J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

TRANSISTORS, BIPOLAR JUNCTION

Based on previous work on silicon *p–n* junctions, the junction transistor was developed during the 1940s by a research group at the Bell Telephone Laboratories. The first *n–p–n* transistor was invented in January 1948 by William Shockley, and was basically a germanium device with two *p–n* junctions separated by a thin layer of *p* -type semiconductor to control the current flow between the *n* -type regions. Over the years after its invention, the bipolar junction transistor (BJT) was progressively developed, with higher voltage handling and current capability, thus becoming in certain applications a replacement of the vacuum tube. Since then the bipolar transistor has been in competition with other solid-state devices invented in the same period or later, which have acquired specific niches of application (1). Today, the bipolar transistor is produced either as a discrete component or as the elementary cell in an integrated circuit, where a single chip can contain up to 10 million transistors.

A general distinction holds in the field of discrete components between BJTs for signal applications at low power, and power BJTs. From the standpoint of behavior, transistors are classified as linear and switching, for use respectively as amplifiers and as on–off devices. The end of the 1980s represented the peak of the great research development on power BJTs; in that period there appeared in the field of power electronics competitive devices that caused a slow decline of interest on BJTs. Power bipolar transistors are today available on the market with the following ratings: single devices, with voltage up to 2000 V, current up to 250 A, and switching speed as high as 40 kHz; monolithic multiple connections of BJTs (Darlington configuration) with respectively 2000 V, 400 A, and 30 kHz. The forward voltage drop ranges between 0.5 V and 1.5 V. Over the 1990s the annual world sales of power BJTs has been stable at about US $1.5 billion.

Basic BJT Operations

The BJT is a three-terminal device with two junctions, the base–collector (BC) junction, and the base–emitter (BE) junction, in which the base current controls the collector current. In Fig. 1 are shown the two different types of BJTs: the *n–p–n*, and the *p–n–p*. In switching applications the common-emitter connection is generally used, which means the emitter is in both the input and the output path. Depending on the polarization of the BC and BE junctions, the BJT can work in four different regions of operations: (1) the forward active region (amplifier), (2) the saturation region (switch on), (3) the cutoff region (switch off), and (4) the reverse active region. In Fig. 2 are reported the bias conditions of the junctions, and accordingly the operating regions on the static *I–V* characteristic of a BJT.

In an *n–p–n* transistor, the current supplied to the base terminal causes the injection of electrons into the base region, which come from the emitter and reach the collector. The following relation holds between the terminal currents:

$$I_E = I_C + I_B \qquad (1)$$

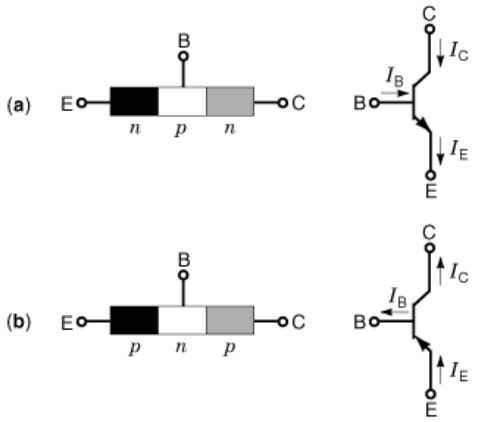

Fig. 1. Schematic representation of the bipolar junction transistor: (a) *n–p–n* and (b) *p–n–p* junctions, and the corresponding electrical symbols.

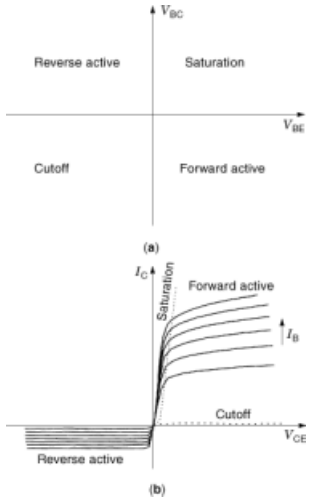

Fig. 2. Operating regions of a biased *n–p–n* BJT: (a) representation of the bias conditions (positive or negative) for both the BC and BE junctions, and related device behavior; (b) static *I–V* characteristics, showing the cutoff, forward active, saturation, and reverse active region.

The current gain is defined as

$$\beta = \frac{I_C}{I_B} \tag{2}$$

By combining Eqs. (1) and (2) we obtain

$$I_C = \frac{\beta}{\beta + 1} I_E \tag{3}$$

Two major quantities affect the current gain β of the transistor: the injection efficiency γ of the emitter (defined as the ratio between the injected electrons from the emitter region and the injected holes from the base region, in the forward-biased BE junction condition), and the base transport factor α_T (or current gain in the common-base connection, defined as the fraction of the total current flowing through the BE junction that reaches the BC junction). In power transistors the current gain β is typically in the range 10 to 50, which means a continuous feeding of high base current in steady-state conditions. On the other hand, at high injection the on-state resistance is noticeably lowered by the augmented density of electrons in the drift region. In those facts lie respectively the main disadvantage and the main advantage of the BJT devices.

Structure and Fabrication Technology

Power BJTs must be designed to withstand high voltage, current, and power. For many applications ruggedness is required, that is, the availability of wide operating areas in forward or reverse bias driving conditions. In addition, a good switching speed is often required. Three technologies are today used to produce power BJTs: epitaxial, multilayer, and planar, with resulting cross sections respectively shown in Fig. 3(a), 3(b), and 3(c). Epitaxial base transistors are characterized by a homogenous, lightly doped, wide base. The main advantage of such devices is their ruggedness; however, they switch slowly and are unable to sustain high voltages. Typical applications are in linear amplifiers and complementary pairs (push–pull), with V_{CE0} (the maximum sustained collector–emitter voltage with an open-circuited base) less than 200 V. Multilayer base (mesa) transistors have both an epitaxial and a graded base. The collector–base depletion layer may extend on both sides of the BC junction, thus allowing free choice of the base and collector widths. They achieve good performances in blocking voltage as well as in switching time, but their manufacturing process is more complicated; moreover, the integration of more than one transistor in the same chip is very difficult. Typical applications are motor control, inverters, resonant converters, and TV deflection circuits, with V_{CB0} up to 2000 V.

Graded base (planar) transistors offer a better tradeoff between blocking voltage and switching time. An actual planar device with 800 V, 5 A (BUL138 type by STMicroelectronics) is shown in Fig. 4. The comb structure shown is used to maximize the ratio of emitter perimeter to emitter area, so as to reduce the crowding of the current under the emitter fingers, as described in more details in the subsequent sections.

In Fig. 5 is depicted a schematic cross section along the line AB of Fig. 4, showing the many layers, that constitute the structure of commercial power BJTs. An n -type collector region N and the high-resistivity (lightly doped) N^- region are grown on the heavily doped substrate N^+ . The N layer, as will be explained in the next sections, is used to improve the reverse secondary breakdown capability of the power BJT. The N^- layer, which is called the drift region, is wide enough to support the breakdown voltage between the collector N^- and the diffused base P^+ . The heavily doped emitter region N^{++} shows the typical finger structure. The main advantage of this structure is that the voltage handling depends only on the collector characteristics. The P^- region is a deep ring used at the edge termination to improve the collector voltage capability.

The doping profiles of a graded-base BJT like that shown in Fig. 5, suitable for high voltage (1000 V), are reported in Fig. 6(a), while the typical thicknesses of the different layers and the doping levels are reported in Table 1.

Reverse Behavior: Off State

Ideal Junction. We can treat the voltage ratings of bipolar transistors starting from two main physical phenomena: (1) avalanche breakdown and (2) sustaining. Limiting factors are punchthrough, reachthrough, and junction curvature. An abrupt plane p–n junction is considered, with constant doping concentration N_A of acceptors in the p -type region and constant doping concentration N_D of donors in the n -type region. We will

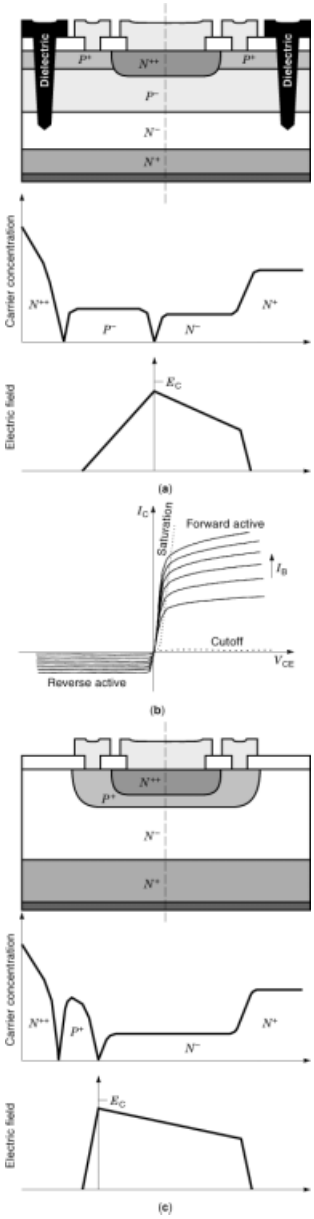

Fig. 3. Vertical cross sections of different power BJT structures: (a) epitaxial base; (b) multilayer base or mesa; (c) planar base.

initially assume unbounded of p - and n -type regions. When a reverse voltage V is applied, an electric field E is generated, which sweeps out free carriers near the junction, thus inducing a space charge region, or *depletion layer*, where ionized acceptors and donors are no longer compensated by mobile majority carriers (holes and

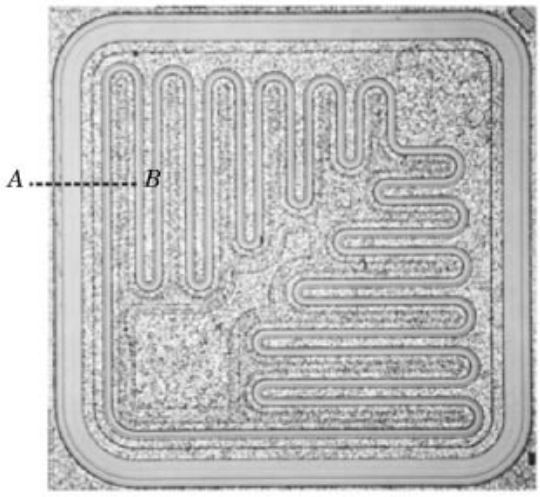

Fig. 4. Top view of an *n–p–n* BJT die, 800 V, 5 A (BUL138 type). Note the fingers, in the emitter region, that are used to avoid current crowding under the emitter. (Courtesy of STMicroelectronics.)

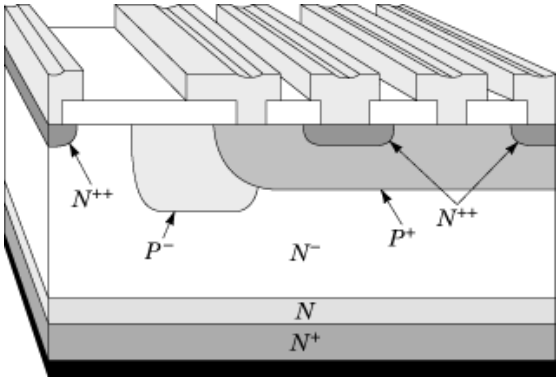

Fig. 5. Simplified cross section of the *n–p–n* BJT shown in Fig. 4 (line AB) having four main layers: substrate (N^+), collector (N), collector (drift region N^-), base (P^+) and edge termination (P^-), and emitter (N^{++}).

electrons respectively). For a plane junction the one-dimensional Poisson's equation applies:

$$\frac{dE}{dx} = -\frac{\rho(x)}{\epsilon} \tag{4}$$

where $\rho(x)$ is the charge density, and $\varepsilon = \varepsilon_0 \varepsilon_r$ is the silicon dielectric constant ($\varepsilon_r = 11.7$, and $\varepsilon_0 = 8.85 \times 10^{-14}$ F·cm^{-1}). The following two assumptions are made: (1) all the dopant atoms in the depletion layer are ionized, and (2) the neutrality condition is verified ($N_A x_{dP} = N_D x_{dN}$ where x_{dP} and x_{dN} are the extensions of the depletion layers respectively in the p - and n -type regions). In such a hypothesis, by properly applying Eq.

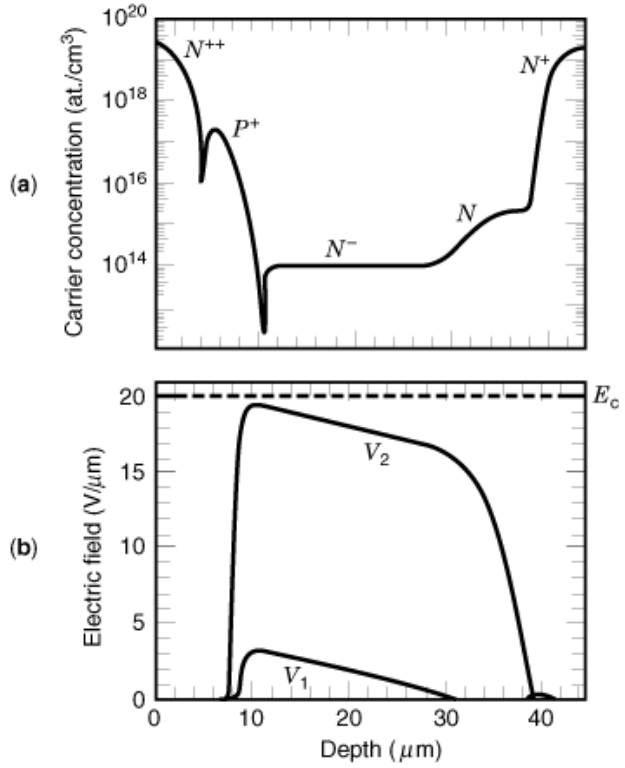

Fig. 6. (a) Doping profiles in a vertical section under the emitter for n–p–n BJT shown in Fig. 5: the base and emitter have graded densities, while the collectors have constant density. (b) Printout of a simulation to evaluate the electric field in the base–collector region for two different bias voltages ($V_2 > V_1$).

Table 1. Typical Thickness and Doping Levels for High-Voltage Planar BJTs

Layer	Thickness (μm)	Resistivity ($\Omega \cdot$ cm)	Peak Doping Level (cm^{-3})
N (collector)	5–15	5–10	
N^- (collector)	80–100	50–80	
P^+ (base)	10–30		10^{17}–10^{18}
N^{++} (emitter)	5–15		10^{19}–10^{20}

(4) we obtain

$$\frac{dE}{dx} = \frac{qN}{\epsilon} \qquad (5)$$

By integration of Eq. (5) we obtain

$$V = \int E\,dx \tag{6}$$

A simulation printout yielded by the application of Eqs. (5)–(6) to a BJT having the doping profiles of Fig. 6(a) is reported in Fig. 6(b) for two applied voltages $V_2 > V_1$.

The reverse breakdown of a $p-n$ junction may be ascribed to one of two phenomena: the Zener effect and the avalanche effect. At very high electric fields ($E > 10^5$ V·cm^{-1}) electrons can be excited from valence to conduction band through the depletion layer by tunneling. Zener breakdown voltages are below about 5 V and only concern junctions heavily doped on both sides. The energy bandgap E_g in silicon decreases with increasing temperature, so the breakdown voltage due to the tunneling effect has a negative temperature coefficient (2).

In the case of the avalanche effect, free carriers are accelerated by the field E . Between two consecutive collisions with the silicon lattice atoms, there is a finite probability for such carriers to acquire energy equal to or greater than the silicon ionization energy, thus creating an electron–hole pair (impact ionization). The electron multiplication factor M_n is defined as the ratio of the electron current after to that before crossing the depletion layer. At some value of the junction electric field ($E = E_{\rm cr}$) impact ionization strongly increases, and the reverse junction current becomes unlimited. In such a case, the multiplication factor M_n greatly increases ($M_n \to \infty$) and the breakdown is caused by high-energy electrons (a similar multiplication factor $M_{\rm p}$ is defined for holes). This situation, which appears as an abrupt increase of the reverse current, is known as avalanche breakdown.

