HETEROJUNCTION BIPOLAR TRANSISTOR

The heterojunction bipolar transistor (HBT) differs from the traditional homojunction bipolar transistor (BJT) in that at least one of its two junctions is formed between dissimilar semiconductor materials. Table 1 lists some commonly used material combinations. Effective transistor action demands that the junctions in the device not be sites for significant electron–hole recombination. This requires good lattice-matching at the junctions and can be achieved either by adjusting the composition of the ternary alloys in GaAs and InP devices, or by allowing the formation of a strained layer, as in the case of Si HBTs with a SiGe base.

The main motivation for studying and developing HBTs is to capitalize on the advantage to device performance that can result from having an emitter material of a wider bandgap than that of the base (1,2). This advantage applies to both direct current (dc) and high-frequency performance and is documented in the section entitled "Wide-Bandgap Emitter." Another advantageous feature of HBTs is that the use of alloys allows compositional grading in the various regions of the device, thereby providing an opportunity to create "alloy fields" to aid the transport of carriers through the device. This further example of "bandgap engineering" is described in the section entitled "Regional Bandgap Engineering." In the case of III–V HBTs, the various layers are grown sequentially by epitaxy; the section entitled "Structural Versatility and OIEC Compatibility" indicates how this provides some flexibility in device structure, and how it also permits the creation of monolithic optoelectronic integrated circuits.

Table 1. Examples of Material Combinations Frequently Used to Form HBTs With Either a Single Heterojunction (SHBT) or Two Heterojunctions (DHBT)[a]

Type	Emitter	Base	Collector	Substrate
SHBT	AlGaAs	GaAs	GaAs	GaAs
SHBT	GaInP	GaAs	GaAs	
DHBT	AlGaAs	GaAs	AlGaAs	
SHBT	InP	GaInAs	GaInAs	InP
SHBT	AlInAs	GaInAs	GaInAs	
DHBT	AlInAs	GaInAs	InP	
SHBT	α-Si	Si	Si	Si
SHBT	SiC	Si	Si	
DHBT	Si	SiGe	Si	

[a] Generically, the devices are grouped according to the substrate material.

The advantages of HBTs over BJTs come at the price of increased complexity of fabrication. The difficulties are illustrated in the section entitled "Fabrication and Performance of HBTs," where representative processes for fabricating HBTs on GaAs, InP, and Si substrates are outlined. Also in this section, the electrical performance of the various classes of HBTs is compared and summarized, with the aim of providing device data which is relevant to the applications discussed in the section entitled "Applications." The applications incline toward high-speed, low-power-consumption digital circuits and high-frequency, high-output-power analog circuits.

The link between devices and circuits is forged by modeling. The section entitled "Modeling" discusses equivalent-circuit models for HBTs that are applicable to dc, transient, and small-signal situations. Quasiballistic transport is briefly alluded to, because it is likely to become important as the basewidth of HBTs shrinks to dimensions of the order of a mean-free-path length.

ADVANTAGES OF THE HBT OVER THE BJT

Wide-Bandgap Emitter

The difference in bandgap between two materials is accommodated at the junction by discontinuities in the conduction and valence bands, in a proportion that is material- and composition-dependent. In GaAs and InP HBTs with abrupt emitter-base junctions, the band offsets between the wide-bandgap material of the emitter and the narrow-bandgap material of the base lead to a spike, ΔE_C, in the conduction band and a

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 1999 John Wiley & Sons, Inc.

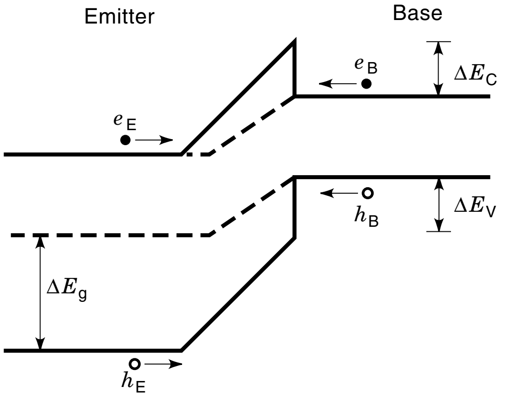


Figure 1. Idealized energy band diagram for an *n–p* heterojunction (solid lines), with the emitter having a wider bandgap than that of the base, and for an *n–p* homojunction (dashed lines). In the heterojunction, the net hole flow is determined by h_E, and the net electron flow is determined by e_B.

step, ΔE_V, in the valence band, as shown in Fig. 1. In this idealized, equilibrium energy band diagram, a homojunction made from the narrow bandgap material is also shown. Regarding the balancing hole flows across the junction at equilibrium, the minority-carrier flow from the emitter, h_E, is unimpeded by the barrier and, consequently, dictates the magnitude of the reverse, majority-carrier flow, h_B. This situation holds for both the homojunction and heterojunction cases. Therefore, perhaps counterintuitively, the valence band offset, ΔE_V, is not, per se, responsible for the reduced back-injection of holes into the emitter of the heterojunction. The reason for the reduction is the much smaller hole concentration in the flow h_E, which is a direct consequence of the wider bandgap in the emitter.

Concerning the equilibrium electron flows, the rate-determining flow is that of the minority carriers from the base, e_B. In the homojunction case this flow is unimpeded, but in the heterojunction case it is diminished by the presence of the band spike. Consequently, the balancing majority-carrier electron flow at equilibrium from the emitter, e_E, is less in the heterojunction case. Thus, relative to a homojunction, the flows, e_E and h_B, are both reduced: the former by $\approx e^{-\Delta E_C/kT}$, and the latter by $\approx e^{-\Delta E_g/kT}$, where ΔE_g is the bandgap difference, k is Boltzmann's constant, and T is temperature. As $\Delta E_C < \Delta E_g$, it is clear that the effect on the holes is more severe. Thus, under the application of forward bias, the back-injection of holes into the emitter is reduced by a greater extent than the forward injection of electrons into the base; this leads to two beneficial effects, namely: (1) a reduction in the minority-carrier charge stored in the emitter under forward bias, and hence a reduction in the emitter-base storage capacitance and, consequently, an improvement in high-speed and high-frequency performance; and (2) an improvement in the electron injection efficiency, which impacts directly and favorably on β, the forward common-emitter current gain of the device.

If the reduction in hole flow is large enough, the emitter doping density, N_E, can be decreased, and the base doping density, N_B, can be increased, while still maintaining β at an acceptably high value. A lightly doped emitter leads to a reduction in the emitter-base junction capacitance, along with an additional improvement in dynamic performance. More importantly, the ability to employ a highly doped base opens up a large number of advantageous possibilities, namely: (1) a reduction in the lateral base resistance, R_B; this improves the power gain at high frequencies, as characterized by a higher unity-power-gain bandwidth, f_{max}; additionally, it reduces emitter-current crowding and also improves the noise performance by reducing the thermal noise in the base; (2) a tolerance of a narrower base in meeting a particular R_B goal; this results in a shorter base transit time and, therefore, an improved high-frequency response, as characterized by a higher unity-current-gain bandwidth, f_T; (3) a reduction in the depletion-region encroachments into the base, leading to higher output conductance and Early voltage, which is often beneficial in analog circuitry, and a reduced susceptibility to device breakdown via punchthrough.

In an HBT, as illustrated in Fig. 1, for example, the barrier to the electron flow from the emitter, e_E, is less than that for the reverse flow of holes from the base, h_B. If this difference in thermal activation energies is ΔE_a, say, then the emitter injection efficiency will depend on $e^{\Delta E_a/kT}$. Thus, the current gain, β, in an HBT decreases with temperature. This has implications for the operation of HBTs at high power densities, as discussed in the subsection entitled "Device Performance."

Regional Bandgap Engineering

In nearly all of the HBTs listed in Table 1, alloy materials are used in some region of the device. By varying the composition of the alloys it is possible to introduce features that enhance performance beyond that of abrupt-junction, regionally uniform devices. Figure 2 shows some possibilities that are discussed in the following subsections.

Emitter Grading. Figure 2(a) illustrates a situation where the wide bandgap material of the bulk emitter is compositionally graded close to the junction with the base. An example would be an $Al_xGa_{1-x}As$ emitter with the mole fraction x varied from 0.3 to 0 over a distance of about 30 nm from a GaAs base. The diagram depicts a linear grading, as would follow from a linear variation in Al mole fraction. Parabolic grading is also often used. In either case, the objective of the emitter grading is to reduce, or eliminate, the conduction-band spike present at abrupt junctions. There are at least two favorable consequences of this. First, because the electron flow across the junction is increased, while the flow of holes is unchanged,

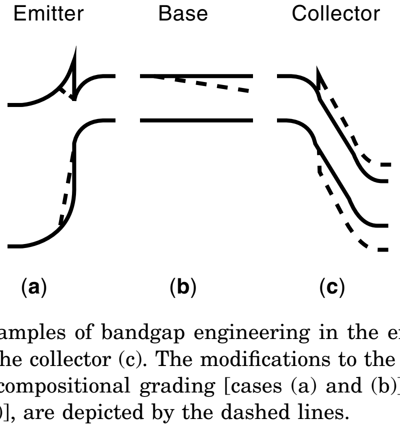


Figure 2. Examples of bandgap engineering in the emitter (a), the base (b), and the collector (c). The modifications to the band diagram due to either compositional grading [cases (a) and (b)] or additional layers [case (c)], are depicted by the dashed lines.

the collector current and the current gain are enhanced. One advantage of this is a lower turn-on voltage—that is, a lower base-emitter bias for a given collector current. This is advantageous for the implementation of HBTs in low-power-consumption circuitry. Second, the ideality factor of the emitter current reduces to unity as thermionic emission takes over from tunneling as the dominant barrier-transport mechanism. This fact, when coupled with the increase in emitter current, leads to a lower dynamic emitter resistance and, when coupled with the increase in collector current, leads to an improvement in transconductance. A possible third benefit of a graded emitter is that electrons are injected into the base at near-equilibrium energies, rather than at the elevated kinetic energies possessed by electrons which have tunneled through a high potential barrier, or have been thermionically emitted over it. While these latter "hot" electrons may make a faster transit of the base, they are more likely in GaAs and InP HBTs to be scattered into the lower-mobility, upper conduction-band valleys on entering the high-field, collector space-charge region. Thus, the overall emitter–collector delay time, τ_{EC}, may in some instances be lower in a graded-junction device (3).

Base Grading. In base-graded n–p–n HBTs the bandgap is progressively reduced from the emitter to the collector by an appropriate variation in the composition of the base material. As shown in Fig. 2(b), the bandgap change is taken up by the conduction band alone. This is because the high hole conductivity precludes any significant variation in the valence band (4). Thus, the grading has the effect of producing an electric field to aid the passage of electrons across the base. The obvious benefit of this is an improvement in the base transit time, τ_B. If, because of the very narrow base that is allowed by having a high base doping density, τ_B is not a major contributor to the overall delay time, τ_{EC}, then base grading can be used to achieve an acceptably low value of τ_B in a wider base. The resulting smaller lateral base resistance enables an improvement in $f_{\max}$.

Base grading can also have a beneficial effect on the current gain, β; the reasons for this are twofold. First, recombination in the bulk quasineutral base, which can be one of the significant contributors to the base current, is diminished, with respect to that in a uniform, narrow-bandgap base, because the alloy field increases the electron velocity and, for a given collector current, this means a reduction in the base charge. Of course, this will reduce the base recombination current only if the minority-carrier lifetime is not overly shortened by using an alloy material for the base. Second, recombination at surface regions of the base is reduced by the tendency of the base alloy field to sweep electrons to the collector—that is, in a direction perpendicular to the base surface. This effect can be particularly beneficial in HBTs with small-dimension emitters, because these devices necessarily have a large emitter perimeter/area ratio.

