

Contents

1	INTRODUCTION	2
2	MICROSCOPIC MODELING OF HBTs	3
2.1	Introduction	3
2.2	Direct solution of the BTE	3
2.2.1	Components of the distribution function	6
2.2.2	Profiles of electron concentration and velocity	7
2.2.3	Collector current	8
3	COMPACT MODELING OF HBTs	10
3.1	Introduction	10
3.2	Compact models for the collector current	10
3.2.1	Graded-emitter HBTs	11
3.2.2	Abrupt-junction HBTs	14
3.3	Compact models for f_T	16
3.3.1	The charge-control method	16
3.3.2	Regional signal-delay expressions	18
3.4	Compact models for f_{\max}	22
3.4.1	Base-collector network	22
3.4.2	Extrapolated f_{\max}	22
3.5	Compact model for large-signal analysis	24
3.5.1	An Ebers-Moll model	24
3.5.2	Thermal considerations	25
4	CONCLUSIONS	27
5	PROBLEMS	29
6	REFERENCES	31
7	FIGURE CAPTIONS	35

1 INTRODUCTION

Epitaxial-layer, bipolar transistors are intrinsically well-suited to high-frequency applications because their critical, physical dimensions are mainly in the direction of the semiconductor film growth, which can be controlled on a near-atomic scale. This is in contrast to field-effect transistors, where the critical dimension of gate length must be determined lithographically.

Among the family of bipolar transistors, heterojunction bipolar transistors (HBTs) are particularly attractive for operation at high frequencies because their employment of a wide-bandgap emitter allows a highly doped base region to be used without compromising the current gain [1]. With a highly doped base, the basewidth can be reduced while still maintaining an acceptable base resistance. A short basewidth leads directly to an improved cut-off frequency, f_T , which, when coupled with the lower base resistance, leads to an improved oscillation frequency, f_{\max} . These, and other, attributes of HBTs have been recently reviewed [2].

In this chapter, some important aspects of modeling high-performance HBTs are discussed. The aims are twofold: *(i)* to gain some insight into the workings of an HBT at the microscopic level; *(ii)* to use this insight to examine, or develop, analytical expressions which may be useful in the engineering-design of high-frequency and high-speed devices.

At the microscopic level, the emphasis is on the collector current density, J_C . This is an important parameter for high-frequency devices because, via the charge-control method, it is involved in the calculation of the overall signal delay time, τ_{EC} , which, in turn, is related to the common-emitter, unity gain, current cut-off frequency, f_T which, in its turn, is related to the unity gain, power cut-off frequency, f_{\max} . These two frequencies are the main performance metrics for a high-frequency device. At the engineering level, the emphasis is on providing compact models for J_C , f_T and f_{\max} . Brief consideration is also given to the incorporation of the compact model for J_C into a large-signal equivalent circuit suitable for the simulation of HBTs in switching applications.

2 MICROSCOPIC MODELING OF HBTs

2.1 Introduction

A complete microscopic model of the HBT would invoke the Boltzmann Transport Equation (BTE) to describe charge flow in the bulk regions of the device, the Schrödinger Wave Equation (SWE) to describe charge concentrations and flows at regions of abruptly changing potential, and Poisson's Equation (PE) to link internal charge distributions to external voltages. Less demanding models might be realized by, for example, using the drift-diffusion equation (DDE) as an approximate solution of the BTE, using the Jeffreys-Wentzel-Kramers-Brillouin (JWKB) approximation as a solution for the SWE, and not requiring these solutions to be consistent with PE.

In this chapter, as an example of a tractable microscopic model for HBTs, an iterative approach is described which solves directly the one-dimensional BTE in a field-free base, and accounts for quantum-mechanical tunneling at the emitter-base junction via the JWKB approximation [3, 4]. Limiting the quantum-mechanical treatment to tunneling through junctions is reasonable as other quantum aspects, such as bound energy states, are unlikely to be of importance because of the near-negligible depth of any potential notches in the device (see Fig. 1). Limiting the bulk-transport treatment to the base is reasonable, at least when computing the collector current density, J_C , because this is the zone of the associated minority-carrier flow. Limiting base transport to the field-free case is also reasonable, at least in instances of uniform doping and composition, because the high doping in the base ensures that low-level injection conditions apply. Limiting the problem to 1-D is reasonable, at least for the calculation of J_C , as the quasi-neutral basewidth, W_B , is short compared to any relevant lateral dimension of a modern HBT.

2.2 Direct solution of the BTE

When subjected to the restrictions listed above, the BTE reduces to [4]:

$$\begin{aligned} v_z \frac{df(z, k, \theta)}{dz} &= C_{\text{in}}(z, k, \theta) - C_{\text{out}}(z, k, \theta) \\ &= C_{\text{in}}(z, k, \theta) - \frac{f(z, k, \theta)}{\tau(k)}, \end{aligned} \quad (1)$$

where v_z is the electron velocity, 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. The latter three properties can be expressed in the non-degenerate case as, respectively:

$$\begin{aligned} C_{\text{in}}(z, k, \theta) &= \frac{1}{(2\pi)^3} \int_{\vec{k}'} f(z, k', \theta') S(\vec{k}', \vec{k}) d\vec{k}' \\ C_{\text{out}}(z, k, \theta) &= \frac{1}{(2\pi)^3} \int_{\vec{k}'} f(z, k, \theta) S(\vec{k}, \vec{k}') d\vec{k}' \\ &= \frac{f(z, k, \theta)}{\tau(k)}, \end{aligned}$$

where $S(\vec{k}, \vec{k}')$ is identified with a particular scattering mechanism and describes the rate of transition of carriers from an occupied state with wave vector \vec{k} to an empty state with wave vector \vec{k}' . The scattering mechanisms considered here are those due to screened-ionized impurities and polar-optical phonons; these are the most important ones for the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ material system, from which the examples in this chapter are drawn. Eqn. (1) 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 f , the distribution function, is not fully specified at any boundary. However, by splitting-up f into forward-going (f^+) and negative-going (f^-) parts, the known partial boundary conditions are sufficient to allow an iterative solution to be obtained [4, 5]. The basic forms of the two components of the distribution are:

$$f^+(z, k, \theta) = f^+(0, k, \theta) e^{-z/v_z\tau} + \int_0^z e^{(z'-z)/v_z\tau} \frac{1}{v_z} C_{\text{in}} dz' \quad (2)$$

$$f^-(z, k, \theta) = f^-(W_B, k, \theta) e^{(W_B-z)/v_z\tau} + \int_{W_B}^z e^{(z'-z)/v_z\tau} \frac{1}{v_z} C_{\text{in}} dz'. \quad (3)$$

For an abrupt junction, electrons are injected by thermionic emission over the barrier, and by tunneling through it, giving rise to a distribution function $f_{TTE}^+(0, k, \theta)$. Additionally, electrons backscattered in the base can be reflected into the forward-going ensemble by the potential spike at the junction, giving rise to a distribution function $f_{RFL}^+(0, k, \theta)$. These features are illustrated in Fig. 2. Thus, the boundary condition at $z = 0$ for the forward-going part of the electron ensemble in an abrupt-junction device is [4]:

$$f^+(0, k, \theta) = f_{TTE}^+(0, k, \theta) + f_{RFL}^+(0, k, \theta), \quad (4)$$

where

$$f_{TTE}^+(0, k, \theta) = \frac{n_E}{N_c} e^{-E_{be}/k_B T} e^{-E_k/k_B T} \mathcal{T}(k, \theta) \quad (5)$$

$$f_{RFL}^+(0, k, \theta) = f^-(0, k, \pi - \theta)[1 - \mathcal{T}(k, \theta)]. \quad (6)$$

In the above, n_E is the equilibrium (Maxwellian) electron concentration in the emitter at the distal edge of the emitter-base space-charge region, from which the injected flux originates, N_c is the effective density of states in the conduction band, E_{be} is the potential energy difference shown in Fig. 2, $k_B T$ is the thermal energy and $\mathcal{T}(k, \theta)$ is the JWKB approximation for the tunneling transmission probability. The electron energy, E_k , is computed here by assuming parabolic bands and spherical constant-energy surfaces, *i.e.*,

$$E_k = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2), \quad (7)$$

where m^* is the electron effective mass.

For graded-emitter junctions, the absence of a dominant potential spike (see Fig. 1) means that $f^+(0, k, \theta)$ can take the form of a simple hemi-Maxwellian.

At the other end of the base, $z = W_B$, a classical homojunction is assumed (see Fig. 2), regardless of the type of emitter-base junction. The collector is considered to be perfectly absorbing, but is also capable of injecting electrons into the base under appropriate bias conditions. The distribution for this flux is taken to be hemi-Maxwellian and, analogously to the emitter case, is given by

$$f^-(W_B, k, \theta) = \frac{n_C}{N_c} e^{-E_{bc}/k_B T} e^{-E_k/k_B T}, \quad (8)$$

where n_C is the equilibrium electron concentration in the collector, and the barrier E_{bc} is shown in Fig. 2.

Substituting the appropriate boundary conditions and collision integrals into (2) and (3) permits an iterative solution for the forward- and backward-going parts of the electron distribution function to be obtained. These can be summed at any position to determine $f(z, k, \theta)$, from which it is straightforward to compute useful parameters such as: carrier concentration, carrier mean velocity and current density.

