

Implementation and quality testing for compact models implemented in Verilog-A

Anindya Mukherjee¹, Andreas Pawlak¹, Michael Schroter^{1,2}, Didier Celi³, Zoltan Huszka⁴

¹CEDIC, Technische Universität Dresden, 01062 Dresden, Germany, ²ECE Department, UC San Diego, La Jolla, USA, ³STMicroelectronics, Crolles, France, ⁴ams AG, A8141 Unterpremstatten, Austria

Abstract --- An overview on the implementation of new physical effects into the heterojunction bipolar transistor compact model HICUM/L2 is presented along with a description of the quality testing procedures performed before its public release for production circuit design in commercial simulators. Related topics such as potential measures for model run time improvements and failures are also discussed. Significant differences in run time for different commercial circuit simulators reflect their different approaches towards compact model implementation and solution of the non-linear circuit equation system.

Index Terms --- bipolar transistors, compact modeling, HICUM, circuit simulation, modified nodal analysis (MNA).

I INTRODUCTION

HICUM Level 2 (HICUM/L2) is a physics-based compact bipolar transistor model for silicon bipolar junction transistors (Si BJTs) and silicon-germanium heterojunction bipolar transistors (SiGe HBTs) [1]. It was recently also successfully applied to InP DHBTs [2]. Each element of the equivalent circuit is associated with mathematical functions describing the corresponding bias and temperature dependence. Since its acceptance as a standard HBT compact model (CM) by the Compact Modeling Coalition (CMC) in 2003 the model formulations have undergone various improvements as well as extensions. The actual version 2.34 includes all known physical effects of even most advanced SiGe HBTs relevant for circuit design. Being a standard compact model in commercial circuit simulators, HICUM undergoes a thorough test procedure before its release. This paper describes testing procedures and demonstrates the due diligence being exercised for releasing new model versions.

The organization of the paper is as follows. Section II presents a brief overview of HICUM/L2 and its implementation. Section III and IV describe the updates GICCR and the NQS effect implementation. Discussions on model runtime comparison using the simulators Keysight ADS and Cadence Spectre, are included in section V. Model testing is described in section VI.

II HICUM/L2

A. Brief overview

HICUM/L2 is based on the so-called Generalized Integral Charge-Control Relation (GICCR) [1],[3],[4]. However, in contrast to the Gummel-Poon model (GPM) [5] as well as the SPICE-GPM (SGPM) [6] and its variants, in HICUM/L2 the (G)ICCR concept is applied consistently without inadequate simplifications and additional fitting parameters (such as the Early voltages). The complete equivalent circuit is given in Fig. 1. It consists of the core model containing the elements of the internal tran-

sistor as well as the many components representing the peripheral and parasitics effects for complete 3D structures. Additionally, the adjunct networks for modeling electro-thermal effects, NQS effects (cf. section IV) and noise-correlation are shown.

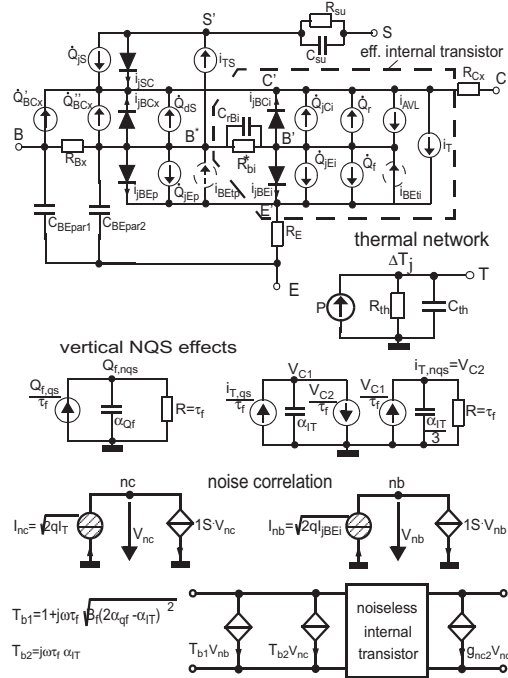


Fig. 1 HICUM/L2 equivalent circuit with adjunct networks for self-heating, vertical NQS effects and noise correlation.

Since reliable design and optimization of high-speed circuits requires accurate modeling mainly of the dynamic transistor behavior, quantities like depletion capacitances and the transit time of mobile carriers as well as the associated charges are considered as basic variables of the model.

