Journal of ELECTRICAL ENGINEERING, VOL. 51, NO. 1-2, 2000, 21–29

A NOVEL EXTRACTION METHOD FOR BJT–PARAMETERS

Róbert Lórencz [*] — **Christian Reckleben** [**] — **Karsten Hansen** [**]

We report on a reliable parallel parameter extraction method of the Spice Gummel-Poon (GP) model. It is based on a non-linear least-squares fit of the measured base- and collector dc-currents and results in the optimal parameter set. The goodness of the fit, a level of confidence in each parameter, the degree of coupling between parameters and the sensitivity of the model with respect to each parameter are calculated.

K e y w o r d s: BJT, Gummel-Poon model, parameter extraction, error-free computation

1 INTRODUCTION

The accuracy of circuit simulations depends on a reliable, robust and unambiguous parameter extraction method. It is as important as accurate models [1] and the parameters should generally be extracted from measured transistor data.

The determination of bipolar transistor parameters with respect to the Spice Gummel-Poon model is usually performed sequentially. Single parameters or small groups of parameters (typically not more than 3) are determined under certain bias conditions of the device [2]. Only a small subset of model equations and data from a restricted area of a device's operating range are used. This extraction scheme does not account for the interactions of parameters or for the effect of parameters in bias regions other than that from which they were obtained. Incorrect extracted parameters will affect others and an error may propagate.

Thus, for example, the parameters of the inner transistor are separated from its series resistances of the base, collector and emitter [3–5]. Hence, the correlation of high-level injection effects with the series resistances is ignored and both effects are certainly overestimated.

AC methods [4] to obtain the series resistances requires additional measurements which are generally very sensitive to parasitic elements of the measurement setup. Thus, using these methods, complicated calibration schemes are involved and the large-signal resistances have to be calculated from the S- or Y-parameters [3].

The purpose of this paper is to introduce a new reliable parallel-extraction method for Spice Gummel-Poon parameters, which overcomes the previous marked tradeoffs. Easy to perform collector and base current dc-measurements are done over a large range of bias conditions. A non-linear least-squares fit of the device characteristics to the model results in the optimal parameter set. This is performed by a highly efficient optimization technique, based on an algorithm proposed by Ward and Doganis [6] for MOS model parameter extraction. Moreover not only the goodness of the fit is calculated, but also a level of confidence for each parameter and the degree of coupling between parameters are evaluated.

The main effort here is the optimized fit of the model to the real device behaviour in order to get the optimal and most realistic circuit simulation fascilities. Nonetheless the values extracted in this manner will have physical relevance, if a physical model is used as in our case.

The paper is organized as follows. Section 2 briefly describes the Gummel-Poon model. Section 3 gives an outline of the extraction algorithm together with the interpretation of the estimates. Section 4 presents results obtained with this approach. Section 5 gives the conclusions derived from this work.

2 THE MODIFIED GUMMEL–POON MODEL

The Gummel-Poon (GP) model [7] in its Spice implementation is still the most often used bipolar-junction transistor (BJT) model in today's bipolar circuit design. This section gives a comprehensive view of the model's DC-characteristic [8].

Based on the bias-dependent charge distribution of the majority carriers Q_B in the base (normalized with the zero-bias base charge), the collector current density i_c through the one-dimensional base is calculated as follows:

$$i_c = i_{ce} - i_{bc} = \frac{isA_n}{Q_B}\left(e^{\frac{V'_{BE}}{nfV_T}} - e^{\frac{V'_{BC}}{nrV_T}}\right) - \frac{isA_n}{br}\left(e^{\frac{V'_{BC}}{nrV_T}} - 1\right) - iscA_n\left(e^{\frac{V'_{BC}}{ncV_T}} - 1\right), \quad (1)$$

[*] Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University, Karlovo nám. 13, 121 35 Praha 2, Czech Republic

[**] Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, D-22 607 Hamburg, Germany

ISSN 1335-3632

with the bias dependent

$$Q_B = \frac{Q_1}{2}\left(1+\sqrt{1+4Q_2}\right), \tag{2}$$
$$Q_1 = \left(1-\frac{V'_{BE}}{var}-\frac{V'_{BC}}{vaf}\right)^{-1},$$
$$Q_2 = \frac{is}{ikf}\left(e^{\frac{V'_{BE}}{nfV_T}}-1\right)+\frac{is}{ikr}\left(e^{\frac{V'_{BC}}{nrV_T}}-1\right)$$

the thermal voltage $V_T = kT/q$ and the normalized emitter area ratio A_n. The voltages V'_{BE} and V'_{BC} are the base-emitter (BE) and base-collector (BC) voltage drops of the inner transistor, see Fig. 1. These voltages differ from terminal voltages due to ohmic drops at the intrinsic series resistances of the emitter, collector and base. The 17 model parameters which are needed for characterizing the static BJT behaviour are included in Table 1.

Q_1 models the increase of majority carrier charge by the shift of the base-side space-charge layer edges from their zero-bias conditions due to increasing V_{BE} and V_{BC}. This is related to the Early effect, which leads to a non-zero slope of i_c in the $i_c - V_{CE}$ output characteristic.

High current-injection effects are incorporated via Q_2. That is the increase in the majority carrier charge required to neutralize the minorities injected into the base from the emitter and collector regions. So Q_2 reduces the current gain $\beta = i_c/i_b$ at high collector currents (Webster effect).

The second essential model equation predicts the base current density:

$$i_b = i_{be}+i_{bc} = \frac{isA_n}{bf}\left(e^{\frac{V'_{BE}}{nfV_T}}-1\right)+iseA_n\left(e^{\frac{V'_{BE}}{neV_T}}-1\right) + \frac{isA_n}{br}\left(e^{\frac{V'_{BC}}{nrV_T}}-1\right)+iscA_n\left(e^{\frac{V'_{BC}}{ncV_T}}-1\right). \tag{3}$$

The BE current i_{be} is modeled by an ideal diode ($nf = 1.0$) in parallel with a non-ideal diode ($ne > 1$). The latter links the generation/recombination process in the base-emitter space charge region and the model. This view can be used equivalently for the BC current i_{bc}.