The results which follow are for $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ HBTs, with the mole fraction, x , being 0.3 for abrupt-junction devices. For graded-emitter devices, the barrier for electron flow in the *npn* devices considered here is determined by the height, and not by the shape, of the conduction-band barrier (see Fig. 1). Thus it is appropriate to set $x = 0.0$ for the purpose of estimating J_C in a graded-emitter HBT when using the microscopic model. For both graded and abrupt devices, the emitter, base and collector doping densities in cm^{-3}

are, respectively, 5×10^{17} , 1×10^{19} and 5×10^{16} . In presenting the results, the basewidth is often normalized to the overall mean-free-path length, l_{sc} , which is computed from the individual energy-dependent scattering lengths for screened, ionized-impurity scattering, (SII), and polar-optical phonon scattering, (POP), as follows:

$$l_{sc} = \frac{1}{n} \frac{2}{(2\pi)^3} \int_{\vec{k}} \left[\frac{1}{l_{sc,SII}(k)} + \frac{1}{l_{sc,POP}(k)} \right]^{-1} f_{MB}(\vec{k}) d\vec{k}. \quad (9)$$

A Maxwell-Boltzmann distribution function, f_{MB} , is used only for the purpose of computing this normalizing value of l_{sc} . Its value here is 46 nm, which is in reasonable agreement with the value estimated from actual time-of-flight measurements on similarly doped material [7].

2.2.1 Components of the distribution function

The components of the distribution function for the case of an abrupt-junction HBT with basewidth $W_B = 1l_{sc}$ are shown in Fig. 3; the distribution function is normalized with respect to $n_E e^{-E_{pk}/k_B T} / N_c$, where E_{pk} is the peak barrier height at the base-emitter junction ($E_{pk}(\text{abrupt})$ in Fig. 1); the energy is normalized to the polar-optical-phonon energy (0.036 eV); the operating voltages are $V_{BC} = 0$ and $V_{BE} = 0.8 V_{bi}$, where V_{bi} is the built-in potential at the emitter-base junction. The carriers injected into the base from the emitter form the ballistic component at $z = 0$; the distribution is sharply peaked in energy and strongly focused about the z -axis ($\theta = 0$). This distinctive shape of the distribution is characteristic of tunneling through a conduction-band spike. The maximum tunnel flux occurs, irrespective of bias, at an energy which is close to 80% of the peak barrier height [8]; below this energy tunneling is reduced by the increasing thickness of the barrier; above this energy tunneling is reduced by the decreasing electron population. The distribution is focused around the z -axis because of the dependence of tunneling on the longitudinal component of energy ($E_z = \hbar^2 k_z^2 / 2m$).

The ballistic component is reduced by scattering as it transits the base. The corresponding scattered component is shown in row 3 of Fig. 3; there is a noticeable step at the phonon energy in this distribution at $z = W_B$ due to the loss of carriers by electron-POP interactions. Notice also that this distribution approaches the form of a hemi-Maxwellian, *i.e.*, it is nearly exponential in shape over the full range of the forward angle. This is somewhat surprising in view of the fact that the base has a width of only one scattering length. Similar behaviour in short-base Si homojunction devices has been reported by others and discussed

in Ref. [9]. In Eqn. (2), the first term on the right-hand side represents those carriers that are still traveling ballistically at some point z : this fraction can be written as $e^{-z/(l_{sc} \cos \theta)}$. Therefore, for $W_B = 1l_{sc}$, it is clear that even for the least favourable scattering case of $\cos \theta = 1$, over 60% of the carriers will scatter before exiting the base. Thus, there is significant scattering from the entire, injected distribution, which produces a near-thermalized, scattered component of electrons at the collector end of a short, GaAs base. Further scattering, which is experienced by the backscattered electrons, serves to drive the scattered component of the distribution even closer to that of a hemi-Maxwellian. A large fraction of the backscattered electrons at $z = 0$ is reflected by the conduction-band spike, which accounts for the near-hemi-Maxwellian nature of the reflected distribution shown in Fig. 3.

The bottom row of Fig. 3 shows both the forward- and backward-directed components of the total distribution. While the backward component builds into a near-hemi-Maxwellian distribution at the emitter end of the quasi-neutral base, the full distribution is far from thermalized at any point in the base.

For the case of the graded-emitter HBT, the components of the distribution are not nearly so interesting. In the absence of a conduction-band spike at the emitter-base interface, the distribution is injected as a near-equilibrium hemi-Maxwellian, and remains in this condition across the base, as any scattering serves only to drive it closer to its equilibrium form.

2.2.2 Profiles of electron concentration and velocity

The electron concentration profiles for the graded-emitter case and various basewidths are shown in Fig. 4. The concentration is normalized to $n_E^* = n_E e^{-E_{pk}/k_B T}$, *i.e.*, to twice the value of the concentration in the hemi-Maxwellian distribution injected over the emitter-base potential barrier of height E_{pk} (*i.e.*, $E_{pk}(\text{graded})$ in Fig. 1). As the basewidth shrinks, the profiles get progressively nonlinear, and the mean gradient decreases as the flow moves from being diffusion-limited to ballistic-limited. This sort of behaviour has been documented before for Si-like bipolar transistors [5], for which acoustic phonons provide the dominant scattering mechanism. Fig. 4 shows that, for the case of a graded-emitter device, the concentration profiles are barely sensitive to the nature of the scattering mechanism. This is a consequence of the forward- and backward-going components of the distribution maintaining near-hemi-Maxwellian forms over the entire basewidth.

The concentration profiles for the abrupt-junction case are shown in Fig. 5. The normal-

ization is again to n_E^* but, note that in this case, normalized values in excess of unity occur. This is because of both forward tunneling through the barrier and reflection of backscattered electrons at the interface (see Fig. 2). The tunneling probability, $\mathcal{T}(k, \theta)$, for the backscattered component, under the conditions of the JWKB approximation, is the same as for the injected component, but the backward tunneling current is orders of magnitude less, because of the huge difference between n_E and $n(0)$.

The mean, forward-velocity profiles for each type of junction are shown in Fig. 6. For the graded-emitter HBT, the hemi-Maxwellian nature of the injected distribution fixes the boundary condition of $v^+(0) = 2v_R$, where $v_R = \sqrt{k_B T / 2\pi m^*}$ is the Richardson velocity. Scattering tends to drive any distribution towards a Maxwellian, as Fig. 3 illustrates, so $v^+(z)$ remains close to $2v_R$ over the entire width of the base in the graded-emitter case. There is a slight increase towards the collector end of the base, where the forward-directed flow is not mitigated by scattering in the immediate vicinity of the absorbing collector [6, 9].

The $v^+(z)$ profiles for the abrupt-junction HBT differ from those for the graded-emitter HBT, not only in magnitude, but also in the fact that they are thickness- and voltage-dependent. In a short-base, abrupt-junction device, $v^+(z)$ is high, in accordance with the highly focused, energetic nature of the tunneling contribution to the injected distribution. As the base thickens, there is more backscattering and, from this component of the distribution, which tends towards a hemi-Maxwellian, reflection into the forward direction occurs at the conduction-band spike. Thus, the mean velocity of the forward component falls and, at $W_B = 10l_{sc}$, the situation is almost identical to that in a graded-emitter transistor, *i.e.*, scattering and reflection essentially destroy the focused, high-energy character of the injected electron distribution.

2.2.3 Collector current

The collector current density as a function of basewidth for the two types of HBT is shown in Fig. 7 [10]. The normalization in this case is to the ballistic current density, which is not the same for the two devices, *i.e.*,

$$\begin{aligned} J_{\text{bal,graded}} &= -q \frac{n_E^*}{2} 2v_R \\ J_{\text{bal,abrupt}} &= -qn_{TTE}^+ v_{TTE}^+ \end{aligned} \quad (10)$$

In the abrupt-junction case, n_{TTE}^+ is the concentration of injected carriers arising from tunneling and thermionic emission, and v_{TTE}^+ is their mean velocity. For the conditions cited in

Fig. 7, it is found that $n_{TTE}^+ v_{TTE}^+ = 17.1 n_E^* v_R$.

The most surprising feature of Fig. 7 is the near constancy of J_C in the abrupt-junction case, at least for basewidths up to about $W_B = 10l_{sc}$. As W_B increases, the backscattered component of the electron distribution increases, but due to specular reflection at the emitter-base junction, there is no contribution of this flux to the net current which, therefore, continues to be dominated by the injected component. It is only when $n(0)$ builds up to the extent that backward tunneling into the emitter becomes significant, that J_C drops appreciably below its ballistic value.

3 COMPACT MODELING OF HBTs

3.1 Introduction

Microscopic models, such as the one described in the previous section, provide quantitative information on the details of carrier transport within a device. This knowledge comes at the expense of a significant investment, both in model development and in program execution time. For engineering device-design purposes, at least in the early iterations of the design cycle, models which are less time-intensive, and more easily mastered, are more likely to find widespread application. Particularly useful are compact models, in which insight into the key factors determining device performance is provided by a set of analytical expressions that relate the terminal behaviour of a device to its composition and layout. Ideally, the simplifications inherent in the analytical expressions can be justified by appeal to the more rigorous results of a microscopic model. In this vein, compact models for J_C are presented here, and are contrasted with the results from the microscopic model considered above. With confidence in the expressions for J_C , it is then possible to proceed to compact models for the related high-frequency performance metrics, f_T and f_{\max} , and also to a model suitable for assessing high-speed performance.

