A Preconditioned Waveform Relaxation Solver for Signal Integrity Analysis of High-Speed Channels

A Preconditioned Waveform Relaxation Solver for Signal Integrity Analysis of High-Speed Channels

H. Hu, S. Grivet-Talocia

Dip. Elettronica e Telecomunicazioni, Politecnico di Torino
C. Duca degli Abruzzi 24, 10129 Torino, Italy
Email: stefano.grivet@polito.it

***Abstract*—This work presents a fast transient solver for Signal Integrity analysis of high-speed channels. We consider general chip-to-chip coupled interconnect structures, including arbitrary discontinuities at chip, package and board level. An external characterization of the interconnect in terms of tabulated scattering frequency samples is first converted to a closed-form macromodel, whose transient effects on input signals can be computed very efficiently through recursive convolutions. When combined with suitable models for drivers and receivers, a large-scale but very sparse system of equations is obtained. The latter is solved by an iterative scheme based on the Generalized Minimal RESidual (GMRES) method, further enhanced by a preconditioner based on Waveform-Relaxation. Contrary to previous formulations, the proposed scheme is guaranteed to converge in few iterations. Numerical examples show that the proposed solver outperforms standard SPICE in terms of runtime, with no loss of accuracy.**

I. Introduction

The transmission of high-speed signals through electrial interconnects may be affected by several degradation effects due to metal and dielectric losses, spurious reflections by impedance discontinuities, and electromagnetic coupling. Such effects are best characterized in the frequency-domain by direct measurements or by electromagnetic simulations, typically cast as tabulated scattering responses over some prescribed bandwidth of interest.

Signal Integrity assessment must be performed in the time domain, in order to verify safe bitwise transmission at the desired bit rate. Eye diagrams and their vertical/horizontal openings are typically used for this verification. Since these eye diagrams must be computed in the time-domain, the channel characteristics should be converted into some simulation model allowing for fast computation of the transient response to long bit sequences.

Three main methods are available for this task. A first class computes transient impulse responses by inverse Fourier transform, and convolves these responses with input signals [2], [3]. This technique allows very fast and efficient simulations but is limited to linear terminations only. Since digital drivers and receivers are inherently nonlinear devices, active research is still ongoing towards inclusion of nonlinear effects, using either transistor-level or behavioral driver/receiver models. A second approach is based on standard SPICE simulations [5], which are performed in time domain by using some closed-form passive macromodel of the channel [4]. This approach may be very accurate, but is generally too slow for the generation of reliable eye diagrams with sufficient statistical information. A third class of methods has recently appeared, based on Waveform Relaxation (WR) [10], [1], [7], [8], [9]. By exploiting the structure of the overall interconnected system, this approach can include nonlinear terminations very efficiently, leading to reliable results in much less time than SPICE. Unfortunately, this approach is not always guaranteed to converge, as pointed in [6]. Even sophisticated over-relaxation schemes [6] may not fix the convergence issue.

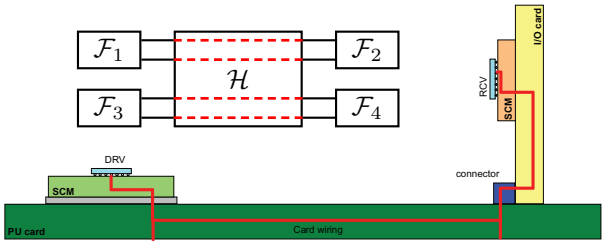

Fig. 1. A high-speed channel and its schematic representation.

This paper formulates a new transient solver for terminated high-speed channels, based on the Generalized Minimal RESidual (GMRES) scheme [11], [12]. It is well known that this scheme, mainly aimed at the solution of large-scale and sparse linear systems, always converges. In this work, we enrich this scheme by a WR-based preconditioner, which leads to convergence in very few iterations. The numerical results obtained for six industrial test cases show that the same level of accuracy of SPICE is obtained in much reduced runtime.

The present formulation is based on a linear model of drivers and receivers. However, its extension to the fully nonlinear case is possible and will be documented in a forthcoming report. Therefore, the formulation in this paper should be intended as a framework and as a basis for the development of general and highly efficient transient channel solvers.

II. Problem Statement and Background

Throughout this work, we consider a P-port (P even) fully-coupled high-speed channel $\mathcal{H}$. The channel $\mathcal{H}$ is supposed to be formed by a set of individual interconnects $\mathcal{H}_q$, $q = 1, \ldots, P/2$, each providing a direct electrical link from a

single-ended driver to a single-ended receiver (the extension to the differential case is trivial and not further commented here). Any two different interconnects $\mathcal{H}_q, \mathcal{H}_{q'}$ with $q \neq q'$ are not electrically connected, but interfere with each other through near-field electromagnetic coupling.

