UDK 621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 23(1993)1, Ljubljana

SECOND-ORDER EFFECTS IN JUNCTION BIPOLAR TRANSISTORS

Zvonko Fazarinc

Keywords: semiconductors, bipolar transistors, bipolar junction transistors, transistor models, computer analysis, Ebers-Moll model, Gummel-Poon model, Early effect, Kirk effect, C language

Abstract: The Gummel-Poon model of bipolar transistor is derived with inclusion of Early and Kirk effects.

Efekti drugega reda v bipolarnih tranzistorjih

Ključne besede: polprevodniki, transistorji bipolarni, BJT transistorji bipolarni junction, modeli transistorjev, analiza računalniška, Ebers-Moll model, Gummel-Poon model, Early efekt, Kirk efekt, C jezik

Povzetek: V prispevku je izveden Gummel-Poon model bipolarnega tranzistorja, ki vključuje Earlyjev in Kirkov efekt.

1. Introduction

The first model of the bipolar junction transistor (BJT) was published in 1954 by J.L.Moll and J.J.Ebers from Bell Laboratories /1/. The model quickly became the workhorse of the industry because it captured the state of the art of the transistor technology of that time in a most elegant way. The model was simple and sufficiently accurate to allow prediction of transistor circuit behaviour based on a few physical parameters.

The transport of charges in a semiconductor is governed by diffusion, drift and recombination. The early transistors had relatively long bases and low current gains, both resulting in a considerable recombination in the base. High level injection and other phenomena associated with drift were hardly a problem of that period. The Ebers-Moll model was based on the diffusion and recombination mechanisms only, and was as such in tune with the technological state of the 50-s and 60-s.

The semiconductor technology was making big strides in the subsequent decades and the BJT bases have shrunk down to below a micrometer. Due to this and in conjunction with constant improvement of material purity the recombination rates were suppressed to theoretical minima. At the same time the quest for ever higher current gains called for high base resistivities and the high-level injection became the limiting mechanism of current gain.

It was until 1970, that H.K.Gummel and H.C.Poon also of Bell laboratories published their Integral Charge Control model of the BJT /2/ which captured the high-level injection mechanism as part of its physical basis. Again the model was a reflection of the state of tehnology. Gummel and Poon have correctly concluded that recombination played a minor role in physical behaviour of transistors and have ignored it in favor of including the drift mechanism in addition to the diffusion. The base current in the Ebers-Moll model was a natural consequence of recombination in the base. With no recombination considered the Gummel-Poon model requires a separate theory for the base current.

Both models, the Ebers-Moll and Gummel-Poon must treat the Early effect /3/, the Kirk effect /3/, the emitter crowding /5/ and other second-order effects as separate problems. This paper integrates the Early and the Kirk effect into the Gummel-Poon model derivation and thereby avoids the present conflict of first deriving the Gummel-Poon model under the assumption of a fixed base length, and then adding the Early and Kirk effect which both arise from a variation of the base length.

2. Definition of Model Parameters

We arbitrarily choose an NPN bipolar transistor as the basis for our analysis. The results are easily expandable to PNP structures. Extensive approximations are necessary to arrive at the basic Gummel-Poon model. A real transistor is a three- dimensional structure while the Gummel-Poon model possesses no spatial dependence at all. We start with a one-dimensional approximation of the transistor and demonstrate how the spatial dependence is integrated out. In process we make the user of the model aware of the many necesary approximations.

At the top of Fig.1 is a schematic depiction of an NPN transistor. The polarities of externally applied biases are in agreement with normal operation while the junction biases at x_1 and x_2 are defined in agreement with convention.