3.2 Compact models for the collector current

The path towards a compact model for J_C starts at the second moment of the BTE [11] or, alternatively, at Eqn.(16) of Chapter 9, which can be written as:

$$J_n = q\mu_n n \mathcal{E} + qD_n \frac{\partial n}{\partial z} + 2\mu_n n \frac{\partial u}{\partial z} - \tau_{\text{sc}} \frac{\partial J_n}{\partial t} \quad (11)$$

where τ_{sc} is an average time related to scattering processes, μ_n is the mobility, D_n is the diffusivity, and u is the average z -directed kinetic energy, defined by the following expressions:

$$\frac{1}{\tau_{\text{sc}}} = \frac{\int [v_z / \tau^*(\mathbf{v})] f(z, \mathbf{v}, t) d\mathbf{v}}{\int v_z f(z, \mathbf{v}, t) d\mathbf{v}} \quad (12)$$

$$\mu_n = \frac{q\tau_{\text{sc}}}{m^*} \quad (13)$$

$$D_n = \left[\frac{2u}{q} \right] \mu_n \quad (14)$$

$$u = \frac{\int [m^* v_z^2 / 2] f(z, \mathbf{v}, t) d\mathbf{v}}{\int f(z, \mathbf{v}, t) d\mathbf{v}} \quad (15)$$

where, with respect to the variables used in Eqns. (1-3), $f(z, \mathbf{v}, t) \equiv f(z, k, \theta, t)$ and $\tau^*(\mathbf{v}) \equiv \tau(k, \theta) = \tau(k)$ for the scattering treatment used here.

To proceed further, the term involving $\partial J/\partial t$ in (11) can be safely dropped in the static case. If the energy-gradient term involving $\partial u/\partial z$ could also be ignored, then the non-equilibrium DDE would result:

$$J_n = q\mu_n n\mathcal{E} + qD_n \frac{\partial n}{\partial z}. \quad (16)$$

The validity of dropping the energy-gradient term can be judged by studying Fig. 8, which shows the concentration gradient (diffusion) and energy-gradient terms from (11) for a field-free base. For both graded and abrupt devices the magnitudes of the two terms are similar in the region close to the collector. This is a consequence of the specified nature of the collector. Its absorbing property leads to an increase in the mean forward-going velocity, and its non-injecting property (in the normal, active mode of operation) leads to a decrease in the mean backward-going velocity. Thus, there is a strong energy gradient in this region which opposes the flow due to the concentration gradient. Apart from this, for graded-emitter HBTs it would appear from Fig. 8(a) to be not unreasonable to build a compact model on consideration of the diffusion term alone. For the abrupt-junction device, it would appear that a similar construction should not be attempted with so much confidence.

If (16) is judged to be appropriate, then a final simplification would follow if the velocity distribution function, $f(z, \mathbf{v}, t)$, could be considered to maintain a form close to its equilibrium Maxwellian value over the base region. Under these circumstances, (16) would reduce to the equilibrium DDE:

$$J_n = q\mu_{n0} n\mathcal{E} + qD_{n0} \frac{\partial n}{\partial z}. \quad (17)$$

From the foregoing discussion and the results of the microscopic model it can be appreciated that (17) may be applicable to graded-emitter HBTs, but not even (16) is likely to be valid for abrupt-emitter HBTs [12].

3.2.1 Graded-emitter HBTs

It is *sine qua non* that high-frequency devices must have small dimensions, so any compact model for J_C must acknowledge the increasing importance of the boundary conditions as the basewidth shrinks. This can be elegantly done by invoking the one-flux method to represent the forward- and backward-going carrier flows at the interfaces to the base [13], as depicted in Fig. 9. The one-flux model is, in effect, a one-speed solution to the near-equilibrium BTE because it treats all fluxes as hemi-Maxwellians with a velocity of $2v_R$. This is a fair

representation of the situation in the base of a graded-emitter device, as the microscopic results of Fig. 6 indicate. The use of a constant velocity means that the energy-gradient term in the BTE (Eqn. 11) can be ignored, and the use of the equilibrium value for this velocity means that the one-flux method is equivalent to solving the equilibrium DDE with appropriate boundary conditions [14]. The relevant boundary conditions were proposed by Hansen [15]. In the derivation that follows, the exit velocity $v^+(W_B)$ is not explicitly set to $2v_R$ in order that the influence of a real, non-absorbing collector may be considered later, *i.e.*,

$$v^+(W_B) = v_{\text{coll}}. \quad (18)$$

It is implicit that the collection velocity v_{coll} does not differ from $2v_R$ to an extent that the conditions for application of the equilibrium DDE are violated. Thus, with reference to Fig. 9, the boundary conditions are:

$$\begin{aligned} J(0) &= -q\left(\frac{n_E^*}{2} - n_L\right)2v_R \\ J(W_B) &= -q(n_R v_{\text{coll}} - \frac{n_C^*}{2} 2v_R) \end{aligned} \quad (19)$$

and

$$\begin{aligned} n(0) &= n_E^*/2 + n_L \\ n(W_B) &= n_C^*/2 + n_R. \end{aligned} \quad (20)$$

For operation in the normal, active mode, $n_C^* \approx 0$, so combining the expressions for J with the field-free, equilibrium version of the DDE (Eqn. 17), and ignoring recombination in the base, gives:

$$J_C = \frac{-qn_E^*}{\left[\frac{W_B}{D_{n0}} + \frac{1}{2v_R} + \frac{1}{v_{\text{coll}}}\right]}. \quad (21)$$

The form of Eqn. (21), with each of the denominator terms being a reciprocal effective velocity, is characteristic of serial-flow transport situations [16, 17]. It is useful in identifying the bottleneck to charge flow [18] which, in this case, can be the base ($v_{\text{eff}} = D_{n0}/W_B$), or either of the two junctions ($v_{\text{eff}} = 2v_R$ for the emitter-base, or $v_{\text{eff}} = v_{\text{coll}}$ for the base-collector). Results are shown in Fig. 10 for various values of the collection, or exit, velocity, v_{coll} .

When $v_{\text{coll}} = 2v_R$, the agreement with the prediction from the microscopic model (Fig.7) is excellent. Similar agreement for the case of Si homojunction bipolar transistors has been

observed and commented upon previously [6, 10, 15]. However, as pointed out in Ref. [10], this agreement should not be construed as evidence that (17) is rigorously valid on a microscopic level. The assumptions made about the point value of D_n , the insignificance of the energy-gradient term, and the departure of v_{coll} from $2v_R$, all give cause for concern as the basewidth shrinks. If one of these assumptions is removed, *e.g.*, by making $v_{\text{coll}} = 1.15 \times 2v_R$, as suggested by the results of the microscopic model, then, as Fig. 10 indicates, (21) actually gives poorer agreement with the results from the microscopic analysis. This suggests that the assumptions made in the analytical model are somehow compensatory [6]. Nevertheless, (21) captures sufficient of the physics and gives adequate agreement with more sophisticated models to be viewed as a useful compact model.

Giving further consideration to v_{coll} , the assumption made in the microscopic model of an absorbing collector may need modification because, in real devices, phenomena such as velocity saturation and velocity overshoot could determine the effective value of the collection velocity. In homojunction devices, this velocity is often taken to be v_{sat} , the high-field saturation velocity [19, pp. 236–237]. In HBTs, because of the high doping density in the base, bandgap-narrowing phenomena are likely to induce a heterojunction at the base-collector interface (see Fig. 1), even if there is no change in material composition, and this may further reduce v_{coll} [20]. Some idea of the importance of these collection-velocity-limiting factors can be obtained by using $v_{\text{coll}} = v_{\text{sat}}$ in Eqn. (21), and also displaying some experimental data from GaAs homojunction devices [21], which behave in the same way as the graded-emitter devices being considered here. As Fig. 10 indicates, setting $v_{\text{coll}} = v_{\text{sat}}$ does significantly reduce the current, but, judging from the experimental data shown, a more realistic value of v_{coll} would appear to lie closer to the absorbing limit of $2v_R$.

Because of the widespread use of commercial, DDE-based simulators, such as MEDICI¹, it is instructive to demonstrate their capabilities for predicting the current in short-base devices. In the heterojunction mode, MEDICI employs current balancing at the interface but, in essence, uses Maxwellian distributions throughout, so the velocity of n_E^* electrons injected into the base, for example, is v_R . This leads to the factor of $1/2v_R$ in (21) being doubled. For the case of $v_{\text{coll}} = v_{\text{sat}}$, this has only a slight effect on the predicted current, as demonstrated by the lowest curve in Fig. 10.

