POLITECNICO DI TORINO
Repository ISTITUZIONALE

Fast Channel Simulation via Waveform Over-Relaxation

Publisher copyright

(Article begins on next page)

09 June 2023

Fast Channel Simulation via Waveform Over-Relaxation

S. Grivet-Talocia, V. Loggia
Dip. Elettronica, Politecnico di Torino
C. Duca degli Abruzzi 24, 10129 Torino, Italy
stefano.grivet@polito.it

Abstract

This paper presents a fast simulation method for long coupled multi-chip interconnects. The channel is represented through a time-domain passive delay-rational macromodel, which is identified from tabulated scattering samples. A two-level Waveform Relaxation (WR) framework is then applied in order to perform fast transient simulation of the terminated channel via an iterative process. A new over-relaxation scheme is introduced for improving the convergence of the WR iterations.

1 Introduction

The electrical verification of point-to-point links connecting different subsystems requires extensive simulation [1, 2]. This analysis is quite demanding, since multiple coupled channels must be simulated concurrently in the time domain, in order to investigate all possible sources of signal degradation including intersymbol interference, crosstalk, jitter, and nonlinear distortions induced by driver and receiver circuitry. Long bit sequences need to be analyzed for each sensitive channel, including multiple aggressor patterns to assess worst-case scenarios.

The most direct simulation approach involves a linear analysis based on convolution, combined with a statistical post-processing. This technique allows very fast and comprehensive analyses. Unfortunately, nonlinear effects of drivers and receivers are only approximated, so that the results may not be fully representative. Alternative approaches based on direct circuit-based (SPICE) simulations allow direct treatment of nonlinear transceiver models. In these simulations, suitable macromodels of the channel are first identified from native tabulated scattering responses and then synthesized as SPICE-compatible netlists. Main drawback of this approach is runtime, which can be considerable for complex channels.

In this work, we propose an alternative approach that is able to preserve SPICE (or better) accuracy, with dramatic speedup in runtime. The technique we pursue in this work is based on a two-level Waveform Relaxation (WR) [3, 8, 9, 11, 12, 13]. One relaxation loop is based on transverse partitioning, aimed at decoupling individual channels though suitable relaxation sources. A second relaxation loop is based on a longitudinal partitioning, aimed at decoupling each individual channel from its terminations. The basic scheme was first presented in [10].

The main contribution of this paper is a new formulation based on a successive over-relaxation [14, 15]. The new formulation is able to fix possible convergence issues of the basic scheme through the introduction of an over-relaxation parameter, which is tuned in a preprocessing phase in order to guarantee the best convergence rate. Although we cannot prove that the over-relaxed scheme will be able to guarantee convergence in the general case, we show its practical effectiveness on a set of industrial benchmarks.

2 Problem statement

We consider a generic chip-to-chip link with P fully-coupled electrical ports (P even), terminated by possibly nonlinear single-ended termination circuits (drivers and receivers). A graphical illustration of the system topology is depicted in the left panel of Fig. 4. The mathematical formulation of the interaction between channel and terminations can be obtained as

$$\begin{cases} \boldsymbol{b}(t) = \mathbf{h}(t) * \boldsymbol{a}(t) \\ F_q\left(a_q(t); b_q(t); t; \frac{d}{dt}\right) = 0, \quad q = 1, \ldots, P \end{cases} \quad (1)$$

where the first row represents the convolution between transient scattering waves $\boldsymbol{a}(t)$ entering the channel and the channel impulse response matrix $\mathbf{h}(t)$, and second row collects all nonlinear and dynamic equations representing drivers and receivers. For convenience, also the terminations are represented in the transient scattering form. Using a more compact operator notation, we have

$$\begin{cases} \boldsymbol{b} = \mathcal{H}\,\boldsymbol{a}, \\ \boldsymbol{a} = \mathcal{F}(\boldsymbol{b}), \end{cases} \quad (2)$$

where the linear operator $\mathcal{H}$ represents the channel, and where the explicit nonlinear operator $\mathcal{F}$ is diagonal.

