THE MATHEMATICS CLINIC

Contract No. 95650

PARAMETER EXTRACTION AND TRANSISTOR MODELS

By

Charles Rykken, Team Leader
Verena Meiser (Fall)
Greg Turner
Qi Wang

Professor Ronald Gribben (Faculty Supervisor)
Professor Mario Martelli (Faculty Consultant)

FINAL REPORT TO

Jet Propulsion Laboratories

June 1985

Conducted at

Claremont Graduate School

This report was prepared for the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration.

FOREWORD

The problem which the Jet Propulsion Laboratory asked the Mathematics Clinic at the Claremont Graduate School to consider was twofold in character. In essence it consisted of the following:

(a) given specified mathematical models of the MOSFET device to extract, from data supplied by J.P.L., the optimal values of the model-dependent parameters;

(b) to assess the sensitivity of the several models to variations of the parameters from their optimal values.

This report describes the approach used, and the conclusions reached, by the Clinic in tackling these two questions. In the event, we confined ourselves to just three MOSFET models, all one-dimensional, and in one of which diffusion (as well as convection) currents are taken into account. Although we feel that significant progress has been made as regards the tasks (a) and (b) it is also our view that much still remains to be done for a fully comprehensive and systematic study.

It is a pleasure to thank all the individuals involved in the successful operation of the Clinic; the student members of the team for their perseverance when difficulties, often mystifying, occurred; Mike Robkin, second year Harvey Mudd College student, who was employed by the Clinic to carry out the bulk of the computing; to Professor Mario Martelli, the Faculty Consultant in the second semester, for his interest, inspiration and enthusiasm; to Professor Hedley Morris, who visited for a short time and whose 1984 Summer Clinic Report (in conjunction with Richard Everson) served as the

basis of the work of our Clinic; to Professor Ellis Cumberbatch who organised the creation of the Clinic; to Joy Marshall, the Mathematics Clinic secretary, for patient typing of unfamiliar, often indecipherable and seemingly endless mathematics; and last, but not least, to Cesar Pina, the liaison link with J.P.L., for his consistent help throughout the year and constant interest in the progress of the work.

A listing of all the programs and subroutines used by the Clinic has been produced as a supplement to this report and is available, upon request, to

Mathematics Clinic
Claremont Graduate School
Claremont, California 91711.

TABLE OF CONTENTS

INTRODUCTION

Chapter 1

The device known as the metal-oxide-semiconductor-field-effect-transistor, or MOSFET, is described in detail in many places. (See e.g. Sze (1981), Morris & Everson (1984)). Briefly, it consists of doped semiconducting material (silicon) to which are connected four terminals (see Figure 1) at the source, drain, bulk substrate and gate. The gate is separated from the main body of the device by a layer of non-conducting material such as silicon dioxide. The silicon has a doping profile, which means that it has been implanted within its crystal structure with impurity atoms of other elements. In this way we suppose that it has been made 'p-type' (by implanting, for example, with boron) in the bulk material substrate and lightly 'n-type' (doped, for example, with phosphorus) in the regions near the source and drain. Then under a sufficiently positive voltage V_{GS} (relative to the source) applied at the gate, an n-type inversion channel will be created in the silicon along which a drain current I_D flows when the drain voltage V_{DS} is sufficiently positive i.e. above some threshold value with respect to the source. The manner in which I_D depends on V_{DS} is illustrated in Figure 2 where typical curves are sketched for given fixed values of V_{GS}. Such curves can be obtained experimentally over a range of MOSFETs of different sizes and properties with good accuracy. The main features consist of a near linear growth of current with V_{DS} in the early stages, during which the MOSFET acts as a linear amplifier, followed by a rapid change to a

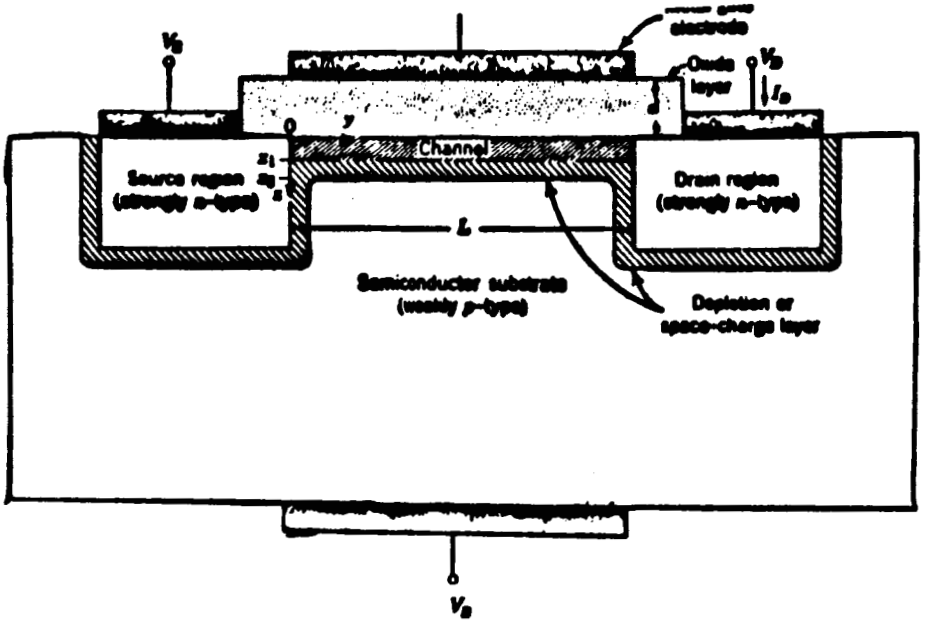


Figure 1

Sketch of n-channel MOSFET

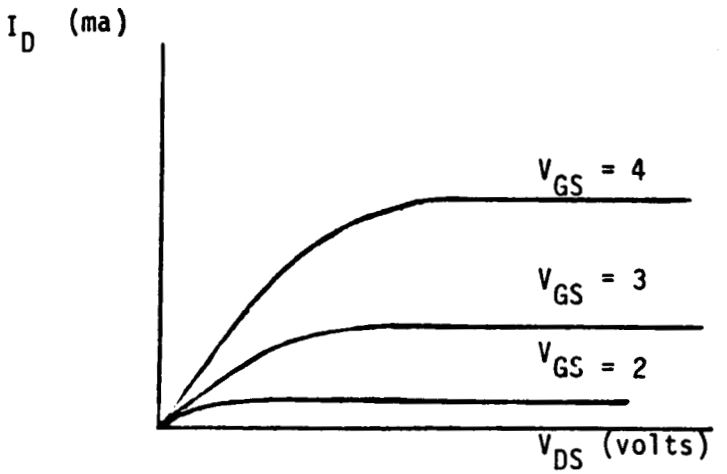


Figure 2

Typical I_D-V_{DS} curves for an n-channel MOSFET;

V_{GS} in volts.

domain in which I_D is nearly constant as V_{DS} increases. Subsequently, the material breaks down electrically and there is a final increase of I_D at very large V_{DS} (the avalanche region).

There have been developed in recent years many attempts at modelling mathematically the physical processes that occur in MOSFET operation. Such processes involve the appropriate Maxwell equation for the electrostatic potential ϕ , the methods of statistical mechanics to express charge densities in terms of ϕ , the Einstein relations to simplify the diffusion currents and Gauss's law to formulate an expression for the current along the channel. This clinic is closely associated with the analysis of such models, which suffer from the disadvantage that many of the physical parameters which enter into their construction cannot be measured, or even defined, with any degree of certainty. Examples of these parameters are the length and width of the channel, the mobility of the carriers within the channel and the degree of doping of the semiconductor: the small dimensions of MOSFETs renders the experimental determination of such quantities most imprecise. Consequently they must be deduced in an indirect manner and it is the principal objective of our report to describe a mathematical and numerical method by which this process can be carried out.

Thus, the 'fit' of a particular model to given data (as plotted on a diagram like that drawn in Figure 2) is optimized with respect to the parameters that the model contains. In this way values are obtained (or 'extracted') for the unknown parameters present in the model. The data is provided by JPL and consists of sets of measurements of I_D over

a range of values of V_{DS} for specified values of V_{GS} and the substrate bias voltage V_{BS}. Once reliable parameter values are known they may be incorporated within circuit simulation programs (see e.g. Vladimirescu and Liu (1980)) to predict behaviour in circuit design. A secondary objective is then concerned with the sensitivity of a model with respect to its parameters. For example, if a particular parameter is changed from its optimal value by, say 10%, how do the resulting I_D-V_{DS} curves deviate from the optimal one?

The MOSFET models that we examine are all one-dimensional models which assume that changes take place much more rapidly across the channel than in directions parallel to it. Thus the expressions for the current are derived on the basis of the 'slowly-varying channel' approximation and we should expect the theory to be more accurate for longer, wider MOSFETs. In fact, the data that we use is for MOSFETs of length from 1.2 μMs to 24 μMs and width from 2.5 μMs to 24 μMs. An interesting question is to determine the variation in accuracy obtained by the models over the different sizes of device. Unfortunately, the particular data sets provided do not permit a fully systematic study of this behaviour for different lengths of MOSFET of fixed width, and vice versa. However, the general trend can be ascertained and some comments in this connection are made in chapters 4 and 5.

The most straightforward of the models neglects diffusion currents compared with the drift currents and the derivation is described in Morris and Everson (1984). There are two variants of this model which we have tested against the data and they differ only in the form assumed for the expressions defining the effective mobility and the effective length of

the channel. The details are given in Chapter 2 and we refer to these models as the Ihantola model and the Spice 2 model. The latter is in a form which is suitable for use in the Spice circuit simulation program.

Now effects of diffusion currents can be important at small values of V_{GS} and so we also consider the simplest model which takes such effects into account. This is a so-called 'charge-sheet' model due to Brews (1978) and is derived in Chapter 2. We shall refer to this as the Brews model.

The methods of optimization used in fitting the model to the data are described in Morris and Everson (1984). (See also Chapter 3 of this report). Roughly speaking the objective function is taken to be the sum of squares of the differences between predicted and measured values of I_D and this quantity is to be minimized with respect to the parameters of the model. The nonlinear optimization is carried out using two different techniques and in practice these operate in tandem. Each uses a subroutine of the IMS library available on VAX. Firstly, a program SARAH which employs a Gauss-Newton method, is used to obtain a minimum with respect to some pre-set convergence criterion. The miniumum is constrained to lie within a hyperplane of the parameter space and is chosen as the deepest arising from a large number of initial 'guesses', so ensuring as far as possible that the value obtained is a global one. The process tends to be slow so, secondly, a program MOSES, using the Levenberg-Marquardt algorithm (essentially a hybrid, steepest descent - Newton scheme) refines this value to a desired (higher) accuracy.

The results obtained by applying the programs SARAH and MOSES to the

three models Ihantola, Spice 2 and Brews are described in detail in Chapter 4. The chief conclusions indicate that, for Ihantola and Spice 2, provided the data for V_{GS}=2 is omitted to remove the apparently important effects of diffusion currents, the accuracy obtained decreases as the model dimensions decrease although even at the smallest model (1.2 μMs x 2.5 μMs) RMS errors of only a few percent are obtained. The same is true of the Brews model, with generally a somewhat larger error, but here there is the important distinction that all V_{GS} values are included. The RMS errors are generally increased for non-zero values of V_{BS} in the Ihantola and Spice 2 models: the Brews model was not adapted to non-zero V_{BS}.

The results for the sensitivity of each model on its parameters are also given in Chapter 4. It is found that each model contains parameters on which it depends rather critically and others to which it is relatively insensitive.

Finally, in Chapter 5, an account is presented of our overall experience in applying the programs SARAH and MOSES. Comparison of performance of the different models is given and their strengths and weaknesses, together with some of the difficulties that were encountered. Suggestions for future development and extension of the parameter extraction technique are offered.

Chapter 2 MATHEMATICAL MODELS

The clinic has studied three one-dimensional mathematical models of the MOSFET, the first two of which are derived directly from that given by Ihantola and Moll (1964) and discussed in detail in Morris and Everson (1984). These models neglect diffusion currents and differ only in the assumed functional form for the effective mobility, μ_{eff}, and effective length, L_{eff}, of the device. We refer to these models as the Ihantola model and the SPICE 2 model and we summarize below in (a) and (b) their relevant formulae. An important aspect of all models is the presence of a number of parameters, P1, P2,...etc., which are not known accurately and, as described in Chapter 3, are to be determined for each model by optimizing the fit of the model to the available empirical data. The formulae then for the first two models are as follows. (Taken from Morris and Everson (1984))

(a) Ihantola Model

$$I_D = P5 \cdot \mu_{eff} \{(V_G - P1 - P7 - \frac{V_{DS}}{2})V_{DS} - \frac{2}{3} P2 \left[(V_{DS} - V_{BS} + P1)^{3/2} - (P2 - V_{BS})^{3/2}\right]\},$$

provided the drain voltage $V_{DS} < V_{DSAT}$ (the saturation voltage).

When $V_D < V_{DSAT}$

$$I_D = \frac{I_{DSAT}}{L_{eff}} \qquad (2)$$

where I_{DSAT} is obtained by substituting

$$V_D = V_{DSAT} = V_{GS} - P1 - P7 + \frac{(P2)^2}{2}\{1 - \left[1 + \frac{4}{(P2)^2}(V_{GS} - P7 - (P2)(P1)^{\frac{1}{2}} + P2(P1 - V_{BS})^{\frac{1}{2}})\right]^{\frac{1}{2}}\} \qquad (3)$$

into (1). The seven parameters defined in the Ihantola model, in terms of physical quantities, are

$$P1 = \frac{2}{\beta} \ell n \left(\frac{N_A}{n_i}\right)$$

$$P2 = \frac{(2\kappa_s q N_A)^{\frac{1}{2}}}{C_{ox}}$$

P3, P4 = parameters used in defining empirical mobility law, see (4)

$$P5 = \frac{C_{ox}\ Z}{L}$$

P6 = parameter used in defining empirical channel length modulation, see (5)

$$P7 = V_{FB}\ .$$

In these expressions N_A is the p-dopant concentration, n_i the intrinsic carrier concentration, β^{-1} the thermal potential kT/q (k = Boltzmann's constant, T= temperature, q = electronic charge), κ_s the semi-conductor permittivity, C_{ox} the oxide capacitance per unit area, Z and L the width and length of the device respectively and V_{FB} the flatband voltage. The expression for μ_{eff} and L_{eff} in this model are chosen to be,

$$\mu_{eff} = \frac{P_3}{1 + P4(V_{GS}-P1-P7-P2\sqrt{P1})}\ , \tag{4}$$

$$L_{eff} = 1 - P6\left[(V_{DS}-V_{DSAT} + P1 - V_{BS})^{\frac{1}{2}} - (P1 - V_{BS})^{\frac{1}{2}}\right] \tag{5}$$

(b) Spice 2 Model

In this model the current in the subsaturation and saturation regions of drain voltage is still expressed in the form (1) and (2) with V_{DSAT} given by (3). However, different and much more complicated empirical expressions are assumed for L_{eff} and μ_{eff} in terms of the parameters. There are nine parameters in this model defined by

P1 = V_{FB}

P2 = N_A

P3,P4,P5 = parameters used in defining mobility, see (7) and (8) ,

P6 = length of channel

P7 = width of channel

P8 = parameter used in defining channel length modulation, see (9) ,

P9 = $(2\kappa_s q N_A)^{\frac{1}{2}}/C_{ox}$.