¹Available from Avant! Corp., www.avanticorp.com

3.2.2 Abrupt-junction HBTs

The results from the microscopic model for J_C in abrupt-junction HBTs, as shown in Fig. 7, indicate that a compact model should focus on describing the tunnel and thermionic-emission currents, as the basewidth has little influence on the charge flow. Accordingly, it is reasonable to start with the JWKB approximation and consider additional simplifications to produce an analytical expression for the interfacial current. As noted in the discussion of Fig. 3, the tunneling flux reaches a maximum at some energy, which turns out to be about $0.8E_{pk}$ [8]. By making a second-order Taylor series expansion about this point, the integral for the tunneling current can be greatly simplified. Further, if the limits to the integration are allowed to extend to $\pm\infty$, rather than being limited by the bounds of the tunneling barrier, a particularly convenient expression materializes [8]:

$$\begin{aligned} J_{TTE} &= J_{TUN} + J_{TE} \\ &= J_0 C e^{-E'_{pk} \tanh(U'_{oo})/(U'_{oo})} + J_0 e^{-E'_{pk}}, \end{aligned} \quad (22)$$

where

$$\begin{aligned} J_0 &= -qn_E v_R \\ C &= \sqrt{\frac{4\pi E'_{pk} \sinh(U'_{oo})(U'_{oo})}{\cosh^3(U'_{oo})}} \\ U'_{oo} &= \frac{\hbar q}{2k_B T} \sqrt{\frac{N_E}{m_E^* \epsilon_E}}, \end{aligned}$$

where N_E , m_E^* , and ϵ_E refer to the emitter parameters of doping density, electron effective mass, and permittivity, respectively, and E'_{pk} is the peak barrier height normalized to $k_B T$.

J_{TTE} from (22) agrees almost exactly with the prediction from the microscopic model, as shown in Fig. 11. Note that the normalization in this figure is not to the ballistic current, as is the case in Fig. 7, but, instead, to the thermionic-emission current:

$$J_{TE} = J_0 e^{-E'_{pk}} = -qn_E^* v_R. \quad (23)$$

Practical, high-frequency HBTs are likely to have $W_B \approx 1l_{sc}$, so the comparison in Fig. 11 suggests that the compact model of Eqn. (22) should prove useful in the prediction of the collector current in these devices.

The remaining curve on Fig. 11 is from MEDICI simulations. In this simulator, for abrupt-junction devices, the interfacial current resulting from the balancing of currents com-

puted from the JWKB approximation is coupled to drift-diffusion currents in the bulk regions of the device [23, 24]. Analytically, the allowance for tunneling, and the presence of a conduction-band spike, can be represented by the addition of two symbols, γ and δ , respectively, to (21):

$$J_C = \frac{-qn_E^*/\delta}{\left[\frac{W_B}{D_{n0}} + \frac{1}{\gamma\delta v_R} + \frac{1}{v_{\text{coll}}}\right]} \quad (24)$$

where $\gamma = J_{TTE}/J_{TE}$ is the ratio of the total current to the thermionic-emission current and $\delta = e^{-\Delta E_n/k_B T}$, with ΔE_n being the barrier height shown on Fig. 1. Typically, γ does not exceed 10, but δ can be as small as 10^{-4} , so $\gamma\delta v_R \ll v_{\text{coll}}$ and, as $W_B \rightarrow 0$, $J_C \rightarrow J_{TTE}$, as desired. Even though Eqn. (24) has a thickness-dependent term, it turns out that the effective base transport velocity, D_{n0}/W_B , is orders of magnitude higher than the effective injection velocity, $\gamma\delta v_R$, so this expression predicts a current that is essentially constant, with the same value as that given by the compact model of (22)².

Some recent results from experimental, abrupt-junction, $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ devices, with $x = 0.25$ [22], are also displayed in Fig. 11. The ratio of the highest value of J_C/J_{TE} to the lowest value is 1.3 for the high-gain devices (TA series), and 1.5 for the TB devices. Over the same thickness range ($W/l_s c \approx 1 - 3$), the corresponding ratio for the graded-emitter devices is 1.9 (see Fig. 10). While this value is higher than those for the abrupt-junction devices, clearly more experimental data is needed before any definite corroboration can be claimed for the theory that abrupt-junction devices exhibit less thickness-dependence of J_C than graded-emitter devices.

²Note that (24) is derived assuming full-Maxwellian distributions, consistent with the treatment in MEDICI. If a hemi-Maxwellian flux approach is used, consistent with the treatment used to derive (21), then (24) is modified slightly (see Problem 3).

3.3 Compact models for f_T

3.3.1 The charge-control method

The frequency response of a transistor is related in some way to the rate at which the charge distribution within the device responds to a change in applied bias. One such relation, from basic charge-control theory [25], links the total emitter-to-collector, signal-delay time, τ_{EC} , to the changes in regional charge, ΔQ_i , that are induced by a small change in collector current, ΔI_C , with the collector-emitter voltage, V_{CE} , held constant :

$$\tau_{EC} = \frac{1}{2\pi f_T} = \sum_i \tau_i = \sum_i \left. \frac{\Delta Q_i}{\Delta I_C} \right|_{V_{CE}}. \quad (25)$$

With appropriately chosen boundaries, Q can refer to the charge of either the electrons or the holes contained within the specified region [20]. Before identifying convenient boundaries for the regions, i , it is instructive to show that (25) is equivalent to the familiar result from the small-signal, hybrid- π , equivalent circuit. To do this involves a bit of calculus:

$$\begin{aligned} Q &= Q(V_{BE}, V_{CB}) \\ \Delta Q &= \left. \frac{\partial Q}{\partial V_{BE}} \right|_{V_{CB}} \Delta V_{BE} + \left. \frac{\partial Q}{\partial V_{CB}} \right|_{V_{BE}} \Delta V_{CB} \\ \Delta Q &= \left. \frac{\partial Q}{\partial V_{BE}} \right|_{V_{CB}} \Delta V_{BE} + \left. \frac{\partial Q}{\partial V_{BC}} \right|_{V_{BE}} \Delta V_{BE} \\ \left. \frac{\Delta Q}{\Delta I_C} \right|_{V_{CE}} &= \left. \frac{\partial Q}{\partial V_{BE}} \right|_{V_{CB}} \left. \frac{\Delta V_{BE}}{\Delta I_C} \right|_{V_{CE}} + \left. \frac{\partial Q}{\partial V_{BC}} \right|_{V_{BE}} \left. \frac{\Delta V_{BE}}{\Delta I_C} \right|_{V_{CE}} \\ &= C_\pi \frac{1}{g_m} + C_\mu \frac{1}{g_m} \end{aligned}$$

where C_π and C_μ are the capacitances associated with charge changes due to changes in V_{BE} and V_{BC} , respectively, and g_m is the transconductance. This is the same as the result obtained from the hybrid- π equivalent circuit when parasitic resistances are neglected [19, p.175]. If these resistances need to be considered, as they are in the full hybrid- π circuit shown in Fig. 12, then they can be factored into the charge-control derivation by distinguishing between terminal voltages and junction voltages³.

A convenient separation of the device into “quasi-neutral” and “space-charge regions” is shown in Fig. 13. Based on this partition, the following regional delay times may be defined:

$$\tau_E = \left. \frac{\Delta Q_E}{\Delta I_C} \right|_{V_{CE}}, \quad (26)$$

³See Problem 5 at the end of this chapter.

$$\tau_B = \left. \frac{\Delta Q_B}{\Delta I_C} \right|_{V_{CE}}, \quad (27)$$

$$\tau_C = \left. \frac{\Delta Q_C}{\Delta I_C} \right|_{V_{CE}}, \quad (28)$$

where τ_E accounts for the signal-delay through the quasi-neutral emitter and the emitter-base space-charge region, τ_B represents the delay of the quasi-neutral base region, and τ_C represents the delay through the collector-base space-charge region and the quasi-neutral collector.

To compute the changes in stored charge, access is needed to some simulator that can represent the entire device, from emitter terminal to collector terminal. In the section on compact models for J_C , it was shown that a numerical simulator, such as MEDICI, can provide adequate results for the collector current, even in short-base devices. It is convenient, therefore, to use this simulator to determine the changes in regional stored charge needed to evaluate the regional delay times. The changes in stored charge can be found simply as the integrated change in electron *or* hole concentration in each region: $\Delta Q_i \equiv q \int_i \Delta n(z) dz = q \int_i \Delta p(z) dz$, where $\Delta n(z)$ and $\Delta p(z)$ represent the concentration changes brought about by the change in current ΔI_C . The values of τ_E , τ_B and τ_C so calculated can then serve as the benchmarks against which compact expressions for the regional delay times can be judged.

Extraction of the regional delay times from the numerical simulations is straightforward, provided the edges of the quasi-neutral base are taken to demarcate the three regions of the device associated with the delay times τ_E , τ_B and τ_C , as shown in Fig. 13(a). In a quasi-neutral region the changes in electron and hole concentration must be equal at each point (see Fig. 13(b)): $\Delta p(z) = \Delta n(z)$ and, furthermore, on integrating entirely across a quasi-neutral region, the total change in stored electron and hole concentrations must be equal (see Fig. 13(c)): $\int_{QN} \Delta p(z) dz = \int_{QN} \Delta n(z) dz$. This region is sharply defined in the base of HBTs because of the very high doping density that exists in this region⁴. Numerically, equality of the electron and hole quantities involved can be defined to exist if the difference between the quantities is a small fraction of either quantity taken alone.

Unfortunately, it is not feasible with this scheme to subdivide τ_E into portions that can be associated with the quasi-neutral emitter and the emitter space-charge layer. This is because such a boundary is blurred by the large number of free charges present during forward-bias

⁴The onset of the space-charge regions at either end of the base can also be identified by the relatively large change in majority-carrier concentration, as required to charge and discharge the majority-carrier, space-charge capacitance.

operation. Thus, if the numerical condition used to define equality of the electron and hole quantities involved is changed by a small amount, the resulting space-charge-region edge in the emitter changes by an unacceptable amount.