The important physical and electrical effects taken into account by HICUM/L2 are briefly summarized below:

- high-current effects (including quasi-saturation and collector current spreading)
- distributed high-frequency model for the external base-collector region
- emitter periphery injection and associated charge storage
- emitter current crowding (through a bias and frequency dependent internal base resistance)
- parasitic (bias independent) capacitances between base-emitter and base-collector terminal
- vertical non-quasi-static (NQS) effects for transfer current and minority charge

- temperature dependence and self-heating
- weak avalanche breakdown in the BC junction
- tunneling in the BE junction
- parasitic substrate transistor
- bandgap differences (occurring in HBTs)
- lateral (geometry) scalability and statistical modeling capability.

B. Model implementation

The original implementation of the first versions was written in FORTRAN including the (time consuming) manual derivation of all required derivatives. The latter makes the implementation model updates to new versions very time-consuming. In contrast, the introduction of Verilog-A (VA) enables a significantly faster model development. VA-compilers translate high-level code into code of a high-level programming language (typically C) that is specific for the different simulator to model interfaces. This allows any compact model with elements compliant to the SPICE simulator standard to be linked to any of the (supported) circuit simulators and has significantly reduced the implementation burden for model developers. Among others, the biggest advantage of using VA is the automated generation of the derivatives required for the Jacobian of the non-linear model equations by the compiler.

A disadvantage of VA is that the implementation of various physical effects is not straightforward and sometimes causes increased model code, since the VA-language is limited to elements from network theories, i.e. current and voltage dependent (dynamic) current and voltage sources. Special care is to be considered when implementing such effects and the subsequent testing before the release. The next section describes certain important effects and their implementation issues.

III Implementation of the GICCR

A. Internal iteration

The transfer current in HICUM/L2 is based on the GICCR. The master equation reads

$$I_T = \frac{c_0}{\int h(x)p(x)dx} \left[\exp\left(\frac{V_{BE}}{V_T}\right) - \exp\left(\frac{V_{BC}}{V_T}\right) \right], \quad (1)$$

which is a closed-form solution of the drift-diffusion equation. Here, c_0 is a constant and

$$Q_{p,T} = q \int h(x)p(x)dx \quad (2)$$

is the weighted (transport related) hole charge with integration boundaries equivalent to the location of the internal transistor contacts E' and C'. For compact modeling, it was useful to divide the weighted hole charge into separate components associated with regions in the transistor

$$Q_{p,T} = Q_{p0} + h_{jEi}Q_{jEi} + h_{jCi}Q_{jCi} + \Delta Q_{f,T}, \quad (3)$$

where Q_{p0} is the zero-bias hole charge, $h_{j(E,C)i}Q_{j(E,C)i}$ the weighted depletion charges and $\Delta Q_{f,T}$ the bias dependent weighted minority charge. Note that the weighted hole charge is normalized by the weight factor h_0 for Q_{p0} (which follows from the transformation of (2) into (3)) in order to avoid an additional parameter. This changes c_0 in (1) to $c_{10} = c_0/h_0$.

While the weighted depletion charges are just voltage dependent and represent the impact to the classical reverse and forward Early effect, the minority charges in (3) also depend on the transfer-current. This finally gives the forward component of the transfer-current,

$$i_{Tf} = \frac{c_{10} \exp(V_{BE}/V_T)}{Q_{p,T,low}(V_{BE}, V_{BC}) + \Delta Q_{f,T}(i_{Tf}, V_{BC})}, \quad (4)$$

and thus an implicit equation for i_{Tf} and $\Delta Q_{f,T}$, respectively. Here, $Q_{p,T,low}$ is the sum of the first three terms in (3). In the SGPM the implicit solution is avoided by simplifying $\Delta Q_{f,T}$ to $\tau_{f0}i_{Tf}/Q_{p0}$, i.e. assuming a current independent transit time, allowing (4) to be solved directly as a quadratic equation. This simplification is not admissible for advanced HBTs, requiring

$$Q_{p,T} - (Q_{p,T,low} + \Delta Q_{f,T}(i_{Tf})) = 0 \quad (5)$$

to be solved iteratively. The iteration is implemented in the official VA-formulation of the model, leading to an internal iteration executed at each iteration step of the circuit iteration. This internal iteration has several drawbacks. First, a number of equations is repeated inside the model evaluation for each circuit iteration step. A typical performance of the internal iteration is given in Fig. 2. Especially for simulations in saturation, a large number of iterations may be required for solving (5). Second, the derivative of the charge with respect to the current (i.e. the weighted transit time) is required to be calculated in the model code, leading to a large number of additional automatically generated derivative terms, of which a large percentage does not have a numerical impact on the solution, but just increases the model runtime.

