MOS Model Parameter Extraction Techniques: A Comparison

C.G. Cahill, K. McCarthy, W.A. Lane

National Microelectronics Research Centre,
University College, Cork, Ireland.

1. SUMMARY

The extraction of MOS model parameters for circuit simulation has received considerable attention in the last few years. The reasons are clear as simulator models have large numbers (10-15) of more or less physically based parameters, which often are difficult or impossible to relate meaningfully to a physical test structure. Thus non-linear optimization function minimization methods have been proposed as an efficient way to obtain these model parameters [1,2,3]. Essentially these techniques treat the parameter extraction problem as a curve fitting exercise in which the parameters are manipulated in some algorithmic fashion, to minimize the residual error between the model prediction and the measured data set.

This paper discusses two approaches to the optimization based extraction problem and compares them from the point of view of "transistor modelling", circuit simulation use and computer usage. Both approaches use the same optimization algorithm but differ in the method of extraction. The first technique, called "global" or "undirected" optimization extracts all parameters simultaneously by minimizing the error over the entire device geometry and data space. The second technique called "directed" or "sequential" extraction, obtains the model parameters from specific regions of the I-V and/or geometry space.

In general, the directed technique proves superior to the undirected in that it produces a model parameter set which is more closely related to physical values, and which behaves more uniformly and repeatably throughout the data space. Although the total residual error is often larger for the directed extraction its useage of computer time is typically more than an order of magnitude less than for a global optimization.

2. APPROACHES AND METHODS

Workers in this field have used a number of minimization algorithms for the extraction problem. A full discussion of these algorithms is beyond the scope of this work but reference to the techniques can be found in e.g. [4]. For the non-linear MOS device equations, varying from exponential to quadratic in nature, stable convergence properties are of major importance. Doganis et al [1] used the Levenberg-Marquardt algorithm for speed of convergence, evaluating the derivatives numerically, while Yang et al [2] used a combination of the modified Gauss and the steepest descent methods for stability and rapid convergence. For general purpose model development and comparison, a highly robust minimization routine, the Simplex algorithm, has been used by the authors [3] and found to be both stable and reasonably efficient for practical extraction problem sizes (several thousand data points).

The Simplex algorithm is an efficient version of a directed search procedure for function minimization and has the advantage that no evaluation of the function derivative is needed. The general function to be minimized is:

$$E(P) = \sum_{W_i, L_i} \left[\sum \frac{|Ids(P) - Ids'|}{\max(Ids', Idmin)} \right] \qquad (1)$$

where: Ids' is the measured drain current for some bias point, Ids(P) is the predicted drain current for some bias point, Idmin is some minimum drain current allowed, chosen usually from measurement system resolution constraints.
(P) is the vector of model parameters.
W_i, L_i are device width and length.

Additionally, the first derivative of current with respect to one or more of the terminal voltages can be included in the summation of differences [5]. This gives control of small signal as well as DC fitting. The value of $P=P_{min}$ for which E(P) is a minimum is considered to be the best fit of the model parameters. In order to find the optimum point in the parameter space, the simplex algorithm applies a series of geometric operations to the parameter set which drives the vector towards the minimum.

Various error functions can be used; the main alternative is the difference of squares. The main difference between the two is likely to be a reduced spread of errors for the difference of squares function.

For the two extraction strategies under consideration the same simplex optimization algorithm is used. Both the previous workers described the application of their algorithms using some form of directed and undirected optimization. Essentially the size of the problem is the main variation in the work; the undirected optimization can be considered a subset of the directed optimization strategy in which all parameters are extracted simultaneously covering the whole data space. The directed optimization breaks down the problem into a series of (ideally) unrelated extraction sequences, each of the sequences is designed to extract the parameters relevant to some part of the I-V/geometry space.

In the undirected extraction problem, other than: i) applying limits to some of the model parameters, ii) choosing a suitable set of initial guesses for the parameter vector, and iii) fixing one of any fully redundant pairs, little set-up work is needed. The application of limits to some or all of the model parameter vector ensures that a reasonable relationship to the physical world is maintained, alleviates the problem of regional parameter redundancy (i.e. some parameters may not be significant for certain combinations of other parameters), and avoids problems of model "unreasonableness" (e.g. negative output conductances may occur for certain values or combinations of values of input parameters).