3.3.2 Regional signal-delay expressions

Traditionally in bipolar transistor analysis and design, and also commonly in HBT studies [26], the following expressions are used to estimate the regional signal-delay times:

$$\tau_E = \frac{\epsilon A}{W_{BE} g_m}, \quad (29)$$

$$\tau_B = \frac{W_B^2}{2D_{n0}} + \frac{W_B}{v_{\text{coll}}}, \quad (30)$$

$$\tau_C = \frac{\epsilon A}{W_{BC}} \left(\frac{1}{g_m} + \frac{W_E}{A\sigma_E} + \frac{W_C}{A\sigma_C} \right) + \frac{W_{BC}}{2v_{\text{sat}}}, \quad (31)$$

where the new symbols are as follows: ϵ is the permittivity; A is the cross-sectional device area; σ_E and σ_C are the conductivities of the emitter and collector materials; W_E and W_C are the quasi-neutral widths of the emitter and collector; W_{BE} and W_{BC} are the space-charge region widths at the emitter-base and collector-base junctions, respectively. The expression for τ_E accounts only for the charging and discharging of the emitter-base junction capacitance via the dynamic resistance, $1/g_m$; the effects of stored free charge in the neutral emitter and in the emitter-base space-charge region are ignored. The expression for τ_B assumes a finite electron velocity v_{coll} at the collector side of the quasi-neutral base. The expression for τ_C accounts for the effects of charging and discharging the collector-base junction capacitance via the dynamic resistance and the parasitic emitter and collector resistances; the delay due to stored free charge within the collector-base space-charge region is also included, with the factor of $\frac{1}{2}$ arising specifically from the simplifying assumption of a uniform electron velocity [19, pp. 238–240]. Equations (30)–(31) are most easily applied when the region widths are estimated by invoking the depletion approximation.

To test the suitability of the above compact expressions for estimating the regional delay times in $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ HBTs, a comparison is made below with results from MEDICI, using the relevant, resident models for mobility, recombination, Fermi-Dirac statistics, and bandgap narrowing. The devices studied here are $n\text{-Al}_{0.3}\text{Ga}_{0.7}\text{As}/p\text{-GaAs}$ HBTs with metallurgical emitter and base widths of 4000 and 1000 Å, respectively, and associated doping densities of $6 \times 10^{17} \text{ cm}^{-3}$ and $3 \times 10^{19} \text{ cm}^{-3}$. The collector and subcollector have widths of

3000 and 4000 Å, respectively, with doping levels of $4 \times 10^{16} \text{ cm}^{-3}$ and $4 \times 10^{18} \text{ cm}^{-3}$. Two devices are considered: an abrupt-junction device and a graded device with 200 Å of linear grading in the emitter. Unless indicated otherwise, the results presented are for $V_{CE} = 2 \text{ V}$ and $J_C \equiv I_C/A = 1 \times 10^4 \text{ A/cm}^2$, which represents a reasonable operating point prior to the onset of high-current effects.

Emitter Delay

Figure 14(a) shows values of τ_E versus J_C for the graded device. In employing (29) to obtain the analytical predictions, the terminal parameter g_m was set equal to that employed in the MEDICI simulation, and W_{BE} was estimated from the depletion approximation. As shown, τ_E is significantly underestimated by (29), with the error increasing at high bias. In general, the failure of (29) can be attributed to two factors, namely: (i) the breakdown of the depletion approximation due to the presence of large numbers of free carriers in the emitter-base, space-charge region, as noted earlier; (ii) the presence of significant stored charge in the quasi-neutral emitter, which is ignored in writing (29). Figure 14(b), which shows $\int_0^z \Delta p(z) dz$ in the metallurgical emitter, illustrates that stored emitter charge is unimportant in the graded device, but is very significant in the abrupt device. This is because the hole barrier is actually lower in the abrupt-junction case, due to the larger value of V_{BE} required to produce a given collector current.

Attempts have been made to improve the accuracy of the analytical expression for τ_E by including terms to account for the above factors [27, 28]. However, the suggested terms are either unwieldy, limited in their bias range of applicability, or require knowledge of poorly defined effective junction velocities. The best that can be offered at present is to note from Fig. 14 that, for graded-junction devices, where stored charge in the quasi-neutral emitter can be neglected, τ_E from (29) can be made to match the simulation results by multiplying by a correction factor of value around two [20].

Practically, the emitter delay is usually the least likely to be of concern because it can be reduced, without adversely affecting the other delay components, by operating at current densities at least twice as high as used in this comparison. Such currents can be obtained at a low enough forward bias to avoid neutral-emitter storage problems, even in abrupt-junction GaAs HBTs, by using, for example, an emitter of InGaP, which has a favourably high bandgap and low conduction-band offset.

Base Delay

In Fig. 10 the importance of the collection velocity, v_{coll} , in determining the collector current in short-base, graded-junction HBTs is apparent. The lower the velocity, the greater the charge needed to carry a given current and, consequently, the larger the stored charge in the base. This effect is embodied in (30) by the term W_B/v_{coll} , which arises analytically from the supposed trapezoidal shape of the minority-carrier charge profile. Fig. 5 indicates that this is not a bad description of the charge profile, even in a short-base, abrupt-junction device. The predictions of (30) for τ_B are shown in Fig. 15. The MEDICI simulations predict $v_{\text{coll}} = 0.28v_{\text{sat}}$, and when this value is used in (30), the agreement between the analytical and numerical results is excellent. What may be surprising is that MEDICI gives the same result for both graded-emitter and abrupt-emitter HBTs. This is because in this type of DDE simulator, at least when the energy-balance option is not operative, the full height and width of the induced heterojunction barrier at the base-collector interface are “felt” by the flux exiting the base. This serves, in our examples, to reduce v_{coll} to the pessimistic value of $0.28v_{\text{sat}}$. In reality, the forward-going flux in a short-base device will be less affected by the presence of a base-collector barrier because its velocity is not v_R , as implied by the full-Maxwellian description of charge concentrations in a classical DDE simulator, but is, as we have seen from the microscopic model, either $2v_R$ in a graded device, or v_{TTE}^+ in an abrupt device. In practical HBTs, therefore, τ_B can be expected to lie closer to the limit given by (30) when employing either of these velocities, rather than when using v_{coll} as predicted by MEDICI.

With short bases and potentially high electron velocities, τ_B is no longer the major contributor to the overall delay in modern bipolar transistors. It can be further reduced by employing compositional grading in the base, in which case a convenient expression for τ_B is⁵:

$$\tau_B = \frac{W_B^2}{bD_{n0}} \left[1 - \frac{(1 - e^{-b})}{b} \right] + \frac{W_B}{v_{\text{coll}}} \frac{(1 - e^{-b})}{b} \quad (32)$$

where $b = \Delta E_g/kT$ and ΔE_g is the bandgap change in the base due to linear compositional grading. The results in Fig. 16 show how useful a graded base could be in mitigating the effect of a low collection velocity.

⁵See Problem 7 at the end of this chapter

Collector Delay

Figure 17 shows values of τ_C versus J_C for the graded device (similar results are obtained for the abrupt-junction device); results from MEDICI are presented along with the predictions of (31). In employing (31), the terminal parameter g_m was set equal to that employed in the MEDICI simulation, and the value of W_E was set equal to the metallurgical emitter width. The value of W_{BC} was computed by employing the usual, one-sided, depletion-approximation in two different circumstances. In the first instance, the space-charge region on the collector side of the junction was assumed to be fully depleted, giving a charge density of magnitude qN_C , where N_C is the collector doping density. In the second instance, the magnitude of the charge density was taken to be $q(N_C - n_c)$, where n_c accounts for the finite electron concentration required to support the collector current, $J_C = -qn_c v_{\text{sat}}$. The results emphasize the importance of accounting for this mobile charge in the space-charge region, even at current-density magnitudes below the onset of the Kirk effect [29], which occurs in these devices when $|J_C| > qN_C v_{\text{sat}} \approx 4 \times 10^4 \text{ A/cm}^2$.

Further subtle effects in the space-charge region can occur due to: (i) the dependence of the electron velocity on both the electric field and the junction self-heating [30]; (ii) the sensitivity of the velocity profile to the electron space-charge density [30, 31]. The former effect can increase τ_C with increasing collector voltage, while the latter can decrease the delay at high currents. For a fully depleted collector these effects can be incorporated in an improved expression for the transit-related term in (31) [30]:

$$\tau_{Cd} = [\kappa_0 + \kappa_2(\Delta T)] \frac{W_{BC}}{2} + \kappa_1 \frac{V_{bi} - V_{BC}}{2} + \kappa_1 \frac{q(N_C - 2\bar{n})W_{BC}^2}{12\epsilon}, \quad (33)$$

where \bar{n} is the average electron density in the space-charge region, ΔT is the rise in junction temperature, and the coefficients κ come from a linear fit to the inverse velocity $vs.$ field relation for GaAs at high field strengths.

As τ_C is presently the most significant of the delay times in high-performance HBTs, there is no shortage of collector designs to reduce this portion of the signal delay. Exploiting velocity overshoot, by engineering a low-field region in the part of the space-charge region proximal to the base, can be worthwhile [32]. From calculations using a phenomenological model, it has been suggested that this improvement can be simply quantified in GaAs junctions by changing the factor of $\frac{1}{2}$ in (31) to $\frac{1}{3}$ [33].