Thus only P1, P3 and P9 appear directly in the list of parameters used in the Ihantola model. The appropriate expressions used for μ_{eff} and L_{eff} are

$$\mu_{eff} = \begin{cases} P3 , & V_{GS} - V_{To} < V_{CRIT} \\ \left(\dfrac{P3 \times V_{CRIT}}{V_{GS} - V_{To}}\right)^{P5} , & V_{GS} - V_{To} > V_{CRIT} , \end{cases} \tag{7}$$

where

$$V_{CRIT} = \frac{P4}{P9}\left(\frac{\kappa_s}{2qP2}\right)^{\frac{1}{2}} , \tag{8}$$

$$V_{To} = P1 + \phi + P9\,\phi^{\frac{1}{2}}, \quad \phi = \frac{2}{\beta}\,\ell n\left(\frac{N_A}{n_i}\right) ,$$

and

$$L_{eff} = \begin{cases} 0.8\ x_D\ , & L^* < 0.8x_D \\ L^*\ , & L^* > 0.8x_D \end{cases} \qquad (9)$$

where $L^* = P6 - x_D \left\{ V_{DS} - V_{DSAT} + \left[1 + \frac{(V_{DS}-V_{DSAT})^2}{16} \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}}$

$$- x_D \left[\left(\frac{x_D P8}{2\mu_{eff}} \right)^2 + V_{DS} - V_{DSAT} \right]^{\frac{1}{2}} - \frac{x_D^2\ P8}{2\mu_{eff}}\ ,$$

$$x_D = \frac{2\kappa_s}{qP2}\ .$$

The Spice 2 model has the property that the expressions for the current and the channel conductance are continuous functions and it is in a suitable form for adaptation to the SPICE circuit simulation program.

(c) Brews Model

There are a number of charge sheet models for the MOSFET but perhaps one of the easiest models to implement is that developed by Brews (1978). The principal assumption of the charge sheet model is that the current travels in a surface of zero depth at the interface between the semiconductor and the gate insulator. This means the inversion layer is assumed to have zero thickness.

In Brews (1978) the expression

$$I_D = qZ\mu_{eff}\, N(y)\frac{d\phi_f}{dy} \tag{10}$$

is derived for the drain current, where N(y) is the carrier density per unit area, $d\phi_f/dy$ is the average quasi-fermi level gradient and the corrdinate y measures distance along the channel from source to drain. Brews next approximates $d\phi_f/dy$ by

$$\frac{d\phi_f}{dy} = \frac{d\phi_s}{dy} - \frac{1}{\beta}\frac{d}{dy}(\ell nN) \tag{11}$$

where $\phi_s(y)$ is the potential along the oxide-silicon interface and it is the second term on the right hand side which is assumed to take into account diffusion current effects. Integration of (11) yields

$$N(y) = N(o)\exp\{\beta\left[\phi_s(y) - \phi_s(o)\right] - \beta\left[\phi_f(y) - \phi_f(o)\right]\}. \tag{12}$$

On estimating $d\phi_f/dy$ between (10) and (12) we obtain the result

$$N(y) = N(o)\exp\{\beta\left[\phi_s(y) - \phi_s(o)\right]\} - \frac{I_D e^{\beta\phi_s(y)}}{kT\mu_{eff}Z}\int_o^y e^{-\beta\phi_s(yo)}dyo. \tag{13}$$

This gives the carrier density in terms of ϕ_s and to determine the electrostatic potential ϕ we have

$$\nabla^2\phi = \begin{cases} 0 & \text{in gate oxide} \\ -\frac{q}{\kappa_s}(p-N_A) & \text{in silicon} , \end{cases} \tag{14}$$

where $p = n_i e^{-\beta\phi + \beta\phi_f}$ is the hole density in the semiconductor. The charge sheet model then assumes a boundary condition at $x = 0$ in the form

$$\kappa_{ox} \frac{d\phi}{dx}\Big]_{x=0_-} - \kappa_s \frac{d\phi}{dx}\Big]_{x=0_+} = -qN(y) , \tag{15}$$

where $x = 0$ is the interface between oxide and silicon, κ_{ox} is the permittivity of the oxide and x is the coordinate measured positive into the silicon.

Now in the long channel approximation the solution of the Poisson equation (14) is simplified by assuming $\nabla^2\phi \approx d^2\phi/dx^2$. Hence, in the silicon, integration of (14) once yields

$$\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 = \frac{q}{\kappa_s}\left\{ -\frac{n_i}{\beta} \exp\left(-\beta\phi + \beta\phi_f \right) - N_A\phi \right\} + \text{constant} .$$

If the constant is chosen to satisfy $\phi \to 0$, $d\phi/dx \to 0$ as $x \to \infty$ and exponentially small terms are neglected, we obtain for the boundary value needed in (15),

$$\kappa_s \frac{d\phi}{dx}\Big]_{x=0_+} = -qN_A L_B \left[2(\beta\phi_s - 1)\right]^{\frac{1}{2}} ,$$

where the Debye length $L_B = (\kappa_s/\beta q N_A)^{\frac{1}{2}}$. More simply in the oxide we find

$$\kappa_{ox} \left.\frac{d\phi}{dx}\right]_{x=o_-} = - C_{ox} (V_{GS} - \phi_s) .$$

Hence, the boundary condition (15) becomes,

$$C_{ox}[V_{GS}-\phi_s(y)] = qN_A L_B \{2[\beta\phi_s(y)-1]\}^{\frac{1}{2}} + qN(y) . \qquad (16)$$

Thus we have derived two equations (13) and (16) for $\phi_s(y)$ and $N(y)$. Brews (1978) suggests a method of eliminating $N(y)$ to obtain a relation between I and ϕ_s. First equations(13) and (16) are differentiated and dN/dy eliminated between them. Then $N(y)$ is eliminated from this result by using (16) again. In this way we obtain the expression

$$I_D = \frac{\mu_{eff}}{\beta} P5 \{ (1 + \beta V_{GS}) (\varphi_{sL} - \varphi_{so}) - \frac{\beta}{2} (\phi_{sL}^2 - \phi_{so}^2)$$

$$- \frac{2P2}{3\beta^{\frac{1}{2}}} \left[(\beta\phi_{sL}-1)^{3/2} - (\beta\phi_{so}-1)^{3/2} \right]$$

$$+ \frac{P2}{\beta^{\frac{1}{2}}} \left[(\beta\phi_{sL}-1)^{1/2} - (\beta\phi_{so}-1)^{1/2} \right] \} ,$$

where the parameters Pi are defined below and $\phi_{so} = \phi_s(y=o)$ and $\phi_{sL} = \phi_s(y=L)$, the source and drain values of the potential respectively.

This is the expression to be used in the Brews model but in the optimization process the current is required for given specified values of the drain and gate voltages. This means that in any model evaluation we must first compute the appropriate values of ϕ_{sL}. We also note that the parameters in the model are expressed in exactly the same form as in the Ihantola model if the same empirical expressions are chosen for μ_{eff} (and L_{eff} if required).

Calculation of ϕ_{so} . The condition applied by Brews here is that ϕ_{so} should be obtained from the one-dimensional Poisson equation when $V_{DS} = 0$. He obtains

$$V_{GS} - \phi_{so} = \frac{P2}{\beta^{\frac{1}{2}}} \left[\beta\phi_{so} - 1 + \exp(\beta\phi_{so} - \beta P1) \right]^{\frac{1}{2}} ,$$

if an exponentially small term is neglected. This is implicit in ϕ_{so} for given V_{GS} but we may write the equation in the alternative form

$$\beta\phi_{so} = \beta P1 - \ell n\,(\beta P2^2) + \ell n\{[\beta(V_{GS}-\phi_{so})]^2 / [1 + (\beta\phi_{so}-1) / \exp(\beta\phi_{so} - \beta P1)]\} .$$

Then Brews suggest the following iteration scheme,

$$\phi_{so}^{(o)} = P1 - \ell n(\beta P2^2)/\beta$$

$$\phi_{so}^{(i+1)} = \phi_{so}^{(o)} + \frac{1}{\beta} \ell n \{[\beta(V_{GS}-\phi_{so}^{(i)})]^2 / [1 + (\beta\phi_{so}^{(i)}-1) \exp(\beta\phi_{so}^{(i)} - \beta P1)]\} ,$$

$$(i \geq 0) .$$

It is this scheme which we use for determining $\phi_{so} = \phi_{so}(V_{GS})$.

Calculation of ϕ_{sL}. After some discussion Brews uses the condition

$$\beta\phi_{sL} = \beta\phi_{so} + \beta V_{DS} + \ell n \left[N(L)/N(0) \right] ,$$

to obtain ϕ_{sL} , where N(L) AND N(0) are obtained from (16). Hence ϕ_{sL} is implicity defined for specified V_{DS} by the equation

$$\phi_{sL} = \phi_{so} + V_{DS} + \frac{1}{\beta} \ell n \left[\frac{\beta^{\frac{1}{2}}(V_{GS}-\phi_{sL}) - P2(\beta\phi_{sL}-1)^{\frac{1}{2}}}{\beta^{\frac{1}{2}}(V_{GS}-\phi_{so}) - P2(\beta\phi_{so}-1)^{\frac{1}{2}}} \right] , \qquad (17)$$

A straightforward iteration scheme for this equation is

$$\phi_{sL}^{(o)} = \phi_{so} + V_{DS} ,$$

$$\phi_{sL}^{(i+1)} = \phi_{sL}^{(o)} + \frac{1}{\beta}\, \ell n\left[\frac{\beta^{\frac{1}{2}}(V_{GS}-\phi_{sL}^{(i)}) - P2(\beta\phi_{sL}^{(i)}-1)^{\frac{1}{2}}}{\beta^{\frac{1}{2}}(V_{GS}-\phi_{so}) - P2(\beta\phi_{so}-1)^{\frac{1}{2}}}\right] , \quad i \geq 0$$

but this fails at large V_{DS} because there the numerator of the log term in (17) tends to zero (near saturation) and hence the perturbation to $\phi_{sL}^{(i)}$ becomes large. This prompted an investigation into the asymptotic form of ϕ_{sL} for large V_{DS} and equation (16) shows that its limiting value ϕ^* is obtained from

$$\beta^{\frac{1}{2}}(V_{GS} - \phi^*) - P2(\beta\phi^* - 1)^{\frac{1}{2}} = 0 . \tag{18}$$

More precisely from (17), for large V_{DS},

$$\beta^{\frac{1}{2}}(V_{GS} - \phi_{sL}) - P2(\beta\phi_{sL}-1)^{\frac{1}{2}} = Ae^{-\beta V_{DS}} + \ldots ,$$

for some A independent of V_{DS}. This is easily transformed into a quadratic equation for ϕ_{sL} having an appropriate solution in the form,

$$\phi_{sL} = \phi^* + Be^{-\beta V_{DS}} + \ldots ,$$

where B is known in terms of A. Substitution of this expression into (17), written more conveniently as

$$\exp\left[\beta(\phi_{sL}-V_{DS})\right] = D\left[\beta^{\frac{1}{2}}(V_{GS}-\phi_{sL}) - P2(\beta\phi_{sL}-1)^{\frac{1}{2}}\right] ,$$

where $D = \exp(\beta\phi_{so}) \,/\, \left[\beta^{\frac{1}{2}}(V_{GS}-\phi_{so}) - P2(\beta\phi_{so}-1)^{\frac{1}{2}}\right]$, gives the value of A. We obtain finally

$$\phi_{sL} = \phi^* + \frac{2(\phi^*-V_{GS})\left[\beta^{\frac{1}{2}}(V_{GS}-\phi_{so}) - P2(\beta\phi_{so}-1)^{\frac{1}{2}}\right]}{\beta^{\frac{1}{2}}(2V_{GS}-P2^2-2\phi^*)} \exp\left[\beta(\phi^*-\phi_{so}-V_{DS})\right] + \ldots , \tag{19}$$

where, from (18),

$$\phi^* = V_{GS} + \frac{P2^2}{2} \pm \frac{P2}{2}\left[P2^2 + 4(V_{GS} - \frac{1}{\beta})\right]^{\frac{1}{2}} ,$$

and the negative sign is required because, for example, $V_{GS} - \phi^*$ must be positive from (18).

In applying the model the criteron was selected whereby the asymptotic expression (19) should be used unless $\phi^* - \phi_{so} - V_{DS}$ is greater than some fixed tolerance (such as $-3/\beta$, $-4/\beta$, ... etc.) and this gave satisfactory results. We chose the same form for μ_{eff} as in the Ihantola model. However, the Brews model is valid in sub-threshold and saturation regions so it was decided not to include any empirical form of channel modulation. This has the effect of reducing the number of parameters to six and in the foregoing it has been assumed that $P6 = V_{FB}$. Through an oversight the parameter V_{FB} (which enters the current through a simple translation of V_{GS}) was omitted in the model except where it enters μ_{eff} . Further modification to remedy this fault and also to include substrate bias effects (V_{BS} is assumed to be zero in the above anlaysis) are proceeding but are not available for this report

Chapter 3 OPTIMIZATION METHODS

Let $\underline{P}$ be the parameter vector associated with the models Ihantola (I), Spice 2 (c) or Brews (b), so that each component of $\underline{P}$ is one of the parameters of (I), (S) or (B) respectively. We make the convention that the i-th component of $\underline{P}$ corresponds to the i-th parameter as it appears in either one of the three equations. Recall that for different values of the gate voltage, V_{GS}, we are provided with several experimentally obtained pairs of values of the source to drain current, I_D, versus the drain voltage V_{Dj},

$$\left(V_{DSi}, I_{Di}\right)_{V_{Gj}} \qquad \begin{matrix} i = 1,2,\ldots,m \\ j = 1,2,\ldots,n \end{matrix}$$

Typically $m = 20$, $n = 4$. Therefore to every $\underline{P}$ we can associate the scalar function

$$F(\underline{P}) = \sum_{i,j} \left[I_{Di} - I^{*}_{Di}(\underline{P})\right]^2 V_{Gj}$$

where $I^{*}_{Di}(\underline{P})$ is the model-predicted current at the drain voltage V_{DS} corresponding to the gate voltage V_{Gj}. Such a scalar function, F, can be constructed for each device for which V_{DS}, I_D values are provided and for each of the three mathematical models of the device response. Our goal is to estimate $\underline{P}$ so that $F(\underline{P})$ is minimized with $\underline{P}$ belonging to a set of physically acceptable vectors. This is recognized as a constrained non-linear least squares problem in the components of $\underline{P}$.