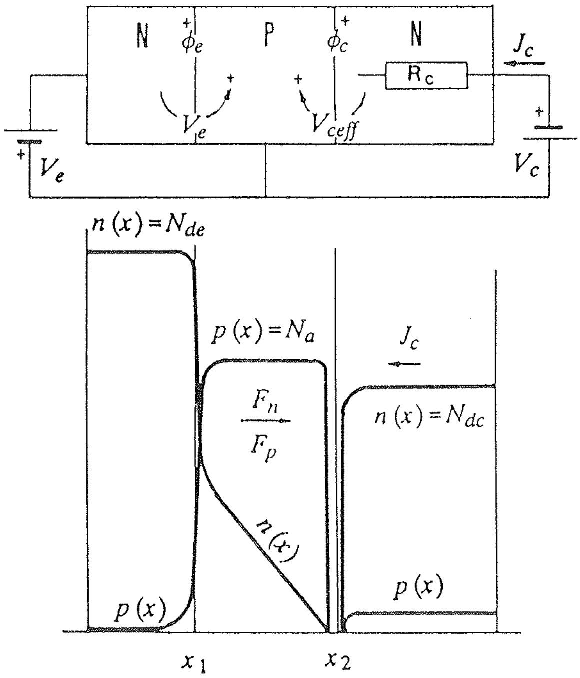

Fig. 1: *NPN Transistor Schematic*

Below the transistor and in relative position to it are shown the internal concentration of electrons n and of holes p as functions of x. The concentrations shown correspond to a positive V_e and a negative V_{ceff}. The junction at x_2 is therefore back-biased.

3. Modeling of High-Level Injection

The high-level injection or the Webster effect is an integral part of the Gummel-Poon model but we will rederive it here in order to lay ground for inclusion of additional second-order effects. The Webster effect originates in the base voltage drop near the emitter junction. This directly reduces the effective bias across the junction from the full value V_e. It happens when the number of minority carriers injected into the base approaches that of the background concentration and results in a significant drop of current gain. The absence of neutrality which accompanies this phenomenon requires that the model include the electric field. We start therefore with the transport equations /6/ in terms of electron and hole flux densities F_n and F_p, respectively, which contain in addition to the diffusion term the drift term as well.

$$F_n(x) = -D_n \frac{\partial n(x)}{\partial x} - \frac{qD_n}{kT} E(x) n(x) \tag{1}$$

$$F_p(x) = -D_p \frac{\partial p(x)}{\partial x} + \frac{qD_p}{kT} E(x) p(x) \tag{2}$$

In (1) and (2) D stands for the diffusivity of respective carriers and $E(x)$ for the electric field. q is the electronic charge, k is the Boltzmann's constant and T is the absolute temperature. F_p in the base's P-region is the majority carrier flux and as such negligible compared to F_n. We take advantage of this, set it to zero, and find from it the electric field $E(x)$. This field is, of course, common to both carriers and we substitute it into the first equation to obtain

$$F_n(x) = -D_n \frac{\partial n(x)}{\partial x} - D_n \frac{\partial p(x)}{\partial x} \frac{n(x)}{p(x)}$$

Multiply both sides by $p(x)$ and obtain

$$F_n(x) p(x) = -D_n \left[\frac{\partial n(x)}{\partial x} p(x) + \frac{\partial p(x)}{\partial x} n(x) \right]$$

The term in brackets is recognized as the derivative of the product $p(x)\, n(x)$ so the integration of the above equation between the two junctions yields

$$\int_{x_1}^{x_2} F_n(x) p(x)\, dx = -D_n n(x) p(x) \Big|_{x_1}^{x_2} \tag{3}$$

The product $n(x)\, p(x)$ is given by the mass action law /7/ everywhere except near the two junctions where the law of the junction applies. Therefore the following must be true

$$n(x_1)\, p(x_1) = n_i^2\, e^{qV_e/kT} \tag{4}$$

$$n(x_2)\, p(x_2) = n_i^2\, e^{qV_{ceff}/kT} \tag{5}$$

where n_i is the intrinsic concentration of the host material. Substitute (4) and (5) into (3) and end up with the following integral equation

$$\int_{x_1}^{x_2} F_n(x) p(x)\, dx = -D_n n_i^2 \left[e^{qV_{ceff}/kT} - e^{qV_e/kT} \right] \tag{6}$$

