}{t_B(\mathrm{Si})} = \frac{2kT}{\Delta E_{g,\mathrm{SiGe}}} \times \left[1-\frac{kT}{\Delta E_{g,\mathrm{SiGe}}}(1-\exp(-\Delta E_{g,\mathrm{SiGe}}/kT))\right]. \tag{3}$$

B. Comparison With GaAs HBT

Ever since the first successful demonstration of SiGe-base transistor more than ten years ago [27], there have been many reports comparing the performance of SiGe-base transistors and GaAs HBTs. Unfortunately, instead of clarifying the merits and limitations of the two transistors, these reports often cause more confusion. For a fair comparison, the devices should be of comparably advanced structure and design rule, and the comparison should be made at the same collector current densities. Once the two transistors have been "normalized" in terms of device structure, design rule, and current density, a fair comparison can be made simply by comparing the intrinsic device parameters [35].

Recently, GaAs HBT devices comparable in structure to the silicon transistor shown in Fig. 1 have been reported [36]. Here we discuss the comparison results for such advanced devices.

Fig. 7 shows the electron drift velocity in silicon and GaAs as a function of electric field [37]. At low fields electrons in GaAs have much higher velocity, and higher mobility, than electrons in Si. However, for fields greater than 10^4 V/cm, which are typical in the base-collector junction depletion region, the velocities are comparable. Some other relevant properties of Si and GaAs, as well as device parameters of a typical SiGe-base transistor and a typical GaAs HBT are compared in Table IV. To first order, the two materials have about the same dielectric constant, suggesting that comparison of the junction capacitances can be made simply by comparing the junction doping concentrations.

For the emitter-base diode, a SiGe-base transistor has the same emitter-base junction capacitance, C_{dBE}, as a Si-base transistor. On the other hand, the emitter-base heterojunction of a GaAs HBT suppresses the base current and allows its emitter to be very lightly doped. Therefore, a GaAs HBT has a much smaller C_{dBE} than a SiGe-base transistor.

For the intrinsic-base region, the emitter-base heterojunction allows the intrinsic base of a GaAs HBT to be much more heavily doped than the intrinsic base of a Si-base transistor. For the parameters shown in Table IV, the base resistance, r_b, of a GaAs HBT is $10\times$ smaller than that of a SiGe-base transistor.

As for the base transit time, SiGe-base bipolar transistor benefits from its graded base bandgap, as discussed earlier. How-

TABLE IV
COMPARISON OF Si AND GaAs, AND SiGe-BASE BIPOLAR AND GaAs HBT

Properties	Si and SiGe-base Tx	GaAs and GaAs HBT
Dielectric constant	11.9	13.1
Energy gap (eV)	1.12	1.424
Typical emitter doping conc. (cm^{-3})	2×10^{20}	5×10^{17}
Typical base doping conc. (cm^{-3})	2×10^{18}	4×10^{19}
Typical base layer resistivity (Ω-cm)	3×10^{-2}	3×10^{-3}
Electron drift mobility (cm^2/V-s)	1500	8500
(at typical base doping conc.)	(~ 200 at $2x10^{18}$ cm^{-3})	(~ 700 at $4x10^{19}$ cm^{-3})

ever, it should be noted that graded base bandgap is also possible in GaAs HBT [38]. Even without base-bandgap grading, GaAs HBT benefits from the higher electron mobility. For the parameters in Table IV and Fig. 6, it can be shown that a SiGe-base transistor with $\Delta E_{g,\mathrm{SiGe}} = 150$ meV has about the same t_B as a GaAs HBT with the same base width but without base-bandgap grading.

For the collector region, the two transistors have about the same collector doping concentration when designed to operate at the same current density. This implies that the two transistors should have about the same C_{dBC} and the same base-collector junction depletion-layer transit time, t_{BC}.