The avalanche breakdown voltage increases with increasing temperature. A simple explanation of this phenomenon can be given. Hot carriers accelerated by high electric field in the depletion layer lose a part of their energy colliding with crystal lattice (by exchanging optical phonons), and the amount of the lost energy increases with the lattice temperature. Consequently, at higher temperature a higher field is required before carriers acquire enough energy to produce an electron–hole pair, and the voltage that generates the critical electric field (avalanche breakdown voltage) increases. The breakdown voltage variation with temperature is generally used to distinguish tunneling from the avalanche breakdown mechanism.

In BJTs the base concentration [Fig. 6(a)] is much greater than the doping level on the collector. The BC junction depletion layer is approximately equal to the collector width $X_{\rm C}$ (reachthrough phenomenon); thus the breakdown voltage $\mathrm{BV_{CBO}}$ (the maximum collector–base voltage with open-circuited emitter) can be calculated by integration of Eq. (4), with the proper boundary conditions, as

$$\mathrm{BV_{CB0}} \approx E_{\rm cr} X_{\rm C} - \frac{qN}{2\epsilon} X_{\rm C}^2 \tag{7}$$

In Eq. (7) $E_{\rm cr}$ is dependent on the dopant concentrations, which determine the slope of E starting from the maximum value E_0 located at the junction. Punchthrough is a possible condition for transistor structures with a homogeneous and low-doped base. The depletion region extends in this case mainly in the base; for low values of the base width $X_{\rm B}$ and low base dopant concentration $N_{\rm A}$, the depletion layer may touch the emitter at some value of the junction field $E_{\rm PT} < E_{\rm cr}$. In such a case an emitter–collector short (via the depleted base) occurs, thus allowing unlimited current flow between the collector and emitter regions.

In $\mathrm{BV_{CE0}}$ polarization the base terminal is disconnected ($I_{\rm B} = 0$) and the BE junction is forward biased; thus the electron current injected by the emitter contact, $I_{n\rm E}$, is reduced to $\alpha_{\rm T}\ I_{n\rm E}$ at the depletion layer boundaries, and a hole current $(1 - \alpha_{\rm T})I_{n\rm E}$ is injected from the base region. Electrons crossing the depletion region create electron–hole pairs by impact ionization, which determine an electron current $M_n \alpha_{\rm T}\ I_{n\rm E}$ at the lower end of the depletion layer, and a hole current $\alpha_{\rm T}(M_n - 1)I_{n\rm E}$ at the upper end. The transistor can then

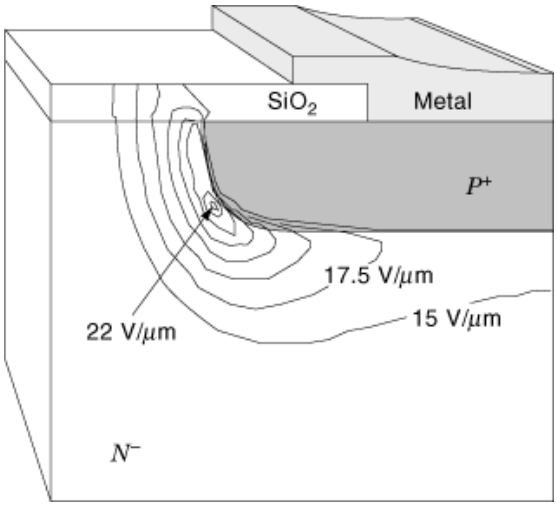

Fig. 7. Contour plot of the electric field, showing the crowding of the lines near the junction curvature. The high electric field in this region can cause premature breakdown.

sustain a current between collector and emitter if

$$\alpha_T(M_n - 1)I_{nE} = (1 - \alpha_T)I_{nE}, \qquad \text{which gives} \quad M_n = \frac{1}{\alpha_T} \quad (8)$$

and in this condition the sustained electric field E_S will be lower than the critical electric field E_{cr} . The breakdown voltage between the collector and emitter will be lower than the breakdown voltage between the collector and base ($BV_{CE0} < BV_{CB0}$) (3). In the case of the base short-circuited to the emitter and collector–emitter voltage V_{CES}, sustaining does not normally occur because the BE junction cannot be forward biased; thus $BV_{CES} = BV_{CB0}$. However, at high avalanche current levels, a voltage drop due to the nonzero value of the base resistance under the emitter can forward-bias the BE junction, thus producing snapback of the BV_{CES} curve toward lower voltages.

Planar Junction. In case the transistor BC junction is planar, edge effects have to be considered; in the curved region the external N^- layer (collector) is wider than the P^+ layer (base), but the overall charge has to be zero. As a consequence, the depletion layer in the N^- side must extend less in the curved region than in the plane region, and equipotential lines are more closely spaced; thus the electric field is higher, so that E_{cr} is reached at a voltage lower than the value of BV_P for the ideal junction.

Edge effects strongly degrade the reverse performance of devices. Figure 7 shows the crowding of the electric field in a planar junction, and in the case shown a premature breakdown was observed at a voltage 300 V lower than $BV_P = 800$ V of the ideal junction.

In BJT design, high-efficiency edge structures are commonly used, and the avalanche breakdown voltage of a BC junction can be substantially augmented, for example, by the use of field plates, a field-limiting ring, a variation of lateral doping (VLD) ring, or a deep ring (4). These techniques allow one to increase the ratio of actual to theoretical breakdown voltage almost to 1. The field plate technique remains one of the most useful approaches to the improvement of the avalanche breakdown characteristics of planar junctions. It consists in extending the junction contact metallization behind the P^+–N^- junction (5). The field-limiting ring consists of a diffused region having the same polarity as the heavily doped region of the reverse-biased junction, but electrically floating. The distance between the ring and the main junction is small enough to prevent the

merging of depletion regions of two (or more) junctions prior to avalanche breakdown. This technique allows an appreciable reduction of the curvature of the depletion region (6). Edge structure design is based mainly on the idea of modifying the charge distribution in edge regions in order to increase the equipotential line spacing. This is usually done by using ion implantation to create a localized low-doped p region (ring), which extends the junction termination and depletes under reverse bias (7, 8).

Today, the most effective solution is based on the VLD technique using aluminum, which is the fastest p dopant in silicon and produces a ring deeper than the main junction (9). In a VLD ring, doping and depth of implanted aluminum decrease toward the ring periphery. This is achieved by the variation of the effective dose introduced in silicon by opening small windows in the silicon oxide. Window dimensions become smaller and smaller toward the junction edge. A printout of a simulation showing the equipotential lines in an edge structure with a VLD ring combined with a metallic field plate is shown in Fig. 8(a), while in Fig. 8(b) the surface electric field on silicon in the ring lateral dimension is also reported. Efficient edge structures greatly broaden the equipotential lines in the ring area, and minimize the peak value of the surface electric field. Edge structure design is widely supported by computer device simulations: two- and three-dimensional solutions of Poisson's equation, with fixed doping distribution, allow one to optimize the reverse behavior, and nowadays planar technology is being extended to more and more high voltage devices.

Forward Behavior: on State

Current Capability. Consider the n–p–n vertical transistor shown in Fig. 9, with a forward voltage V_{BE} and a reverse voltage V_{BC} applied to the junctions. We assume that the current throughout the emitter and collector is perpendicular to the device surface. The voltage V_{BC} is assumed much lower than $\mathrm{BV_{CB0}}$, so that multiplication effects are negligible and the relation $\mathrm{M}_n = M_p \approx 1$ holds. The main current components involved in the transistor regions are the electron ($I_{n\mathrm{E}}$) and hole ($I_{p\mathrm{E}}$) currents crossing the forward-biased BE junction. The ratio between electron and hole currents is the injection efficiency:

$$\gamma = \frac{I_{n\mathrm{E}}}{I_{p\mathrm{E}}} \tag{9}$$

The electrons flowing in the base are partially recombined with holes, and a hole current $I_{p\tau}$ is injected into the base contact. Then, the current $I_{n\mathrm{E}}$ is reduced to $I_{n\mathrm{C}}$ in the collector, and the transport factor α_{T} is defined as

$$\alpha_{\mathrm{T}} = \frac{I_{n\mathrm{C}}}{I_{n\mathrm{E}}} = \frac{I_{n\mathrm{E}} - I_{p\tau}}{I_{n\mathrm{E}}} \tag{10}$$

Correspondingly, the currents at the base, collector, and emitter terminals are defined as depicted in Fig. 9. At high currents, when the high injection condition prevails, the current gain β of the device reduces to two terms, and is given by

$$\frac{1}{\beta} = \frac{1}{\beta_\tau} + \frac{1}{\beta_\gamma} \tag{11}$$

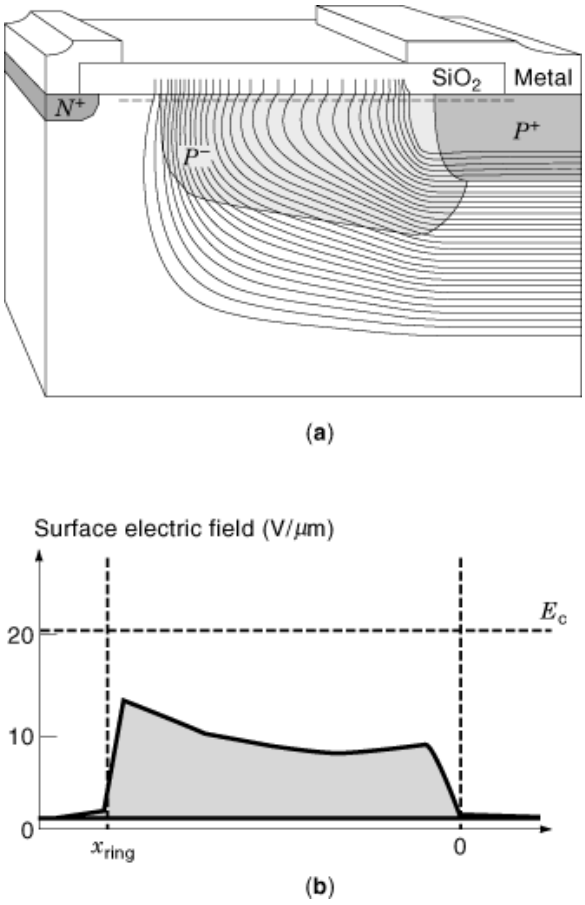

Fig. 8. The presence of an edge structure, with a deep variation of lateral doping (VLD) ring, creates quite a good distribution of the potential lines. (a) Cross section with a simulation plot of potentials; (b) simulation plot of the electric field on the surface against the lateral dimension of the ring.

where

$$\beta_\tau = \frac{4D_n\tau}{X^2} \tag{12}$$

and

$$\beta\gamma = \frac{4qD_n^2A_\mathrm{E}}{I_\mathrm{C}} \cdot \frac{Q_\mathrm{E}/D_\mathrm{E}}{X^2} \tag{13}$$

In Eqs. (12)–(13) τ is the base minority carrier lifetime; D_n is the electron diffusivity, which is constant in high-injection conditions; A_E is the emitter area; $Q_\mathrm{E}/D_\mathrm{E}$ is the emitter efficiency (or emitter Gummel number); and X is the effective base thickness. Figure 10 is a log–log plot of the current gain β versus the collector current for a typical power BJT. In Fig. 11 are shown the static output I–V characteristics of a power BJT for

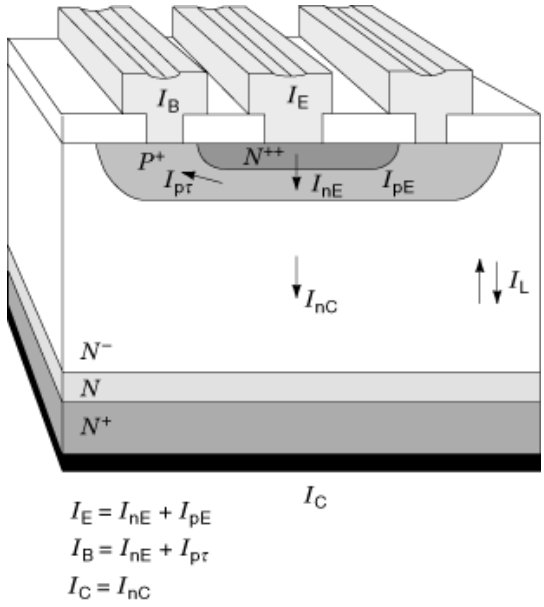

Fig. 9. Main current components in an *n–p–n* BJT shown in cross section: I_E, I_B and I_C are the terminal currents and are given by the contributions of one or more internal components; I_L is the leakage current with forward or reverse applied voltage.

several base current values; the knee current I_k, derived from Eqs. (11) (13), is the current capability, for a defined base current, at the boundary between the saturation and quasisaturation zones:

$$I_k = \frac{qD_nA_EQ_E/D_E}{X}\left[\sqrt{\left(\frac{X}{2\tau}\right)^2 + \frac{4I_B}{qA_EQ_E/D_E}} - \frac{X}{2\tau}\right] \quad (14)$$

The current capability I_k depends on the minority carrier lifetime τ, the emitter Gummel number Q_E/D_E, and the effective base thickness X . The emitter Gummel number is a key parameter in power BJTs, since its magnitude strongly influences the device current ratings. The emitter injection efficiency γ can be expressed as

$$\gamma = \frac{I_{nE}}{I_{pE}} = \frac{D_BQ_E}{D_EQ_B} = \frac{D_B\int_0^{X_E} N(x)\,dx}{D_E\int_0^{X_B} N(x)\,dx} \quad (15)$$

where D_B and D_E are the average values of the electron and hole diffusion constants, X_B is the base width, X_E is the emitter width, and Q_E and Q_B are the charges due to the ionized impurities respectively in the emitter and in the base regions. The base charge Q_B increases if either the base width or the base doping level is increased. The variation of Q_E is more complex: when the emitter doping level is lower than 1.85×10^{19} cm^{-3}, Q_E increases together with the emitter doping concentration; however, when the doping level exceeds that value, Q_E falls due to bandgap narrowing and Auger recombination (10). At high emitter doping the donor levels lying in the bandgap near the conduction band edge become degenerate; the effect is a lowering of the

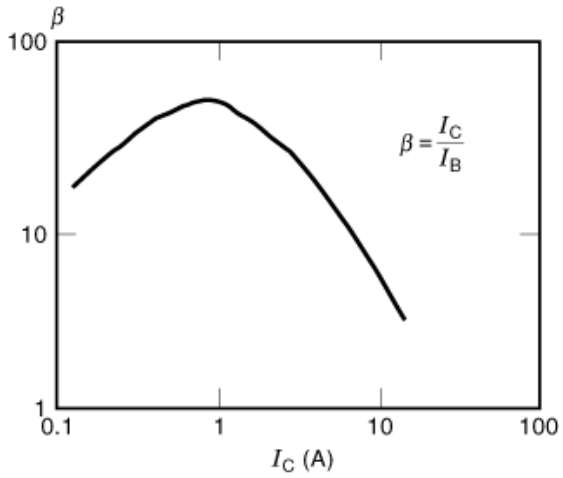

Fig. 10. The current gain β varies strongly as function of the collector current I_C . The trace shown refers to an *n–p–n* BJT, 600 V, 8 A.

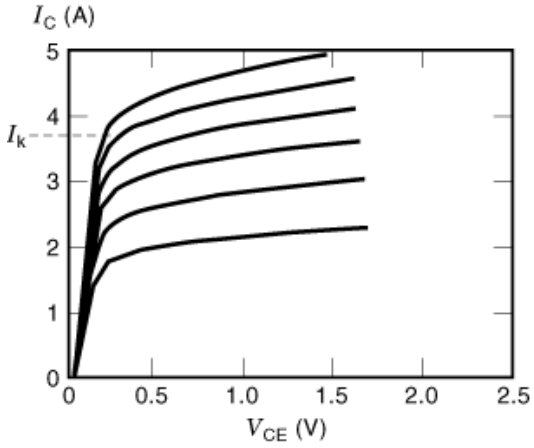

Fig. 11. Static output *I–V* characteristics of a BJT with the knee current on the curve at a constant I_B .

conduction band edge by some amount ΔE_g, known as *bandgap narrowing*. The quantity ΔE_g is a function of the emitter doping N_E and affects the product of the volumetric concentrations of holes and free electrons, which is no longer expressed by $pn = n^2_{i0}$, but by ${n_i}^2$ defined as ${n^2}_i = n^2_{i0} \exp(\Delta E_{g/kT})$.

