Solid-State Electronics Vol. 33, No. 6, pp. 705–718, 1990
Printed in Great Britain.

0038-1101/90 $3.00 + 0.00

IMPACT IONIZATION IN SILICON: A REVIEW AND UPDATE

W. MAES, K. DE MEYER* and R. VAN OVERSTRAETEN
IMEC vzw, Kapeldreef 75, 3030 Heverlee, Belgium

(Received 4 December 1989; in revised form 23 January 1990)

Abstract—After defining the multiplication factor and the ionization rate together with their interrelationship, multiplication and breakdown models for diodes and MOS transistors are discussed. Different ionization models are compared and test structures are discussed for measuring the multiplication factor accurately enough for reliable extraction of the ionization rates. Multiplication measurements at different temperatures are performed on a bipolar *NPN* transistor, and yield new electron ionization rates at relatively low electrical fields. An explanation for the spread of the experimental values of the existing data on ionization rate is given. A new implementation method for a local avalanche model into a device simulator is presented. The results are less sensitive to the chosen grid size than the ones obtained from the existing method.

1. INTRODUCTION

Impact ionization is an important charge generation mechanism. It occurs in many semiconductor devices and it either determines the useful characteristic of the device or it causes an unwanted parasitic effect. Let us first consider bipolar devices. The breakdown voltage V_b of a silicon p–n diode is caused by impact ionization if V_b is larger than about 8 V. The temperature coefficient of V_b is then positive. For breakdown voltages smaller than about 5 V, the breakdown is caused by tunneling. The temperature coefficient of V_b is then negative. In diodes with a breakdown voltage between 5 and 8 V, both mechanisms contribute to the breakdown and V_b is more temperature insensitive, improving the properties of a Zener diode, as a reference voltage.

The operation of some devices is based on avalanche generation. This is the case for a thyristor and for an impatt diode. In an avalanche photodetector, the internal multiplication is used to increase the signal to noise ratio of the total detector system, although avalanche generation produces its own noise due to its statistical behaviour, as explained by McIntyre[1] and Webb[2].

Avalanche generation also plays an increasing role in MOS devices. By scaling down the geometrical dimensions while keeping the supply voltage constant, the electrical field increases and therefore impact ionization plays a more important role in device degradation due to hot-carrier effects and bipolar parasitic breakdown. Indeed the hot-carrier degradation phenomena are caused by the electron–hole pairs generated in the high-field drain region of the MOSFET resulting in substrate and in gate currents. Due to hot-electron trapping in the oxide, the threshold voltage V_T of the transistor will shift in time and cause reliability problems. Substrate currents can cause bipolar action (latch-up or snap-back) of the MOS transistor. Also breakdown of the drain–substrate junction becomes more important because, due to scaling, the substrate doping must increase in order to avoid punch-through in the device. Many of these problems can be overcome by using LDD-structures, which have a lower electrical field at the drain side, resulting in a lower generation rate. Computer simulation on a specific example shows that a reduction of the peak electrical field by about 40% results in a reduction of the generation rate by a factor of one thousand.

In Section 2, the ionization rate and the multiplication factor together with their interrelationship are defined. In Section 3, different multiplication and breakdown models are discussed while in Section 4 different ionization models are compared. In Section 5, the parameter extraction procedure for fitting the physical and empirical parameters of the avalanche generation model to the multiplication data on a bipolar transistor at different temperatures is studied. New ionization rate data is given. In Section 6, we discuss the implementation of an avalanche model into a device simulator. A new implementation method, insensitive to the mesh, is introduced.

2. DEFINITION OF THE IONIZATION RATE AND OF THE MULTIPLICATION FACTOR

The ionization rate α is defined as the number of electron–hole pairs generated by a carrier per unit distance travelled. It is different for electrons (α_n) and for holes (α_p). Impact ionization can only occur when the particle gains at least the threshold energy for ionization E_i from the electrical field. From the

*Senior Research Associate of the Belgian National Fund for Scientific Research.

application of the laws of conservation of energy and of momentum at a collision event, it can be derived that a minimum energy of $1.5\,E_g$, with E_g being the bandgap, is needed if the effective masses of both holes and electrons are assumed equal. A large spread of experimental values for E_i exists. In general, the ionization rates ($\alpha_{n,p}$) depend on the probability for the carriers to reach this threshold energy and this is not only a function of the local electrical field but also of the "history" of the particle; a "non-local" ionization model thus should be used. As a first approximation a local model is given in which the ionization rate depends only on the local electrical field. The most commonly used local avalanche generation model is the empirical expression of Chynoweth[3], which is similar to the expression for ionization rates in gases:

$$\alpha_{n,p} = \alpha_{n,p}^{\infty} \exp\left(\frac{-b_{n,p}}{E_J}\right) \tag{1}$$

with $\alpha_{n,p}^{\infty}$ and $b_{n,p}$ the ionization coefficients and E_J the electrical field component which can be expressed mathematically as:

$$E_J = \frac{\mathbf{E}\cdot\mathbf{J}}{|J|}. \tag{2}$$

As can be observed, $\alpha_{n,p}^{\infty}$ is the maximum number of carriers that can be generated per unit distance at very high electrical fields. The avalanche generation term is used in the current continuity equations as:

$$\nabla J_{n,p} = \pm q\,(G_{n,p} - R_{n,p}) \tag{3}$$

with $J_{n,p}$ the current density for electrons respectively for holes and with $G_{n,p}$ the generation and $R_{n,p}$ the recombination term for electrons and for holes. The plus sign has to be used for holes and the minus sign for electrons. The generation term for avalanche generation can be written as:

$$G_{n,p} = \frac{\alpha_n |J_n|}{q} + \frac{\alpha_p |J_p|}{q}. \tag{4}$$

It is through these equations that the avalanche generation term is built into a device simulator.

The multiplication factor M is defined as the ratio of the current of one kind (electrons or holes) $J_{\mathrm{OUT},(n,p)}$ coming out of a volume to the current of the same kind, $J_{\mathrm{IN},(n,p)}$, flowing into that volume:

$$M_{n,p} = \frac{J_{\mathrm{OUT},(n,p)}}{J_{\mathrm{IN},(n,p)}}. \tag{5}$$

In one dimension, the relationship between multiplication and avalanche generation can easily be found by solving the above continuity equation, neglecting the recombination term and knowing that:

$$J_{\mathrm{TOT}} = J_n(x) + J_p(x) \tag{6}$$

with J_{TOT} the total current which must be constant in space. For electrons, as incident carriers[4–6]:

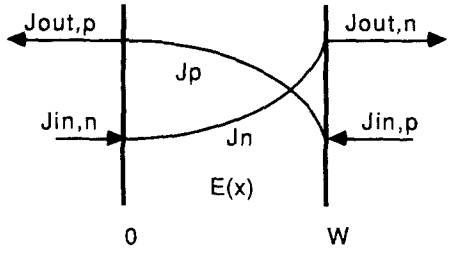

Fig. 1. Simple one-dimensional picture for defining multiplication within a certain region W where a high electrical field exists. The sum of hole and electron current must be constant.

$$M_n = \frac{1}{1 - \int_0^W \alpha_n \exp\left(-\int_0^x (\alpha_n - \alpha_p)\,dx'\right) dx} \tag{7}$$

and for holes as incident carriers:

$$M_p = \frac{1}{1 - \int_0^W \alpha_p \exp\left(-\int_x^W (\alpha_p - \alpha_n)\,dx'\right) dx} \tag{8}$$

where W is the width of the avalanche region as shown in Fig. 1. If a hole and electron current are coming into this region simultaneously, the multiplication of both current components must be added.

Breakdown is obtained when $M \to \infty$ or when:

$$\int_0^W \alpha_n \exp\left(-\int_0^x (\alpha_n - \alpha_p)\,dx'\right) dx = 1. \tag{9}$$

Extracting the ionization coefficients from multiplication measurements is not trivial. Indeed, in eqn (7) the integrals cannot be obtained analytically when the electrical field is not constant, which is the case in most devices. The numerical calculation of the integral can be simplified assuming that $\gamma = (\alpha_p/\alpha_n)$ is a constant as a function of field, which is valid if the electrical field does not change to much as pointed out by Van Overstraeten and De Man in Refs [7,8]. If the field changes appreciably, the γ ratio should be taken at the highest field value, because multiplication occurs predominantly there. Another difficulty is the fact that the electron and hole ionization coefficients are always appearing together in the equation even when the injected current is of only one type. Parameter extraction therefore is difficult. Yet another source of error is the calculation of the electrical field as a function of distance because the exact doping profile, especially at the junction, must be known. This doping profile can be obtained from C–V measurements as is explained in Refs [4,8,9] or directly out of SRP or SIMS measurements. It should be born in mind, however, that each of these techniques suffers from inaccuracies. The determination of the peak electrical field must be very accurate because of the strong dependence of the generation rate on the electrical field.