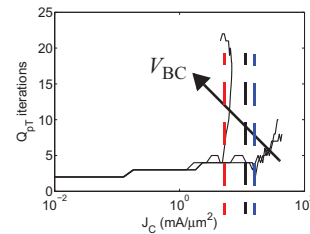


Fig. 2 Number of iterations for the internal loop of HICUM/L2 for $V_{BC} = [-0.5; 0; 0.5]$ V; vertical dashed lines (blue, black & red) indicate $J_C(f_{T,peak})$, for $V_{BC} = -0.5, 0$ & 0.5 V respectively.

B. Alternative approach without the iteration

An approach to overcome this issue was first presented in [7]. As mentioned before, with a VA-compiler the derivatives of all variables in the code with respect to the

solution variables are generated. While in standard nodal analysis, node voltages are the only solution variables, the MNA adds also the current through voltage sources to the set of solution variables. Thus calculating the weighted hole charge from such a current, the required derivatives are generated directly from the model compiler.

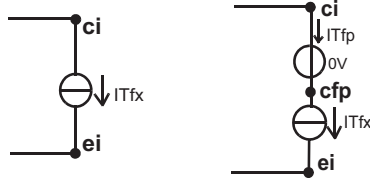


Fig. 3 Equivalent circuit for the transfer current sources when removing the internal loop from HICUM/L2.

Fig. 3 provides the equivalent circuit when removing the internal iteration loop from HICUM/L2. A voltage source is added in series to the controlled current source of the transfer current. For the time being, only the forward transfer current component is considered. In the compact model formulation, the weighted hole charge (3) is calculated from the current of the new voltage source rather than from the internal iteration variable. Since the voltage source is connected in series to the current source, the currents are forced to be the same by Kirchhoff's law. By using this manipulation of the equivalent circuit, the internal iteration is removed from the model and (4) is solved at the circuit iteration level. This significantly reduces the executed and compiled code by around 39%.

Results for the implementation of the removed internal loop are given in Fig. 4 for standard characteristics and usual operating regions.

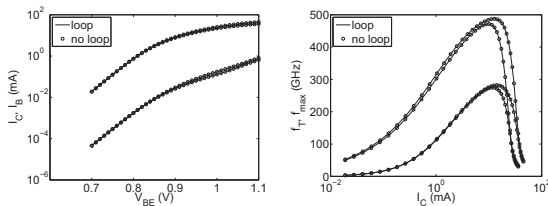


Fig. 4 Forward Gummel curves and f_T and f_{max} for $V_{BC} = [0; -0.5]$ V simulated with the original model and the model without internal iteration.

Interestingly, for the given example, the number of circuit iterations did almost not change as shown in Fig. 5. Thus, the required iterations due to series resistances were sufficient for also solving (5), or in other words this equation has no bad influence on the circuit convergence. Hence, the number of mathematical operations performed before reaching the operating point is also significantly reduced.

However, from the MNA approach, using voltage sources not only adds the additional node *cfp* (cf. Fig. 3) but also an additional solution variable of this branch to the system of equations. Thus, for each instance of the model in a circuit, the order of the equation system (i.e.

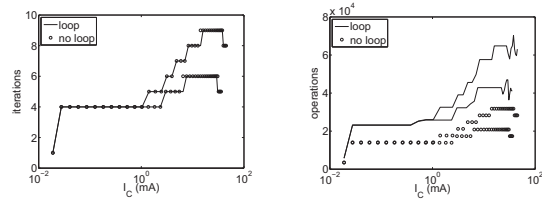


Fig. 5 Number of circuit iterations and performed mathematical operations in the same conditions as in Fig. 1.

the number of independent variables) is now increased to 19 instead of 15¹ as one additional node and corresponding solution variable (current) is needed for each of the forward and reverse transfer current. Thus, even for highly optimized (sparse) matrix solvers, the computational cost for one circuit iteration increases by roughly 30% (assuming most of the nodes are connected to transistors).

