POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Two-Level Waveform Relaxation Approach for System-Level Power Delivery Verification

(Article begins on next page)

15 May 2024

A Two-Level Waveform Relaxation Approach for System-Level Power Delivery Verification

Alessandro Moglia*, Antonio Carlucci*, Stefano Grivet-Talocia*,
Scott Mongrain§, Sid Kulasekaran§, Kaladhar Radhakrishnan§
*Dept. Electronics and Telecommunications, Politecnico di Torino, Italy
§Intel Corporation, Chandler, AZ, USA
stefano.grivet@polito.it

Abstract—This paper considers a complete power delivery network model of a multicore processing system, including per-core voltage regulation loops through Fully Integrated Voltage Regulators. Based on a nonlinear descriptor formulation of the system equations, we propose a transient solver based on a two-level Waveform Relaxation iteration. The convergence properties and the scalability of this solver when implemented on a parallel computing architecture are investigated. Numerical results show fast convergence and excellent scalability properties.

I. INTRODUCTION

This work is part of a research effort by the Authors towards the development of quasi-real time transient power integrity verification of modern microprocessor systems for High-Performance Computing (HPC) and Artificial Intelligence (AI) applications. Such structures are equipped with possibly more than a hundred computing cores. The latter are fed by a power delivery network connecting the platform Voltage Regulation Module (VRM) to chip activity models through board and module interconnects, loaded by a large number of decoupling capacitors. A fine-grained stabilization of the supply voltage is required at the core level, which is achieved through a bank of Fully Integrated Voltage Regulators (FIVRs) in Buck configuration, including switching circuitry and LC filtering networks in a feedback configuration with dedicated controllers.

The objective of this work is to efficiently evaluate the transient voltage at suitably defined ports on each core, induced by realistic switching current loads. The large-scale nature of this problem makes a direct SPICE simulation very challenging or even unfeasible, since accurate models of all system parts are to be simulated concurrently, including high-accuracy electromagnetic models of board and package.

Previous attempts to reduce the computational cost were mainly directed towards eliminating redundancy of the models, through ad hoc Model Order Reduction approaches (MOR). A Krylov projection framework was applied in [1], and a black-box parameterized macromodeling framework was developed in [2]. In both cases, results clearly demonstrated that MOR can significantly speedup system simulation, resulting from an accuracy-controlled compression of the system equations.

In this work, we consider an alternative approach that focuses on the transient solver. Rather than solving the fully coupled system equations, a two-level partitioning scheme is applied to separate PDN contributions associated to individual cores (transverse partitioning) and board/package from the rest of per-core network (longitudinal partitioning). These subsystems are solved independently, possibly in parallel on a multithread computing platform, in a significantly reduced runtime. Equivalence with the original full system is recovered through insertion of suitable relaxation sources, which are updated in successive iterations through a Waveform Relaxation (WR) scheme. WR methods are well known [3]–[6] and have been extensively studied. This work fits in the framework introduced in [7], although for a different application.

This paper documents preliminary results on convergence and runtime on prototype implementations of two WR schemes. Numerical results computed for an Intel-based enterprise server with 60 computing cores show convergence to engineering accuracy levels in less than 4 iterations, with excellent parallel efficiency and scalability.

II. FORMULATION

The formulation of the equations that represent the dynamic behavior of the system represented in Fig. 1 is discussed in [1] and [2]. We do not repeat here all derivations, which lead to the following differential-algebraic (DAE) system

$$\dot{x} = \mathbf{A}x + \mathbf{B}_w w + \mathbf{B}_u u \tag{1a}$$

$$z = \mathbf{C}_z x + \mathbf{D}_{zw} w + \mathbf{D}_{zu} u \tag{1b}$$

$$v^o = \mathbf{C}_v x + \mathbf{D}_{vw} w + \mathbf{D}_{vu} u \tag{1c}$$

$$w = \mathbf{\Delta}(d)z \tag{1d}$$

$$e = \mathbf{N}v^o - \mathbf{V}_{\text{ref}} \tag{1e}$$

$$\dot{x}_{\mathcal{K}} = \mathbf{A}_{\mathcal{K}}x_{\mathcal{K}} + \mathbf{B}_{\mathcal{K}}e \tag{1f}$$

$$d = \mathbf{C}_{\mathcal{K}}x_{\mathcal{K}} \tag{1g}$$

The input vector u includes all current excitations loading all cores, and the output vector v^o collects the corresponding voltages to be monitored. Variables $w = (i_1; v_2)$ and $z = (v_1; i_2)$ collect the currents and voltages at the interface between FIVR switches and input network (index $_1$) and between switches and output networks (index $_2$). The corresponding
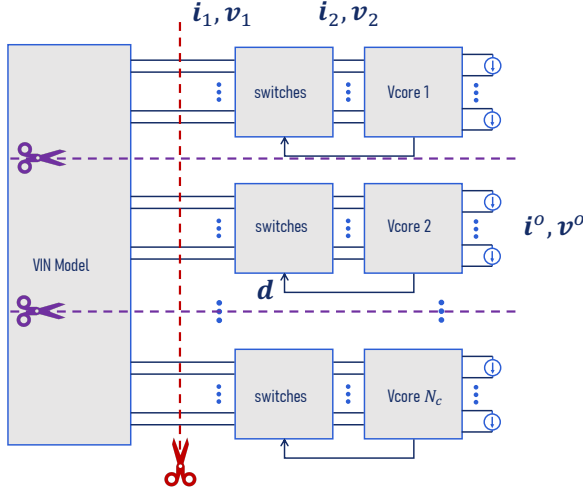
Fig. 1. Schematic description of the power delivery network addressed in this work with annotation of the main variables. Feedback loops include comparison with voltage references and controllers (not shown). Purple and red cutlines correspond to transverse and longitudinal partitioning, respectively.