A channel example is depicted in Fig. 1, showing that the considered channel topology is possibly complex, including all kind of discontinuities caused by via fields, connectors, packages and irregular routing. For this reason, our starting point is a channel characterization in terms of its sampled scattering matrix $\widehat{\mathbf{H}}_l \in \mathbb{C}^{P\times P}$ at the discrete frequencies $\omega_l,\ l = 1,\ldots,L$, which may be obtained either from direct measurement or 2D/3D electromagnetic characterizations. Figure 1 provides a schematic view of the system for the case $P = 4$.

We will consider in this work six real chip-to-chip links in industrial products (scattering data courtesy of IBM) with $P = 18$ ports and consisting of a victim channel connecting ports 9 and 10 surrounded by eight aggressor channels. These structures will be labeled "Case I" to "Case VI" in the following. Figure 1 provides a schematic illustration of the channel routing for Case I.

Drivers and receivers providing the terminations to the channel, denoted as $\mathcal{F}_i, i = 1,\ldots,P$, are supposed to be independent and decoupled from each other. Such terminations may be available as transistor-level circuits or behavioral models. In typical applications, such structures are characterized by strong nonlinear characteristics, which calls for transient Signal Integrity analysis. Therefore, our main objective in this work will be to convert the frequency-domain channel characterization into a form that is suitable for transient analysis, and to propose a highly efficient transient solver.

A. Delayed-Rational Macromodels

We start by casting the channel characteristic equations in a form that is suitable for direct transient analysis. To this end, we convert the sampled frequency-domain scattering responses to a closed-form Delayed-Rational Macromodel (DRM)

$$H^{i,j}(s) = \sum_{m=0}^{M^{i,j}} \sum_{n=1}^{N_m^{i,j}} \frac{R_{mn}^{i,j}}{s - p_{mn}^{i,j}} e^{-s\tau_m^{i,j}} + D^{i,j} \qquad (1)$$

where s is the Laplace variable, i, j denote output and input port, respectively, corresponding to the selected scattering response, and $\tau_m^{i,j}$ are delays corresponding to the various arrival times of the signal reflections caused by internal and port discontinuities. Each delay term is multiplied by a low-order rational coefficient in pole-residue form, who is responsible for approximating dispersion, attenuation, and distributed coupling effects along the interconnect. The numerical technique leading to (1) is a delay-rational approximation algorithm based on a generalization of the Vector Fitting scheme [4] to include delay terms in the model structure. This algorithm has been extensively validated in [5], an illustrative example is provided in Fig. 2. The main advantage of (1) with respect to a pure rational form is compactness and efficiency in numerical simulations [4].

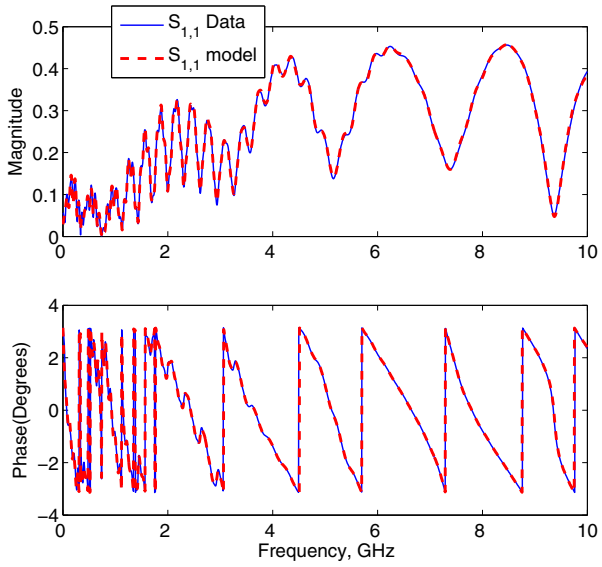

Fig. 2. Macromodeling of a high-speed channel (Case VI). Comparison between original scattering data and Delayed-Rational Macromodel response S_{11}.