A computer solution of (6) for any arbitrary distribution of $p(x)$, given the other parameters, is quite feasible and may prevail in the future as a means of circuit analysis. At this time the computational effort still becomes excessive if thousands of transistors are involved even in view of the present day computer performance level. Therefore, we are forced to eliminate the last remaining spatial coordinate, i.e., x from our model. This we do by making the assumption of space charge neutrality which is in direct conflict with the initial goal of deriving a model subject to internal electric fields. The fact is that we cannot solve the integral in (6) for the general case by any known means. A consoling circumstance is that a very small charge unbalance is sufficient to maintain the electric fields in the region of interest and the neutrality condition can be justified on that basis. This says that sum of all charges in the region of interest must be equal to zero. In addition to mobile charges $n(x)$ and $p(x)$ we must consider the ionized impurities which results in the commonly encountered neutrality condition

$$N_d(x) - N_a(x) + p(x) - n(x) = 0 \qquad (7)$$

An alternative expression for the minority carrier flux density $F_n(x)$ is in the terms of their density $n(x)$ and velocity $v_n(x)$

$$F_n(x) = n(x)\, v_n(x)$$

A combination of (7) with the above produces

$$p(x) = \frac{F_n(x)}{v_n(x)} + N_a(x) - N_d(x) \qquad (8)$$

Before we substitute (8) back into (6) we recognize that the donor density N_d in the base P-region of our NPN transistor must be zero and that in absence of recombination the minority flux density in the base $F_n(x)$ must be independent of position x and equal to the collected flux density F_c. Consequently (6) translates into the following quadratic equation

$$F_c^2 \int_{x_1}^{x_2} \frac{dx}{v_n(x)} + F_c \int_{x_1}^{x_2} N_a(x)\, dx + D_n n_i^2 \left[e^{qV_{ceff}/kT} - e^{qV_e/kT}\right] = 0$$

The solution of the quadratic yields

$$F_c = -\frac{\int_{x_1}^{x_2} N_a(x)dx}{2\int_{x_1}^{x_2} \frac{dx}{v_n(x)}} \left[1 - \sqrt{1 + \frac{4D_n n_i^2 (e^{qV_e/kT} - e^{qV_{ceff}/kT})}{\left[\int_{x_1}^{x_2} N_a(x)dx\right]^2} \int_{x_1}^{x_2} \frac{dx}{v_n(x)}}\right]$$

The integral of $dx/v_n(x)$ is the transit time from x_1 to x_2 which we denote by τ_n. The transit time in the base is primarily controlled by diffusion despite the electric field-presence. We will therefore use the diffusion aproximation of the transit time (8)

$$\tau = \frac{(x_2 - x_1)^2}{2D_n} = \frac{w^2}{2D_n}$$

Finally we are forced to assume that the concentration of acceptors in the base $N_a(x)$ is independent of x, admittedly not a good approximation. All of this results in the highly simplified expression for the collector flux density

$$F_c = -\frac{D_n N_a}{w}\left[1 - \sqrt{1 + 2(n_i/N_a)^2 (e^{qV_e/kT} - e^{qV_{ceff}/kT})}\right]$$

The collector current density J_c is the negative of the electron flux density multiplied by q

$$J_c = \frac{qD_n N_a}{w}\left[1 - \sqrt{1 + 2(n_i/N_a)^2 (e^{qV_e/kT} - e^{qV_{ceff}/kT})}\right] (9)$$

Equation (9) is the principal result of the work by Gummel and Poon if w is replaced by the unbiased base length w_{bo} and the effective collector junction bias V_{ceff} is raplaced by the applied collector-base voltage V_c. We will give this expression an interpretation later.

4. Base-Width Modulation

Parameters w and V_{ceff} can be expressed in terms of some physical parameters illustrated in Fig.2. This depicts the electric field for our idealized structure from Fig.1 with the unbiased case shown in dashed lines.