However, there is a limit to the base grading, beyond which β starts to decrease. The situation is illustrated in the idealized energy band diagram of Fig. 3. Increasing the bandgap of the base material at the emitter-base junction serves to increase the built-in junction potential, V_{bi}, to reduce the band offsets, and to make the junction become more and more like a homojunction made from wide-bandgap material. The result is that the net electron flow into the base is reduced, whereas the net hole flow into the emitter is unchanged. These effects conspire to reduce the current gain.

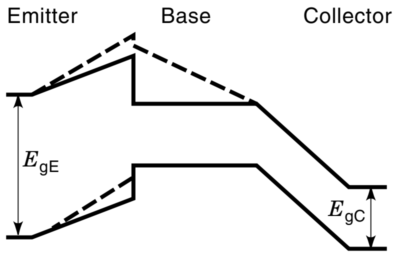


Figure 3. Idealized energy band diagram illustrating the effect of grading the base (dashed lines) in a SHBT with fixed bandgaps in the emitter (E_{gE}) and collector (E_{gC}).

In a DHBT, a further limitation to the permissible amount of base grading is imposed by the necessity to trade-off some of the base grading against the height of the barrier to electron flow into the collector. The situation is illustrated in the idealized energy band diagram of Fig. 4. For a given bandgap in the base at the emitter end of the device, decreasing the base bandgap at the collector end increases the amount of base grading, but also increases the barrier to electron flow into the collector. Both β and τ_B are affected by this tradeoff. Calculations indicate that for a DHBT that has an emitter and collector made from $Al_{0.3}Ga_{0.7}As$, along with a linearly graded $Al_xGa_{1-x}As$ base with $x = x_{be} = 0.15$ at the emitter end of the base, τ_B is minimized with $x = x_{bc} \approx 0.1$ at the collector end of the base (5).

Collector Engineering. There are two main motivations for modifying the collector region of a SHBT, namely: (1) to reduce the signal-delay time, τ_{CSCR}, of electrons in transit across the collector-base space-charge region and (2) to increase the breakdown voltages, BV_{CEO} and BV_{CBO}.

Historically, in bipolar transistors the major transit-time component of τ_{EC} has been τ_B. In modern HBTs, the ability to realize very narrow bases and, to a lesser extent, to implement compositionally graded bases, has led to significant reductions in τ_B, with the result that the dominant transit-related time is τ_{CSCR}. Clearly, reducing the width of the base-collector space-charge region, W_{CSCR}, would shorten the collector signal-delay time, but, if this were accomplished by increasing the collector doping density, then the breakdown voltage would be adversely affected. Trading-off speed for improvement in breakdown performance, and vice versa, pres-

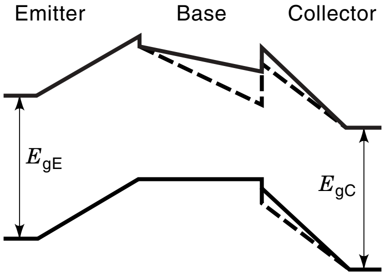


Figure 4. Idealized energy band diagram illustrating the effect of increasing the base grading in a DHBT by decreasing (dashed lines) the bandgap of the base material at the collector end of the base.

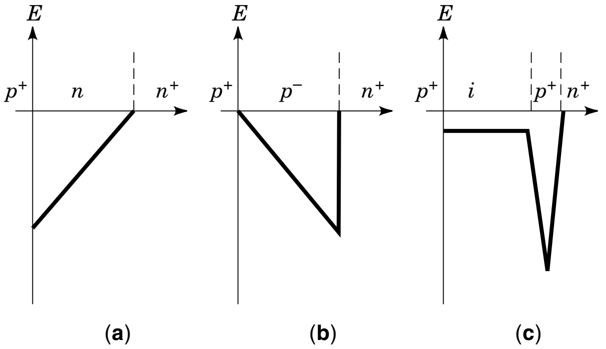


Figure 5. Electric-field profiles in different collector structures. The p^+-base is on the left and the collectors are classed as (a) conventional, (b) inverted, and (c) intrinsic.

ents opportunities for novel collector designs (6). Two structures intended to reduce τ_{CSCR} in GaAs devices are shown in Fig. 5. In the "inverted-field" structure of Fig. 5(b) (7), the field pattern of a conventional collector [Fig. 5(a)] is reversed, with the intention of delaying the onset of electron transfer to the lower-mobility, upper-conduction bands. This transfer occurs more readily in a conventional collector, as electrons immediately enter a high-field region, in which they are accelerated and rendered more likely to scatter into the upper valleys. This phenomenon is more likely to occur in GaAs than in InP because of the smaller separation between the Γ and L valleys (3). In the "intrinsic" structure shown in Fig. 5(c), the high-field zone is restricted to a narrow p^+–n^+ region, and the field remains favorably low in the remaining, weakly doped part of the collector. The first GaAs HBT to register an f_{T} in excess of 100 GHz was fabricated using such a collector structure (7).

Turning now to breakdown voltage considerations, the attainment of acceptably high values is difficult in HBTs employing collectors made from narrow-bandgap material, for which the ionization coefficients for electrons and holes are generally high. This is the situation in InP HBTs that have collectors of lattice-matched $\mathrm{Ga_{0.47}In_{0.53}As}$, which has a bandgap of only 0.75 eV. Solutions to this problem involve making all, or part, of the bulk of the collector from a wider bandgap material, such as InP, for which the impact-ionization coefficients are low. When all of the collector is InP, the device is a DHBT. When the collector comprises GaInAs near to the base and then InP for the remainder of the layer, as illustrated in Fig. 2(c), the device is labeled as a composite-collector HBT. This design is a refinement of the DHBT, with the objective being to use just enough n–GaInAs in the collector to ensure that the conduction-band spike at the base–collector junction is reduced below the level of the conduction band in the base. In this way, the stored base charge, which strongly influences τ_{B} and β, and the collector current are not adversely affected by the presence of the conduction-band spike.

Structural Versatility and OEIC Compatibility

HBTs, at least those of the III–V material variety, are invariably fabricated from a stack of epitaxial layers grown on a semi-insulating substrate. The latter feature has the beneficial effect of eliminating the parasitic collector-substrate capacitance, which is present in most BJT structures. It also provides an "inert" platform, which may allow fabrication with equal ease of emitter-up and collector-up structures. Collector-up structures are of interest because the smaller dimension of the collector leads to a reduction in collector-base junction capacitance (8). Their simultaneous employment with emitter-up devices in ICs could also provide some flexibility in circuit design, particularly regarding interconnectibility of components. However, for both devices to be formed from a single stack of epitaxial layers, each device would necessarily be a DHBT, and a truly reversible device, akin to the MOSFET, would only be realizable if equal doping densities for emitter and collector could be tolerated. A more realistic use of collector-up devices might be as the sole transistor type in exceptionally high-speed ICs (9).

The fact that many semiconductor lasers and photodiodes are formed epitaxially from layers of materials similar to those used in HBTs leads naturally to the concept of monolithic optoelectronic integrated circuits (OEIC). Inevitably, there are difficulties associated with the simultaneous realization of high-quality optical and electronic devices from a given stack of epitaxial layers. However, in optoelectronic integrated-receiver front ends, for example, good results have been obtained when using the base–collector of an InP HBT as the photodiode (10,11). An example of this embodiment is shown in Fig. 6; the top-illuminated photodiode is constructed from the HBT layers by removing the emitter layers and adding an antireflection coating.

FABRICATION AND PERFORMANCE OF HBTS

HBTs using GaAs or InP substrates, or SiGe base layers, have all progressed beyond the experimental-device stage to the point that they are being included in, or considered for, commercial circuits. Examples of the fabrication procedures of HBTs representative of these three material systems are discussed in this section. Also presented are some device-performance metrics that are relevant to the use of HBT circuits in the high-end applications considered in the section entitled "Applications."

GaAs HBTs

An example of a manufacturable process for AlGaAs/GaAs HBTs is illustrated in Fig. 7 (12). The epitaxial layers comprising the device can be grown by MBE, MOCVD, or CBE,

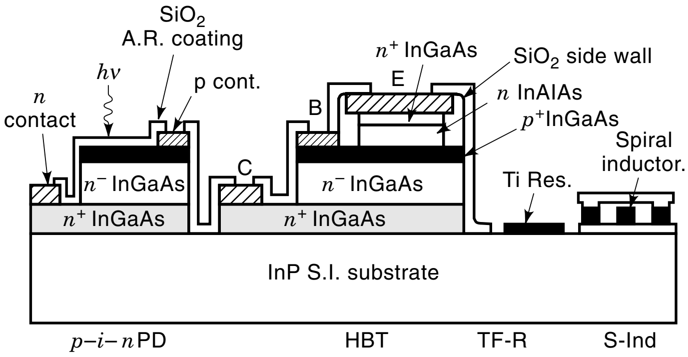
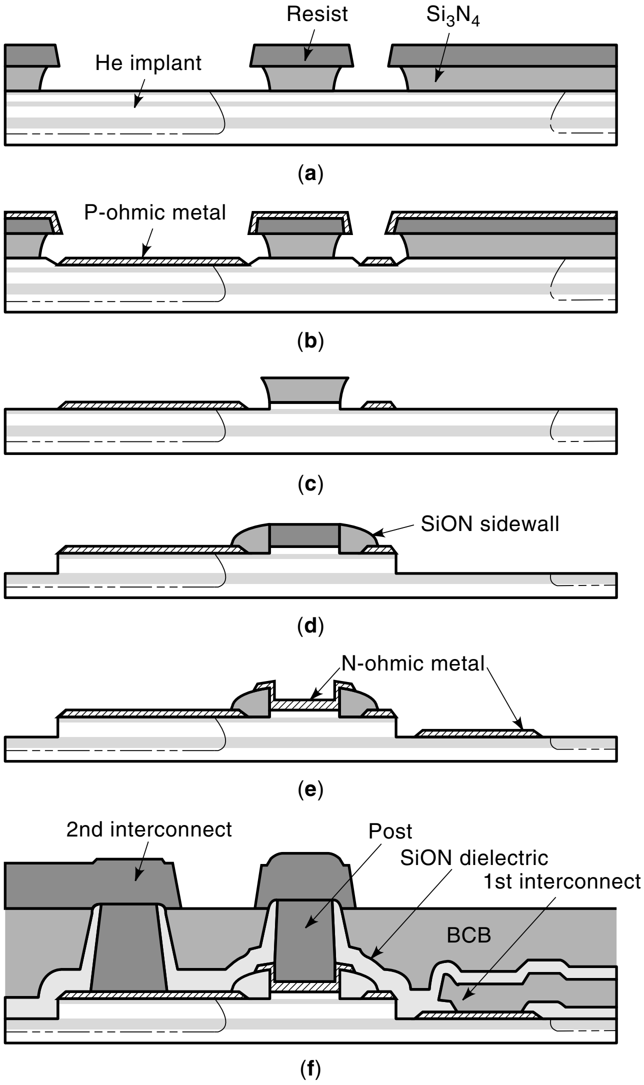


Figure 6. Schematic cross section of the elements of a p–i–n/HBT OEIC photoreceiver (11).



He isolation implant and dummy emitter formation.

Selective dry etch to AlGaAs emitter and P-ohmic metal deposition.

Field Si_3N_4 removal and second etch to AlGaAs emitter.

Thick sidewall deposition and RIE etch to sub collector.

Selective removal of dummy emitter and N-ohmic metal deposition.

Back end processing:
- SiON deposition and via etch
- NiCr resistor deposition
- ME1 deposition.
- ME1 to ME2 post deposition
- BCB planarization and ME2 deposition

Figure 7. Fabrication sequence for an AlGaAs/GaAs/GaAs HBT (12). In the finished device, the first interconnect is to the subcollector, the second interconnect is to the base, and the emitter is under the post metallization.

and are shown schematically on the figure by the alternating dark and light bands. From the top, the layers are: the heavily doped cap to facilitate ohmic contacting to the emitter metal, the emitter, the base, the collector and the sub-collector.