The diffusion length L_{pE} of the holes in the emitter is governed by the minority carrier lifetime. In a heavily doped emitter, the minority carrier lifetime is limited by two recombination processes: a phonon-assisted indirect recombination process through deep impurity levels in the energy gap (Shockley–Read–Hall recombination, lifetime τ_{SRH}) and a direct band-to-band recombination process that transfers excess energy to mobile electrons (Auger recombination, lifetime τ_A) (11). The effective lifetime in the emitter is now $1/\tau = 1/\tau_{SRH} + 1/\tau_A$. For a doping level higher than 2×10^{19} cm^{-3}, the Auger lifetime falls off more rapidly because τ_A varies inversely with the squared ionized doping level, whereas τ_{SRH} varies approximately inversely with

the doping level. In this condition the diffusion length of holes is less than the emitter width:

$$L_{p\mathrm{E}} = (D_{p\mathrm{E}}\tau)^{1/2} < X_{\mathrm{E}} \qquad (16)$$

so that Q_{E} is further reduced because Auger recombination limits the integration width to $L_{p\mathrm{E}}$:

$$Q_{\mathrm{E}} = \int_0^{L_{p\mathrm{E}}} N(x) \exp\left(-\frac{\Delta E_{\mathrm{g}}}{kT}\right) \exp\left(-\frac{X}{L_{p\mathrm{E}}}\right) dx \qquad (17)$$

Accordingly, the effective value of Q_{E} changes with the emitter surface concentration and it has a maximum when the emitter surface concentration is about 2×10^{19} cm^{-3} .

Effective Base Thickness. In power BJTs the current gain is also affected by phenomena associated with the BC junction. These phenomena result in an enlargement of the base width at high current density, and in the diffusion of minority carriers through this extended base region, with a subsequent decrease of the current capability (12). When a forward current flows, resulting in transport from the emitter to the collector, $n(x)$ is the concentration of mobile electrons traversing the collector depletion layer; at high current density it is larger than the fixed donor density N_{D}, and the charge density is expressed by:

$$\rho(x) = q[N_{\mathrm{D}} - n(x)] \qquad (18)$$

At sufficiently high collector voltages, the electric field in the depletion region exceeds 10^4 V/cm and the carriers move at saturated velocity v_{s} (10); the collector current density is given by $J_{\mathrm{C}} = qv_{\mathrm{s}}\, n(x)$. As consequence Eq. (18) becomes

$$\rho(x) = q\left(N_{\mathrm{D}} - \frac{J_{\mathrm{C}}}{qv_{\mathrm{s}}}\right) \qquad (19)$$

It is this space charge that sustains the collector–base voltage. The Poisson equation for depletion region when the collector current is flowing can be written as

$$\frac{dE(x)}{dx} = \frac{q}{\epsilon}\left(N_{\mathrm{D}} - \frac{J_{\mathrm{C}}}{qv_{\mathrm{s}}}\right) \qquad (20)$$

The slope of the electric field depends now on the current density J_{C}, and in case it is zero the peak of the electric field is located at the BC junction (10). Equation (20) allows the calculation of the extended base width. At a current density $J_{\mathrm{C}} = J_0 = qv_{\mathrm{s}}\, N_{\mathrm{D}}$, the field becomes uniform, since its slope calculated by Eq. (20) is zero (Fig. 12). As soon as the collector current increases beyond J_0, the slope changes, and at a certain value the field becomes zero at $X = 0$. If V_{BC} is the voltage drop in the depleted collector region of width X_{N-}, then the electric field at the N^- N^+ interface is given by

$$E(X_{N^-}) = -\frac{2V_{\mathrm{BC}}}{X_{N^-}} \qquad (21)$$

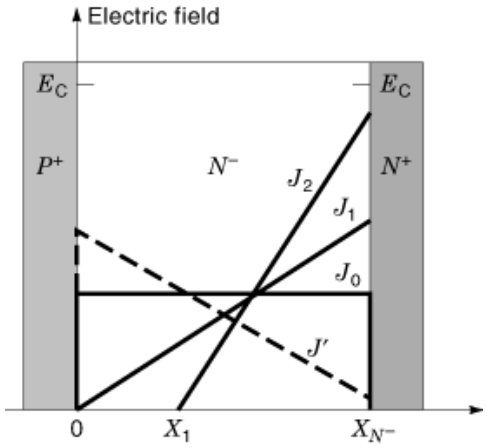

Fig. 12. Electric field profiles versus the collector width at different collector current densities and fixed collector voltage. The lines J_0, J_1 and J_2 ($J_0 < J_1 < J_2$) represent the electric field $E(x)$ as a function of the distance x from the P^+ border, at different injection levels, while J' is obtained without injection.

The current density giving this electric field is

$$J_1 = qv_s\left(N_D + \frac{2\epsilon V_{BC}}{X_{N^-}}\right) \tag{22}$$

As the collector current surpasses J_1, the slope dE/dx increases and the electric field extends through a reduced width ($x = X_N - X_1$) in the collector, since the area under the lines of the electric field in Fig. 12 is constant and equal to the voltage V_{BC} . The electric field at N^- N^+ interface is

$$E(X_{N^-}) = -\frac{2V_{BC}}{X_{N^-} - X_1} \tag{23}$$

where X_1 is the width of the neutral region representing the collector-induced base. The total base thickness now becomes $X = X_B + X_1$. The value of the base widening X_1 can be calculated by using Eqs. (20) (22), with a constant value V_{BC} . By evaluation of the areas under the electric field distribution curves in Fig. 12, we obtain

$$X_I = X_{N^-}\left[1-\left(\frac{J_1 - qv_sN_D}{J_C - qv_sN_D}\right)^{1/2}\right] = X_{N^-}\left[1-\left(\frac{J_1 - J_0}{J_C - J_0}\right)^{1/2}\right] \tag{24}$$

Output Characteristics. The current–voltage static output characteristic (at a fixed base current) of a high-voltage power BJT can be divided into three main zones having different slopes, as shown in Fig. 13(a). For a given base current, the device characteristic exhibits a saturation region (formed by the points between the origin and point 1), a quasisaturation region (between 1 and 3), and an active region (beyond 3). The current capability I_k and the output resistance R_{out} (defined as $R_{out} = V_{CE}/I_C$) depend on both the collector doping profile and the thickness. The BC junction is forward biased ($V_{BC} > 0$) over the saturation and quasisaturation regions. The voltage V_{BC} decreases from point 1 to point 3, where it becomes zero, and

beyond point 3 the base–collector junction is reverse biased, as the device is then in its active region. When the BJT is working in its active region, the density of electrons injected into the base region must be zero at the BC junction. At point 3, the collector series resistance depends on both the resistivity and the thickness of the two collector layers N^- and N ; the effective base thickness coincides with the metallurgical base width. The current capability I_k is given by Eq. (14), where we set $X = X_B$. Based on the above analysis, at point 3 the output resistance and the knee current are

$$R_{out} = R_p + \rho_{N^-}\frac{X_{N^-}}{A_E} + \rho_N\frac{X_N}{A_E} \tag{25}$$

$$I_k \propto \frac{1}{X_B} \tag{26}$$

where ρ_N- and ρ_N are the collector resistivities, and R_p is the parasitic series resistance, which includes the internal device resistance, lead, and connection resistances.

In case the device is operating in its quasisaturation region, at lower voltage V_{CE}, the BC junction becomes forward biased and electrons are injected into the base also from the collector. Moreover, holes are injected from the base into some part of the collector region, which now behaves as an extended base. In this extended base the quasineutrality condition can be applied. The mobile carrier density $n(x)$ in the collector space-charge region becomes larger than the fixed impurity density N_D, and the resistivity decreases rapidly. The base-widening phenomenon largely accounts for the fact that the effective base region may be defined as the quasineutral region where the emitter injects carriers. At point 2 the collector region N determines the collector series resistance, and the effective base is the physical base region plus the collector region [Fig. 13(b)]. The current capability I_k is still given by Eq. (14), where $X = X_B + X_{N-}$. According to the above consideration, at point 2 the output resistance and the knee current are

$$R_{out} = R_p + \rho_N\frac{X_N}{A_E} \tag{27}$$

$$I_k \propto \frac{1}{X_B + X_{N^-}} \tag{28}$$

At point 1, free carriers also invade the collector region N [Fig. 13(b)], and the collector voltage drop is zero, while the mobile carrier density $n(x)$ remains higher than the fixed impurity density in N . In such a condition, the collector is completely invaded and the effective base thickness is the physical base region plus the two collector regions N^- and N, which means $X = X_B + X_{N-} + X_N$, and the current capability is further decreased. Now, at point 1 the output resistance and the knee current are

$$R_{out} = R_p \tag{29}$$

$$I_k \propto \frac{1}{X_B + X_{N^-} + X_N} \tag{30}$$

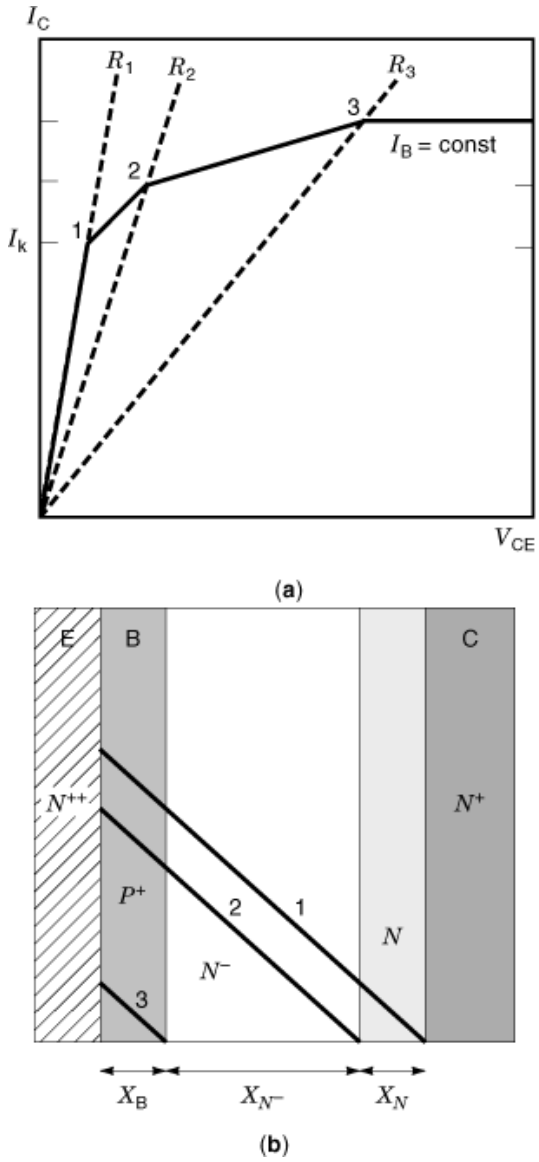

Fig. 13. Linearized static output *I–V* characteristic of a power BJT. (a) 0–1, saturation; 1–3, quasisaturation; beyond 3, active behavior. (b) Stored charge in the different regions under different saturation conditions.

The values of the output resistance previously determined are valid if the relationship between mobility and electric field is $E = \mu v$ and the mobility μ is considered constant. Actually, the electron mobility in the collector region decreases at high electric fields E_C, and this produces a resistivity variation modifying the voltage drop V_{CE} in the collector, and leading to quasisaturation. The static *I–V* characteristics obtained by the use of variable collector mobility show that the quasisaturation spreads out the R_3 line, especially at large collector current (13).

Temperature Dependence of Current Gain. The temperature dependence of the common-emitter current gain β can be explained by assuming that all the variation is due to the bandgap narrowing effect and the high-injection phenomenon (14). At low current density, the low-injection approximation can be used for the collector current, as the density of electrons injected into the p region is small compared to the density of majority carriers. In such a condition, there is negligible disturbance of the charge neutrality in this p region; thus the electrons move only by diffusion. In other words, the entire applied voltage appears across the junction, and the injected electrons in the p region cannot move by drift (i.e., the electron density is small compared with the concentration of holes, so $p \approx N_A$). The collector and base currents are

$$I_C = \frac{qA_E n_i^2(T)}{Q_B/D_B} \exp\left(\frac{qV_{BE}}{kT}\right) = I_{C0} \exp\left(\frac{qV_{BE} - E_g^B}{kT}\right) \quad (31)$$

$$I_B = \frac{qA_E n_i^2(T)}{Q_E/D_E} \exp\left(\frac{qV_{BE}}{kT}\right) = I_{B0} \exp\left(\frac{qV_{BE} - E_g^E}{kT}\right) \quad (32)$$

where $E_g{}^B$ and $E^E{}_g$ are the bandgaps respectively in the base and emitter regions. The current gain (at low current density) calculated from Eqs. (31)–(32) is

$$\beta = \beta_0 \exp(-\Delta E/kT) \quad (33)$$

where

$$\Delta E = E_g^B - E_g^E > 0 \quad (34)$$

At low current density the current gain β increases with increasing temperature, and this is due to the energy gap reduction of a heavily doped emitter [Fig. 14(a)]. For a moderately or lightly doped emitter (concentration below $\approx 5 \times 10^{19}$ cm^{-3}), ΔE is very small and the temperature effect on the current gain is reduced [Fig. 14(b)]. At high current density the electron concentration equals the acceptor density N_A ; thus charge neutrality is violated and $p \approx N_A + N$. In this condition, in order to maintain charge neutrality, the holes in the p region must have the same concentration gradient as the electrons. That produces an electric field in the direction of the flow of electrons (drift component). The voltage drop in the p region becomes comparable with the junction drop, so that it reduces the effective applied voltage (10). In this condition the base current is again given by Eq. (32), while the collector current is

$$I_C = I'_{C0} \exp\left(\frac{qV_{BE} - E_g^B}{2kT}\right) \quad (35)$$

Thus, for high current density the current gain β is

$$\beta = \beta_0 \exp\left(\frac{E_g^E - \Delta E - qV_{BE}}{2kT}\right) \quad (36)$$

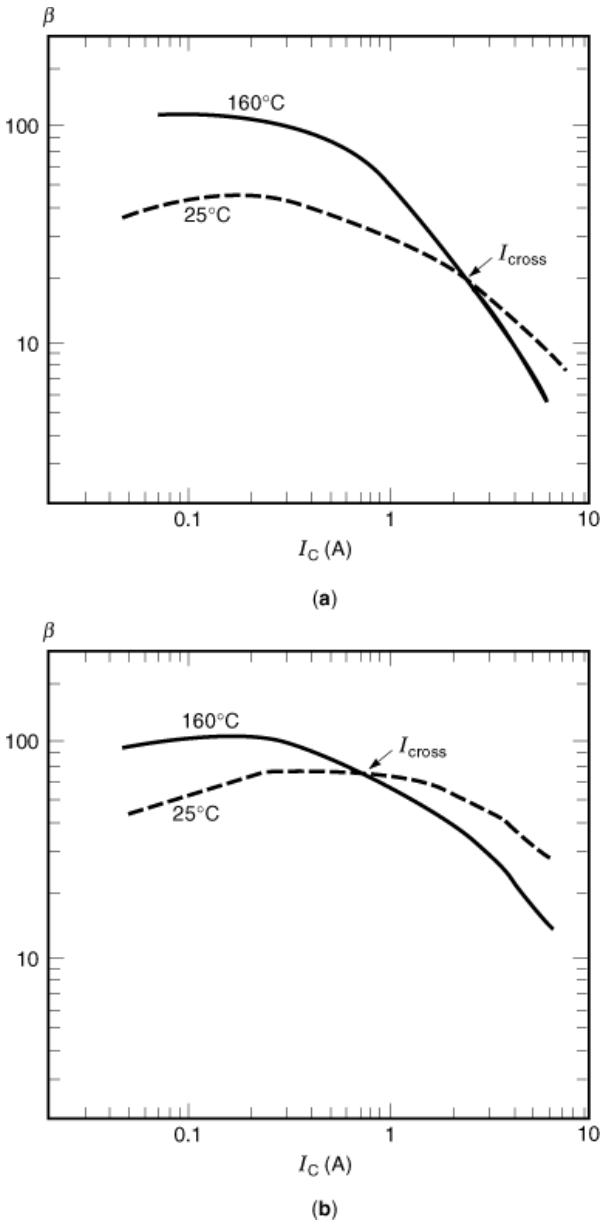

Fig. 14. Temperature dependence of current gain. (a) Device with heavily doped emitter (concentration greater than 10^{20} cm^{-3}). (b) Device with lightly doped emitter (concentration 2×10^{19} cm^{-3}). The value of the collector cross-current is marked in each case.

and in this case the temperature dependence of β is also influenced by the collector current value. For low to medium current in Eq. (36), we have $E^{E}{}_{g} - \Delta E > qV_{BE}$, and $d\beta/dT > 0$, while at high current $E^{E}{}_{g} - \Delta E < qV_{BE}$ and $d\beta/dT < 0$. As soon as $E^{E}{}_{g} - \Delta E = qV_{BE}$, the current gain does not depend on the temperature,

[Fig. 14(a)]. The value of the collector current at which $d\beta/dT = 0$ is lower when the emitter is lightly doped, because of the reduced value of ΔE [Fig. 14(b)].