As pointed out by Beni and Capasso[10] there is also a difference between the increase in current and

the increase in the number of generated electron–hole pairs because

$$M_{n,p} = \frac{J_{\mathrm{OUT},(n,p)}}{J_{\mathrm{IN},(n,p)}} = \frac{V_{\mathrm{OUT}}}{V_{\mathrm{IN}}} \frac{N_{\mathrm{OUT}}}{N_{\mathrm{IN}}} \tag{10}$$

with $V_{\mathrm{IN}}(V_{\mathrm{OUT}})$ the carrier velocities and $N_{\mathrm{IN}}(N_{\mathrm{OUT}})$ the carrier concentrations at the input (output) of the avalanche region. The multiplication factor is due only to pair production (increase in number of carriers), if the carrier velocity is assumed constant over the avalanche region. This is not always true and the difference becomes more pronounced for relatively low electrical fields and for highly doped regions. Only measurements of the multiplication factor on a P–I–N structure would eliminate the influence of the carrier velocity. If multiplication is simulated with a device simulator assuming a constant mobility, the result cannot directly be related to the number of generated electron–hole pairs because the velocity increases linearly with electrical field.

Different approaches to predict the multiplication factor and the breakdown voltage for a given doping profile and structure will be presented in the next section.

3. MULTIPLICATION AND BREAKDOWN MODELS

Instead of using elaborate numerical calculations, it often is useful to have an empirical expression which predicts the multiplication factor of a structure as a function of applied voltage.

3.1. Bipolar diodes

As an example, Miller[5] proposed an empirical formula for the total multiplication factor of a diode as a function of applied voltage:

$$1 - \frac{1}{M} = \left(\frac{V}{V_{\mathrm{b}}}\right)^n \tag{11}$$

with V_{b} the breakdown voltage, n an empirical parameter and M defined as:

$$M = \frac{I(V)}{I_0} \tag{12}$$

with I_0 the saturation current of the diode[11]. This relationship holds well for relatively low breakdown voltages, but there is a significant departure for breakdown voltages with an internal field strength below about 350 kV/cm. Leguerre[12] used the empirical Miller expression to predict the multiplication factor near breakdown by relating the slope factor n to the different materials and substrate dopings.

Fulop[13] proposed the following empirical expression for the ionization rate:

$$\alpha = C(E)^g \tag{13}$$

with C and g constants which are given for silicon. Using this expression, the breakdown voltage for abrupt and graded junctions can be calculated analytically from the multiplication integral.

Other expressions for the diode breakdown voltage as a function of the doping profile were proposed by several authors[14–17].

Also relations exist to calculate the breakdown voltage of planar junctions with curved edges[18–22] or with field plates[23,24]. The avalanche injection in gate-controlled diodes[25–29] and some transient effects[30] are also studied.

3.2. MOS transistors

Finding the values of the multiplication factor in the drain region of an MOS transistor as a function of gate length and of applied voltage is an important but difficult task. Avalanche multiplication generates the substrate current and part of the gate leakage current. Predicting the multiplication factor analytically is difficult because the electrical field in the drain region has a 2-D character. The validity of a local avalanche generation model as implemented in most device simulators is questionable, since the voltage drop over the avalanche region is rather small compared to (E_i/q), as will be discussed in Section 4 and since the gradient of the electrical field is very high at the drain side. Both these conditions violate the applicability of a local model as is explained more extensively in Section 4. A large number of papers deal with the prediction of MOSFET breakdown[31–49].

3.3. Test structures for measuring ionization rates

An important topic is the choice of the test structure used to measure the multiplication factor and to extract the ionization rates. The most commonly used structure is a reverse-biased diode in which photons generate the carriers necessary to initiate the ionization process[8,9,50]. This method is, however, not sensitive enough for measuring very low values of multiplication and thus for determining ionization coefficients at low electrical fields. For this reason, it is more appropriate to use a test structure which has an internal gain factor for the generated current. Sayle[51] proposed a JFET structure where the initiating current comes from the channel and the generated current is measured in the gate. The problem here is the uncertainty about the electrical field distribution in the avalanche region of the channel. Another interesting device described in Section 5 is the bipolar transistor where the base–collector junction acts as the avalanche region. This structure has two major advantages: the current injected by the emitter can be controlled exactly and second the multiplication can be measured β times more accurately. β is the current gain of the transistor. This is because the avalanche generated current is superimposed on the base current, which is β times smaller than the current injected by the emitter and which initiates the generation process. Slotboom[52]

used a bipolar structure for determining the ionization rate and also a CCD structure with which very small multiplication factors can be detected because the charge packet is multiplied in a serial shift register.

4. THE DIFFERENT AVALANCHE IONIZATION MODELS

In a local model, the ionization rate depends only on the local electrical field and can be written as

$$\alpha(x_i) = f[E(x_i)]. \tag{14}$$

For a local model to be valid, two conditions must be fulfilled: (i) every individual carrier has to sample a large number of values throughout the whole momentum probability distribution within a distance that is small compared to the mean distance between ionizing encounters and (ii) this distance also has to be small compared to the distance in which the electrical field changes by an appreciable amount. The first condition implies that the total voltage drop over the avalanche region must be several times (E_i/q). If one of the above conditions is not satisfied, the "history" of the particle should be taken into account for calculating the ionization probability at any location.

The objective of the theory of impact ionization is to derive the probability for an electron to gain from the electrical field an energy at least equal to the threshold energy E_i. In the diffusion model of Wolff[53], the electron gains the energy gradually through many collisions. In solving the Boltzmann transport equation, he retains the first two terms in the Legendre polynomial expansion of the energy distribution function, and obtains

$$\alpha \propto \exp\left(\frac{-A}{E^2}\right) \tag{15}$$

where A is a constant.

Shockley[54] argues that the electrons gaining enough energy from the field are the ones that are lucky enough to avoid collisions. This theory yields:

$$\alpha \propto \exp\left(\frac{-E_i}{qlE}\right) \tag{16}$$

with l the mean free path.

In the theory of Baraff[55,56], higher-order terms of the Legendre polynomial expansion of the distribution function could be retained by using a truncation technique based on the principle of maximum anisotropy. He uses three independent parameters, namely the mean free path, the threshold energy and the energy loss per phonon collision. He also assumes a constant mean free path for both optical phonon and impact ionization scatterings. In the low field region, this theory agrees with Shockley's model and in the high field region with the results of Wolff.

Based on the work of Okuto and Crowell[57] and of Baraff, Thornber[58] derived an expression for the ionization rate valid for all field strengths:

$$\alpha(E) = \left(\frac{qE}{E_i}\right)\exp\left(\frac{-E_{if}}{E\left(1+\frac{E}{E_r}\right)+E_{kT}}\right) \tag{17}$$

with E_r the field for which phonon energy is reached in one mean free path, E_{kT} the field for which the thermal energy kT is reached in one mean free path, and E_{if} the field for which the threshold energy for ionization is reached in one mean free path. The expression reduces to:

$$E < E_{kT} \quad \alpha(E) \approx \left(\frac{qE}{E_i}\right)\exp\left(\frac{-E_i}{kT}\right)$$

(Thermal) (18)

$$E_{kT} < E < E_r \quad \alpha(E) \approx \left(\frac{qE}{E_i}\right)\exp\left(\frac{-E_i}{E}\right)$$

(Shockley[54]) (19)

$$E_r < E \quad \alpha(E) \approx \left(\frac{qE}{E_i}\right)\exp\left(\frac{-E_r E_i}{E^2}\right)$$

(Wolff[53]) (20)

for different electrical field ranges.

Thornber used the data published by several authors to extract parameters for his model. He found a good agreement between the field parameters for the data of Van Overstraeten and De Man[4], Woods[59] and Grant[60] for electrons. For holes, the agreement is not very good. For the effective threshold energy, he found high values (3.6 eV for electrons and 5–6.2 eV for holes). This effective threshold energy thus lies well above the nominal one. With respect to the exact values of the parameters, some criticism should be made. Thornber indeed calculated the parameters from $\alpha_{n,p}$ data, extracted from the multiplication data using the Chynoweth approximation. The correct way would be to use the Thornber expression in the multiplication integral and to fit it to the measured multiplication data.