A dummy dielectric emitter (Si_3N_4) is employed for self-alignment to the base metal, which completely surrounds the emitter in this case [Figs. 7(a) and 7(b)]. A deep He^+ ion implantation serves to deactivate the collector layer under most of the base metal area, thereby eliminating most of the extrinsic base–collector capacitance. Obtaining a good ohmic contact to the thin, *p*-type base, without allowing metal penetration through to the collector, is a crucial step. The procedure here is to first etch-off the emitter cap and then deposit Pd–Zn–Pt–Au–Pd, which is then alloyed through the AlGaAs to the p^+ base. The penetration depth during alloying is controlled by the thickness of the Pd layer below the Pt barrier. The next step [Fig. 7(c)] is to etch-off the cap layer between the edge of the base metal and the emitter stack. This defines the emitter dimensions and, very importantly, leaves an AlGaAs "shelf" layer on top of the base, thereby reducing the surface recombination velocity in this peripheral region. This helps maintain a high β in small-dimension devices. The peripheral region is protected from the subsequent etch, which exposes the subcollector [Fig. 7(d)], by a SiON sidewall spacer, which also provides alignment tolerance for the emitter metallization. The fabrication sequence is terminated by providing two levels of metal interconnects and a planarizing layer of benzocyclobutene (BCB).

In the HBTs described above from NORTEL, Canada, the base dopant is carbon. This element is also used by other HBT manufacturers, e.g., Rockwell and Northrup Grumman in the United States, principally because of the reproducibility and reliability that are usually attributed to its minimal migration during subsequent processing steps. Another base-doping element, Be, is often considered to be inferior in this regard (13). However, the California company TRW, which appears to be the world's largest supplier of HBTs, currently uses Be at concentrations up to 2×10^{19} cm^{-3}, and it reports excellent reliability—for example, a median time to failure of $>10^8$ h for discrete devices (14) and of $>10^7$ h for an HBT MMIC (X-band logarithmic amplifier) (15). The trick to successful Be doping appears to be the encouragement of substitutional doping on Ga sites, by growing in a sufficient As overpressure to ensure that there are relatively few As vacancies (16,17).

A significant event in GaAs HBT development has been the replacement of the AlGaAs emitter by GaInP, at least in laboratory-prototype devices. Oxygen complexes, which are often incorporated in Al-containing emitters, are less prevalent when using GaInP. This reduces the emitter–base–junction recombination current and allows current gains of $\beta > 1$ to be maintained down to very low current densities (18). In fact, near-ideal base and collector I–V characteristics over about six decades of collector current have been reported (19). Given the importance attached to surface recombination elsewhere in this article, it is remarkable that this result was obtained with an unpassivated GaAs surface. The suggestion in Ref. 19 is that when junction recombination is reduced by using a GaInP emitter, the short minority carrier lifetime in the heavily doped base (4×10^{19} cm^{-3}) causes quasi-neutral-base recombination to dominate over surface recombination, resulting in the observed near-ideal base characteristic.

InP HBTs

A typical process sequence for an InP HBT is illustrated schematically in Fig. 8 (20). Three mesa etches are required: one to define the emitter, one to define the base and collector regions, and one to etch-down to the semi-insulating substrate in order to provide device isolation. This latter feature gives the device a distinctively different look from the GaAs HBT of Fig. 7, and it is necessary because the InP-material system lacks an ion-implantation damage process capable of rendering GaInAs sufficiently resistive to act as an isolator. To reduce the parasitic base–collector capacitance in a mesa structure the base–collector mesa must be made as small as possible, necessitating tight self-alignment of the base and emitter metallizations. This is achieved by a slight undercutting of the emitter metal contact, which allows this metal to serve as a mask for the medial edge of the subsequently deposited base metal. The etch used in the undercutting must also be selective in order not to destroy the thin base material; for GaInAs bases this can present a problem (20). If the collector is InP, a selective etch can also be used to define the base–collector mesa. If the collector is GaInAs, the same material as the subcollector, then this convenient etch-stop method cannot be employed. However, the etching is less critical than in the case of the emitter-mesa formation because the subcollector is relatively thick. The device structure resulting from a triple-mesa process is highly nonplanar, but can be planarized, to facilitate device interconnection and terminal access, by using a suitable polyimide. This coating also provides some passivation of the exposed GaInAs surfaces.

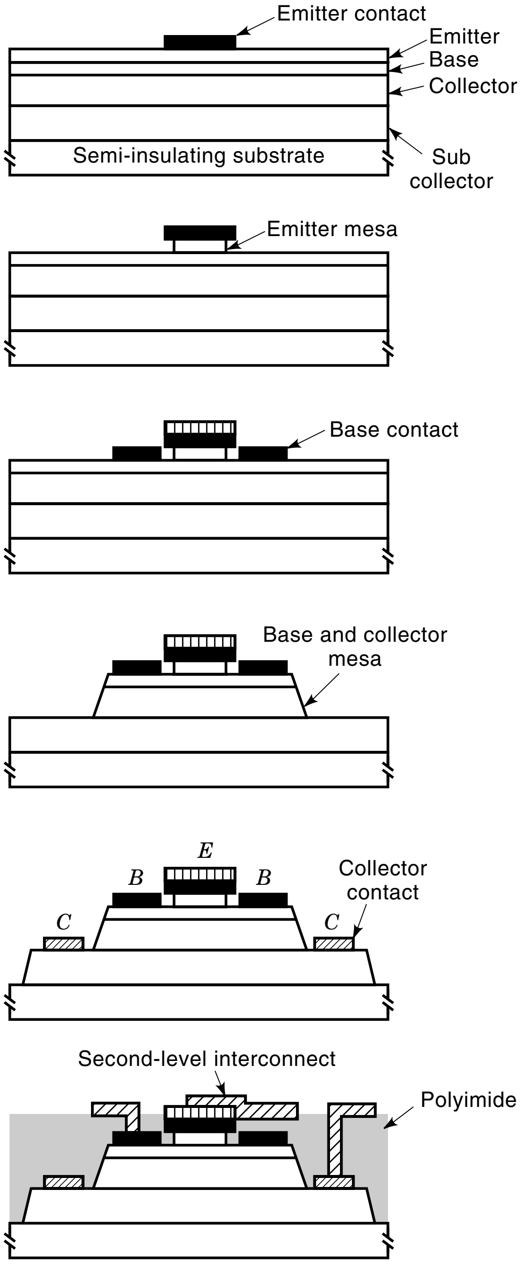


Figure 8. Fabrication sequence for an AlInAs/GaInAs/InP or GaInAs HBT (20).

Si HBTs

Wide-bandgap-emitter Si HBTs have been made by using a Si base and either amorphous Si or SiC emitters, as well as by using a SiGe base with an Si emitter. The latter configuration is the most developed, and shows the most promise for near-term applications; for these reasons it is discussed here.

Unlike the junctions in the III–V HBTs discussed in the preceding subsections, the SiGe/Si junction is not near-perfectly lattice-matched. However, if the epitaxial film is not too thick and the growth temperature is not too high, the SiGe base layer will conform to the Si collector material on which it is grown. Such a layer is referred to as being pseudomorphic and commensurately strained. The strain in the SiGe film breaks the six-fold degeneracy of the conduction band minima and the two-fold degeneracy of the valence band maxima; these phenomena can be exploited to improve the electron mobility (reducing τ_B) and the hole mobility (reducing R_B) (4). Furthermore the strain decreases the SiGe bandgap, fortuitously improving the bandgap differential between the base and the subsequently deposited Si emitter. The critical thickness—beyond which the SiGe film becomes unstable, and sufficiently defective for the associated rise in recombination centers to render it useless for bipolar applications—is inversely proportional to the Ge content or, for compositionally graded films, to the integrated Ge content. Bearing in mind the necessity of restricting the basewidth to less than 100 nm in order to maintain an acceptable τ_B, the average Ge content should not be greater than about 15% (21). Maintenance of the pseudomorphic nature of the film, by not allowing it to relax and assume its bulk lattice constant, is dependent on minimizing the exposure of the film to subsequent high-temperature environments.

The constraints imposed by the considerations of SiGe-film stability lead to a Si/SiGe HBT fabrication process which is more complicated than those described previously for III–V HBTs. A review of the procedures used to date to achieve high-performance Si/SiGe HBTs can be found in Ref. 22, along with a detailed description of the technology that is employed at IBM, USA to produce HBTs in a manner compatible with standard Si-CMOS processing. Here we describe a technology from NEC, Japan which is geared toward high-performance bipolar circuitry and therefore has the same goal as that of the processes described earlier for III–V HBTs. A partial process sequence is illustrated in Fig. 9 (23,24). To prepare for the SiGe growth and to begin the emitter definition, the top dielectric film (Si_3N_4) and the large-grain p^+ polysilicon layer shown in Fig. 9(a) are etched and the edge of the resulting feature is covered with a Si_3N_4 sidewall. The bottom dielectric (SiO_2) is then etched and allowed to laterally undercut the sidewall. From this point on, attention must be paid to the thermal budget, so as not to destroy the pseudomorphic nature of the SiGe base. This layer is grown on the exposed Si substrate, simultaneously with a polySi film which descends from the underside of the polySi layer exposed by the earlier lateral etch. Growth is stopped when the two growing films touch [see Fig. 9(b)]. Growth is carried out by CVD in a cold-wall, ultrahigh-vacuum (UHV) system at about 650 °C. Because films grown by this selective-epitaxial-growth (SEG) process are very sensitive to the condition of the Si surface, the collector is formed subsequently, using ion implantation. The remaining fabrication steps comprise (23,24) (1) coating of the Si_3N_4 sidewall with borosilicate glass (BSG), (2) doping of the external SiGe base by driving-in boron from the glass at 800 °C for 10 min, (3) deposition of phosphorus-doped polysilicon for the emitter, and (4) drive-in of the emitter at 950 °C for 10 s. The finished device, complete with BSG-trench isolation, is shown in Fig. 9.

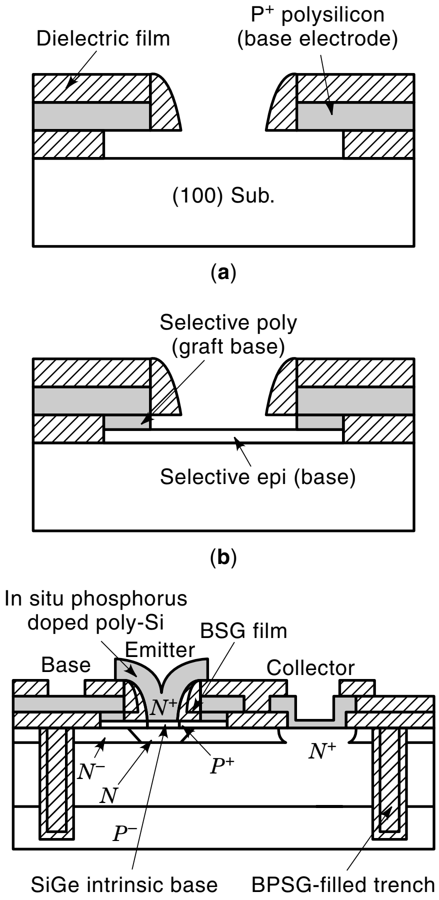


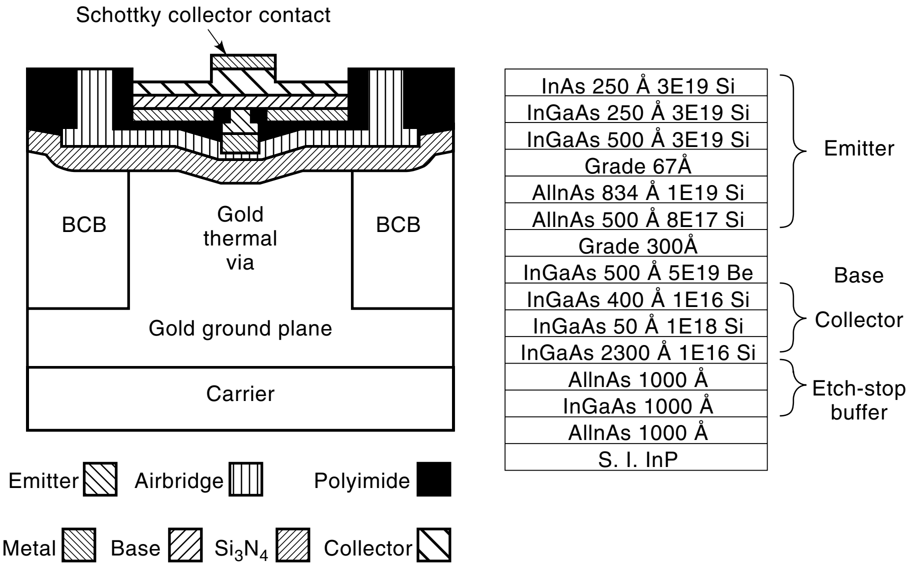
Figure 9. Partial fabrication process for a Si/SiGe/Si HBT [(a) and (b)], and schematic cross-section of the finished device (24).