The impulse response $\mathbf{h}(t)$ corresponding to (1) is a linear combination of delayed exponential terms

$$h^{i,j}(t) = \sum_{m=0}^{M^{i,j}} \sum_{n=1}^{N_m^{i,j}} R_{mn}^{i,j} e^{p_{mn}^{i,j}(t-\tau_m^{i,j})} u(t-\tau_m^{i,j}) + D^{i,j}\delta(t) \quad (2)$$

as readily obtained by analytical inverse Laplace transform. The time-domain application of the channel operator to some input signal $a(t)$ results in a convolution with (2). We will be interested in the solution over some interval $t \in [0, T]$, and we will discretize this time interval with a fixed time step δt. Denoting with $t_k = k\delta t$, if we extract a single impulse response term $h(t) = e^{p(t-\tau)}u(t-\tau)$ from (2) we have the general first-order approximation

$$b(t_k) = (h * a)(t_k) \simeq \alpha_0 b(t_{k-1}) + \sum_{\nu=0}^{2} \beta_\nu a(t_{k-\nu-\bar{k}}) \qquad (3)$$

where $\bar{k} = \lfloor \frac{\tau}{\delta t} \rfloor$ and with suitable precomputed coefficients [10]. This expression shows that convolution with (2) can be computed very efficiently through a three-term recursive formula.

In the following, we will denote with $\boldsymbol{a}(t)$ and $\boldsymbol{b}(t)$ the vectors collecting all incident and reflected transient scattering waves from the channel interface ports, and with $\mathbf{a}$ and $\mathbf{b}$ the vectors stacking all time samples of the transient scattering waves $\boldsymbol{a}(t_k), \boldsymbol{b}(t_k)$ for $k = 0,\ldots,K$ with $K = T/\delta t$. With this compact notation, the channel response can be described in abstract terms as

$$\mathbf{b} = \mathcal{H}\,\mathbf{a}\,, \qquad (4)$$

where $\mathcal{H}$ is a linear convolution operator based on (2). This operator can be interpreted as a very large $(KP) \times (KP)$

matrix, which requires however only $O(KP)$ operations for its application to some vector $\mathbf{a}$ thanks to (3).

A similar formulation can be applied to the channel terminations, which can be represented as

$$\mathbf{a} = \mathcal{F}(\mathbf{b})\,. \tag{5}$$

The above abstract form will be particularized in this work to the simpler linear form

$$\mathbf{a} = \mathcal{G}\mathbf{b} + \mathcal{Q}\mathbf{u}\,, \tag{6}$$

where $\mathcal{G}, \mathcal{Q}$ are large and very sparse matrices which can be automatically obtained from a Modified Nodal Analysis (MNA) of the termination circuits, and $\mathbf{u}$ collects all time samples of the independent sources within the terminations. Whenever the linear structure of the termination equations is not relevant, the more general form (5) will be used in the presentation.

B. Two-Level Waveform-Relaxation

Channel (4) and termination (5) equations form a coupled system, whose efficient solution is our main objective. Contrary to standard circuit solvers such as SPICE, we do not apply here a time-stepping process to compute individual time samples, but we consider the whole waveforms $\mathbf{a}$ and $\mathbf{b}$ as unknowns. These will be solved within a Waveform Relaxation framework by refining an initial guess through iterations. In this section, we recall some background results from [10], which will be used as starting points for our developments.

We first separate the channel recursive convolution operator into two parts as

$$\mathcal{H} = \mathcal{D} + \mathcal{C} \tag{7}$$

where $\mathcal{D}$ is a block-diagonal operator collecting direct transmission and reflection coefficients of individual channels, and operator $\mathcal{C}$ represents all the crosstalks. It is expected that $\|\mathcal{C}\| \ll \|\mathcal{D}\|$ for a well-designed channel, since the cumulative effect of crosstalk should not deteriorate signal transmission excessively. This consideration led in [10] to a two-level waveform-relaxation scheme as

$$\begin{cases} \mathbf{b}_{\mu,\nu} = \mathcal{D}\,\mathbf{a}_{\mu,\nu-1} + \boldsymbol{\theta}_{\mu-1}\,, \\ \mathbf{a}_{\mu,\nu} = \mathcal{F}\left(\mathbf{b}_{\mu,\nu}\right)\,, \\ \boldsymbol{\theta}_{\mu} = \mathcal{C}\,\mathbf{a}_{\mu,\bar{\nu}}\,, \end{cases} \tag{8}$$

where iteration indexes μ and ν correspond to *outer transverse* and *inner longitudinal* relaxation, respectively. The WR iterations start with $\mu = 1$ by setting the initial conditions $\mathbf{a}_{1,0} = \mathbf{0}$ and $\boldsymbol{\theta}_0 = \mathbf{0}$ (no inter-channel coupling considered in the first outer iteration). Then, the first two equations in (8) are applied iteratively for $\nu = 1, \ldots, \bar{\nu}$. This process, called inner or longitudinal relaxation, solves decoupled channels through iterative forward evaluation of channel and termination equations, requiring no matrix inversion (linear case) or Newton-Raphson iterations (nonlinear case). This technique is recognized as a fixed-point iteration. After a total number of $\bar{\nu}$ inner iterations, the available solution estimate is used in the third equation to estimate the crosstalk contributions and define the outer *relaxation source* $\boldsymbol{\theta}_\mu$. Then, the overall process is repeated for $\mu = 1, 2, \ldots$ until convergence.