We assume that a Delay-Rational Macromodel (DRM) is available for the channel [5, 4]. In the Laplace-domain, a DRM can be expressed as

$$H^{i,j}(s) = \sum_{m=0}^{M^{i,j}} Q_m^{i,j}(s) e^{-s\tau_m^{i,j}} + D^{i,j},$$

$$Q_m^{i,j}(s) = \sum_{n=1}^{N_m^{i,j}} \frac{R_{mn}^{i,j}}{s - p_{mn}^{i,j}} \quad (3)$$

where i, j denote a particular element of the scattering transfer matrix, $\tau_m^{i,j}$ are suitable delays, and $Q_m^{i,j}(s)$ are rational coefficients. The identification of (3) can be performed from tabulated scattering matrix samples $\widehat{\boldsymbol{S}}_l \in \mathbb{C}^{P \times P}$ avaiable at the discrete frequencies ω_l, $l = 1, \ldots, L$ through Delayed Vector Fitting (DVF) or the Delayed Sanathanan-Koerner (DSK) iterations, see [5, 4]. Model passivity can also be checked end enforced [6, 7]. The main advantage of DRM operators is that their time-domain application can be cast as a delayed recursive convolution, whose numerical evaluation has a cost that scales only linearly with the number of time steps to be computed [10]. Thus, the time-domain evaluation of the first row in (2) is extremely efficient.

3 Inner and Outer Waveform Relaxation

The main disadvantage of (1) or (2) is that, for each time step, there is an instantaneous coupling between channel and terminations. Within a time-stepping framework, this would
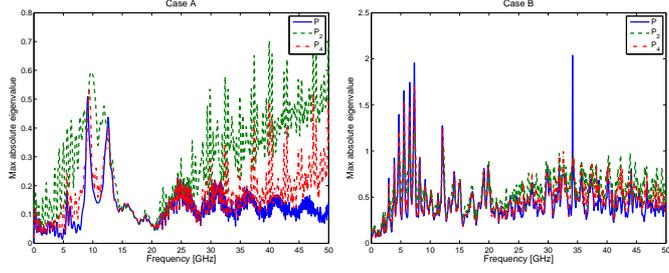
Figure 1: Spectral radius of operator $\mathbf{P}_\mathcal{I}$ for $\mathcal{I} = 2, 4, \infty$ plotted versus frequency for Cases A and B.

Figure 2: Evolution of the inner (line) and outer (dots) loop errors through iterations for case A (left) and B (right).

require the concurrent solution of the fully-coupled channel together with its terminations using a nonlinear solver. This is what standard SPICE solvers do. We use a completely different approach here. Our unknowns are vectors collecting all time samples of all port variables, rather than individual time samples. These unknowns are computed collectively through an iterative scheme that refines an initial estimate via a Waveform Relaxation (WR) approach.

We start by reviewing the two-level WR scheme presented in [10]. A first level exploits a decoupling process between different channels, by decomposing the channel operator as $\mathcal{H} = \mathcal{D} + \mathcal{C}$, where $\mathcal{D}$ collects all transfer matrix entries representing direct transmission and reflection coefficients. and $\mathcal{C}$ collects all near and far end crosstalks. A transverse partitioning and "outer" relaxation loop with index μ is obtained by delaying the application of the coupling operator $\mathcal{C}$ by one iteration. This can be expressed as

$$\begin{cases} \boldsymbol{b}_\mu = \mathcal{D}\, \boldsymbol{a}_\mu + \boldsymbol{\theta}_{\mu-1}\,, \qquad \boldsymbol{\theta}_{\mu-1} = \mathcal{C}\boldsymbol{a}_{\mu-1} \\ \boldsymbol{a}_\mu = \mathcal{F}(\boldsymbol{b}_\mu)\,, \end{cases} \tag{4}$$

to be solved for $\mu = 1, 2, \ldots$ with a suitable initial condition, e.g., $\boldsymbol{\theta}_0 = \mathbf{0}$. The evaluation of the relaxation source $\boldsymbol{\theta}_{\mu-1}$ is performed via recursive convolutions. The above outer relaxation is motivated by the small correction that is expected at each iteration as an effect of the inter-channel couplings $\mathcal{C}$, which for a well-designed link are small. This condition should also ensure fast convergence of the WR iterations.