The SiGe base layers in the HBTs from IBM are also grown using a UHV/CVD system, but in this case it is of the hot-wall variety (25). This equipment has been developed specifically for blanket Si and SiGe epitaxy and is consistent with a low-temperature, commercially feasible process. In this process, hydrogen-passivated Si wafers are admitted to the system, in which a vacuum of around 10^{-9} torr is maintained. The residual gas is predominantly hydrogen; other species, which may be chemically active with silicon, are not present at sufficient partial pressures to violate the hydrogen passivation of the wafer. Films are subsequently deposited under vacuum by CVD at temperatures in the range 400–500 °C. Precise dimensional control, of the order of 1–2 atomic layers, is possible and satisfies the need of the UHV/CVD process to be competitive with ion implantation, which is the benchmark for doping control in conventional Si transistor technology (22).

Device Performance

The purpose of this subsection is to briefly mention some of the high-performance HBTs that have been reported in the literature and to relate the attained performance to (a) the information on device structure and fabrication presented earlier and (b) the properties of the materials constituting the device. While an HBT with one particularly outstanding figure of merit may not be a practical device because of its poor performance as judged by another metric, the numbers presented below should give some idea of the relative strengths of GaAs, InP, and Si HBTs and should help in understanding why HBTs are well-suited to the applications discussed in the section entitled "Applications."

It is well known that GaInAs, as used for the base material in InP HBTs, has many desirable properties (4). For example, with respect to GaAs and Si, GaInAs possesses higher electron mobility, higher electron peak velocity, higher electron saturation velocity, and, compared to GaAs, a higher separation of the Γ and L conduction band minima. These features are conducive to high-frequency and high-speed performance and have led to GaInAs-base HBTs with exceptional values of f_T [165 GHz (26)], f_{max} [277 GHz (9)] and τ_{pd} [12 ps (27)], where τ_{pd} is the propagation delay measured for a current-mode-logic (CML) gate. The high-f_T device used an abrupt InP/GaInAs emitter–base junction to launch high-energy electrons into the base, and it used a narrow collector with a low V_{CB} bias to reduce scattering into the upper conduction band valleys. The low-τ_{pd} device used a graded AlInAs emitter. The high-f_{max} device is an interesting version of the collector-up structure and was achieved by removing the grown epi-



(a) (b)

Figure 10. Schematic cross section (a) and layer structure (b) of a high-performance collector-up HBT (9).

taxial-layer stack from its original InP substrate [see Fig. 10(b)] and then mounting the inverted stack on a gold ground plane and GaAs carrier substrate [see Fig. 10(a)].

The high-speed and high-frequency performance of Si HBTs is also impressive and probably owes much to the fact that device designers are able to capitalize on the unrivaled maturity of silicon technology in general. A useful design variable in Si HBTs is the Ge profile in the base. A triangular profile, increasing toward the collector, provides an aiding field for electron transport and has been employed to achieve an impressive f_T of 113 GHz (28) and an ECL-τ_{pd} of 20 ps (22). Increasing the Ge content in the base close to the collector also reduces the depletion-region excursion into the base (via the permittivity difference), thereby helping increase the early voltage, V_A. Contrarily, a high Ge content near to the emitter leads to an increased equilibrium, minority-carrier electron concentration at the emitter–base junction, and hence a higher β. Thus, it would be expected that some compromise profile, such as a trapezoid, might be used to obtain a high βV_A product, which is desirable for many analog applications. However, the best results reported to date for βV_A are from HBTs with a triangular profile (22).

Another beneficial property of GaInAs is its low bandgap, which results in a higher minority-carrier electron concentration for a given base doping density. When combined with a graded-emitter junction, a significantly lower emitter-base turn-on voltage, V_{TO}, can be achieved. For example, V_{TO} values have been reported that are 780 mV lower than in graded AlGaAs/GaAs HBTs and that are 220 mV lower than in Si BJTs (27). A low V_{TO} permits lower supply voltages to be used, resulting in less power consumption and a more favorable delay-power product (see section entitled "Digital Circuitry").

GaInAs also has a very low surface recombination velocity that permits achievement of near-ideal I–V characteristics over a large range of forward bias (27). As mentioned in the section entitled "GaAs HBTs," GaAs-base HBTs can match this I–V performance if GaInP emitters are employed. No matter how it is achieved, minimization of the recombination velocity at the base surface is also desirable to reduce $1/f$ noise. This is important in broadband amplifiers and in oscillators employing up-conversion to the oscillation frequency (29).

Of the three substrate materials considered in this article, GaAs has the poorest thermal conductivity. This makes GaAs HBTs prone to transient local heating when the collector current is switched to high densities. When taken in conjunction with the negative temperature coefficient of β, which is a feature of AlGaAs/GaAs HBTs (30), this can lead to a negative differential output conductance (31), and even to a drastic collapse of β in unballasted, multifinger devices (32). However, under stable conditions, GaAs HBTs are well-suited to high-temperature operation on account of the suppression of thermal generation of carriers by virtue of the relatively large bandgap of this material. A stable gain of $\beta = 35$ has been measured at 350°C in an AlGaAs/GaAs HBT with a graded emitter to improve electron injection, and with an emitter mole fraction of 0.45 to maximize the direct bandgap in the emitter, and hence minimize the net hole flow from the base (33). Hole back-injection has been identified as one of the principal mechanisms for gain reduction in AlGaAs/GaAs HBTs operating at elevated temperatures (30). In this regard, the slightly higher bandgap of GaInP can only help in maintaining a high value of β and may be a contributory factor to the observed temperature independence of the gain at moderate current densities over the range 25 to 300°C (18).

The large bandgap of GaAs is also helpful in the attainment of acceptable breakdown voltages. For example, values of BV_{CEO} in the range of 9 to 15 V are typical of Rockwell's HBTs in which the collectors have a doping density of 3×10^{16} cm^{-3} and thicknesses in the range 300 to 700 nm (34). Comparable breakdown voltages in InP HBTs can be obtained with composite-collector structures (35) and with DHBTs (36). For really high breakdown voltages, such as might be needed in very high output-power circuits, the GaInP/AlGaAs/GaInP structure, with a graded base to eliminate the conduction-band discontinuity at the collector, may be of interest. Prototype devices have yielded $BV_{CEO} = 45$ V (37).

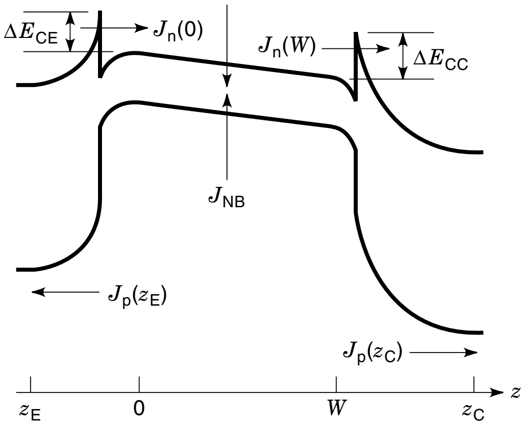


Figure 11. Energy band diagram of a general, base-graded DHBT, in the normal, active mode of operation, showing the current components of the intrinsic device.

MODELING

Modeling (as presented in this section) serves mainly to link the physics and structure of HBTs (as described in the sections entitled "Advantages of the HBT over the BJT" and "Fabrication and Performance of HBTs") to the applications of HBT circuits (as discussed in the section entitled "Applications"). To this end, emphasis is placed on an analytical model of the HBT and its relationship to electrical equivalent circuits for HBTs. The circuits are intended to cover the important operating modes of HBTs, namely: dc, large-signal transient, and high-frequency, small-signal ac. The models assume that transport in the bulk regions of the device is by drift and diffusion. Although this is valid for most present-day devices, it may not be so for future devices with very short bulk regions. To briefly address modeling in this situation, the topic of quasiballistic transport is introduced at the end of this section.

Dc Model

The intrinsic current components in a general HBT are illustrated in Fig. 11. The electron transport current at an edge of the quasineutral base, $J_n(0)$ for example, is most conveniently formulated by balancing the interfacial currents (tunneling and thermionic emission) at the emitter-base junction, with the drift-diffusion current in the base (38). The latter current depends also on the balancing of currents at the base–collector junction, so $J_n(0)$ and $J_n(W)$ are linked. This interaction between the junctions is perhaps most elegantly captured by the Ebers–Moll relations, that is,

$$\begin{aligned} J_E &= -a_{11}(e^{V_{BE}/V_t} - 1) + a_{12}(e^{V_{BC}/V_t} - 1) - J_p(z_E) \\ J_C &= a_{21}(e^{V_{BE}/V_t} - 1) - a_{22}(e^{V_{BC}/V_t} - 1) - J_p(z_C) \end{aligned} \tag{1}$$

where V_t is the thermal voltage and the J_p terms are the hole currents shown in Fig. 11. The Ebers–Moll coefficients, a_{ij}, are used to characterize solely the electronic components of the currents (39), rather than, as is the case for BJTs, both the electron and hole currents. This separation of the carrier currents is convenient for HBTs, in which tunneling is important because the electron flows, $J_n(0)$ and $J_n(W)$, do not follow the same ideal Boltzmann dependence (e^{V/V_t}) as the hole currents. The Ebers–Moll coefficients for the general case of a DHBT with graded junctions and a graded base are listed in Ref. 40, and those for the case of an abrupt junction, graded-base SHBT, assuming Shockley boundary conditions at the base–collector junction, are listed in Ref. 41. The bias dependence of the Ebers–Moll coefficients comes in through the barrier heights ΔE_{CE} and ΔE_{CC} (see Fig. 11) and the tunneling factors γ_E and γ_C, which express the ratio of the thermionic-emission current to the total interfacial current (38). It is convenient to compute the tunneling current by using the WKB approximation for the barrier transparency, with the shape of the barrier being dependent on the amount of junction grading present (38).

The equations presented above form a physics-based compact model, making them useful for device analysis and design, and also for an equivalent-circuit representation that can be used for circuit simulation. Here we focus on incorporation of the equations into SPICE, which is probably the most widely used simulator for electronic circuits.