Setting up the directed optimization sequences is considerably less straight forward and requires reasonably detailed understanding of the model equations and parameter interactions. Establishing a software extraction system that is flexible enough to cater for all the file manipulation, parameter manipulation and extraction strategy documentation in a robust and "user friendly" fashion has proved a substantial task. The program DEPS [6] can be instructed to extract parameters in a highly flexible and efficient manner. It is used in this study to perform both directed and undirected optimizations, the undirected optimizations being simply treated as a single extraction sequence.

Details of the extraction strategy chosen for SPICE level 3 [7] MOS model is given in figure 1. These are by no means the only possible sequences. As is shown, a variety of device sizes and measurements are used to obtain data pertinent to the relevant parameter extraction sequence. In general the strategy is to fit the important, large geometry and geometry independent parameters (eg. low field mobility, substrate concentration, etc.) first using a device large

Figure 1

Example of the directed optimisation procedure

for SPICE level 3

NSUB and PHI are evaluated from VT vs Vbs of the large device.

THETA and NFS are optimised to the I_Vgs data for the large device.

THETA is optimised to the I-Vds data of the large device.

GAMMA is optimised to the I_Vgs data of the large device at several Vbs.

XJ and LD are optimised to I_Vgs data of the wide devices using a range of lengths. VMAX is also included.

DELTA is optimised to the I_Vgs data for the long channel devices with a range of device widths.

ETA, KAPPA and VMAX are optimised to I_Vds data for the wide devices using a range of lengths and Vbs values.

enough that edge effect can be ignored, then bring in secondary or geometry dependent parameters using device sizes and bias ranges selected to emphasize these effects. Some parameters (e.g. VMAX in the SPICE models) may be allowed to vary in the optimization at more than one point in the set of extractions, as a good initial guess is needed to fix some other parameter and then find a final value that gives a best average fit. This is largely to compensate for the problem that parameters are not always decoupled sufficiently to allow unrelated extraction sequences. Both DC and conductance (slope) fitting can be applied at any time to the optimization sequence.

3. RESULTS AND COMPARISONS

(a) Device Fitting

In this test the data used was taken from a set of polysilicon gate N-well CMOS devices. A total of 13 p-type and 13 n-type device geometries were measured with effective lengths ranging from 2.7um to 17.7um and effective widths from 56um to 8um. All measurements were repeated for 3 values of substrate bias 0, 2.5 and 5.0 V. IDS-VDS data was collected at 100mV V_{DS} with V_{GS} ranging from 0 to 6V or 15V for V_{DS} and V_{GS}, depending on device channel length. Eleven steps in V_{DS} and 9 in V_{GS} were used for all devices. V_T vs V_{BS} data was also measured for the devices using 20 points, 10 between 0V and 1V and 10 between 1V and 10V. This was because the devices tested have far from uniform channel doping characteristics so that more data is used at low V_{BS} to capture the initial high V_{BS} - V_T variation.

During the extractions for the SPICE level 3 P and N models used as the example in this comparison, the V_{TO}, LD, DW, TOX and UO were predetermined from electrical methods. This could obviously be included as extra optimization sequences in the directed extraction. However in this case, it was felt to be a fairer and more useful test of the two strategies if the constraints normally applied to the SPICE models (i.e. that some directly measurable physical meaningfulness be associated with key circuit design parameters), were applied for this test. The same initial guess of the parameter vector was used for both directed and undirected optimization.

Figure 2 and 3 shows full voltage range I_{DS} - V_{DS} curves for N and P channel CMOS devices using both extraction strategies. Initial errors for P and N-channel devices were both approximately 50%, evaluated over the whole geometry/IV data space. In all cases, for the devices shown, both peak and average errors are reduced for the directed optimization compared to the undirected case. This will obviously not be