A second "inner" relaxation loop is also introduced in order to avoid the concurrent solution of individual (decoupled) channels and their terminations. A longitudinal partition is applied in order to separate channels from terminations, and port variable estimates are refined through a longitudinal relaxation with index ν, whereas channel and termination equations perform updates to the interface variables alternating in time. Formally, this can be represented as

$$\begin{cases} \boldsymbol{b}_{\mu,\nu} = \mathcal{D}\, \boldsymbol{a}_{\mu,\nu-1} + \boldsymbol{\theta}_{\mu-1}\,, \\ \boldsymbol{a}_{\mu,\nu} = \mathcal{F}(\boldsymbol{b}_{\mu,\nu})\,, \end{cases} \tag{5}$$

to be solved for any fixed μ for $\nu = 1, \ldots, \mathcal{I}_\mu$ with a suitable initial condition, e.g., the available solution estimate at previous transverse relaxation step $\boldsymbol{a}_{\mu,0} = \boldsymbol{a}_{\mu-1,\mathcal{I}_{\mu-1}}$. Further details on this scheme, denoted as WR-LPTP (Waveform Relaxation via Longitudinal and Transverse Partitioning) are available in [10].
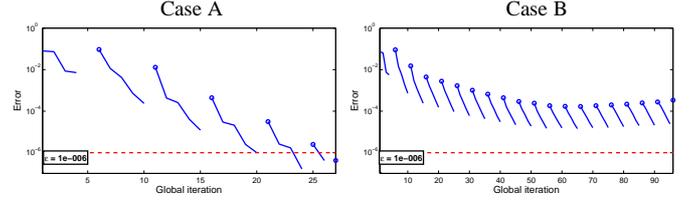
Convergence of inner and outer loops is detected by monitoring the respective residual norms

$$\xi(\mu, \nu) \;=\; ||\boldsymbol{a}_{\mu,\nu} - \boldsymbol{a}_{\mu,\nu-1}|| \tag{6}$$
$$\delta(\mu, \mathcal{I}_\mu) \;=\; ||\boldsymbol{a}_{\mu,\mathcal{I}_\mu} - \boldsymbol{a}_{\mu-1,\mathcal{I}_{\mu-1}}||\,. \tag{7}$$

When below a prescribed tolerance ϵ, these norms are used as stopping conditions for the iterations.

A more formal convergence analysis can be performed in the frequency domain by assuming linear terminations with passive scattering matrix $\mathbf{\Gamma}$, such that $||\mathbf{\Gamma D}|| < 1$. In this case, the operator mapping one iteration onto the next can be explicitly constructed as

$$\mathbf{P}_\mathcal{I} = \mathbf{P} + (\mathbf{\Gamma D})^\mathcal{I}\,(\mathbf{I} - \mathbf{P})\,, \tag{8}$$

with $\mathbf{P} = \lim_{\mathcal{I} \to \infty} \mathbf{P}_\mathcal{I} = (\mathbf{I} - \mathbf{\Gamma D})^{-1}\,(\mathbf{\Gamma C})$ and where for simplicity a constant number $\mathcal{I}$ of inner iterations are assumed, independently on the outer iteration μ. Under such hypothesis, the error of the iterative solution at the outer iteration $\mathcal{K}$ with respect to the exact solution reads

$$\boldsymbol{\mathcal{E}}_{\mathcal{K},\mathcal{I}} = \boldsymbol{A}_{\mathcal{K},\mathcal{I}} - \boldsymbol{A}_{\text{exact}} = -\mathbf{P}_\mathcal{I}^\mathcal{K}\boldsymbol{A}_{\text{exact}}\,. \tag{9}$$

As a result, a necessary and sufficient condition for convergence is the unitary boundedness of the spectral radius of the iteration operator

$$\rho_{\max}\{\mathbf{P}_\mathcal{I}\} < 1\,, \tag{10}$$

which must hold for all frequencies.

