The generation of three-dimensional bipolar transistor models for circuit analysis

by F.-Y. Chang
L. F. Wagner

The results of a two-dimensional bipolar numerical device-analysis program are processed by identifying three regions of the transistor: the intrinsic transistor, the sidewall transistor, and the extrinsic base collector diode. The key parameters which describe each of these regions are extracted using a linking program and fed into a quasi-three-dimensional device analysis program referred to here as the Model Generation Program (MGP). The MGP first generates a large equivalent-circuit distributed network which simulates the three-dimensional geometry of an actual transistor. This distributed network is then analyzed using the existing Advanced Statistical Analysis Program (ASTAP) for circuit analysis. Finally the MGP extracts parameters from the ASTAP analysis to characterize the elements of a lumped-model equivalent circuit, which is then suitable for the circuit design of large-scale integrated circuit chips. The MGP is sufficiently flexible that transistors with a variety of geometries can be generated without repeating the two-dimensional analysis.

Introduction

Advanced technology development requires long lead times from when a new transistor design is conceived until it is reduced to practice. Since there are many options to consider before a design is finalized, and because subsequent changes can be costly, it is desirable to evaluate the consequences of these options well in advance. There are many trade-offs whereby process changes can improve one aspect of circuit performance but degrade another. It is therefore necessary to be able to generate accurate device models, so that the impact of process changes on circuit performance can be objectively evaluated.

We have available an accurate two-dimensional process modeling capability (SAFEPRO [1]) for simulating profiles based on a proposed new technology. We also have available a state-of-the-art two-dimensional device-analysis program (2DP [2]) which can analyze devices on the basis of these profiles and provide us with internal details of device operation which are not otherwise available.

When we analyze the results of the two-dimensional device-analysis program, we find that the detailed carrier densities and current flow patterns actually provide too much information. It is necessary to develop methods to extract meaningful information from these results. In addition a real transistor is three-dimensional, but a full three-dimensional device analysis would be too costly. Thus it is necessary to develop a procedure for scaling the two-dimensional results so that accurate circuit models for a variety of real three-dimensional transistors can be constructed.

The strategy we have chosen addresses these considerations by identifying three unique regions of the two-dimensional transistor. These are a central filamentary transistor, a sidewall transistor, and an extrinsic base diode. These regions are characterized and subsequently combined to form a large distributed network equivalent circuit representing the full three-dimensional transistor. A final lumped equivalent circuit is then generated based on an analysis of the distributed network.

Transient behavior is modeled by adding capacitance to the distributed and lumped equivalent circuits. The two-dimensional program 2DP can only do a static analysis; however, it computes capacitance in the quasi-static approximation from changes in carrier concentrations resulting from small changes in transistor bias [2].

In this way we have combined capabilities for process modeling, device analysis, and model generation to form a powerful set of tools that can affect the course of the design process by providing significant information about circuit impact [3].

The Model Generation Program (MGP), which generates these equivalent-circuit models, consists of four segments as follows:

1. The results of the two-dimensional bipolar numerical device-analysis program (2DP) are first processed with a linking program (MGPLINK), which extracts key parameters from each of the transistor regions and sorts them for later use and for plotting.
2. This information is processed by Part A of the MGP (MGP-A). MGP-A is a FORTRAN program that generates a large equivalent-circuit distributed network which simulates the three-dimensional geometry of the actual transistor.
3. This distributed network is then analyzed using the Advanced Statistical Analysis Program (ASTAP).
4. Finally the FORTRAN program MGP-B uses parameters from the ASTAP analysis to characterize the elements of a lumped-model equivalent circuit which is suitable for the design of large-scale integrated circuit chips.

In this paper we discuss each of the above program segments and conclude by giving some examples of MGP output.

The linking program

The Model Generation Program Link (MGPLINK) is a FORTRAN program which analyzes the two-dimensional numerical simulation results and extracts the parameters necessary to describe each of the three transistor regions. The MGPLINK program automatically creates an input file for Part A of the quasi-three-dimensional Model Generation Program (MGP-A). In addition, as a by-product of this analysis, the MGPLINK program creates files for plotting

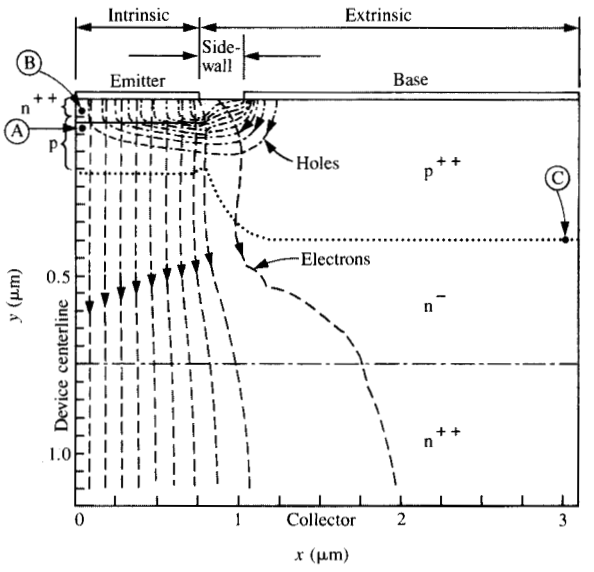

Figure 1

Contour plot of electron and hole current flow from the 2DP simulation: Each contour line encloses 10% of the electron or hole current flow. The vertical scale is expanded by about a factor of two for clarity.

the information which characterizes each of the three regions of the transistor.

In this section we describe the parameters which are extracted from the 2DP program and the reasons behind their choice.

• *Extracting parameters from the two-dimensional device analysis*

When one considers the current flow from the two-dimensional device analysis as shown in the contour plot [4] in **Figure 1**, it is easy to see that the electron current in the center of the device flows essentially in one dimension. In fact, since the current flow is nearly vertical in most of the n-region under the emitter, the characterization of this central region applies very closely to the entire region under the emitter. This is the motivation for considering a central filamentary transistor.