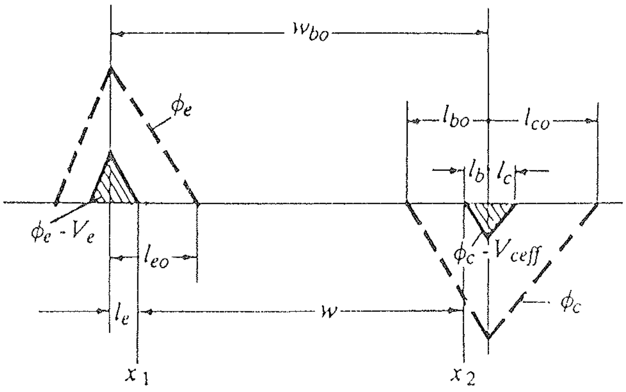

Fig. 2: Electric Field Inside the Transistor.

The area under the field curve equals the voltage across the particular junction. For the unbiased case we have in the emitter and collector junctions, respectively, the following built-in voltages

$$\phi_e = \frac{kT}{q}\ln\frac{N_{dc} N_a}{n_i^2} \quad \text{and} \quad \phi_c = \frac{kT}{q}\ln\frac{N_{dc} N_a}{n_i^2} \qquad (10)$$

In terms of these it is easy to extract by simple geometric considerations for similar the following expression for the diffusion width of our transistor

$$w = w_{bo} - l_{eo}\sqrt{1 - V_e/\phi_e} - l_{bo}\sqrt{1 - V_{ceff}/\phi_c} \qquad (11)$$

$$l_{eo} = \sqrt{\phi_e \varepsilon / q(N_a + N_{de})} \quad \text{and} \quad l_{bo} = \sqrt{\phi_c \varepsilon / q(N_a + N_{dc})}$$

Expression (11) contains the base length modulation caused by the emitter and collector voltage. The latter is, of course, the Early effect and if we substituted (11) back into (9) we would have included the Early effect in the Gummel-Poon model. But we have promised to include the Kirk effect as well so we must now focus on it before we finalize the model.

As the readers might know, the Kirk effect is also referred to as "base push-out", i.e., a base lengthening effect.

It has its origin in the resistive drop in the collector region for high currents. As the collector current rises some of the applied collector voltage V_c is expended in the collector resistance and the remaining junction voltage V_{ceff} may be diminished to the point where no more collection of the arriving electrons takes place in the junction. The electrons start accumulating and the diffusion length of the base increases. This is illustrated in Fig.3.

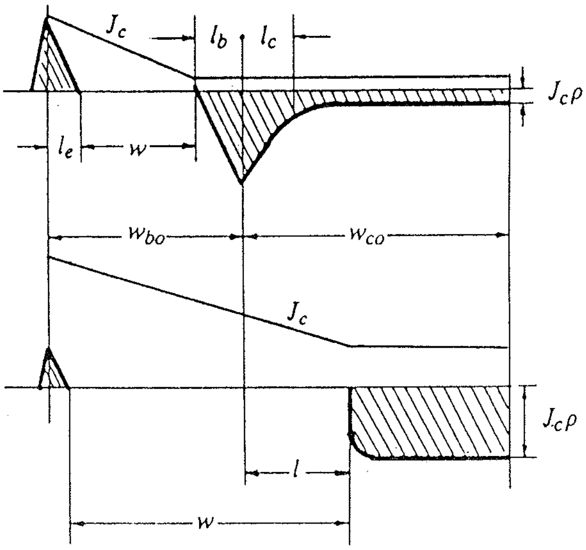

Fig.3: Illustration of the Base Push-out.

In it we see the shape of the electric field for moderate collector current at the top and for high current on the bottom. What the two cases have in common is the area under the field on the collector side which represents the collector voltage. One can observe the lengthening of the base region through which the electrons must diffuse as the current increases. The net effect of this is a reduced current gain but more importantly a quadratic increase of transist time, which reduces the hihg frequency cutoff of the transistor in inverse proportion. With this much physical insight we should be able to drive the effective collector junction bias V_{ceff} and the diffusion base length w. We focus on the collector side and try to find the lengths l_b or l as the case may be.