The cutoff frequency is given by

$$f_T = \frac{1}{2\pi}\left(\tau_F + \frac{kT}{qI_C}(C_{\mathrm{dBE}} + C_{\mathrm{dBC}})\right)^{-1} \quad (4)$$

and the maximum oscillation frequency is given by

$$f_{\max} = \left(\frac{f_T}{8\pi r_b C_{\mathrm{dBC}}}\right)^{1/2} \quad (5)$$

where $\tau_F = (t_B + t_E + t_{\mathrm{BE}} + t_{\mathrm{BC}}) \cong (t_B + t_{\mathrm{BC}})$ is the sum of the transistor transit times [4]. As discussed earlier, C_{dBE} is smaller for a GaAs HBT than for a SiGe-base transistor, but the two transistors should have about the same C_{dBC} and about the same t_{BC}. Also, for the parameters shown in Table IV, the two transistors have about the same t_B, while r_b is much smaller for the GaAs HBT. Therefore, when compared at the same collector current densities, we should expect a GaAs HBT to show higher f_T and $f_{\max}$ than a SiGe-base transistor, if the transistors are comparably advanced.

Fig. 8 is a plot of f_T as a function of collector current density for a SiGe-base transistor [39] and a GaAs HBT [36]. Both transistors are comparably advanced. Fig. 8 clearly shows that the GaAs HBT has higher f_T than the SiGe-base transistor. Moreover, the larger energy gap of GaAs means that GaAs HBT is much less susceptible to base-collector junction avalanche effect. Therefore, a GaAs HBT can be designed to operate at much higher collector-current densities than a SiGe-base transistor and still meets the breakdown voltage requirements. That is, GaAs HBTs can be scaled down to smaller dimensions than Si-base or SiGe-base transistors.

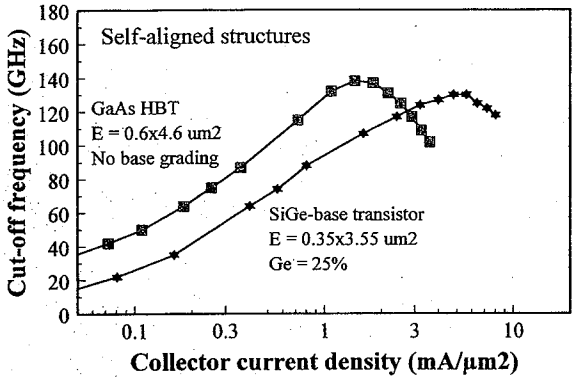


Fig. 8. f_T as a function of collector current density for a GaAs HBT (after [36]) and a SiGe-base bipolar transistor (after [39]) of comparably advanced device structures.

Compared to GaAs HBTs, the real advantage of SiGe-base bipolar transistors is their compatibility with silicon VLSI processes. SiGe-base bipolar transistors can be low cost, if produced in large volumes, and can be readily integrated with CMOS devices on the same chip to produce SiGe-base BiCMOS.

VI. SUMMARY

A brief account of the development of advanced silicon bipolar in IBM Research has been given. A modern SBT typically has the following salient features: 1) self-aligned polysilicon base contact; 2) polysilicon emitter; 3) pedestal collector; and 4) deep-trench isolation. Replacing the Si-base by SiGe-base greatly improves the current gain, Early voltage, and frequency response of the transistor. However, both the Si-base and the SiGe-base transistors are limited in scaling by the conflicting requirements of maintaining acceptable base-collector junction avalanche and minimizing base widening.

The commonly practiced SiGe-base transistor is not a true HBT, but a graded-base-bandgap bipolar transistor. As a result, a GaAs HBT inherently has higher f_T and $f_{\max}$ than a SiGe-base transistor when compared at same current densities. Furthermore, GaAs HBT is inherently more scaleable than SBT because of the larger energy gap of GaAs.

REFERENCES

[1] D. D. Tang, P. M. Solomon, T. H. Ning, R. D. Isaac, and R. E. Burger, "1.25 μm deep-groove-isolated self-aligned bipolar circuits," *IEEE J. Solid-State Circuits*, vol. SC-17, pp. 925–931, 1982.