Various iterative methods exist to expedite this minimization or vector-optimization process. One, called Steepest Descent (McCormick (1983)), searches for a minimum in the direction of the negative gradient of $F(\underline{P})$

(a function of several variables decreases most rapidly along the direction of the negative gradient vector) and then adjusts for step length along this vector. This process is repeated for each successive iteration. Steepest Descent is quite stable - convergence is assured; however, convergence is too slow for practical use. A faster method is Newton's Method (McCormick (1983)) which relies on the Taylor expansion of the error function with respect to P. A modification of this, known as Gauss-Newton, converges rapidly but lacks the stability of Steepest Descent (i.e., convergence is not guaranteed). The algorithm known as Levenberg-Marquardt (Levenberg (1944), Marquardt (1963), McCormick (1983)) is an interpolation between Gauss-Newton and Steepest Descent in that search direction and step length are modified simultaneously For this reason, its stability and rapid convergence inherent from each of the two previously mentioned methods, Levenberg-Marquardt is recognized to be one of the most efficient algorithms available. For a more detailed numerical analysis of the iterative (algebraic) formulation and convergence considerations of these gradient methods see Morris and Everson (1984).

Also described in detail in this report are the two primary programs. SARAH and MOSES. They use IMSL subroutines invoking the above described Gauss-Newton and Levenberg-Marquardt methods to expedite the minimization procedure. SARAH is given a number of initial values or starting points from which to search over a prescribed hyper-rectangle, S, of the parameter space. The selection of the extreme points of the hyper-rectangle is guided by an a priori estimate of the range of the different parameters. SARAH sifts through the data employing a constrained Gauss-Newton method to locate what is hoped to be the global minimum. The multi-dimensional parameter space is seeded with minima; by sorting through a wide range of data SARAH

locates a number of these minima and then selects the deepest as the global minimum. This, then, is fed to MOSES which uses an unconstrained Levenberg-Marquardt algorithm to improve upon the convergence and accuracy of the vector $\underline{P}$ to minimize the error function $F(\underline{P})$. It should be emphasized that the convergence to such a vector (within a specified tolerance) subsequent to the MOSES program does not offer assurance that this indeed is the global minimum. In fact there is no clear cut way of determining this. Moreover the value of $\underline{P}$ provided by MOSES may no longer be a vector $\underline{P}$ belonging to the set, S, of acceptable vectors (see Section 5 for a more extensive discussion of this issue). However, since MOSES improves upon the constrained minimization provided by SARAH, the minimum obtained, according to the specified criteria of tolerance, usually lies within S or it is sufficiently close to it.

Program MOSES stops its search whenever one of the following tolerance criteria is satisfied:

(a) On two successive iterations the parameter estimates agree, component by component, to within a specified number of significant digits

(b) The norm of the gradient vector is within a specified tolerance

(c) On two successive iterations the error function $F(\underline{P})$ differs by some prescribed small amount ε,

i.e., $\left|F(\underline{P}_{K+1}) - F(\underline{P}_K)\right| < \varepsilon$.

The size of the Root Mean Square (RMS) error expressed in percentage terms is chosen as the main accuracy criterion for deciding whether or not

a given value of $\underline{P}$ is acceptable. Recall that the RMS error is a quantification of accuracy based upon an averaging of error distributed over all data points, say N. The formula for RMS error is given by:

$$\text{RMS ERR} = \left[\left(\sum_{j=1}^{N} \left| \frac{(I^{*}_{Dj}(\underline{P}) - I_{Dj})}{I_{Dj}} (100) \right|^{2}\right)/N\right]^{1/2} .$$

Less than ten percent RMS error from either SARAH or MOSES for a given device is regarded as a satisfactory result (RMS MOSES < RMS SARAH).

Chapter 4 RESULTS

This chapter is divided into three sections giving the main results for the three models tested, Ihantola, Spice 2 and Brews respectively. Comparisons between the models are presented and discussed in Chapter 5.

(a) IHANTOLA MODEL

The Ihantola model provides accurate results, at least when $V_G \neq 2$ and $V_{BS} = 0$. The accuracy remains surprisingly good even for small MOSFETs. Table 1 gives the relevant data concerning the Ihantola model.

As anticipated, the RMS error corresponding to devices of greater channel length and width is substantially less than that associated with smaller devices. That is to say, accuracy of the mathematical model's fit to the data (to currents produced experimentally) provided by extraction of the appropriate estimated parameter vector $\underline{P}$, generally increases as the channel's length and width increase. Table 2, extracted from Table 1, confirms this result.

Notice that the best fit is usually provided when length/width = 1.

TABLE 1

RMS ERROR/PARAMETER VALUES FOR IHANTOLA MODELS

DATA SET	DIM μMs		P A R A M E T E R S P1	P2	P3	P4	P5	P6	P7	RMS ERR%
101	1.2 x 2.5	(S)	.40	1.3	1.1	.50	.014	.051	-.40	5.7
		(M)	1.2	3.4	1.1	.11	.0065	.083	-5.0	3.9
102	2.5 x 2.5	(S)	.40	1.1	.61	.17	.0080	.044	-.40	4.6
		(M)	.95	1.7	.61	.083	.0060	.054	-2.1	4.5
103	1.2 x 5.0	(S)	.40	1.2	1.2	.50	.036	.049	-.40	7.1
		(M)	1.5	3.7	1.2	.11	.017	.087	-6.0	4.9
104	2.5 x 5.0	(S)	.40	1.1	.75	.22	.020	.044	-.40	4.7
		(M)	2.0	2.0	.74	.093	.014	.066	-4.4	4.6
N12	13.5 x 13.5	(S)	.45	.81	.93	.036	.0049	.021	.39	0.25
		(M)	.59	.84	.93	.036	.0049	.023	.63	0.22
N13	13.5 x 4.5	(S)	.40	.87	.67	.033	.0018	.022	-.33	0.30
		(M)	.36	.86	.67	.033	.0018	.021	-.26	0.31
AN11	3.0 x 3.0	(S)	.40	1.3	1.1	.18	.0062	.065	-.40	3.2
		(M)	1.5	2.3	1.1	.066	.0043	.090	-3.7	1.9
AN12	24.0 x 24.0	(S)	.41	.88	.65	.040	.0063	.014	-.21	0.24
		(M)	.40	.88	.65	.040	.0063	.014	-.19	0.24
AN13	3.0 x 24.0	(S)	.40	1.2	.64	.22	.087	.071	-.40	3.2
		(M)	1.3	2.1	.62	.086	.061	.094	-3.2	1.9
AN21	2.5 x 2.5	(S)	.40	1.1	.60	.17	.0083	.045	-.40	2.8
		(M)	1.5	1.9	.60	.078	.0061	.062	-3.2	1.7
AN22	2.5 x 5.0	(S)	.40	1.1	.78	.22	.020	.045	-.40	3.2
		(M)	2.1	2.1	.78	.092	.014	.069	-4.6	1.8
AN23	2.5 x 10.0	(S)	.40	.93	.60	.19	.046	.041	-.40	2.1
		(M)	1.0	1.3	.60	.13	.039	.052	-1.9	1.5

All data for VBS = 0 , VG $\neq$ 2

For SARAH: NSIG = 2 , NSRCH = 100 ,

For MOSES: NSIG = 3 .

TABLE 2

DATA SET	CH L/CH W	RMS SARAH/RMS MOSES
101	1.2/2.5	5.7/3.9
102	2.5/2.5	4.6/4.5
AN21	2.5/2.5	2.8/1.7
AN23	2.5/10	2.1/1.5
N13	13.5/4.5	0.30/0.31
N12	13.5/13.5	0.25/0.22
AN12	24/24	0.24/0.24

These above results neglect VG = 2 , which, if included, produces unacceptable errors (90% RMS or greater). Both non-diffusive models, Ihantola and Spice 2, neglect consideration of this gate voltage in the scheme of extracting parameters (see Chapter 5 for further discussion).

Another factor influencing accuracy involves the number of initial values or starting guesses. The SARAH program prompts for specification of the number of initial values - then distributes these (internal to the program) as starting points with which to commence the search for minima over the hyper-rectangle of seven dimensional parameter space. It would seem intuitively clear that a larger field of initial starting points would produce greater accuracy in obtaining an optimal parameter estimate; certainly an increase in the number of starting guesses improves the probability of locating the global minimum. And whether or not the global minimum is in fact the minimum SARAH ultimately calculates is not

particularly relevant. SARAH's process of sifting through the seven dimensional parameter space in search of a number of minima - then selecting the deepest as the global minimum - would indicate that a larger assortment of starting values provides greater flexibility in the choice of this deepest minimum which is eventually handed over to the MOSES program. SARAH, however, consumes a great deal of time in the process of computing the optimal parameter estimate. Our initial selection of just one starting value necessitated allowing for large blocks of turnaround computing time. To expedite this process with greater efficiency data sets for devices of varying channel length and width were run continuously in batch mode. This avoided having to input each set individually and wait for long periods of time before running the next set. In this fashion we were able to obtain the more accurate results from SARAH using up to 100 initial values.

A sensitivity analysis was performed with respect to each of the seven parameters of the Ihantola model. If we view our parameter extraction process as essentially that of best-fitting response curves to experimental data then in this light it is prudent to consider the following situation. Suppose we have obtained our parameter $\underline{P}$ enabling minimization of the error function $F(\underline{P})$ and therefore a best-fit of the response curve to the data. In what way does a small perburbation of just one of the parameters (P_i, the ith component of the parameter vector $\underline{P}$, $1 \leq i \leq 7$) effect changes in the accuracy of the response curve's fit to the data?

A graphic analysis provides visual insight to the fashion with which a ten percent increment and decrement alter the fit of these response curves. It is found that a $\pm 10\%$ perturbation of the first parameter,

the doping parameter, indicates a rather significant change of the I_D-V_{DS} curve (from the curve with optimal parameter estimates) for gate voltages of 3, 4 and 5. In this respect we regard P1 as quite sensitive. In contrast, we find similarly that the same perturbation of the sixth parameter, P6 , results in an insignificant change of the response curves indicative of great insensitivity.

In Figures 3-6 graphs are drawn which provide an illustration of the model's sensitivity (for two particular data sets at zero substrate bias) to small perturbations of the seven parameter components. For the data set corresponding to the long-channel device (24 x 24μMs), P3 and P5 are found to be the most sensitive; a 10% change (increase and decrease) in the value of these parameters produces close to 10% RMS error of the model to experimental data (RMS error for the optimal parameter set was 0.24%. The effect of such a change in P3 is illustrated in Figure 3. The same perturbation of P6 , illustrated in Figure 4, seen as the least sensitive of the parameters, produced 1.25% RMS error. Similarly, for data set AN21 (as an example of a shorter channel device, 2.5 x 2.5 Ms, a 10% increase and decrease of P7 results in a 15.5% and 17.3% RMS error - a significant deviation from the model's best-fit of 1.7% RMS error.

Again, P6 is seen to be the least sensitive of the seven parameters. A 10% perturbation of this parameter produces a 1.8% RMS error of the model to experimental data - less than a 0.01% change from the model's best-fit. The behavior with respect to these two parameters is depicted in Figures 5 and 6.

Eigenvalues of the Hessian matrix provide further information with

regard to the sensitivity analysis. However, MOSES failed to output eigenvalues consistently for reasons presently not understood. This question of sensitivity to (or redundancy of) a given component of $\underline{P}$ deserves further investigation. Future analysis should aim to provide a complete picture of this important issue (see Chapter 5 for additional comments).

Table 2 also compares parameter components, P1 - P7 , (with listed RMS percentage error) from the optimization of the twelve data sets tested. Notice that the parameter values for data sets 102 and AN21, two different experimental sets of data provided for devices of the same dimension (2.5 x 2.5 μMs), closely coincide. Similarly, parameter values for data sets 104 and AN22 (both 2.5 x 5.0 μMs) are roughly equivalent (although the MOSES estimate of the vector $\underline{P}$ for data sets 102 and 104 has more than twice the RMS error than the corresponding parameter estimates for AN21 and AN22 respectively). Examination of the parameter component pertaining to the capacitance, P5 , reveals increment of the device's width (i.e. equal length, varying width) to be proportional to (similarly increasing) values of this particular parameter. For example, data sets 101 and 103 of dimension 1.2 x 2.5 μMs and 1.2 x 5.0 μMs respectively correspond to P5 values of .0065 and .017. And data sets 102 and 104 (2.5 x 2.5 μMs and 2.5 x 5.0 μMs) have P5 values of .0060 and .014. Similarly for AN21 - AN23 (2.5 x 2.5 μMs, 2.5 x 5.0 μMs, and 2.5 x 10.0 μMs) values produced are .0061, 014 and .039. And for N13 and N12 (13.4 x 4.5 μMs and 13.5 x 13.5μMs) P5 is .0018 and .0049. A three-fold increase in the device's width produces a similar increase in the parameter value

(consistent with what we might expect on physical grounds).

Finally, we note that the greatest accuracy of parameter estimation (i.e. the least RMS error) occurs when $V_{BS} = 0$. Increasingly negative V_{BS} values yields greater error and diminished accuracy. Table 3 itemizes Ihantola RMS error for various V_{BS} :

DATA SET	$V_{BS} = 0(M)$	$V_{BS} = 2(S)$	$V_{BS} = -2.5(S)$	$V_{BS} = -5(S)$
101	3.9			
102	4.5			
103	4.9			
104	4.6			
AN11	1.9	4.2		13
AN12	.24	.82		11
AN13	1.9	3.9		7.5
AN21	1.7	3.5		7.6
AN22	1.8	4.1		6.5
AN23	1.5	2.3		4.0
N12	.22	.68		3.3
N13	.31	.92		5.4
121			9.4	
122			5.2	
123			11.0	
124			5.6	

TABLE 3

RMS percentage errors for the Ihantola model for each data set and different values for V_{BS}, $(V_G \neq 2)$.