Additionally the internal series resistances R_E, R_C and R_B of each transistor terminal are considered. Emitter and collector resistances R_E, R_C are regarded as simple, linear elements, whereas the base resistor is bias dependent. Emitter current crowding effects and the conductivity modulation of the base contact region at high-level injection can be modeled using one of the following equations:

$$R_B = rbm + \frac{rb-rbm}{Q_B} \tag{4}$$

or

$$R_B = rbm + 3\,(rb+rbm)\,\frac{\tan(z)-z}{z\,\tan^2(z)}, \tag{5}$$

where z is a function of the base current i_b. Both (4) and (5) yield rb for low base currents and tend towards rbm for increasing i_b. Typically $rbm < rb$.

Table 1. DC-parameters of the GP model

nf	forward emission coeff.	–
is	saturation current	A/μm
bf	ideal forward current gain	–
vaf	forward early voltage	V
ikf	corner for forward current gain roll-off	A/μm
ne	B–E leakage current emission coeff.	–
ise	B–E leakage saturation current	A/μm
nc	B–C leakage current emission coeff.	–
isc	B–C leakage saturation current	A/μm
nr	reverse emission coeff.	–
br	ideal reverse current gain	–
var	reverse early voltage	V
ikr	corner for reverse current gain roll-off	A/μm
rb	zero-bias base resistance	Ω
rbm	minimum base resistance	Ω
R_C	collector resistance	Ω
R_E	emitter resistance	Ω

Since the emitter area is reduced due to further device scaling, the series resistances are increasing. For that reason the influence of the resistances on the device behaviour and also on parameter extraction must not be neglected. Particularly at high level currents the ohmic drops are the sole cause of i_c and i_b deviating from the ideal model (1) and (3).

Thus the current-voltage relationships $i_c = f(V_{BE}, V_{BC})$ and $i_b = f(V_{BE}, V_{BC})$ of the BJT terminals is modeled by a set of two implicit exponential-equations.

3 EXTRACTION METHOD

A. The optimization method

The measured data samples for base and collector currents $I_{B,i}$ and $I_{C,i}$ are fitted during the extraction process to the model equations $i_b(\boldsymbol{x}_i, \boldsymbol{p})$ (3) and $i_c(\boldsymbol{x}_i, \boldsymbol{p})$ (1), by optimizing the parameter vector $\boldsymbol{p}$. For that, the relative error

$$f_i(\boldsymbol{p}) = \frac{i_c(\boldsymbol{x}_i,\boldsymbol{p}) - I_{C,i}}{\max(I_{C,i}, I_{C,\min})}$$
$$f_{i+n}(\boldsymbol{p}) = \frac{i_b(\boldsymbol{x}_i,\boldsymbol{p}) - I_{B,i}}{\max(I_{B,i}, I_{B,\min})} \tag{6}$$
$$\text{for } \; i = 1,\ldots,n$$

is calculated in each data point i. These nonlinear residuals are included in the $2n$-dimensional weighed residual vector $\boldsymbol{f}(\boldsymbol{p})$. The $\boldsymbol{x}_i$ are vectors of the applied terminal voltages as independent variables and n is the number of measured data samples. $I_{C,\min}$ and $I_{B,\min}$ are weights

provided by the user. At currents above $I_{C,\min}$ or $I_{B,\min}$ the relative error is used, otherwise the absolute error is used, which is scaled by $I_{C,\min}$ or $I_{B,\min}$.

The task of the optimization algorithm is minimizing the nonlinear least squares function

$$\Phi(\boldsymbol{p}) = \|\boldsymbol{f}(\boldsymbol{p})\|^2 = \boldsymbol{f}^\top(\boldsymbol{p})\boldsymbol{f}(\boldsymbol{p}) = \sum_i^{2n}(f_i(\boldsymbol{p}))^2, \quad (7)$$

with respect to the g-dimensional vector $\boldsymbol{p}$. This minimizes the rms error in the fit of the model to the measured data.

At each optimization step during the extraction process, the model is evaluated at each data point i and the vector $\boldsymbol{f}(\boldsymbol{p})$ is calculated. The parameter vector $\boldsymbol{p}$ is then adjusted to reduce $\|\boldsymbol{f}(\boldsymbol{p})\|^2$. The modified Marquardt algorithm [nas] is used for the adjustment of $\boldsymbol{p}$ to minimize $\Phi(\boldsymbol{p})$ in the optimization process.

Minimizing a nonlinear least squares function of several variables is based on a Taylor series expansion of the error vector $\boldsymbol{f}(\boldsymbol{p})$:

$$\boldsymbol{f}\left(\boldsymbol{p}^{k+1}\right) \approx \boldsymbol{f}\left(\boldsymbol{p}^{k}\right) + \mathbf{J}\boldsymbol{\Delta}\boldsymbol{p}^{k}. \quad (8)$$

Where

$$\boldsymbol{\Delta}p^k = \boldsymbol{p}^k - \boldsymbol{p}^{k+1}$$

is the updated difference vector for the parameter vector $\boldsymbol{p}^k$ at the k-th step of optimization and

$$\mathbf{J} = \left[\boldsymbol{\nabla}_p \boldsymbol{f}^\top(\boldsymbol{p})\right]^\top$$

is the Jacobian matrix $(2n \times q)$ of the residual vector $\boldsymbol{f}(\boldsymbol{p})$ with elements $J_{i,j} = \partial f_i(\boldsymbol{p})/\partial p_j$. $\boldsymbol{\nabla}_p$ is the matrix operator of the first derivative with respect to the elements of $\boldsymbol{p}$.

Linearization of the error vector $\boldsymbol{f}(\boldsymbol{p})$ by a Taylor series expansion involves a systematic error into the iteration process. This error is due to the truncation of the Taylor vector series and becomes negligible for small $\boldsymbol{\Delta}p^k$ at the end of the extraction process.

The modified Marquardt algorithm for nonlinear least squares combines the steepest descent and Gauss-Newton method [6], [9]. The updated parameter vector $\boldsymbol{p}^{k+1}$ at the $(k+1)$-th iteration step is given by:

$$\boldsymbol{p}^{k+1} = \boldsymbol{p}^k + (\mathbf{J}^\top\mathbf{J} + \lambda^k(\mathbf{J}^\top\mathbf{J})_{jj})^{-1}\mathbf{J}^\top\boldsymbol{f}\left(\boldsymbol{p}^k\right), \quad (9)$$

where the λ^k is a non-negative real constant.