Thermal Ratings

The main parameters of BJTs depend strongly on the temperature. Since circuit designers must specify all the electrical and mechanical parameters, a prerequisite is to know the significance of both the parameters and the other quantities determining the device thermal behavior: the maximum junction temperature, the thermal resistance, and the maximum power (15). Manufacturers specify two different maximum temperatures: the maximum storage temperature $T_{\text{smax}}(\approx 200^\circ\text{C})$, and the maximum junction temperature under working conditions, T_{jmax} (150° to 175°C). The former is the temperature a device cannot exceed while it is not operating. It only depends on the materials used for manufacturing the BJTs, and more specifically depends on the die-to-package soldering material. The temperature T_{jmax} is the value a device cannot exceed while it is operating, and is lower than the maximum storage temperature; manufacturers do not guarantee safe behavior of devices that are working behind the maximum junction temperature.

Thermal Resistance. By applying a constant power P_0 to a BJT, in steady state conditions a constant temperature gradient will be established between the chip and the case. The energy flow per unit time, calculated by the one-dimensional Fourier thermal equation, is given by

$$P_0 = \sigma_{\text{th}} A \frac{dT}{dx} \tag{37}$$

where σ_{th} is the thermal conductivity, and A is the section of the device orthogonal to the energy flow. Let us assume σ_{th} is not temperature-dependent, then by integration of the first-order differential equation in Eq. (37), and application of the boundary conditions, we obtain

$$T_{\text{j}} - T_{\text{c}} = P_0 \frac{L}{\sigma_{\text{th}} A} \tag{38}$$

where T_{j} and T_{c} are respectively the junction and case temperatures, and L is the width of the material interposed. The ratio between the junction-to-case temperature gradient and the applied power is defined as the junction-to-case thermal resistance:

$$R_{\text{thjc}} = \frac{T_{\text{j}} - T_{\text{c}}}{P_0} = \frac{L}{\sigma_{\text{th}} A} \tag{39}$$

Inspection of Eq. (39) allows us to establish an analog of Ohm's law, with a correspondence between thermal and elecctrical parameters, and thus to treat a thermal circuit like a corresponding electric one. Since in actual working conditions a device is mounted on a heat sink, it is useful to define the junction-to-ambient thermal resistance R_{thja}, which is the sum of two different contributions: the junction-to-case and the case-to-ambient thermal resistances:

$$R_{\text{thja}} = \frac{T_{\text{j}} - T_{\text{a}}}{P_0} = R_{\text{thjc}} + R_{\text{thca}} \tag{40}$$

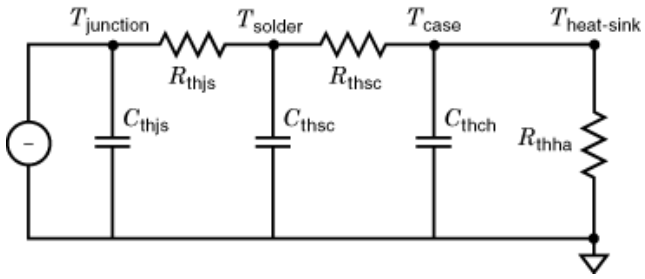

Fig. 15. Equivalent electric circuit for simulating the thermal behavior of power devices. This model assumes three layers.

However, to treat the transient behavior of the thermal flow we must also take heat storage into consideration by defining a thermal capacitance:

$$C_{\text{th}} = \frac{dQ}{dT} = \frac{P_0\,dt}{dT} \tag{41}$$

This concept leads in turn to that of the thermal impedance Z_{th}, valid for transient behavior, which replaces the thermal resistance R_{th} valid in steady-state conditions (16). In the case of an actual BJT in operating conditions with its package, a more complete model should include the four different layers (silicon die, solder preform, case, heat sink) with four different thermal resistances (R_{thjs}, R_{thsc}, R_{thch}, R_{thha}), and four thermal capacitances (C_{thjs}, C_{thsc}, C_{thch}, C_{thha}). However, the effect of the solder preform can be often neglected as a first approximation.

According to the simplified analogy between the thermal and electrical laws, the electrical equivalent scheme shown in Fig. 15, with a three-layer model, can be used to simulate the thermal behavior (3). The thermal parameter values are reported in the specification sheets supplied by the manufacturers, thus allowing circuit designers to make an equivalent electrical simulation of their specific application by using circuit simulators such as PSpice.

Maximum Power. Once the maximum junction temperature is fixed, the maximum applied power which can be transmitted to the ambient is given by

$$P_{\max} = \frac{T_{\text{j max}} - T_{\text{a}}}{R_{\text{thja}}} \tag{42}$$

where T_{jmax} is the highest temperature the BJT can survive at any point of its structure. In working condition the power loss in a bipolar transistor is proportional to the collector curprent density, the collector–base voltage, and the active emitter area:

$$P = A_{\text{E}} V_{\text{CB}} J_{\text{C}} \tag{43}$$

If the current density could be considered uniform within the chip structure, it would be easy to calculate of the die size needed to obtain a desired power capacity. Unfortunately, the distribution of the collector current density is dependent on the base–emitter voltage, as can be deduced from inspection of Eq. (31). The BJT designers must take account of unequal sharing of the current inside the device structure, and also of the temperature dependence of both V_{BE} and I_{C} . In particular, V_{BE} of Eq. (31) decreases about 2.3 mV/°C in the range of operating temperatures, while the variations of I_{C0} follow the variation of the density of intrinsic

carriers, N_i . By considering the BJT as composed of several elemental BJTs in parallel connection, from application of Eq. (31) it clearly appears that unequal sharing either of V_{BE} or of T_j will cause unequal distribution of the collector current density. Thus the maximum power that the transistor can supply must be decreased, since T_{jmax} can be locally reached at some internal hot point, thus generating a positive feedback between temperature and current, as will be described in more detail in the next sections. Uniformity of V_{BE} sharing within the inner structure is mainly dependent on the voltage drops across the base–emitter metallization or across the internal resistance. On the other hand, unequal distribution of T_j as a starting effect can arise if voids are present in soldering material or if the crystal structure has defects.

Safe Operating Areas

The safe operating area (SOA) method is a simple way to represent, in a plot of the collector current versus the collector–emitter voltage, the locus of points where the bipolar transistors can operate, over the short or the long term, without destructive failure. However, the maximum ratings for the current (I_{Cmax}), the breakdown voltage (BV_{CE0}), and the power (P_{max}) are not sufficient to determine an SOA. In fact, secondary breakdown can occur even within the area delimited by the above maximum ratings. Since several mechanisms can originate secondary breakdown, a distinction is made depending on the operating conditions:

- *Forward secondary breakdown* can occur when a BJT is driven into the on state. In this case, the base current is entering the base terminal, and the maximum collector current I_{SB} in the SOA reduces it to a smaller area, which is termed the forward bias safe operating area (FBSOA).
- *Reverse secondary breakdown* can occur when a BJT enters the off state with an inductive load. In this case it is the energy stored in the inductance that limits the SOA; the result is termed the reverse bias safe operating area (RBSOA), and the energy causing the secondary breakdown is denoted by E_{SB} .

Forward Secondary Breakdown. If a power pulse $P = I_C V_{CE}$ is applied to a BJT, a junction temperature increase ΔT will arise, which causes an increase ΔI_C of the collector current. The rise of the collector current will lead to an incremental power dissipation $\Delta P = \Delta I_C V_{CE}$. The BJT can dissipate an incremental power given by $\Delta P' = \Delta T/Z_{th}$; thus if $\Delta P' \geq \Delta P$, a thermal equilibrium can be established, while on the contrary if $\Delta P' < \Delta P$, a thermal runaway will start destroying the transistor if the power supply is not immediately switched off.

Taking into account for the above considerations, a stability factor S can be defined:

$$S = \left| \frac{\Delta P}{\Delta P'} \right| = Z_{th} V_{CE} \frac{\Delta I_C}{\Delta T} \tag{44}$$

A BJT is said to be *thermally stable* if $S \leq 1$. This definition is based on the assumption that the dissipated power density is uniform within the BJT structure and the temperature over the silicon-die–heat-sink interface is uniform. Actually, this is never the case, and when an *n–p–n* BJT is forward biased on the BE junction, most of the electron base current flows laterally under the emitter towards the base contact. Thus the current crowding effect appears (Fig. 16), that is, the current may be concentrated in a small area. A similar effect may take place due to lack of homogeneity in the emitter–base junction region: for example, the presence of pointlike crystal defects or base–emitter debiasing in inner zones, which can cause voltage drop along the emitter finger metallization.

A conventional approach used to model this phenomenon is to divide the power BJT into several elementary transistors connected in parallel (Fig. 17), and to define a local stability factor S_i for the ith device

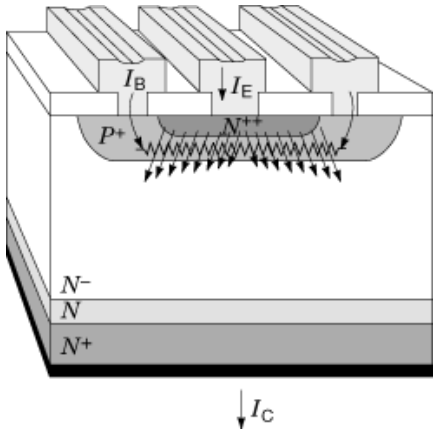

Fig. 16. Current density distribution under the emitter finger in the on state.

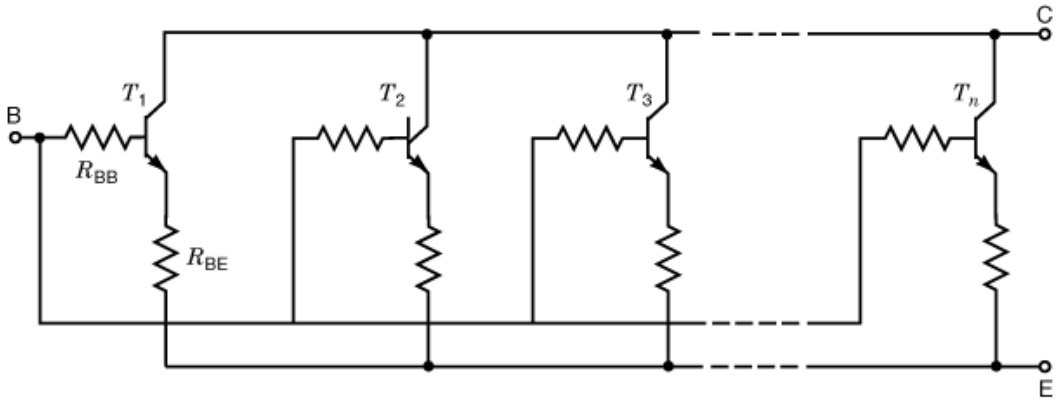

Fig. 17. Equivalent circuit of a bipolar transistor with multiple connection of several elementary devices. Also shown are the base and emitter ballast resistances.

$i = 1, \ldots, n$. In case any one of these local stability factors exceeds the unity, the BJT will fail. The calculation of the local stability factor S_i as a function of the physical and electrical parameters requires one to consider the collector current of the elementary transistor:

$$I_{Ci} = \mathrm{f}\,(T_i) = I_{COi}(T_i) \exp\left(\frac{q V_{BEi}}{k T_i}\right) \qquad (45)$$

If R_{BE} and R_{BB} are respectively the emitter and the base ballast resistances of the ith device, the voltage drop in the BE junction of this transistor can be calculated as a function of the applied base–emitter voltage V_{BB} (Fig. 18) by

$$V_{BEi} = V_{BB} - I_{Ci}\left(R_{BE} + \frac{R_{BB}}{\beta_i(T)}\right) \qquad (46)$$

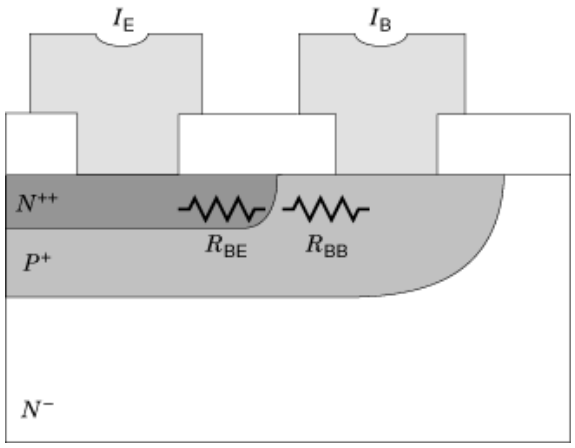

Fig. 18. Integrated ballast resistance. (Equations 48, 49) give the dependence of the ballast resistance on the physical and geometrical quantities.

where $\beta_i(T)$ is the current gain of the elementary transistor, which is considered dependent on the temperature. By combining Eqs. (44) (46) we obtain the expression for the ith stability factor:

$$S_i = \left(1 - \frac{T_c}{T_i}\right) \frac{\frac{kT}{q} \ln\left(\frac{I_{C0i}}{I_{Ci}}\right) + \left(I_{Ci} \frac{R_{BB}}{\beta_i(T)}\right)\left(\frac{\Delta E}{kT}\right)}{\frac{kT}{q} + I_{Ci}\left(R_{BE} + \frac{R_{BB}}{\beta_i(T)}\right)} \qquad (47)$$

Insertion of integrated emitter and base resistances is the easiest way to improve the thermal stability. In fact, in actual devices the power BJT emitters consist of multiple fingers to enhance the current density value. It is almost impossible to maintain uniform current distribution without the use of emitter resistive ballasting, thus providing each elementary emitter with a small series resistance. Such resistances are usually integrated within the device. The sheet resistance of the base (R_{sB}), and emitter (R_{sE}) layers, together with the emitter perimeter P_E, and the distance d_{cj} from the contacts to the junction, allows us to easily calculate R_{BE} and R_{BB} (Fig. 18) by

$$R_{BE} = \frac{d_{cj} R_{sE}}{P_E} \qquad (48)$$

$$R_{BB} = \frac{d_{cj} R_{sB}}{P_E} \qquad (49)$$

Looking at the diffusion process, a big improvement of thermal stability is reached by using a lightly doped emitter. In this case ΔE is reduced and the positive feedback between the collector current and the temperature diminishes at low current. Finally, any imperfections in the crystal or in the solder material, such as voids, will cause local current crowding and thus a local increase of thermal resistance, limiting the SOA.