[2] G. P. Li, T. H. Ning, C. T. Chuang, M. B. Ketchen, D. D. Tang, and J. Mauer, "An advanced high-performance trench-isolated self-aligned bipolar technology," *IEEE Trans. Electron Devices*, vol. ED-34, pp. 2246–2254, 1987.

[3] "Special issue on bipolar and BiCMOS/CMOS devices and technologies," *IEEE Trans. Electron Devices*, vol. 42, Mar. 1995.

[4] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[5] S. P. Gaur, "Performance limitations of silicon bipolar transistors," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 415–421, 1979.

[6] M. Takagi, K. Nakayama, C. Tevada, and H. Kamioko, "Improvement of shallow base transistors technology by using a doped polysilicon diffusion source," *J. Jpn. Soc. Appl. Phys. (Suppl.)*, vol. 42, pp. 101–109, 1972.

[7] J. Graul, A. Glasl, and H. Murrmann, "Ion implanted bipolar high performance transistors with POLYSIL emitter," in *IEDM Tech. Dig.*, 1975, pp. 450–454.

[8] ——, "High-performance transistors with arsenic-implanted poly emitters," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 491–495, 1976.
[9] T. Sakai, Y. Sunohara, Y. Sakakibara, and J. Murota, "Stepped electrode transistor," in *Ext. Abst. 8th Int. Conf. on Solid State Devices*, Tokyo, Japan, 1976, pp. 13–14.
[10] T. Sakai, Y. Sunohara, H. Nakamura, and T. Sudo, "A 100 ps bipolar logic," in *ISSCC Tech. Dig.*, 1977, pp. 196–197.
[11] J. A. Bondur and H. B. Pogge, "Method for forming isolated regions of silicon utilizing reactive ion etching," U.S. Patent 4 104 086, Aug. 1, 1978.
[12] H. B. Pogge, "Deep-groove isolation," presented at the Recent Newspaper, ECS Ext. Abst., Boston, MA, May 1979.
[13] H. N. Yu, "Transistor with limited-area base-collector junction," U.S. Patent Re. 27 045 re-issued, Feb. 2, 1971.
[14] H. N. Ghosh, K. G. Ashar, A. S. Oberai, and D. DeWitt, "Desigh and development of an ultralow-capacitance high-performance pedestal transistor," *IBM J. Res. Develop.*, vol. 15, pp. 436–441, 1971.
[15] L. Zielinski and G. C. Schwartz, "Reactive ion etching," in *ECS Extended Abstracts*, vol. 75-1. Toronto, ON, Canada, 1975, pp. 117–118.
[16] H. A. Clark, "Plasma processing at moderate vacuum," in *ECS Ext. Abst.*, vol. 75-1. Toronto, ON, Canada, 1975, pp. 119–121.
[17] T. H. Ning and R. D. Isaac, "Effect of emitter contact on current gain of silicon bipolar devices," in *IEDM Tech. Dig.*, 1979, pp. 473–476.
[18] ——, "Effect of emitter contact on current gain of silicon bipolar devices," *IEEE Trans. Electron Devices*, vol. ED-27, pp. 2051–2055, 1980.
[19] T. H. Ning, R. D. Isaac, P. M. Solomon, D. D. Tang, and H. N. Yu, "Self-aligned npn bipolar transistors," in *IEDM Tech. Dig.*, 1980, pp. 823–824.
[20] T. H. Ning, R. D. Isaac, P. M. Solomon, D. D. Tang, H. N. Yu, G. C. Feth, and S. K. Wiedmann, "Self-aligned bipolar transistors for high-performance and low-power-delay VLSI," *IEEE Trans. Electron Devices*, vol. ED-28, pp. 1010–1013, 1981.
[21] A. K. Kapoor and D. J. Roulston, Eds., *Polysilicon Emitter Bipolar Transistors*. New York: IEEE Press, 1989.
[22] C. T. Kirk, Jr., "A theory of transistor cutoff frequency (f_T) falloff at high current densities," *IEEE Trans. Electron Devices*, vol. ED-9, pp. 164–174, 1962.