The parameter vector $\boldsymbol{p}$ is considered to be optimal, when

$$\frac{|p_j^k - p_j^{k+1}|}{|p_j^{k+1}|} < \epsilon, \quad j \in \{1, 2, \ldots, q\}, \quad (10)$$

where $\epsilon = 10^{-5}$. This is the stoppage criterion of the optimization process.

The iteration process strongly depends on errors in the numerical calculation of the Jacobian and on the matrix inversion of the approximation used in (9). The equation set (9) is typically ill-conditioned and therefore a potential source of errors, which slows down the convergence process or makes it completely impossible. To avoid numerical errors in solving a linear equation set, the error-free calculation based on finite number residue arithmetic [10] is used, see Appendix.

B. Interpretation of the estimates

Computing only the optimal parameter vector $\boldsymbol{p}^{ex}$ as a result of minimizing the error function $\Phi(\boldsymbol{p})$ cannot state anything about its reliability and precision. Ill-determined parameters, those which are strongly correlated and have a large uncertainty, are potential sources of errors in the extraction process and lead to ambiguous parameter sets.

The statistical attributes of each extracted parameter p_j^{ex} are calculated at the end of the fitting procedure using estimation theory. This additional information allows the assessment of the extracted parameters about the accuracy and dependences of the parameters. All this information contains the covariance matrix $\mathbf{V}$ of the estimated parameter vector $\boldsymbol{p}^{ex}$.

The approximate covariance matrix in the case of nonlinear problems has different forms depending upon the estimation method. But in general the expressions are similar to comparable linear estimation cases. For negligible change ($\epsilon = 10^{-5}$) in any parameter, nonlinear estimation is reduced to linear estimation

$$\mathbf{J}^\top\mathbf{J}\boldsymbol{\Delta}\boldsymbol{p}^{ex} = \mathbf{J}^\top\boldsymbol{f}(\boldsymbol{p}^{ex}). \quad (11)$$

In this case the covariance matrix $\mathbf{V}$ (dim. $q \times q$) of $\boldsymbol{p}^{ex}$ is given [11]:

$$\mathbf{V} = \mathrm{cov}(\boldsymbol{p}^{ex}) = (\mathbf{J}^\top\mathbf{J})^{-1}\mathbf{J}^\top\boldsymbol{\Psi}\mathbf{J}(\mathbf{J}^\top\mathbf{J})^{-1}$$

Assuming the elements in $\boldsymbol{f}(\boldsymbol{p})$ are noncorrelated and have constant variances in both currents $I_{C,i}$ and $I_{B,i}$, then the covariance matrix $\boldsymbol{\Psi}$ of the weighed error vector $\boldsymbol{f}(\boldsymbol{p}^{ex})$ can be expressed as a diagonal matrix with the elements

$$\Psi_{ii} = \frac{2\boldsymbol{f}^\top(\boldsymbol{p}^{ex})\boldsymbol{f}(\boldsymbol{p}^{ex})}{2n-g} = \begin{cases} \sigma_{I_C}^2, & i = 1, \ldots, n \\ \sigma_{I_B}^2, & i = n, \ldots, 2n, \end{cases} \quad (12)$$

where σ_{I_C} and σ_{I_B} are the standard deviations of the relative errors $f_i(\boldsymbol{p})$.

A figure of the correlation is given by the elements of the correlation matrix $\mathbf{C}$:

$$C_{jl} = \frac{V_{jl}}{\sqrt{V_{jj}}\sqrt{V_{ll}}}. \quad (13)$$

Along with an estimation of the parameter p_j^{ex} a confidence interval give some indication of how far the estimate may be expected to be from the true parameter p_j^{tr}. The confidence interval for p_j^{tr} is expressed as [11]

$$p_j^{tr} \in \langle p_j^{ex} - \delta p_j^{ex} \, , \, p_j^{ex} + \delta p_j^{ex} \rangle \, , \tag{14}$$

where

$$\delta p_j^{ex} = t_{1-\varphi,\mu} \sqrt{V_{jj}} \, ,$$

with $t_{1-\varphi,\mu}$ being the Student t distribution evaluated at some confidence level φ (ie 95 %) and $\mu = n - g$ the number of degrees of freedom. The estimation of the parameter p_j^{ex} we got from the minimization of $\Phi(\boldsymbol{p})$ is such that the true parameter p_j^{tr} is within the confidence interval with $\varphi = 95\,\%$ probability. The uncertainty U_{p_j} of p_j^{ex} is defined as [12]

$$U_{p_j} = \frac{\delta p_j^{ex}}{|p_j^{ex}|} \cdot 100 \; [\%] \, . \tag{15}$$

The uncertainty of a parameter is also a valuable measure for the quality of that model part in which the parameter is used.

At the end of the optimization process the mean normed sensitivity coefficients of the optimal parameters are calculated:

$$\begin{aligned} S_{p_j}^{i_c} &= \frac{1}{n} \sum_i^n \Big| \frac{\partial(i_c(\boldsymbol{x}_i, \boldsymbol{p}))}{\partial p_j} \frac{p_j}{i_c(\boldsymbol{x}_i, \boldsymbol{p})} \Big| \\ S_{p_j}^{i_b} &= \frac{1}{n} \sum_i^n \Big| \frac{\partial(i_b(\boldsymbol{x}_i, \boldsymbol{p}))}{\partial p_j} \frac{p_j}{i_b(\boldsymbol{x}_i, \boldsymbol{p})} \Big| \, . \end{aligned} \tag{16}$$

$S_{p_j}^{i_c}$ and $S_{p_j}^{i_b}$ give the sensitivity of the collector and base current model equations (1) and (3) with respect to parameter perturbations. $S_{pj} = \frac{1}{2}(S_{pj}^{i_c} + S_{pj}^{i_b})$ is the mean sensitivity of the complete model with respect to the parameter p_j. The parameters with a high sensitivity coefficient should feature a small uncertainty.