The above scheme, denoted as WR-LPTP in [10], presents some major advantages. The overall system is partitioned into separate parts, which are solved independently using forward evaluations of their constitutive equations. Therefore, this approach is very well suited to parallel implementation and deployment on multicore hardware. The results in [6] show that with optimized code, the WR-LPTP can outperform standard SPICE solvers by 2–3 orders of magnitude. Also, full control over accuracy is possible, up to a desired error threshold, as the validations in [10] show.

Unfortunately, there are also some disadvantages, so that this simulation speed comes with a cost. The main problem is convergence, which is not always guaranteed. The main objective of this work is in fact to present a scheme that preserves efficiency but is able to guarantee unconditional convergence. We first report a brief convergence analysis in Sec. II-C. Then, we present our new proposed numerical scheme in Sec. III.

C. Linear convergence analysis

The convergence of the WR-LPTP schemes can be assessed in the frequency domain by supposing linear termination conditions (6). In the following, we will denote with $\mathbf{H}(j\omega) = \mathbf{D}(j\omega) + \mathbf{C}(j\omega)$ and $\boldsymbol{\Gamma}(j\omega)$ the frequency-dependent scattering matrices of channel (split into its block-diagonal and remainder parts) and terminations, respectively, and with boldface uppercase letters $\boldsymbol{A}(j\omega)$ the frequency-domain scattering waves at the channel ports. The explicit argument $(j\omega)$ will be omitted.

As shown in [6], the error between the solution estimate $\boldsymbol{A}_{\mu,\bar{\nu}}$ at the outer iteration μ and the exact solution $\boldsymbol{A}_{\text{exact}}$ is

$$\boldsymbol{\mathcal{E}}_{\mu,\bar{\nu}} = \boldsymbol{A}_{\mu,\bar{\nu}} - \boldsymbol{A}_{\text{exact}} = -\mathbf{P}_{\bar{\nu}}^{\mu}\boldsymbol{A}_{\text{exact}}\,, \tag{9}$$

where

$$\mathbf{P}_{\bar{\nu}} = \mathbf{P} + (\boldsymbol{\Gamma}\mathbf{D})^{\bar{\nu}}\,(\mathbf{I} - \mathbf{P}) \tag{10}$$

and

$$\mathbf{P} = (\mathbf{I} - \boldsymbol{\Gamma}\mathbf{D})^{-1}\,(\boldsymbol{\Gamma}\mathbf{C})\,. \tag{11}$$

Convergence is guaranteed, equivalently, $\boldsymbol{\mathcal{E}}_{\mu,\bar{\nu}} \to 0$ for $\mu \to \infty$, if the spectral radius (the magnitude of the largest eigenvalue) of operator $\mathbf{P}_{\bar{\nu}}$ is unitary bounded as

$$\rho_{\max}\{\mathbf{P}_{\bar{\nu}}\} < 1 \tag{12}$$

at any frequency ω.

The convergence condition (12) is not always verified in practical cases. For instance, although the crosstalk elements are usually negligible at low frequencies, there might be frequency bands over which these elements are even larger than direct transmission and refection coefficients. At those frequencies, the interconnect behaves like a directional coupler by transferring energy from one channel to another. This situation prevents convergence of the WR-LPTP, as observed in [6]. For this reason, modified WR schemes including constant or even frequency-dependent optimal over-relaxation

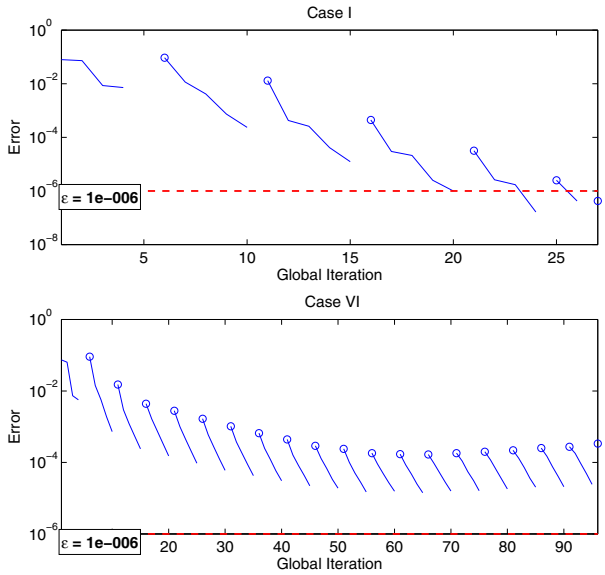

Fig. 3. Evolution of the inner (continuous line) and outer (dots) loop errors through iterations for Cases I and VI