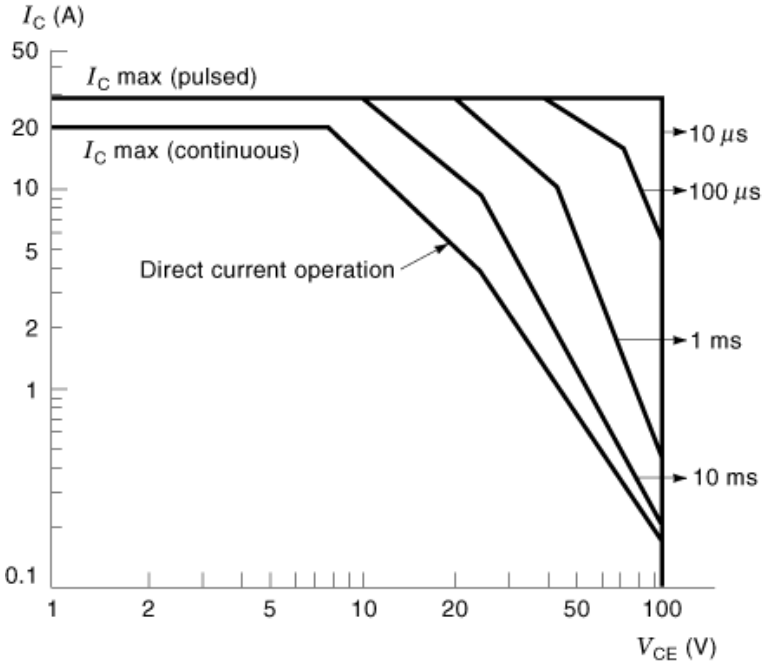

Fig. 19. Forward bias safe operating area (FBSOA) for a power transistor. The inner area is for dc operation; for pulsed operation the areas increase and the switching times decrease.

Equations (45) and (47) clarify the representation of the FBSOA; a typical plot is shown in Fig. 19. By inspection of Fig. 19, it is possible to distinguish at least three different operating conditions represented by:

- A first segment, where the collector current is constant, representing the maximum current which the emitter wire can withstand
- A second segment, where the power is constant
- A third segment, the locus of points where the stability factor equals one.

Very often curves for pulsed operation are also plotted, and the meaning of these curves can be easily understood if the concept of thermal impedance (appropriate to transient behavior) is used instead of thermal resistance (appropriate to steady state).

Reverse Secondary Breakdown. Forward-bias secondary breakdown is satisfactorily explained by thermal instability. In the case of secondary breakdown occurring when the BE junction is reverse biased, as during device turnoff with inductive loads, the initiation of failure is attributed to electrical instability and to the so-called secondary breakdown. The base current with reverse-biased emitter of an *n–p–n* BJT flows from the emitter towards the base contact: thus the BE junction becomes more forward biased in its inner regions than at its periphery, and the center injects more heavily (Fig. 20). Between the emitter and base terminals a reverse voltage V_B is applied, and a reverse current I_B flows within the base. This base current produces a voltage drop across the local base resistance r_B . The BE junction at the center of the emitter becomes forward biased, thus producing current injection: in turn, the active area dramatically decreases and the maximum value of the current density can exceed the average value given by the ratio I_C/A_E . During switching off with an inductive load, high values of the collector current I_C and collector emitter voltage V_{CE} will be simultaneously

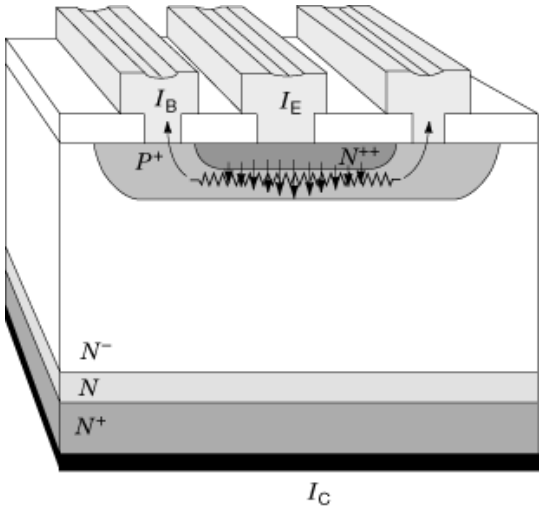

Fig. 20. Current density distribution under the emitter finger during the turnoff transient. The emitter center is affected by current crowding.

present. Usually V_{CE} is so high that the carrier velocity reaches the saturated value ($v_s \approx 8 \times 10^6$ cm/s for electrons and $v_s \approx 4 \times 10^6$ cm/s for holes), and then a traveling electron concentration will modify the concentration of the fixed donors in the collector region (assuming the case of a n–p–n transistor). Equation (20) integrated over the collector region gives

$$E(x) = E(0) + \frac{q}{\epsilon}\left(N_D - \frac{J_C}{q v_s}\right) x \qquad (50)$$

By setting a constant value for V_{CE}, integration of Eq. (50) gives

$$V_{CE} = \int_0^x E(x)\,dx \approx \text{const} \qquad (51)$$

When the collector current density reaches the value $J_C = J_0 = q v_s N_D$, the electric field constant, while for higher values ($J_C > J_0$) its maximum value is located at the N^-–N^+ interface (Fig. 12). This phenomenon is responsible for the base widening effect causing the current gain collapse at high current values. In particular, it leads to an increase of the electrical field at the bottom of the collector that can exceed, when a high voltage is applied, the critical value E_{crit} that causes the avalanche breakdown. Let J_{crit} be the collector current density at which E_{crit} is reached. It is worth noting that this current-induced avalanche breakdown can occur at any voltage V_{CE}, provided that J_C is sufficiently high. An approximate expression for the critical current density is (3)

$$J_{crit} = 2\epsilon v_s \frac{V_{CE}}{X_d^2} + q v_s N_D \qquad (52)$$

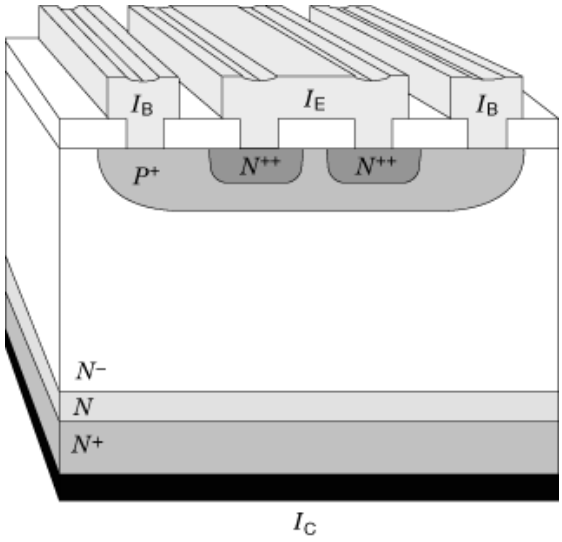

Fig. 21. Cross section of the hollow emitter structure of a BJT, used to limit the effect of current crowding under the emitter finger.

where $X_d = X_{N^-} - X_1$ is the depletion extension measured from the interface (Fig. 12).

It is very important to avoid current crowding due to debiasing of the centers of the emitter fingers, due to pinched base resistors (Fig. 20). Since the crowding diminishes the effective emitter area, BJTs are designed with special emitter shapes for examples, hollow emitters (Fig. 21) or cellular structures (17). With these emitter structures, the current concentration at the device center is impeded. Another advantage claimed for these emitter types is the reduction of the transistor fall time, because the base charges are stored closer to the base contacts and thus can be removed faster than charges at the emitter center. Also, the collector doping profile and thickness influence the robustness in switching conditions, so a good way to avoid device failure may be either to increase the collector thickness or to add one collector layer N with doping level intermediate between those of the collector and the substrate. Both the doping level and the thickness of this N region are adjusted to minimize the electric field at the N–N^+ interface, thus avoiding premature failure.

Switching Transients

When power BJTs are used as switches, it is desirable for them to have three main characteristics that ensure their effectiveness in working conditions: small voltage drop in the on state, very small current leakage in the off state, and very fast switching from on state to off state and vice versa (18). As the first two requirements have already been treated, the last one will be now taken into consideration. Although an inductive load is the most frequent actual case, first of all we will refer to a resistive load, since that case is helpful for describing the physical phenomena inside the silicon while avoiding some circuitry complications. Subsequently, the case of inductive load will be treated to describe the energy losses.

The basic circuit to determine the electrical switching times is shown in Fig. 22(a), and the relevant waveforms in Fig. 22(b). The electrical switching times are defined as follows:

- delay time t_d : from the application of I_{B1} to $I_C = 0.1\, I_{max}$;
- rise time t_r : from to $I_C = 0.1\, I_{max}$ to $I_C = 0.9\, I_{max}$;
- storage time t_s : from the application of I_{B2} to $I_C = 0.9\, I_{max}$;
- fall time t_f : from $I_C = 0.9\, I_{max}$ to $I_C = 0.1\, I_{max}$;

where $I_{max} = V_{CC}/R_1$. The total switch-on and switch-off times are defined as

$$t_{on} = t_d + t_r \tag{53}$$

$$t_{off} = t_s + t_f \tag{54}$$

To explain the switching transients from a physical point of view, we will refer to the charge-control model (19). The basis of this model is that at any instant of time the variation rate of the stored charges in the effective base X is equal to the difference between the external base current $I_{B(ex)}$ supplied and the base current needed to allow a certain collector current $I_C(t)$:

$$\frac{dQ_s}{dt} = I_{B(ex)} - \frac{I_C(t)}{\beta} \tag{55}$$

The time-dependent collector current will be

$$I_C(t) = \frac{Q_s}{\tau_F} = \frac{2D_B Q_s}{X^2} \tag{56}$$

where τ_F is the transit time of the carriers in the effective base. Taking into account these new concepts, the switching times can be defined as follows. Upon the application to the base of the turn-on current I_{B1}, both the collector current $I_C(t)$ and the charge Q_s increase, with a delay. The delay time t_d is the time to charge the BE junction from $V_{BE(off)}$ to $V_{BE} = 0$, and depends on the input capacitance. The rise time t_r is the time that the current requires (starting from the active region) to reach the quasisaturation region [i. e., $I_C(t) = I_0$ in Fig. 23, at which the base widening starts]:

$$t_r = \beta_f \frac{X_B^2}{2D_B} \tag{57}$$

where $\beta_f = I_{max}/I_{B1}$ is called as the forced current gain of the device (forced by the external circuit) and generally is much lower than β . The quasi rise time t_{qr} is the time required by the current $I_C(t)$ to reach the full saturation (from I_0 to I_{max} in Fig. 23):

$$t_{qr} = \beta_f \frac{(X_B + X_N + X_{N^-})^2}{4D_B} \tag{58}$$

In this condition the BJT is in the on state, and the effective base is completely invaded while the carrier concentration profile is given by line 1 in Fig. 13(b). As long as a negative base current I_{B2} is applied, the device undergoes the switch-off transient. Its behavior can be divided into three time intervals. In the first, the storage time t_s, the device is still in saturation and the collector current maintains the constant value I_{max} :

$$t_s = \beta_{sat} \frac{(X_B + X_N + X_{N^-})^2}{4D_B f(I_{B2}, I_0, \beta_f, \beta_{sat})} \tag{59}$$

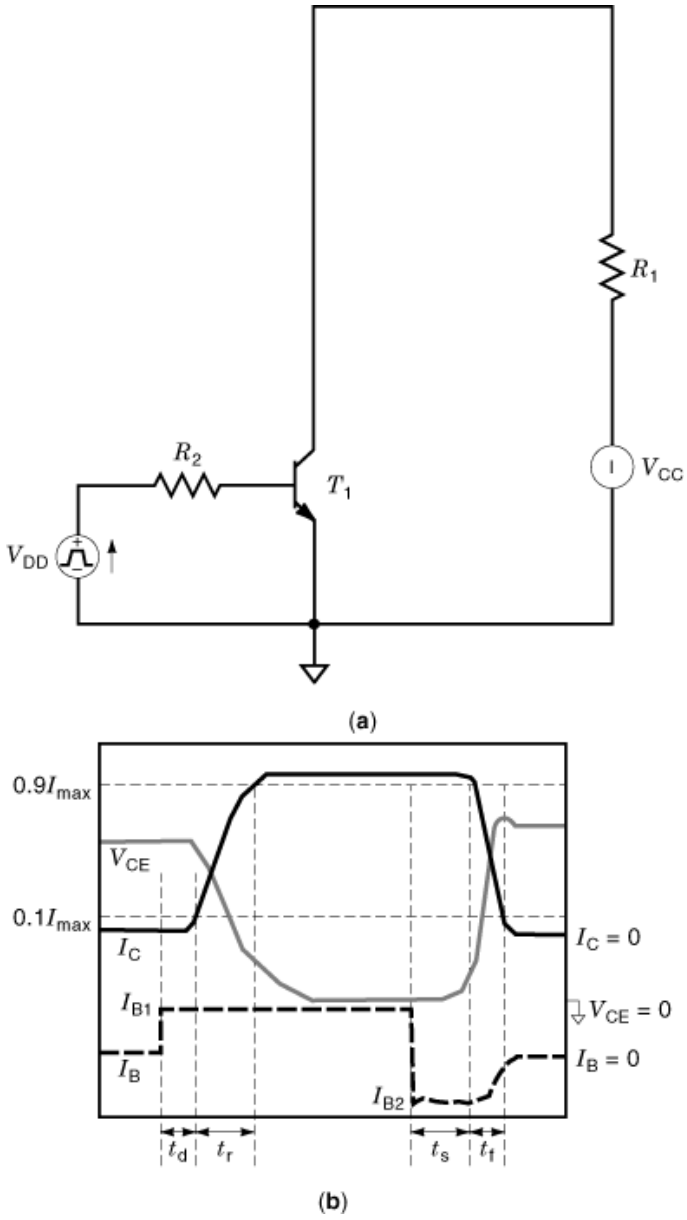

Fig. 22. Determination of the switching times of power BJTs for a resistive load: (a) electric circuit (b) voltage and current waveforms during an on–off cycle.

where $f(I_{B2}, I_0, \beta_f, \beta_{sat})$ takes into account the dependence on the driving condition and oversaturation. In the second time interval, the quasistorage time t_{qs}, the device is in its quasisaturation region. The collector current goes to I_0 from I_{max} ; the excess carriers present in the collector region are removed [in Fig. 13(b) this behavior means moving from line 1 to line 3, where the saturation phase is finished]. In case of hard base

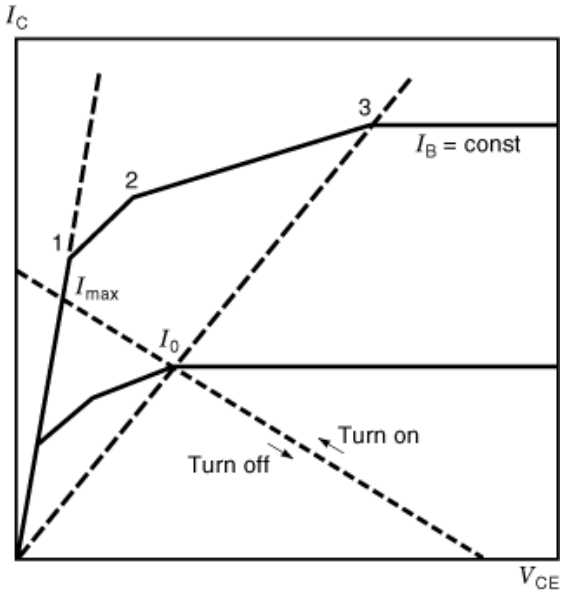

Fig. 23. Trajectory of the switching transient for a resistive load plotted in the *I–V* plane; the crossing of the load line with the output characteristic is also displayed.

driving conditions, i.e., $\beta_{sat} > I_0/I_{B2}$, the quasistorage time can be defined as

$$t_{qs} = \frac{I_0}{I_{B2}} \frac{(X_B + X_N + X_{N^-})^2}{4D_B} \qquad (60)$$

From inspection of Eqs. (55) 60, we can draw the conclusion that both low thickness of the base and collector regions and low saturated gain are required to reduce the storage and quasistorage times. Finally, in the last time interval the fall time t_f, the device is in its active region and the base region is completely swept out, and meanwhile the collector current is gradually reduced to zero:

$$t_f = \frac{I_0}{I_{B2}} \frac{X_B^2}{2D_B} \qquad (61)$$

An important role is played by the minority carrier lifetime; in particular, the off time increases at increasing carrier lifetime. The fall time is greatly affected by two-dimensional effects. The electron–hole plasma in saturation is supplied by the base contacts; therefore, a time is required for carriers to propagate into the emitter strips and to reach the central regions. During a turnoff transient, the carriers are swept out much faster at the edge than at the center of the emitter fingers. The time required to sweep out these carriers also leads to an increased fall time (much larger than the one expected from one-dimensional analysis). The influence of the dimensions of the emitter fingers on the fall time is also confirmed by measurements performed on devices with identical vertical structure, but with different strip widths: 300 μm and 150 μm. The storage and quasistorage times are identical, while t_f is half as long in the narrow-fingered device (20).