Because of the additional bias dependencies, the equivalent circuit model for the intrinsic HBT differs from that of the BJT in that some of the diodes must be replaced by current sources (see Fig. 12). In this circuit the diodes IPE and IPC represent the hole currents in Eq. (1). The current sources INE and INC account for neutral-base recombination and are given by

$$\text{INE} = A_E(a_{11} - a_{21})(e^{V_{BE}/V_t} - 1) \tag{2}$$

$$\text{INC} = A_E(a_{22} - a_{12})(e^{V_{BC}/V_t} - 1) \tag{3}$$

where A_E is the emitter area. The remaining current source represents the transport current:

$$\text{ICT} = A_E a_{12}(e^{V_{BE}/V_t} - e^{V_{BC}/V_t}) \tag{4}$$

In writing the expression for ICT, use has been made of the fact that $a_{12} = a_{21}$, which holds true if the compositional dependencies of the effective densities of states in the graded base can be ignored. It is easy to show that Fig. 12 is equivalent to Eq. (1). To turn the intrinsic model of Fig. 12 into an extrinsic one, extra elements need to be added to represent the junction recombination-generation currents and, if necessary, recombination currents at insufficiently passivated surfaces. Regarding the interior recombination-generation currents, the mechanisms of importance in III–V materials are indirect nonradiative (SRH), radiative, and Auger. All of these can be represented in the equivalent circuit by additional diodes (42), as illustrated in Fig. 13.

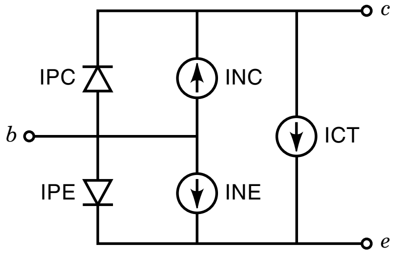


Figure 12. Direct-current electrical equivalent circuit for the intrinsic HBT.

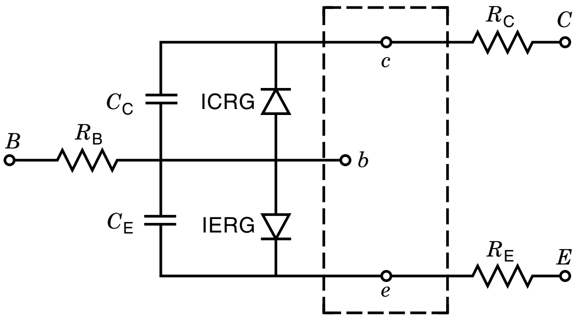


Figure 13. Large-signal electrical equivalent circuit for the HBT. The dashed box is the intrinsic circuit of Fig. 12. The diodes represent recombination-generation currents at the respective junctions.

If the recombination current in the base is small compared to the collector current, and the gradient of the hole quasi-Fermi level in the base is zero [which is not unreasonable given the high base doping density (4)], then the collector current can be written as (43)

$$\mathrm{ICT} = A_{\mathrm{E}} \frac{-q}{N_{\mathrm{B}}} \left[\frac{e^{V_{\mathrm{BE}}/V_{\mathrm{t}}} - e^{V_{\mathrm{BC}}/V_{\mathrm{t}}}}{\Theta_{\mathrm{B}} + \Theta_{\mathrm{EB}} + \Theta_{\mathrm{BC}}} \right] \tag{5}$$

where

$$\Theta_{\mathrm{B}} = \int_0^W \frac{dz}{D_{\mathrm{B}} n_i^2(z)}$$
$$\Theta_{\mathrm{EB}} = 1/(n_i^2(0) v_{\mathrm{E}} e^{-\Delta E_{\mathrm{E}}/kT})$$
$$\Theta_{\mathrm{BC}} = 1/(n_i^2(W) v_{\mathrm{C}} e^{-\Delta E_{\mathrm{C}}/kT})$$

where $n_i(z)$ is the spatially dependent intrinsic carrier concentration in the base, D_{B} is the electron diffusivity, and the mean velocities v refer to electrons in the base, either at the emitter end, v_{E}, or at the collector end, v_{C}.

The three terms in the denominator of this equation can be viewed as being related to the transport of electrons across the three important regions of the device, namely, the emitter-base junction (Θ_{EB}), the quasineutral base (Θ_{B}), and the base-collector junction (Θ_{BC}). This interpretation is convenient for assessing which part of the device is the "bottleneck" for carrier transport (44). In the usual treatments of homojunctions, it is implicitly assumed that transport across the junctions occurs infinitely quickly; that is, $\Theta_{\mathrm{EB}} = \Theta_{\mathrm{BC}} = 0$. This leads to base-dominated transport—that is, drift and diffusion. However, in an abrupt-junction AlGaAs/GaAs HBT, transport across the emitter-base junction can be the rate-determining process, with Θ_{EB} being several orders of magnitude higher than both Θ_{B} and Θ_{BC} (43). Under these circumstances, the expression for the interfacial transport current can be considerably simplified and even written in a diode-like form, that is, (45),

$$\mathrm{ICT} = A_{\mathrm{E}} J_{0\,\mathrm{tunn}} e^{V_{\mathrm{BE}}/n_1 V_{\mathrm{t}}} + A_{\mathrm{E}} J_{0\,\mathrm{therm}} e^{V_{\mathrm{BE}}/V_{\mathrm{t}}} \tag{6}$$

where the J_0 terms are voltage-independent saturation currents. In this formulation the nonideal-Boltzmann-like voltage dependency of the tunnel current is neatly accommodated in the diode ideality factor, n_1.

Large-Signal Model

To turn the dc equivalent-circuit model into a large-signal model, such as would be useful for evaluation of the switching performance of HBTs, it is straightforward to add the parasitic resistances and junction capacitances (see Fig. 13). The resistance and capacitance values are geometry-dependent. Expressions for R_{E}, R_{B}, and R_{C} for a typical pyramidal structure can be found in Ref. 41. The junction capacitance contributions to C_{E} and C_{C} follow from basic electrostatic considerations (46). For operation in the normal active mode, the storage capacitance contribution to C_{C} demands a value for the SPICE parameter TF, the forward transit time. Because the base transit time, τ_{B}, and the base–collector depletion-region signal-delay time, τ_{CSCR}, are both important in modern HBTs, a useful general expression for TF is

$$\mathrm{TF} = \frac{L_{\mathrm{B}}^2}{D_{\mathrm{B}}} \left[\frac{t\cosh(tW) + \frac{g}{2}\sinh(tW)}{te^{(g/2)W}} - 1 \right] + \frac{\sinh(tW)}{te^{(g/2)W} S_{\mathrm{C}}} + \frac{\lambda W_{\mathrm{CSCR}}}{v_{\mathrm{C}}} \tag{7}$$

where the first two terms constitute the contributions to τ_{B} (40) and the third term is τ_{CSCR}, with the factor λ having a value of 0.5 for the uniform velocity case, or 0.4 if some concession to velocity overshoot needs to be made (47); often one equates v_{C} to v_{sat}, the electron saturation velocity. In the expression for τ_{B}, which reduces to the well-known $W^2/2D_{\mathrm{B}} + W/v_{\mathrm{sat}}$ for a uniform-base SHBT, L_{B} is the diffusion length in the base, $g = \Delta E_{\mathrm{gb}}/kTW$, ΔE_{gb} is the bandgap change in the base due to compositional grading, $t = \sqrt{g^2 L_{\mathrm{B}}^2 + 4}/2L_{\mathrm{B}}$ and $S_{\mathrm{C}} = v_{\mathrm{C}} \gamma_{\mathrm{C}} e^{-\Delta E_{\mathrm{C}}/V_{\mathrm{t}}}$.

High-Frequency, Small-Signal Model

In the hybrid-π equivalent circuit used to model HBTs at high frequencies (48), the conventional, common-base current gain, α_0, is modified to account for additional phase shifts associated with the base transit time, τ_{B}, and the base–collector signal-delay time, τ_{CSCR}, that is,

$$\alpha' = \alpha_0 \left[\frac{\sin(\omega\tau_{\mathrm{CSCR}})}{\omega\tau_{\mathrm{CSCR}}} \right] \exp\{-j\omega[(1-m)\tau_{\mathrm{B}} + \tau_{\mathrm{CSCR}}]\} \tag{8}$$

where ω is the radian frequency and m is an empirical fitting factor.

Considerable simplification of the small-signal analysis can be achieved by using the single-zero approximation for α':

$$\alpha' \approx \alpha_0 \{1 - j\omega[(1-m)\tau_{\mathrm{B}} + \tau_{\mathrm{CSCR}}]\} \tag{9}$$

which will agree with the more involved form in Eq. (8) when

$$\omega < \frac{1}{3}[(1-m)\tau_{\mathrm{B}} + \tau_{\mathrm{CSCR}}] \tag{10}$$

With this simplification, and splitting-up both the base resistance, r_{bb}, and the collector–base junction capacitance, C_{bc}, into n parts, the hybrid-π circuit is as shown in Fig. 14. In this circuit, $r_{\mathrm{bb}} = \sum_{k=1}^{n} r_k$ and $C_{\mathrm{jc}} = \sum_{k=1}^{n} C_k$, the parasitic emitter and collector resistances, r_{ee} and r_{cc}, are modeled as

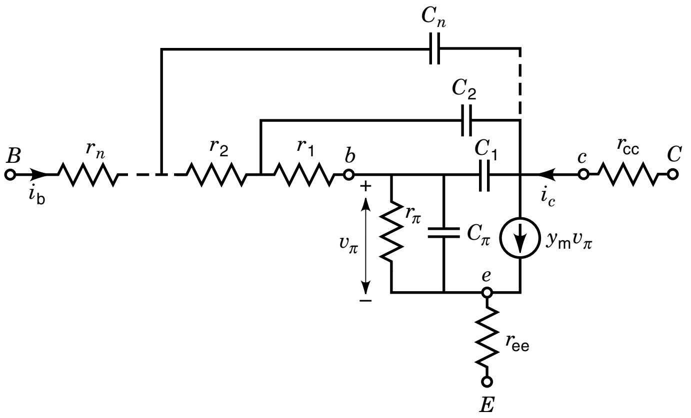


Figure 14. A simple, general-form, small-signal equivalent circuit for the HBT. The definitions of the elements are given in the text (52).

lumped elements, C_π is the sum of the forward storage capacitance and the base-emitter junction capacitance, $r_\pi = r_e/(1 - \alpha_0)$, where r_e is the dynamic emitter resistance, and the transadmittance is expressed as $y_m = g_m - j\omega C_{dc}$, where g_m is the transconductance and $C_{dc} = g_m[(1 - m)\tau_B + \tau_{CSCR}]$ is the portion of the forward storage capacitance associated with the collector lead.

Traditionally, and especially for common-emitter applications, the best choice of m (to minimize the overall error in the extrinsic characteristics) is taken to be that which yields the proper phase of the common-emitter current gain at high frequencies ($\omega \sim 1/\tau_B$). For a uniform-base device, this best choice is $m = 5/6$, as originally suggested by Pritchard (49), although $m = 2/3$ is sometimes used (50,51). For other devices, the value of m depends on the nature of the vertical transport (e.g., drift or quasiballistic) through the base at high frequencies.

The best choice of values for the elements $r_1, r_2, \ldots, r_n$ and $C_1, C_2, \ldots, C_n$ will depend on the exact physical structure of the HBT in question; appropriate values for a conventional HBT structure are developed in Ref. 52. In many modern high-performance HBTs [e.g., see Ref. (53)], the total base resistance, r_{bb}, is comparable to the parasitic resistances, r_{cc} and r_{ee}, and to $1/g_m$, which is essentially the dynamic resistance of the device. This means that these terms cannot be neglected in deriving a simple expression for f_{max}. Starting from the circuit of Fig. 14 and performing some laborious algebraic manipulation, it can be shown that f_{max} can be well-approximated by

$$f_{\text{max}} = \sqrt{\frac{f_T}{8\pi (RC)_{\text{eff}}}} \tag{11}$$

where $(RC)_{\text{eff}}$ is dependent not only on r_{bb} and C_{bc}, as is the case for conventional BJTs, but also on r_{ee}, r_{cc}, and $1/g_m$ (50).