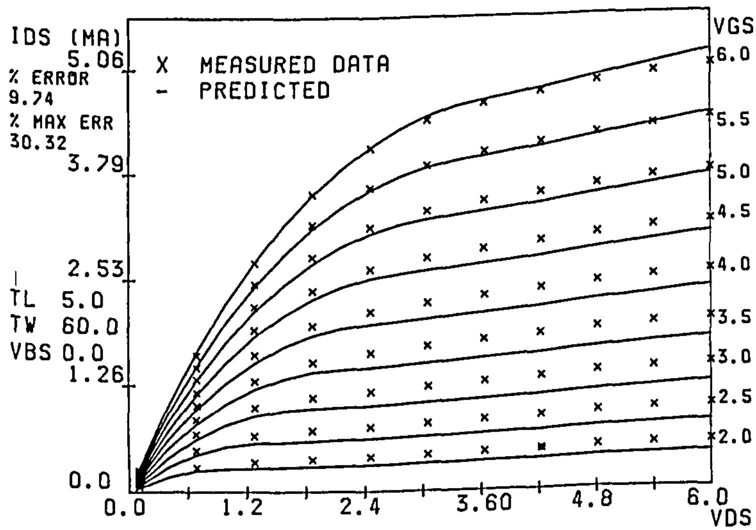

Fig 2(a). Sample I-V curve obtained for N channel device using undirected optimisation procedure.

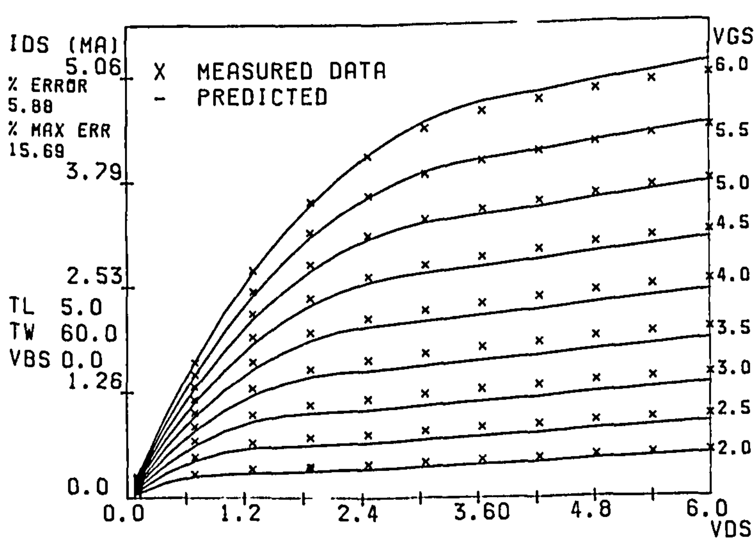

Fig. 2(b). Sample I-V curve obtained for P channel device using directed optimisation procedure.

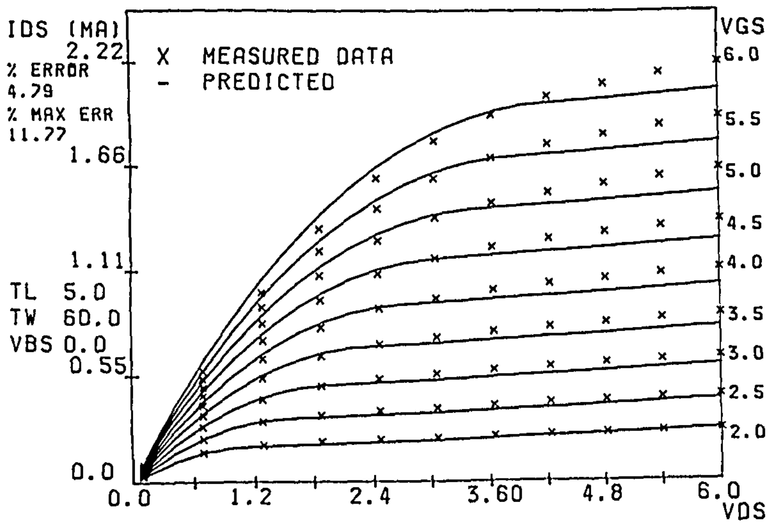

Fig. 3(a). Sample I-V curve obtained for P channel device using undirected optimisation procedure.