In contrast to this, the current flow in the sidewall region is highly two-dimensional; the hole flow is perpendicular to the electron flow. For this reason we are forced to consider the sidewall region as an independent unit.

In the extrinsic base region the current flow is small, except when the transistor is saturated or in the inverse mode. Under these conditions the forward-biased base-to-collector current under the extrinsic base is again approximately one-dimensional. The extrinsic base-collector diode is thus treated as a one-dimensional filament.

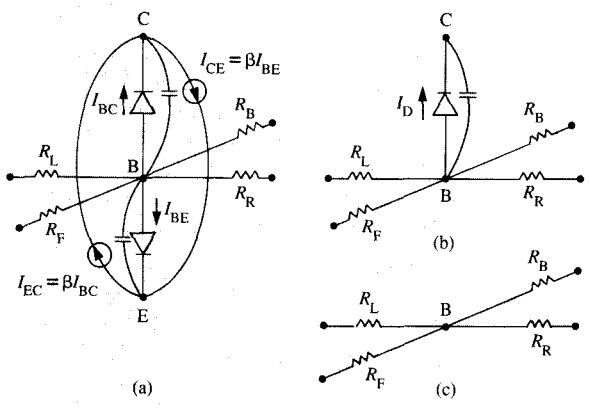

Figure 2

Elementary equipment circuits in the distributed model: (a) Equivalent circuit used for the central intrinsic transistors and the sidewall transistors. Note that the transport model as illustrated in Fig. 3 is used. The base diodes use a standard diode equation, Eq. (1), while the resistors and beta values are current-dependent. The capacitances combine a voltage-dependent depletion part with a current-dependent neutral component. (b) Equivalent circuit used for the extrinsic base diodes. (c) Equivalent circuit used for polysilicon over oxide. These resistors are taken to be constant.

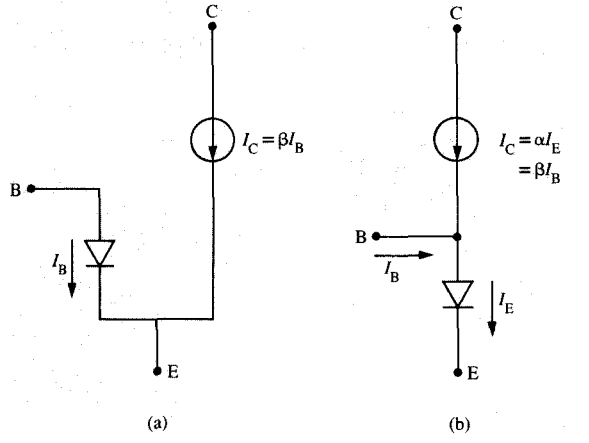

Figure 3

The transport model and the original Ebers-Moll model compared: The transport model (a) separates hole and electron current flow. This leads to a simpler formulation than the original Ebers-Moll model (b), in which the base-emitter diode must describe both hole and electron current flow. The transport model is used for forward and inverse mode in Figs. 2(a) and 11.

The following paragraphs describe the characterization of the central filamentary transistor, the sidewall transistor, and the extrinsic base diode. This is followed by a discussion of our interpretation of the numerical capacitance calculations and their splitting into neutral and depletion components and into intrinsic and extrinsic parts.

The filamentary transistor

The filamentary transistor, as discussed above, is derived only from current flows in the central part of the two-dimensional simulation. The equivalent circuit in **Figure 2(a)** shows the elements which must be characterized. Note that this circuit is in the transport form, **Figure 3(a)**; that is, the resistors and diodes represent only hole current, while the beta-dependent sources represent the transport of electrons from the emitter to the collector. This is different from the original Ebers-Moll form [5] where the diodes represent both electron and hole current flow, **Figure 3(b)**. The justification for this choice is discussed below.

The sheet resistance and the beta values are assumed to be functions of the transistor operating point. The characterization of each of the elements in the filamentary transistor follows.

The four resistors R_R, R_L, R_F, and R_B represent resistance to hole current flow resulting from the sheet resistance of the intrinsic base. The sheet resistance, as a function of bias, is obtained directly from 2DP by integrating the product of the hole mobility and concentration to obtain the conductivity, and then taking the reciprocal. The electron concentration does not enter this computation since only hole flow is being modeled by this part of the equivalent circuit.

In order to characterize the base-emitter diode and the beta-dependent current sources, it is necessary to identify the corresponding features in the two-dimensional simulation.

The base voltage is identified with the hole quasi-Fermi level. This is nearly constant in the base region along the vertical central filament. However, for an automatic numerical calculation this definition must be more precise. For this reason a vertical search is made to find the maximum hole density. Since this should be the highest conductivity point, the hole quasi-Fermi energy at this point is taken to be the base voltage. This point is marked A in Fig. 1.

The base current enters the filament distributed along the side; it then turns and flows vertically into the emitter. Thus the base current density increases as one moves vertically toward the emitter. In the emitter some of the holes recombine with electrons, so that the vertical hole current density reaches a maximum and then decreases toward the emitter. The vertical hole current density at this maximum is then taken to be the hole current represented by the base-emitter diode. The point at which the hole current is measured is marked B in Fig. 1.

The electron current density is nearly constant throughout the filament, but again, to be precise, the electron current density is also evaluated at the point of maximum hole current density. The forward beta is then simply the ratio of electron to hole current density at this point.

Figure 4(a) shows a typical semi-logarithmic plot of electron and hole current density vs hole quasi-Fermi potential obtained in this way. Note that this filamentary base current, the hole current density, forms almost a perfect straight line on this plot. There is no curve at high bias from a series *I-R* voltage drop, since the characterization is done internally inside the transistor via the simulation. In contrast, the curve for the collector current, the electron current density, is curved and approaches the base current at high forward bias. Physically this results from the Kirk effect [6]. This is illustrated in **Figures 5** and **6**, which show how the base pushes out at large current density, causing the transistor beta to decrease. These observations justify as simpler the transport form for the equivalent circuit [Fig. 3(a)] chosen above.

The base-collector diode and the inverse beta are evaluated in a similar manner from inverse-mode 2DP simulations.