The other drawback is that the solution control parameters for the circuit equations needs to be adapted carefully in order to avoid numerical issues. This is demonstrated in Fig. 6. Results for a 21 stage ring-oscillator [8] are given comparing the performance of the model with and the one without internal iteration. While for appropriately chosen small time-steps (a) both models provide the same results, reducing the time-steps (b) can lead to numerical artifacts. For larger tail-currents, these artifacts can strongly impact the results (c) or even lead to a complete failure of the simulation (d). Note, in all cases the original model succeeded and the issue is not related to NQS effects, i.e. it occurred with and without using the corresponding models. While the simulation parameters can be adapted to provide results with the new model approach, backward compatibility is not ensured. Overall, this approach cannot be included into the model release

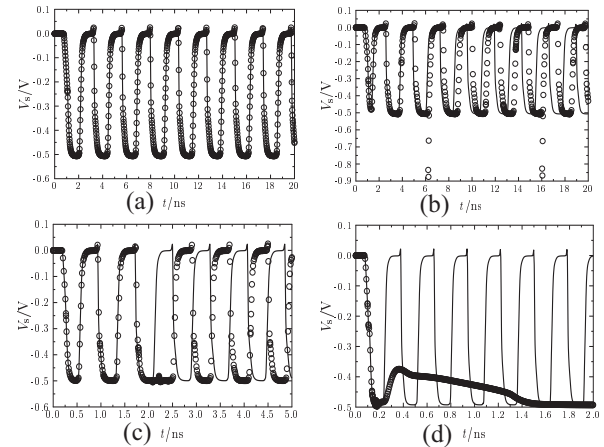


Fig. 6 Ring-oscillator simulation results. In all plots, solid lines represent the original model and symbols the model without internal loop. (a) and (b) represent the same tail current (300 μ A) with (a) 1000 time-steps and (b) 500 time-steps. (c) with $I_0 = 1$ mA and (d) with $I_0 = 3$ mA are given with 300 and 400 time-steps, respectively.

1. Including terminals, internal nodes and the thermal node.

This comparison finally shows, that promising model approaches may not always lead to the desired results. Therefore, careful testing is required also with more complicated test cases.

IV NQS Effect

For high frequencies or fast switching processes, the minority charge (Q_f) and the transfer current (I_T) are reacting delayed w.r.t the voltage across the BE junction; this is known as vertical¹ non-quasi static (NQS) effect [1]. The main task of the compact model is to provide a consistent formulation for the time and frequency domain simulation, which requires theoretical treatment of the simulation of continuity and transport equation. Assuming small-signal operation, the 2nd-order differential equation can be solved in one dimension and in frequency domain. Truncating the solution after the 2nd frequency term reads,

$$\underline{X}_{nqs} = X_{qs}(1 - (\omega\tau_2)^2 - j\omega\tau_1) \quad (6)$$

with ($\underline{X} = I_T, Q_f$), which is valid at each bias point.

Converting (6) to a slightly different form (with $s=j\omega$)

$$\frac{\underline{X}_{nqs}(s)}{\underline{X}_{qs}} \equiv \frac{1}{1 + A_1s + A_2s^2}, \quad (7)$$

which is valid for not too high frequencies, allows to apply Laplace transformation in order to obtain the corresponding time-domain approximation,

$$A_2 \frac{d^2 x_{nqs}}{dt^2} + A_1 \frac{dx_{nqs}}{dt} + x_{nqs} = x_{qs}(t), \quad x=(i_T, Q_f) \quad (8)$$

For general large-signal operation usually no analytical solution exists. However, observation from device simulation and measurements at medium and high-current densities for both sinusoidal and pulse signals with several 100mV amplitude indicates that there is no fundamental difference in the NQS behavior. In other words, the terminal currents react delayed w.r.t. the input voltage. Hence, the 2nd-order differential equation can still be applied to accomplish the delay. In practice, transient simulation consists of a sequence of circuit equation system solutions at discrete time points. The change between time points is generally small enough to consider the theoretical solution as a sufficiently accurate representation, if just the time constants at each time (i.e. bias) point are inserted. In other words, the linearized delay transformation form (8) can be applied to each time point. The accuracy of this approach has been demonstrated numerous times by examples ranging from device simulation to production circuit design.