3.4 Compact models for f_{\max}

3.4.1 Base-collector network

The cut-off frequency, f_T , relates to the current gain of a transistor. While this is an important property of a device, of more practical significance is the ability of a transistor to produce a power gain at high frequencies. The appropriate figure of merit is f_{\max} , which has long been described by the compact expression [19, p.177]:

$$f_{\max} = \sqrt{\frac{f_T}{8\pi RC}}, \quad (34)$$

where RC is some base-collector time constant, as discussed below. f_{\max} refers to the frequency at which the power gain becomes unity, when determined by extrapolation at -20 dB/decade from lower frequencies. It is relevant to employ an extrapolated figure of merit because, in reality, electronic instrumentation is presently incapable of directly measuring the high values of f_{\max} which are characteristic of high-performance HBTs.

In early transistors it was permissible to view the base resistance, r_b , and the collector-base junction capacitance, C_{jc} , as lumped elements, *i.e.*, $RC = r_b C_{jc}$. More recently, the distributed nature of the base-collector RC network has been accounted for by choosing an effective value for this time constant, namely: $RC = (r_b C_{jc})_{\text{eff}}$ [34]. In the equivalent circuit of Fig. 12, for example, r_b and C_{jc} are each shown as being distributed into n components. The case of $n = 3$ for a conventional, mesa-style HBT, is worth considering briefly as the three components relate directly to obvious features of the physical structure. With reference to Fig. 18, the three features are the intrinsic, extrinsic and contact regions of the device, and are represented by (r_{bi}, C_{jci}) , (r_{bx}, C_{jcx}) and (r_{bc}, C_{jcc}) , respectively. It can be shown⁶ that, at high frequencies, the appropriate base-collector network is given by the circuit in Fig. 18 (b), where r_{cv} is the purely vertical part of the base contact resistance [35]. In this network, the three components of r_b and C_{jc} to substitute into Fig. 12 are clearly identifiable.

3.4.2 Extrapolated f_{\max}

If Eqn.(34) is to remain useful for modern, high-performance HBTs, the RC time constant in this compact expression must be further modified. Modification is necessary because of violation of one of the assumptions on which the original derivation of (34) is based, namely: that the collector and emitter resistances, r_{cc} and r_{ee} , and the dynamic resistance, $1/g_m$, are

⁶See Problem 8 at the end of the chapter.

negligible compared to the base resistance. This assumption is invalid for high-performance HBTs due to the employment of very highly doped base regions.

To find the appropriate form for RC in order that (34) can be applied to high-performance HBTs, it is necessary to start with an equivalent circuit, such as in Fig. 12, and move towards the small-signal parameters on which (34) is based [36], systematically dropping terms of relatively minor importance. The approach has been carefully documented [35, 37], and leads to compact expressions for RC , which can then be substituted into (34) to compute f_{\max} . When f_{\max} is based on the extrapolation of Mason's unilateral gain, U , to 0 dB [38], the relevant expression for RC is:

$$RC = (RC)_{\text{eff}}^U = (r_b C_{jc})_{\text{eff}}^U + [2\pi f_T r_{cc} C_{jc}] \left(r_{ee} + \frac{1}{g_m} \right) C_{jc}, \quad (35)$$

where $(r_b C_{jc})_{\text{eff}}^U$ accounts for the distributed nature of the base resistance and the collector capacitance [34]. This expression has been tested against results from the SPICE simulation of the circuit in Fig. 12 for the case of a high-performance device in which $r_{ee} = 6.2 \Omega$ and $r_{cc} = 16 \Omega$ [35]. Excellent agreement was demonstrated. By contrast, if the new terms in (35) are neglected, $(RC)_{\text{eff}}^U$ is underestimated by about 30%.

It can be appreciated from (35) that the collector resistance must be minimized if extremely high values of f_{\max} are to be obtained. It is noteworthy in this regard that the world-record device described at the end of the following chapter has negligible r_{cc} , due primarily to its employment of a Schottky barrier collector contact.

3.5 Compact model for large-signal analysis

In the previous two sections we considered the small-signal modeling of HBTs, such as would be appropriate for high-frequency amplifiers, for example. In this section we briefly examine large-signal modeling, which has relevance for applications in high-speed switching, for example.

3.5.1 An Ebers-Moll model

The well-known Ebers-Moll representation of a three-terminal bipolar device is a convenient starting point for the formulation of a large-signal model for HBTs. In fact, the traditional model for BJTs applies without modification to graded-emitter HBTs, at least in those cases where there is no significant conduction-band spike due to bandgap narrowing (see Fig. 1). This is because the electronic collector current, J_C , has the same ideal Boltzmann dependence ($e^{qV/k_B T}$) as the hole currents considered in the intrinsic model. This follows from Eqn.(21), where

$$n_E^* = n_E e^{-E_{pk}/k_B T} = n_E e^{-qV_{bi}/k_B T} e^{qV_{BE}/k_B T}. \quad (36)$$

On the other hand, in an abrupt-junction HBT, the presence of a large conduction-band spike at the emitter-base junction causes the electron- and hole-injection currents to have different voltage dependencies. In an equivalent circuit, such as the one shown in Fig. 19, this phenomenon can be represented by using ideal diodes for the hole currents (IPC and IPE), a current source based on non-ideal diodes for the collector electron current (ICT) and non-ideal diodes for the quasi-neutral base current (INC+INE). Details of this circuit are given elsewhere [2, 40]: here we concentrate on the representation of ICT in an abrupt-junction HBT.

Although quasi-neutral base recombination is an important contributor to the base current in some HBTs [22], it can usually be ignored in computing J_C , as we have done throughout this chapter. Under these circumstances, and for short-base transistors, Eqn.(24) yields:

$$J_C \approx -\gamma q n_E^* v_R, \quad (37)$$

which is then precisely equal to $J_{TUN} + J_{TE}$, as given by Eqn.(22). There are two ways in which the conduction-band spike contributes to the non-ideal nature of these components of J_C . Firstly, because E_{pk} is now determined by the potential energy difference across only the

emitter side of the junction, rather than across the entire junction, the thermionic emission component of the current is not dependent on the full value of V_{BE} . Instead, elaborating on Eqn.(23), it is given by [8]:

$$J_{TE} = J_0 e^{-E'_{pk}} = J_0 e^{-qN_{rat}V_{bi}/k_B T} e^{-qN_{rat}V_{BE}/n_2 k_B T}, \quad (38)$$

where $N_{rat} = \epsilon_B N_B / (\epsilon_B N_B + \epsilon_E N_E)$, and the new diode ideality factor is $n_2 = 1/N_{rat}$. In reality, because of the very high base doping density in HBTs, $n_2 \approx 1.0$, so this deviation from ideality is barely significant. However, the second factor, which concerns the tunneling current, is significant. The deviation from ideality of J_{TUN} is due to not only the splitting of the potential drops at the junction, but also to the fact that the tunneling current is carried by electrons of energy less than E_{pk} . These effects are embodied in Eqn.(22) which, when written out in more detail, leads to another diode current, namely [8]:

$$J_{TUN} = J_0 C^* e^{(-qN_{rat}V_{bi}/k_B T)(\tanh U'_{oo}/U'_{oo})} e^{qV_{BE}/n_1 k_B T}, \quad (39)$$

where C^* is C from Eqn.(22) evaluated at the desired V_{BE} , and the diode ideality factor is $n_1 = U'_{oo}/N_{rat} \tanh(U'_{oo})$. Thus, when using Fig. 19 for an abrupt-junction HBT, ICT ($=J_C \cdot \text{Area}$) can be represented by a current source controlled by the voltage across two parallel diodes, with characteristics described by Eqns.(38) and (39).

Regarding the resistive components RB, RE and RC in Fig. 19, these are the parasitic resistances equivalent to r_b , r_{ee} and r_{cc} , respectively, introduced previously in the small-signal equivalent circuit of Fig. 12. Similarly, the capacitances CE and CC are equivalent to C_π and C_μ , respectively. In a SPICE implementation of the large-signal equivalent circuit, the storage component of CE would be represented by (ICT.TF/ V_t), where V_t is the thermal voltage and TF is the forward transit time. A reasonable value for the latter parameter can be obtained by adding the base- and collector-transit times, as given by Eqns.(32) and (33).

3.5.2 Thermal considerations

In practical n - p - n HBTs the barrier height presented to electrons attempting to enter the base is either E_{pk} , in the case of thermionic emission, or somewhat less than this if tunneling is the dominant flow mechanism (see Eqn.(22)). In either case, this barrier is less than the corresponding barrier height, say E_{val} , which inhibits holes from being back-injected into the emitter. Thus, the emitter injection efficiency is approximately $\propto \exp(E_{val} - E_{pk}/k_B T)$.

Because there is very little quasi-neutral-base recombination in high-performance HBTs, it follows that the common-emitter current gain, β , will fall with temperature. This can present a problem in devices built upon substrates of poor thermal conductivity, such as GaAs. Transient local heating can occur when the collector current is switched to high densities, and this would lead to a negative differential output conductance, which is usually not desirable.