Recently also the lucky electron model of Shockley has been extended and made suitable for use in a non-local model. Ridley[61–63] differentiated between the rate of momentum and of energy relaxation, and proposed a lucky drift model which is intermediate between the Shockley and the Wolff model. His results agree well with the ones of Baraff over four orders of magnitude. This model was further improved by Woods[64,65] and Marsland[66] by introducing the concept of "soft threshold energy". Childs[67,68] also includes the effect of field variation over the "dead" space. Chen and Tang[69] use the theory of Keldysh[70], who solved Boltzmann transport equation involving phonon scattering and impact ionization scattering. Chen and Tang derive an explicit expression for the energy mean free path which allows the calculation of Keldysh's energy distribution function. They find a good agreement with the theory of Baraff and of Ridley. They have only one fitting parameter $l(E_i)$, characterizing the

continuous transition from the phonon assisted impact ionization to the phononless impact ionization. This mean free path parameter has reasonable values (around 70 Å for electrons and 50 Å for holes).

To obtain correct simulation results, a non-local generation model should be used in a device simulator because the history of the particle has to be taken into account. Attempts are made to use the Monte Carlo technique within a device in the high field region[71]. The recent models discussed above have also tried to account for the "dark space" using the lucky drift theory as a post-processing option. As an example for a non-local model, Childs[68] calculates the probability that an electron acquires the ionization threshold energy after travelling a distance z_i as:

$$P(z_i) = \int_0^{z_i} \exp\left(\frac{-z}{\lambda}\right) \times \exp\left(-\int_z^{z_i}\left(\frac{z_i - z'}{\lambda_e(E)}\right)\frac{dz'}{\lambda}\right)\frac{dz}{\lambda} \quad (21)$$

where λ and λ_E are the mean free paths for momentum and energy-relaxing collisions. The ionization rate is therefore characterized by summing the probabilities for electrons to drift to the ionization threshold energy as they travel along the channel.

The disadvantage of this method is that it already requires an expression for the electrical field along the channel in order to be able to evaluate the probabilities. Other authors[72,73] use other probability functions and show simulation results on MOS devices.

Various authors have extracted coefficients for the Chynoweth model from their multiplication measurements, e.g. McKay[74,75], Miller[5], Chynoweth[3,76], Lee[77], Moll[78,79], Ogawa[80], Van Overstraeten[8], Grant[60], Dalal[81] and Woods[59]. Their results are summarized in Fig. 2 for electrons and in Fig. 3 for holes (from Ref. [82]). Additionally the results are compared with theoretical calculations of Baraff[55,56] with material constants coming from Sze[83,84]. Also shown are the theoretical limits published by Okuto[57,85] which assume that all the energy the carriers can gain from the field is used to generate other carriers when the threshold energy for ionization is supposed to be 1.6 eV for electrons and 1.8 eV for holes as predicted by Hauser[86]. All authors find different parameter couples $(\alpha_{n,p}^{\infty}, b_{n,p})$ for the Chynoweth expression to fit their data; as it will be explained in Section 5, this can be caused by a correlation which seems to exist between the parameters α_n^{∞} and b_n for a given data set. The parameter values of Van Overstraeten and De Man are mostly used and seem to give the best simulation results. They extracted their parameters out of a large set of different diodes which automatically optimizes their parameter set for a large number of electrical field profiles as also explained by Mânduteanu[87].

Slotboom[52] extracted avalanche ionization coefficients for the Chynoweth expression at the surface of an MOS transistor where the generation is smaller due to the roughness of the interface.

Expressions for the temperature dependence of the ionization rate are given by Crowell and Sze[88]. They are in good agreement with the theoretical calculations of Baraff[55]. As the temperature increases, the probability for the carriers to reach the threshold energy becomes smaller, and the breakdown voltage increases. The temperature dependence

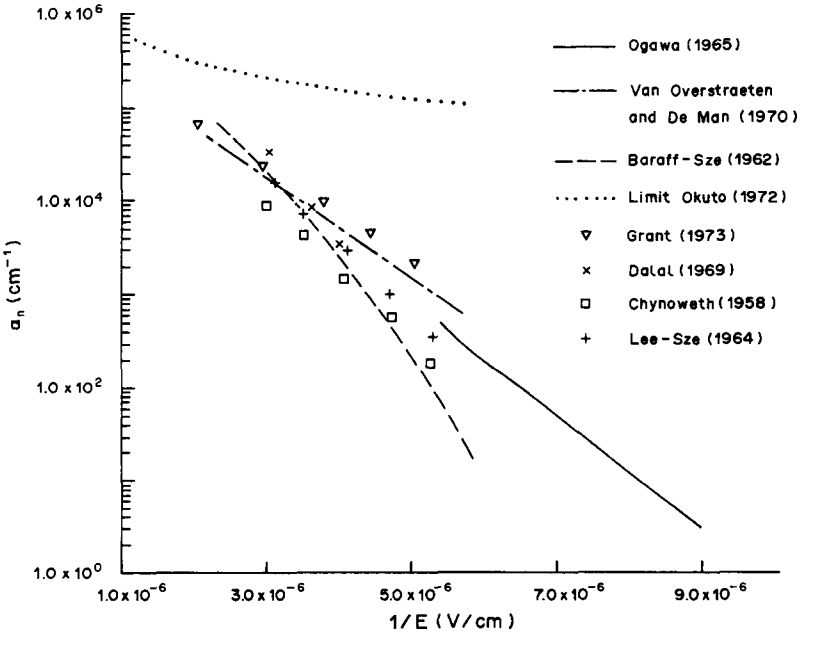

Fig. 2. Log(α_n), given by the Chynoweth model, as a function of electrical field, extracted by various authors.

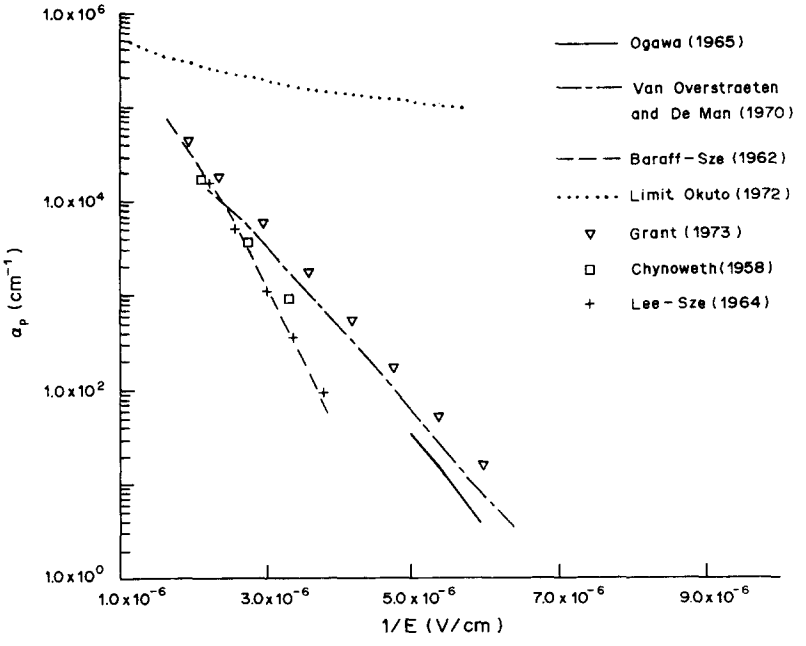

Fig. 3. Log(α_p), given by the Chynoweth model, as a function of electrical field, extracted by various authors.

of the ionization rates is smaller at higher electrical fields. Sutherland[89] found another approximation for the theoretical Baraff curves.

A dependence of ionization rate on crystal orientation has not been found[74,77,90].

5. MULTIPLICATION MEASUREMENTS ON A BIPOLAR TRANSISTOR. EXTRACTION PROCEDURE FOR THE IONIZATION RATE

5.1. General remarks about parameter extraction

From experimental values for the multiplication factor and from an ionization model, it is possible to extract the ionization coefficients from the measured data set. Parameter extraction can be seen as the link between the real physical world (measurements on test structures) and the simulation world (the ionization model with its parameter values in a device simulator). The task of the extraction program is to find a parameter set for the model which minimizes an error function. This error function is in most cases the root mean square value of the relative differences defined as:

$$\Phi = \sum_{i=1}^{N} \left(\frac{S_i(\boldsymbol{\theta}) - M_i}{M_i} \right)^2 \frac{1}{N} \tag{22}$$

with M_i the measured value of the multiplication factor for the biaspoint i, $S_i(\boldsymbol{\theta})$ the calculated multiplication factor for a given parameter vector $\boldsymbol{\theta}$ of the ionization model for bias condition i while N is the number of datapoints. Minimizing this function gives the optimal parameter set for the given data set and given error criterion. It is sometimes preferable to choose a different criterion. If, for example, the measured data set extends over several orders of magnitude, it is more desirable to take the deviation of the logarithm of the measured quantity. If the obtained coefficient values are implemented in a device simulator, the user must be aware that his simulation results are only valid for the field interval out of which they were extracted.