An additional cause of large fall time is the presence of large emitter areas used to bond the emitter wire. In fact, this region is several times the width of the emitter strips. The carriers stored under this zone are the

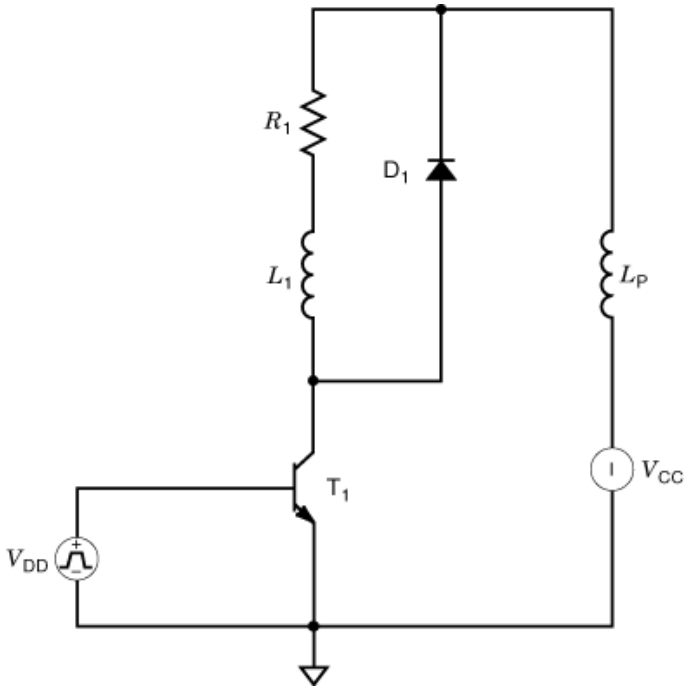

Fig. 24. Electric circuit for the determination of the switching times of power BJTs on inductive load, with a freewheeling diode across the inductor.

last to be removed during turnoff. To avoid this influence, special structures are used that avoid the presence of emitter areas under the zone dedicated to the wire bonding.

Energy Losses. To understand the quantities concerned in energy losses in a BJT when it is used as a switch on an inductive load, it is useful to analyze the typical test circuit drawn in Fig. 24, and also the current and voltage waveforms shown in Fig. 25(a) and 25(b). The main components of the circuit are a freewheeling diode D_1, the switching device T_1, an inductive load L_1 with its parasitic resistance R_1, and the parasitic inductance of the wires, L_p ; the energy supply is V_{CC}, and the base driving voltage is V_{DD} (21). Looking at the waveforms, it is possible to split the transient into three different regions with respect to losses: turn-on, turn-off, and conduction time. The respective energies can be calculated by

$$E_{on} = \int_0^{t_{on}} V_{CE} I_C \, dt \tag{62}$$

$$E_{off} = \int_0^{t_{off}} V_{CE} I_C \, dt \tag{63}$$

$$E_{con} = V_{CE(sat)} I_C t_{con} \tag{64}$$

At increasing switching frequencies E_{on} and E_{off} become the most important losses.

Energy Losses at Turn-on. First of all we will deal with the energy losses during the turn-on transient, focusing on Fig. 25(a). It is supposed that the current flowing through the inductance L_1 is different from zero at the instant when the base current is supplied. The time interval during the commutation with high values of current and voltage strongly contributes to the energy losses at turn-on. During this transient it is possible to distinguish three different zones on the waveforms:

- A first zone where the current rises with a slope $dI_C/dt \approx (V_{CC} - V_{CE})/L_p$ depending only on the circuit parameters, and the collector voltage remains unchanged at V_{CC} .
- A second zone where the collector current crosses its on-state value increasing with the previous slope. During this time interval (t_1) the voltage remains nearly unchanged at V_{CC} . After the t_1 elapses, the current reaches its peak value, and in a time interval t_2 goes to its on state value. The peak value of the current and t_2 depend on the characteristics of the freewheeling diode D_1—in particular, on its reverse recovery time t_{rr} and reverse recovery peak current $I_{RR(max)}$. Until the beginning of t_2, the collector voltage stays at about V_{CC}, starting to decrease after the current peak. The value of t_2 and the slope of the voltage during t_2 are imposed by D_1 .
- A third zone where the current reaches the on-state value and the voltage slowly decreases to $V_{CE(sat)}$ (dynamic saturation voltage).

It is interesting to note that the BJT energy losses only depend on the third zone, and the transistor parameters mainly affecting the losses are those linked to the forward conduction, heavy saturation, and quasisaturation. Thus high values of the current gain, low base and collector thickness, and high minority carrier lifetimes will produce lower dynamic $V_{CE(sat)}$ and less energy losses (22).

Energy Losses at Turn-off. To turn off the transistor all minority carriers must be removed from the effective base induced by the high current values, thus a reverse biasing of the base terminal is needed. In Fig. 25(b) is represented the switch-off transient. In a first time interval (the storage time), when the stored charges maintain the transistor at full saturation, the loss of energy is small, and the power loss can be neglected, especially at high frequencies. After this interval, the most important energy losses will take place. By inspection of Fig. 25(b) it is possible to distinguish at least three zones:

- A first zone where a voltage tail is present. All aspects of quasisaturation effects play a major role; thus high values of the current gain, low base and collector thickness, and high minority carrier lifetimes will produce a small voltage tail.
- A second zone where the current falls and the collector voltage increases towards V_{CC} . The slope of V_{CE} depends only on the transistor parasitic capacitances. For a time interval equal to the forward recovery time of D_1, V_{CE} can exceed V_{CC}, since the diode for a certain time t_{fr} exhibits a forward voltage V_{fp} greater than the modulated value V_f (18, 23).
- A third zone where a current tail is present. All processes involving the storage of carriers will cause current tail; thus an incorrect emitter layout or large thickness of base and collector will produce large energy losses. All the expedients useful to reduce the distance from the stored carriers to the recombination contact will help to eliminate or reduce the current tail; for example, emitter layouts like those described to improve the RBSOA are useful in this case too.

A correct understanding of the mechanism responsible for the voltage or current tails is important both for the transistor design and for the circuit design, so that by changing the circuit parameters one can modify the shape of the waveforms during switch-off. Three main different cases may happen during a switch-off, as depicted in Fig. 26:

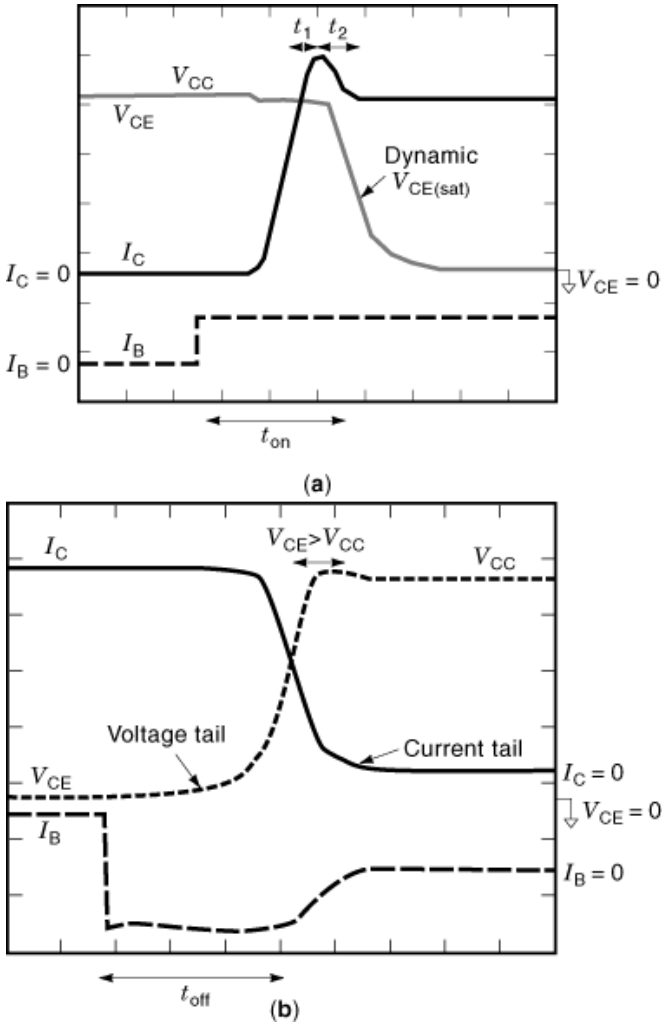

Fig. 25. Experimental traces of the voltage and current waveforms during a switching transient on an inductively loaded power BJT: (a) turn-on, (b) turn-off.

- The BC junction is switched off before the BE one [Fig. 26(a)]. The BJT goes out of saturation, and a voltage tail appears. Low minority carrier lifetimes and low current gain lead to large energy losses.
- The BE junction is switched off before the BC one [Fig. 26(b)]. The emitter current must then compensate the recovery of the BC junction, and a current tail will appear. Of course, long minority carrier lifetimes and long distances to recombination contact lead to large energy losses. The circuit designer must choose the correct ratio between $I_{B(on)}$ and $I_{B(off)}$ to set the corresponding circuits in the best condition. Sometime it is useful to insert an inductance in series with the base terminal in order to delay the current extraction from the base region.
- The BE and the BC junctions are switched off at the same time [Fig. 26(c)]. This is the ideal condition to minimize the energy losses; the transistor is in full saturation as long as possible, and the base–collector recovery time is as low as possible.

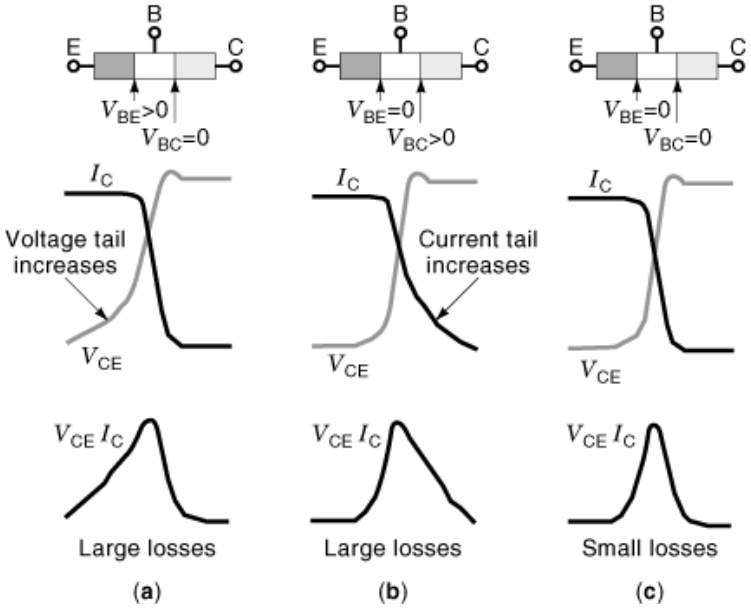

Fig. 26. Energy losses during a turn-off transient correlated with the bias conditions of the BE, and BC junctions: (a, and b) large energy losses due to the nonsimultaneous zero voltage bias of the BE and BC junctions; (c) good switching with small energy losses.

Assisted switching by auxiliary *RC* snubbing circuits can be used to reduce the power losses of the device; the aim of the capacitor is to reduce the voltage slope across the device during the switch-off. Thus losses are moved from the BJT to the resistor, which can dissipate energy more safely than a transistor (23).

On-state Losses. Only at low frequency does the energy loss in the on state constitute a significant component of the total losses; its value depends on the transistor's V_{CE} in the on state. The main aspects of on-state characteristics have already been reported in a previous section.

Multiple Connections of BJTs

Monolithic Darlington Configuration. The integration scale of power BJTs is very poor for traditional three-pin devices. The part count does not exceed about ten, except for hybrid power modules and the fast-growing smart technologies. Very often, inside a three-pin power BJT a pair of devices are integrated in the so-called monolithic Darlington (MD) configuration, Fig. 27. Looking at the electrical circuit, it is easy to find an expression for the gain current of the whole MD:

$$\beta = \frac{I_C}{I_{B1}} = \frac{\beta_1 I_{B1} + \beta_2(1+\beta_1) I_{B1}}{I_{B1}} = \beta_1 + \beta_2 + \beta_1\beta_2 \approx \beta_1\beta_2 \tag{65}$$

The effect of the resistors R_1 and R_2 drawn in Fig. 27 has not been taken into account. The aim of R_1 and R_2 is to avoid amplification of the base–emitter leakage current. These resistors can be designed very easily in

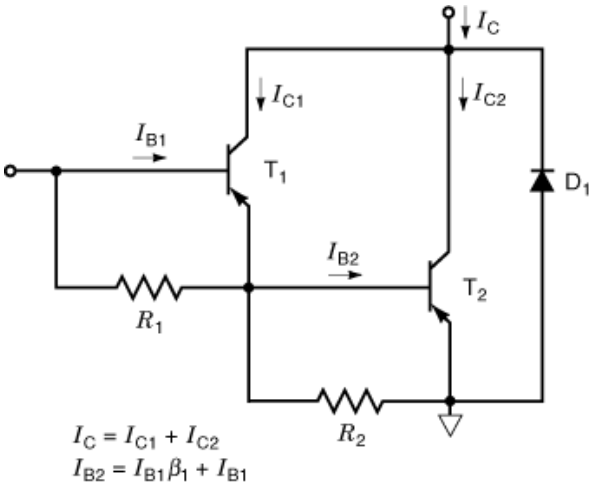

Fig. 27. Electric circuit of a Darlington connection. The antiparallel diode is integrated in a monolithic structure to allow reverse current flow in bridge applications.

several ways; for example, one can realize in the same design a resistor and a collector–emitter diode that can be used as a freewheeling diode in many applications (D_1 in Fig. 27).

As the frequency of the application increases, it is evident that a simple MD as previously considered cannot be used as a switch. In fact, whereas from the driver stage T_1 it is possible to remove minority carriers from the base terminal, from the final stage t_2 minority carriers must be removed by the resistor R_2 or by recombination in the base of t_2 . Since the lifetime of minority carriers is typically greater than 10 μs, for a faster MD it is necessary to integrate the speedup diode D_2 shown in Fig. 28(a). To avoid the introduction of more complicated technological steps, like the use of polysilicon layers, typically the diode D_2 is an *n–p–n* BJT with the BE junction taking the role of the diode D_2 [Fig. 28(b)]. The current gain of this parasitic transistor must be very low, because during the switch-off the minority carriers from the base of T_2 must be removed by the BE junction of such a BJT. A high value of the current gain of this BJT, besides being useless for carrier removal from T_2, would actually be dangerous, since it would conduct high current density at high collector emitter voltage. In practice, a deeper base with higher concentration is integrated inside the base of T_2 to form the base of this parasitic BJT, namely T_3 in Fig. 28(b). The same effect can be obtained by using a very low dopant concentration in the emitter of the parasitic transistor T_3 (24).