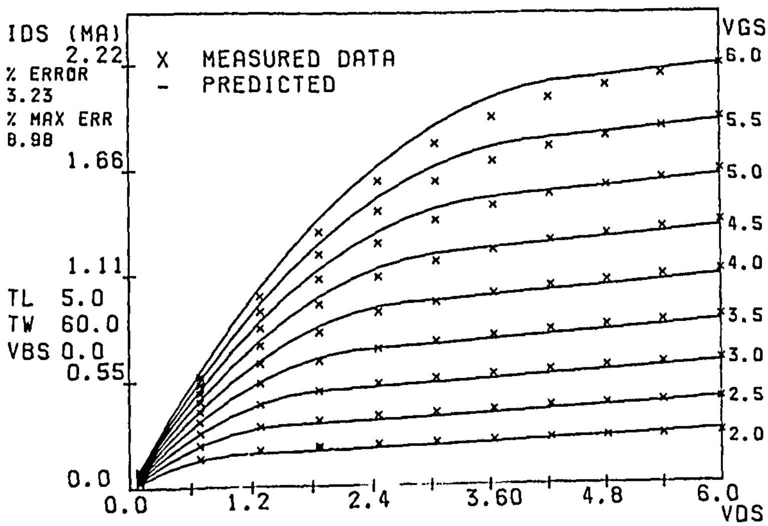

Fig. 3(b). Sample I-V curve obtained for P channel device using directed optimisation procedure.

universally true, indeed, as Table I shows, the overall DC and GDS errors are reduced by 5½% and 4% respectively for the directed optimization P-channel case but are greater by about 1 and 2% respectively for the N-channel case. Table I also shows the times taken to perform the parameter extraction. Undirected optimization for this data set and initial guess is 12 times slower for the P-channels and 27 times slower for the N-channels giving the directed optimization a decisive advantage from a CPU time point of view. The generally poorer performance of the P-channel model (DC error > 10%) compared with the N-channel model (DC error < 7.5%) is due to the difficulty that the SPICE model has in predicting P-channel behaviour. The N-well P-channel device is especially difficult due to the highly non-uniform impurity profile observed in this type of device. The undirected optimization is probably stuck at a local minimum for this test. Restarting with a new initial parameter vector could yield a better fit.

Table I

Comparison of Extraction Strategies

Strategy	Final Residual		CPU Time
	DC Error	GDS Error	
Undirected P-ch	15.8%	49%	33 min
Directed P-ch	10.4%	45%	2.7 min
Undirected N-ch	6.6%	34.3%	98 min
Directed N-ch	7.4%	37%	3.6 min

While the undirected N-channel extraction gave a lower global residual error, it has been observed that this is often achieved at the expense of poorer local region fitting. In other words, the data set used to extract the undirected set of parameters may weight the model towards a best fit in a particular region (e.g. high current, high V_{DS}, V_{GS}). Some examples of this are shown, for the test extraction described above, in figures 4, 5 and 6.

Figure 4 shows the measured and calculated V_T vs. V_{BS} behaviour for a 56/2.7um effective geometry N-channel device. While neither extraction method gets a particularly close fit - peak errors are 110mV and 170mV at 10V back bias for directed and undirected methods respectively - the directed extraction strategy is clearly an improvement on the undirected case. Similar behaviour is observed for other geometries.

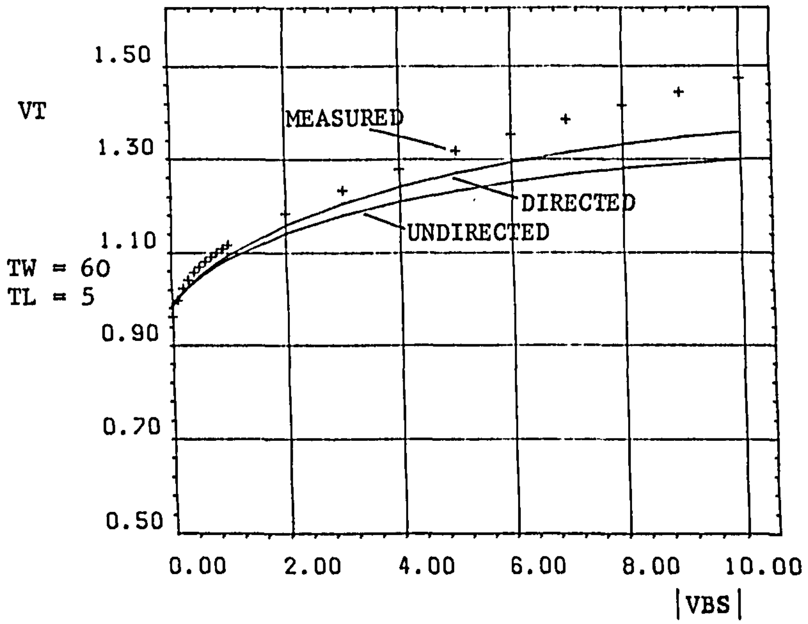

Fig. 4. A comparison of threshold voltage prediction for an N channel device using models obtained from the directed and undirected optimisation techniques.