TABLE 4
RMS ERROR/PARAMETER VALUES FOR SPICE 2 MODEL
All data for $V_{BS} = 0$, $V_G \neq 2$
For SARAH : NSIG = 2, NSRCH = 100 ,
For MOSES : NSIG = 3

DATA	DIM µMs Length x Width		PARAMETERS P1	P2	P3	P4	P5	P6	P7	P8	P9	RMS ERRORS %
101	1.2x2.5	(S)	-1.9	3.5	1.2	0.46	0.50	0.97	3.0	8.6	2.0	5.2
		(M)	-2.1	2.8	1.2	0.40	0.52	0.80	3.0	14	2.2	4.8
102	2.5x2.5	(S)	-1.9	4.9	0.81	1.0	0.47	2.1	2.7	8.6	1.7	4.8
		(M)	-1.9	4.1	0.80	0.95	0.55	1.8	2.7	56	1.7	4.8
103	1.2x5	(S)	-1.4	24	0.83	1.5	0.5	1.4	4.3	1.5	1.4	8.4
		(M)	-1.3	26	0.91	1.4	0.73	1.3	4.6	1.5	1.5	6.9
104	2.5x5	(S)	-1.6	7.6	1.1	0.94	0.25	2.1	4.8	1.5	1.5	4.9
		(M)	-1.6	7.5	1.1	1.0	0.29	2.1	4.8	2.3	1.5	4.9
N12	13.5x3.5	(S)	-0.78	0.27	0.63	0.78	0.13	16	11	1.5	0.8	0.67
		(M)	-0.78	0.27	0.63	0.78	0.13	16	11	1.5	0.8	0.67
N13	13.5x4.5	(S)	-0.93	3.5	1.2	1.1	0.064	15	3.8	1.5	1.0	0.78
		(M)	-0.93	3.5	1.2	1.1	0.064	15	3.8	1.5	1.0	0.78
AN11	3 x 3	(S)	-1.5	29	1.0	0.37	0.24	3.4	3.1	1.5	1.5	3.3
		(M)	-1.5	29	1.0	0.37	0.24	3.4	3.1	1.5	1.5	3.3
AN12	24x24	(S)	-1.1	18	0.67	3.3	0.11	28	20	1.5	1.1	1.1
		(M)	-1.1	18	0.67	3.3	0.11	28	20	1.5	1.1	1.1
AN13	3 x 24	(S)	-1.5	1.5	1.2	8.6	0.43	2.4	29	1.5	1.4	4.6
		(M)	-1.6	1.5	1.3	8.6	0.43	2.4	29	1.5	1.5	4.3
AN21	2.5x2.5	(S)	-1.9	34	0.45	1.7	0.16	2.2	2.1	1.5	1.6	2.4
		(M)	-1.9	34	0.45	1.7	0.16	2.2	2.1	1.5	1.6	2.4
AN22	2.5x5	(S)	-2.0	28	0.51	2.0	0.26	2.3	5.9	1.5	1.6	1.9
		(M)	-2.0	28	0.51	2.0	0.26	2.3	5.9	1.5	1.6	1.9
AN23	2.5x10	(S)	-2.0	3.2	1.1	0.83	0.20	2.0	12	8.6	1.6	1.1
		(M)	-1.9	2.0	1.1	0.72	0.33	1.6	12	77	1.5	0.92

In common with the Ihantola model it was found that large errors arise from incorporating the $V_G = 2$ data. Hence the RMS errors quoted in this table all refer to cases with $V_G = 2$ excluded and for $V_{BS} = 0$. The general pattern of results follows that for the Ihantola model described in the previous section. The result for long MOSFETs generally show small RMS errors and these increase to a maximum for the shorter MOSFETs. A similar trend with regard to the width of device can be seen from data sets AN21, AN22, AN23 where for a length (2.5 μMs) the RMS error decreases as the width increases (see Table 5).

TABLE 5

Variation of RMS percentage error with width for fixed length $L = 2.5$ μMs , $V_{BS} = 0$

DATA	W(μMS)	RMS (%)
AN21	2.5	2.4
AN22	5.0	1.9
AN23	10.0	0.92

The Spice 2 model contains nine parameters as described in Chapter 2. Included in Table 4 are the values of these parameters, extracted to two significant figures, by programs SARAH and MOSES from the different data sets. There are two pairs of data sets relating to different devices of the same dimensions, namely AN22 and 104 (2.5 x 5.0 μMs) and AN21 and 102 (2.5 x 2.5 μMs), but the results show that different RMS errors are obtained for each member of a pair. Moreover, there can be wide variations in the optimal values of the parameters. In this respect P2 and P8 are particularly noteworthy examples. Now P2 is the doping parameter

defined by (6) and a wide variation over different devices is understandable. However, P8 is a parameter (V_{max}) representing the maximum drift velocity of the carriers and contributes to the channel length shortening (Vladimirescu and Liu(1980)). The fact that it takes such widely differing values for similar devices would appear to point to a shortcoming of the model.

The parameters P6 and P7 , representing the length and width of the channel respectively, were, on physical grounds, constrained to lie within ±20% of the actual device dimensions during operation of the program SARAH. As shown in Table 4 the unconstrained optimal values predicted by MOSES are generally close to the SARAH values.

An important parameter is P1 , the flatband voltage V_{FB} , and in all cases this turns out to have a negative value (for $V_{BS} = 0$) of magnetude of O(1) . In some cases run with $V_{BS} = -5$ a positive V_{FB} results.

The effect of changing V_{BS} from zero to different negative values is summarized in Table 6 (analogous to Table 3 for the Ihantola model), where RMS percentage errors are quoted from running the SARAH and MOSES programs with each of the data sets. The same general trend is observed as before, that the fit of the model to the data becomes less accurate as $|V_{BS}|$ is increased.

The sensitivity analysis with respect to the parameters, described in part (a) for the Ihantola model, was also carried out for the Spice 2 model for the same two data sets for a 'large' MOSFET (24 x 24 μMs) and a "small' MOSFET (2.5 x 2.5 μMs). The results are consistent in the sense

that the most sensitive and the least sensitive parameters are the same for each size of MOSFET. A sensitive parameter is P9 (coresponding to the parameter P2 of the Ihantola model) and this is illustrated for both data sets in Figures 7 and 8. By contrast the model is insensitive to the parameter P4, occurring in the mobility law, and the associated graphs are drawn in Figures 9 and 10.

TABLE 6

RMS percentage errors for the Spice 2 model for each data set and different values of V_{BS} ($V_G \neq 2$) .

DATA SET	V_{BS} = 0	V_{BS} = -2	V_{BS} = -2.5	V_{BS} = -5
101	4.8			
102	4.8			
103	6.9			
104	4.9			
AN11	3.3	4.3		16
AN12	1.1	1.6		17
AN13	4.3	4.0		11
AN21	2.4	4.5		11
AN22	1.9	3.8		6.6
AN23	0.92	7.1		6.2
N12	0.67	1.3		4.0
N13	0.78	1.5		5.8(S)
121			7.0	
122			5.1	
123			9.4(S)	
124			6.1	

(C) BREWS MODEL

The principal objective in setting up the Brews model is to create a model that includes the effects of diffusion current so that the model can predict drain current for low values of gate voltage. The data with which we worked had four gate voltages, viz. 2, 3, 4 and 5. With the Ihantola

and Spice 2 models very poor fit to the V_{GS} = 2 data prompted us to exclude these data sets when trying to fit the model to the data. When this is done a fairly low RMS error results. The Brews model is much better than Ihantola or Spice 2 at finding a fit to data that includes V_{GS} = 2 but in general does not result in lower RMS error values than the Ihantola and Spice 2 when V_{GS} = 2 is excluded.

As might be expected the Brews model is noticeably longer in terms of CPU time spent in calculating the optimal parameter set. This is due primarily to the iteration schemes used to calculate the voltages at the end points of the charge sheet. Accurate figures are not possible as no software exists at the time of this work to measure actual CPU time.

A summary of the results obtained from the Brews model is given in Table 7 in analogy with Table 1 and Table 4 for the other models. From Tabel 7 one feature stands out rather strongly, namely that frequently MOSES gives little or no improvement in the SARAH results. In fact in two cases (101, AN13) the RMS error actually increased. This situation occurred also from time to time in the development phase of the other models where it was explained in terms of an inconsistent use of the Library routine parameter NSIG (the number of significant figures to which the model parameters must agree on successive iterations in order to stop the program) but in this case NSIG=2 for SARAH and NSIG=3 for MOSES. Hence it is expected that the RMS error in MOSES should generally be less than that in SARAH. Further numerical experimentation to investigate this phenomenon is indicated.

TABLE 7

RMS ERROR/PARAMETER VALUES FOR BREWS MODEL

DATA SET	DIMμMs		P1	P2	P3	P4	P5	P6	RMS ERR%
101	1.2 x 2.5	(S)	.12	1.2	.72	.49	.022	.49	7.8
		(M)	.072	1.3	.72	.48	.022	.47	8.5
102	2.5 x 2.5	(S)	.35	.78	.69	.40	.012	-.15	6.1
		(M)	.17	.91	.70	.21	.0084	-.39	5.4
103	1.2 x 5.0	(S)	.11	1.2	1.1	.48	.035	1.1	8.9
		(M)	.044	1.2	1.1	.43	.034	.98	8.9
104	2.5 x 5.0	(S)	.15	.89	.80	.25	.022	-.26	5.5
		(M)	.15	.89	.80	.25	.022	-.26	5.5
N12	13.5 x 13.5	(S)	.17	.77	.60	.062	.0086	1.3	2.5
		(M)	.17	.77	.60	.062	.0086	1.3	2.5
N13	13.5 x 4.5	(S)	.11	1.2	.61	.47	.0071	-.40	2.4
		(M)	.22	.70	.60	.060	.0025	-.96	0.94
AN11	3.0 x 3.0	(S)	.15	1.1	.79	.18	.0092	-.34	5.6
		(M)	.15	1.1	.79	.18	.0093	-.36	5.4
AN12	24.0 x 24.0	(S)	.28	.76	.82	.060	.0057	-.25	0.56
		(M)	.28	.76	.82	.060	.0057	-.25	0.56
AN13	3.0 x 24.0	(S)	.21	.90	1.2	.19	.037	1.3	5.3
		(M)	.17	.93	1.2	.17	.036	1.3	5.5
AN21	2.5 x 2.5	(S)	.34	.78	1.3	.48	.0073	-.34	5.7
		(M)	.18	.92	1.3	.25	.0051	-.73	4.6
AN22	2.5 x 5.0	(S)	.15	.90	1.1	.26	.017	-.40	4.9
		(M)	.15	.90	1.1	.27	.017	-.40	4.9
AN23	2.5 x 10.0	(S)	.21	.69	1.1	.22	.025	.74	3.6
		(M)	.21	.69	1.1	.22	.025	.74	3.6

All data for V_{BS} = 0 , VG = 2 included

FOR SARAH: NSIG = 2, NSRCH = 100

FOR MOSES: NSIG = 3

We note from Table 7 that data sets for MOSFETs of the same size have approximately the same RMS error. For example, the data sets 102 and AN21 (2.5 x 2.5 μMs) have RMS errors of 5.4% and 4.6% respectively. Similar results hold for the 104 and AN22 data sets (2.5 x 5 μMs). Further, note that as the channel width increases for fixed length (data sets AN21, AN22, AN23) progressively decreasing error are obtained. This is to be expected as the increasing width would make the one-dimensional model assumption used in the Brews model more valid.

A sensitivity analysis was performed using the AN12 and AN21 data sets. Table 8 gives the RMS percentage errors which result from a ±10% change in each of the optimal parameters in turn.

TABLE 8

RMS percentage errors for parameter variation of ±10% from the optimum

	AN12 (RMS=0.56%)		AN21 (RMS=4.6%)	
MULTIPLIER	0.9	1.1	0.9	1.1
P1	4.06	3.56	6.05	4.29
P2	6.76	6.04	9.51	6.62
P3	9.92	10.13	9.42	12.76
P4	1.77	1.62	7.84	5.39
P5	9.92	10.13	9.42	12.76
P6	0.60	0.55	5.23	4.38

In Table 8 it can be seen that the most sensitive parameters are 3 and 5 but it is informative to note that the changed RMS errors are identical for P3 and P5. The parameter showing the least amount of change is P6 and again we note that increasing the value of P6 results in an improved RMS error in both the AN12 and AN21 data sets. This is the only place

where an improvement of fit results from a perturbed parameter.

Again as with the other models the sensitivity results are depicted graphically. For each data set quoted in Table 8 the fit for ±10% variations in the most sensitive (P5) and least sensitive (P6) parameters is displayed in Figures 11-14.

As mentioned in Chapter 2 (c) the Brews model used here has not been adapted to include $V_{BS} \neq 0$. Further a correction to include flat band effects more fully should be incorporated into the model.

Chapter 5 CONCLUSION AND DISCUSSION

Three topics are presented in this last chapter.

(a) An overview of the three models used by the team to extract the relevant parameters of a MOSFET;

(b) Relevant features of the sensitivity problem and trends of all models with respect to it;

(c) Suggested lines of future inquiry on the basis of our team knowledge of the behaviour of the devices, of the limitations of the proposed models and of the complexity of the required numerical investigations.

(a) General Overview

The Ihantola model (as implemented by the SARAH and MOSES programs) provides a more accurate estimation of the parameter vector $\underline{P}$ than Spice 2. Comparing long-channel devices, we note that corresponding to dimensions of 24 x 24 μMs, 13.5 x 13.5 μMs and 13.5 x 4.5 μMs the RMS Ihantola MOSES errors at $V_{BS} = 0$ are 0.24, 0.22 and 0.31 percent respectively. For Spice 2 the errors are 1.1, 0.67 and 0.78. For the Brews Charge-Sheet model we have 0.40, 2.5 and 0.94 percent with data at $V_{GS} = 2$ included in the parameter extraction. For the two non-diffusive models, Ihantola and Spice 2, we exclude evaluation at the lowest gate voltage since parameter estimates of these two models at this particular gate voltage increase the RMS error beyond acceptable accuracy (up to 80 and 90 percent). However, comparison exclusive of evaluation at $V_{GS} = 2$ for the non-diffusive models substantiate Ihantola as somewhat more accurate than Spice 2 also for short channel devices. For example for devices

of dimensions 1.2 x 2.5 μMs, 2.5 x 2.5 μMs and 2.5 x 5 μMs the respective Ihantola/Spice 2 RMS MOSES errors are 3.9/4.8, 1.7/2.4 and 4.6/4.9 percent.

The consistency of MOSES improving upon the RMS error of SARAH (as expected) in all three models is well established (with two unexplained exceptions for the Brews model, see Chapter 4(c)). However, in some instances MOSES fails to converge (to a particular $\underline{P}$) within the set maximum number of model evaluations (2000); instead, although improving upon SARAH's result, iterations exhibit a slow oscillatory type of behavior. When the routine is interrupted after the oscillatory behavior is observed, the resulting parameter estimate is well within acceptable accuracy and (as noted) improves upon the SARAH estimate.

Also, for reasons not understood, eigenvalues of the Hessian (significant in ascertaining the sensitivity of a model to a perturbation of one or more of its parameter's components, see (b) below) frequently fail to be output in MOSES with no apparent regularity or consistency in all of the models tested.

Intrinsic to the SARAH program are the variables NSIG and NSRCH, NSRCH prompts for a specification of initial values or starting guesses for SARAH to begin its search over the multi-dimensional parameter space for the global minimium. In all three models a larger value of NSRCH produced greater RMS accuracy. 100 was the typical value specified for NSRCH: a test with NSRCH equal to 300, however, did not provide greater accuracy. NSIG is a tolerance criterion that sets the number of significant figures of each component of $\underline{P}$ which are required to coincide in two successive iterations. In Ihantola, Spice 2 and Brews NSIG was set to be 2 for SARAH

and 3 for MOSES. The improvement of Sarah accuracy gained from setting NSIG=2 (from previous trials using NSIG=1) demanded in turn more CPU time. This was true again for all three models incorporated into SARAH and MOSES as subroutines.