In some low-frequency applications it is possible to integrate into an MD a protection against reverse secondary breakdown, by means of a clamp diode D_Z connected between the base and collector of the output stage, as shown in Fig. 29. To understand the aim of this clamping diode it is useful to consider a typical switch-off on an inductive load. As soon as the MD is switched off, the energy stored in the coil will produce an increase of the voltage at the collector of the device, until the breakdown voltage of D_Z is reached. The current flowing through D_Z in breakdown condition will turn on T_2, thus changing the stress on T_2 from the reverse condition (outgoing base current) to the forward condition (incoming base current). This condition is much more favorable for a BJT. As in the previous case, to make the clamp integration as cheap as possible, D_Z is not actually a zener diode but a BJT (in the simplest case) or several BJTs in some other equivalent configuration, whose current gain is higher than those of both T_1 and T_2 . As a consequence the voltage BV_{CE0} of this transistor will be lower than the voltage BV_{CE0} of T_1, as well as of T_2 . All the previous solutions for MDs can be also integrated in three-stage configurations. However, the speedup diodes and the clamping diodes

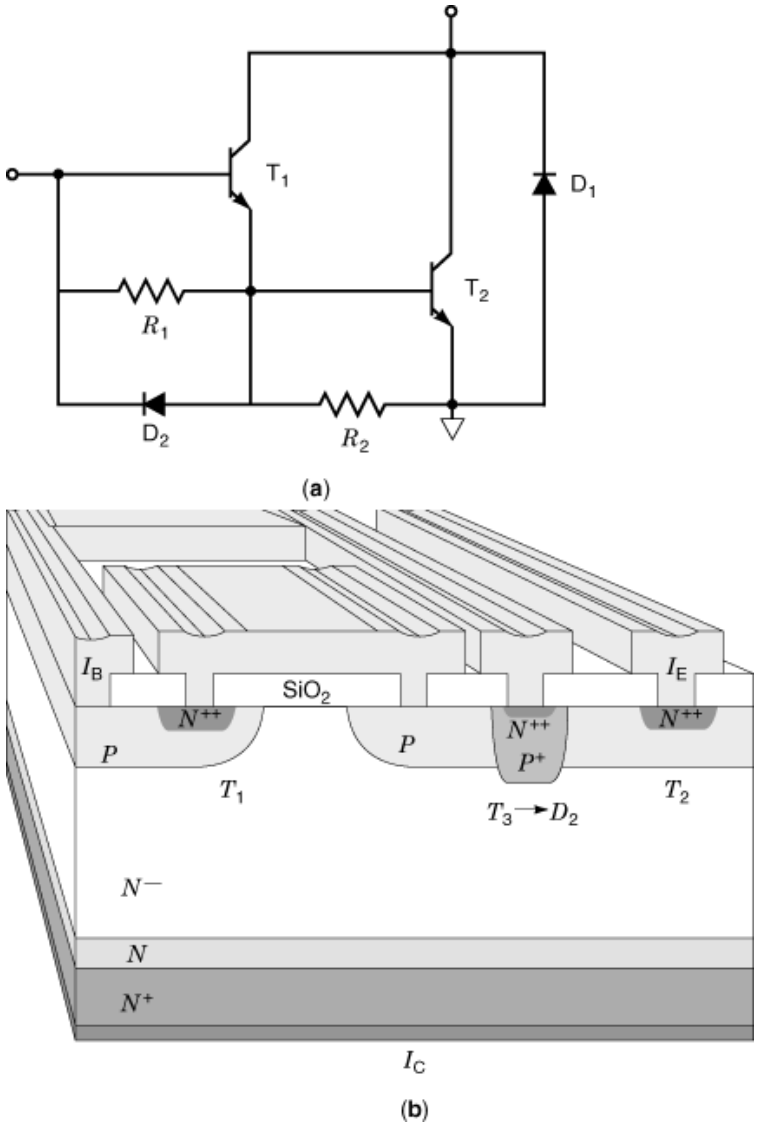

Fig. 28. Darlington configuration with an integrated speedup diode: (a) electric circuit, (b) cross section of an *n–p–n* BJT. In practice, the speedup diode is integrated as an *n–p–n* BJT with a current gain much lower than 1.

cannot be integrated at the same time, since during the breakdown of the clamping diode the current would flow through the speedup diode, thus undoing its benefits.

Series and Parallel Connections. Series and parallel connections of BJTs are used to augment respectively the voltage-handling and the current-carrying capability of single devices. In order to avoid destructive failure, unmatched series-connected devices, even if they have the same part count, must be protected against the voltage imbalance, due to the parameter spread, that appears in both static and dynamic conditions. A common solution is the connection of *RC* snubber circuits across each device, which slows down the switching transient and equalizes the static voltage, thus ensuring safe behavior. Parallel connections of BJTs are

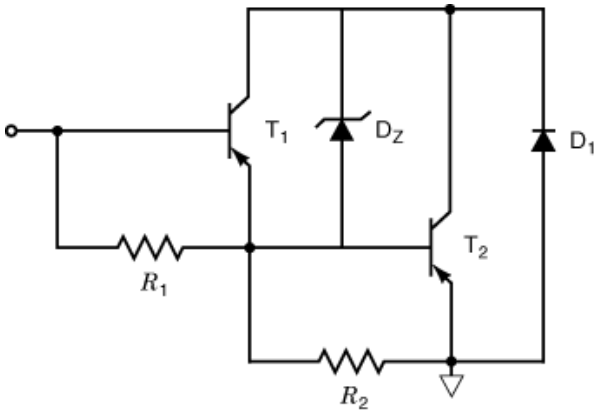

Fig. 29. Equivalent electric circuit of a Darlington configuration with an integrated collector–base clamp.

intrinsically unstable due to the negative temperature coefficient of the on-state voltage drop $V_{CE(sat)}$. In fact, if the devices share current at all unequally, the heating decreases the on-state resistances, so aggravating the initial current imbalance, and eventually causing the failure of the overcharged device. Resistors and coupled inductors connected in the emitter paths can ensure stability by the negative feedback on the base–emitter feed voltages.

Modeling and Simulation

The task of modeling solid-state electron devices can be faced by two different methodologies. Device designers use well-established methods based on physical equations that are applicable in the same form to all devices once the specific geometry, the boundaries, and the doping profiles are determined. Such a method, which requires sophisticated software tools and expensive hardware, is very time-consuming both for steady-state and for transient simulations. On the other side, it gives accurate information about the electric field and current flows inside the device structure, and moreover allows one to easily change the design parameters that are strictly tied to the production process. From the point of view of device design it is a valuable tool, especially for interior analysis.

The designers of power converters prefer less time-consuming tools. In fact, as they need to observe the overall behavior of electron devices and other electromechanical devices, which are part of the power converters as well as the control circuits, great effort is devoted to simulating several switching commutations at once. Unfortunately, this is incompatible with the physical-equation approach. Obviously, a reduction in the simulation time entails a reduction of the accuracy with respect to a single commutation, but the system behavior is obtained. This second path is followed by the packages (25, 26) that use the Ebers–Moll and Gummel–Poon models. However it must be pointed out that these packages, do not describe the behavior of power transistors as well as that of integrated low-power BJTs.

According to the different needs, as above outlined, a lot of commercial software tools are available on the market. Sophisticated packages, devoted to the computer-aided design of semiconductor devices, have as a cornerstone a set of main equations that allow simulations in three dimensions. The complexity often can be reduced by exploiting symmetry, and in such a case the simulation is sped up significantly. Basically five fundamental equations are considered (27).

Poisson's Equation. Poisson's equation for a homogeneous lattice having a dielectric constant ε can be written as

$$\nabla E = \frac{\rho}{\epsilon} \tag{66}$$

which reduces in one dimension to

$$\frac{dE}{dx} = \frac{\rho}{\epsilon} \tag{67}$$

with p the volumetric charge density given by

$$\rho = q(p - n + N_{\mathrm{D}} - N_{\mathrm{A}}) \tag{68}$$

where p, N, N_{D}, and N_{A} are the volumetric concentrations of holes, free electrons, and donor and acceptor ions.

Current Density Equations. By considering a one-dimensional model for motion within a doped crystal, one obtains a well-known form of the equation for the current density of holes:

$$J_p = q\mu_p E - qD_p \frac{dp}{dx} \tag{69}$$

where $\mu_p = q\tau/2m_p$ is the mobility of holes and $D_p = (kT/q)\mu_p$ is the hole diffusivity.

The density equation for free electrons is given by

$$J_n = q\mu_n E - qD_n \frac{dn}{dx} \tag{70}$$

where $\mu_n = q\tau/2m_n$ is the electron mobility and $D_n = (kT/q)\mu_n$ is the electron diffusivity.

Continuity Equations. The carrier concentrations p and n in a semiconductor volume can vary with time due to the spatial variations of the hole density J_p and electron density J_n, which are caused by the generation and recombination of electrons and holes. Depending on the energy absorbed, a valence electron can leave the valence band and either reach the conduction band or be trapped inside the bandgap. The transfer from the valence band to the conduction band is called direct generation (a), while the intermediate state is called hole emission (b). The transition from the trapped state to the conduction band is called electron emission (c). The recombination phenomena can be considered as the inverses of the generation ones and are called respectively direct recombination (a′), hole capture (b′), and electron capture (c′). According to the above analysis the continuity equations for holes and electrons are respectively

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial x} + (G_{\mathrm{th}} + g - R)_{\mathrm{h}} \tag{71}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + (G_{\mathrm{th}} + g - R)_{\mathrm{e}} \tag{72}$$

where $G_{\mathrm{th}} + g$ is the thermal and nonthermal generation per unit volume and time, R is the recombination per unit volume and time, and $G_{\mathrm{th}} + g - R$ is the net generation rate of holes or free electrons.

The above set of equations is solved in commercial tools by several standard discretization methods. The device geometry is implemented by a mesh where the nodal points can be adapted according to the doping profiles, the electrode positions, and the peculiar features of the specific device. By varying the electrode bias, Eqs. (66) (72) are solved numerically at each nodal point.

Behavioral Models. From the above general equations, through the introduction of simplifying hypotheses to reduce the complexity of the problem, BJT equations that hold with some approximation both in thermal equilibrium and under biased conditions are derived in order to obtain analytical models (25). A well-known model is the Ebers–Moll one (28). The equations for the currents at the BJT terminals in steady-state conditions are

$$I_C = (I_{CC} - I_{EC}) - \frac{I_{EC}}{\beta_R} = I_s\left[\exp\left(\frac{qV_{BE}}{kT}\right) - \exp\left(\frac{qV_{BC}}{kT}\right)\right] - \frac{I_s}{\beta_R}\left[\exp\left(\frac{qV_{BC}}{kT}\right) - 1\right] \tag{73}$$

$$I_B = \frac{I_{CC}}{\beta_F} + \frac{I_{EC}}{\beta_R} = \frac{I_s}{\beta_F}\left[\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right] - \frac{I_s}{\beta_R}\left[\exp\left(\frac{qV_{BC}}{kT}\right) - 1\right] \tag{74}$$

$$I_E = -\frac{I_{CC}}{\beta_F} - (I_{CC} - I_{EC}) = \frac{I_s}{\beta_F}\left[\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right] - I_s\left[\exp\left(\frac{qV_{BE}}{kT}\right) - \exp\left(\frac{qV_{BE}}{kT}\right)\right] \tag{75}$$

The major limitations of such a model are: (1) the resistances of the base, emitter, and collector regions are not included; (2) the modulation of the base width (Early effect), that is, the influence of the base–collector voltage on the saturation current, is neglected; (3) the forward (β_F) and reverse (β_R) current gains are considered independent of the current value, and the high-injection effects are neglected. The Ebers–Moll dynamic large-signal model of BJTs can be derived from the static model by allowing for the effect of charge storage by including quantities modeling the junction, diffusion, and substrate parasitic capacitances. The behavior in the forward-biased condition for small signals can be analyzed by introducing some changes to make the model dynamic.

The Gummel–Poon model (19) removes some limitations on the above model, more precisely with respect to the base width modulation and the variation of the forward current gain. Its version for large signals, which is based on the Ebers–Moll model, allows for the variation of the transit time and for the distributed base–collector capacitance. Moreover, the small-signal model can be derived similarly to the Ebers–Moll model by including new quantities, which have different definitions. As an example, in Fig. 30 is reported the equivalent circuit of the Gummel–Poon model as implemented in the simulation package SPICE (25, 26). However, the Gummel–Poon model also does not exactly fit a power BJT, since it does not take into account emitter current crowding, base pinch, and quasisaturation effects. The last aspect has been treated in Refs. 13, and 29 by

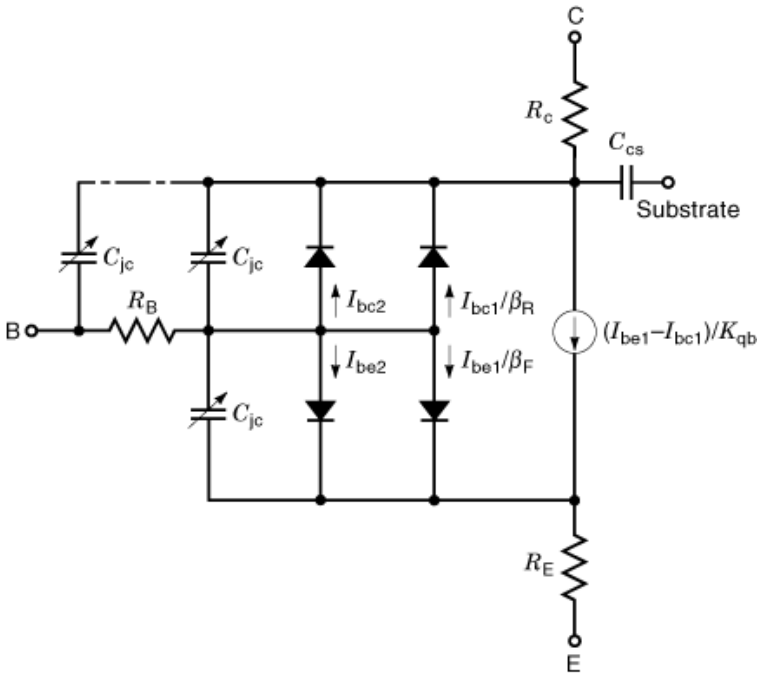

Fig. 30. Equivalent circuit of the Gummel–Poon BJT model implemented in SPICE.

introducing some more parameters in the BJT model. Interested readers can also find in Ref. 29 a detailed analysis of the behavioral models.

Trends in BJT Development

The growth of sales of discrete power BJT is already finished, and the market for their is likely to be flat in the next years. Other well-consolidated power devices with different technologies, such as power MOSFETs and IGBTs, or new ones integrating the power and control stages on the same chip, have already taken the place of power BJTs in many applications. However, in the range of medium-voltage high-frequency, discrete power transistors are still a viable solution, and in the field of resonant inverters too. Looking at these applications, and at their typical configurations, such as push–pull and half bridge, two major peculiarities are required of BJTs: (1) very tight control of the storage time is mandatory, since the reliability and the output power are strictly dependent on it; (2) at startup or during some fault conditions (load short circuited or sudden load step), the collector–emitter voltage can exceed the rated breakdown voltage; thus the devices should be able to withstand significant current levels in breakdown condition. The most popular technique used to control the BJT storage time is to trim the minority carrier lifetime by introducing a lifetime killer during the diffusion process (heavy ions, or damage to the crystal by electron irradiation).

In general, circuit designers have protected against the above misbehavior by inserting appropriate circuits such as the Baker clamping network (23) or current and voltage transient snubbers. However, the tendency of BJT manufacturers is to integrate into the device as many safety functions as possible that otherwise would have to be performed by external auxiliary circuitry. Moreover, interesting new possibilities are continually arising for the use of process steps coming from more advanced technologies, such as large-

scale integration (LSI) and MOS—in particular, at present, the monolithic cascode configuration and the antisaturation network integrated on a single chip.

BJT and MOSFET as Monolithic Cascode-Connected Switches. The combination of semiconductor switches having different technologies to build a power switch has been exploited to combine their advantages while avoiding their disadvantages. Among these, we can mention the cascode connection of two discrete devices as depicted in Fig. 31(a) for a MOSFET–BJT pair. However, a complementary zener diode is required between the BJT base and the MOSFET source to provide the turnoff path for the collector current in emitter-opened conditions. The BJT is designed as a high-voltage high-power switch, and the MOSFET as a very fast low-voltage (some tens of volts) device. The advantages of emitter-open (cascode) switching are: (1) the turnoff is speeded by substantial reduction of the storage time and enhancement of the fall time, which in turn allows higher operating frequencies; (2) the forward voltage drop is lower than in a single device specified for sustaining the same off-state voltage; (3) the reverse-biased secondary breakdown of the BJT is practically eliminated, due to the mechanism of switching through the BJT base, which avoids current crowding under the emitter region; (4) a negative voltage supply for the drive circuitry to extract the charge from the base is not required, and only a small amount of energy need be provided by a voltage source to drive the MOSFET.