Quasi-Ballistic Transport

As device dimensions continue to shrink in the never-ending quest for higher speed, transport in the shortened bulk regions of the HBT can no longer be faithfully described by the classical processes of drift and diffusion (54). The base of modern HBTs is one such region whose width is approaching that of a mean-free-path length and, therefore, is a region in which quasi-ballistic transport can be expected to prevail. Some of the implications of this different transport mechanism can be appreciated by solving the one-dimensional Boltzmann transport equation in the field-free case. The appropriate equation for the electron distribution f is

$$\begin{aligned} v_z \frac{df(z,k,\theta)}{dz} &= C_{\text{in}}(z,k,\theta) - C_{\text{out}}(z,k,\theta) \\ &= C_{\text{in}}(z,k,\theta) - \frac{f(z,k,\theta)}{\tau(k)} \end{aligned} \tag{12}$$

where v_z is the electron velocity in the z-direction, k is the magnitude of the electron wavevector and is directed at an angle θ to the z axis, C_{in} and C_{out} are the incoming and outgoing collision integrals, respectively, and τ is the scattering lifetime.

Equation (12) is a first-order ordinary differential equation and, in principle, can be solved using an integrating factor. The difficulty lies in not being able to evaluate the integration constant because the full distribution function, f, is not known at any boundary. However, by splitting-up f into forward-going (f^+) and negative-going (f^-) parts, the known partial boundary conditions for carriers injected at either end of the base are sufficient to allow an iterative solution to be obtained (55,56). Figure 15 illustrates the situation for an abrupt-junction SHBT (57). Electrons from a Maxwellian distribution at the emitter edge of the emitter-base depletion region are injected via tunneling and thermionic emission into the base. This distribution, $f^+_{\text{TTE}}(0, k, \theta)$, can be evaluated by using, for example, the WKB approximation for the tunneling transmission probability, $\mathscr{T}(k, \theta)$. At the other end of the base, the hemi-Maxwellian distribution of electrons, $f^-(W, k, \theta)$, injected from the collector is also known. These two boundary conditions can then be used to solve iteratively the two equations resulting from Eq. (12), that is,

$$\begin{aligned} f^-(z,k,\theta) &= f^-(W,k,\theta)e^{(W-z)/v_z\tau(k)} \\ &\quad + \int_W^z e^{-(z-z')/v_z\tau(k)} \frac{C_{\text{in}}(z,k,\theta)}{v_z}\, dz' \end{aligned} \tag{13}$$

$$\begin{aligned} f^+(z,k,\theta) = {} & \{f^+_{\text{TTE}}(0,k,\theta) + f^-(0,k,\theta)[1 - \mathscr{T}(k,\theta)]\}e^{-z/v_z\tau(k)} \\ & + \int_0^z e^{-(z-z')/v_z\tau(k)} \frac{C_{\text{in}}(z,k,\theta)}{v_z}\, dz' \end{aligned} \tag{14}$$

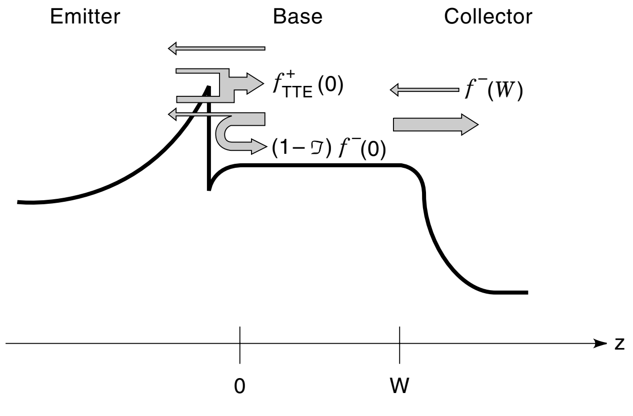


Figure 15. Illustration of the electron flows in an abrupt-junction SHBT operating in the normal, active mode (55).

The form of C_{in} depends on the scattering mechanisms involved. Here we present results for screened ionized-impurity scattering and polar optical-phonon scattering, the two dominant processes in GaAs. The SHBT considered is an AlGaAs/GaAs device with a base of width approximately equal to one mean-free-path length. The various components of the distribution at the two ends of the base are shown in Fig. 16 (57).

The ballistic component comprises electrons that have not yet scattered. Note how focused this distribution is on account of the strong dependence of tunneling on the barrier width. The reflected component is due to electrons that have been scattered and returned to the emitter–base interface and then reflected forward from that interface. The near-hemi-Maxwellian nature of this component at $z = 0$ indicates how successful the scattering has been in thermalizing the electrons. It is a surprising result in view of the narrowness of the base. The slight step at the optical phonon energy is due to the strong emission of phonons. The scattered component comprises electrons that have scattered but are still moving toward the collector. The total distribution, which shows both f^+ and f^-, is clearly far from Maxwellian, indicating the inappropriateness of trying to use the drift-diffusion formalism to describe transport in such a device.

Figure 16 gives a useful microscopic view of transport in a short base. On the quantitative side, the knowledge of $f(z)$ can be used to obtain profiles of the carrier concentration, velocity, and current. From the integrated charge density and the current, the base transit time, τ_B, can be computed. Results are shown in Fig. 17 (57) and are compared with results from a homojunction device (obtained by replacing the AlGaAs emitter in Fig. 15 by a similarly doped GaAs emitter). Also shown in the figure are results from the classical expression, as given in the text immediately following Eq. (7). Figure 17 shows that even in the homojunction case, where the injected distribution is initially hemi-Maxwellian, the classical expression overestimates τ_B. This is because nonequilibrium conditions prevail over a significant portion of the narrow base and are manifest as an increase in mean velocity that is nearly independent of bias (56). In the HBT case, the focused distribution shifts to higher energies as the forward bias increases, leading to a progressively smaller transit time.

APPLICATIONS

Why bother with HBTs? This is a legitimate question to ask in a CMOS-dominated era. The miniscule static-power consumption and physically small size of its transistors render CMOS the technology of choice for applications demanding circuits with very high levels of integration. The CMOS juggernaut has pushed Si homojunction BJTs into smaller markets, where the traditional benefits of bipolar over MOS (4)—principally, speed, transconductance and high output power—can be exploited. HBTs are worth bothering with because they can outperform BJTs. For example, they can be faster and less power-hungry, which are good attributes for digital applications; they can operate at higher signal frequencies and exhibit improved linearity, which are good attributes for analog applications; they can operate at high temperatures, making them useful in high-output-power applications. They are also amenable to monolithic integration with optical devices. Some of the applications able to capitalize upon these qualities are discussed below. It can be concluded that HBTs have a significant role to play in specialized, high-performance electronic and optoelectronic circuitry up to medium levels of integration.

Digital Circuitry

A simple comparison of the gate length of advanced CMOS transistors, say 0.5 μm, with the basewidth of advanced bipolar transistors, say 0.05 μm, suffices to indicate the speed advantage offered by bipolar. Add to this the fact that the dynamic power consumption in CMOS gates is directly proportional to the switching frequency, then it is possible to see why bipolar transistors are likely to be the devices of choice for digital medium-scale-integrated circuits operating at clock rates above about 10 GHz (27). ECL, perhaps in its CML form where the emitter buffering stage is absent, has the desired low propagation delay and power consumption per gate to make it a practical logic family at these high frequencies.

Even at their present, relatively immature state of technological development, HBTs are beginning to challenge Si BJTs, as the delay-power plot of Fig. 18 shows (27). On this figure, Hitachi's Si BJT has presently the best figure-of-merit (25 fJ), achieved partly by using silicon-on-insulator to eliminate the collector–substrate capacitance, as is done naturally in HBTs grown on semi-insulating substrates. By scaling their AlInAs/GaInAs HBT from its present emitter dimensions of $2 \times 2\ \mu\text{m}^2$ to $0.5 \times 1.25\ \mu\text{m}^2$, Hughes (now Raytheon) of California predict that they can match this speed-power figure. This company has been encouraged by its progress in InP HBT development to investigate a range of CML circuits, including latches, flip-flops, and level shifters, for applications in high-performance frequency synthesizers. For example, excellent speed–power performance has been demonstrated in static-divider circuits [see Fig. 19 (27)]. The place to be on this figure is in the top left-hand corner, a spot which is occupied by a graded-junction AlInAs/GaInAs HBT. The graded nature of the emitter is significant because it leads to a lower turn-on voltage, which contributes to the lower power consumption. Also noteworthy is the fact that the graded-emitter versions of this transistor are more reliable than their abrupt-junction counterparts (20). This is useful because it is difficult to make reproducible abrupt junctions, where steplike transitions in composition must occur, ideally, within one atomic spacing.

Another significant digital application of HBTs is in optoelectronic communications equipment, where large-signal switching is performed in, for example, laser drivers, decision circuits, multiplexers, and demultiplexers (58). The world's first commercial, 10 Gb/s, high-capacity, synchronous optical-transport system (OC192) utilizes electronic circuitry based on GaAs HBTs (59). Rockwell has reported an AlGaAs/GaAs HBT 8-channel, 1:8 data multiplexer capable of utilizing input clock rates up to 3 Gsps (60). High-performance transmitters have been reported in which a quantum-well laser and AlGaAs/GaAs HBT driver circuitry have been monolithically integrated, using selective area epitaxy to grow separately the different layers for the two device types (61).

Analog Circuitry

In analog circuit applications, desirable device features include: high transconductance, high output impedance (equiva-