The sidewall transistor

In the sidewall region of the transistor, it is not easy to identify filaments which, when combined, properly represent current flow. For this reason the sidewall region is taken as a single unit.

The equivalent circuit for the sidewall region is the same as for the central filamentary transistor except that the elements represent a unit of perimeter instead of area and thus have different dimensions. Similarly, it is not as straightforward to characterize these elements from information in the 2DP simulation. An important consideration for the characterization of these elements is the need to make a concrete definition which can be easily implemented numerically.

For the implementation we have chosen, the resistors are characterized by an average sheet resistance in the sidewall region. This bias-dependent average is obtained by taking the 2DP sheet resistance values over the sidewall to be short resistors connected in series.

The base voltage is found by taking a numerical average of the hole quasi-Fermi potential over the same region. This average is evaluated on grid points at the same depth used in evaluating the base voltage of the central filamentary transistor.

As can be seen from the contour plot in Fig. 1, the sidewall hole current largely flows laterally from the extrinsic base to the intrinsic base under the emitter. The base current for the sidewall equivalent circuit is then taken to be the total hole current crossing the side of the metallurgical base-emitter junction.

The electron current spreads out from its vertical flow in the sidewall region, but it becomes orders of magnitude smaller towards the extrinsic base. The collector current for the sidewall equivalent circuit is then taken to be the sum of the electron current crossing the metallurgical base collector from the beginning of the sidewall completely through the extrinsic base.

The logarithmic plots **[Figure 4(b)]** of base and collector currents vs base voltage defined in this way have properties similar to those for the intrinsic filamentary transistor. Thus the transport model [Fig. 3(a)] is still a good representation for the sidewall.

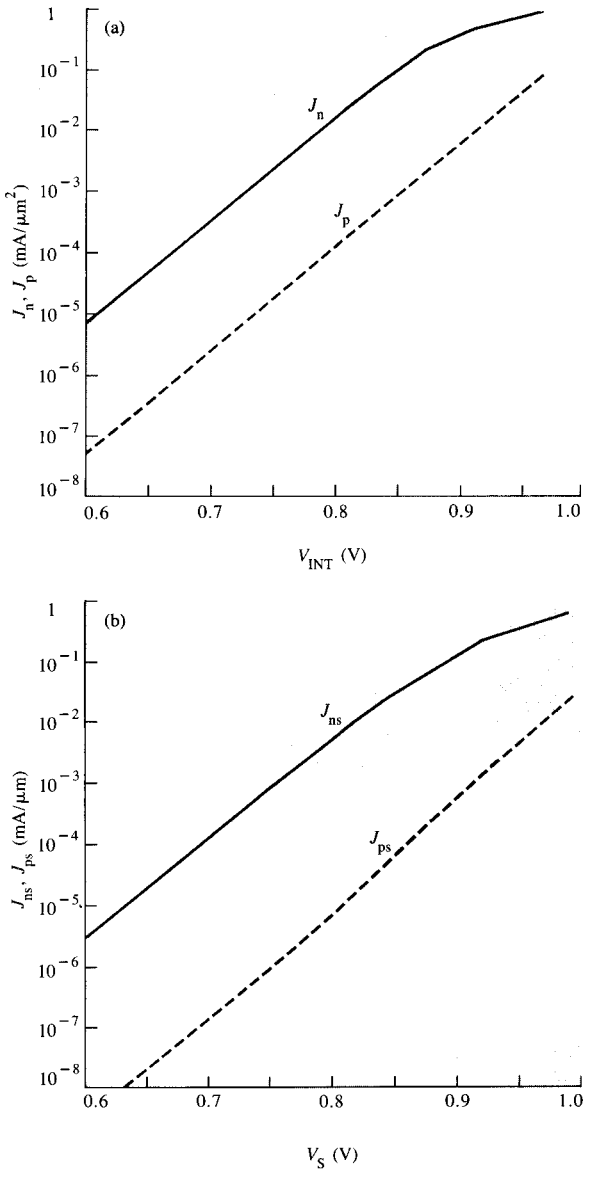

Figure 4

Internal electron (J_n) and hole (J_p) current densities: (a) For the central filamentary transistor the hole current density forms a nearly perfect straight line, thus justifying the choice of the transport model. (b) The hole current density for the sidewall transistor shows similar approximately straight-line behavior. Note that the units are different because the sidewall element represents a unit of perimeter.

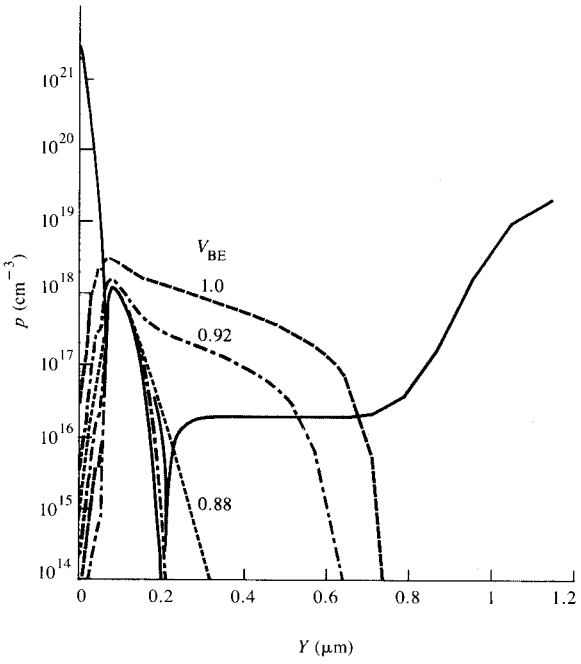

Figure 5

Hole concentration: The hole concentration at the center of the transistor is plotted for several values of V_{BE} superimposed on the net doping concentration. Note how the base pushes out at high bias due to the Kirk effect.