Strong correlation of parameters hampers unambiguous determination of individual parameters. Therefore ill-determined parameters are set to a default value and the data are refitted using the remaining parameters. For this, the worst parameter is chosen by comparing the uncertainties and sensitivities of the strongly correlated parameters. That one with the largest uncertainty together with a small sensitivity is set to a default value. The sensitivity coefficient is taken into account in order to avoid an elimination of an important parameter which is strongly correlated to an uncertain one. Thus any bad effects the ill-determined parameters might have had are removed and the remaining parameter set better characterizes the measured device and has a physical relevance.

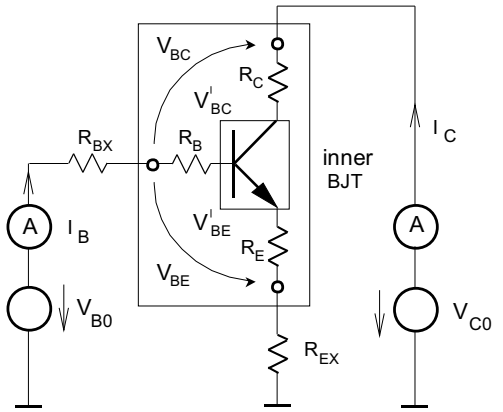

Fig. 1. Configuration used to perform the DC characteristics of the BJTs

4 EXTRACTION OF BJT–PARAMETERS

A. Structures and Measurement

To verify the proposed method 12 GHz f_T poly-emitter npn transistors have been used which are fabricated with an 0.8 μm BiCMOS process. Three structures with drawn emitter dimensions ($W_E \times L_E$) of $0.8 \times 3\,\mu\text{m}^2$ (npn1), $0.8 \times 45\,\mu\text{m}^2$ (npn2) and $0.8 \times 60\,\mu\text{m}^2$ (npn3) are used. The emitter length L_E is specified as the normalized area ratio $A_n = L_E$. The BJT npn1 has single base, collector and emitter contacts, whereas the other possess interdigitated multiple emitter and base strips located within double collector contacts. In order to get reliable contact resistances to perform homogeneous and reproducible measurements all transistors are bonded in standard IC packages.

The DC characteristics of base and collector currents are performed with the semiconductor analyzing system HP 4156A. In order to get a meaningful parameter estimation the measured data should result in a low uncertainty U_p together with a relevant importance S_p of each parameter p_j. To avoid high-current stress degradation effects [13] like instabilities in the forward current gain and emitter resistance degradation, the maximum emitter current density is below $1.25\,\text{mA}/\mu\text{m}^2$. These effects are not implemented in the Spice/Spectre model and therefore falsify the extracted parameters. Figure 1 depicts the measurement structure and introduces the utilized voltages.

In practice it is more valuable to use V_{B0} as an independent variable instead of I_B because of more reliable (small uncertainty) extraction results for the series resistances. Nonetheless, for the extraction tool it is not relevant whether constant currents or voltages are used as independent variables. It only uses a data vector in each measurement point, which contains I_B, I_C, V_{B0} and V_{C0} as input.

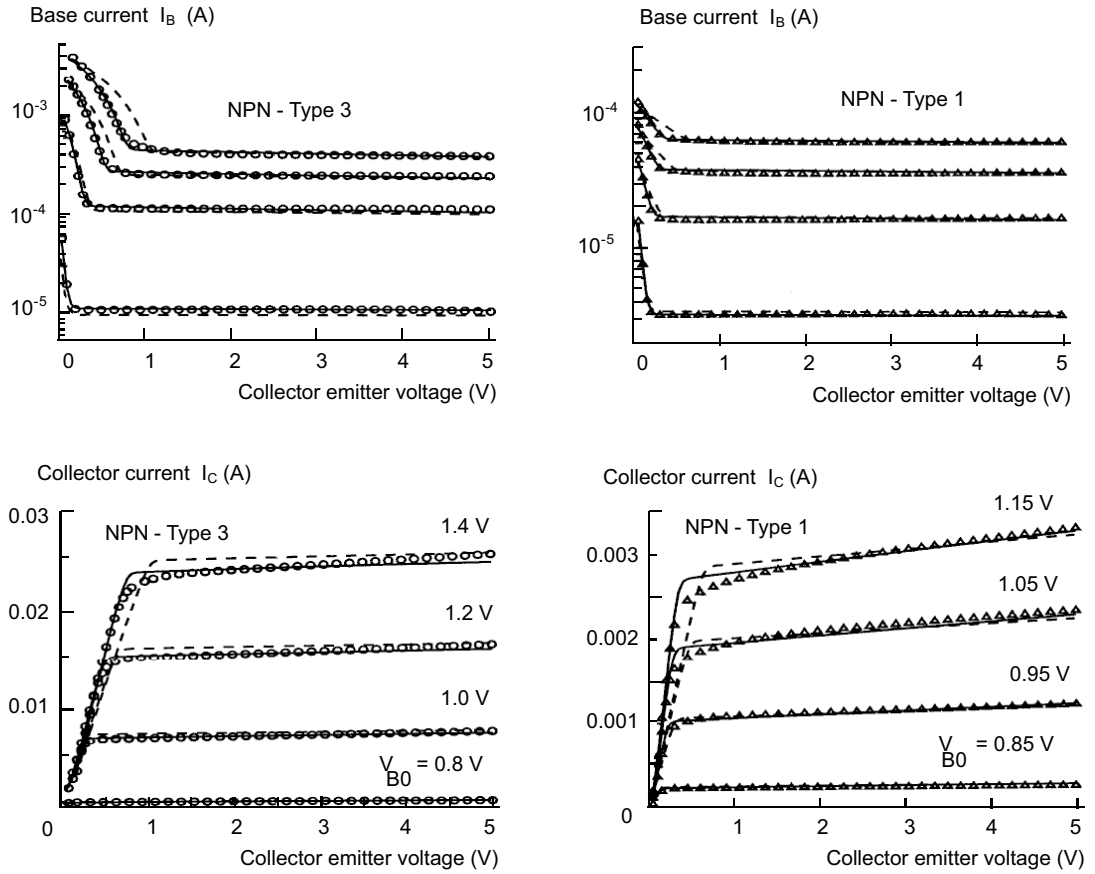

Fig. 2. Comparison of the measured (∘) output characteristic with calculated curvatures using the default parameter set (dashed line) and the extracted parameter set (solid line) for two transistor structures npn1 and npn3.

Table 2. Typical Extraction Result of npn1.