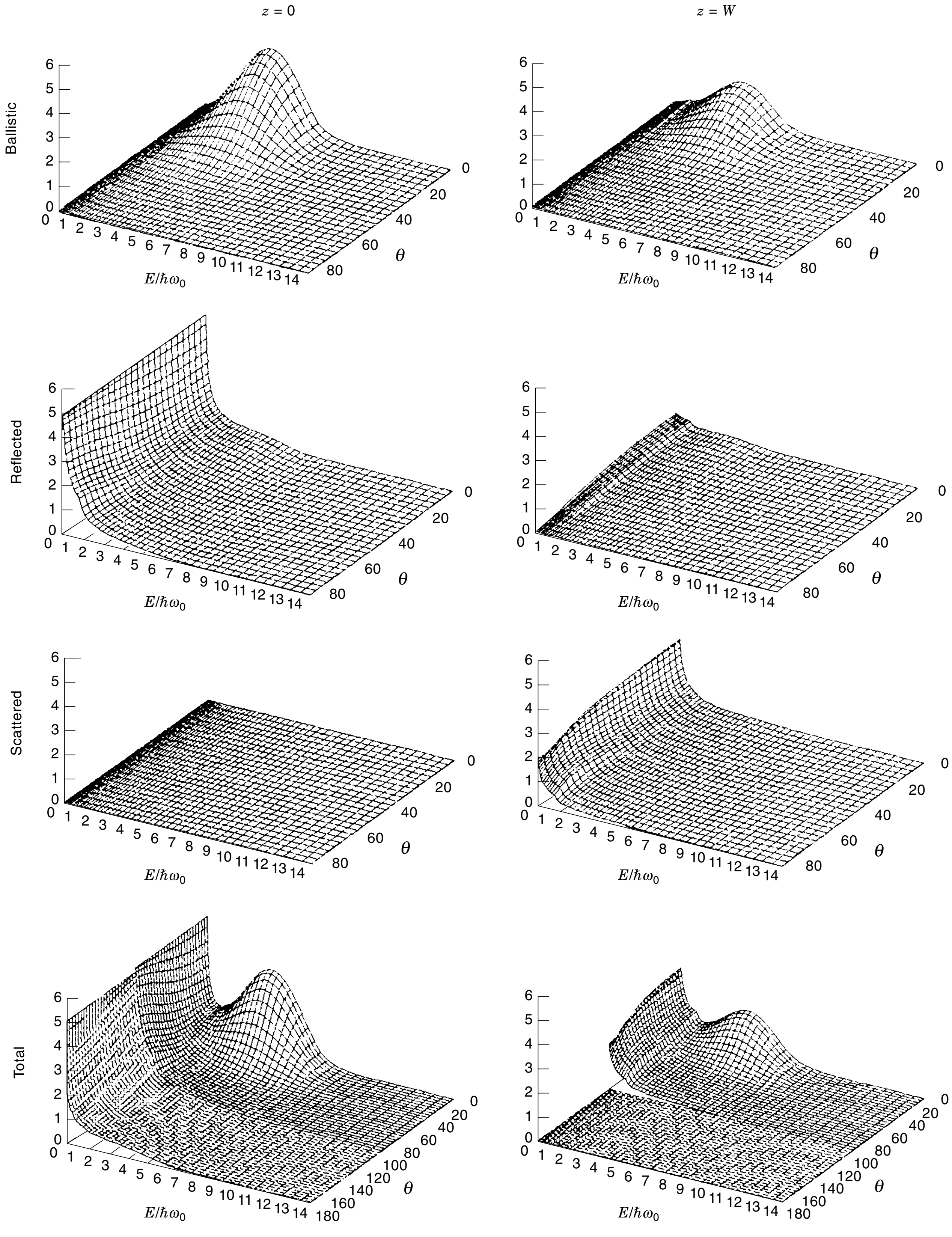


Figure 16. The various components of the electron distribution at the extreme ends of the base in an abrupt-junction AlGaAs/GaAs HBT in which scattering is via screened ionized impurities and polar-optical phonons. The distribution function is normalized to that pertaining to injection from a homojunction, and the energy is normalized to the phonon energy. $V_{BE} = 0.8V_{bi}$ and W is one mean-free-path length (57).

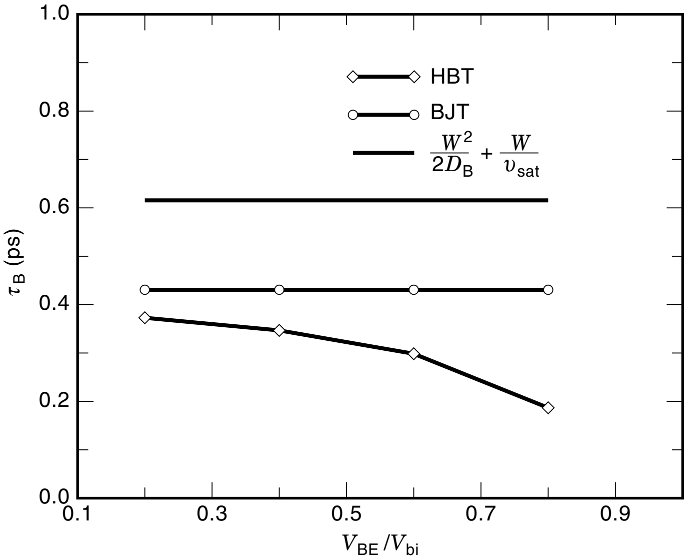


Figure 17. Comparison of τ_B values for W equal to one mean-free-path length as computed using: the classical drift-diffusion expression (thick solid line); quasiballistic analysis for a GaAs BJT (circles); quasiballistic analysis for an AlGaAs/GaAs SHBT (diamonds) (55).

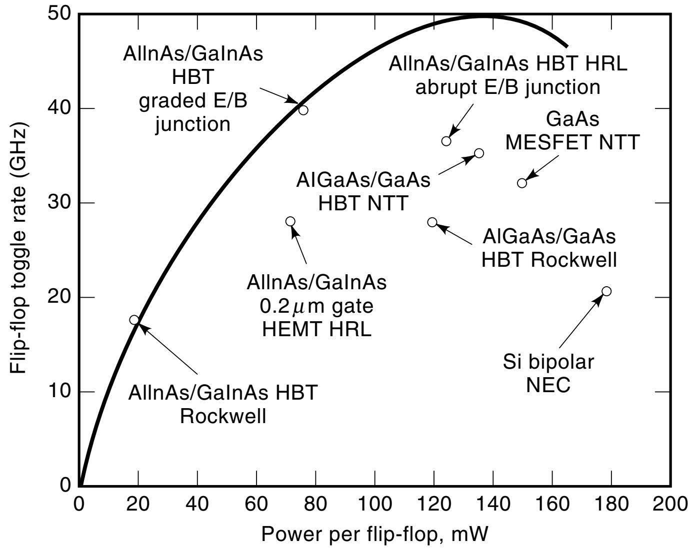


Figure 19. Speed-power comparison of graded-junction AlInAs/GaInAs HBT static-divider circuits with other technologies (27).

lent to a high Early voltage in bipolar transistors), good device matching (turn-on voltage and gain), and low noise. Generally, bipolar transistors are superior to FETs in realizing these attributes, which has led to the Si BJT being the transistor of choice for many analog applications. A challenge to this transistor is now occurring in high-end applications, where the advantages of HBTs over BJTs have led to significantly improved performance.

One such application, which has been targeted by some HBT manufacturers [IBM (62), TRW (63) and Hughes (64)] to demonstrate the device's analog capabilities, is wide-band A-D conversion. The demands placed on devices in this application are severe and include (63): minimization of parasitic capacitance to achieve wide-bandwidth performance without a large power-consumption penalty; very fast forward transit time to reduce voltage-settling time in the presence of capacitive loads; good matching of turn-on voltage and gain between transistors in order to obtain acceptable ADC threshold uniformity (dc linearity); breakdown voltges sufficiently high to accommodate large dynamic range signals; and high β and output impedance (Early voltage, V_A) to minimize the circuit complexity required to meet a given linearity specification.

To date, TRW (California, USA) has achieved impressive ADC results, with both GaAs and InP HBTs, in the form of an effective number of bits (ENOB) of 3.5 when operating under Nyquist conditions at 1 GHz (63). The bit resolution of advanced ADCs is presently around 4 to 6 (63). At higher levels, the device count and circuit complexity become issues, leading to the requirement of a low-voltage, low-power technology. This would seem to leave SiGe and InP HBTs as the main contenders. Significant in this regard is IBM's demonstration of a SiGe HBT 12 bit ADC with a transistor count of ≈2000, a power dissipation of 750 mW from a 5 V supply, and successful operation at 1.3 Gsps (62). However, impressive ADC results have also been obtained with AlGaAs/GaAs HBTs; Rockwell (58) has incorporated an 8 bit, 2.4 Gsps ADC into a

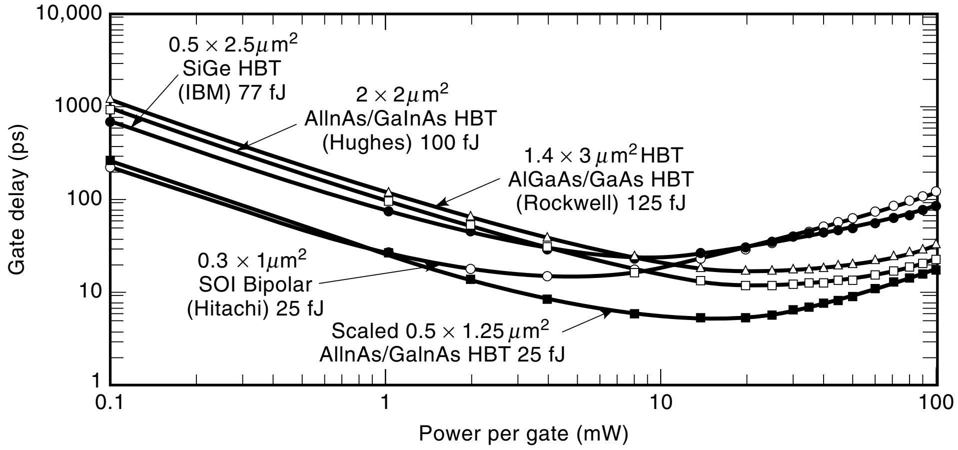


Figure 18. Delay–power relationship for a CML gate in a variety of state-of-the-art technologies (27).

prototype digital radar receiver, which has a 3 GHz analog input bandwidth and the capability of driving 50 CML gates.

Other analog applications where HBTs have performance advantages over rival homojunction bipolar and MESFET technologies are: low-noise oscillators and amplifiers (60); DACs (61); high-output-power microwave and millimeter wave monolithic integrated circuitry (65); and low-power-consumption wireless systems (62), particularly in cellular telephones, where the added attributes of operational capability at high-frequency and high output power make HBTs the transistor of choice (66). In fact, in the general area of portable, wireless communications systems, a wide variety of products based on HBT radio-frequency ICs is already available commercially, including: linear and low-noise amplifiers, quadrature modulators, and up-converters and mixers (66). A specific—and impressive—example of HBT implementation at the time of writing (February 1998) is in the cellular telephones of Nokia, Finland. Their 6100-series products employ HBT amplifiers which power-down when not transmitting, enabling about 400 hours of listening time and about 8 hours of talking time per battery charge (67).

Perhaps in the future, we will see HBTs playing the role of "master transistor" in complex, high-frequency systems-on-a-chip, in the same way that MOS transistors currently perform this function in lower-frequency CMOS designs (63). HBTs have already demonstrated, separately, the ability to enable RF, A/D, and digital functions, so the goal of integrating these functions into a single-chip system is not unreasonable. HBTs also lend themselves to integration with optical devices in mixed-signal circuitry, particularly in optical receiver front ends, both for single-wavelength operation (10,11) and for wavelength-division-multiplexed systems (68). When this optoelectronic capability is added to the analog, digital, and mixed-signal electronic attributes of HBTs, the vision of a highly versatile, truly multiple-function chip becomes an exciting prospect (69).

ACKNOWLEDGMENT

The author sincerely thanks the following people: A. R. St. Denis for supplying the figures relating to quasiballistic transport and for commenting on the manuscript; M. Vaidyanathan for supplying some of the material used in the section on small-signal modeling and for commenting on the manuscript; M. K. Jackson and N. G. Tarr for commenting on the manuscript; D. C. Streit for supplying information on TRW's devices and products; P. J. Zambardi for supplying information on Rockwell's devices and products; R. K. Surridge for his support of the author's work and for supplying information on Nortel's products; R. H. Walden for supplying reprints of conference papers concerning devices and circuits from Hughes; and P. A. Houston for stimulating and informative discussions.