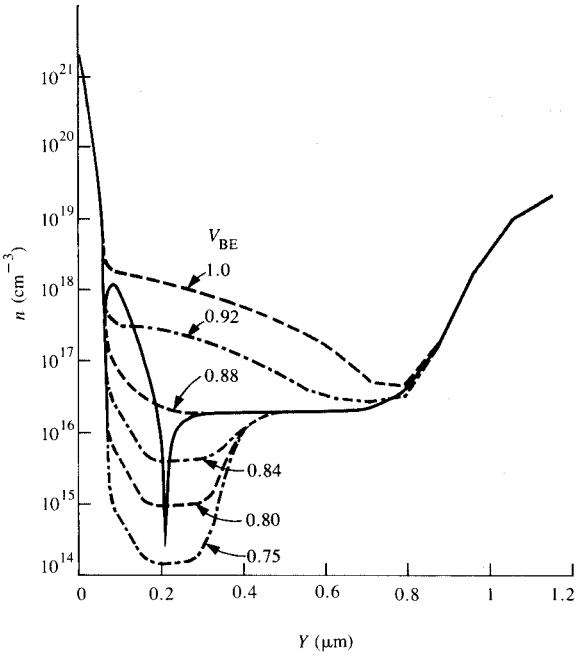

Figure 6

Electron concentration: The electron concentration corresponding to the hole concentration in Fig. 5 is plotted.

There is not a good definition for sidewall inverse mode currents. The inverse mode for the sidewall transistor is at present only estimated or not characterized at all. This makes little difference since the total inverse transistor base current is largely dominated by the extrinsic base-collector diode.

The extrinsic base diode

The inverse collector-base current becomes again approximately one-dimensional as one moves into the extrinsic base region. The extrinsic base-collector diode is then characterized by the outermost vertical filament in the 2DP simulation using the equivalent circuit in **Figure 2(b)**.

The polysilicon over the transistor extrinsic base is treated only as a boundary condition in the 2DP simulation. Therefore, the extrinsic base-resistor values are obtained by taking the calculated silicon sheet resistance, which is bias-dependent, to be in parallel with the known sheet resistance of the base polysilicon.

There are no *I-R* voltage drops in the simulation for the diode, so the diode voltage is just the external base-collector voltage from the 2DP simulation. The diode current is taken to be the sum of the hole and electron current densities evaluated at the metallurgical collector-base junction, point C in Fig. 1. Again this characterization gives a good straight-line log (I) vs V plot.

Extrinsic resistance

The extrinsic resistance of polysilicon over oxide is represented by the equivalent circuit in **Figure 2(c)** consisting of only distributed resistors. These resistors have constant values determined by the known value of sheet resistance.

Capacitance

The 2DP program not only computes base-emitter and base-collector junction capacitances, but also identifies components of capacitance according to the region of the transistor (intrinsic or extrinsic) and according to type (depletion or neutral) [2]. The split of the emitter and collector capacitances into their depletion and neutral parts is shown in **Figures 7** and **8**.

For the purpose of the above equivalent circuits, the depletion (or space charge) component of the capacitance is treated as being dependent on the junction voltage. Physically this component arises because of the variation of the space charge depletion region width with junction voltage.

The neutral (or diffusion) component is treated as being electron- (or collector-) current-dependent. This component results from the increase in both hole and electron densities at high currents where the base pushes out due to the Kirk effect [6]. When this occurs, the base and the epitaxial regions maintain approximate charge neutrality, hence the name neutral capacitance.

For the intrinsic filamentary transistor, the intrinsic components of capacitance are normalized by area and included in the equivalent circuit.

No special component of capacitance is computed by 2DP for the sidewall transistor area. Therefore the extrinsic capacitance must be treated differently. First, the depletion component is normalized by area and allocated to the sidewall transistor and the extrinsic base diode according to their respective areas. Second, the neutral component, under forward transistor bias where the base-collector junction is reverse-biased, arises because of the high electron current density coming through the sidewall region. This component is therefore normalized by emitter perimeter and assigned only to the sidewall transistor. In inverse mode the base-collector junction is forward-biased and even the neutral component is normalized by area.

• *Operation of the linking program*

The Model Generation Program Link (MGPLINK) automatically creates an input file for Part A of the quasi-three-dimensional Model Generation Program (MGP-A). In addition, as a by-product of this analysis, the MGPLINK program creates files for plotting the information which characterizes each of the three regions of the transistor.

At each bias point of the 2DP program, selected information is stored in an intermediate file. The MGPLINK program sequentially reads this information and sorts it into arrays which are outputted into files for MGP input and for plotting.

The passed dataset

The ASTAP circuit analysis program describes nonlinear elements in one of two ways, either with analytical expressions or by the use of linear interpolation between the points of tables. Most of the extracted information to be read by the MGP consists of tables which list the various values of beta, resistance, and capacitance as functions of the transistor operating point. In addition, when simple expressions are available, some fitting takes place in the MGPLINK program. In these cases only the parameters of the expressions are needed by the MGP. Examples of the fitting are the base-emitter and base-collector diodes which have nearly straight-line semi-log plots. A saturation current I_0 and an ideality factor n are fit:

$$I = I_0(e^{qV/nkT} - 1). \quad (1)$$

Also, the depletion capacitances which are assumed to be voltage-dependent are fit to the expression

$$C = \frac{C_0}{\left(1 - \frac{V}{V_0}\right)^n} . \quad (2)$$

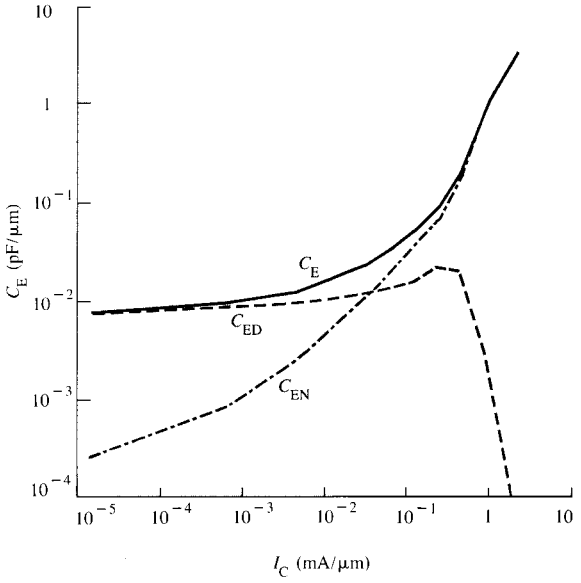

Figure 7

Base-emitter capacitance: The forward-biased base-emitter capacitance is shown with the split into depletion (C_{ED}) and neutral (C_{EN}) components.