Good starting values for the fitting process are also important because the residual plane defined by eqn (22) can have more than one minimum, especially if a large number of parameters are fitted simultaneously. Depending on the starting values for the iteration, the fitting algorithm will go to the nearest minimum even if this is a local instead of the global minimum. We used our general purpose parameter extraction program SIMPAR[91] for the extraction procedure.

Next to finding the optimal parameter values, another important aspect is the calculation of the sensitivity of the different parameters, especially if these values are used for explaining physical phenomena. The shape of the residual plane around the minimum tells us how much a parameter may vary in order to increase the residual by a certain amount. It also informs us on how the parameters are correlated to each other. This can be calculated by approximating the residual function by an ellipsoid around the minimum as explained briefly in Appendix A. We will apply this theory to the measurements on our bipolar test structure.

5.2. Experimental results

A special *NPN* bipolar test structure was developed for the purpose of measuring multiplication factors at relatively low electrical fields. As already described in Section 3, the bipolar transistor is an appropriate structure for measuring low multiplication factors

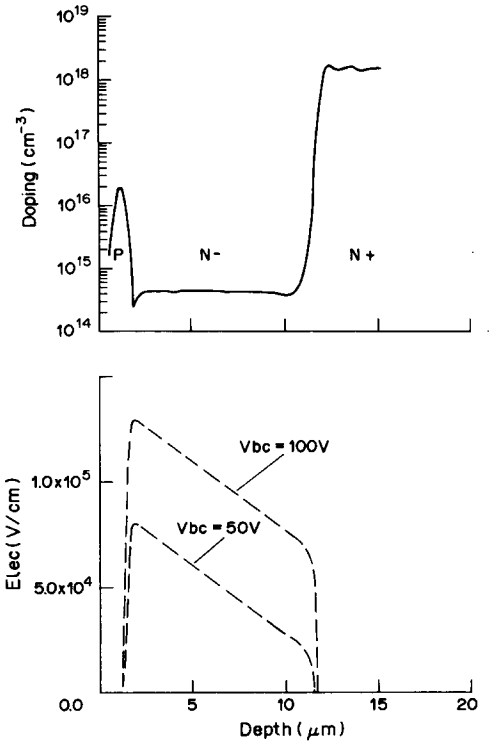

Fig. 4. Doping profile of the bipolar transistor obtained with SRP measurement together with the calculated electrical field in the base–collector region for two different values of V_{bc}.

because of its internal gain factor. The transistor is processed on an N^-/N^+ epi-wafer. The doping profile of the base-collector region, obtained with SRP-measurements, together with the calculated electrical field distribution for two different V_{BC} bias voltages, is shown in Fig. 4. This structure allows the measurement of the multiplication factor at relatively low electrical fields (8.0×10^4–1.5×10^5 V/cm). The N^- region has a doping level of about 4.0×10^{14} cm^{-3} and extends over almost 8 μm. The structure breaks down at the edges for a V_{BC} value of about 100 V.

For the measurement, a constant emitter current is injected and the base current is measured for different

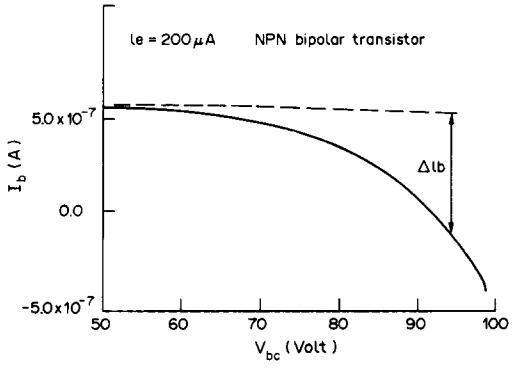

Fig. 5. Measured base current as a function of V_{bc} for a constant emitter current of 200 μA. ΔI_b is a measure for the multiplication factor, as given by eqn (23).

Table 1. Extracted ionization parameters for the bipolar transistor measurements at different temperatures

Temp (°C)	α_n^∞	b_n	Residual (%)
10	3.010×10^5	1.135×10^6	0.5
40	3.318×10^5	1.174×10^6	1.5
70	2.990×10^5	1.186×10^6	0.9
100	3.593×10^5	1.229×10^6	3.4
130	2.951×10^5	1.225×10^6	0.3
160	2.428×10^5	1.216×10^6	0.6

Although the fitting residual is very low, no direct relationship between the fitted parameter values as a function of temperature is observed.

V_{BC} values. The value of the current is not critical as long as the mobile carriers do not disturb the electrical field in the collector junction. The electrons injected into the collector space-charge region create electron–hole pairs. The holes travel back and the current they generate is superimposed on the base hole current, causing the measured base current to decrease as indicated in Fig. 5. This decrease in base current ΔI_B, is a direct measure of the generated current if the influence of base modulation is negligible, which is the case for our structure. The decrease ΔI_B can be expressed as:

$$\Delta I_B(V_{BC}) = I_E[M(V_{BC}) - 1] \tag{23}$$

with I_E the injected current in the emitter. These measurements were performed at different temperatures between 10 and 160°C, allowing us to investigate the temperature dependence of the ionization coefficients.

Our parameter extraction program SIMPAR[91] finds a very low sensitivity for the hole ionization coefficients which appear in the multiplication integral. This is normal because the ionization probability for holes is much lower than for electrons, especially for low electrical fields. For this reason, only the ionization rates for electrons are extracted. The extraction results at different temperatures are summarized in Table 1. It can be observed that the fitting error is very low. The dependency of the coefficients on temperature will be discussed in the next subsection. Around room temperature (40°C), the results of Van Overstraeten and De Man[8] and the newly obtained ionization coefficients are shown in Fig. 6. The new results extend the older data to lower electrical fields. The agreement with the extrapolated Van Overstraeten and De Man data is excellent. Also the recent data of Slotboom *et al.* in the intermediate electrical field range are shown. They used also a bipolar transistor to determine the low multiplication values and their data are also in agreement with our data.

5.3. Discussion of the obtained extraction results

As shown in Figs 2 and 3, different authors find strongly different ionization coefficients as optimal parameter values for their measurements. A possible explanation for this discrepancy can be the interaction between the two coefficients α^∞ and b for a

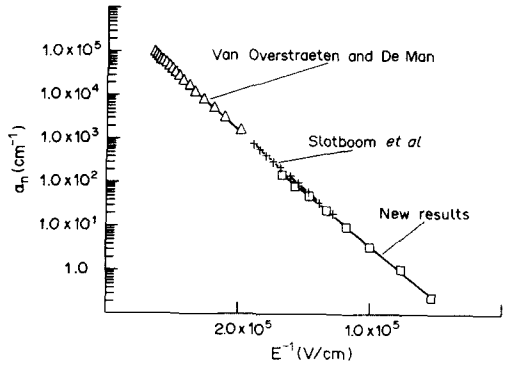

Fig. 6. This plot shows the Chynoweth expression as a function of electrical field for electrons with the ionization coefficients obtained by Van Overstraeten and De Man[8], Slotboom[52] and our new results.

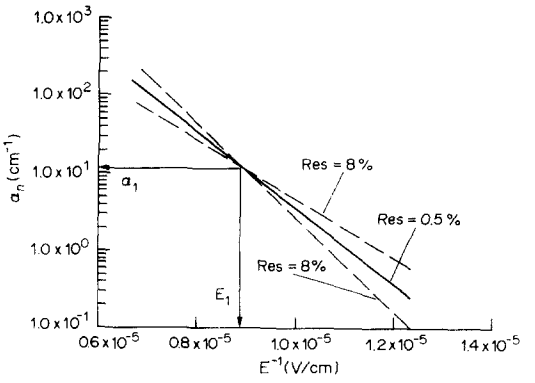

Fig. 8. The full line on this figure shows α_n vs $1/E$ for the fitted ionization coefficients of the Chynoweth expression. The dashed lines indicate the boundaries for the variation of the model parameters, in order to have a residual of less than 8%.

given data set. Because there is always an uncertainty in the measurements, different parameter combinations can give a fitting error which falls within this uncertainty region. As an illustration, we take our measurements on the bipolar transistor at room temperature (40°C) and fit the parameter b_n for different fixed values of α_n'. The relationship shown in Fig. 7 is obtained. On this line and between the points marked by U and W the residual is always less than 8%. This 8% residual value is chosen rather arbitrarily but can be considered as realistic for the uncertainty in the measurements. The best fitting result (residual of about 0.5%) is indicated by the point V. It must be pointed out that this "low-residual" line is well defined, meaning that small deviations for the parameter values from this line give large fitting errors.