[23] D. D. Tang, K. P. MacWilliams, and P. M. Solomon, "Effects of collector epitaxial layer on the switching speed of high-performance bipolar transistors," *IEEE Electron Device Lett.*, vol. EDL-4, pp. 17–19, 1983.
[24] D. D. Tang and P. M. Solomon, "Bipolar transistor design for optimized power-delay logic circuits," *IEEE J. Solid-State Circuits*, vol. SC-14, pp. 679–684, 1979.
[25] P. M. Solomon and D. D. Tang, "Bipolar circuit scaling," in *ISSCC Tech. Dig.*, 1979, pp. 86–87.
[26] P.-F. Lu and T.-C. Chen, "Collector-base junction avalanche effects in advanced double-poly self-aligned bipolar transistors," *IEEE Trans. Electron Devices*, vol. 36, pp. 1182–1188, 1989.
[27] S. S. Iyer, G. L. Patton, S. S. Delage, S. Tiwari, and J. M. C. Stork, "Silicon-Germanium base heterojunction bipolar transistors by molecular beam epitaxy," in *IEDM Tech. Dig.*, 1987, pp. 74–876.
[28] B. S. Meyerson, "Low-temperature silicon epitaxy by ultra-high vacuum/chemical vapor deposition," *Appl. Phys. Lett.*, vol. 48, pp. 797–799, 1986.
[29] G. L. Patton, J. M. C. Stork, J. H. Comfort, E. F. Crabbé, B. S. Meyerson, D. L. Harame, and J. Y.-C. Sun, "SiGe-base heterojunction bipolar transistors: Physics and design issues," in *IEDM Tech. Dig.*, 1990, pp. 13–16.
[30] D. L. Harame, J. H. Comfort, J. D. Cressler, E. F. Crabbé, J. Y.-C. Sun, B. S. Meyerson, and T. Tice, "Si/SiGe epitaxial-base transistors—Part I: Materials, physics, and circuits," *IEEE Trans. Electron Devices*, vol. 42, pp. 455–468, 1995.
[31] ——, "Si/SiGe epitaxial-base transistors—Part II: Process integration and analog applications," *IEEE Trans. Electron Devices*, vol. 42, pp. 469–482, 1995.
[32] T. C. Chen, E. Ganin, H. Stork, B. Meyerson, J. D. Cressler, J. Warnock, D. Harame, G. Patton, G. P. Li, C. T. Chuang, and T. H. Ning, "Submicrometer Si and Si-Ge epitaxial-base double-poly self-aligned bipolar transistors," *IEEE Trans. Electron Devices*, vol. 38, pp. 941–943, 1991.
[33] F. Sato, T. Hashimoto, T. Tatsumi, H. Kitahata, and T. Tashiro, "Sub-20 psec ECL circuits with 50 GHz f_{max} self-aligned SiGe HBTs," in *IEDM Tech. Dig.*, 1992, pp. 397–400.
[34] C. T. Chuang, K. Chin, J. M. C. Stork, G. L. Patton, E. F. Crabbé, and J. H. Comfort, "On the leverage of high-f_T transistors for advanced high-speed bipolar circuits," *IEEE J. Solid-State Circuits*, vol. 27, pp. 225–228, 1992.
[35] T. H. Ning, "Trade-offs between SiGe and GaAs bipolar ICs," in *Proc. Fourth Int. Conf. Solid State and Integrated-Circuit Technology*, Beijing, China, 1995, pp. 434–438. Invited Paper.
[36] T. Oka, K. Hirata, K. Ouchi, H. Uchiyama, K. Mochizuki, and T. Nakamura, "InGaP/GaAs HBT's with high-speed and low-current operation fabricated using WSi/Ti as the base electrode and buried SiO_2 in the extrinsic collector," in *IEDM Tech. Dig.*, 1997, pp. 739–742.
[37] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981.
[38] J. R. Hayes, F. Capasso, A. C. Gossard, R. J. Malik, and W. Wiegmann, "Bipolar transistor with graded band-gap base," *Electron. Lett.*, vol. 19, pp. 410–411, 1983.
[39] K. Oda, E. Ohue, M. Tanabe, H. Shimamoto, T. Onai, and K. Washio, "130-GHz f_T SiGe HBT technology," in *IEDM Tech. Dig.*, 1997, pp. 791–794.