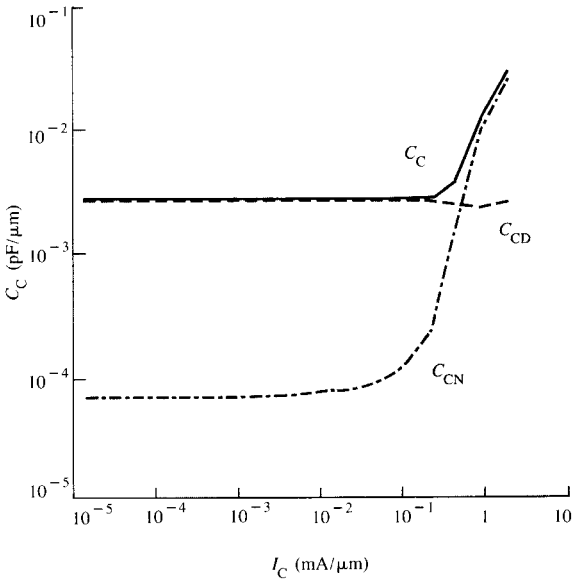

Figure 8

Base-collector capacitance: The base-collector capacitance, also divided into depletion (C_{CD}) and neutral (C_{CN}) components, is shown. Here the junction is reverse-biased. The increase in the neutral component is due to the base pushout from the Kirk effect.

Plotting features

The MGPLINK program sorts output from the 2DP two-dimensional simulation. This makes it easy to generate files

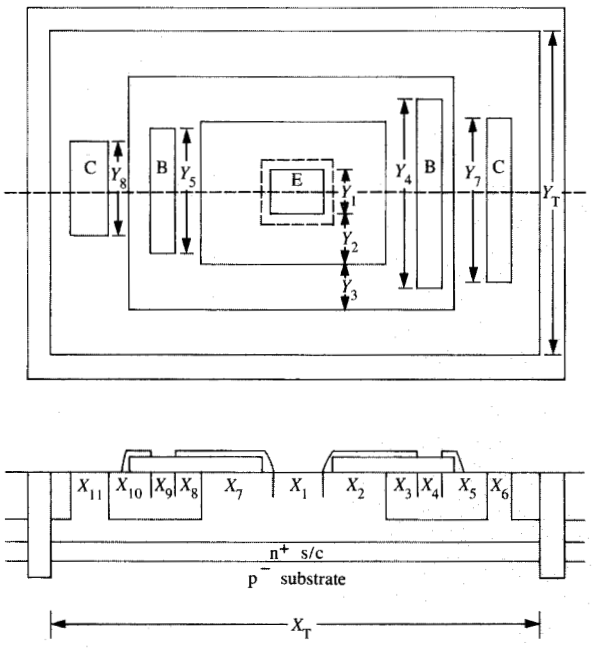

Figure 9

General transistor geometry used for MGP input: The input is sufficiently general that a base or emitter can be omitted from either side.

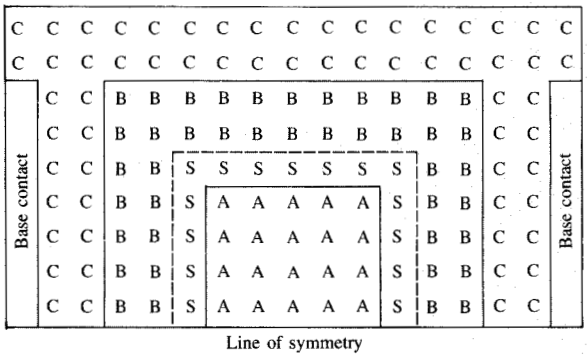

Figure 10

Typical array of models in the distributed network: The letters A, B, and C refer to the equivalent circuits in Figs. 2(a), 2(b), and 2(c) respectively. The letter S refers to the sidewall version of model A. These models are connected together and analyzed with ASTAP.

for plotting various quantities from the 2DP program vs depth, lateral position, or operating point. These plots have proved to be of great value for understanding the physical operation of the transistor. Figures 4 through 8 were made in this manner.

The Model Generation Program

Parts A and B of the Model Generation Program (MGP) provide the final step, putting together in a unified way the results of the two-dimensional process simulation (SAFEPRO) and device analysis (2DP) programs to produce useful models for circuit analysis. Input information, extracted from the 2DP program by the linking program (MGPLINK), describes central filamentary transistor elements, a sidewall transistor section, and an extrinsic base diode. These units are represented by equivalent circuits, then replicated many times in Part A of the MGP to create a large distributed equivalent-circuit model of a three-dimensional transistor. This equivalent-circuit is then analyzed using the Advanced Statistical Analysis Program (ASTAP) in a dc circuit analysis mode. Part B of the MGP then extracts parameters from the results of this analysis and forms a simplified equivalent circuit model of the transistor which is suitable for circuit design using many transistors.

The same parameters can be used in MGP-A to describe the internal elements of the distributed model for all usual transistor geometries. Thus, various horizontal geometries can be analyzed without rerunning SAFEPRO, 2DP, or MGPLINK. The computer time required by the MGP is small compared to that required by the SAFEPRO or the 2DP programs, so that a large number of equivalent-circuit models of transistors with various geometries can easily be generated as required for circuit analysis.