1. Note, the dynamic emitter current-crowding is often labeled as lateral NQS effect.

Upto HICUM/L2 v2.1, the NQS effect in I_T and Q_f was implemented according to (7) and (8) with the corresponding delay times $\tau_{Qf} = \alpha_{Qf}\tau_f$ and $\tau_{IT} = \alpha_{IT}\tau_f$, where α_{Qf} and α_{IT} are model parameters. However, a VA based implementation does not allow access to the previous internal time steps of the simulator. The alternative to realize the 2nd-order differential equations is through an adjunct LCR network or its corresponding gyrator equivalent (cf. Fig. 1). The gyrator network avoids the use of inductors, which is advantageous in certain simulators. Unfortunately, a straight forward use of the adjunct network is not possible since the bias-dependence of the delay times creates non-physical derivatives in the corresponding small-signal network.

In order to overcome this issue, the equations of the adjunct network had to be modified according to Fig. 1. Using KCL at node V_{C1} and V_{C2} , one obtains the following equations in case of a bias dependent transit time τ_f ,

$$\left(\frac{I_{T,qs} - V_{C2}}{\tau_f}\right) = \frac{d}{dt}(\alpha_{IT}V_{C1}) \quad (9)$$

and

$$\left(\frac{V_{C1} - V_{C2}}{\tau_f}\right) = \frac{d}{dt}\left(\frac{\alpha_{IT}}{3}V_{C1}\right) \quad (10)$$

Above equations effectively make the capacitive elements *bias-independent* and hence, remove the generation of undesired derivatives. The same trick has been used for the minority charge related network, where a first-order RC network (cf Fig. 1) turned out to be sufficient for describing its NQS effect. It is to be mentioned that for improving convergence in the actual VA implementation the delay capacitances are properly scaled by the low current forward transit time model parameter τ_0 . The comparison of the various implementations and cases is shown in Fig. 7.

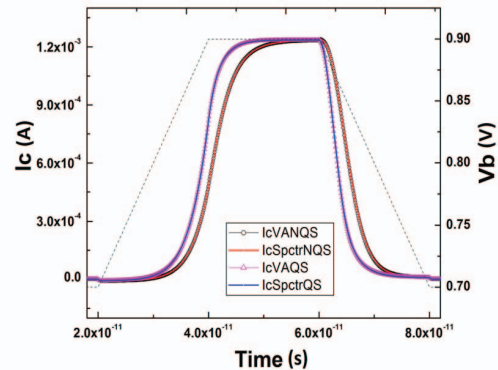


Fig. 7 Time dependence of collector current in response to an input pulse (dotted line): comparison between HICUM-VA code with NQS adjunct networks, built-in model in SPECTRE with (8). For reference, the corresponding quasi-static results are also inserted.

A. Oscillator simulations

Accurate compact models are a prerequisite for reliable circuit design. However, it is also necessary to evaluate and validate the model at the circuit level in addition to regular rigorous model testing. For the purpose of model verification, one can in principle consider any circuit block that is sensitive to model errors. One case of non-oscillatory behavior of a Colpitts oscillator circuit designed at 75 GHz using the built-in HICUM model from a foundry design library has been reported by a user.

Being actually a large signal circuit, NQS effects have been suspected to cause the above issue. The problem has been investigated using the corresponding process model card and HICUM/L2 v2.32. The basic Colpitts oscillator circuit [8], with both common-collector (CC) and common-emitter (CE) configuration has been used for this investigation. The tank circuit elements were carefully designed to ensure oscillation at 75GHz.

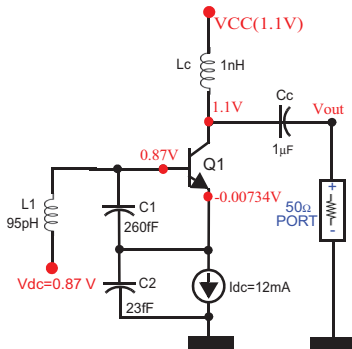


Fig. 8 Common-collector (CC) Colpitts Oscillator

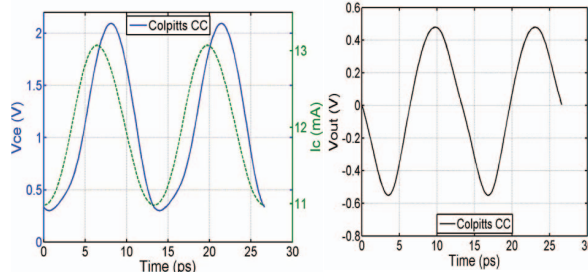


Fig. 9 Time domain waveform of simple CC-Colpitts oscillator

Testing of the CC-Colpitts oscillator (cf. Fig. 8 and Fig. 9) did not exhibit any issues regardless of enabling NQS effects or not.