The top sketch in Fig.3 which representes a moderate current case suggests that most of the applied collector bias appears across the collector-base junction. The resistive drop in the collector region of resistivity ρ_c is

$$J_c\rho_c(w_{co}-l_c) \approx J_c\rho_c w_{co}$$

with w_{co} defined in Fig.3. The approximation is justified for all cases of interest. The remaining collector junction bias is therefore

$$V_{ceff} = -(V_c - J_c\rho_c w_{co}) \qquad (12)$$

From (11) we get for the length l_b

$$l_b = l_{bo}\sqrt{1+(V_c - J_c\rho_c w_{co})/\phi_c} \qquad (13)$$

Expression (13) applies as long as l_b remains real. It attains a zero value for

$$V_c - J_c\rho_c w_{co} = -\phi_c \quad \text{or} \quad J_{c0} = \frac{V_c+\phi_c}{\rho_c w_{co}} \qquad (14)$$

Hence (13) applies for $J_c < J_{c0}$. When J_c gets larger we have at hand the second case in Fig.3 which is easily solved for l as

$$l = w_{co} - \frac{V_c+\phi_c}{J_c\rho_c} \quad \text{for} \quad J_c > J_{c0} \qquad (15)$$

We can now summarize this discussion by expressing the effective base width w and the effective collector bias V_{ceff} applicable to (9) as follows

Case: $J_c \le J_{c0}$ (16)

$$w = w_{bo} - l_{bo}\sqrt{1+(V_c - J_c\rho_c w_{co})/\phi_c}$$

$$V_{ceff} = -(V_c - J_c\rho_c w_{co})$$

Case: $J_c > J_{c0}$ (17)

$$w = w_{bo} + w_{co} - (V_c+\phi_c)/J_c\rho_c$$

$$V_{ceff} = 0$$

Let us return now to expression (9), multiply and divide it by

$$1+\sqrt{1+2(n_i/N_a)^2(e^{qV_e/kT} - e^{qV_{ceff}/kT})}$$

and obtain

$$J_c = 2J_s\frac{w_{bo}}{w}\,\frac{e^{qV_e/kT} - e^{qV_{ceff}/kT}}{1+\sqrt{1+2(n_i/N_a)^2(e^{qV_e/kT} - e^{qV_{ceff}/kT})}} \qquad (18)$$

In (18) we have introduced the reverse saturation current density J_s

$$J_s = \frac{qD_n}{w_{bo}}\,\frac{n_i^2}{N_a} \qquad (19)$$

5. Base Current Model

To complete the model we must add the base current. For our NPN example this consists of holes flowing in the emitter-base diode and is made up of two components. One is due to carriers injected from the base into the emitter and the other has its origin in the transition region where considerable recombination takes place. The reverse injection into the emitter is simply expressed as the diode current

$$\frac{qD_p}{w_e}\frac{n_i^2}{N_{de}}(e^{qV_e/kT}-1) = J_1(e^{qV_e/kT}-1)$$

J_1 is best determined experimentally because of ill defined parameters N_{de} and particularly w_e. The second component of the base current arises from the recombination in the junction. The Shockley-Read-Hall model /9/ relates the net recombination current J_r to some more or less known parameters

$$\frac{\partial J_r(x)/q}{\partial x} = \frac{p\,n - n_i^2}{\tau_p n + \tau_n p}$$

Furthermore, within the junction the product of p and n must obey the law of the junction

$$p(x)n(x) = n_i^2\, e^{qVe/kT}$$

If we make the assumption that $p = n$ in the transition region (not readily justifiable) we obtain

$$n = p = n_i\, e^{qVe/2kT}$$

Substitute this into the differential equation for J_r, integrate over the transition region width w_t, and up with the following expression for the recombination component of the base current

$$J_r = q\, w_t \frac{n_i^2(e^{qV_e/kT}-1)}{n_i e^{qV_e/2kT}(\tau_p+\tau_n)}$$

Because many of the parameters in the above are not well known we lump them into a measurable parameter J_2. We also recognize that for normal forward biases the exponential is much larger than 1 and the latter can be neglected. This produces for the recombination component of the base current the expression

$$J_r = J_2\, e^{qVe/2kT}$$

The base current is then composed of the sum of the two components

$$J_b = J_1(e^{qV_e/kT}-1) + J_2\, e^{qV_e/2kT} \tag{20}$$

6. Model Interpretation

Table 1 below lists the set of parameters used for model evaluation.