Two industrial benchmarks will be used to illustrate the proposed WR schemes. Case "A" is a fully-coupled 18-port channel providing the electrical link between CPU and an IO card. Case "B" is a similar 18-port channel connecting two boards through a connector. Both cases are courtesy of IBM. Figure 1 reports the frequency-dependent spectral radius of iteration operator $\mathbf{P}_\mathcal{I}$ for the two cases. We see that case A is expected to converge, since condition (10) holds. Case B is instead expected to have problems, since there are some frequencies for which condition (10) is violated. The results confirm these propositions. Figure 2 reports the inner and outer error estimates (6)-(7) through iterations (with $\mathcal{I} = 4$) for the two cases. The error for case A converges below the adopted stopping threshold, whereas the outer error for case B blows up. For case A, the transient results are in perfect agreement with SPICE, as Fig. 3 shows. However, a 1000-bit SPICE simulation requires about 23 minutes, whereas the same deck is solved by WR in only 31 seconds, with a $44\times$ speedup. No validation is possible for case B, since the transient waveforms are not even bounded through iterations.
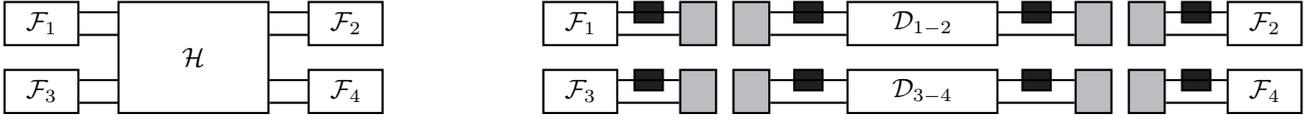
Figure 4: Graphical illustration of terminated channel (left) and WR-SOR partitioning scheme. Dark (light) gray boxes denote outer (inner) relaxation sources, respectively.

4 Successive Over-Relaxation

In Section 3, we observed that there may be cases where the standard WR-LPTP algorithm does not converge, since the spectral radius of the iteration operator exceeds one and condition (10) does not hold. We now construct a new relaxation scheme that overcomes these difficulties. This new algorithm is denoted as WR-SOR and uses a Successive Over-Relaxation technique, as described in [14, 15].

The starting point is system (2), where operator $\mathcal{H}$ is split into its (block-) diagonal part $\mathcal{D}$ and remainder $\mathcal{C}$

$$\begin{cases} \boldsymbol{b} - \mathcal{D}\boldsymbol{a} - \mathcal{C}\boldsymbol{a} &= 0 \\ \boldsymbol{a} - \mathcal{F}(\boldsymbol{b}) &= 0 \,. \end{cases} \tag{11}$$

We multiply both equations by an over-relaxation parameter η (to be determined) and we add $(\boldsymbol{b} - \mathcal{D}\boldsymbol{a})$ and $(\boldsymbol{a} - \mathcal{F}(\boldsymbol{b}))$ to both sides of first and second equation, respectively, obtaining

$$\begin{cases} \eta\boldsymbol{b} - \eta\mathcal{D}\boldsymbol{a} - \eta\mathcal{C}\boldsymbol{a} + \boldsymbol{b} - \mathcal{D}\boldsymbol{a} &= \boldsymbol{b} - \mathcal{D}\boldsymbol{a} \\ \eta\boldsymbol{a} - \eta\mathcal{F}(\boldsymbol{b}) + \boldsymbol{a} - \mathcal{F}(\boldsymbol{b}) &= \boldsymbol{a} - \mathcal{F}(\boldsymbol{b}) \,. \end{cases} \tag{12}$$

Rearranging the various terms and introducing the outer relaxation index μ, we obtain

$$\begin{cases} \boldsymbol{b}_\mu &= \mathcal{D}\boldsymbol{a}_\mu + \boldsymbol{\theta}_{\mu-1} \\ \boldsymbol{a}_\mu &= \mathcal{F}(\boldsymbol{b}_\mu) + \boldsymbol{\varphi}_{\mu-1} \,, \end{cases} \tag{13}$$

where $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\varphi}_\mu$ are the outer relaxation sources, defined as

$$\begin{cases} \boldsymbol{\theta}_\mu &= (1-\eta)(\boldsymbol{b}_\mu - \mathcal{D}\boldsymbol{a}_\mu) + \eta\mathcal{C}\boldsymbol{a}_\mu \,, \\ \boldsymbol{\varphi}_\mu &= (1-\eta)(\boldsymbol{a}_\mu - \mathcal{F}(\boldsymbol{b}_\mu)) \,. \end{cases} \tag{14}$$