Tak H. Ning (M'75–SM'81–F'87) received the Ph.D. degree in physics from the University of Illinois at Urbana-Champaign in 1971.

He joined IBM Thomas J. Watson Research Center, Yorktown Heights, NY, in 1973, where he is currently an IBM Fellow.

Dr. Ning has made contributions to silicon bipolar, CMOS, and DRAM technologies, for which he has been recognized with numerous awards, including the 1991 IEEE Jack A. Morton Award and the 1989 IEEE Electron Devices Society J. J. Ebers Award. He is a member of the National Academy of Engineering, and a Fellow of the American Physical Society.

6.3.3.2 BASE WIDENING AT HIGH CURRENTS

At high current densities, the assumption of Δn being small compared to N_C is no longer valid, and the above equations cannot be used to estimate the base-widening effect. With the mobile-charge concentration comparable to or larger than the fixed ionized-impurity concentration, the depletion approximation is certainly not valid. Furthermore, the excess electrons in the n-type collector can produce a substantial electric field in the collector, according to Eq. (6.4), and the classical concept of a well-defined junction boundary in the base–collector diode is no longer valid. Also, in order to maintain quasineutrality, the excess electrons induce an excess of holes in the n-type collector. The region of the collector with excess holes becomes an extension of the p-type base. In other words, the base region widens into the collector region, until it reaches the subcollector where the excess electron concentration is small compared with the n-type doping concentration. As a result, the high-field region, originally located at the physical base–collector junction, is relocated to near the collector–subcollector intersection (Poon *et al.*, 1969). The numerical simulation results (Poon *et al.*, 1969) shown in Fig. 6.11 illustrate clearly the effects of base widening at high currents. They show that the relocation of the high-field region is accompanied by a buildup of excess electrons and holes in the collector region.

It is instructive to estimate the collector current density at which substantial base widening occurs. The saturated velocity v_{sat} for electrons in silicon is about 1×10^7 cm/s, as indicated in Fig. 2.9. At low collector currents, the maximum electron concentration in the n-type collector region is equal to the collector doping concentration N_C. The maximum electron current density that can be supported by an electron concentration N_C is $J_{\max} = q v_{sat} N_C$. When the injected electron current density approaches $J_{\max}$, there is an increase of the electron concentration in excess of N_C in order to support the injected electron current flow. As the excess electrons build up, there is a buildup of excess holes in order to maintain quasineutrality, and a relocation of the high-field region. The results shown in Fig. 6.11 suggest that significant base widening starts at a collector current density of approximately $0.3 J_{\max}$. This value is consistent with the reported peak cutoff-frequency data for modern VLSI bipolar devices (Crabbé *et al.*, 1993). Thus, ***to avoid significant base widening, a bipolar transistor should not be operated at collector current densities approaching $J_{\max}$.*** For a relatively high N_C of 10^{17} cm^{-3}, $J_{\max}$ is about 1.6 mA/μm^2. To avoid significant base widening, J_C should be less than about 0.5 mA/μm^2.