Table 1

Parameter	Value	Physical Meaning
n_i	$1.45\text{x}10^{10}/\text{cm}^3$	Intrinsic Concentration
N_a	$1\text{x}10^{16}/\text{cm}^3$	Base Impurity Density
N_{dc}	$1\text{x}10^{15}/\text{cm}^3$	Epi Concentration
D_n	$36\ \text{cm}^2/\text{s}$	Electron Diffusivity
w_{bo}	$1\text{x}10^{-4}$ cm	Metallurgical Base
w_{co}	$1\text{x}10^{-3}$ cm	Epi Thickness
β_{max}	100	Maximum Current Gain
$V_{\beta=1}$	0.1 Volt	Low End Bias for β=1

All the necessary quantities can be calculated from the eight parameters given in the table. The reverse saturation current density from (19) evaluates to $J_s = 1.2x10^{-9}$ A/cm^2. The base current component J_1 can be defined in terms of the maximum current gain β_{max} which can be measured. J_1 is then simply equal to J_s/β_{max} as is evident from Fig.4. Also evident from Fig. 4 is the fact that the second component of the base current determines the value of the emitter-base bias $V_{\beta=1}$ for which the current gain passes through unity. For the chosen value of 0.1 Volt we get from the intercepts of Fig.4 the value $J_2 = J_s\, e^{q0.1/2kT}$. All the other quantities are computed from formulas (10), (11), (14), and (16) through (20) while the large signal current gain β is evaluated as the ratio of J_c to J_b.

The results are plotted in Fig.4 versus the emitter-base bias V_e for $V_c = 0$. Intercepts of currents and of biases indicated in the graph have proven useful in determination of model parameters from measurements.

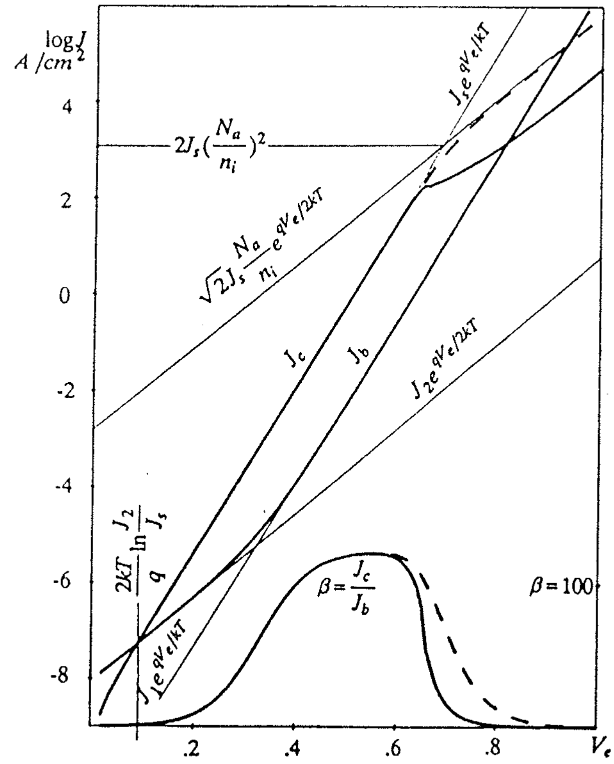

Fig. 4: Modeled Currents and Current Gain.

