Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 16/09

Bratislav Tasić, Jos J. Dohmen, Rick Janssen, E. Jan W. ter Maten, Theo G.J. Beelen, Roland Pulch

Fast Time-Domain Simulation for Reliable Fault Detection

March 2016

http://www.math.uni-wuppertal.de
Fast Time-Domain Simulation for Reliable Fault Detection

Bratislav Tasić, Jos J. Dohmen, Rick Janssen
NXP Semiconductors
5656 AE Eindhoven, the Netherlands
{bratislav.tasic,jos.j.dohmen, rick.janssen}@nxp.com

E. Jan W. ter Maten
Bergische Universität Wuppertal
42119 Wuppertal, Germany
termaten@math.uni-wuppertal.de

Theo G.J. Beelen
Eindhoven University of Technology
5600 MB Eindhoven, the Netherlands
t.g.j.beelen@tue.nl

Roland Pulch
Ernst-Moritz-Arndt-Universität Greifswald
17487 Greifswald, Germany
roland.pulch@uni-greifswald.de

Abstract—Imperfections in manufacturing processes may cause unwanted connections (faults) that are added to the nominal, “golden”, design of an electronic circuit. By fault simulation we simulate all situations: a huge number of new connections each with many different values, up to the regime of large deviations, for the newly added element. We also consider “opens” (broken connections). A strategy is developed to efficiently simulate the faulty solutions until their moment of detection. We fully exploit the hierarchical structure of the circuit. Fast fault simulation is achieved in which the golden solution and all faulty solutions are calculated over the same time step.

I. INTRODUCTION

Designs in nanoelectronics often lead to large-size nonlinear simulation problems. Here DC simulation followed by time domain simulation are the kernel steps in the design process. Industry demands the provisions of variability to guarantee quality and yield. This can be done by carefully setting up Monte Carlo simulations. To guarantee yield, i.e., estimating the probability of failures, even special Monte Carlo Methods have been designed [1], [20]. One has to simulate in the tail of the probability density functions. When one thinks about variation of parameters, this is in the regime of large deviations. One may also speak of 6-σ variations (assuming a normal density distribution). We have used these techniques to guarantee yield under harsh conditions for Static Random Access Memories (SRAMs). An SRAM is used as a building block for the construction of large Integrated Circuits (ICs). To ensure that a digital bit cell in SRAM does not degrade the yield (fraction of functional devices) of ICs with Megabits of memory, very small failure probabilities $P_{\text{fail}} \leq 10^{-10}$ are necessary. To simulate such probabilities, regular Monte-Carlo simulations require too much computing time, so more advanced techniques are required. Importance Sampling is a technique that provides sufficiently accurate results and is relatively easy to implement [12], [13], [18]. Using this, a speed-up of several orders can be achieved when compared to regular Monte Carlo methods [1], [5]. And, more important, we were able to optimise the SRAM active column, while guaranteeing that the failure probability remained below $10^{-10}$ by imposing a stochastic constraint [6], [20].

A refinement here is the use of a response surface model technique [21], which can reduce the cpu costs of evaluating many Monte Carlo samples. A new class in this area is provided by recent techniques from Polynomial Chaos Expansions (PCEs), where the expansions effectively provide a response surface facility [14], [29]. The PCEs are most accurate where the involved probability density functions are larger than some threshold [24]. Hence, when applying these techniques to estimating very small tail probabilities a careful mix of new evaluations and results obtained from evaluating the response surface model has to be taken into account [15], [16].

After having applied all these kinds of simulations and making the layout design, the manufacturing process still may introduce a completely new effect. Here we do not mean the effect of geometrical inaccuracies in widths and lengths of several components: these effects can be analysed in advance by proper techniques from sensitivity analyses. Here especially, the PCEs offer great benefit in automatically providing mean and standard variation of each local quantity at any moment of time [14], [29]. The new phenomenon comes from imperfections in manufacturing processes that may cause unwanted connections (faults). They can downgrade the performance of the carefully designed circuit. To analyse their impact we add artificial, faulty elements to the nominal, "golden", design of an electronic circuit. By fault simulation we simulate all situations: a huge number of new connections and each with many different (resistance) values, up to the regime of large deviations. We also consider "opens" (broken connections). A strategy is developed to efficiently simulate the faulty solutions until their moment of detection: i.e., when at the measurement time moment and location the difference between the faulty solution and the golden solution exceeds a certain threshold. We store the results in a database. This database is of help to first externally diagnose a faulty IC and to identify the candidate circuit submodels where the
fault may have happened. After that the IC can be studied further internally. This can help to improve next productions. Moreover, the collection of simulations can also be helpful as a priori check before layouting. Note that essentially we are looking to the weak spots in the circuit. In our approach the manufacturing process is the immediate cause of the problem. However it can also show up later, due to effects of ageing of the design, or by stress effects due to heating. It is also related to other network problems, e.g., in analysing traffic behaviour in a city when suddenly a road is blocked, or when a new connection pops up. Our approach can be extended to energy distribution networks, sewage systems, and even to networks that are not constant of size in time.