A simple way of accounting for thermal effects is to add the electrical equivalent of a thermal network to the main electrical equivalent circuit discussed in the previous subsection. The sub-circuit shown in Fig. 19 comprises a current source $IPW \equiv I_C V_{CE} (W)$, a resistor R_{TH} representing the thermal resistance ($^{\circ}C/W$), a capacitor such that $C_{TH} R_{TH}$ is equivalent to the thermal time constant (s), and a dc voltage source V_{SUB} which represents the substrate temperature ($^{\circ}C$) [41]. Thus, the voltage at node X of the sub-circuit corresponds to the temperature of the device. This temperature can then be used to recompute the temperature-dependent model-parameter values for the elements in the main electrical equivalent circuit. In this manner, some thermo-electric feedback is included in the large-signal simulation.

4 CONCLUSIONS

From the material presented in this chapter on the modeling of short-base, high-performance HBTs, the following observations can be made:

1. Microscopic modeling is able to show that, over most of the base region of a graded-emitter device, the electron distribution function maintains a near-equilibrium form, and the concentration-gradient term in the BTE dominates over the energy-gradient term. These factors suggest that the equilibrium DDE, with appropriate boundary conditions, may adequately describe base transport in these devices. For abrupt-emitter HBTs this is not the case.
2. Guided by the microscopic results, compact models for the collector current in both abrupt and graded devices can be developed. The different dependencies of J_C on basewidth for these two types of device can be largely reproduced by the compact models.
3. Compact models for the signal delay time can adequately represent the contributions to f_T from delays in the base and collector regions of HBTs.
4. A compact model for f_{\max} can be derived which takes into account the effects of emitter and collector resistances, as demanded in the modeling of modern HBTs exhibiting low values of base resistance.
5. It is possible to incorporate the compact model for J_C into a simple Ebers-Moll, large-signal equivalent circuit which can aid in the evaluation of the high-speed switching performance of HBTs.

Acknowledgement

Over the last 10 years it has been my privilege to work on HBT modeling issues at UBC with a group of very talented graduate students: S. Ho, O.-S. Ang, A. Laser, H. Zhou, B. Ghodsian, J. Feng, S. Searles, M. Vaidyanathan, A. St.Denis and S. Fathpour. My present

understanding of the subject is due, in large part, to the endeavours of these gentlemen. I sincerely thank them all. I particularly acknowledge Drs. St.Denis, Vaidyanathan and Searles, whose work has contributed directly to the material in this chapter.

5 PROBLEMS

1. Derive Eqn. (24) for the collector current in an abrupt-junction HBT by balancing the diffusion current in the base against both the net current at the emitter-base junction and the exit current ($-qn(W_B)v_{\text{coll}}$) at the base-collector junction.
2. Is it of any consequence that the equilibrium diffusivity is used in Eqn. (24), rather than some effective diffusivity that would account for not only the non-Maxwellian nature of the distribution (see Fig. 3), but also for the presence of an appreciable charge flow due to changes in the electron kinetic energy (see Fig. 8)?
3. Eqn. (24) can be improved if a hemi-Maxwellian flux approach is used in the derivation of J_C , in the same manner as described in the development of (21). The key to the derivation is a proper accounting for the flux that is reflected back into the base from the energy barrier at the emitter-base junction (see Fig. 2).

Show that the above considerations lead to:

$$J_C = -q\frac{n_E^*}{2} [1 - \gamma + 2/\delta] / \left[\frac{W_B}{D_{n0}} + \frac{(2-\gamma\delta)}{\gamma\delta 2v_R} + \frac{1}{v_{\text{coll}}} \right].$$

Note that this equation reduces to (21) in the case of a homojunction, and that it reduces to (24) if $\gamma\delta \ll 2$, as will usually be the case for a heterojunction when an appreciable band spike is present.

4. In Fig. 15 it is shown that the the presence of a barrier at the base-collector junction, through its influence on v_{coll} , has a large effect on τ_B . However, from Eqn. (24), it can be shown that this barrier barely affects the collector current density in a short-base, abrupt-emitter HBT.

Is there a contradiction here? Explain?

5. Amend the derivation of $\frac{\Delta Q}{\Delta I_C} \Big|_{V_{CE}}$ at the beginning of the section entitled “The charge-control method” so that the effects of parasitic resistance in the emitter and collector are taken into account.

Confirm that the result obtained is identical to the expression for $1/(2\pi f_T)$ resulting from the hybrid- π equivalent circuit in Fig. 12, for the case of a lumped representation of the collector capacitance.

6. Establish the conditions under which the differential charge formulation for the emitter-collector delay, $\Delta Q/\Delta I_C$, is equivalent to the integral formulation, Q/I_C .
Are these conditions met for an abrupt-junction HBT?
7. Derive Eqn. (32) for the delay in the linearly graded base region of an abrupt-junction HBT.
8. Starting from the schematic cross-section of an HBT shown in Fig. 18, draw a general equivalent circuit for the base-collector RC network, and show that it reduces to the circuit given in Fig. 18(b).

References

- [1] H. Kroemer, “Theory of a wide-gap emitter for transistors”, *Proc. IRE*, vol.45, 1535-1538, 1957.
- [2] D.L. Pulfrey, “Heterojunction bipolar transistor”, *Wiley Encyclopedia of Electrical and Electronics Engineering*, J.G. Webster, Ed., New York: John Wiley & Sons Inc., vol.8, 690-706, 1999.
- [3] A.R. St.Denis, “A one-dimensional solution of the Boltzmann Transport Equation with application to the compact modeling of modern bipolar transistors”, *Ph.D. thesis*, University of British Columbia, 1999.
- [4] A.R. St.Denis and D.L. Pulfrey, “Quasiballistic transport in GaAs-based heterojunction and homojunction bipolar transistors”, *J. Appl. Phys.*, vol.84, 4959-4965, 1998.
- [5] A.A.Grinberg and S.Luryi, “Diffusion in a short base”, *Solid-State Electron.*, vol.35, 1299-1309, 1992.
- [6] A.R.St.Denis and D.L.Pulfrey, “An analytical expression for the current in short-base transistors”, *Solid-State Electron.*, vol.38, 1431-1436, 1995.
- [7] E.S. Harmon, M.L. Lovejoy, M.R. Melloch, M.S. Lundstrom, T.J. de Lyon and J.M. Woodall, “Experimental observation of minority carrier mobility enhancement in degenerately doped p-type GaAs”, *Appl. Phys. Lett.*, vol.63, 536-538, 1993.
- [8] S. Searles, D.L. Pulfrey and T.C. Kleckner, “Analytical expressions for the tunnel current at abrupt semiconductor-semiconductor heterojunctions”, *IEEE Trans. Electron Dev.*, vol.44, 1851-1856, 1997.
- [9] M.A. Stettler and M.S. Lundstrom, “A microscopic study of transport in thin base silicon bipolar transistors”, *IEEE Trans. Electron Dev.*, vol.41, 1027-1033, 1994.
- [10] D.L. Pulfrey, A.R. St.Denis and M. Vaidyanathan, “Compact modeling of high-frequency, small-dimension bipolar transistors”, *Proc. IEEE COMMAD*, Perth, Australia, pp. 81-85, 1998.

- [11] M.S. Lundstrom, *Fundamentals of Carrier Transport*, Vol.X, *Modular Series on Solid State Devices*, Reading, Massachusetts: Addison-Wesley, 1990.
- [12] F. Assad, K. Banoo and M.S. Lundstrom, "The drift-diffusion equation revisited", *Solid-State Electron.*, vol.42, 283-295, 1998.
- [13] S. Tanaka and M.S. Lundstrom, "A compact HBT device model based on a one-flux treatment of carrier transport", *Solid-State Electron.*, vol.37, 401-410, 1994.
- [14] M. Vaidyanathan and D.L. Pulfrey, "An appraisal of the one-flux method for treating carrier transport in modern semiconductor devices", *Solid-State Electron.*, vol.39, 827-832, 1996.
- [15] O. Hansen, "Diffusion in a short base", *Solid-State Electron.*, vol.37, 1663-1669, 1994.
- [16] A. Marty, G. Rey and J.P. Bailbe, "Electrical behavior of an *n*p*n* GaAlAs/GaAs heterojunction bipolar transistor", *Solid-State Electron.*, vol.22, 549-557, 1979.
- [17] M.A. Stettler and M.S. Lundstrom, "A detailed investigation of heterojunction transport using a rigorous solution to the Boltzmann Transport Equation", *IEEE Trans. Electron Dev.*, vol.41, 592-600, 1994.
- [18] A.R. St.Denis, D.L. Pulfrey and A. Marty, "Reciprocity in heterojunction bipolar transistors", *Solid-State Electron.*, vol.35, 1633-1637, 1992.
- [19] D.J. Roulston, *Bipolar Semiconductor Devices*, New York: McGraw-Hill, 1990.
- [20] D.L. Pulfrey, S. Fathpour, A.R. St.Denis, M.Vaidyanathan, W.A. Hagley and R.K. Surridge, "Applicability of the traditional compact expressions for estimating the regional signal-delay times of HBTs", *J. Vac. Sci. Techn. A*, vol.18, no.2, 2000.
- [21] E.S. Harmon, M.R. Melloch, M.S. Lundstrom and F. Cardone, "Thermal velocity limits to diffusive electron transport in thin-base *np*⁺*n* GaAs bipolar transistors", *Appl. Phys. Lett.*, vol.64, 205-207, 1994.
- [22] R.E. Wesler, N. Pan, D.-P. Vu, P.J. Zampardi and B.T. McDermott, "Role of neutral base recombination in high gain AlGaAs/GaAs HBTs", *IEEE Trans. Electron Dev.*, vol.46, 1599-1607, 1999.