In Fig. 8, the typical α_n vs $1/E$ plot is shown together with the boundaries of the region of uncertainty. The full line gives us the Chynoweth expression with the best fitted parameter set (point V in Fig. 7) while the dashed lines indicate the Chynoweth expression for points U and W in Fig. 7. It is the full line that also is shown on Fig. 6. All these lines which can be derived out of the relationship found between the two parameters in Fig. 7 can all be accepted as correct if an uncertainty of 8% is taken. They all go through one point (α_1, E_1). This point depends on the measured data set and of course on the test structure. The effect of this uncertainty region for the residual on the variation of the ionization coefficients can be understood as follows: because the ionization model is integrated over the whole electrical field distribution to obtain the multiplication factor, it can be intuitively understood that other coefficients which give a higher (lower) generation rate at high fields and a lower (higher) generation rate at lower fields than the curve obtained with the best fitted coefficients, give almost the same error residual for a given test structure.

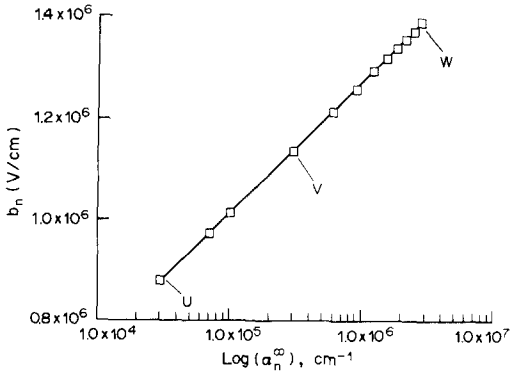

Fig. 7. All parameter couples lying on the indicated line between U and W have a residual of less than 8% while the point V gives the lowest residual (0.5%). This line is sharply defined for the given test structure and data set, meaning that small deviations from it will result in a very high residual value.

To reduce this region of uncertainty, measured data should be collected over a large number of different test structures as was done by Van Overstraeten and De Man[8]. As an illustration, we keep the parameter α_n' fixed to their value at 7×10^5 and fit the parameter b_n for the different temperatures. The results are summarized in Table 2. As can be observed, the fitting error is still very reasonable and the obtained values for b_n now increase monotonically with temperature. Also the extracted value for b_n around room temperature (40°C) compares well with the value found by Van Overstraeten and De Man[8].

Table 2. The same extraction procedure as explained for Table 1 but the parameter α_n' was kept fixed at 7×10^5 and b_n was fitted

Temp (°C)	α_n'	b_n	Residual (%)
10	7×10^5	1.23×10^6	3.0
40	7×10^5	1.258×10^6	3.0
70	7×10^5	1.282×10^6	3.3
100	7×10^5	1.305×10^6	4.3
130	7×10^5	1.323×10^6	3.5
160	7×10^5	1.337×10^6	4.3

The fitting residual is still low and extracted values for b_n increases monotonically with temperature.

To see whether the relationship between α^∞ and b also holds for another test structure and for higher electrical fields, a diode profile is simulated and the multiplication data are generated with the coefficients given in Ref. [8]. We find an uncertainty relation between α and $1/E$ totally similar to the one given in Fig. 8, proving that our considerations about the relationship between α^∞ and b are general. This explains the rather large discrepancy between the experimental α vs $1/E$ results of several authors, as shown in Figs 2 and 3. In contrast with Van Overstraeten and De Man, most authors measured their ionization rates on a very limited number of diodes and using one device structure. Within a reasonable uncertainty of the measurements, the α vs E relation can be rotated around a certain point (α_1, E_1) as shown in Fig. 8. Taking this consideration into account, most experimental results can be made to agree rather well.

The question arises whether it would be possible to predict the characteristics of the residual function around the minimum mathematically. This can indeed be done by approximating this function by its first three Taylor terms as explained in Appendix A, i.e. the region around the optimum point is approximated by an ellipsoid. The direction and the lengths of the axes of this ellipsoid give a good idea for the sensitivity and correlation of the different parameters for small perturbations of the fitting residual. If we apply this theory to the bipolar transistor measurements at room temperature, the results given in Table 3 are obtained for a residual perturbation Δ of 1%. This table shows e.g. that the parameter α_n^∞ may be varied relatively by 0.3% in order to increase the overall residual with $\Delta = 1\%$. The sensitivity of the parameter b_n is even higher since it can only vary by 0.05% to yield a Δ of 1% on the residual.

The factor ρ in Table 3 is an indication of the correlation between the two parameters; a strong correlation exists when $|\rho|$ is close to 1. To have an idea about the real behaviour of the residual function, the function was calculated for our bipolar multiplication measurements and visualized as a function of the two parameters. This three-dimensional plot is shown in Fig. 9 for a small parameter interval. It illustrates clearly that there exists a "valley" where the residual is low instead of only one correct parameterset. It can be proven that the appearance of a pronounced minimum (rather than a slowly varying valley) is more likely if the range of experimental data is increased.

Table 3. Allowed parameter variation for a region of uncertainty of 1% on the multiplication data set of the bipolar transistor at 40°C

	α_n (new)	α_n (Van Overstraeten–De Man, 1970)
α_n^∞	3.318×10^5	7.030×10^5
b_n	1.174×10^6	1.231×10^6
Φ	1.54%	
Δ	1%	
$\Delta\alpha_n^\infty$	922	
$\left(\frac{\Delta\alpha_n^\infty}{\alpha_n^\infty}\right)$	0.3%	
Δb_n	528	
$\left(\frac{\Delta b_n}{b_n}\right)$	0.05%	
ρ	0.6	

Remark the strong sensitivity for both parameters especially for b_n.

Taking a reasonable measurement error into account and performing a sensitivity analysis of the extracted parameters as explained in this section, we find that most experimental results can be made to agree rather well and that the accuracy of the parameters can be increased by enlarging the experimental conditions for which they are extracted (more structures, larger electrical field range). The latter is the reason for which the parameters of Van Overstraeten and De Man are probably the most accurate ones.

6. IMPLEMENTATION OF AN AVALANCHE MODEL IN A DEVICE SIMULATOR

Since most device simulators use local generation models, we will discuss in this section the implementation of such a model in the equations of the simulator. Using a device simulator with an incorporated avalanche model causes the CPU time to increase drastically. Small changes of the potential over the grid points during the iteration loop can cause a relatively large change in the number of generated carriers because of the strong dependence of the generation term on the electrical field.

It can even happen that the simulated test structure breaks down during an intermediate iteration step causing the program to diverge. This can be overcome by simulating the test structure first without the

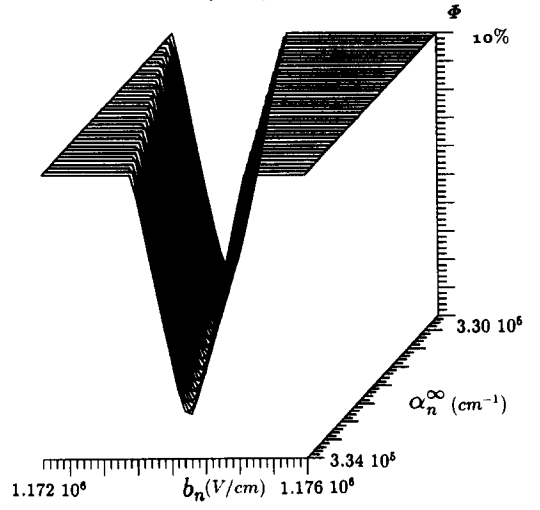

Fig. 9. Calculated residual plane for the parameters α_n^∞ and b_n for the multiplication data set of the bipolar transistor at 40°C. The maximum value for the residual is clipped at 0.1 (or 10%). It is clear that the optimum parameter set is not a point but is rather a narrow valley.