In the following sections we briefly summarise the main algorithmic steps in time domain circuit simulation. These will help to understand how Fast Fault Simulation can be done.

II. TIME INTEGRATION OF CIRCUIT EQUATIONS

The electronic circuit equations can be written as [8], [9]

$$\frac{d}{dt} q(x) + j(x) = s(x, t).$$

(1)

Here $s(x, t)$ represents the specifications of the sources. The unknown $x = x(t)$ consists of nodal voltages and of currents through voltage defined elements. We assume that $q(0) = 0$, and $j(0) = 0$. At $t = 0$, the initial solution is given by the DC solution, $x(0) = x_{DC}$, which satisfies $j(x_{DC}) = s(x_{DC}, 0)$. For time integration in circuit simulation we consider the BDF1, or Euler Backward method. Assuming time points $t_{k+1} = t_k + h_k$ ($k \geq 0$) with stepsizes $h_k$ and approximation $x^n$ at $t_n$, BDF1 calculates $x^{n+1}$ by

$$\frac{q^{n+1} - q^n}{h_n} + j^{n+1} = s^{n+1}.$$  

(2)

Here $q^k = q(x^k)$, $j^k = j(x^k)$, for $k = n, n + 1$, and $s^{n+1} = s(x^{n+1}, t_{n+1})$. The system is solved by a Newton-Raphson procedure. For efficient direct methods to solve the intermediate linear systems, see [3]. A fixed Jacobian can reduce the number of LU-decompositions, but, in general, it will increase the number of iterations and thus the number of (costly) evaluations. Also, in the case of circuit simulation, the assembly of the matrices does not need much more effort when compared to the function evaluations. However, a fixed decomposed Jacobian can be efficient within some Picard-iteration [17] in solving a linear system, or, more general, in using it as preconditioner within GMRES. When changing stepsizes during time integration similar remarks apply.

In the case of an hierarchical linear solver one can profit from hierarchical bypassing [8], which we will also exploit in this paper. When applying it also in the time integration, it even supports a first form of multirate time-integration [28].

Fig. 1. Examples of disturbing bridges. Connections with small resistors values can lead to large defects to the solution.

III. FAULT SIMULATION

We first consider the effect of adding faulty, linear elements to the circuit. E.g., in [4], [26], [27] we added linear bridges (resistors) to the circuit. For each fault, only one element is added to the original, golden circuit. It may mean a new connection, while also different values are considered, see Fig. 1.

In [27] a novel time-integration was introduced where all problems were integrated over the same time step: first the fault-less, golden solution was determined at the next time step. Next, all faulty problems were integrated, see Fig. 2. Hence, effectively, a parameter loop is placed inside the time integration.

Fig. 2. At each time step, first the golden solution is determined, next all faulty problems are integrated.

The hierarchical structure was enhanced, such that the hierarchical solver could deal with all new elements: note that some new connections may violate the hierarchical structure of the golden circuit. A clever software solution to this was developed and is reported in [27]. By this, also the faulty problems could benefit from an enhanced form of hierarchical bypassing [8].

The golden solution at each new time point provides an estimate for the solution of a faulty problem (in addition to the one using extrapolation by Nordsieck vectors, see [8] and references cited there). Each faulty problem uses the stepsize of the golden solution as a maximum one. When the faulty solution really needs a stepsize that is significantly smaller than used by the golden solution, the traditional time integration is invoked, even without bypassing, until the time moment of synchronization with the golden solution. Time integration of a faulty problem is only needed until the moment when its faulty solution really differs from the golden solution. This check is done only at specific time moments. Then the
faulty problem is marked as detectable and taken out of the list and its results are stored. Time integration is continued on the next time intervals for both the golden solution and the reduced list of faulty problems, until the next moment for checking deviations.