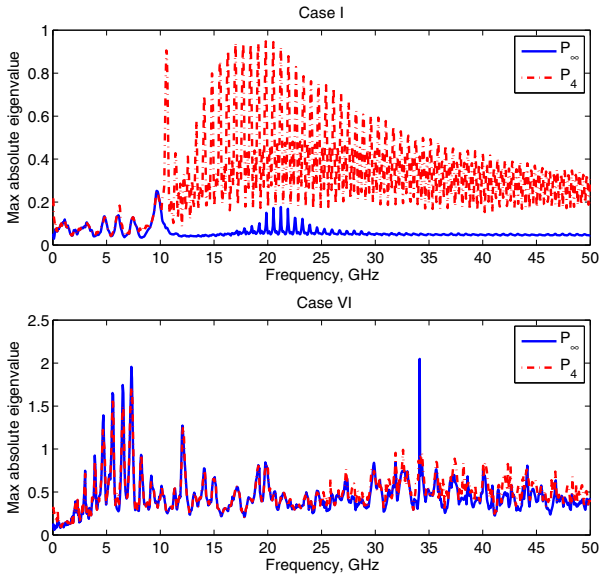

Fig. 4. Spectral radius of operator $\mathbf{P}_{\bar{\nu}}$ for $\bar{\nu} = 4$ and $\bar{\nu} = \infty$ plotted versus frequency for Cases I and VI.

have been proposed in [6]. However, those formulations still do not guarantee convergence a priori.

Top and bottom panels of Fig. 3 illustrate the above convergence issues on two test cases under consideration, excited by drivers with 40Ω internal impedance at odd-numbered ports, and loaded by 1 pF capacitors at the even-numbered ports. The figures depict the error between two successive inner iterations (solid lines) and the error between two outer iterations (dots), using a global iteration count. For Case I (top panel), both inner and outer iterations converge. For Case VI, although the inner iteration loop always converges, the outer transverse relaxation loop diverges in few iterations.

This behavior is easily explained by computing the spectral radius of the iteration operator (10), depicted in top and bottom panels of Fig. 4 for Case I and VI, respectively. This quantity is less than one at all frequencies for Case I, implying convergence. It is instead larger than one in various limited frequency bands for Case VI, which is sufficient to make the WR-LPTP scheme inapplicable.

Finally, we remark that the convergence condition (12) depends on both channel $\mathbf{H}$ and termination $\mathbf{\Gamma}$ scattering matrices. So, the WR-LPTP scheme might converge for a given channel with some terminations and diverge for the same channel with different terminations. This randomness is obviously not acceptable for a transient solver. Therefore, we need a robust formulation that is guaranteed to converge unconditionally. This formulation is the key contribution of this work and is presented next.

III. A Hybrid WR-GMRES Simulator

Our proposed scheme is based on the well-known Generalized Minimal Residual (GMRES) scheme, here applied with the WR-LPTP scheme as a preconditioner. We start by recalling the basic GMRES framework in Sec. III-A, showing how our channel simulation problem fits in this framework. Then, we combine GMRES and the WR-LPTP schemes into our proposed hybrid solver in Sec. III-B.

A. GMRES: Generalized Minimal Residual

The GMRES method [11], [12] is a very popular iterative algorithm to solve general systems

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{13}$$

where $\mathbf{A}$ is any full-rank $m \times m$ non symmetric matrix. The technique works by picking an initial guess $\boldsymbol{x}_0$ of the solution, computing the residual $\mathbf{r}_0 = \boldsymbol{b} - \mathbf{A}\boldsymbol{x}_0$, and finding a refined estimate of the solution $\boldsymbol{x}_n = \boldsymbol{x}_0 + \boldsymbol{z}_n$ at the n-th iteration, where

$$\boldsymbol{z}_n = \arg \min_{\boldsymbol{z} \in \mathcal{K}_n} \|\mathbf{r}_0 - \mathbf{A}\boldsymbol{z}\|_2 \,, \tag{14}$$

where

$$\mathcal{K}_n = \text{span}\left\{\boldsymbol{r}_0, \boldsymbol{A}\boldsymbol{r}_0, \boldsymbol{A}^2\boldsymbol{r}_0, \ldots, \boldsymbol{A}^{n-1}\boldsymbol{r}_0\right\} \tag{15}$$