• *Loading the Model Generation Program*

The initial choice of information generated from the 2DP program requires a tradeoff between detail and computer time. We have chosen to emphasize the transistor forward mode; thus we compute about 30 forward and 10 reverse operating points using the 2DP program. Since this requires a large amount of computer time, it is often desirable to use fewer bias points. In this case the transistor simulated is usually a small perturbation on one for which a full set of bias points has been generated. For this purpose we can choose a partial set of 12 bias points. The MGPLINK program produces an incomplete dataset from these points; however, it is a simple matter to edit a complete dataset and substitute the perturbed information.

In addition to the information from the numerical simulation, the MGP needs data about polysilicon sheet resistance and subcollector parameters to describe the resistors in parts (b) and (c) of Fig. 2. The dimensions of a specific transistor horizontal geometry complete the specification of the MGP input.

It is also possible to edit the MGP input dataset at this stage to use experimental data to account for discrepancies between simulation and experiment. It is the goal, however, of our modeling effort to verify the internal models in our process simulation and device analysis programs so that this will not be necessary.

• *Setting up the distributed model*

A general layout of a typical transistor is shown in **Figure 9**. In order to provide flexibility to the circuit designers, the MGP can omit either base or either collector contact, or extend the base metallization over the silicon.

The various equivalent circuits which make up the distributed model have already been described in the second section and Fig. 2. These equivalent circuits are laid out and connected in a large array, as illustrated in **Figure 10**, and then analyzed using the ASTAP circuit analysis program. This layout can make use of bilateral or quadrilateral symmetry of the transistor to reduce computation time. The integers which specify the numbers of elementary models in the distributed array are adjustable, with the exception of the sidewall transistor. Since the sidewall transistor was characterized as a single unit, the rows of sidewall transistors cannot be replicated in the same manner as the other elements. Typically we use a 5 × 5 array for the intrinsic region of a small-geometry transistor. The total distributed array typically consists of about 1000 to 2000 components, including those in the equivalent-circuit models. The CPU time for the analysis of this array is approximately 5 to 20 minutes.

• *Running the distributed model*

The distributed model is run in forward and inverse dc mode with a range of bias from 0.6 to 1.0 volt applied to the base-emitter or base-collector of the distributed network. This covers the operating points of interest to circuit designers, and also covers the range where high current phenomena occur.

The distributed network can also be run in ac or transient mode to compare with, and verify that, the final lumped model gives accurate transient results; however, this step is not necessary to generate models.

• *Extracting parameters from the distributed model*

The ASTAP analysis of the distributed model is carried out for a range of forward- and inverse-mode bias points. At this point we are faced with the same problem that existed at the completion of the two-dimensional analysis program (2DP). It is necessary to extract useful information from the distributed model in order to set up the final lumped-model equivalent circuit. The lumped model must represent a major simplification over the distributed model, since it will be used in circuit simulations with hundreds of transistors and other elements.

The equivalent circuit generated by MGP-B for the final lumped model is shown in **Figure 11**. The choice of the transport form, as discussed in the second section, is motivated by similar considerations. The physical relation between the elements of this equivalent circuit and regions of the transistor is illustrated schematically in **Figure 12**.

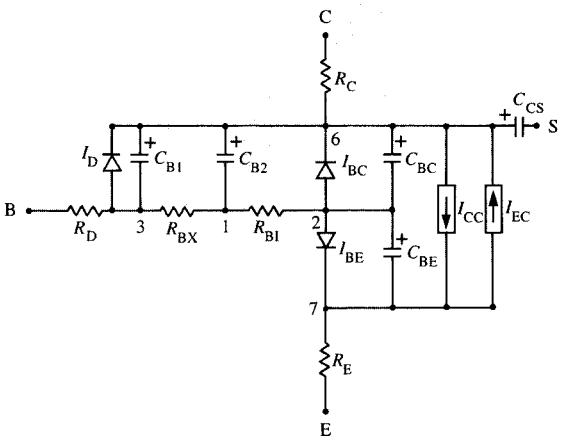

Figure 11

Final lumped equivalent circuit model: Note that the transport formulation, Fig. 3, is used here.

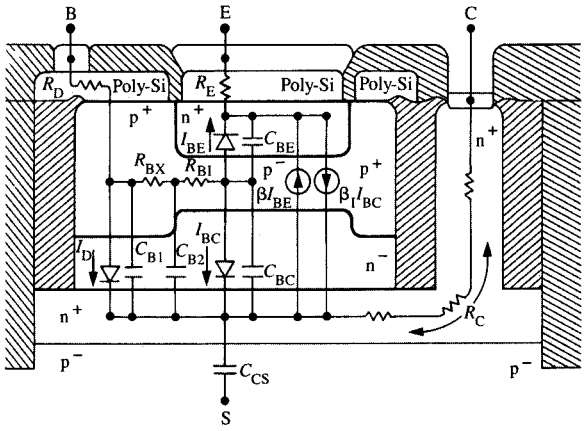

Figure 12

Physical meaning of equivalent circuit elements: The equivalent circuit model is superimposed on a schematic transistor cross section (not to scale).

The analysis of the distributed model to obtain values for the lumped-model elements makes as much use as possible of internal information from the distributed model. For example, the V_{BE} vs I_B output from the distributed model is least-square fit to obtain the saturation current and ideality factors that characterize the base-emitter diode. The situation is better here than when experimentally measuring a transistor since we have access to the internal information in the ASTAP simulation.

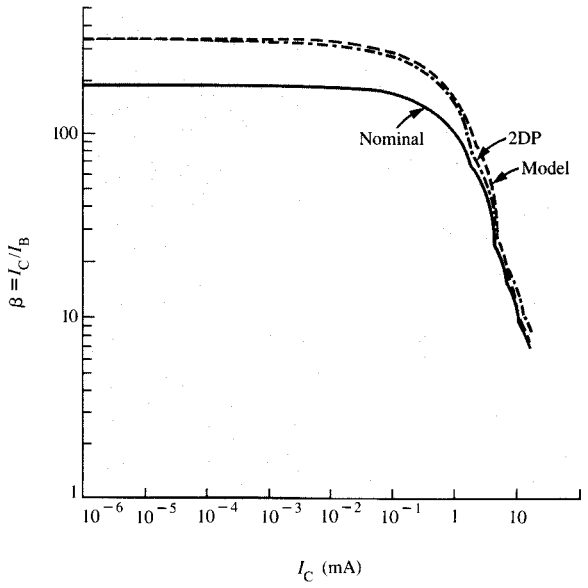

Figure 13

Effect of a base doping change on transistor beta: The curve labeled "Nominal" is a nominal beta generated by the MGP. The curve labeled "2DP" is generated by the MGP from an abbreviated set of 2DP runs for a profile with 30% less base doping. The curve labeled "Model" results from a linearized statistical variation in the final lumped model, as described in the text.