System (13) is a generalization of (4), which can be obtained by setting $\eta = 1$ in (13)-(14). The introduction of the overrelaxation adds an outer relaxation source $\boldsymbol{\varphi}_\mu$ also to the equation corresponding to the channel terminations, causing a negligible

Figure 3: Case A, comparison between WR and SPICE results.

overhead both in computation and storage. The scheme is consistent, since setting $\mu \to \infty$ leads to the original system (2).

At any outer iteration, the actual solution is performed by applying a second level of (longitudinal) partitioning, as for the WR-LPTP scheme, with suitable relaxation sources and inner iteration index ν. The resulting WR-SOR scheme reads

$$\begin{cases} \boldsymbol{b}_{\mu,\nu} = \mathcal{D}\boldsymbol{a}_{\mu,\nu-1} + \boldsymbol{\theta}_{\mu-1} \,, \\ \boldsymbol{a}_{\mu,\nu} = \mathcal{F}(\boldsymbol{b}_{\mu,\nu}) + \boldsymbol{\varphi}_{\mu-1} \,, \end{cases} \tag{15}$$

The right panel of Fig. 4 provides a graphical illustration of this scheme. At the initialization stage $\mu = 0$ we set

$$\begin{cases} \boldsymbol{\theta}_0 &= 0 \\ \boldsymbol{\varphi}_0 &= \mathcal{F}(\boldsymbol{0}) \,. \end{cases} \tag{16}$$

The convergence of inner and outer loops is detected as for the WR-LPTP algorithm, by monitoring the respective residual norm estimates $\xi_{\mu,\nu}$ and δ_μ as in (6)–(7).

The convergence of the WR-SOR scheme can be assessed by a frequency-domain analysis as for the WR-LPTP scheme. The resulting expression for the iteration operator is

$$\mathbf{P}_{\mathcal{I},\eta} = \mathbf{I} - \eta \left[\mathbf{I} - (\mathbf{\Gamma}\mathbf{D})^{\mathcal{I}} \right] (\mathbf{I} - \mathbf{P}) \,. \tag{17}$$

The error of the iterative solution with respect to the exact solution can also be formally computed as

$$\boldsymbol{\mathcal{E}}_{\mathcal{K},\mathcal{I}} = \boldsymbol{A}_{\mathcal{K},\mathcal{I}} - \boldsymbol{A}_{\text{exact}} = -\eta\mathbf{P}_{\mathcal{I},\eta}^{\mathcal{K}}\boldsymbol{A}_{\text{exact}} \tag{18}$$

Therefore, convergence and consistency of the WR-SOR scheme are guaranteed if $\rho_{\max}\{\mathbf{P}_{\mathcal{I},\eta}\} < 1$.

Operator $\mathbf{P}_{\mathcal{I},\eta}$ is parameterized by η. It is thus possible to find a value η_{opt} such that the spectral radius of $\mathbf{P}_{\mathcal{I},\eta}$ is minimized and the convergence rate is optimal. We first provide the feasibility conditions for this optimization. For any fixed frequency ω, let us denote

$$\lambda_q(\omega) \in \text{eig}\{\mathbf{\Lambda}(\omega)\} \quad q = 1, \dots, P \tag{19}$$

the generic eigenvalue of matrix

$$\mathbf{\Lambda} = \left\{ \left[\mathbf{I} - (\mathbf{\Gamma}\mathbf{D})^{\mathcal{I}} \right] (\mathbf{I} - \mathbf{P}) \right\} \tag{20}$$

such that $\mathbf{P}_{\mathcal{I},\eta} = \mathbf{I} - \eta\mathbf{\Lambda}$ from (17). The WR-SOR scheme will converge if