Recently the cascode has been manufactured as a monolithic device with a high-voltage n–p–n BJT and a low-voltage vertical power MOS integrated in a sandwich structure, Fig. 31(b). The major advantage over a discrete component realization is in the use of the same silicon area, which has been achieved by integrating the low-voltage power MOS inside the emitter area of the high-voltage BJT, and in manufacturing the zener diode in the same structure (30). Applications of such a device are in lighting, TV deflection circuits, and motor control. To improve the current gain of the bipolar part, two or three stage connections on the same chip (MD and multistage MD) have been developed, thus decreasing the power consumption on the BJT base, although in such cases the advantage in RBSOA is no longer achievable. Monolithic cascode devices with 1500 V and tens of amperes, with operating frequency higher than 100 kHz and with forward voltage drop of 1 V to 2 V are available. In Fig. 32(a) is drawn the simplified drive circuit for the cascode and the current values during a turnoff transient on inductive load. In Fig. 32(b) are reported the traces of the collector and base currents during the turnoff transient.

Antisaturation Network. One advantage of the Baker clamp network used on the base drive of BJTs is the speedup of the turnoff transient through the action of an antisaturation diode connected between the collector and the driving terminal (23). In fact, the presence of such a diode, and others suitably connected, allows one to maintain the base–emitter voltage equal to the collector–emitter voltage in on-state conditions, thus avoiding the occurrence of saturation voltage between the collector and emitter. A lower saturation voltage means a lower storage time and thus faster switching at the expense of conduction power losses. The antisaturation diode should be designed as a fast recovery diode (with recovery time less than the storage time of the BJT), with a reverse voltage capacity equal to that of the main BJT. Antisaturation networks have been implemented on a single-chip power BJT, using a p–n–p lateral transistor extending out from the monolithic integration of a freewheeling diode as shown in Fig. 33. Such a technique is effective only in applications where the current flowing in the freewheeling diode does not cause spurious turn-on.

Bipolar Transistor with Fast-Recovery Integral Diode. In many applications devoted to supplying inductive loads, with half or full bridge schemes, the switching devices have antiparallel-connected diodes to provide the capability of bidirectional current flow through the bridge legs. These freewheeling diodes, in applications operating in switch mode at high frequencies, should be fast recovery diodes in order to have low reverse recovery current, which adds to the load current at turnon, thus increasing the current peak and the power loss of the main switches. The freewheeling diodes usually are discrete components, externally connected by the user, or integrated in switching modules by the manufacturer in devices for high-current applications, thus minimizing the layout parasitic inductances.

However, a diode can be integrated in the structure of a BJT, and its reverse recovery behavior can be enhanced by metal doping or electron irradiation. The use of electron irradiation or high-temperature diffusion

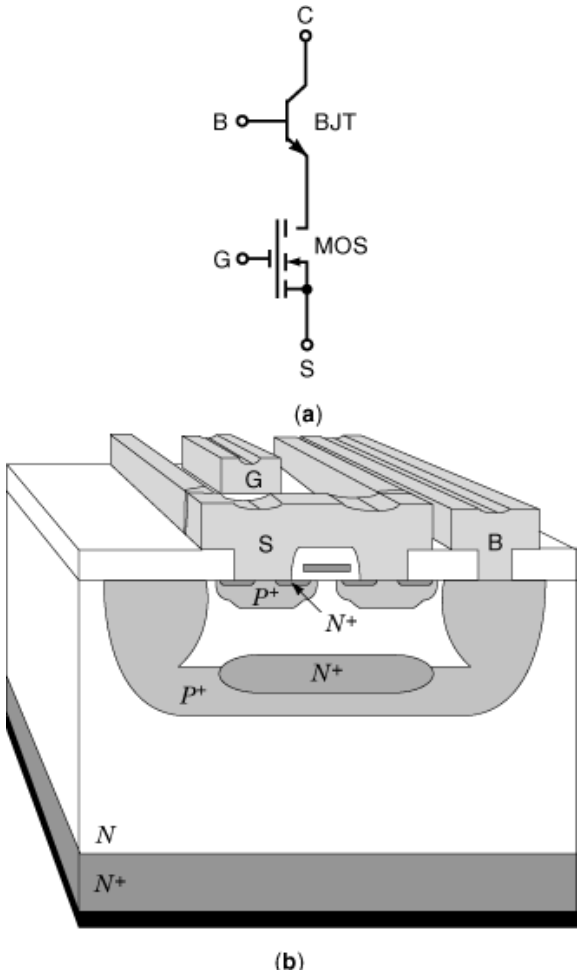

Fig. 31. Monolithic emitter switching: (a) equivalent electric connection of the device; (b) cross section of an actual integrated emitter switching. The N^+ emitter of the power BJT part is buried, and the MOS device is overlapped on the BJT area and has vertical current flow.

of metal impurities produces lifetime reduction in the whole device structure, thus causing a significant degradation in the BJT output characteristics. Such a degradation of the performances can be avoided only if lifetime reduction is selectively achieved in the area of the integrated diode. Precise control of vertical and lateral profiles of atoms in silicon has been obtained in a high-voltage BJT, having a fast recovery diode in a monolithic structure, with the prediffusion source realized by ion implantation of platinum in an area as small as 200 $\times$ 200 μm. The switching transient of the diode has been improved without affecting the static characteristics of the main device (31). In this case, the value of the stored charge Q_{RR} is reduced from 3.8 μC in the undoped device to 0.3 μC in the platinum-doped device, while the current capability I_k is reduced from 3.5 A to 2.8 A.

Unclamped Inductive Switching. Manufacturers verify the ruggedness of their own devices under electrical stresses by submitting them to severe test conditions in order to guarantee users against destructive failures in actual circuits. Turnoff commutation in unclamped inductive switching (UIS) is a benchmark of

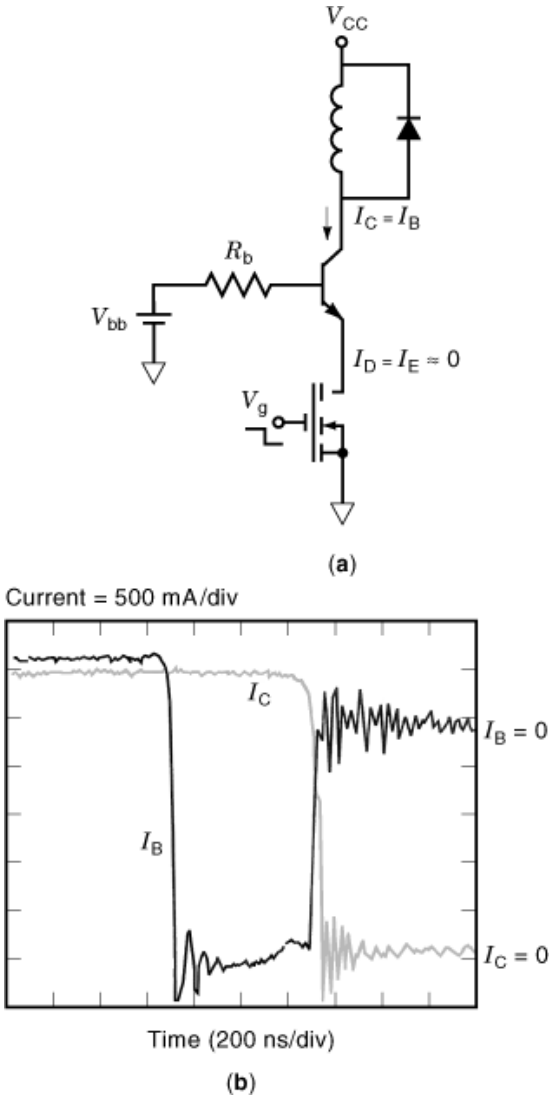

Fig. 32. A cascode device test circuit, and the resulting current traces during a turn-off transient: (a) practical circuit with an inductive load, current values during a turn-off, and simplified drive circuits; (b) traces of the base and collector currents during the turn-off transient.

such ruggedness, since the energy stored in the magnetic field of inductors and leakage inductances stresses the device strongly (32). The UIS circuit test is in principle similar to that shown in Fig. 24, except that the freewheeling diode D_1 is missing. The objectives of such a test can be summarized as understanding (1) the inherent ability of the device to survive a pulse of energy in extreme conditions, (2) the influence of the device structure and the ways to assure safe operation.

Today, BJTs are designed to withstand current of several amperes in breakdown. For example, a transistor with $BV_{CB0} = 1600$ V can withstand a current peak up to 5 A at 1800 V, and a transistor with $BV_{CB0} = 900$ V can withstand a current peak up to 7 A at 1100 V, as is shown in Fig. 34 for an experimental test in UIS

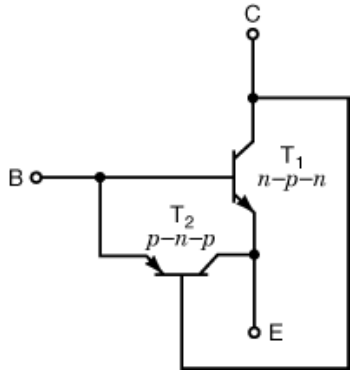

Fig. 33. Electric circuit of a power BJT (T_1) with an integrated antisaturation diode. The diode is actually integrated as a BJT (T_2).

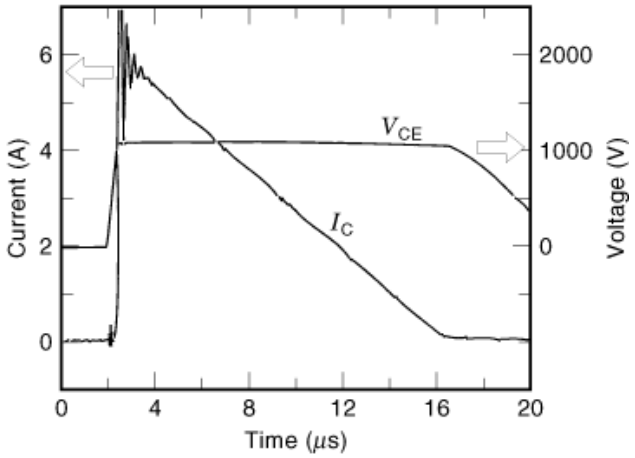

Fig. 34. Transient behavior during turn-off on an unclamped inductive load; during the turn-off the voltage across the device remains at the breakdown value at the rated current for a suitable time.

conditions. It is possible to bear such severe conditions by ensuring very high boundary efficiency in reverse biasing. Control of boundary edges, by using techniques like multiple rings, resurf rings, or variations on lateral doping concepts, allow the critical electric field to be reached uniformly under the main junction, far from the boundary, thus reducing the current density peak during the breakdown.

BIBLIOGRAPHY

1. M. S. Adler *et al.* The evolution of power device technology, *IEEE Trans. Electron Devices*, **ED-31**: 1570–1590, 1984.
2. S. M. Sze *Physics of Semiconductor Devices*, 2nd ed., New York: Wiley, 1981.
3. P. Antognetti (ed.) *Power Integrated Circuits: Physics, Design and Applications*, New York: McGraw-Hill, 1986.
4. B. J. Baliga *Power Semiconductor Devices*, Boston: PWS Publ., 1995.
5. F. Conti M. Conti Surface breakdown in silicon planar diodes equipped with field plate, *Solid State Electron.* **15**: 95–105, 1972.
6. A. Blicher *Field-effect and Bipolar Power Transistor Physics*, New York: Academic Press, 1981.

7. S. M. Teray B. W. Grabowsky Junction breakdown analysis of a floating field–limiting ring structure using a PISCES-2B devices simulator, *Solid State Electron.* **35**: (2), 201–205, 1992.
8. A. Scandura, et al. Diffusion and electrical behavior of Al implanted into capped Si, *J. Electrotechn. Soc.*, **140** (7): 2057–2062, 1993.
9. R. Stengl U. Gosele Variation of lateral doping: A new concept to avoid high voltage breakdown of planar junction, *Proc. Int. Electron Devices Mtg.* 1985, pp. 154–157.
10. S. K. Ghandi *Semiconductor Power Devices*, New York: Wiley, 1987.
11. H. J. Deman The influence of heavy doping on the emitter efficiency of a bipolar transistor, *IEEE Trans. Electron Devices*, **ED-18**: 833–835. 1971.
12. G. Rey F. Dupy, J. P. Bailb A unified approach to the base widening mechanism in bipolar transistors, *Solid State Electron.* **18**: 863–866, 1975.
13. G. M. Kull, et al. A unified circuit model for bipolar transistors including quasi-saturation effects, *IEEE Trans. Electron Devices*, **ED-32**: 1103–1113, 1985.
14. R. V. Martinelli The temperature dependence of the dc base and collector currents in silicon bipolar transistors, *IEEE Trans. Electron Devices*, **ED-23**: 1218–1224, 1976.
15. D. J. Brenner A technique for measuring the surface temperature of transistors by means of fluorescent phosphors, NBS Tech. Note 591, July 1971.
16. *Thermal Resistance Measurements of Conduction Cooled Power Transistors*, EIA Recommended Std. RS-313-B (revision of RS-313-A), October 1975.
17. L. Fratelli G. F. Vitale On the reverse bias safe operating area of power bipolar transistors during inductive turn-off, *Solid State Electron.* **37** (2): 275–287, 1994.
18. M. H. Rashid *Power Electronics: Circuits, Devices, and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1994.
19. H. K. Gummel H. C. Poon An integral charge control model of bipolar transistors, *Bell Syst. Tech. J.*, **49**: 827–851, 1970.
20. G. P. Chiavarotti M. Conti, S. Luciani The turn-off time of high voltage epitaxial transistors, *Rev. Phys. Appl.* **13**: 737–739, 1978.
21. *Le Transistors de Puissance dans la Conversion d'Energie, Tome II—de la Compréhension à la Réduction des Pertes* (in French), Thomson-CSF, Division Semiconducteurs Discrètes, 1983.
22. *The Power Transistor in Its Environment, Vol. III—Switching Times of Power Transistors*, Thomson-CSF, Division Semiconducteurs Discrètes, 1984.
23. N. Mohan T. M. Undeland, W. P. Robbins *Power Electronics: Converters, Applications, and Design*, 2nd ed., New York: Wiley, 1995.
24. D. Ballarò *et al. Darlington device with an ultra light weight emitter speed up transistor*, US Patent No. 5,083,182, filed Oct. 24, 1989.
25. L. W. Nagel SPICE 2: A computer program to simulate semiconductor circuits, Memorandum ERL-M520, Electronics Research Lab., Univ. of California, Berkeley, 1975.
26. *PSpice*, MicroSim Corp., Irvine, CA 92718.
27. U. Cilingiroglu *Systematic Analysis of Bipolar and MOS Transistors*, Boston: Artech House, 1993.
28. J. J. Ebers J. L. Moll Large-signal behavior of junction transistors, *Proc. IRE*, **42**: 1767–1769, 1954.
29. P. Antognetti G. Massobrio *Semiconductor Device Modeling with SPICE*, New York: McGraw-Hill, 1988.
30. F. Frisina G. Ferla *Integrated high voltage bipolar power transistor on low voltage MOS power transistor structure in the emitter switching configuration*, US Patent No. 5,065,213, filed Dec. 21, 1988.
31. S. Coffa *et al.* Power bipolar transistor with a fast recovery integrated diode, *IEEE Trans. Electron Devices*, **43**: 836–839, 1996.
32. K. Fischer K. Shenai Dynamics of power MOSFET switching under unclamped inductive loading conditions, *IEEE Trans. Electron Devices*, **43**: 1007–1015, 1996.

FERRUCCIO FRISINA
ALFONSO PATTI
MARIO SAGGIO
STMicroelectronics
ANGELO RACITI
University of Catania