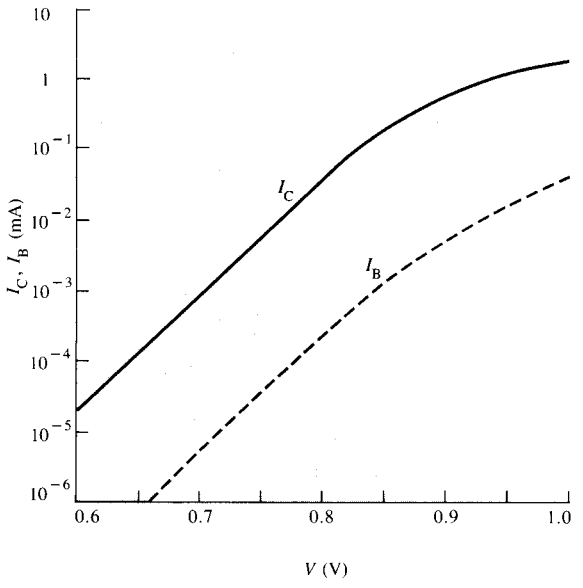

Figure 14

Lumped model base and collector current: This and the next two figures represent the characteristics of the lumped model generated by the MGP for a typical advanced transistor. The characteristics were generated using ASTAP to evaluate the model. In practice the circuit designer would combine many such models to obtain the performance of large circuits.

The corners of a real transistor are rounded, not square as assumed in setting up the distributed model. In addition, statistical variations resulting from image tolerance and mask alignment must be accounted for. These effects are introduced when making the transition from the distributed to the lumped model. At this stage all quantities in the distributed model are normalized by area or perimeter and then adjusted for the actual values in the lumped model. The principal corrections for a minimum-size transistor are to the area and perimeter of the emitter.

Extracting resistance values

The base current enters the intrinsic base from all sides, since the resistive extrinsic polysilicon base surrounds the emitter. In this situation not all portions of the boundary have the same potential. Therefore, in order to properly account for the resistance, we have chosen a power dissipation formulation whereby the equivalent resistance in the lumped model will dissipate the same power as the corresponding resistances in the distributed model. Thus the power dissipated in each transistor region is divided by the square of the current flowing into that region to find the equivalent resistance. These resistance values are then placed in a table as a function of electron current density for the final ASTAP equivalent-circuit model in Fig. 11.

Final capacitance formulation

The capacitance enters the lumped model in two ways. First, the voltage-dependent depletion capacitance is determined by multiplying the expression in Eq. (2) by the appropriate junction area. In principle, the sum of the depletion capacitances could be used. There is little inaccuracy introduced by not doing this, since at low bias, where depletion capacitance dominates, all the elements have essentially the same voltages.

The current-dependent neutral component of the capacitance is derived by summing all of the corresponding capacitances in the distributed network. A table based on the appropriate current density is then created for the ASTAP lumped model.

• *Examples of process optimization*

The procedure of generating an ASTAP transistor model based on two-dimensional process and device simulation makes it possible to evaluate the circuit impact of process-related perturbations. A new model can be generated for a perturbed process with a set of only 12 bias points from the 2DP program. For example, **Figure 13** illustrates the increase in transistor beta which results from a 30% decrease in the integrated base doping. Note that at high currents the percentage increase becomes much less. This effect can be

understood as a manifestation of base pushout, the Kirk effect ([6] and Figs. 5 and 6). At high current densities the injected holes and electrons dominate the built-in behavior related to the transistor profile. The lumped model can be made to have a statistical dependence on integrated base doping by linearizing the low current increase in beta and introducing a rolloff factor which becomes small at high current densities. For example,

$$\beta = \beta_0(1 + a \cdot f \cdot d), \tag{3}$$

where a is a constant, f is the rolloff factor, and d is the change in integrated base doping. In the ASTAP program d can be made a random variable.

In a similar manner all components of the equivalent circuit in Fig. 11 can be made to have a statistical dependence on integrated base doping, epitaxial concentration, epitaxial depth, sidewall thickness, and mask alignment and image tolerance.

• *Model validity*

The lumped model can be checked in two different ways. First, a model for a purely two-dimensional hypothetical transistor can be generated. When this is done, the resulting dc characteristics and capacitance values agree extremely well with the 2DP program from which the model was generated. This gives us confidence that the extension to realistic three-dimensional structures is accurate.

The second method is to compare the lumped model with a variety of advanced experimental transistors. There are many process uncertainties and theoretical assumptions that have gone into the model generation procedure. In general the comparison with experiment gives good agreement. Reference [2] discusses efforts which are being made to evaluate and improve this agreement by examining the underlying physics which is incorporated into the 2DP program. At this time it appears that we are near our goal of a complete set of programs with a predictive modeling capability.

Note that even without detailed agreement with experiment, our predictive modeling procedure would still be extremely valuable in evaluating the differences in circuit performance which would result from hypothetical process or geometry changes.

Conclusions

The output of the lumped equivalent-circuit model generated by the MGP for a simulated advanced transistor is illustrated in **Figures 14** through **16**, showing the *I-V* curves, the total base resistance, and the total emitter and collector capacitances.

We would like to emphasize the importance of the Model Generation Program (MGP). This program is released to product designers in lieu of a transistor specification for the design and evaluation of integrated circuit chips. Thus the MGP is not the last step. It is followed by a hierarchy of design and modeling efforts on a circuit level, a system level, and even a machine level. These efforts, however, are beyond the scope of this group of four papers.