BIBLIOGRAPHY

1. W. Shockley, Circuit element using semiconductor material, U.S. Patent No. 2,569,347, 1951.
2. H. Kroemer, Theory of a wide-gap emitter for transistors, *Proc. IRE,* **45**: 1535–1538, 1957.
3. A. F. J. Levi, Nonequilibrium electron transport in heterojunction bipolar transistors, In B. Jalali and S. J. Pearton (eds.), *InP HBTs: Growth, Processing and Applications,* Boston: Artech House, 1995, pp. 89–133.
4. P. M. Asbeck, Bipolar transistors, In S. M. Sze (ed.), *High-Speed Semiconductor Devices,* New York: Wiley, 1990, pp. 335–397.
5. O.-S. Ang and D. L. Pulfrey, The cut-off frequency of base-graded and junction-graded AlGaAs DHBTs, *Solid-State Electron.,* **34**: 1325–1328, 1991.
6. H. F. Chau et al., Breakdown-speed considerations in InP/InGaAs single- and double-heterostructure bipolar transistors, *IEEE Trans. Electron Devices,* **40**: 2–8, 1993.
7. T. Ishibashi and Y. Yamauchi, A possible near-ballistic collection in an AlGaAs/GaAs HBT with a modified collector structure, *IEEE Trans. Electron Devices,* **35**: 401–404, 1988.
8. C. G. Fonstad, Consideration of the relative frequency performance potential of inverted heterojunction *n–p–n* transistors, *IEEE Electron Device Lett.,* **5**: 99–100, 1984.
9. B. Agarwal et al., A 277 GHz $f_{\max}$ transferred-substrate heterojunction bipolar transistor, *IEEE Electron Device Lett.,* **18**: 228–231, 1997.
10. R. H. Walden et al., High-speed InP-based optoelectronic integrated receiver front-ends using heterojunction bipolar transistors and base-collector photodiodes. In *Proc. 12th Norchip Seminar,* Gothenburg, 1994, pp. 106–112.
11. K. Yang et al., Design, modeling and characterization of monolithically integrated InP-based (1.55 μm) high-speed (24 Gb/s) *p–i–n*/HBT front-end photoreceivers, *IEEE J. Lightwave Technol.,* **14**: 1831–1838, 1996.
12. T. P. Lester et al., A manufacturable process for HBT circuits. In *Proc. Int. Symp. GaAs and Related Compounds,* 1993, pp. 449–454.
13. F. Fantini et al., Reliability and degradation of HEMTs and HBTs. In *Technol. Dig. 21st WOCSDICE,* 1997, pp. 11–16.
14. D. C. Streit et al., Comparison of MOCVD and MBE for GaAs-AlGaAs HBT manufacturing. In *Technol. Dig. Int. Conf. GaAs Manufacturing Technol.,* 1997, pp. 162–165.
15. F. M. Yamada et al., High-reliability GaAs HBT monolithic microwave amplifier, *Technol. Dig. IEEE MTT-S,* 141–144, 1997.
16. D. C. Streit et al., High-reliability GaAs-AlGaAs HBT's by MBE with Be base doping and InGaAs emitter contacts, *IEEE Electron. Device Lett.,* **12**: 471–473, 1991.
17. N. Jourdan et al., Heavily doped GaAs(Be)/GaAlAs HBTs grown by MBE with high device performance and high thermal stability, *IEEE Trans. Electron. Devices,* **39**: 767–770, 1992.
18. W. Liu et al., Temperature dependences of current gains in GaInP/GaAs and AlGaAs/GaAs heterojunction bipolar transistors, *IEEE Trans. Electron. Devices,* **40**: 1351–1353, 1993.
19. W. Liu and S.-K. Fan, Near-ideal I–V characteristics of GaInP/GaAs heterojunction bipolar transistors, *IEEE Electron. Device Lett.,* **13**: 510–512, 1992.
20. M. Hafizi and W. E. Stanchina, Device and circuit fabrication, device characteristics and reliability. In B. Jalali and S. J. Pearton (eds.), *InP HBTs: Growth, Processing and Applications,* Boston: Artech House, 1995, pp. 135–194.
21. J. C. Bean, Materials and technologies for high-speed devices. In S. M. Sze (ed.), *High-Speed Semiconductor Devices,* New York: Wiley, Inc., 335–397, 1990.
22. D. L. Harame et al., Si/SiGe epitaxial-base transistors—Part I: Materials, physics and circuits, *IEEE Trans. Electron Devices,* **42**: 455–468, 1995.
23. F. Sato et al., A super self-aligned selectively grown SiGe Base (SSSB) bipolar transistor fabricated by cold-wall type UHV/CVD technology, *IEEE Trans. Electron Devices,* **41**: 1373–1378, 1994.

24. F. Sato et al., Sub-20 ps ECL circuits with high-performance super self-aligned selectively grown SiGe Base (SSSB) bipolar transistors, *IEEE Trans. Electron Devices,* **42**: 483–488, 1995.
25. B. S. Meyerson, UHV/CVD growth of Si and SiGe alloys: Chemistry, physics and device applications, *Proc. IEEE,* **80**: 1592–1608, 1992.
26. Y. K. Chen et al., Subpicosecond InP/InGaAs heterostructure bipolar transistors, *IEEE Electron Device Lett.,* **10**: 267–269, 1989.
27. J. F. Jensen, L. M. Burns, and W. E. Stanchina, High-speed InP HBT circuits. In B. Jalali and S. J. Pearton (eds.), *InP HBTs: growth, processing and applications,* Boston: Artech House, 1995, pp. 265–315.
28. E. F. Crabbé et al., Vertical profile optimization of very high frequency epitaxial Si- and SiGe-base bipolar transistors, *Technol. Dig. IEEE IEDM,* 83–86, 1993.
29. B. Jalali, Device physics and modeling. In B. Jalali and S. J. Pearton (eds.), *InP HBTs: Growth, Processing and Applications,* Boston: Artech House, 1995, pp. 229–263.
30. C.-M. S. Ng, P. A. Houston, and H.-K. Yow, Analysis of the temperature dependence of current gain in heterojunction bipolar transistors, *IEEE Trans. Electron Devices,* **44**: 17–24, 1997.
31. S. Tiwari, S. L. Wright, and A. W. Kleinsasser, Transport and related properties of (Ga,Al)As/GaAs double heterostructure bipolar junction transistors, *IEEE Trans. Electron Devices,* **34**: 185–194, 1987.
32. W. Liu et al., Current gain collapse in microwave multifinger heterojunction bipolar transistors operated at very high power densities, *IEEE Trans. Electron Devices,* **40**: 1917–1927, 1993.
33. K. Fricke et al., AlGaAs/GaAs HBT for high-temperature applications, *IEEE Trans. Electron Devices,* **39**: 1977–1981, 1992.
34. P. J. Zampardi, Rockwell Corporation, Thousand Oaks, California, USA, private communication, June 1997.
35. J. F. Jensen et al., *Technol. Dig. IEEE GaAs IC Symp.,* 224–227, 1994.
36. J. F. Jensen et al., High speed InP-based HBT integrated circuits. In *Proc. SPIE High-Speed Electronics and Optoelectronics,* Vol. 1680, 1992, pp. 2–11.
37. B. C. Lye et al., GaInP/AlGaAs/GaInP double heterojunction bipolar transistors with zero conduction band spike at the collector. In *Technol. Dig. 21st WOCSDICE,* 1997, pp. 124–125.
38. A. A. Grinberg et al., An investigation of the effect of graded layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors, *IEEE Trans. Electron Devices,* **31**: 1758–1765, 1984.
39. J. J. X. Feng et al., A physics-based HBT SPICE model for large-signal applications, *IEEE Trans. Electron Devices,* **42**: 8–14, 1995.
40. J. J. X. Feng, Large-signal SPICE models for heterojunction bipolar transistors and lasers, MASc. thesis, University of British Columbia, 1994.
41. S. M. Ho and D. L. Pulfrey, The effect of base grading on the gain and high-frequency performance of AlGaAs/GaAs heterojunction bipolar transistors, *IEEE Trans. Electron Devices,* **36**: 2173–2182, 1989.
42. S. Searles and D. L. Pulfrey, An analysis of space-charge-recombination in HBTs, *IEEE Trans. Electron Devices,* **41**: 476–483, 1994.
43. A. R. St. Denis, D. L. Pulfrey, and A. Marty, Reciprocity in heterojunction bipolar transistors, *Solid-State Electron.,* **35**: 1633–1637, 1992.
44. M. S. Lundstrom, Boundary conditions for *pn* heterojunctions, *Solid-State Electron.,* **27**: 491–496, 1984.
45. S. Searles, D. L. Pulfrey, and T. C. Kleckner, Analytical expressions for the tunnel current at abrupt semiconductor-semiconductor heterojunctions, *IEEE Trans. Electron Devices,* **44**: 1851–1856, 1997.
46. W. Liu, *Handbook of III-V Heterojunction Bipolar Transistors,* New York: Wiley, 1998.
47. H. Zhou and D. L. Pulfrey, Computation of transit and signal-delay times for the collector depletion region of GaAs-based HBTs, *Solid-State Electron.,* **35**: 113–115, 1992.
48. A. P. Laser and D. L. Pulfrey, Reconciliation of methods for estimating $f_{\max}$ for microwave heterojunction bipolar transistors, *IEEE Trans. Electron Devices,* **38**: 1685–1692, 1991.
49. R. L. Pritchard, *Electrical Characteristics of Transistors,* New York: McGraw-Hill, 1967, p. 274.
50. K. Kurishima, An analytical expression of $f_{\max}$ for HBT's, *IEEE Trans. Electron Devices,* **43**: 2074–2079, 1996.
51. M. B. Das, High-frequency performance limitations of millimeter-wave heterojunction bipolar transistors, *IEEE Trans. Electron Devices,* **35**: 604–614, 1988.
52. M. Vaidyanathan and D. L. Pulfrey, Extrapolated $f_{\max}$ of heterojunction bipolar transistors, submitted for publication.
53. M.-C. Ho et al., High-performance low-base-collector capacitance AlGaAs/GaAs heterojunction bipolar transistors fabricated by deep ion implantation, *IEEE Electron Device Lett.,* **16**: 512–514, 1995.
54. M. S. Lundstrom and S. Datta, Physical device simulation in a shrinking world, *IEEE Circuits and Devices Mag.,* **6** (4): 32–37, 1990.
55. A. A. Grinberg and S. Luryi, Diffusion in a short base, *Solid-State Electron.,* **35**: 1299–1309, 1992.
56. A. R. St. Denis and D. L. Pulfrey, An analytical expression for the current in short-base transistors, *Solid-State Electron.,* **38**: 1431–1436, 1995.
57. A. R. St. Denis and D. L. Pulfrey, A microscopic view of quasi-ballistic transport in HBTs, submitted for publication.
58. K. Runge et al., AlGaAs/GaAs HBT IC's for high-speed lightwave transmission systems, *IEEE J. Solid-State Circ.,* **27**: 1332–1339, 1992.
59. R. K. Surridge, Nortel, Ottawa, Canada, private communication, March 1997.
60. Rockwell Corporation, California, USA, Technical information, http://www.risc.rockwell.com/converters.
61. D. B. Slater, Jr. et al., Monolithically integrated SQW laser and HBT driver via selective OMVPE regrowth, *IEEE Photonics. Technol. Lett.,* **5**: 791–794, 1993.
62. D. L. Harame et al., Si/SiGe epitaxial-base transistors—Part II: Process integration and analog applications, *IEEE Trans. Electron Devices,* **42**: 469–482, 1995.
63. B. P. Wong and B. K. Oyama, Analog-to-digital converters using III-V HBTs. In B. Jalali and S. J. Pearton (eds.), *InP HBTs: Growth, Processing and Applications,* Boston: Artech House, 1995, pp. 317–350.
64. J. F. Jensen et al., A 3.2-GHz second-order delta–sigma modulator implemented in InP HBT technology, *IEEE J. Solid-State Circuits,* **30**: 1119–1127, 1995.
65. B. Bayraktaroglu, Highly robust GaAs cascode HBTs for microwave and millimeter-wave applications. In *Technol. Dig. 21st WOCSDICE,* 1997, pp. 120–121.
66. RF Micro Devices, Inc., North Carolina, USA, A world of wireless technology, *Designer's Handbook,* 1997.
67. Nokia Group, Finland, Technical information, http://www.nokia.com.
68. R. H. Walden, A review of recent progress in InP-based optoelectronic integrated circuit receiver front-ends, *Technol. Dig. IEEE GaAs IC Symp.,* 255–257, 1996.

69. W. E. Stanchina et al., An InP-based HBT fab for high-speed digital, analog, mixed-signal, and optoelectronic ICs, *Technol. Dig. IEEE GaAs IC Symp.,* 31–34, 1995.

D. L. PULFREY
University of British Columbia

HETEROJUNCTION DEVICES. See HETEROSTRUCTURE DEVICES.