Figures 5(a) and (b) show the lowest measured current (V_{GS} = 2V) I_{DS}-V_{DS} curve for the same N-channel device described above. Figure 5(a) shows the fit for the undirected case which is clearly worse than the fit for the directed case in fig. 5(b) (DC error 25% compared with 11.5%). The conductance is also very poorly modelled for the undirected case, a factor of critical importance in analogue circuit modelling.

A further example is given in figures 6(a) and (b) which show I_{DS}-V_{GS} curves for a large geometry N-channel device (W_{eff}/L_{eff} - 56/17.7) and again the directed extraction, figure 6(b), gives a clearly better fit (1.5% vs. 3.6%) to the data than the undirected method. The model parameters controlling mobility rolloff are apparently overestimated for the undirected case. Again similar behaviour is observed for other devices.

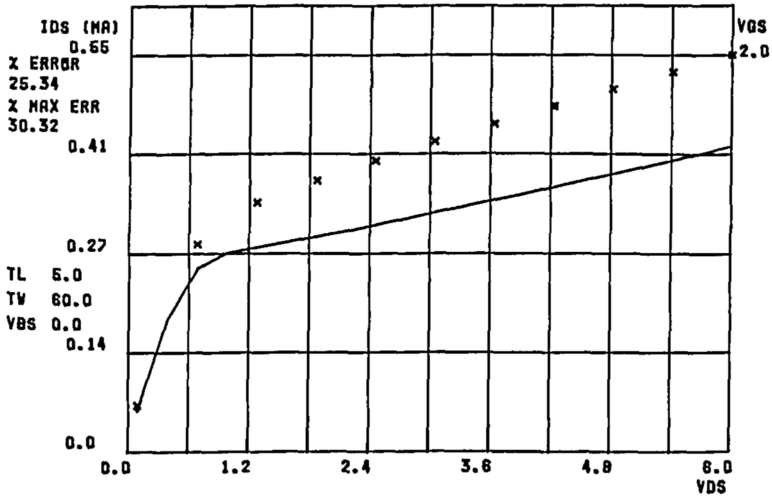

Fig. 5(a). Sample I-VDS curve for low VGS obtained from the undirected procedure.

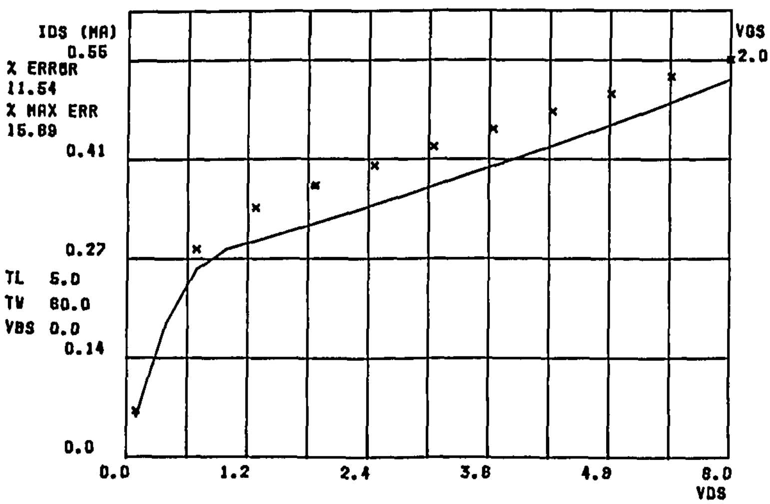

Fig. 5(b). Sample I-VDS curve for low VGS obtained from the directed procedure.