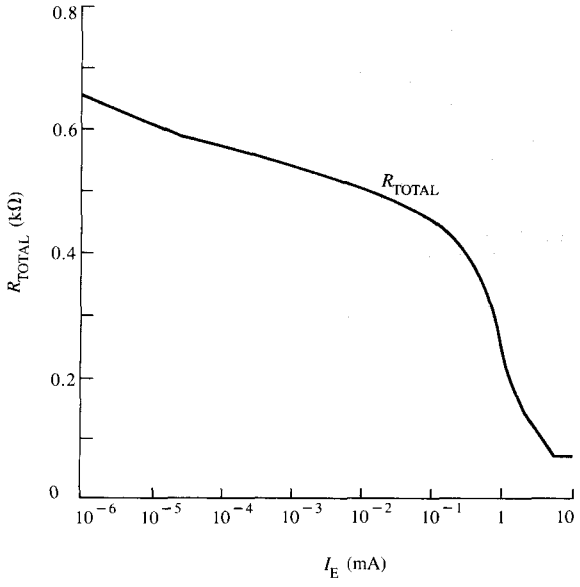

Figure 15

Lumped model total base resistance: This is the sum of R_D, R_{BX}, and R_{BI} in Fig. 11.

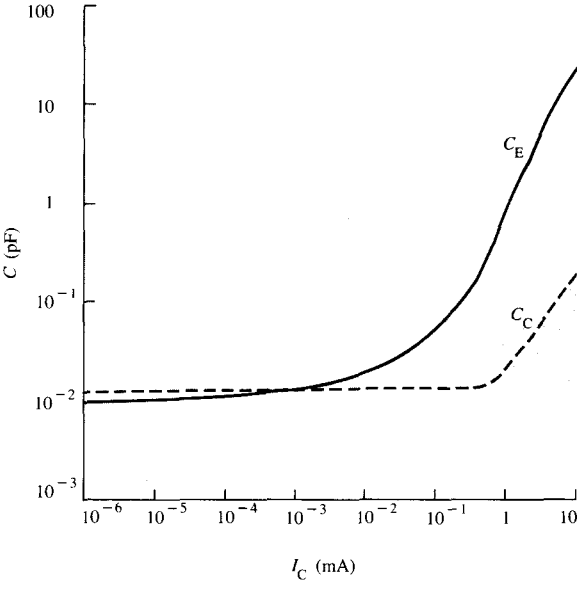

Figure 16

Lumped model base-emitter and base-collector capacitance: The base-collector capacitance is the sum of C_{BI}, C_{B2}, and C_C in Fig. 11.

In this paper we have discussed the third step in the evolution of transistor models generated in a predictive fashion from process modeling with SAFEPRO, through two-dimensional device physics modeling with 2DP, to the quasi-three-dimensional device modeling of MGP. These techniques have been extremely valuable in our advanced bipolar technology development.

References and note

1. R. R. O'Brien, C. M. Hsieh, J. S. Moore, R. F. Lever, P. C. Murley, K. W. Brannon, G. R. Srinivasan, and R. W. Knepper, "Two-Dimensional Process Modeling: A Description of the SAFEPRO Program," *IBM J. Res. Develop.* **29**, 229–241 (1985, this issue).
2. S. P. Gaur, P. A. Habitz, Y.-J. Park, R. K. Cook, Y.-S. Huang, and L. F. Wagner, "Two-Dimensional Device Simulation Program: 2DP," *IBM J. Res. Develop.* **29**, 242–251 (1985, this issue).
3. R. W. Knepper, S. P. Gaur, F.-Y. Chang, and G. R. Srinivasan, "Advanced Bipolar Transistor Modeling: Process and Device Simulation Tools for Today's Technology," *IBM J. Res. Develop.* **29**, 218–228 (1985, this issue).
4. This plot was generated using a program written by P. C. Murley.
5. J. J. Ebers and J. L. Moll, "Large-Signal Behavior of Junction Transistors," *Proc. IEEE* **42**, 1761–1772 (1954).
6. C. T. Kirk, "A Theory of Transistor Cutoff Frequency (ft) Fall-off at High Current Density," *IEEE Trans. Electron. Devices* **ED-9**, 164–174 (1962).

Received October 1, 1984; revised November 12, 1984

Fung-Yuel Chang *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Chang has been with the IBM General Technology Division in East Fishkill since 1968. He is currently a Visiting Associate Professor of Electrical Engineering at Columbia University, where he teaches courses on bipolar device modeling and computer-aided circuit analysis. Prior to his sabbatical leave from IBM, he was a senior engineering manager of the Bipolar Device Design and Modeling Department. Dr. Chang received a B.S. in electrical engineering from National Cheng Kung University, Taiwan, and an M.S. in electronics from the National Chiao Tung University, Taiwan. In 1964 he earned an M.S. and in 1968 an Eng.Sc.D. degree in electrical engineering from Columbia University. At IBM he has taught courses on bipolar device modeling; his areas of interest include distributed RC networks, coupled transmission lines, integrated circuits, and computer-aided circuit design. Dr. Chang is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

Lawrence F. Wagner *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Wagner is an advisory engineer in the Device Design and Modeling Department. He joined the IBM General Technology Division in East Fishkill in 1981 and has since worked on the modeling of Schottky barrier diodes and bipolar transistors. Dr. Wagner received a B.S. from the Massachusetts Institute of Technology in 1960. From 1961 to 1968 he worked for the U.S. Army Electronics Command in Fort Monmouth, New Jersey, investigating the design of micropower logic circuits. Subsequently, as a National Science Foundation Fellow, he received an M.S. in 1969 and a Ph.D. in 1981 from Stanford University, California. His doctoral research included the first direct observation of silicon surface states using ultraviolet photoelectron spectroscopy. He has also completed a postdoctoral assignment at the University of Hawaii, where he modeled the angular dependence of X-ray photoelectrons. Dr. Wagner is a member of the American Physical Society, the Institute of Electrical and Electronics Engineers, and Sigma Xi.