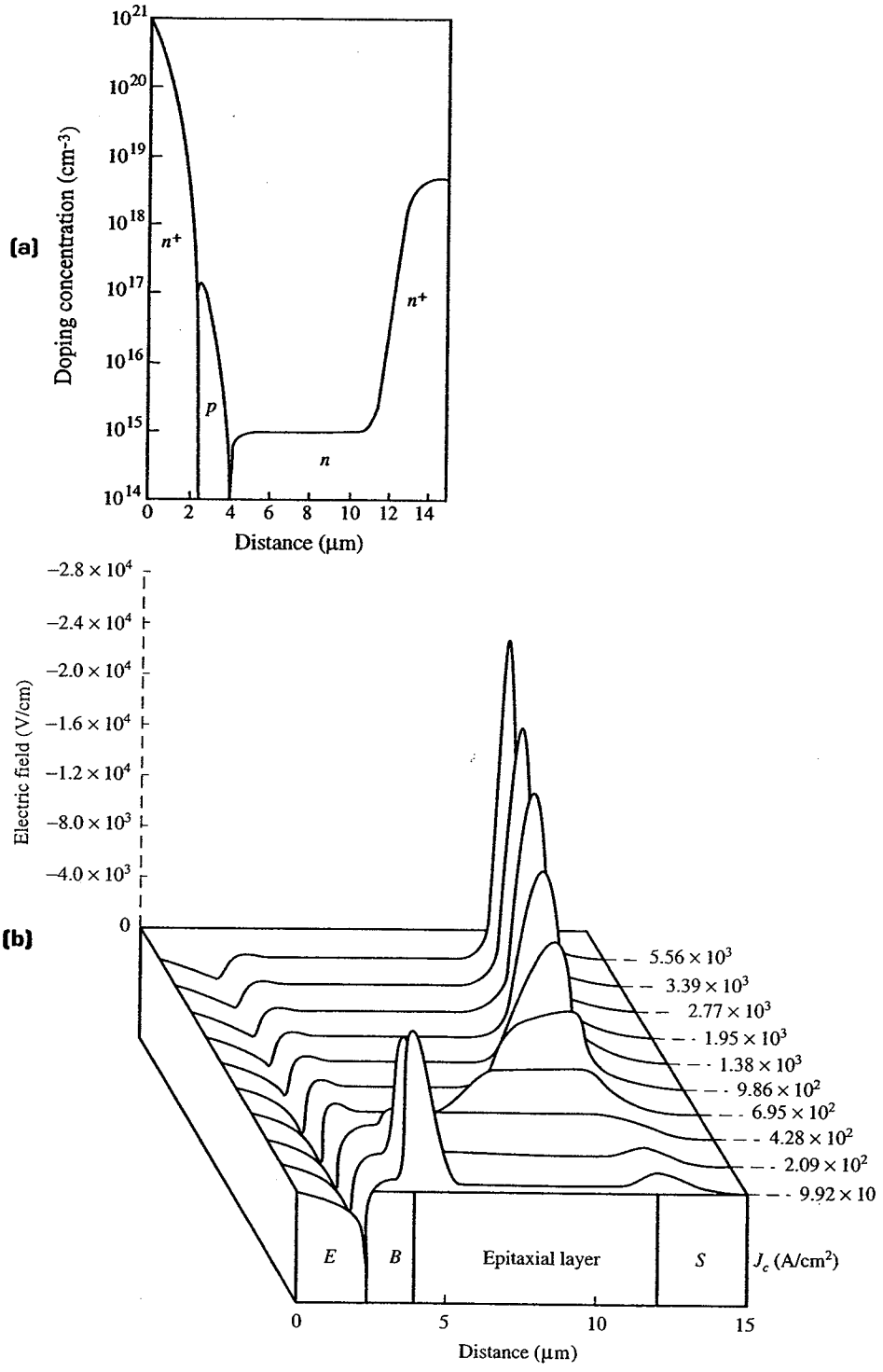
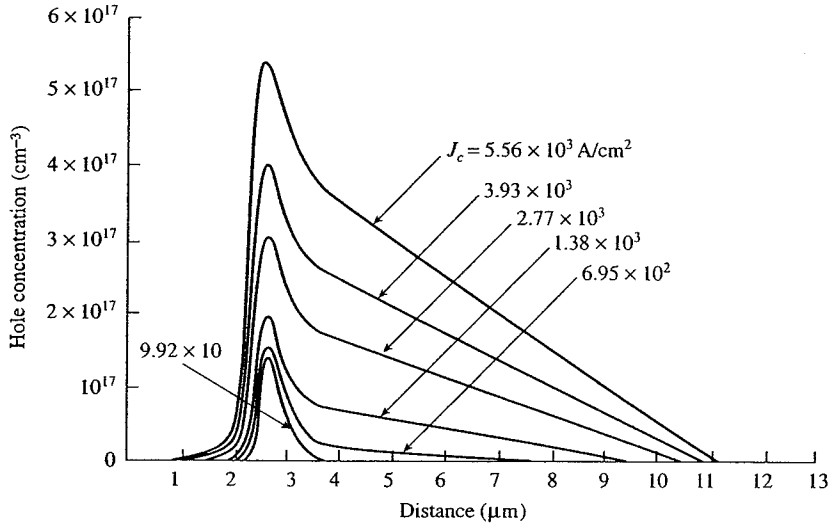
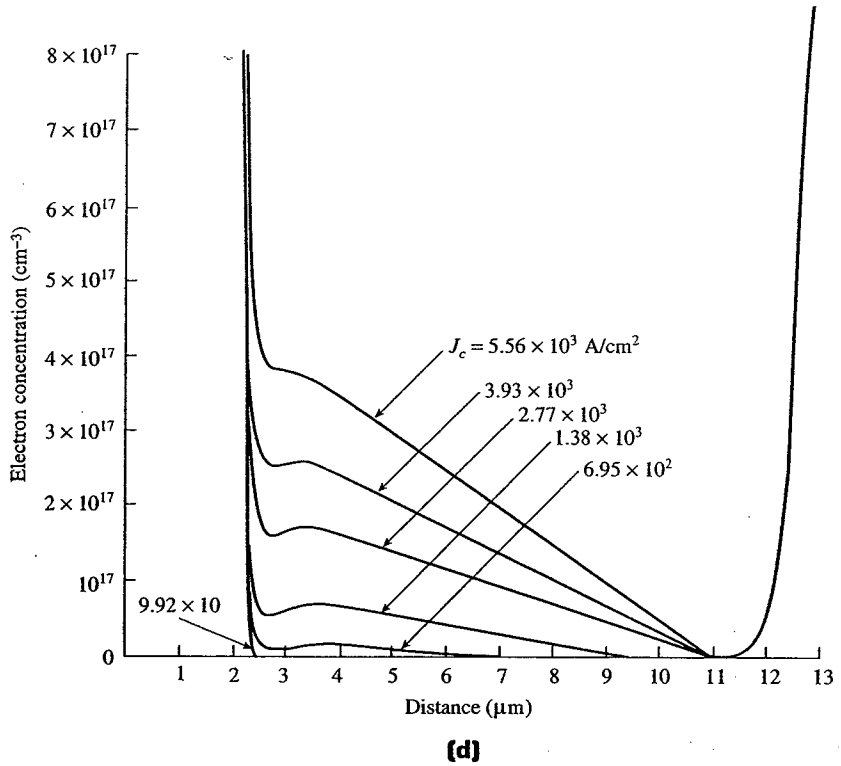


FIGURE 6.11. Numerical simulation results showing the effects of base widening in an n–p–n transistor at high collector current densities: (a) the doping profiles of the device simulated, (b) relocation of the high-field region from the physical base–collector junction to the collector–subcollector intersection, (c) buildup of excess holes in the collector, and (d) buildup of excess electrons in the collector. (After Poon *et al.*, 1969).



(c)



(d)

FIGURE 6.11. (*Continued*)