Number of datapoints = 990						*npn1*								
$\sigma_{tot} = 0.0261$ $\sigma_{I_C} = 0.0285$ $\sigma_{I_B} = 0.0235$						*Correlation matrix*								
p	$Value$	$U_p\%$	S_p	$S_p^{i_c}$	$S_p^{i_b}$	is	bf	vaf	ikf	rb	R_E	R_C	ne	nr
is	1.395×10^{-18}	0.42	0.365	0.379	0.349									
bf	112.574	1.07	0.435	0.362	0.498	+0.094								
vaf	22.156	2.28	0.060	0.080	0.028	+0.505	+0.185							
ikf	1.294×10^{-3}	1.95	0.145	0.190	0.076	−0.068	−0.907	+0.060						
rb	658.67	9.41	0.056	0.046	0.064	+0.291	−0.150	+0.115	+0.164					
R_E	7.912	1.99	0.253	0.217	0.284	−0.211	−0.038	−0.111	+0.013	−0.964				
R_C	42.34	2.61	0.052	0.069	0.026	+0.128	−0.015	+0.159	+0.102	+0.259	−0.258			
ne	1.618	0.48	0.613	0.259	0.828	+0.195	−0.917	+0.004	+0.816	+0.180	+0.009	+0.041		
nr	1.045	2.83	0.069	0.104	0.036	+0.059	−0.098	−0.035	+0.069	+0.588	−0.575	−0.311	+0.078	
nc	1.01	1.20	0.227	0.328	0.125	−0.070	+0.097	+0.028	−0.068	−0.589	+0.575	+0.344	−0.081	−0.997

The thermal voltage V_T is obtained by measuring the temperature T using a PT-100 sensor, thermally connected with the IC package.

The intrinsic device along with parasitic capacitances, resistances and inductances of the bond wires and socket leads result in oscillations of the base and collector currents at certain bias conditions. In the measured DC characteristics the oscillations result in a significant increase of the currents which eventually reaches the compliance of the analyzer rapidly.

To avoid the oscillations and stabilize the device, resistive loading of the input (external resistance tied to the base terminal in series R_{BX}) and negative feedback (external resistance tied to the emitter in series R_{EX}) are used. Both added resistors can easily be eliminated during the extraction process.

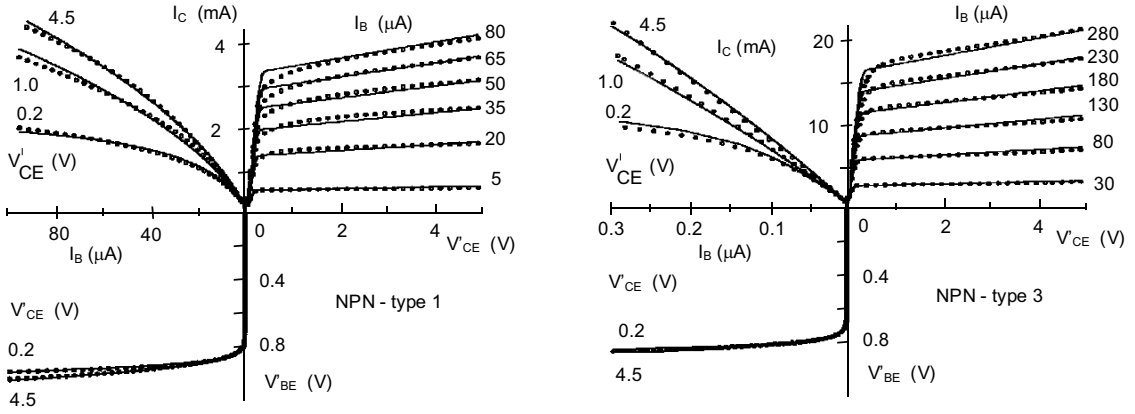

Fig. 3. Output and input characteristics of npn1 and npn3. Measured data (∘) using base current I_B as independent variable and simulated data (solid line) using the same extracted parameter set as in Fig. 2. The external resistors are eliminated from the data.

B. Extraction Example and Discussion

In the first example the extraction of 10 parameters $is, bf, ne, vaf, ikf, rb, R_E, R_C, nr$ and nc is demonstrated. The measured BJT output characteristics cover the whole operating range ($0.8V < V_{B0} < 1.4V$ and V_{C0} is swept from $0.1V$ to $5V$) in which the device will be used for analog IC circuit design, see Fig. 2. Note, that the characteristics in Fig. 2 are still affected by the presence of the external resistors R_{BX} and R_{EX}, thus the measured values V_{B0} and V_{C0} are used.

In this bias regime the reverse parameters br, var, ikr and isc are of little significance because of the almost reverse biased BC diode. The influence of the BC current is sufficiently adapted with nr and nc. Since $V'_{BE} > 0.7V$, the influence of the BE space charge recombination and with it ise is also negligible. $nf = 1.0$ is assumed because of an abrupt BE junction, the other parameters are set to default values.

Figure 2 compares the measured base and collector currents (∘) with the modeled currents using a default parameter set (broken line) as initial values for the optimization process and using the extracted parameter set (solid line) for the test structures npn1 ($\sigma_{tot} = 2.6\%$) and npn3 ($\sigma_{tot} = 3.3\%$). The wide bias range gives the extracted parameter set a global range of validity. In total about 1000 data points of each transistor are used. Only a small portion of them is displayed in Fig. 2. These examples require $k = 8$ Marquardt iteration steps to obtain the convergence criterion (10). The improvement due to the optimization is clear.

The right half of Table 2 gives the corresponding statistical information of the measurement of npn1. The correlation matrix contains the dependences of the parameters in the BJT model for this specific data set. It clarifies the need of a parallel extraction process for the inner transistor parameters and the series resistances, since their correlation cannot be neglected. Not even the reverse (nr and nc) and the forward parameters of the inner transistor are strictly linear independent. This is only true for $V_{BC} = 0$ — just one bias condition in the global range of interest.