Table 9 compares accuracy (RMS percentage error) of the Ihantola and Spice 2 models at different substrate biases (at V_{BS}=0, Brews is included). We can see that an increasingly negative substrate bias produces greater error and that the Ihantola model performs generally better than the Spice 2 model over all data sets at the V_{BS} values tested.

TABLE 9

Comparative RMS errors for Ihantola, Spice 2 and Brews models at different substrate biases

	V_{BS}=0			V_{BS}=-2		V_{BS}=-5	
DATA	(I)	(S)	(B)	(I)	(S)	(I)	(S)
101	3.9	4.8	8.7				
102	4.5	4.8	5.4				
103	4.9	6.9	8.9				
104	4.6	4.9	5.5				
N12	.22	.67	2.5	.68	1.3	3.3	4.0
N13	.31	.78	.94	.92	1.5	5.4	5.8
AN11	1.9	7.4	5.4	4.2	4.3	12	16
AN12	.24	1.1	.56	.82	1.6	11	17
AN13	1.9	5.4	5.5	3.9	4.0	7.7	11
AN21	1.7	2.4	4.6	3.5	4.5	7.6	11
AN22	1.8	1.9	4.9	4.1	3.8	6.5	6.6
AN23	1.5	.92	3.6	2.3	7.1	4.0	6.2

(I)=IHANTOLA (S)=SPICE 2 (B)=BREWS

(b) Sensitivity

An analysis of sensitivity is possible by looking at the eigenvalues of the Hessian, the matrix of second partial derivatives, of the sum of squares function at an optimal parameter set.

Let

$$S(\underline{P}) = \sum_{j=1}^{N} [\, I_j^*(\underline{P}) - I_j \,]^2$$

be the sum of squares function where $\underline{P} = (P1,\ldots,Pk)$ is the k-dimensional parameter vector. Then a Taylor expansion of $S(\underline{P})$ gives

$$S(\underline{P} + \delta\underline{P}) = S(\underline{P}) = S(\underline{P}) + \nabla S(\underline{P}).\delta\underline{P} + (\delta\underline{P})^T H(\underline{P})\delta\underline{P} + o(|\delta\underline{P}|^2)$$

where $H(\underline{P})$ is the k x k Hessian matrix, i.e. $H_{ij}(\underline{P}) = \partial^2 S/\partial P_i \partial P_j$ where $H_{ij}(\underline{P})$ is the (i,j) entry in the matrix. Now at a minimal point, $\underline{P}^*$, $\nabla S(\underline{P}^*) = 0$. (In our calculations $|\nabla S(\underline{P}^*)| \approx 10^{-11}$) which is much smaller than the typical largest eigenvalue of the Hessian which was approximately 10^{-1}). Using this result we find that

$$S(\underline{P}^* + \delta\underline{P}) - S(\underline{P}^*) = (\delta\underline{P})^T H\, \delta\underline{P} + o(|\delta\underline{P}|^2)$$

Dividing both sides by $|\delta\underline{P}|$, the norm of the perturbing term, we obtain a formula for the relative change

$$\frac{S(\underline{P}^* + \delta\underline{P}) - S(\underline{P}^*)}{|\delta\underline{P}|} = (\delta\underline{P})^T H\,(\underline{P}^*)\left[\frac{\delta\underline{P}}{|\delta\underline{P}|}\right] + o(|\delta\underline{P}|) \qquad (20)$$

This result is quite illuminating as the Hessian is first of all a symmetric real matrix so that its eigenvalues are all real and the corresponding eigenvectors are orthogonal. But since $\underline{P}^*$ is assumed to be a minimum of $S(\underline{P})$ it follows that all the eigenvalues of $H(\underline{P})$ must be non-negative. If we further assume $\underline{P}^*$ to be an isolated minimum then eigenvalues are strictly positive and $H(\underline{P})$ is a positive definite quadratic form. It then follows that the left hand side of (20) would take on isolated maxima when $\delta\underline{P}$ were an eigenvector and the maximum values would be $\lambda_i \, |(\delta\underline{P})_i|$ where $(\delta\underline{P})_i$ is an $i^{\underline{th}}$ eigenvector and λ_i is its associated eigenvalue. Thus from equation (20) for $|\delta \underline{P}| = 1$ the maxima would occur at the normalized eigenvectors with the maximal value as the eigenvalue for that eigenvector. But a more useful relationship is to see how (20) can be expressed when $\delta\underline{P}$ is just a change in a single parameter coordinate. If we have the eigenvectors and eigenvalues of $H(\underline{P}^*)$ this is a fairly straightforward computation. For let J denote the matrix whose $i^{\underline{th}}$ column is the $i^{\underline{th}}$ normalized eigenvector of $H(\underline{P}^*)$ and let Λ denote the diagonal matrix whose $i^{\underline{th}}$ diagonal element is the $i^{\underline{th}}$ eigenvalue. The $H(\underline{P}^*) = J \Lambda J^T$. Let $\delta \underline{P}^j(\theta) = (0, 0, \ldots, \theta, 0, \ldots, 0)$ where θ occupies the $j^{\underline{th}}$ position.

Then

$$Sn_j(\theta) \equiv \frac{S(\underline{P}^* + \delta\underline{P}^j(\theta)) - S(\underline{P}^*)}{|\theta|} = (\delta\underline{P}^j(\theta))^T \, J \Lambda J^T \, \underline{e}_j$$

where $e_j = \dfrac{\delta\underline{P}^j(\theta)}{|\theta|}$, $|\theta| \neq o$.

Multiplying this right hand side out in detail gives

$$Sn_j(\theta) = \theta \sum_{i=1}^{k} \lambda_i J_{ij}^2 \quad .$$

So it is seen that the term $\sum_{i=1}^{k} \lambda_i J_{ij}^2$ measures the relative change of the sum of squares function in the $j^{\underline{th}}$ coordinate direction.

We did experience some difficulty in obtaining the eigenvalues and eigenvectors of the Hessians in running MOSES with some data sets. The IMSL subroutine would run into a floating point overflow and the program would terminate. This occurred in all models. The precise reason for this was not found but it is suspected that the smallest eigenvalue may have been too small for the precision of the machine. Another possibility is that the condition number, which is the ratio of the largest to the smallest eigenvalue, is too large, resulting in numerical instability in the calculation of the eigenvectors and their associated eigenvalues.

(c) Suggestions for future work

As a result of the experience gained from the clinic's work on the parameter extraction process for the MOSFET device several directions in which further investigations could usefully be made become apparent. Some of these correspond to relatively simple variations in the conditions under which the programs are run: there was insufficient time for the team to include these in its work. Others involve more major excursions into different aspects of MOSFET modelling. We outline our views on these matters under separate headings below.

(i) Role of V_{BS}

It is known that the operation of a MOSFET depends critically upon the nature of the inversion layer: upon how many carriers are in it, upon their mobility, etc. Therefore for an accurate description of how the MOSFET works we need the inversion layer carrier density per unit area N_e .

The effect of body-to-source reverse bias, V_{BS} , on N_e at fixed gate-to-source bias is a reduction in qN_e . Therefore a larger gate bias will be needed to cause inversion.

It does not seem that this feature of the MOSFET is properly represented in the Ihantola and Spice 2 models and this indicates why the RMS error increases to unacceptable levels when they are used to fit data with body-to-source bias of -2 and -5 . (see Table 9).

It should be possible to include V_{BS} in these models in a more realistic way, so that the accuracy of the parameter extraction process will not be affected by necessary changes in V_{BS} .

(ii) Weighting and CLEAK

The programs SARAH and MOSES possess the facility for giving more weight to some data points relative to others in optimising the sum of squares. In all cases run by the clinic equal weight was attached to all data points. Similarly a parameter CLEAK is chosen for each operation of the program, and is defined by

$$d_j = \max_{j=1,..,N} (|I_j| , \text{CLEAK}) ,$$

where d_j appears in the sum of squares,

$$S^*(\underline{P}) = \sum_{j=1}^{N} w_j \left(\frac{I_j^*(\underline{P}) - I_j}{d_j}\right)^2 ,$$

w_j is the weighting factor and N is the number of data points. Thus by appropriately choosing CLEAK, S^* can be made to denote a relative sum of squares, an absolute sum of squares or a mixture of the two. The choice of CLEAK in the clinic's work was such that $d_j = 1$.

Now it was found that much improvement of fit is obtained in the Ihantola and Spice 2 models if the data for $V_{GS} = 2$ is excluded. Further, the fit tends to be worse for lower values of V_{DS} . It may be that these problems could be tackled by varying the w_i so that low V_{DS} and V_{GS} values carry greater weight than others. Similar advantages may be gained by altering the CLEAK parameter although we note, as mentioned in Morris and Everson (1984), that at very low currents it is more appropriate to use the absolute sum of squares to avoid very large errors. It is recommended that this area be investigated.

(iii) Role of constraints

The constraints on the parameters, imposed in the SARAH program, can be adjusted and the chosen limits can have an important effect on the running of the program. Over-relaxed constraints can produce estimated parameters from SARAH which make the results from the MOSES program less predictable. In extreme cases, the model evaluation may fail due, for example, to negative square roots occurring. Such modifications are associated with the choice of the parameter NSRCH in SARAH (number of starting points) and a systematic study of the effect of variation of constraints and NSRCH

would be of interest. In fact, NSRCH is the only control that one has in finding the global minimum and repetition of results for two values of this parameter (differing say by a factor two) was taken as satisfactory evidence that the global minimum had in fact been attained in the search area. This procedure was only carried out in a few cases, however.

The related question of choosing 'realistic' values of parameters is also important. For example, P6 and P7 in the Spice 2 model (length and width parameters) were constrained to be within ±20% of the device measurements and the resulting predictions led MOSES to an improved fit of the model to the data. In other cases the choice is much less clear cut and the parameter V_{FB} seemed particularly difficult to estimate.

(iv) Non-convergence

It was found on a number of data sets that Ihantola and Spice 2 would not converge using the MOSES program. What seemed to occur was that the algorithm was causing certain parameters to oscillate slowly around a closed orbit. It would be worthwhile to obtain a clear analysis of the weakness of the Levenberg-Marquardt algorithm in its application to this problem.

(v) Other models

The clinic concentrated on three one-dimensional models of the more straightforward type. No study was made of more elaborate models such as the Pao-Sah model or possibly two-dimensional models. These are computationally expensive models but are expected to be more accurate. An

analysis of these models might well pay dividends in a more complete understanding of the device behaviour as well as showing more explicitly the limitations of the one-dimensional models which were studied.

(vi) Inverse theory

There is a branch of mathematics that is concerned with the analysis of parameter estimation problems (as well as inverse eigenvalue problems, inverse scattering problems and many others). Our problem, of MOSFET parameter estimation, has its most general formulation in terms of this theory. It is possible that framing the problem within such a formulation could produce some practical results. From preliminary literature searches no publications were located on this topic in spite of the fact that numerous applications of this theory in other engineering disciplines have met with some success. An attempt to accomplish this analysis will be made by a member of the clinic over the summer of 1985 and results, if any, will be made available to those who might be interested.

REFERENCES

1) Brews, J. R., A Charge-sheet Model of the MOSFET. Solid State Electronics, Vol. 21, pgs. 345-355, (1978).

2) Ihantola, H.K.J. and J.L. Moll, Design Theory of a Surface Field-Effect Transistor. Solid State Electronics, Vol. 7, pgs 423-430, (1964).

3) Levenberg, K., A Method for the Solution of Certain Nonlinear Problems in Least Squares. Quarterly of Applied Mathematics, Vol. 2, pgs. 164-168, (1944).

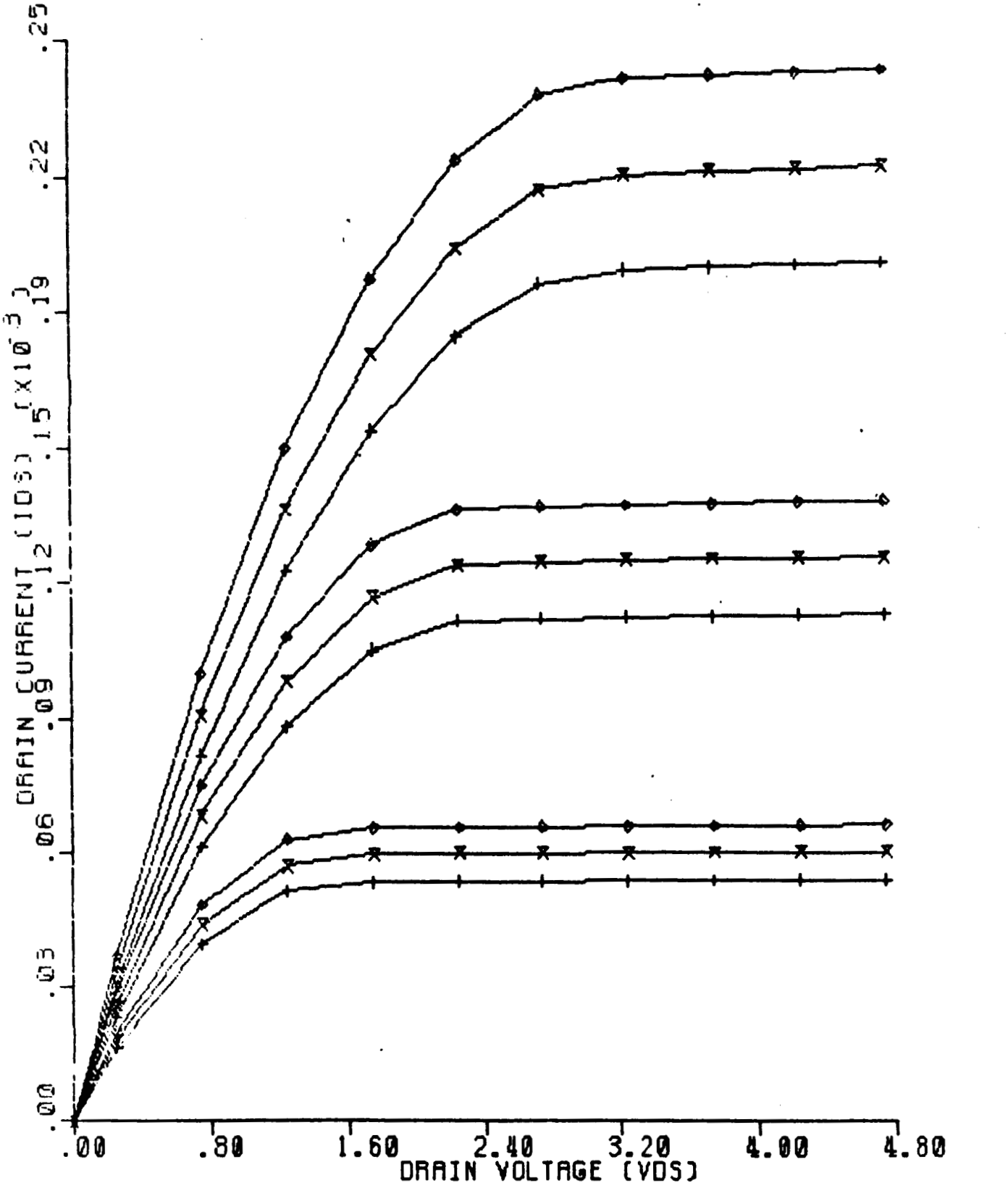
4) McCormick, Garth P., Nonlinear Programming: Theory, Algorithms and Applications. New York: John Wiley & Sons, (1983).