$$|1 - \eta\lambda_q(\omega)| < 1 \quad \forall\omega, q \,. \tag{21}$$

A straightforward derivation leads to the feasibility condition

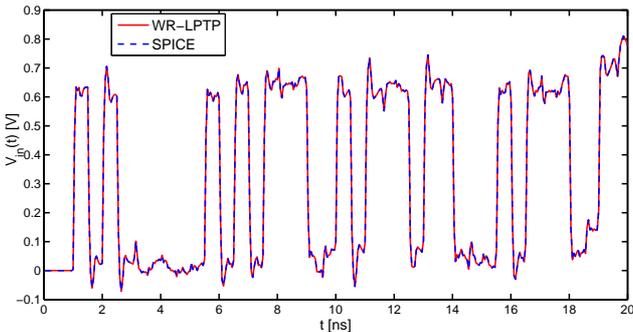$$0 \leq \eta \leq \check{\eta}_{\min} = \min_{\omega,q} \frac{2\cos\phi_q(\omega)}{\rho_q(\omega)} \tag{22}$$
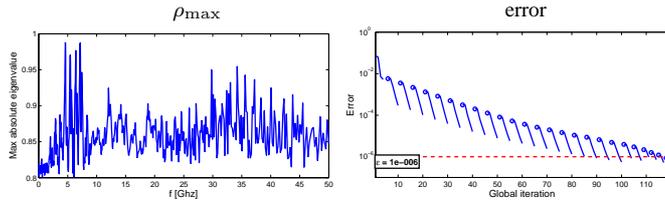
Figure 5: Case B. Left: spectral radius of the over-relaxed iteration operator with $\eta = \eta_{\text{opt}}$. Right: evolution of the inner (continuous line) and outer (dots) loop errors through the corresponding WR-SOR iterations.
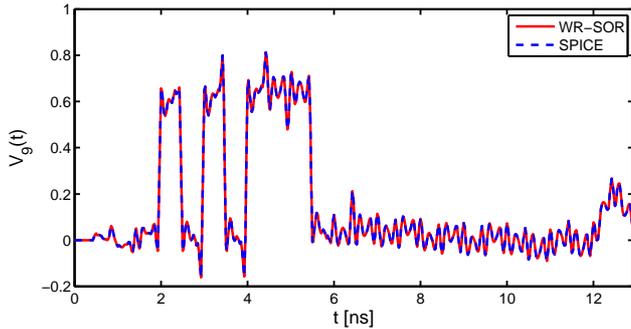

Figure 6: Case B: comparison between WR-SOR and SPICE results.

with ρ_q and ϕ_q denoting magnitude and phase of λ_q. Condition (22) is feasible only if $\Re\lambda_q > 0 \; \forall q, \omega$, leading to a non-vanishing range for the overrelaxation parameter. The optimal value for η is then found by minimizing the spectral radius of the iteration operator, i.e.,

$$\eta_{\text{opt}} = \arg\min_{\omega,q} |1 - \eta\lambda_q(\omega)| . \tag{23}$$

5 Results

In this section, we show that the WR-SOR scheme is able to fix the convergence problems that were observed in Section 3 when applying the WR-LPTP scheme to case B. The optimal over-relaxation parameter was first computed through a preprocessing analysis by minimizing the spectral radius of the iteration operator throughout the frequency range of interest. The resulting ρ_{max}, depicted in the left panel of Fig. 5, results strictly unitary bounded. Running the WR-SOR scheme on case B leads to the evolution of the iteration error depicted in the right panel of Fig. 5. As expected this error converges below the prescribed stopping threshold. Finally, we validated the WR-SOR scheme with a direct SPICE simulation. The results for one of the termination voltages of case B are depicted in Fig. 6, showing excellent correlation.

In conclusion, the proposed scheme is able to improve or guarantee convergence in some cases that are probelmatic for standard preexisting WR schemes. A general methodology for the optimization of convergence rate of the over-relaxed WR scheme has been presented. The numerical results are indeed encouraging, although the proposed methodology is still not able to guarantee the convergence in the general case. Future work will be devoted to further generalizations and to on-the-fly adjustment of the over-relaxation parameter.