Note that the correlation of nr and nc is almost 1 in this bias regime. Since $is/br = 1.52 \times 10^{-18}\mathrm{A}/\mu\mathrm{m}$ and $isc = 1.6 \times 10^{-18}\mathrm{A}/\mu\mathrm{m}$ have a small difference only, both parts of the base collector current i_{bc} (Equ. (3)) have nearly the same influence on i_b and i_c. Thus the parameters nr and nc are almost linearly dependent and their impact on the model is nearly the same. In the case of a correlation coefficient next to 1, it is recommended (see chapter 3) to fix one of the concerned parameters and repeat the extraction process. Doing this and fixing $nr = 1.0$ (abrupt BC pn-junction) results in $nc = 1.097$, whereas the other parameters are not significantly affected. Decoupling can also be achieved by a selection of additional data points in the saturation region ($V'_{BC} > 0$), to emphasize the significant bias conditions for nr and nc. The first part of Table 2 contains the relative error σ_{tot} of the complete data set and the model. This error is divided into a part of the collector σ_{I_C} and a part of the base current σ_{I_B}. For each optimized parameter p_j the uncertainty U_p and the sensitivity coefficient S_p is given. The latter is also divided in its base and collector current parts.

Using the default values the rms deviation is 8.5 % ($\sigma_{I_C} = 10\,\%$ and $\sigma_{I_B} = 6.6\,\%$). This error is minimized during the extraction process and results in $\sigma_{tot} = 2.6\,\%$ for the optimal parameter set.

The main error occurs in the transition region (where $V'_{BC} \approx 0$) from the saturation into the linear region of the collector current i_c, see Fig. 2. It becomes significant for $V_{B0} > 1V$, and is -of stronger influence for small devices (npn1: $\sigma_{I_C\,\max} = 7.6\,\%$ and npn3: $\sigma_{I_B\,\max} = 5.1\,\%$). Thus fitting and parameter extraction only at the bias condition $V_{BC} = 0$ is more uncertain and unreliable and therefore should be avoided.

Table 3. Extraction Result for all Forward Parameters of npn1.

Number of datapoints = 3458						*npn1*								
$\sigma_{tot} = 0.0447$ $\sigma_{I_C} = 0.0518$ $\sigma_{I_B} = 0.0364$						*Correlation matrix*								
p	*Value*	$U_p\,\%$	S_p	$S_p^{i_c}$	$S_p^{i_b}$	is	bf	vaf	ikf	rb	R_E	R_C	ne	nr
is	1.428×10^{-18}	0.30	0.424	0.429	0.419									
bf	108.91	0.38	0.356	0.224	0.488	+0.717								
vaf	23.829	3.35	0.027	0.043	0.012	+0.325	+0.284							
ikf	1.33×10^{-3}	1.36	0.081	0.117	0.045	−0.355	−0.554	+0.246						
rb	593.86	6.87	0.031	0.024	0.037	+0.162	−0.079	+0.132	+0.065					
R_E	8.324	1.24	0.175	0.167	0.182	−0.116	−0.145	−0.110	−0.027	−0.920				
R_C	39.46	2.34	0.026	0.039	0.012	+0.061	−0.034	+0.147	+0.174	+0.508	−0.515			
ne	2.850	2.64	0.074	0.000	0.148	+0.029	−0.213	+0.016	+0.134	+0.027	+0.023	+0.017		
nr	1.002	0.01	0.327	0.386	0.254	−0.117	−0.099	−0.112	+0.034	+0.013	+0.022	−0.478	+0.046	
nc	1.114	0.05	0.101	0.073	0.129	+0.254	+0.151	+0.120	−0.062	+0.009	−0.002	+0.289	−0.05	−0.827
ise	$1.906 \times \cdot 10^{-16}$	12.98	0.014	0.000	0.028	+0.029	−0.201	−0.015	+0.126	+0.025	+0.022	+0.014	+0.996	−0.385

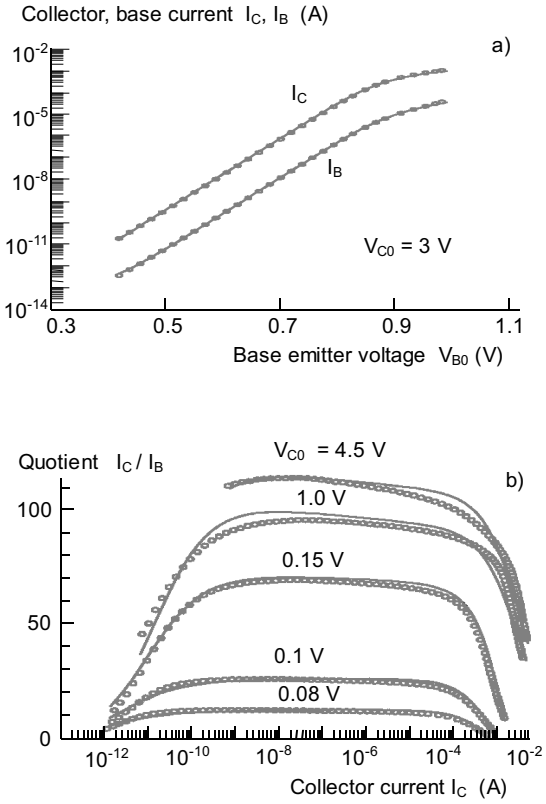

Fig. 4. Comparison of the measured characteristic (∘) with simulated curves using the extracted parameter set (solid line) for a) Gummel plot data and b) the current gain β. The external resistors R_{BX} and R_{EX} are eliminated.

Since constant R_C is used in the GP model, it is not possible to fit the transition region properly, where collector resistance modulation becomes relevant [14]. Using for example the more sophisticated inter-company model VBIC95, the onset of the quasi-saturation is modeled reasonably, just as the deep saturation behaviour. The VBIC95 models the R_C modulation in the intrinsic epitaxial region depending on the collector current and V'_{BC}. However it shows a wired, unphysical behaviour of the differential resistance [1].

The estimation of the series resistance is tested by adding successively two different test resistors at each transistor terminal in series, measuring the DC characteristics and extracting both data sets. Using this method, excellent accuracy of the extraction process for resistor parameters is demonstrated for differences below $1\,\Omega$ in R_E and R_C, whereas the accuracy of R_B is more than $10\,\Omega$.