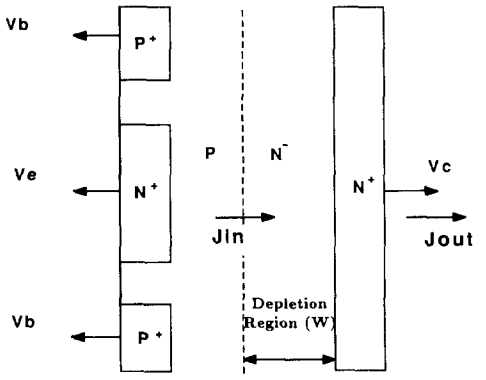

Fig. 10. Definition of the simulated bipolar structure with our 2-D device simulator PRISM. The collector–base junction consists of a lowly doped region where the electrical field is constant over W = 1 μm width. The hole and electron currents can be calculated at different values of V_{BC}. The recombination term is taken as zero.

avalanche term. Then these intermediate results can be used as a starting point for the simulation with the avalanche generation term included.

Another source of error is the discretization of the structure. If the discretization mesh is coarse, the results can be erroneous even when the program converges. This is clearly illustrated with the simulation of a simple *NPN* bipolar transistor as shown in Fig. 10. The problem is essentially one-dimensional so the computer simulation can easily be compared with exact analytical calculations. The 1 μm wide collector region contains a lightly-doped N^- region where the electrical field is constant. For the simulation, the recombination term is excluded, allowing us to calculate exactly both the current injected by the emitter and the base hole current as a function of applied base–collector voltage. The simulated results and the analytical calculations are compared in Fig. 11. When the high-field collector region is divided into ten mesh points in the direction of the current, the simulated current vs base–collector voltage agrees very well

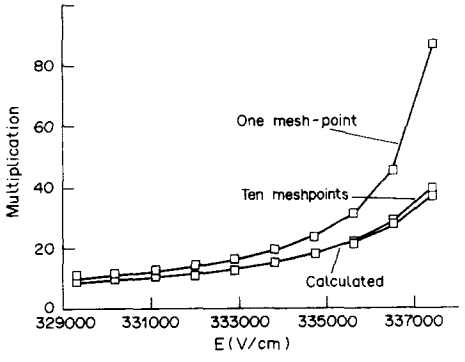

Fig. 11. The simulated multiplication factors on the structure defined in Fig. 10 as a function of electrical field over the base–collector junction by discretizing this region respectively with one meshpoint and with ten meshpoints in the direction of the current. The computed results are compared with analytically calculated values.

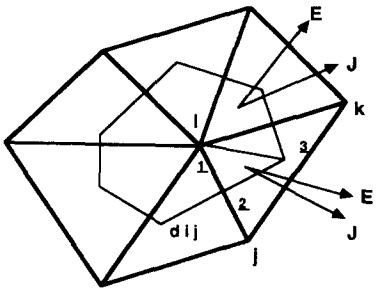

Fig. 12. A piece of a finite difference mesh around the node *i*. A cell which contains all near neighbour points can be constructed around the node. For each sub-triangle of the cell, a different current and electrical field can be determined.

with the analytically calculated values. If this same collector region is discretized with only one meshpoint, the multiplication deviates from the analytical solution and the deviation becomes larger as the electrical field increases. The conclusion is that even when there is a region with a very low doping level where the voltage drop is linear, a fine grid is required to correctly simulate the current increase due to avalanche generation.

A practical approach is to derive a method to implement a local generation model in a way which is almost grid independent. To illustrate the problem, the way the generation term is used in most device simulators is studied. The current continuity eqn (3) can be discretized using the divergence theorem, and referring to the 2-D finite element mesh of Fig. 12, in the following way for node i:

$$\sum_j J_{n,p}^{i,j} d_{i,j} = \sum_{\text{neighbours}} G_{n,p}^{i,j,k} \left(\frac{A_{i,j,k}}{3} \right) \tag{24}$$

with $A_{i,j,k}$ the area and G the generation rate over the whole mesh triangle. The sum must be taken over all neighbour nodes j of node i. The currents are defined on the lines between the nodes with $d_{i,j}$ the length of the bisectors of the different node lines. Around each node, a cell is created which contains all points nearest to that node. That is the reason why only one third of the area of all neighbouring mesh-triangles must be taken into account in the generation term of the pertinent node. The left side of this equation is solved simultaneously with all other equations in an internal coupled loop of the device simulator. The right-hand side (generation term) must be recalculated between every iteration. It is this calculation which introduces the sensitivity to the chosen mesh. This term is implemented in most 2-D device simulators by

$$G_{n,p} \approx \frac{\alpha_{n,p}(E_J)|J_{n,p}|}{q}. \tag{25}$$

The advantage of this formula is its simplicity, but the drawback is that the result is only correct for very small-sized triangles. This difficulty can be overcome by taking the direction of the current J and the shape

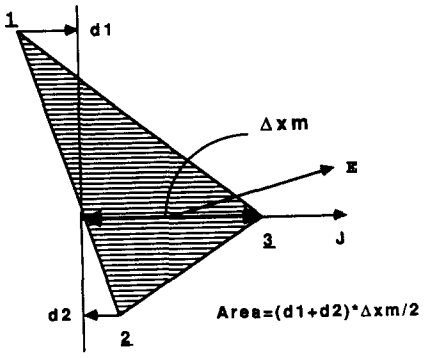

Fig. 13. The different variables used to determine the generation rate over a triangle obtained from integration between the boundaries d_1 and d_2 which depend on the direction of the current.

of the triangle into account in calculating the generation term. Also the effect of "secondary" electrons and holes which can produce other pairs within the triangle must be included.

The mesh triangle can be divided into different sub-triangles by constructing a nearest region for every node as pictured in Fig. 12. When the electrical field and the current direction are assumed constant over such a sub-triangle (as indicated on Fig. 12 by the coordinates [1, 2, 3]), the exact generation rate in the direction of the current can be calculated analytically as a function of distance. This distance varies linearly perpendicular to the current direction and by integrating between the boundaries d_1 and d_2 as visualized in Fig. 13, the exact generation rate over the triangle can be found. For a more detailed explanation of this calculation, we refer to Ref. [92]. The formula obtained is given below. It should be compared with expression (25).

$$G_n = \left[\frac{\log\left(\dfrac{\alpha_n - \alpha_p}{\alpha_n - \alpha_p \exp[(\alpha_n - \alpha_p)\Delta X_M]} \right)}{\alpha_p \Delta X_M} - 1 \right] \times * \left(\frac{2|J_n|}{\Delta X_M q} \right), \quad (26)$$

with ΔX_M the maximum distance over the triangle in the direction of the current. This result shows that the generation rate depends only on the area of the triangle and on ΔX_M, which on its turn depends on the direction of the current and on the shape of the triangle. In this way the generation rates of the different sub-triangles in which node i is embedded (see Fig. 12) can be summed to give the total generation term for node i for the incoming electron and hole current.

To investigate the impact of this new implementation method, we calculate the generation term for an incoming current of unit value on a right-angle triangle shape as a function of the current direction over the triangle. Figure 14 shows the plot of this generation rate as a function of the current angle. Due to symmetry considerations, one can easily

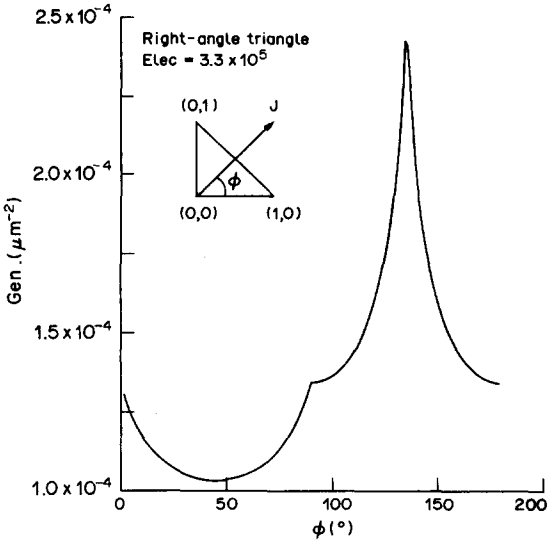

Fig. 14. The calculated generation rate for a right-angle triangle as a function of the current angle. The peak generation rate occurs when the current is parallel to the hypotenuse of the triangle.

observe that the generation rate at 0° for the current is the same as for 90°. It is also important to remark that the generation rate gives a peak at 135° for the current direction when the current path is parallel to the hypotenuse of the triangle where the maximum ΔX is found for the whole triangle resulting in a maximum of generation. It is easy to see that the generation term can differ appreciably depending on the current direction which was not taken into account with the old generation model.