denotes the n-th order Krylov subspace associated to pair $(\mathbf{A}, \boldsymbol{r}_0)$. Condition (14) corresponds to finding a particular solution that minimizes the euclidean norm of the residual $\mathbf{b} - \mathbf{A}\boldsymbol{x}_n$. This solution is found by projecting the original linear system (13) onto the Krylov subspace $\mathcal{K}_n$ and finding the corresponding optimal solution as a least-squares problem. Practical implementations of the GMRES scheme are based on the Arnoldi algorithm [13] to construct an orthonormal basis for $\mathcal{K}_n$ and to project the original system onto this space.

The GMRES scheme is particularly well-suited for problems characterized by a possibly large but sparse system matrix. No matrix inversion or factorization is performed, rather

only the construction of the Krylov subspace (15) is needed, which in turns only requires matrix-vector multiplications.

Our transient channel simulation problem requires to find the solution of coupled equations (4) and (6), resulting in the following linear system

$$(\mathcal{I} - \mathcal{G}\mathcal{H})\mathbf{a} = \mathcal{Q}\mathbf{u}\,. \tag{16}$$

As discussed in the foregoing sections, the application of matrix operators $\mathcal{G}$ and $\mathcal{H}$ can be performed very efficiently, since these operators correspond to recursive convolutions. So, application of matrix operator $(\mathcal{I}-\mathcal{G}\mathcal{H})$ to a given vector $\mathbf{a}$ can be performed very fast. These considerations justify the idea of applying the GMRES scheme to solve (16). We proceed as follows.

1) To setup the GMRES iterations, we need an initial estimate of the solution $\mathbf{a}_0$. This estimate is computed as the result of the first ($\mu = 1$) WR-LPTP iteration. Practically, the inter-channel couplings $\mathcal{C}$ are neglected, and the various individual channels are solved with their terminations through iterative application of the first two equations in (8). Note that this inner iteration always converges if the channel model is passive, see [10].
2) Successively refined solution estimates are computed as $\mathbf{a}_n = \mathbf{a}_0 + \boldsymbol{\delta}_n$, where $\boldsymbol{\delta}_n$ are elements of the Krylov subspace
$$\mathcal{K}_n = \text{span}\left\{\mathbf{r}_0, (\mathcal{I} - \mathcal{G}\mathcal{H})\mathbf{r}_0, \ldots, (\mathcal{I} - \mathcal{G}\mathcal{H})^{n-1}\mathbf{r}_0\right\} \tag{17}$$
and $\mathbf{r}_0 = \mathcal{Q}\mathbf{u} - (\mathcal{I} - \mathcal{G}\mathcal{H})\mathbf{a}_0$.
3) Iterations are stopped when the residual norm is less than a prescribed threshold, $\|\mathbf{r}_n\| < \epsilon$, where $\mathbf{r}_n = \mathcal{Q}\mathbf{u} - (\mathcal{I} - \mathcal{G}\mathcal{H})\mathbf{a}_n$.

Supposing that GMRES stops after $\bar{n}$ iterations, the number of required operations is approximately $KP\bar{n}(\bar{n} + 3N_{\text{avg}})$, where N_{avg} is the average number of coefficients $R^{i,j}_{mn}$ among all macromodel responses in (1), while the storage requirement is approximately $(\bar{n}+2)KP$. So, both CPU and memory costs scale linearly with the number of time samples to be computed. Since the number of required bits may be large for a reliable statistical analysis of signal degradation effects, e.g. in terms of eye diagram opening, the above memory requirements may still be excessive. For this reason, in this work we adopt a *restarted* version of the GMRES scheme, which essentially locks to a fixed maximum size $n_{\max}$ the Krylov subspace $\mathcal{K}_n$. If the required number of iterations is larger, the last added vector to the space is taken as the seed for the generation of a new refined Krylov subspace. This procedure is standard, we refer the Reader to [12] for details. In this work, we set $n_{\max} = 10$ in all numerical tests.

B. Preconditioning

We now discuss our strategy for reducing the number of required iterations $\bar{n}$ for convergence. In fact, we do not know a priori how many iterations will be needed, and theoretical results do not help in this respect, since guaranteed convergence takes place only after a number of iterations equal to the matrix size. However, a small number of iterations can be achieved by properly reducing the condition number $\kappa(\mathbf{A})$ of the system through a suitable preconditioning process.