References

[1] Beyene, W.T.; "Applications of Multilinear and Waveform Relaxation Methods for Efficient Simulation of Interconnect-Dominated Nonlinear Networks," *Advanced Packaging, IEEE Transactions on*, vol. 31, no. 3, pp. 637–648, Aug. 2008.

[2] Beyene, W.T.; Madden, C.; Jung-Hoon Chun; Haechang Lee; Frans, Y.; Leibowitz, B.; Ken Chang; Namhoon Kim; Ting Wu; Yip, G.; Perego, R.; "Advanced Modeling and Accurate Characterization of a 16 Gb/s Memory Interface," *Advanced Packaging, IEEE Transactions on*, vol. 32, no. 2, pp. 306–327, May 2009.

[3] F. Y. Chang, "The generalized method of characteristics for waveform relaxation analysis of lossy coupled transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 2028-2038, Dec. 1989.

[4] A. Charest, M. Nakhla, R. Achar, "Delay Extracted Stable Rational Approximations for Tabulated Networks With Periodic Reflections," *IEEE Microwave and Wireless Components Letters*, Vol. 19, No. 12, Dec 2009, pp. 768–770.

[5] A. Chinea, P. Triverio, S. Grivet-Talocia, "Delay-Based Macromodeling of Long Interconnects from Frequency-Domain Terminal Responses," *IEEE Transactions on Advanced Packaging*, Vol. 33, No. 1, pp. 246–256, Feb. 2010.

[6] A. Charest, M. Nakhla, R. Achar, "Scattering Domain Passivity Verification and Enforcement of Delayed Rational Functions," *IEEE Microwave and Wireless Components Letters*, Vol. 19, No. 10, Oct. 2009, pp. 605–607.

[7] A. Chinea, S. Grivet-Talocia, P. Triverio, "On the performance of weighting schemes for passivity enforcement of delayed rational macromodels of long interconnects," in *IEEE 18th Conf. on Electrical Performance of Electronic Packaging and Systems*, Portland (Tigard), Oregon, Oct. 19–21, 2009.

[8] E. Lelarasmee, "The Waveform Relaxation Method for Time Domain Analysis of Large Scale Integrated Circuits: Theory and Applications", *EECS Department University of California, Berkeley Technical Report No. UCB/ERL M82/40 1982.*

[9] Nakhla, N.M.; Ruehli, A.E.; Nakhla, M.S.; Achar, R.; "Simulation of coupled interconnects using waveform relaxation and transverse partitioning," *Advanced Packaging, IEEE Transactions on*, vol. 29, no. 1, pp. 78–87, Feb. 2006

[10] V. Loggia and S. Grivet-Talocia, "A two-level waveform relaxation approach for fast transient simulation of long high-speed interconnects," in *IEEE 19$^{\text{th}}$ Topical Meeting on Electrical Performance of Electronic Packaging and Systems (EPEPS 2010), Austin, TX*, October 24–27, 2010.

[11] J.K.White and A.L.Sangiovanni-Vincentelli, *Relaxation Technique for the Simulation of VLSI Circuits*. Norwell, MA: Kluwer Academic, 1987.

[12] A.Lumsdaine, M.W.Reichelt, J.M.Squyres, J.K.White, "Accelerated Waveform Methods for Parallel Transient Simulation of Semiconductor Devices", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 15, N. 7, July 1996, pp. 716–726.

[13] M.J.Gander, A.E.Ruehli, "Optimized waveform relaxation solution of electromagnetic and circuit problems", in *IEEE 19th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, 25–27 Oct. 2010, Austin, TX, USA, pp. 65–68.

[14] Y. Saad, "Iterative Methods for Sparse Linear Systems", SIAM, 2003.

[15] M.W.Reichelt, J.K.White, J.Allen, "Optimal convolution SOR acceleration of Waveform Relaxation with application to parallel simulation of semiconductor devices", *SIAM J. Sci. Comput.*, Vol. 16, No. 5, Sept. 1995, pp. 1137–1158.