7. CONCLUSIONS

After defining the terms multiplication factor and avalanche generation rate, different multiplication and avalanche generation models have been discussed. Most authors use the Chynoweth expression to fit their multiplication data. When comparing the different experimental parameter sets, an important spread is observed. We have presented new electron multiplication measurements on a bipolar transistor at relatively low electrical fields and at different temperatures for injection by electrons. The data are in agreement with those of Van Overstraeten and De Man at higher electrical fields and those of Slotboom *et al.* at intermediate electrical fields. It is shown that an interdependence between the two parameters α^{∞} and b of the Chynoweth expression exists for a given data set. This can be translated into a region of uncertainty for the obtained results. These considerations can explain the discrepancy of the experimental results among different authors. Most authors use a limited number of measurements such that within a reasonable uncertainty, their relation of $\log(\alpha)$ vs $1/E$ can be rotated around a point (α_1, E_1). The coefficients of Van Overstraeten and De Man can be

considered accurate because they are based on a large number of measurements on different diodes giving α values in different ranges of the electrical field.

For very high electrical fields, the Chynoweth expression is not adequate to describe the avalanche generation. Baraff developed a theory valid in the different field regions. Also the Thornber model combines the Shockley and Wolff model into one expression, making it valid over a wide range of electrical fields but good parameter values are not yet available. Recently the lucky electron model has been extended and made more suitable for use in a non-local model. For simulating MOSFET transistors, a local generation model cannot always be used because of the low voltage drop and strong electrical field gradients in the avalanche region. General non-local ionization models are however difficult to implement in a device simulator because the "history" of the particle has to be taken into account.

Implementing a local avalanche generation model in a device simulator is a delicate task because of the strong dependence of the carrier generation as a function of electrical field. This results in a slow convergence of the algorithm and as a consequence in a large amount of CPU time. The existing implementation methods of a local generation model in a device simulator often give results which depend on the choosen discretization mesh size. In this paper, a method is proposed with which we calculate exactly the generation rate over one mesh-triangle of a 2-D finite element mesh and by which the simulation results become much more insensitive to the chosen grid size.

REFERENCES

1. R. J. McIntyre, *IEEE Trans. Electron* **ED-13,** 164 (1966).
2. P. P. Webb, R. J. McIntyre and J. Conradi, *RCA Rev.* **35,** 234 (1974).
3. A. G. Chynoweth, *Phys. Rev.* **109,** 1537 (1958).
4. R. Van Overstraeten and H. De Man, *Electron. Lett.* **3,** 10 (1967).
5. S. L. Miller, *Phys. Rev.* **105,** 1246 (1957).
6. J. L. Moll, *Physics of Semiconductor Devices.* New York (1964).
7. J. L. Moll and R. Van Overstraeten, *Solid-St. Electron.* **6,** 147 (1963).
8. R. Van Overstraeten and H. De Man, *Solid-St. Electron.* **13,** 583 (1969).
9. F. Van De Wiele, R. Van Overstraeten and H. De Man, *Solid-St. Electron.* **13,** 25 (1970).
10. G. Beni and F. Capasso, *Phys. Rev. B* **19,** 2197 (1979).
11. G. V. Mânduteanu, *IEEE Trans. Electron Devices* **ED-32,** 2492 (1985).
12. R. Leguerre, *Solid-St. Electron.* **19,** 875 (1967).
13. W. Fulop, *Solid-St. Electron.* **10,** 39 (1967).
14. P. Spirito, *IEEE Trans. Electron Devices* **ED-21,** 226 (1974).
15. Y. Okuto and C. R. Crowell, *Solid-St. Electron.* **18,** 161 (1974).
16. J. Urgell and L. R. Leguerre, *Solid-St. Electron.* **17,** 239 (1974).
17. I. Masanori *et al.*, *Solid-St. Electron.* **30,** 969 (1987).
18. C. Basavana Goud and K. N. Bhat, *Solid-St. Electron.* **30,** 787 (1987).
19. S. M. Sze and G. Gibbons, *Solid-St. Electron.* **9,** 831 (1966).
20. J. Ueda and N. Totsuka, *Solid-St. Electron.* **28,** 1245 (1985).
21. C. Basavanagoud and K. N. Bhat, *IEEE Electron Device Lett.* **EDL-6,** 276 (1985).
22. V. Temple and M. Adler, *IEEE Trans. Electron Devices* **ED-22,** 910 (1975).
23. A. S. Grove, O. Leistiko and W. W. Hooper, *IEEE Trans. Electron Devices* **ED-14,** 157 (1967).
24. A. Rusu, O. Pietrăreanu and C. Bulucea, *Solid-St. Electron.* **23,** 473 (1980).
25. C. Bulucea, *Solid-St. Electron* **18,** 363 (1975).
26. C. Bulucea, *Solid-St. Electron.* **18,** 381 (1975).
27. J. F. Verwey *et al.*, *Solid-St. Electron.* **20,** 689 (1977).
28. P. Rutter, *Solid-St. Electron.* **23,** 441 (1980).
29. K. Hwang and D. Navon, *IEEE Trans. Electron Devices* **ED-31,** 1126 (1984).
30. H. Egawa, *IEEE Trans. Electron Devices* **ED-13,** 754 (1966).
31. A. Schütz, S. Selberherr and H. Pötzl, *IEEE Trans. CAD* **CAD-1,** 77 (1982).
32. A. Schütz, S. Selberherr and H. Pötzl, *Solid-St. Electron.* **25,** 177 (1982).
33. F. Hsu *et al.*, *IEEE Trans. Electron Devices* **ED-29,** 1735 (1982).
34. T. Toyabe *et al.*, *IEEE Trans. Electron Devices* **ED-25,** 825 (1978).
35. S. Ochi *et al.*, *IEEE Trans. Electron Devices* **ED-26,** 399 (1980).
36. I. Yoshida *et al.*, *IEEE Trans. Electron Devices* **ED-27,** 395 (1980).
37. M. Declercq and J. D. Plummer, *IEEE Trans. Electron Devices* **ED-23,** 1 (1976).
38. J. Matsunga *et al.*, *Proc. 11th Conf. Solid-State Devices*, p. 93 (1979).
39. C. M. Wu and K. W. Yeh, *IEEE Electron Devices Lett.* **EDL-3,** 245 (1982).
40. W.-S. Feng, T. Y. Chan and C. Hu, *IEEE Electron Device Lett.* **EDL-7,** 449 (1986).
41. K. Brennan and K. Hess, *IEEE Electron Device Lett.* **EDL-7,** 86 (1986).
42. A. Goetzberger and E. H. Nicollian, *J. appl. Phys.* **38,** 4582 (1967).
43. E. Sun, J. Moll, J. Berger and B. Alders, *IEDM Proc.*, p. 478 (1978).
44. T. Tsuchiya and S. Nakajima, *IEEE Trans. Electron Devices* **ED-32,** 405 (1985).
45. F. K. Baker and J. R. Pfiester, *IEEE Trans. Electron Devices* **35,** 2119 (1988).
46. C. Chang *et al.*, *IEEE Electron. Device Lett.* **9,** 588 (1988).
47. B. A. McDonald, *IEEE Trans. Electron Devices* **ED-17,** 871 (1970).
48. S. P. Li, E. T. Bates and J. Maserjian, *Solid-St. Electron.* **19,** 235 (1976).
49. J. M. Higman, I. C. Kizilyalli and K. Hess, *IEEE Electron Device Lett.* **9,** 399 (1988).
50. A. G. Chynoweth and K. G. McKay, *Phys. Rev.* **108,** 29 (1957).
51. W. Sayle and P. Lauritzen, *IEEE Trans. Electron Devices* **ED-18,** 58 (1971).
52. J. W. Slotboom, G. Streutker, G. J. T. Davids and P. B. Hartog, *IEDM Proc.*, p. 494 (1987).
53. P. A. Wolff, *Phys. Rev.* **95,** 1415 (1954).
54. W. Shockley, *Solid-St. Electron.* **2,** 35 (1961).
55. G. A. Baraff, *Phys. Rev.* **128,** 2507 (1962).
56. G. A. Baraff, *Phys. Rev.* **133,** A26 (1964).
57. Y. Okuto and C. R. Crowell, *Phys. Rev. B* **6,** 3076 (1972).
58. K. K. Thornber, *J. appl. Phys.* **52,** 279 (1981).