Consider the general system (13) and premultiply both sides by a nonsingular matrix $\mathbf{M}$,

$$\mathbf{M}\mathbf{A}\boldsymbol{x} = \mathbf{M}\boldsymbol{b}\,. \tag{18}$$

Clearly, the solution of (18) is the same solution of (13). Matrix $\mathbf{M}$, the *preconditioner*, should be chosen so that

- the condition number results $\kappa(\mathbf{M}\mathbf{A}) \ll \kappa(\mathbf{A})$,
- application of $\mathbf{M}$ requires a negligible overhead.

The first requirement is best achieved when $\mathbf{M}$ is a good approximation of the inverse $\mathbf{A}^{-1}$. In fact, if $\mathbf{M} = \mathbf{A}^{-1}$, the modified system matrix in (18) becomes the identity, and the solution requires no iterations at all. Since the inverse of sparse matrices is in general full, it is usually not possible to choose $\mathbf{M} \simeq \mathbf{A}^{-1}$, since preconditioning would require the same computational effort than the solution of the original system. So, a good compromise between the two above conflicting requirements must be found.

The structure of our channel operator $\mathcal{H}$ helps in defining a good preconditioner for (16). In fact, we know that $\mathcal{H} = \mathcal{D}+\mathcal{C}$, where the coupling terms $\mathcal{C}$ can be considered as a second-order effect with respect to the dominant contribution from $\mathcal{D}$. So, the basic idea is to consider the following approximation

$$(\mathcal{I} - \mathcal{G}\mathcal{H})^{-1} \simeq (\mathcal{I} - \mathcal{G}\mathcal{D})^{-1} \tag{19}$$

and to use the right-hand side as a preconditioner. The resulting preconditioned system reads

$$(\mathcal{I} - \mathcal{G}\mathcal{D})^{-1}(\mathcal{I} - \mathcal{G}\mathcal{H})\mathbf{a} = (\mathcal{I} - \mathcal{G}\mathcal{D})^{-1}\mathcal{Q}\mathbf{u}\,. \tag{20}$$

Practical application of this preconditioner requires the solution of system

$$(\mathcal{I} - \mathcal{G}\mathcal{D})\mathbf{x} = \mathbf{y}\,, \tag{21}$$

where $\mathbf{y}$ is the generic vector in the original Krylov subspace construction (17) and $\mathbf{x}$ is the vector to be used for the construction of the Krylov subspace of the preconditioned system. This vector is computed by the Waveform Relaxation iteration

$$\begin{cases} \mathbf{z}_\nu = \mathcal{D}\,\mathbf{x}_{\nu-1}\,, \\ \mathbf{x}_\nu = \mathcal{G}\mathbf{z}_\nu + \mathbf{y}\,, \end{cases} \tag{22}$$

with $\nu = 1, \ldots, \bar{\nu}$. The resulting overhead for application of the preconditioner thus results minimal, since only $\bar{\nu}$ WR iterations are performed. Moreover, no inter-channel couplings are involved in these iterations, thus allowing naive parallelization of the overall scheme for further simulation speedup.

IV. Numerical results

We start by showing the effectiveness of the proposed preconditioner for the reduction of GMRES iterations. Figure 5 shows the error evolution of the of WR-GMRES solution estimates through iterations with and without preconditioner for Case VI. It can be seen that, with our proposed preconditioner, less iterations are required and the error decreases faster, resulting in an overall simulation speedup.

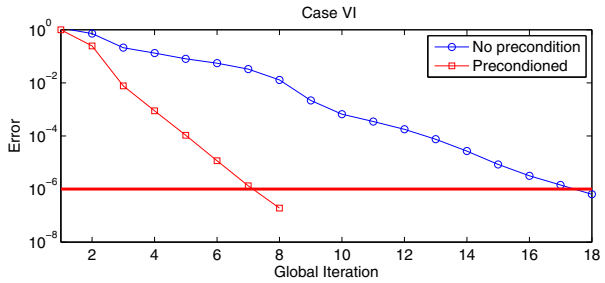

Fig. 5. Iterative errors for GMRES with and without preconditioner.

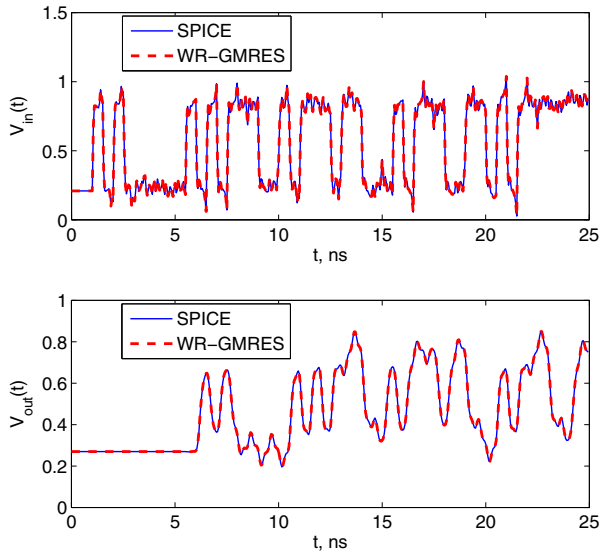

Fig. 6. Comparison of transient results of WR-GMRES and SPICE for Case VI.