The collector current density and the current gain have been calculated once from (18) using (16) and (17) for w and a second time using $w = w_{bo} - l_{bo}\sqrt{(1 + V_c/\Phi_c)}$

which ignores the Kirk effect. Dashed lines have been used in this latter case. A slight degradation of current gain is noticeable at high currents when Kirk effect is considered. As the collector bias is increased the degradation is still there it only occurs at higher currents. In Fig.5 the current gain is graphed for three different collector biases once with inclusion and once without the Kirk effect. The latter is plotted in dashed lines.

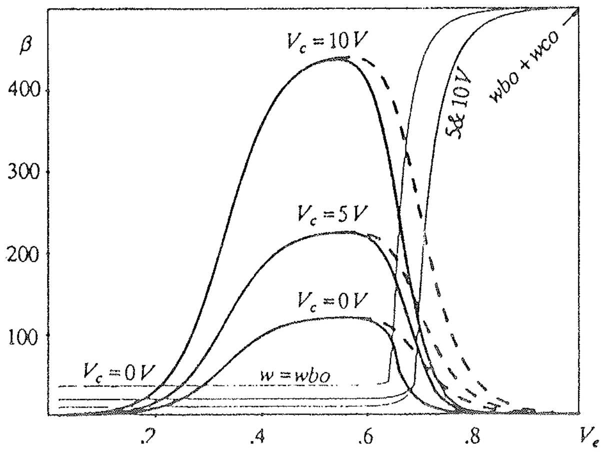

Fig. 5: Current Gain and Base Width for Different V_c.

The main problem caused by base push-out is not the current gain degradation but the degradation of high frequency performance. In Fig.5 the base width *w* is plotted for the same three collector biases as β and the dramatic effect is inescapable. The high frequency cutoff is inversely proportional to the square of the base width across which the carriers must diffuse and consequently the inclusion of the Kirk effect is of crucial importance where the high frequency performance is at issue. Finally in Fig.6 the common emitter characteristics of our sample transistor are shown for moderate bias conditions to ilustrate the Early effect.

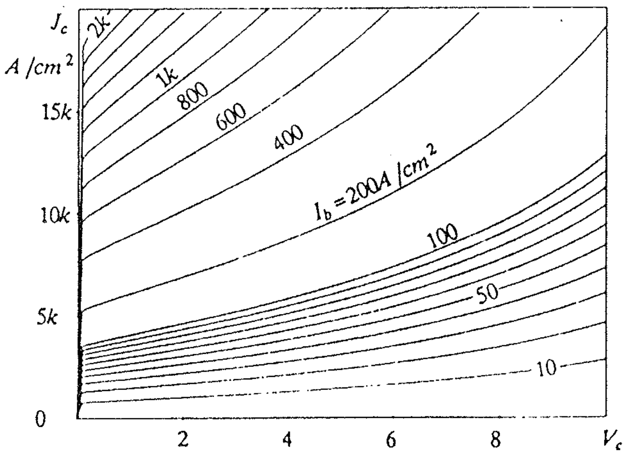

Fig. 6: Common Implementation in C

7. A Computer Implementation in C

C-language is gradually making inroads into scientific computing and it seems appropriate to show one implementation of our Gummel-Poon model in this language. The notation used is adhered to as closely as permitted by the language. Dots are used to mark those variables whose values may be assigned by the user.

//Part 1: Assigned parameters

```
epsilon = 1e-12; q = 1.6e-19; wbo = 1e-4;
wco = 1e-3; Na = 1e16; Ndc = 1e15;
ni = 1.45e10; mun = 1450;
Vc = ..... (positive); Ve = ...... (positive);
temperature = ..... (positive in Kelvin);
decimator = ..... (.03) used in the example.
```