For modeling the base resistance of npn1 (5) is used, whereas (4) is applied for the large types npn2 and npn3. This choice leads to the smallest uncertainty U_{rb} for each transistor type. Nevertheless, the estimation of rb is involved with the largest $U_p > 9\,\%$ of all parameters. This points out another limitation of the GP model. Equation (4) accounts only for emitter crowding effects and (5) is suitable only for base conductivity modulation. But, typically both effects together with base pushout in high level injection influence the base resistance dependence on bias [3]. However the importance of rb is sufficiently small ($S_{rb} < 0.055$), thus the possible inaccuracy of rb has no dramatical impact on the DC model.

Figure 3 compares the GP model (solid line) using again the above extracted parameter set with measured data (∘). In this case the data are measured with constant I_B as independent values instead of V_{B0} and the external resistors R_{BX} and R_{EX} are eliminated after measuring. The rms deviation of measurement and GP model has the same quality as in Fig. 2 ($\sigma_{tot} = 2.7\,\%$ for npn1 and $\sigma_{tot} = 3.3\,\%$ for npn3). Hence the external resistors have no influence on the extraction process.

C. Extraction of all forward parameters

A second example demonstrates the flexibility of the extraction tool. Parameter ise is extracted together with the previous parameter set. The extraction result for all forward parameters is shown in Table 3. This demands measurement points at low currents I_B and I_C, thus

the bias range is extended and Gummel-plot input data are used ($0.25V < V_{B0} < 1.25V$ and $80mV < V_{C0} < 5V$). The number of data points is approximately tripled, which leads to a significant increase of the computation time, approximately by a factor of five.

Figures 4 a) and b) show a comparison of measured data ($\circ$) and GP model generated plots using the extracted parameter set (solid line). An unsparing exposure of the GP model boundaries is depicted in the $\beta(I_C)$ curvatures. Here the maximum relative error exceeds 13 % for currents I_C below 1 mA. This can be corrected by using bias dependent Early voltages vaf and var for modeling the depletion charges at both sides of the base region [14]. Note, that the measured data is not optimized with respect to the $\beta(I_C)$-plot.

5 CONCLUSIONS

In this work we have introduced a new method for the extraction of BJT model parameters with respect to the GP equations. The optimized parameter set is valid for a global bias regime of the device. A level of confidence in each parameter is given to estimate the reliability and the goodness of the fit. The knowledge of the variance, correlation and sensitivity is valuable in eliminating unnecessary or redundant parameters, providing a more accurate parameter set. Otherwise this information can be efficiently used for selecting a useful bias range of the input data and also for model improvements or process control.

The provided examples show good agreement of the GP equations with the measured real device behaviour, bearing in mind the limitations of the model. Thus the GP model still seems useful for circuit simulation in the low frequency regime, particularly in consideration of the relative small number of parameters.

In this work the minimum data needed to extract the parameters is treated. Of course, additional measurements (*eg* small-signal S-parameters) can be carried out and can also be used for the extraction.

APPENDIX

Expression (9) represents a typical ill-conditioned set of linear equations; $cond\,(\mathbf{A}) = \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\| > 10^{44}$ for $g > 7$. In this work the error-free parallel algorithm of [10] is adapted for sequential use. Below, the main features of this algorithm are introduced briefly.

We can rewrite (9) in the form

$$\mathbf{A}\boldsymbol{b} = \boldsymbol{y}\,, \tag{17}$$

where $\mathbf{A} = (\mathbf{J}^\top\mathbf{J} + \lambda^k(\mathbf{J}^\top\mathbf{J})_{jj})$, $\boldsymbol{b} = \boldsymbol{\Delta p}^k$ and $\boldsymbol{y} = \mathbf{J}^\top\boldsymbol{f}(\boldsymbol{p}^k)$.

The size of a modulus M used in an error-free algorithm based on modular arithmetic to solve (17) is given as [10]

$$M > 2\max\left\{\begin{array}{l} g^{\frac{g}{2}}M(\mathbf{A})^g, \\ g(g-1)^{\frac{(g-1)}{2}}M(\mathbf{A})^{g-1}M(\boldsymbol{y}) \end{array}\right\},$$

where $M(\boldsymbol{y}) = \max|y_i|$, $M(\mathbf{A}) = \max|a_{ij}|$ and $i,j \in \langle 1,g\rangle$.

Since the modulus M estimated in this way is a very large number, its application to calculations would not be practical. Therefore, a known set of moduli $m_1, m_2, \ldots, m_r$ are used, which must fulfil the following conditions:

- $\prod_k^r m_k \geq M$
- $m_1 < m_2 < \cdots < m_r$
- $m_k \in \{primes\}$, where $k \in \langle 1,r\rangle$.

The chosen moduli are applied in the algorithm based on the Gauss-Jordan elimination with pivoting performed in modular arithmetic [10]. All steps are almost identical in both the classical and in the proposed algorithm. The only difference is that in the latter the operations of addition, negation, multiplication and inversion are carried out with integers in the appropriate residual class code (RCC) without introducing rounding errors.

Based on [10], the modified algorithm for solving linear equations in modular arithmetics is given as follows:

I. Do:
$$\begin{aligned} \boldsymbol{y}^{int} &:= \tfrac{\boldsymbol{y}}{\min}2^{e+1} \\ \mathbf{A}^{int} &:= \tfrac{\mathbf{A}}{\min}2^{e+1} \end{aligned}$$
II. For $k = 1, 2, \ldots r$ do:
$$\begin{aligned} \boldsymbol{y}^{m_k} &:= \boldsymbol{y}^{int})mod_{m_k} \\ \mathbf{A}^{m_k} &:= \mathbf{A}^{int})mod_{m_k} \end{aligned}$$
Do: (GJ-elimination) $mod(m_k)$
$$\begin{aligned} \boldsymbol{b}^{m_k} &:= \left(\mathbf{A}^{m_k}\boldsymbol{y}^{m_k}\right)^{-1} mod(m_k) \\ D^{m_k} &:= |\mathbf{A}^{m_k}| mod(m_k) \end{aligned}$$
III. Do: Conversion from the RCC
$$\begin{aligned} \boldsymbol{b}^{int} &:= conv(\boldsymbol{b}^{m_1}, \boldsymbol{b}^{m_2}, \ldots, \boldsymbol{b}^{m_r}) \\ D^{int} &:= conv(D^{m_1}, D^{m_2}, \ldots, D^{m_r}) \end{aligned}$$
IV. Do:
$$\boldsymbol{b} := \tfrac{\boldsymbol{b}^{int}}{D^{int}}$$