59. M. H. Woods, W. C. Johnson and M. A. Lampert, *Solid-St. Electron.* **16,** 381 (1973).
60. W. N. Grant, *Solid-St. Electron.* **16,** 1189 (1973).
61. B. K. Ridley, *J. Phys. C: Solid-State Phys.* **16,** 3378 (1983).
62. B. K. Ridley, *J. Phys. C: Solid-State Phys.* **16,** 4733 (1983).
63. B. K. Ridley, *Semicond. Sci. Technol.* **2,** 116 (1987).
64. R. C. Woods, *IEEE Trans. Electron Devices* **ED-34,** 1116 (1987).
65. R. C. Woods, *Appl. Phys. Lett.* **52,** 65 (1988).
66. J. S. Marsland, *Solid-St. Electron.* **30,** 125 (1987).
67. P. A. Childs, *J. Phys. C: Solid-State Phys.* **20,** L243 (1987).
68. P. A. Childs, *NASCODE V—Proc. fifth Int. Conf. Numerical Analysis of Semiconductor Devices and Integrated Circuits*, pp. 17–19, Dublin (1987).
69. Y. Z. Chen and T. W. Tang, *J. appl. Phys.* **65,** (1989).
70. L. V. Keldysh, *Sov. Phys. JETP* **21,** 1135 (1965).
71. S. Bandyopadhyay, M. Klausmeier-Brown, C. Maziar, S. Datta and M. Lundstrom, *IEEE Trans. Electron Devices* **ED-34,** 392 (1987).
72. T. Thurgate and N. Chan, *IEEE Trans. Electron Devices* **ED-32,** 400 (1985).
73. R. Kuhnert, C. Werner and A. Schütz, *IEEE Trans. Electron Devices* **ED-32,** 1057 (1985).
74. K. G. McKay and K. B. McAfee, *Phys. Rev.* **91,** 1079 (1953).
75. K. G. McKay, *Phys. Rev.* **94,** 877 (1954).
76. A. G. Chynoweth, *J. appl. Phys.* **31,** (1964).
77. C. A. Lee *et al.*, *Phys. Rev.* **134,** A761 (1964).
78. J. J. Moll and N. I. Meyer, *Solid-St. Electron.* **3,** 155 (1961).
79. J. J. Moll and N. I. Meyer, *Solid-St. Electron.* **6** (1963).
80. T. Ogawa, *Jap. J. appl. Phys.* **6,** 473 (1963).
81. V. L. Dalal, *Appl. Phys. Lett.* **15** (1969).
82. S. Selberherr, S. Schütz and H. Pötzl, *Two Dimensional MOS Transistor Modeling, Summer Course on VLSI Process and Device Modeling*, Katholieke Universiteit Leuven, Belgium (1983).
83. S. M. Sze and G. Gibbons, *Appl. Phys. Lett.* **5** (1966).
84. S. M. Sze, *Physics of Semiconductor Devices.* Wiley, New York (1981).
85. Y. Okuto and C. R. Crowell, *Phys. Rev. B* **10** (1974).
86. J. R. Hauser, *J. appl. Phys.* **37,** 507 (1966).
87. G. V. Mânduteanu, *Int. J. Electron.* **56,** 555 (1984).
88. C. R. Crowell and S. M. Sze, *Appl. Phys. Lett.* **9,** 242 (1966).
89. A. D. Sutherland, *IEEE Trans. Electron Devices* **ED-27,** 1299 (1980).
90. V. M. Robbins *et al.*, *J. appl. Phys.* **58,** 4614 (1984).
91. W. Maes, K. De Meyer and L. Dupas, *IEEE Trans. CAD* **CAD-5,** 320 (1986).
92. Esprit Report: Esprit 962E-17, Progress in Work package 1: Physical models and Validation (1989).
93. Y. Bard, *Non-linear Parameter Estimation.* Academic Press, New York (1974).

APPENDIX

Sensitivity Calculation on Extracted Parameter Values

In general, data fitting and parameter extraction are carried out mathematically by minimizing a specified objective function $\Phi(\boldsymbol{\theta})$. For any proposed analytical multiplication model, the objective function can be specified as follows,

$$\Phi = \sum_{i=1}^{N} \left(\frac{S_i(\boldsymbol{\theta}) - M_i}{M_i} \right)^2 \frac{1}{N} \tag{27}$$

with M_i the measured multiplication factor, N the number of data points and $S_i(\boldsymbol{\theta})$ the calculated multiplication factor for a given model parameter vector $\boldsymbol{\theta}$. The extraction program calculates a best fit parameter vector $\boldsymbol{\theta}^*$ but due to the experimental uncertainties, there is no reason to prefer $\boldsymbol{\theta}^*$ above any other value of $\boldsymbol{\theta}$ for which

$$|\Phi(\boldsymbol{\theta}) - \Phi(\boldsymbol{\theta}^*)| \leqslant \epsilon \tag{28}$$

where ϵ is an allowed fitting error caused by experimental uncertainties. These uncertainties can have different origins. A first error source can come from noise on the measurements. It can also arise from the fact that the model itself is not perfect in describing the given physical phenomena or that the test structure is not correctly characterized. All these error sources create an uncertainty interval for the different parameters.

One way of calculating this region of uncertainty is by approximating the ϵ indifference region by an ellipsoid. Therefore, the problem of parameter uncertainty is now converted into the problem of the determination of the length and orientation of the ellipsoidal axes. The residual function is expanded in a Taylor series keeping only the first three terms:

$$\Phi(\boldsymbol{\theta}) = \Phi(\boldsymbol{\theta}^*) + \mathbf{J}^{*\mathrm{T}}\delta\boldsymbol{\theta} + \tfrac{1}{2}\delta\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}^*\delta\boldsymbol{\theta}, \tag{29}$$

with $\boldsymbol{\theta}$ the parameterset, $\delta\boldsymbol{\theta} = (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$ while $\boldsymbol{J}^*$ and $\boldsymbol{H}^*$ are the Jacobian (first derivatives with respect to the parameters) and the Hessian matrix (second derivatives with respect to the parameters) evaluated in the minimum $\boldsymbol{\theta}^*$. This is clearly illustrated in Fig. A1 which shows the error function for one parameter and in Fig. A2 showing the contour plots for two parameters. Because the Taylor expansion is evaluated in the minimum of the residual plane, the Jacobian matrix will be zero. The Hessian matrix is defined as

$$H_{i,j} = \left[\frac{\partial^2 \Phi}{\partial\theta_i^* \partial\theta_j^*} \right]. \tag{30}$$

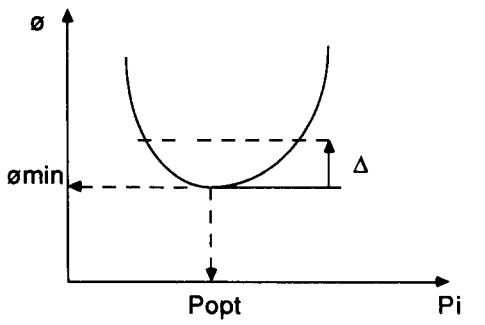

Fig. A1. This simple picture illustrates the shape of the residual plane around the optimal parameter value together with a region of uncertainty Δ.

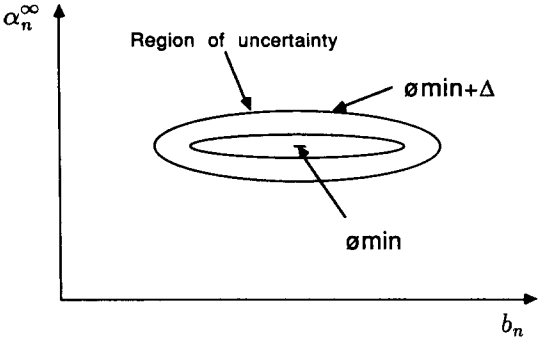

Fig. A2. Visualization of the contour plots showing the region of uncertainty for the two parameters α_n^∞ and b_n. This region is assumed to be an ellipsoid which has the dimension of the number of parameters.

What we are looking for is how much we can vary a parameter for a given increase (or uncertainty) in residual. For this purpose we use the definition of the covariance matrix

$$\mathbf{V} \approx 2\epsilon \mathbf{H}^{*-1}, \tag{31}$$

with ϵ the allowed uncertainty interval on the residual.

This interval can be estimated by using the matrix $\mathbf{V}$ defined above

$$\Delta\theta_i = \sqrt{V_{i,i}}, \tag{32}$$

and also the coupling or correlation between two different parameters can be expressed as

$$\rho_{i,j} = \frac{V_{i,j}}{\sqrt{V_{i,i}V_{j,j}}}. \tag{33}$$

For a maximum value of 1 for $|\rho|$, the correlation between the parameters θ_i and θ_j is very strong. A value for $|\rho|$ around 0 shows practically no correlation. More detailed information about this subject can found in Ref. [93].