As in [27] we consider the faulty elements represented by adding linear conductivities to the circuit\(^1\). In [26] also the effect of additional linear capacitors was considered. However, the main interest is in adding linear conductivities.

\[ j(x(t,p), p) = j_0(x(t,p)) + p u^T x(t,p), \]  

(3)

where \( p = 1/R \), with resistance \( R \). Fault Analysis consists of simulations for a large number of pairs of vectors \((u, v)\) and various values of \( p \), and comparing the result \( x(t, p) \) of (1) at specific time points with the "golden" solution \( x(t) = x(t, 0, 0) \) of the fault-free circuit, corresponding with \( p = 0 \). The golden solution \( x(t) = x(t, 0, 0) \) uses \( j(x(t, 0, 0)) = j_0(x(t, 0)) \). If the deviation between \( x(t, p) \) and \( x(t) \) exceeds some threshold, the fault triple (\( u, v, p \)), is marked as detectable and taken out of the list.

Clearly, for each fault, we have a new contribution \( p u^T x(t,p) \) as low-rank modification to the system of the golden solution, added to \( j_0 \), see (3). Here \( p = 1/R \geq 0 \) is just a scalar, by which the \( p \)-sensitivity ‘matrix’ \( \dot{x}_p(t,p) = \frac{\partial x(t,p)}{\partial p} \) reduces to a vector. Let \( x^p_k = x^p(k) \approx x(k, p) \), \( k = n, n + 1 \), be the numerical approximations of the faulty system and \( \dot{x}_p^k \) be the corresponding sensitivities. Then with \( C^k_p = \frac{\partial q(x^p_k)}{\partial x} \), \( G^k_p = \frac{\partial \dot{x}_p^k}{\partial x} \) (including the effect of the rank-one term with the factor \( p \)) and \( S^p_p = \frac{\partial n(x^p_k, t_k)}{\partial x} \), we obtain by sensitivity analysis [27]

\[ A^{n+1}_p x^{n+1} = -u v^T x^{n+1} + \frac{1}{h_n} C^p_n x^p_n, \]  

(4)

\[ A^{n+1}_p = \frac{1}{h_n} C^{n+1}_p + G^{n+1}_p - S^{n+1}_p. \]  

(5)

For \( p = 0 \), (4) gives the limit sensitivity \( \dot{x}_p^k = \ddot{x}_0^k \) for the golden, fault-free solution \( x^k = x^k_0 \) for \( k = n, n + 1 \)

\[ A^n x^n = -u v^T x^n + \frac{1}{h_n} C^n x^n, \]  

(6)

\[ A^n = \frac{1}{h_n} C^{n+1} + G^{n+1} - S^{n+1}, \]  

(7)

where \( C^k = C^k_0 \) \((-n, n + 1)\), \( G^{n+1} = G^{n+1}_0 \) and \( S^{n+1} = S^{n+1}_0 \) and thus \( A^{n+1} = A^{n+1}_0 \). By Taylor expansion we additionally have

\[ x^p_k = x^k + p \dot{x}^k + \mathcal{O}(p^2) \quad (k = n, n + 1). \]  

(8)

The golden solution \( \ddot{x}^k \) satisfies the linearized equations of the fault-free circuit up to a term \( R \) that indicates the deviation from linearity (note that in (1) we did assume that \( q(0) = 0 \) and \( j(0) = 0 \))

\[ A^{n+1} x^{n+1} = r(t_{n+1}, x^n, x^{n+1}), \]  

(9)

where \( r(t_{n+1}, x^n, x^{n+1}) = s^{n+1} + \frac{1}{h_n} C^n x^n + R, \) in which \( s^{n+1} = s(x^{n+1}, t_{n+1}) \). With (8) and (6) this gives [26], [27]

\[ \begin{aligned} \left[ A^{n+1} + p u v^T \right] x^p_n &= r(t_{n+1}, x^n, x^{n+1}) + \frac{1}{h_n} C^n (x^n_p - x^n) + \mathcal{O}(p^2 + \frac{p^2}{h_n}) \quad (10) \\
 &= r(t_{n+1}, x^n, x^{n+1}) + \mathcal{O}(p^2 + \frac{p^2}{h_n} + \frac{p}{h_n}). \end{aligned} \]  

(11)

This invites for applying the Sherman-Morrison formula [10]. Let \( A_w = p u \) and \( A_y = C^n x^n \). Then the predictions for \( x^{n+1} \) become

\[ x^{n+1}_p = x^{n+1} - \frac{v^T (x^{n+1} + y)}{1 + v^T w} w, \]  

(12)

\[ x^{n+1} = x^{n+1} + y - \frac{v^T}{1 + v^T w} w. \]  