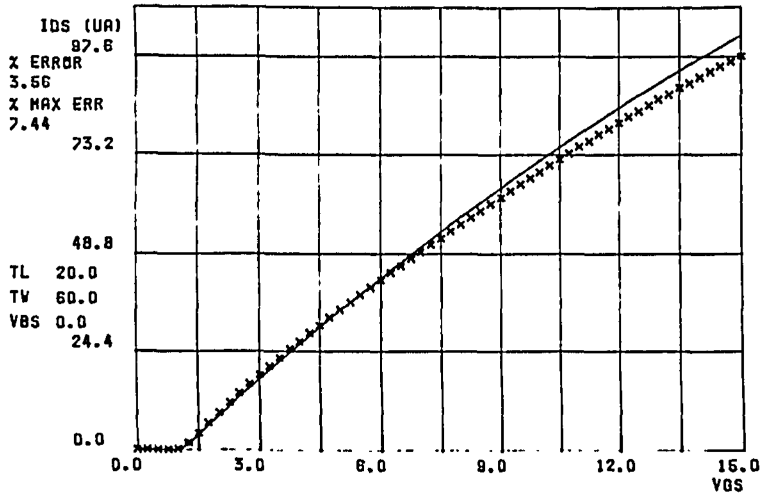

Fig. 6(a). Sample I-VGS curve for N channel device using undirected optimisation (VDS = 0.1V).

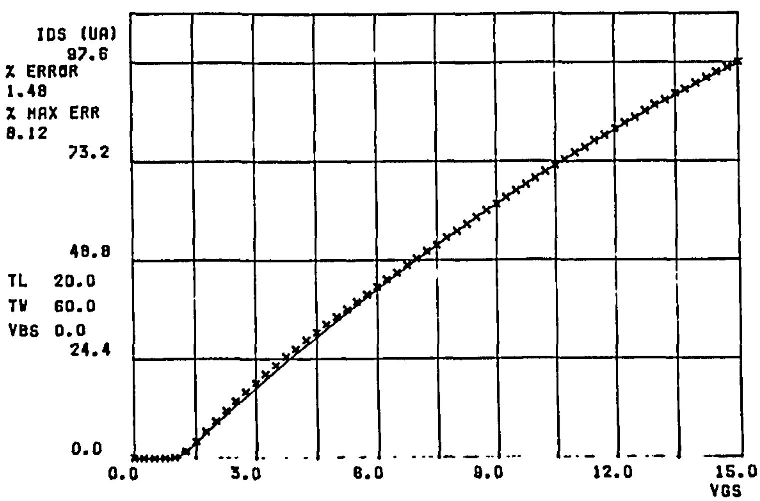

Fig. 6(b). Sample I-VGS curve for N channel device using directed optimisation (VDS = 0.1V).

(b) Circuit Modelling

The SPICE level 3 models extracted above were used to simulate a classical, N-ch input, 2-stage CMOS operational amplifier in open loop configuration. The DC transfer function was also obtained for the circuit using point-by-point DC measurement. Single N and P channel models were used for all devices (i.e. no alteration of model parameters for different geometry devices). Figure 7 shows the measured and predicted transfer curves in the high gain region. The curves have been translated laterally to compensate for input offset voltage but are otherwise as measured or simulated.

While neither simulation is particularly close to the measured gain, the directed simulation is clearly much better. Measured gain is about 1660, while the directed extraction set predicts a gain of 3680 and the undirected set, 6330. Most of the devices that control gain (given by the ratios of transconductance to output conductance for any pair of amplifying devices) have larger than minimum geometries. The SPICE level 3 model has particularly poor scaling of output conductance with channel length which gives rise to most of the GDS errors shown in Table I. This is the main reason for the poor prediction of the op-amp gain with this model. In other work performed by the authors, improving the models' conductance, modelling prediction or simply fitting models only to the device sizes in question (thus reducing the models' generality) produces far superior gain predictions using these parameter extraction techniques.

4. CONCLUSION

Two methods of extracting MOSFET model parameters have been compared. While both use an optimization algorithm to achieve some "best fit" parameter vector, wide differences exist in their performance and ease of use. The very general technique of extracting all parameters simultaneously, undirected optimization, is very CPU intensive and results are closely linked to the distribution of data throughout the I-V and geometry space. However, once a model has been coded and appropriate limits placed on important parameters the extraction is otherwise straightforward.