In step I the conversion of real data $\mathbf{A}$ and $\boldsymbol{y}$ to integer numbers $\mathbf{A}^{int}$ and $\boldsymbol{y}^{int}$ is performed, where the values $\min = \min\{\mathbf{A}, \boldsymbol{y}\}$ and e is the word length of mantisa in bits. In such a way numerical damage of input data by their conversion to integer numbers is minimized. In step II operations in modular arithmetic are carried out. On the converted data (from integer to residual class code RCC) $\mathbf{A}^{m_k}$ and $\boldsymbol{y}^{m_k}$ operations of Gauss-Jordan elimination are executed with pivoting in modular arithmetic. The result of the computing process is the vector $\boldsymbol{b}^{m_k}$ and determinant D^{m_k} represented in the RCC. In the next step vector $\boldsymbol{b}^{m_1}, \boldsymbol{b}^{m_2}, \ldots, \boldsymbol{b}^{m_r}$ and determinant $D^{m_1}, D^{m_2}, \ldots, D^{m_r}$ from the RCC to real numbers are converted. The solution $\boldsymbol{b}$ of (7) is given by the fraction expressed in the last step IV.

References

[1] de GRAAFF, H. C.: State-of-the-Art in Compact Modelling with Emphasis on Bipolar RF Circuit Design, Proc. ESSDERC '97, Stuttgart, 1997.

[2] SADOVNIKOV, A. D.—ROULSTON, D. J.—CELI, D.: Extraction of SPICE BJT Model Parameters in BIPOLE3 Using Optimization Methods, IEEE Trans. on CAD of Integrated Circuits and Systems **15** No. 11 (1996), 1332–1339.

[3] FUSE, T.—SASAKI, Y.: Trans. on Electron Devices **42** No. 3 (1995), 534–539.

[4] PARK, J. S.—NEUGROSCHEL, A.: Parameter Extraction for Bipolar Transistors, IEEE Trans. on Electron Devices **36** No. 1 (1989), 88–95.

[5] SISCHKA, F.: A Method to Determine the SPICE-Parameters for Bipolar Junction Transistors, AEÜ **39** No. 4 (1985), 225–232.

[6] WARD, D. E.—DOGANIS, K.: Optimized Extraction of MOS Model Parameters, IEEE Trans. on CAD of Integrated Circuits and Systems **CAD-1** No. 4 (1982), 163–168.

[7] GUMMEL, H. K.—POON, H. C.: An Integral Charge Control Model of Bipolar Transistors, Bell Syst. Tech. J. **49** (1970), 827–852.

[8] GEIGER, R. L.—ALLEN, P. E.—STRADER, N. R.: VLSI Design Techniques for Analog and Digital Circuits, McGraw-Hill, New York, 1990.

[9] NASH, T. C.: Compact Numerical Methods for Computers, John Wiley & Sons, New York, London, Sydney, Toronto, 1977.

[10] MORHÁČ, M.—LÓRENCZ, R.: A Modular System for Solving Linear Equations Exactly, I. Architecture and Numerical Algorithms, Computers and Artificial Intelligence **11** No. 4 (1992), 351–361.

[11] BECK, J. V.—ARNOLD, K. J.: Parameter Estimation in Engineering and Science, John Wiley & Sons, New York, London, Sydney, Toronto, 1977.

[12] MACHALA, C. F.—PATTNAIK, P. C.—YANG, P.: An Efficient Algorithm for the Extraction of Parameters with High Confidence from Nonlinear Models, IEEE Electron Device Letters **7** No. 4 (1986), 214–218.

[13] CARROLL, M. S.—NEUGROSCHEL, A.—SAH, C.-T.: Degradation of Silicon Bipolar Junction Transistors at High Forward Current Densities, IEEE Trans. on Electron Devices **44** No. 1 (1997), 110–117.

[14] McANDREW, C. C. et al.: The Vertical Bipolar Inter-Company Model, IEEE J. of Solid-State Circuits **31** No. 10 (1996), 1476–1483.

Received 26 July 1999

Róbert Lórencz (Ing, CSc) was born in Prešov, Slovakia, in 1957. He graduated from the Electrical Engineering Faculty of the Czech Technical University in Prague in 1981. From 1983 to 1990 he worked on general problems of data acqusition in Nuclear Physics in the Institute of Physics, Slovak Academy of Sciences. He obtained his PhD degree in 1990 at the Institute of Measurement and Measuring Methods, Slovak Academy of Sciences. He spent two study and research visits in Germany, in 1990-1991 in the Gesellschaft für Schwerionenforschung GSI in Darmstadt and in 1996-1998 in the Deutsches Elektronen-Synchrotron DESY in Hamburg. Since 1998 he is with the Faculty of Electrical Engineering, Czech Technical University, Prague. His research interests include data acqusition methods, parameter estimations, error-free numerical algorithms, parallel algorithms, design of microprocessor systems.

Christian Reckleben (Dipl-Ing) born in Göttingen, Germany in 1964. He received the Dipl-Ing degree in electrical engineering from the Technical University Munich in 1992. From 1992 to 1995 he worked in the area of process development and IC design for MEMS. Since 1995 he is with Deutsches Elektronen-Synchrotron DESY, Hamburg, where he is involved in analog IC design for multi-gigabit applications and ultra low noise sensor interfaces for high energy physics and x-ray applications.

Karsten Hansen (Dr-Ing) was born in Flensburg, Germany, in 1961. He received the Dipl-Ing and PhD degree in electrical engineering, both from the Technical University Braunschweig, Braunschweig, Germany, in 1988 and 1993, respectively. While at the TU Braunschweig he worked in the monolithic integration of laser diodes with transferred-elctron devices for high speed optical communication. In 1994, he investigated the optical and electronic properties of heteroepitaxial InP/Si for InP-on-Si-based integrated optoelectronic circuits. Since 1994 he is with the Deutsches Elektronen-Synchrotron DESY in Hamburg, where he was involved in the development of the H1 silicon-strip detector system. Since 1995 he is the head of the micro- and optoelectronic department. His main activities are in the field of monolithic and hybrid integrated electronic and optoelectronic circuits for applications in the high-energy and x-ray physics.