5) Marquardt, D.W. An Algorithm for Least Squares Estimation of Non-linear Parameters. SIAM Journal of Applied Mathematics, Vol. 11, pgs. 431-441, (1963).

6) Morris, H.C. & R. Everson, The Modelling of Short Channel Devices for use in VLSI. Report to Jet Propulsion Laboratory, Claremont Colleges, Mathematics Clinic, (1984).

7) Sze, S.M., Physics of Semiconductor Devices. 2nd Edition, New York: John Wiley & Sons, (1981)

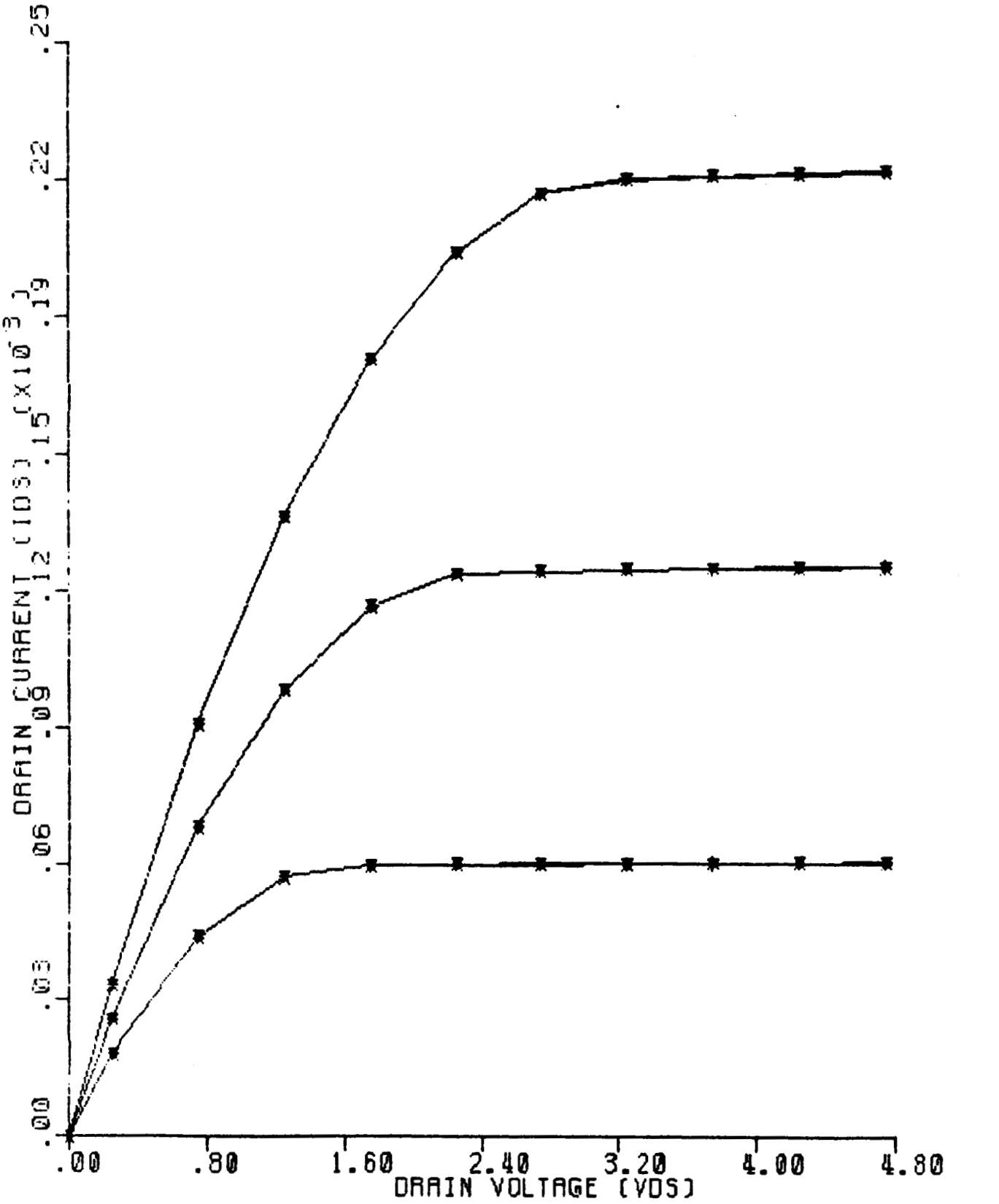
8) Vladimirescu, A. & S. Lin, The Simulation of MOS Integrated Circuits Using Spice 2. Memorandum No.UCB/ERL M80/7, Electronics Research Laboratory, University of California, Berkeley. (1980).



LENGTH = WIDTH = 24.0MU PARM(3)*(1.1 & 0.9)

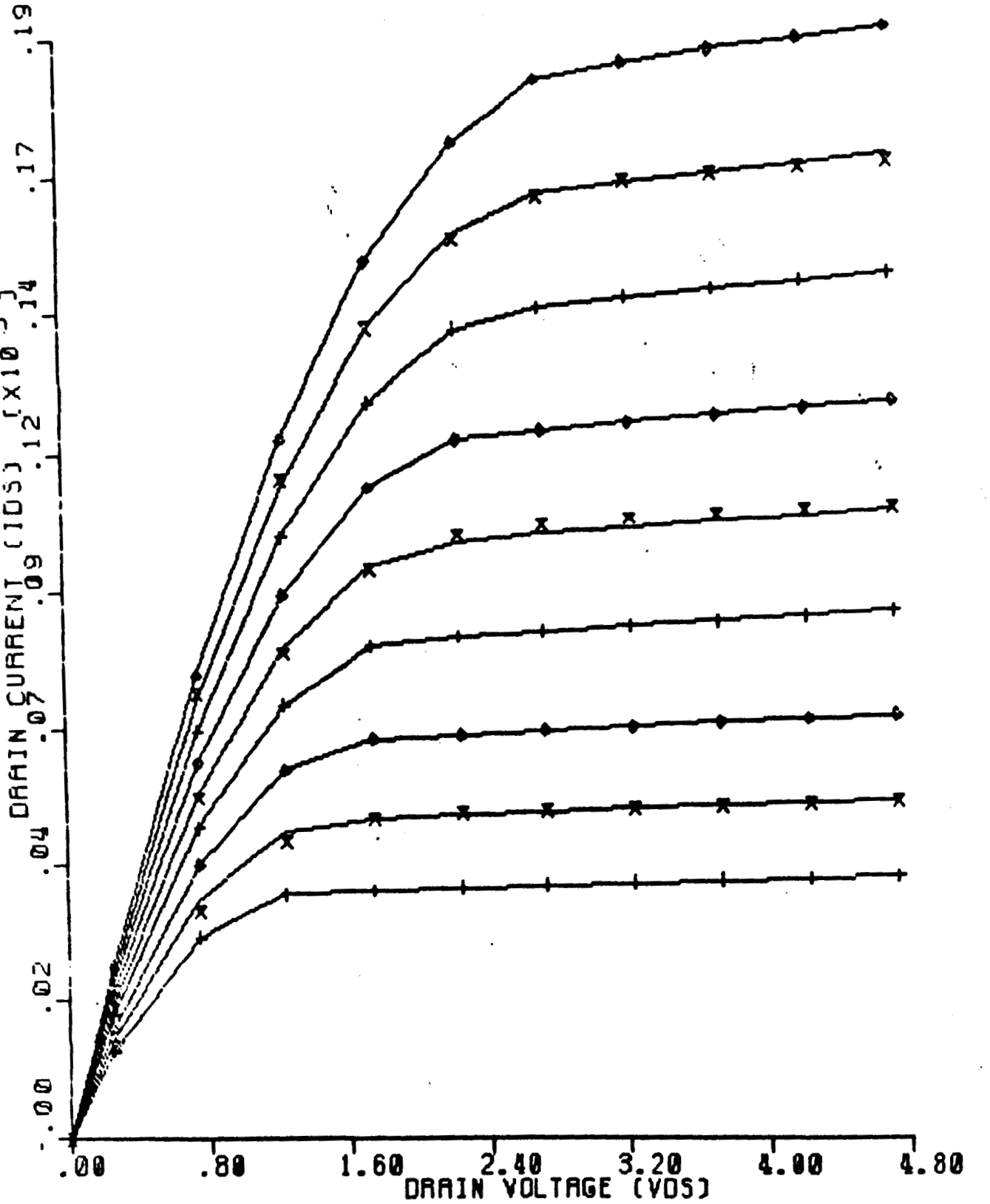
IHANTOLA MODEL

Figure 3



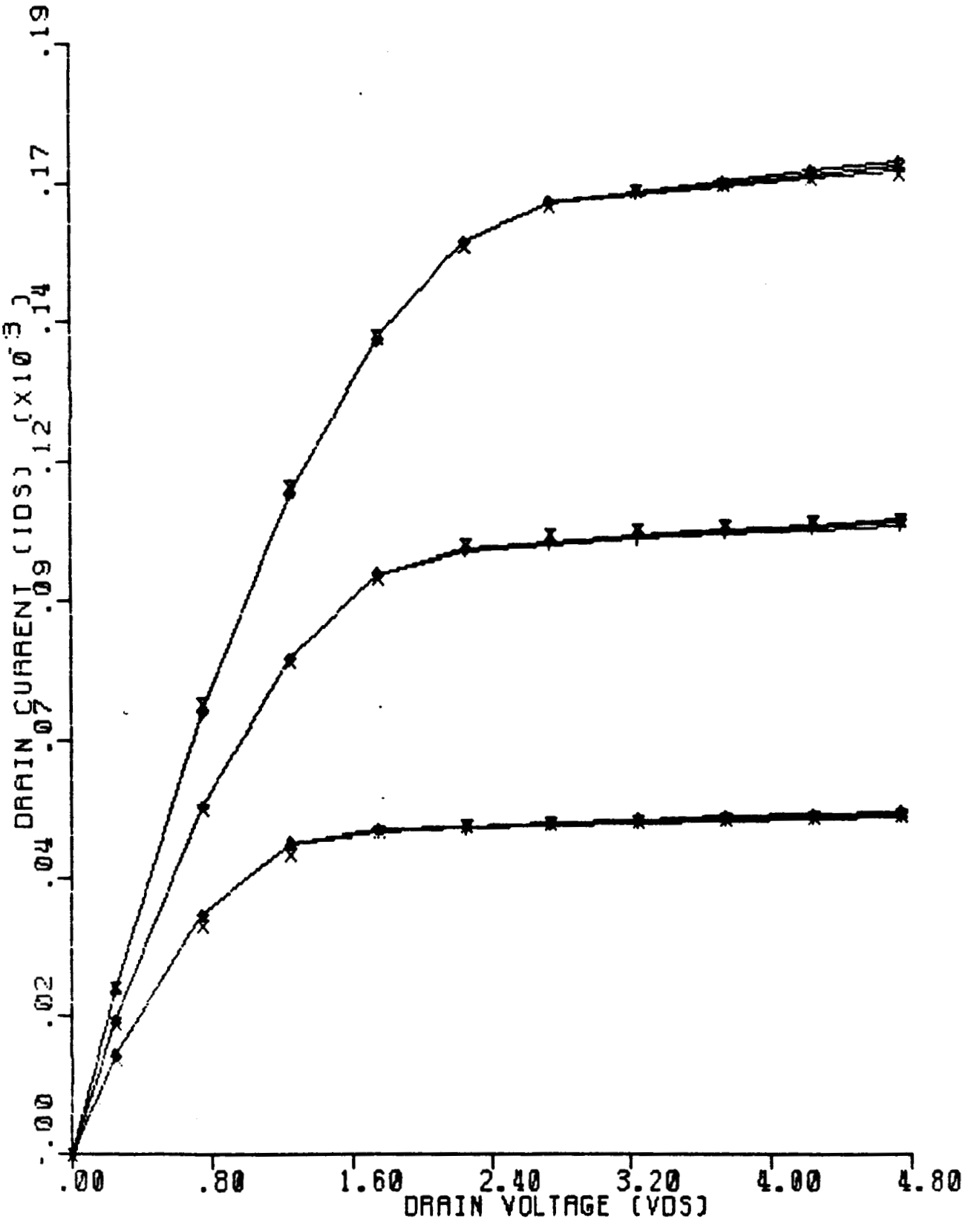
LENGTH = WIDTH = 24.0MU PARM(6)*(1.1 & 0.9)
IHANTOLA MODEL