- [23] A.A. Grinberg, M.S. Shur, R.J. Fischer and H. Morkoç, “An investigation of the effect of graded layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors”, *IEEE Trans. Electron Dev.*, vol.31, 1758-1765, 1984.
- [24] K. Yang, J.R. East and G.I. Haddad, “Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition”, *Solid-State Electron.*, vol.36, 321-330, 1993.
- [25] H.K. Gummel and H.C. Poon, “An integral charge control model of bipolar transistors”, *Bell System Technical Journal*, vol.49, 827-852, 1970.
- [26] W. Liu, *Fundamentals of III-V Devices*, p.265, New York: John Wiley & Sons, 1999.
- [27] J.J. Liou, F.A. Lindholm and D.C. Malocha, *J. Appl. Phys.*, vol.63, 5015-5022, 1988.
- [28] G.-B. Gao, J.-I. Chyi, J. Chen and H. Morkoç, *Solid-State Electron.*, vol.33, 389-390, 1990.
- [29] J.C.T. Kirk, *IRE Trans. Electron Dev.*, vol.9, 164-174, 1962.
- [30] L.H. Camnitz and N. Moll, “An analysis of the cut-off frequency behavior of microwave heterostructure transistors”, S. Tiwari, Ed., New York: IEEE Press, pp.21-45, 1993.
- [31] Y. Betser and D. Ritter, “Reduction of the base-collector capacitance in InP/GaInAs heterojunction bipolar transistors due to electron velocity modulation”, *IEEE Trans. Electron Dev.*, vol.46, 628-633, 1999.
- [32] C.M. Maziar, M.E. Klausmeier-Brown and M.S. Lundstrom, “A proposed structure for collector transit-time reduction in AlGaAs/GaAs bipolar transistors”, *IEEE Electron Device Lett.*, vol.EDL-7, 483-485, 1986.
- [33] 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.*, vol.35, 113-115, 1992.
- [34] M. Vaidyanathan and D.J. Roulston, “Effective base-collector time constants for calculating the maximum oscillation frequency of bipolar transistors”, *Solid-State Electron.*, vol.38, 509-516, 1995.

- [35] M. Vaidyanathan and D.L. Pulfrey, “Extrapolated f_{\max} of heterojunction bipolar transistors”, *IEEE Trans. Electron Dev.*, vol.46, 301-309, 1999.
- [36] R.L. Pritchard, *Electrical Characteristics of Transistors*, New York: McGraw-Hill, Ch.8, 1967.
- [37] M. Vaidyanathan, “Compact models for the high-frequency characteristics of modern bipolar transistors”, *Ph.D. thesis*, University of British Columbia, 1998.
- [38] S.J. Mason, “Power gain in feedback amplifier”, *Trans. IRE*, vol.CT-1, 20-25, 1954.
- [39] S. Searles and D.L. Pulfrey, “An analysis of space-charge-region recombination in HBT’s”, *IEEE Trans. Electron Dev.*, vol.41, 476-483, 1994.
- [40] J.J.X. Feng, D.L. Pulfrey, J. Sitch and R.K. Surridge, “A physics-based HBT SPICE model for large-signal applications”, *IEEE Trans. Electron Dev.*, vol.42, 8-14, 1995.
- [41] P. Roblin, personal communication, Jan. 2000.

Figure 1: Conduction-band profiles for abrupt- and graded-emitter $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ HBTs operating under similar forward-bias conditions. In the abrupt case $x = 0.3$; in the graded case x is linearly changed over a distance of $0.02 \mu\text{m}$. The base region extends from $0.40 - 0.45 \mu\text{m}$. Note the presence of a small barrier at the base-collector junction. This is caused by bandgap narrowing in the base region. This phenomenon is also responsible for the slight spike remaining in the conduction band at the emitter-base junction of the graded device. Note also the two different definitions of the peak barrier height E_{pk} , depending on whether the emitter-base junction is abrupt or graded.

Figure 2: Montage illustrating the injected and reflected electron fluxes at the emitter-base junction, and the collected and back-injected fluxes at the base-collector junction [4]. Note that W in the figure is equivalent to W_B in the text.

Figure 3: Normalized components of the electron distribution function at the ends of the quasi-neutral base of an abrupt-junction HBT with $V_{BE} = 0.8V_{bi}$ and $V_{BC} = 0$ [4]. For the forward direction, $0 \leq \theta < \pi/2$, and for the backward direction, $\pi/2 < \theta \leq \pi$. Note: $W \equiv W_B$.

Figure 4: Basewidth dependence of the carrier profile for two different sets of scattering mechanisms for a graded-emitter HBT under operating conditions of $V_{BE} = 0.8V_{bi}$ and $V_{BC} = 0$ [3]. Note: $W \equiv W_B$.

Figure 5: Normalized carrier concentration profiles in an abrupt-junction HBT for various basewidths and the same bias conditions as in Fig. 4. The abscissae are z/W_B and the ordinates are $n(z)/n_E^*$ [3]. Note: $W \equiv W_B$.

Figure 6: Normalized, mean forward-velocity profiles for both graded-junction HBTs (solid lines) and abrupt-junction HBTs (dashed lines), for three different basewidths and the same bias conditions as in Fig. 4 [4]. Note: $W \equiv W_B$.

Figure 7: Normalized current density as a function of basewidth for both graded- and abrupt-junction HBTs operating at $V_{BE} = 0.8V_{bi}$ and $V_{BC} = 0$ [10]. For both devices the current is normalized to its ballistic value.

Figure 8: Contributions to the collector current of the concentration-gradient term (dashed line) and the energy-gradient term (stippled line), as specified by the second and third terms, respectively, in Eqn. (11). The simulations are for $W_B = 1l_{sc}$ and $J_C \approx 50 \text{ Acm}^{-2}$. (a) Graded-junction HBT. (b) Abrupt-junction HBT. [Data courtesy of A. St.Denis]

Figure 9: Incoming and outgoing fluxes at the junctions to the base of a graded-emitter HBT [6]. Here, $v^-(0) = 2v_R$, as is appropriate for a hemi-Maxwellian distribution. Note: $W \equiv W_B$.

Figure 10: Normalized collector current as a function of normalized basewidth for a graded-emitter HBT, from Eqn. (21) for different values of v_{coll} . The curves are for operating conditions of $V_{BE} = 0.8V_{bi}$ and $V_{BC} = 0$. The circles are experimental data, and are plotted using $l_{sc} = 52.5 \text{ nm}$ [21]. This data has been scaled to coincide with the predicted curves at the long-basewidth limit.

Figure 11: Normalized collector current as a function of normalized basewidth for an abrupt-emitter HBT. The curves are for operating conditions of $V_{BE} = 0.8V_{bi}$ and $V_{BC} = 0$. The data points are experimental results [22], and are plotted using the same mean-free-path length as in Fig. 10. The circles refer to high-gain devices (TA series) and the squares to normal-gain devices (TB series). The data has been extracted from information given in Ref. [22], and corresponds to operation at the same bias conditions as used for the model calculations. This data has been scaled so that the current for the thinnest high-gain device coincides with the current predicted by the microscopic model.

Figure 12: Hybrid- π equivalent circuit for an HBT [35]. The total collector-base junction capacitance is given by $C_{jc} = \sum_{i=1}^n C_i$, and, for operation in the active mode, is equal to C_μ , as introduced following Eqn.(25).

Figure 13: (a) Conceptual partition of an HBT into quasi-neutral and space-charge regions (shaded). (b) Sample plots of the local charge changes, and (c) corresponding plots of the integrated charge changes, versus position z for $J_C = 3 \times 10^4 \text{ Acm}^{-2}$. For (b) and (c) the MEDICI simulations were for a graded-emitter device with the metallurgical base extending from $0.4 - 0.5 \mu\text{m}$ and a sub-collector commencing at $0.9 \mu\text{m}$ [20].

Figure 14: (a) Emitter delay τ_E as a function of J_C for the graded-emitter HBT. (b) Profiles of cumulative delay in the τ_E region, for both graded and abrupt devices [20].

Figure 15: Base delay τ_B as a function of quasi-neutral basewidth W_B [20]. The curves are from Eqn. (30) for various values of v_{coll} . Note that $v_T \equiv 2v_R$.

Figure 16: Base delay as a function of Al mole fraction at the emitter end of the base for a linearly graded-base, abrupt-junction HBT. The results are from Eqn. (32) for various values of v_{coll} . $W_B = 1l_{sc}$, $D_n = 75 \text{ cm}^2\text{s}^{-1}$ [21], $V_{BE} = 0.8V_{bi}$.

Figure 17: Collector delay as a function of J_C for the graded-emitter device [20].

Figure 18: (a) Schematic cross-section of a mesa-type HBT, identifying the components of base resistance and collector capacitance. (b) Reduced equivalent circuit for the base-collector network [35].

Figure 19: Equivalent circuit for large-signal simulations. The only components not described in the text are the diodes IRC and IRE: these represent space-charge-region recombination currents [39].