However, in CE mode (cf. Fig. 10) the circuit oscillates only if the NQS-model is turned-off. Fig. 11 shows the time domain waveform of the oscillator in CE mode without NQS effect.

The above investigations show that the oscillation start-up with NQS effects appears to depend on the circuit configuration. Built-in models in design-kits are basically translated C-code from VA, which is then further optimized to fit the actual model integration interface of each simulator.

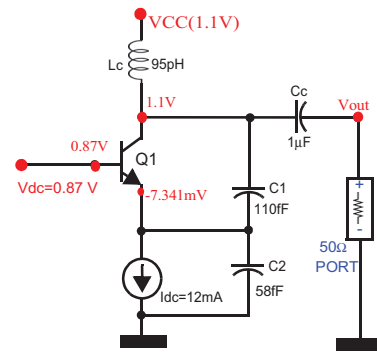


Fig. 10 Common-emitter (CE) Colpitts Oscillator

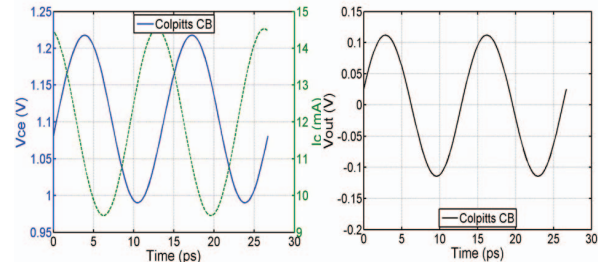


Fig. 11 Time domain waveform of simple CE-Colpitts oscillator

This sometimes creates implementation issues for a particular simulator. Hence, the oscillator issue requires EDA vendors' attention too, as a comparison of the results from VA implementation of NQS effects vs. those of the direct coding may shed some light on actual reason.

V Runtime comparison

Normally, simulators capable of handling VA code also come with an integrated VA-model compiler. These compilers take high-level (VA) code as input, translate it to C code, adapt it to the different simulator interfaces, and then compile the C code into a DLL. This allows any compact model to be linked to any of the (supported) circuit simulators. However, as different simulators use different model compilers, the execution time in practice is different. The most reliable way to compare the model run time is to use suitable benchmark circuits.

Table 5: Model run time comparison

HICUM/L2v2.33 with SH & NQS effect ON	
CPU Time	
Simulator A	32min 39 sec
Simulator B	2min 13 sec

A CML ring oscillator with 21 gates has been selected for such purpose. The current density is chosen corresponding to peak transit frequency (f_T) of the technology

and HICUM/L2 v2.33 VA code is used for circuit simulation. The table-[5] shows the CPU time comparison between simulator “A” and simulator “B”, while considering both self-heating (SH) and NQS-effect of the model [9]

The table-[5] shows “*simulator B*” to be around 10 times faster than “*simulator A*”, although the results are the same. This clearly demonstrates that run-time related to circuits is not a model issue but depends on (i) the model compiler associated with the simulator, (ii) how the code is compiled and linked it to the model library, and (iii) how the simulator creates and solves the circuit equations.

VI Model Release Testing Procedure

Releasing a new production model version is an important event which involves significant testing. Any new effect which has been implemented must go through the specific test procedure to ensure proper implementation of new effects [10]. Testing starts with the 1D modelcard and then moves on with other modelcards including the various effects. Normally 10 different cases are considered for model testing:

- 1D transistor
- internal transistor (= 1D plus internal base resistance)
- transistor with vertical NQS effect ON
- transistor with lateral NQS effect ON
- transistor with correlated noise
- transistor with impact of substrate transistor
- transistor with substrate coupling
- complete 3D transistor
- complete 3D transistor with self-heating ON
- new effects turned off for checking backward compatibility