(13)

Here (12) is based on (11), which implicitly uses arguments from sensitivity analysis. The prediction (13) is based on (10), which explicitly determines the effects of the sensitivity quantities by solving for \( y \). Note that (10) may be a more accurate alternative than (11). However, for simplicity, we just used (11), after ignoring the \( \mathcal{O}(\cdot) \) terms at the right-hand side. For expressions when dealing with capacitors we refer to [26]. The advantage of the right-hand side in (11) is that it is independent of the solution \( x^k_0 \) at the previous time steps. Of course, when followed by further Newton-Raphson iterations, \( x^n \) is still needed. To judge the accuracy of the linear sensitivity prediction the nonlinear solver evaluates the circuit at the sensitivity solution and updates the solution. The difference in the initial sensitivity solution and the nonlinear update is a measure for the truncation error.

If we just stick to the prediction, we may calculate the prediction of the fault at a few selected time points, which significantly reduces the work load for the fault sensitivity analysis.

Finally, we shortly describe the modeling of faulty "opens". We consider a faulty resistor, with value \( R \), in series with another, linear resistor, with value \( r \), where \( r \) is a correct resistor in the golden circuit. Clearly, this introduces an extra node \( n_e \). If the golden system used \( R(n_1, n_2) = r \), the faulty system uses \( R(n_1, n_e) = R, \) \( R(n_e, n_2) = r \). The voltage at this new node \( n_e \) can be simply eliminated by noting that \( v(n_e) = (r v(n_1) + R v(n_2))/(r + R) \). Doing this directly, the remaining system can be formulated as in (3) in which \( p = R/(r + R) \). If \( R \rightarrow \infty \) we obtain an "open" between the nodes \( n_1 \) and \( n_e \) and \( v(n_e) \rightarrow v(n_2) \). In [27] we introduced an extra port to model bridges between models. This extra node can also become functional in providing the extra node.

For modeling a broken joint (or weld) at a node \( n \), it is, mathematically, convenient to first split the node \( n \) into two nodes \( n_1 \) and \( n_2 \), with a simple voltage source in between for the golden circuit: \( E(n_1, n_2) = 0 \). Clearly this satisfies our assumption \( j(0) = 0 \). The faulty system uses \( R(n_1, n_e) = R, \)

\[ A^{n+1}_p x^{n+1} = r(t_{n+1}, x^n, x^{n+1}), \]  

(9)

\[ \text{Note that for voltage-defined resistors we need two rank-one updates to describe the total contribution [8].} \]
$E(n_1, n_2) = 0$. We assume local coordinates that correspond with $v(n_1), v(n_2), i(E)$. We deduce that the faulty system perturbs the golden system with $u_i v_i^T (R) + u_2 v_2^T (R)$, in which, in local coordinates, $u_1^T = (1, 0, 0), v_1^T = (-1/R, 1/R, 1)$, $u_2^T = (0, 0, 1), v_2^T = (1, -1, R)$. This can be treated in a similar way as before.

IV. RESULTS

The algorithm for Fast Fault Simulation has been implemented in Pstar\(^2\). For fault detections in DC-simulations a significant speed-up (> 100) was obtained by exploiting bypassing and abandoning only, but inclusion of sensitivity analysis appeared essential to get significant speed-up for a broad class of problems during transient simulation. Table I shows the speed-up by including sensitivity prediction for a LIN Converter IP Block (first part), as well as for a nonlinear control DAC (second part). Clearly, the linear sensitivity estimate offers an interesting speed-up. The application of nonlinear corrections by Newton iterations reduces this effect. For the LIN Converter IP Block the effect of more iterations remains quite bounded (with 100 iterations still a speed-up of more than 10 was found, see [27]). For the nonlinear control DAC until 5 iterations a speed-up of 10 was obtained.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>LIN Converter IP Block</th>
<th>#iterations per step $\Delta t$</th>
<th>CPU Time [sec]</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard AS/DOTSS</td>
<td></td>
<td>-</td>
<td>100437</td>
<td>1</td>
</tr>
<tr>
<td>Linear Sensitivity</td>
<td></td>
<td>0</td>
<td>458</td>
<td>219</td>
</tr>
<tr>
<td>Nonlinear Correction</td>
<td></td>
<td>5</td>
<td>2341</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Control DAC</th>
<th>#iterations per step $\Delta t$</th>
<th>CPU Time [sec]</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard AS/DOTSS</td>
<td></td>
<td>-</td>
<td>52513</td>
<td>1</td>
</tr>
<tr>
<td>Linear Sensitivity</td>
<td></td>
<td>0</td>
<td>916</td>
<td>58</td>
</tr>
<tr>
<td>Nonlinear Correction</td>
<td></td>
<td>1</td>
<td>4808</td>
<td>11</td>
</tr>
</tbody>
</table>