//Part 2: Computed Variables

```
kT = 1.38e-23* temperature;
Dn = mun*kT/q;
phic = log(Na*Ndc/(ni*ni))*kT/q;          //φc
rhoc = 1/(q*mun*Ndc);                     //ρc
Ibo = agrt(phic*epsilon/(q*(Na+Ndc);      //Ibo
Ib = Ibo*sqrt(1+Vc/phic);                        //Ib
Jco = (phic+Vc)/(rhoc*wco);               //Jco
Js = 2*Dn*ni*ni/(wbo*Na);                 //Js
J1 = Js/betamax;                          //J1
J2 = Js*exp(q/(2*kT)*Vbeta0);             //J2
niNaSquared = 2*ni*ni/(Na*Na);            //2(ni/Na)^2
```

//Part 3: Tentative evaluation with approximate values

```
tentativeW = wbo-Ib;
Vceff = - Vc;
A = exp(Ve)-exp(Vceff);
Jc = 2*Js*wbo/tentativeW*A
      /(1+sqrt(1+niNaSquared*A));
Jb = J1*exp(q*Ve/kT) + J2*exp(q*Ve/(2*kT));
```

//Part 4: Final evaluation with filtred feedback

```
if(Jc <= Jco)
{Vceff = - (Vc-Jc*rhoc*wco);
W = wbo - Ibo*sqrt(1-Vceff/phic);
{
else
{Vceff = 0{;
tentativeW = wbo+wco-(Vc+phic)/(Jc*rhoc);
```

```
w = (tentativeW-w)*decimator + w;
}
A = exp(Ve)-exp(Vceff);
Jc =    2*Js*wbo/w*A
        /(1+sqrt(1+niNaSquared*A));
```

One may wonder why so many steps for a relatively simple set of equations. If the transistor is embedded in reactive circuitry, nothing can change abruptly and one can evaluate the model equations without any complications. But a true DC model presented here experience sudden changes of parameters. The transition from normal operation to Kirk state, for example, is very abrupt and is dependent on the collector current. The collector current, on the hand, is exponentially dependent on V_{ceff}. We are consequently faced with a high gain feedback system of essentially infinite bandwidth. The behaviour of such systems is well known and therefore we have introduced a filter into the feedback loop. The "decimator" has a value less than unity and provides a bandwidth limitation to the system. If the Kirk effect is of no consequence Part 4 can be omitted all together resulting in dashed curves of Fig. 4 and 5.

8. Conclusion

The Gummel-Poon model has been revisited with the purpose of including the Early and Kirk effects earlier in the derivation. The result is a simpler model with fewer parameters that can be derived from the process and from simple measurements. The Kirk effect sets in abruptly and calls for softening of the response unless the external circuitry provides the cushion. If the high frequency response is not of importance, only the basewidth modulation associated with the Early effect should be retained. The current gain deterioration due to the Kirk effect can be neglected in such cases. A sample computer program cast in C-like language shows how one or both base modulation effects can be implemented in the model.

Reference

/1/ Ebers, J.J., and Moll, J.L., "Large-Signal Behaviour of Junction Transistors," Proc. IRE, 42, No.12 (December 1954), pp.1761-1772.

/2/ Gummel, H.K., and Poon, H.C., "An Integral Charge Control model of Bipolar Transistor," B.S.T.J., (May-June 1970), pp.827- 852.

/3/ Early, J.M., "Effects of Space-Charge Layer Widening in Junction Transistor," Proc. IRE, 40, (November 1952), pp.1401- 1406.

/4/ Kirk, C.T., "A Theory of Transistor Cutoff Frequency Falloff at High Current Densities," IRE Trans. Electron Devices, ED-9, No.2 (March 1962), pp.164-174.

/5/ Sze, S.M., "Semiconductor Devices", John Wiley & Sons, 1986, p.126.

/6/ Muller, S.R., and Kamins, "Device Electronisc for Integrated Circuits," Second Edition, John Wiley & Sons, p.220.

/7/ ________, pp.16-18
/8/ ________, p.335
/9/ ________, pp.226-228

prof. Zvonko Fazarinc, Ph.D
Stanford University, ret.
Hewlett-Packard Laboratories, ret.
880 La Mesa Drive
Menlo Park, CA 94025
fax (415) 854 8574

Prispelo: 05.2.93 ***Sprejeto: 25.2.93***