The accuracy of the WR-GMRES simulation is demonstrated for Case VI, for which the standard WR-LPTP scheme does not converge. Figure 6 shows a comparison of transient results for the port signals of the victim channel (ports 9 and 10) obtained by proposed scheme and SPICE. The results are undistinguishable on this scale.

Finally, we illustrate the efficiency of the proposed solver. We consider a long pseudo-random bit sequence (a total of 1000 bits) and we report in Table I the runtime required by SPICE and by the WR-GMRES scheme for all benchmark cases. All simulations converge as expected, and the overall runtime is much reduced with respect to SPICE. The WR-GMRES simulation time results approximately twice the runtime of the WR-LPTP scheme, except for the cases II and VI for which WR-LPTP is not applicable due to lack of convergence.

TABLE I
COMPARISON OF SIMULATION TIMES (SECONDS) REQUIRED BY SPICE, WR-LPTP AND PROPOSED WR-GMRES SCHEME.

Case	SPICE	WR-LPTP	WR-GMRES
I	1611	79	168
II	1746	—	217
III		62	135
IV	1115	56	98
V	1740	68	159
VI	1715	—	256

V. CONCLUSIONS

We have presented a hybrid transient solver for Signal Integrity analysis high-sped channels. This solver is based on a formulation of the well-known GMRES scheme, combined with a preconditioner based on system partitioning and longitudinal Waveform Relaxation. Numerical results show that the same accuracy of SPICE is obtained by our proposed solver in much faster runtime, with guaranteed convergence. Further work is ongoing to generalize the scheme for the inclusion of general nonlinear driver and receiver circuits, as typically found is real-world systems.

REFERENCES

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

[2] W.T.Beyene, C.Madden, J.-H.Chun, H.Lee, Y.Frans, B.Leibowitz, K.Chang, N.Kim, T.Wu, G.Yip, R.Perego, "Advanced Modeling and Accurate Characterization of a 16 Gb/s Memory Interface," *IEEE Trans. on Advanced Packaging*, vol. 32, no. 2, pp. 306–327, May 2009.

[3] G.Balamurugan, B.Casper, J.E.Jaussi, M.Mansuri, F.O'Mahony, J.Kennedy, "Modeling and Analysis of High-Speed I/O Links," *IEEE Trans. on Advanced Packaging*, vol. 32, no. 2, pp. 237–247, May 2009.

[4] 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.

[5] A. Chinea, S. Grivet-Talocia, H. Hu, P. Triverio, D. Kaller, C. Siviero, M. Kindscher, "Signal Integrity Verification of Multi-Chip Links using Passive Channel Macromodels," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, pp. 920–933, June 2011.

[6] S. Grivet-Talocia, and V. Loggia, "Fast Channel Simulation via Waveform Over-Relaxation," in *15th IEEE Workshop on Signal Propagation on Interconnects, Napoli, Italy*, May 8–11, 2011.

[7] 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.

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

[9] N.Nakhla, A.E.Ruehli, M.S.Nakhla, R.Achar, C.Chen, "Waveform Relaxation Techniques for Simulation of Coupled Interconnects With Frequency-Dependent Parameters", *IEEE Transactions on Advanced Packaging*, Vol.30, N.2, 2007, pp. 257–269.

[10] V.Loggia, S.Grivet-Talocia, H.Hu, "Transient simulation of complex high-speed channels via Waveform Relaxation", *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, pp. 1823–1838, November 2011.

[11] Y.Saad, M.H.Schultz, "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems," *SIAM J. Sci. Stat. Comput.*, Vol. 7, No. 3, July 1986

[12] B.Richard, B.Michael, F.C.Tony, D.James, M.D.June, D.Jack, E.Victor, P.Roldan, R.Charles, Henk Van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 1993. Available: http://www.siam.org/books.

[13] W.E.Arnoldi, "The principle of minimized iterations in the solution of the matrix eigenvalue problem," *Quarterly J. of Applied Mathematics*, Vol. 9, pp. 17–29, 1951.