TABLE I

**Speed-up by including sensitivity prediction. Top: a LIN Converter IP Block, #faults=412. Bottom: a Control DAC, #faults=100. See also [27].**

Currently, NXP Semiconductors’ Fast Fault Algorithm is the best in the world in this area [27]. It can identify locations on a chip that are probably affected by very tiny manufacturing inaccuracies and thus causing faulty behaviour at predefined time points for measurements. Fig. 3 gives further results.

Further speed-up scenarios are currently considered by initiating the fault later, see Fig. 4. If one can simply skip the initial integration of the faults until some $t_1 > 0$, for a large collection of faults no initial simulations have to be made. By this, additional orders of speed-up can be achieved. The scenarios differ in how the fault is started: suddenly, or using a smooth start-up, similar as for the source-stepping-by-transient method as described in [27]. Because of the many faults that are possible, a short start-up is a balance between efficiency

---

2Pstar: in-house circuit simulator of NXP Semiconductors.
Fig. 5. DC Monte Carlo simulations for a TJA1050 High-Speed CAN Transceiver. In blue the original nr of Newton iterations per independent Monte Carlo run is shown. In red we see the reduced nr of iterations due to better predictions obtained with the techniques described before. For both approaches we obtained the same results of the output quantities.

$v_i(t) = \sum_{k=1}^{K} w_k x(t, p_k) \phi_i(p_k)$, where the quadrature formula provides an approximation. This expansion is a so-called generalized Polynomial Chaos expansion, using polynomials $\phi_i(p)$ that are orthogonal with respect to some probability density function in the parameter space for $p$. A library like [2] provides software when using Gauss-Hermite and Laguerre polynomials. For Fast Fault Simulation, where parameter values are positive, one may think about an exponential decay (for an infinite range, see Fig. 1: here one generates Laguerre polynomials), or a beta density distribution function with shape parameters $(\alpha, \beta)$ (when considering a finite range for $p$; here one generates Jacobi-polynomials). In the traditional, non-intrusive, implementation for each value $p_k$ a complete time sweep is made. After completion of this the expansion is available and then also additional information like sensitivity. When we change the order of the loops, one can determine at the next time level the $K$ deterministic solutions (in which one can exploit the sensitivity estimate, as described before for the Fast Fault Simulation Algorithm). This variant provides sensitivity by the expansion as soon when it is needed. Clearly, this makes SC intrusive, but this variant is probably faster than the default, non-intrusive, implementation.

An alternative approach that maintains the non-intrusive character as long as possible is to use two SC sweeps: a first sweep is applied that uses a coarse quadrature rule. This is a traditional non-intrusive implementation. It is just intended to have sensitivity available for the second sweep in which a higher accurate quadrature formula is used. Now, during the time integration one has sensitivity results available from the expansion resulting as postprocessing after the first sweep and which is provided by just a call to a procedure in the UQ-Library [2]. This minimizes the implementation and make sensitivity available where it is needed. It is related to multifidelity uncertainty quantification [22], [23], [25], [30].

VI. Conclusion

We demonstrated details of a successful algorithm for Fast Fault Simulation. The speed-ups were obtained by a combination of different techniques (hierarchical simulation, bypassing, on-the-fly-reduction of the list of faults, prediction of neighbouring problems by implicit sensitivity) together with elegant enhancement of the hierarchical modeling using extra ports. Currently, NXP Semiconductors’ algorithm for Fast Fault Simulation is the best in the world in this area [27]. NXP Semiconductors can identify locations on a chip that are probably affected by very tiny manufacturing inaccuracies and thus causing faulty behaviour at predefined time points for measurements. Options to obtain further speed-ups have been identified (and have been confirmed for first tests). The algorithm indicated how we can also speed up Uncertainty Quantification by Stochastic Collocation for Polynomial Chaos expansion. This is currently further investigated.

ACKNOWLEDGMENT

This work is part of the FP7 Collaborative Project nanoCOPS, which is supported by the European Union in the FP7-ICT-2013-11 Programme under Grant Agreement Number 619166 (Project nanoCOPS - nanoelectronic COupled Problems Solutions). For further details see http://www.fp7-nanocops.eu/ and [11].

REFERENCES