Figure 4



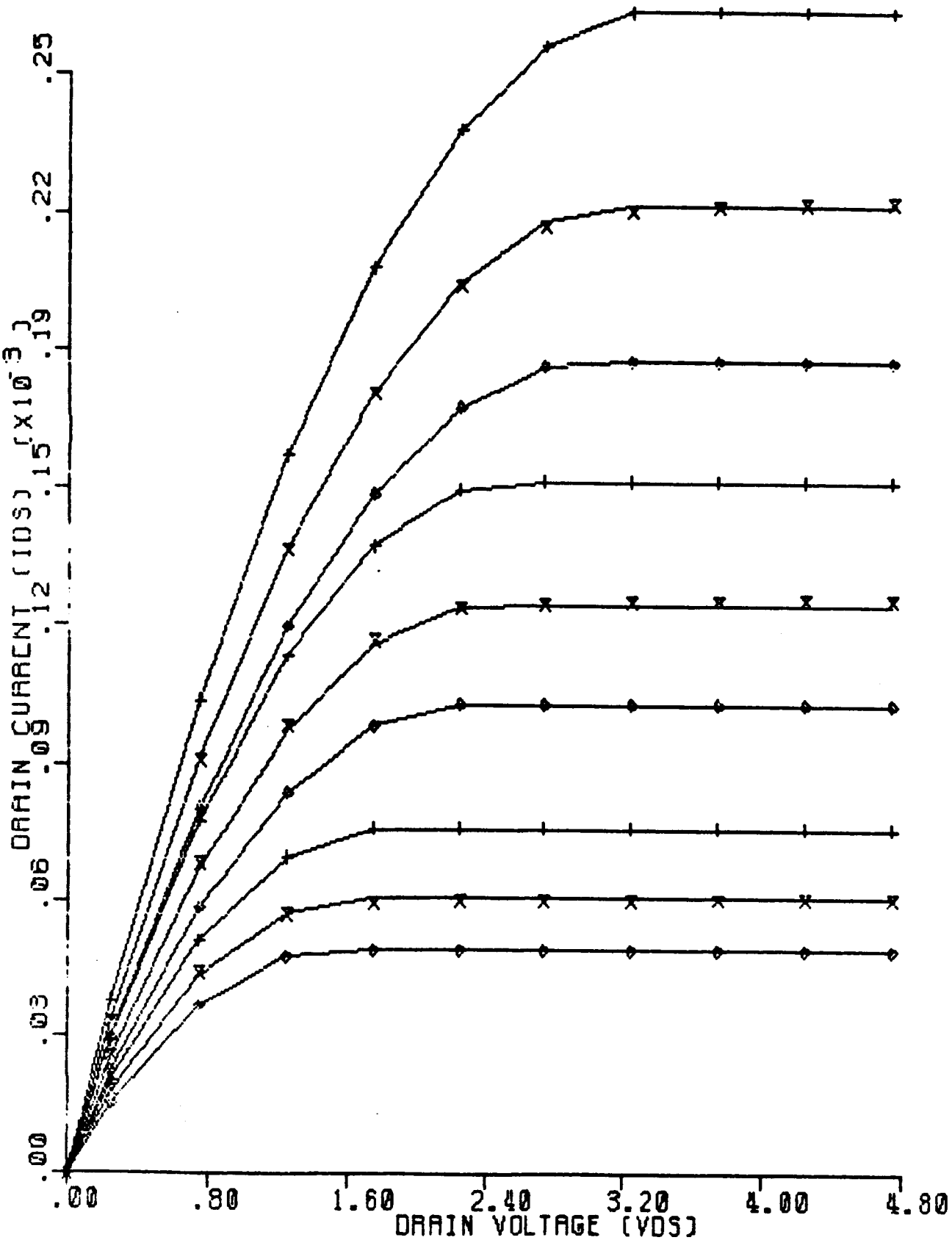
LENGTH = WIDTH = 2.5MU PARM(7)=(1.1 & 0.9)
IHANTOLA MODEL

Figure 5



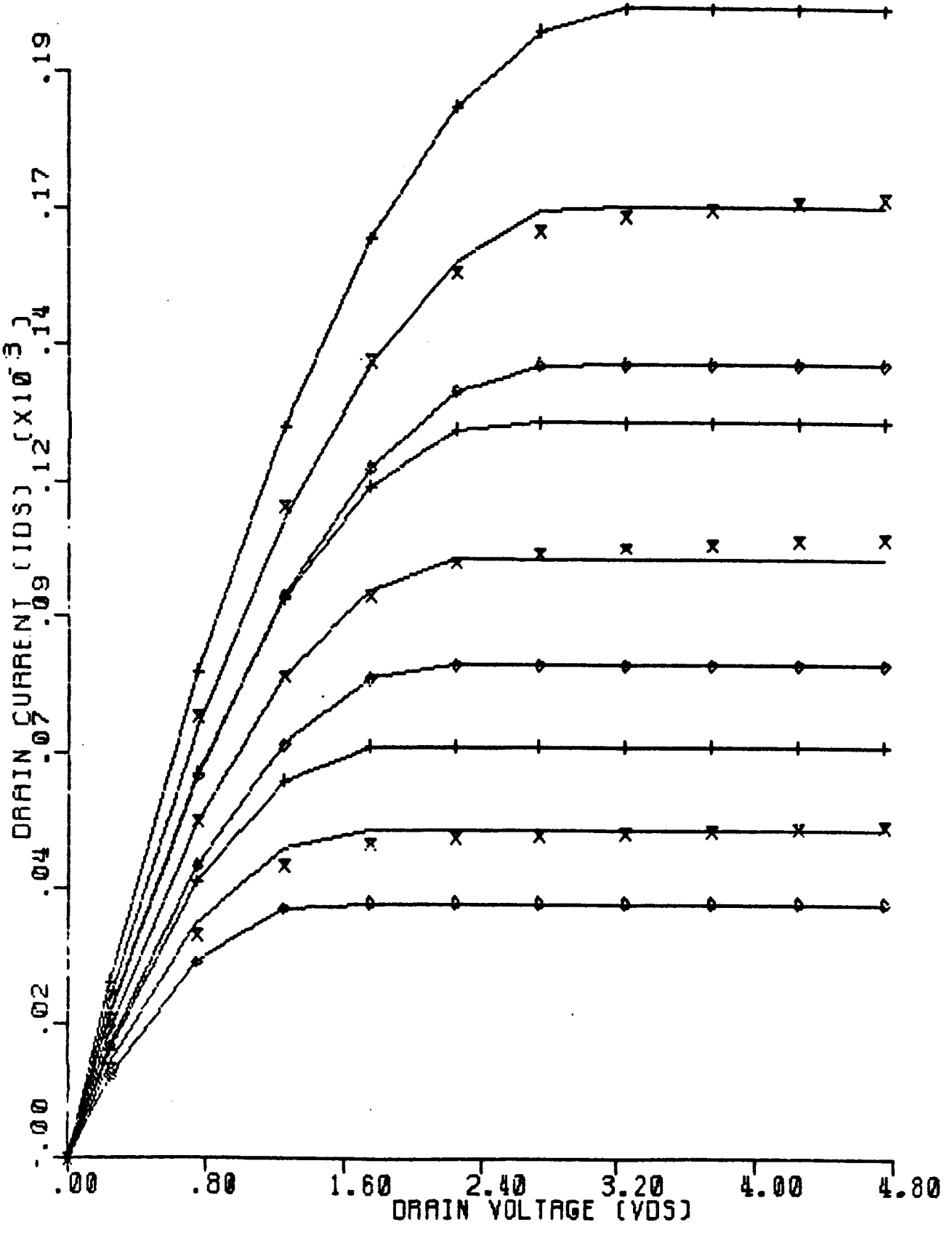
LENGTH = WIDTH = 2.5MU PARM(6)*(1.1 & 0.9)
IHANTOLA MODEL

Figure 6



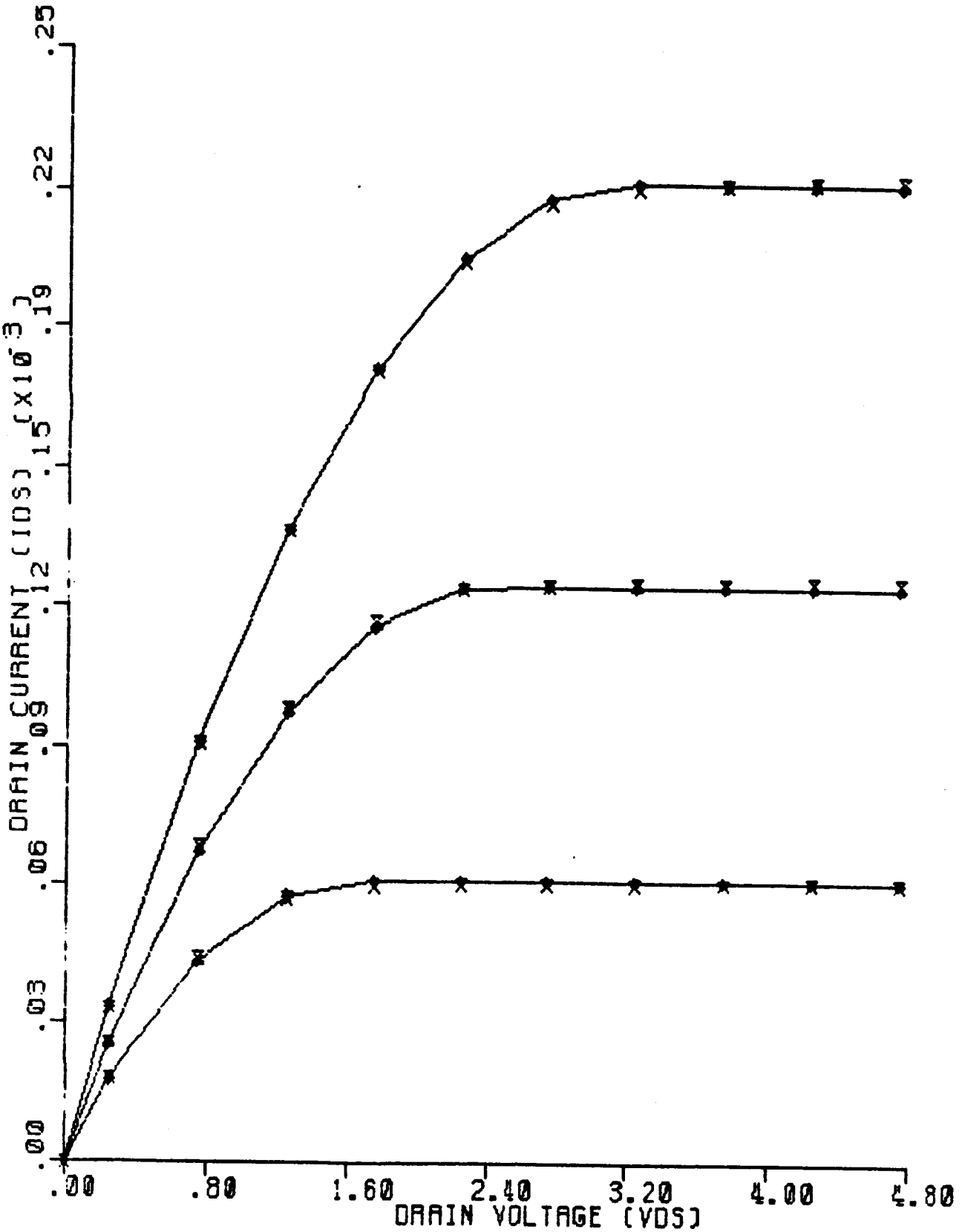
LENGTH = WIDTH = 24.0MU PARM(9)=(1.1 & 0.9)
SPICE2

Figure 7



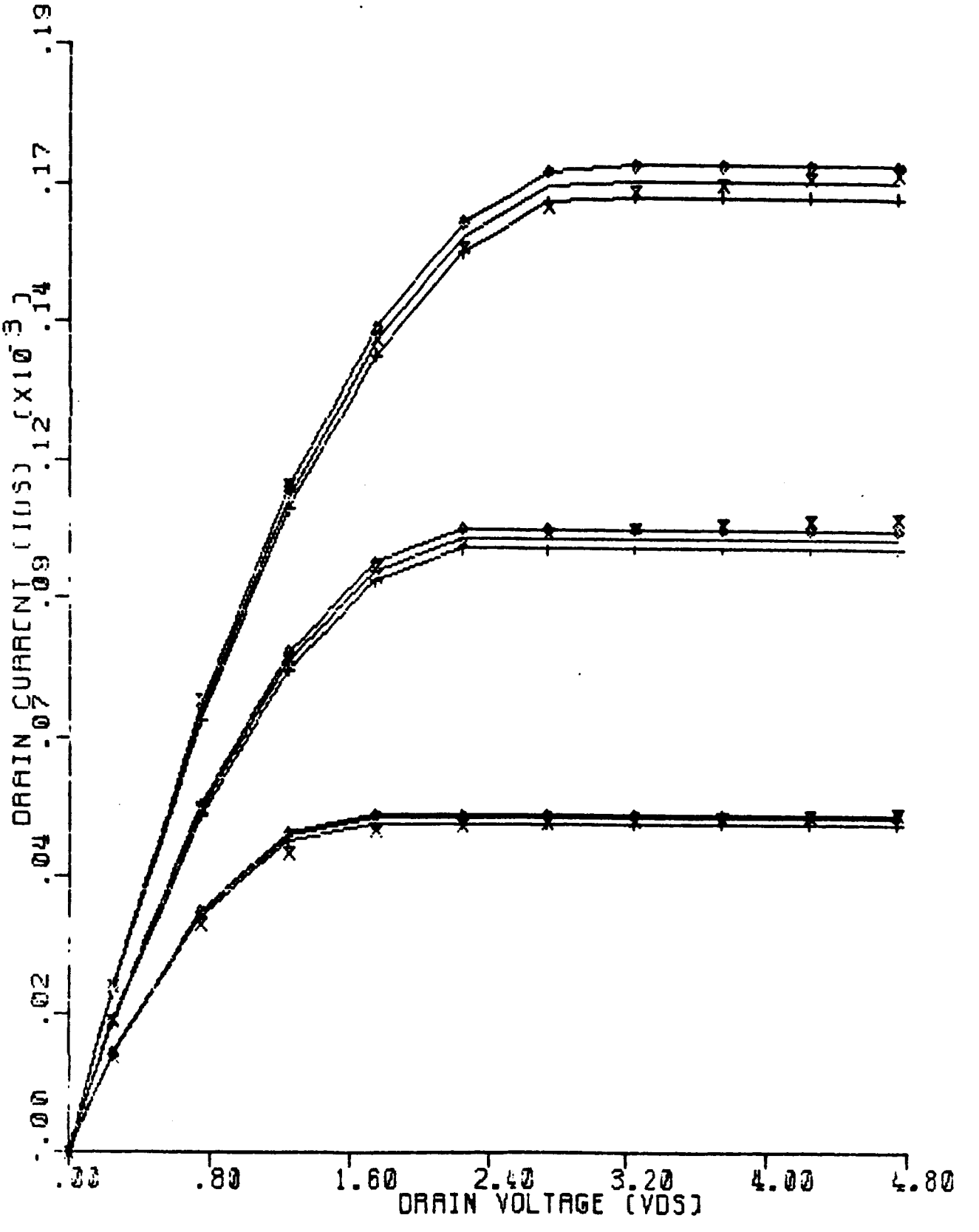
LENGTH = WIDTH = 2.5MU PARM(9)*(1.1 & 0.9)
SPICE2 MODEL