Directed extraction of parameters in sequence is 10-30 times more efficient from a CPU time viewpoint and can produce model parameter vectors which give better prediction of current in localized I-V regions than the undirected case. Though the undirected extraction can give a lower global residual error, this is often at the expense of

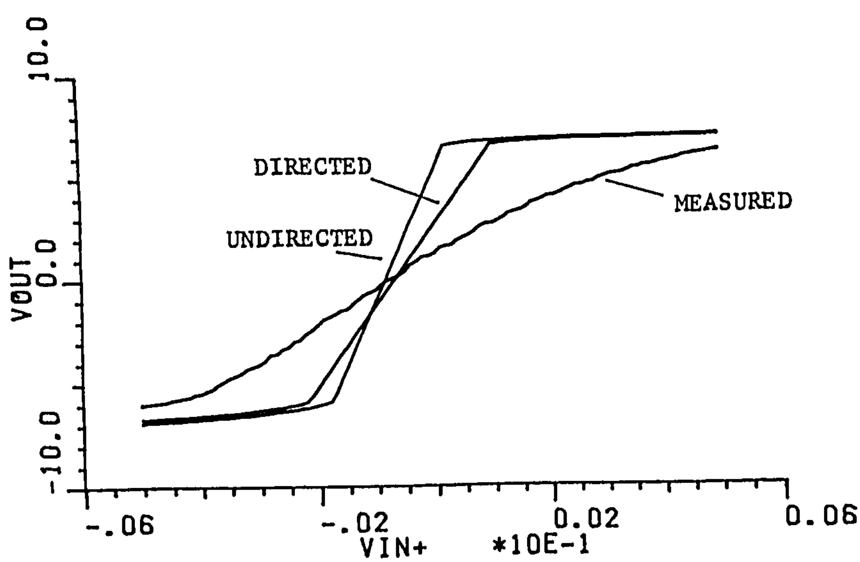

Fig. 7. Comparison of op-amp open loop prediction using the directed and undirected models.

poorer local region matching which means that the model's generality is compromised from a circuit design point of view. One possible problem with sequential parameter extraction is the requirement that the model's parameters be highly independent or non-interacting with each other. This is often not the case for existing MOS models (SPICE level 3 being a case in point). However, a carefully thought out extraction package can get around this problem by allowing multiple attempts at extraction with different constraints applied, or to different data sets, to arrive at a parameter value which gives a best fit over the widest range of devices.

Given the CPU efficiency of the directed extraction procedure, this technique has recently been applied at this location to statistical modelling of CMOS parameters to take account of wafer variation. Successful prediction of the mean and standard deviation of such performance parameters as current level, inverter gain and switching levels has been achieved for simple circuits.

5. REFERENCES

1. K. Doganis, D.L. Scharfetter, "General Optimisation and Extraction of IC Device Model Parameters", IEEE Trans. Elec. Dev. Vol. ED-30 No. 9, Sept. 1983, pp 1219-1228.

2. P. Yang, P.K. Chatterjee, "An Optimal Parameter Extraction Program for MOSFET Models", IEEE Trans. Elec. Dev. Vol. ED-30 No. 9 Sept 1983, pp 1214-1219.

3. P. Conway, C. Cahill, W.A. Lane, S.U. Lidholm "Extraction of MOSFET Parameters Using the Simplex Direct Search Method". IEEE Trans. CAD., Vol. CAD-4 No. 4 Oct 1985, pp 694-698.

4. R.W. Daniels, "An Introduction to Numerical Methods and OptimizationTechniques". North Holland, New York, 1978.

5. Y. Tsividis, G. Masetti, "Problems in Precision Modelling of the MOS Transistor for Analog Applications". IEEE Trans. CAD., Vol. CAD-3, No. 1, January 1984.

6. C.G. Cahill, W.A. Lane, "Extraction of MOS Model Parameters Using a Directed Parameter Optimization Sequence", to be submitted to IEEE Trans. CAD, for publication.

7. A. Vladimirescu and S. Liu "The Simulation of MOS Integrated Circuits using SPICE2", Memo No. UCB/ERL M807 Feb 1980. Univ. Calif. Berkeley.