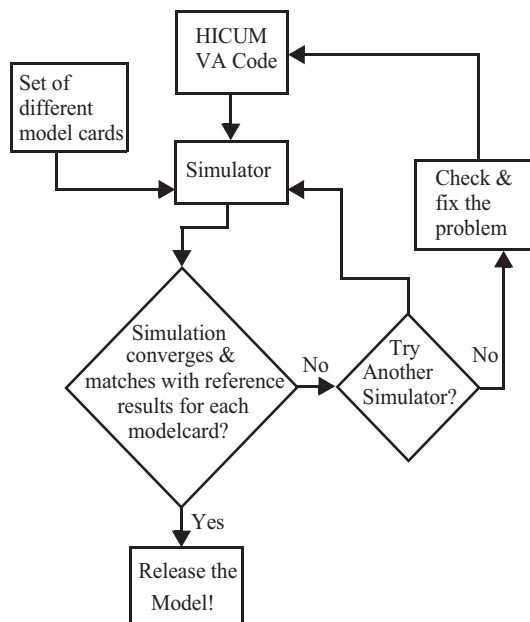


Fig. 12 Simplified model testing flow-chart.

The regular testing procedure includes several model cards with different physical effects. The flowchart shown in Fig. 12 gives a rough overview about the model testing procedure

The entire above mentioned procedure is implemented in MATLAB using the CMC QA-suite. The latest test results can be found in [11], [12], [13].

VII Conclusions

Careful testing of new model versions is required. As shown in the example regarding the internal iteration, terminal results in standard test simulations may show very good results, while more elaborate tests in actual circuits may show significant issues preventing a new formulation or approach to be implemented.

References

- [1] M. Schroter, A. Chakravorty, *Compact Hierarchical Bipolar Transistor Modeling with HICUM*, World Scientific Publishing Company.
- [2] T. Nardman, S. Lehmann, M. Schröter, R. Driad, “Application of HICUM/L0 to InP DHBTs using single-transistor parameter extraction”, Proc. IPRM, May 22-26, Berlin, Germany, 2011.
- [3] H.-M. Rein, H. Stübing, and M. Schröter, “Verification of the Integral Charge-Control Relation for high-speed bipolar transistors at high current densities”, IEEE Trans. Electron Dev., Vol. 32, pp. 1070-1076, 1985.
- [4] M. Schröter, M. Friedrich, and H.-M. Rein, “A generalized Integral Charge-Control Relation and its application to compact models for silicon based HBTs”, IEEE Trans. Electron Dev., Vol. 40, pp. 2036-2046, 1993.
- [5] H.K. Gummel and H.C. Poon, “An Integral Charge-Control Model for Bipolar Transistors”, BSTJ, Vol. 49, 1970, pp. 827-852.
- [6] P. Antognetti and G. Massobrio, “Semiconductor Device Modeling with SPICE”, McGraw-Hill, 1988.
- [7] Z. Huszka, E. Seebacher, “Removing the Internal Iteration Loop from Hicm/L2”, 9th HICUM Workshop, Germany, 2009.
- [8] S. Voinigescu, “*High-Frequency Integrated Circuits*”, The Cambridge RF and Microwave Engineering Series
- [9] Didier Céli, Jean Remy, Salim El Ghoul, “HICUM/L2 runtime evaluation with and without internal iteration”, Report CMC Meeting, 2Q-2015.
- [10] M. Schroter, A. Pawlak, A. Mukherjee, “HICUM/L2: A geometry scalable physics-based compact bipolar transistor model”, available on https://www.iee.et.tu-dresden.de/iee/eb/hic_new/hic_doc.html
- [11] A. Pawlak, M. Schröter, J. Krause, D. Céli, N. Derrier, “HICUM/L2 v2.3 Parameter Extraction for Advanced SiGe-Heterojunction Bipolar Transistors”, in Proc. IEEE BCTM, 2011, pp. 195 - 198, 2011.
- [12] A. Pawlak, M. Schröter, A. Fox, “Geometry scalable model parameter extraction for mm-wave SiGe-heterojunction transistors”, in Proc. IEEE BCTM, pp. 127-130, 2013.
- [13] A. Pawlak, M. Schröter, “An Improved Transfer Current Model for RF and mm-Wave SiGe(C) Heterojunction Bipolar Transistors”, IEEE Trans. Electron Dev., Vol. 61, No. 8, pp. 2612-2618, 2014.