Figure 8



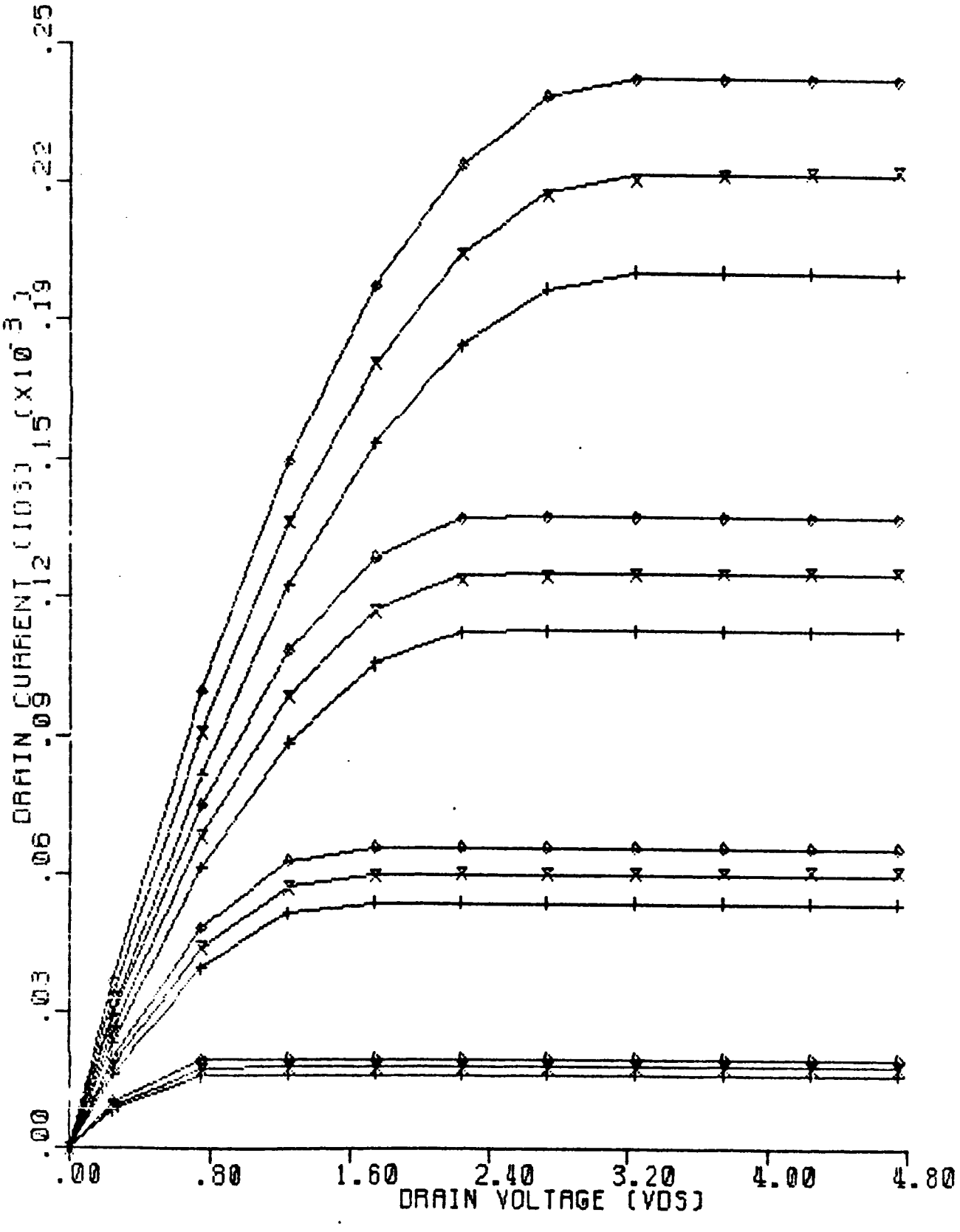
LENGTH = WIDTH = 24.0MU PARM(4)*(1.1 & 0.9)
SPICE2 MODEL

Figure 9



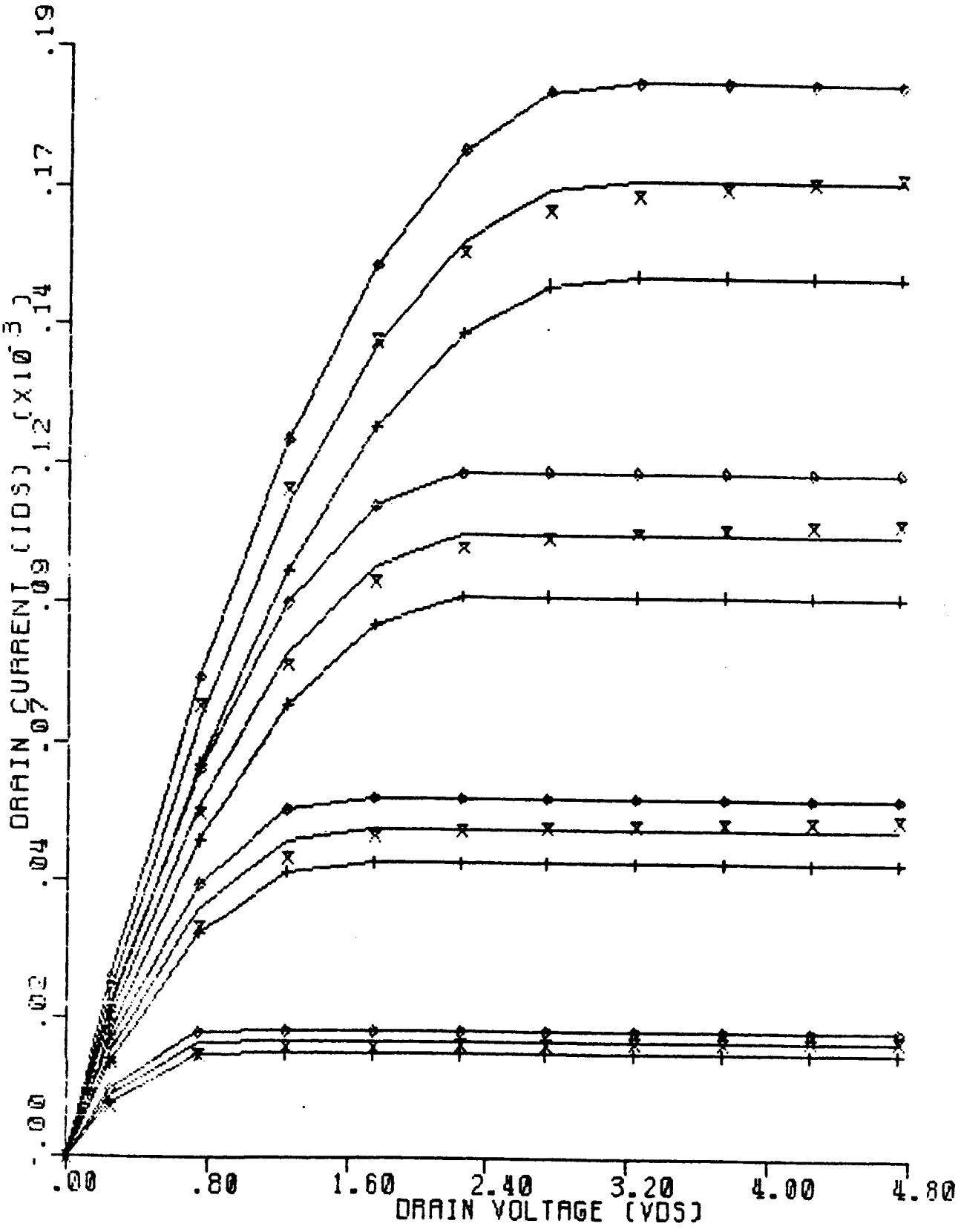
LENGTH = WIDTH = 2.5MU PARM(4)*(1.1 & 0.9)
SPICE2 MODEL

Figure 10



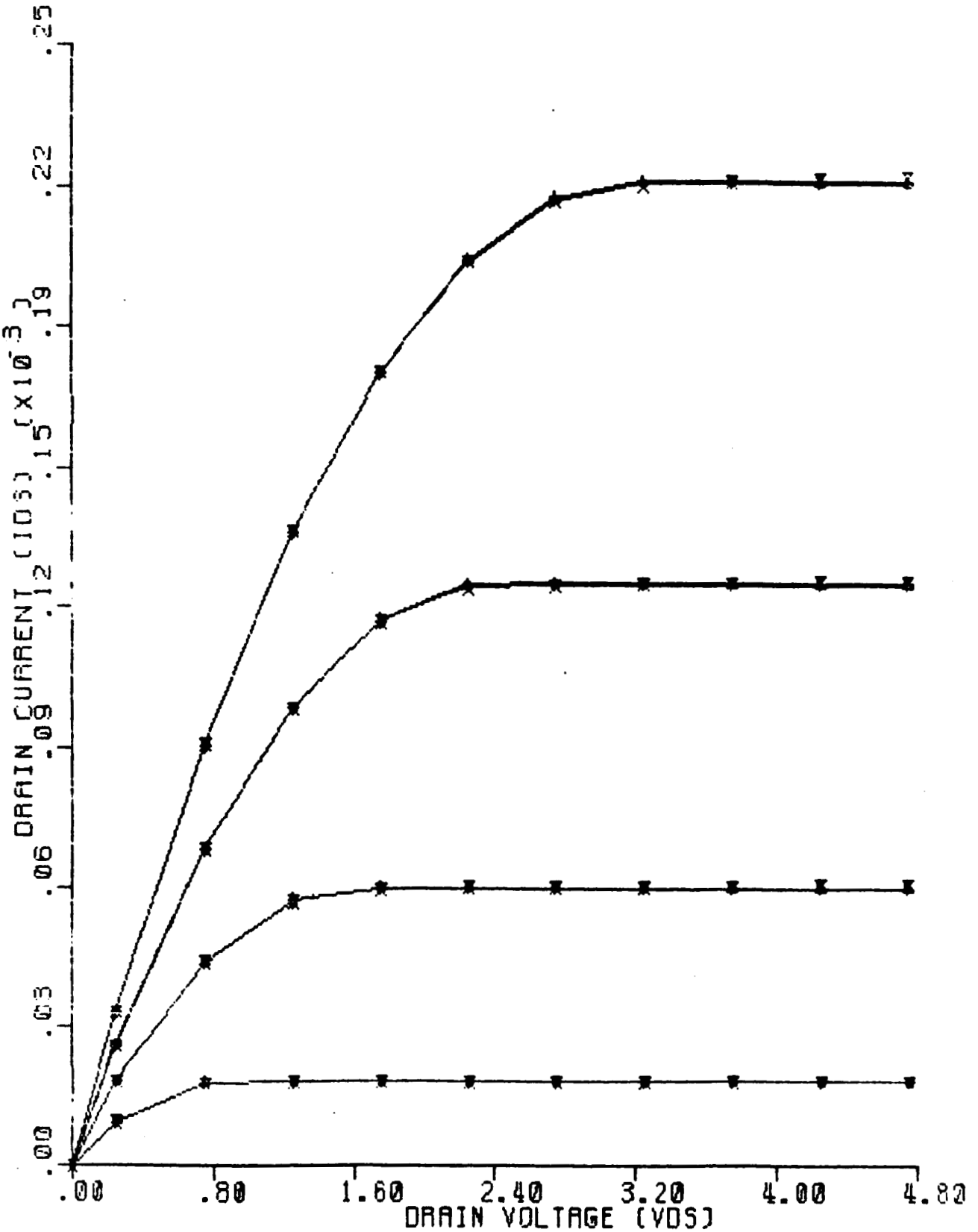
LENGTH = WIDTH = 24.0MU PARM(5)*(1.1 & 0.9)
BREWS MODEL

Figure 11



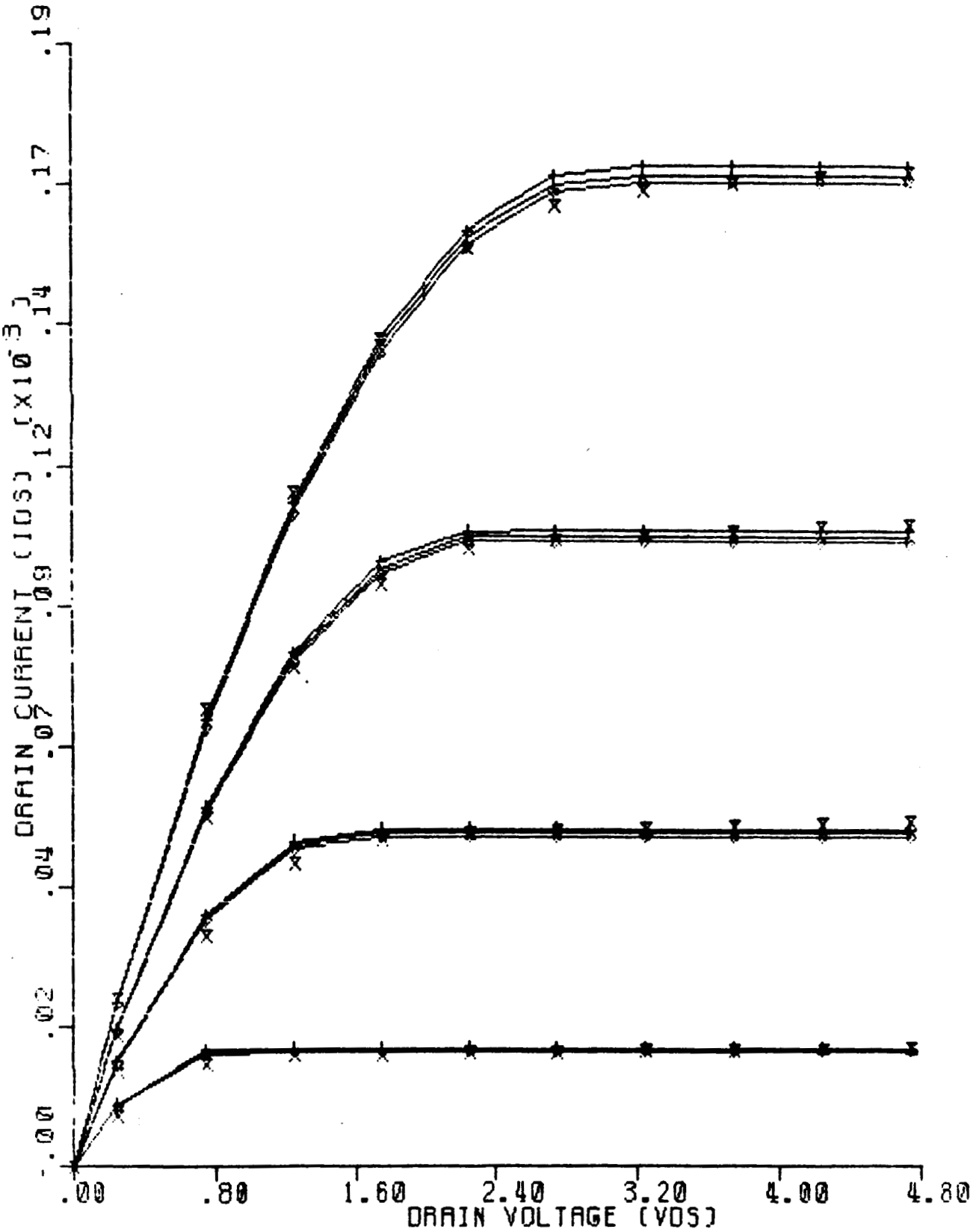
LENGTH = WIDTH = 2.5MU PARM(5)*(1.1 & 0.9)
BREWS MODEL

Figure 12



LENGTH = WIDTH = 24.0MU PARM(6)*(1.1 & 0.9)
BREWS MODEL

Figure 13



LENGTH = WIDTH = 2.5MU PARM(6)*(1.1 & 0.9)
BREWS MODEL

Figure 14

ERRATUM

The results obtained in the Brews model (Model 6) assumed an incorrect dependence on the parameter V_{FB}. Necessary corrections amount to replacing V_{GS} by $V_{GS} - V_{FB}$ everywhere after label 501 in SUBROUTINE BREWS (nine substitutions).