equation (1d) represents the averaged characteristics of the FIVR switches, which are represented as a bank of ideal transformers with turn ratios collected in vector d. The latter correspond to the duty cycle signals provided instantaneously by the closed-loop controllers (1f)-(1g). Note that matrix $\Delta(d)$ is linear in each duty cycle component d_k of each k-th core. Vector e in (1e) collects the voltage mismatches at the output with respect to per-core voltage references V_{ref}. The first two equations (1a)-(1b) collect all models of board, package, decaps, as well as all individual core models (Buck inductors, on-chip capacitances, and chip PDN models) in state-space form. Such models are obtained by first processing with a passive rational fitting engine [8] the frequency responses of unloaded board and package interconnects, followed by termination of decaps ports with corresponding decap models. The latter operation is performed by assembling all individual state-space models of the various subblocks.

III. TRANSIENT SOLVERS

The direct transient solution of the full system (1) can be performed by applying any DAE discretization method. Here we adopt a modification of the Backward Euler method based on uniform time stepping $t_k = k \cdot \delta t$, where the nonlinear term (1d) is discretized with a mixed explicit-implicit strategy so that no explicit solution of large-scale nonlinear algebraic equations is required at each time step. This scheme will be the reference for the numerical results.

A. Circuit partitioning and Waveform Relaxation

Figure 1 illustrates the circuit partitioning strategy supporting the proposed Waveform Relaxation solver. The purple line (Transverse Partitioning, TP) neglects the coupling between the portions in the input model directly connected through the switches to different cores. These couplings are reintroduced

as series voltage sources at the input-switch interface, which are used as relaxation terms. The red line (Longitudinal Partitioning, LP) separates on a per-core level the input model from switches and output models. Also in this case, the connection between the separated blocks is established through relaxation sources placed at the disconnected ports. All this procedure is standard and is well documented, e.g. in [6], [7].

The resulting two-level partitioning separates the original system in decoupled blocks with additional relaxation source terms. At the algebraic level, this partitioning corresponds to extracting for each individual subcircuit the corresponding blocks in the global state-space matrices, so that the solution of an individual subsystem does not require any variable from the other decoupled parts. Waveform relaxation is then setup as a double nested loop that performs outer (transverse) and inner (longitudinal) iterations. A pseudocode is listed in Algorithm 1, where a placeholder ξ is used to denote any variable at the interface of the various partitions.

Algorithm 1 Basic WR-LPTP iteration scheme

1: Find initial conditions (nominal DC solution)
2: Partition circuit and initialize relaxation sources
3: **for** $\mu = 1$ to $\mu_{\max}$ **do**
4:     **for** $\nu = 1$ to $\nu_{\max}$ **do**
5:         Solve all subblocks for interface variables $\xi_{\mu,\nu}$
6:         Update inner relaxation sources
7:         **if** $||\xi_{\mu,\nu} - \xi_{\mu,\nu-1}||_\infty < \epsilon$ **then**
8:             Break
9:         **end if**
10:     **end for**
11:     Update outer relaxation sources
12:     **if** $||\xi_\mu - \xi_{\mu-1}||_\infty < \epsilon$ **then**
13:         Break
14:     **end if**
15: **end for**

B. Convergence

Convergence is here analyzed by considering three scenarios

- **TP only**: in this setting, the blocks separated by purple lines in Fig. 1 are not decoupled and assumed to be solved exactly. Convergence is analyzed by considering transverse relaxation only. Intuitively, this scheme is expected to converge in very few iterations since the couplings between different cores through the input network is expected to be very small. The solution of the decoupled cores is a good proxy of the overall system solution, provided that the entire system is well designed.
- **LP only**: similarly, in this setting only the splitting based on the red line is considered in Fig. 1. In particular, the entire Input model is left unpartitioned, and convergence is analyzed by considering longitudinal partitioning only. Also in this case a fast convergence is expected since the entire PDN aims at stabilizing the voltages through the feedback controls. This stabilization effectively helps convergence, since the difference between successive
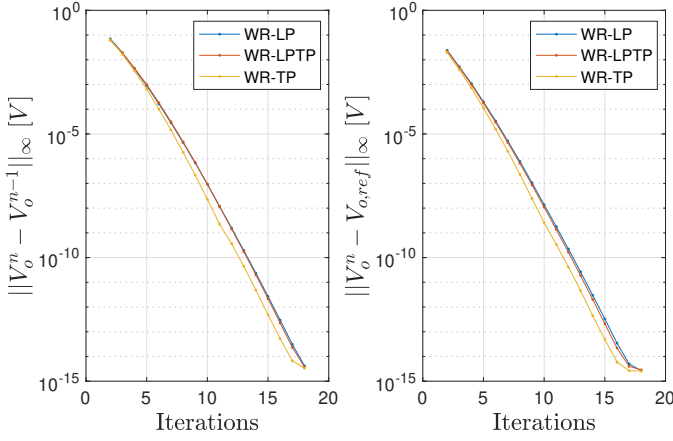
Fig. 2. Convergence of Waveform Relaxation schemes

Fig. 3. Runtime of Waveform Relaxation schemes. The dashed lines provide a reference ideal scaling factor.

iterations induced by inexact values of the relaxation sources is automatically reduced by the system dynamics.

- **LPTP**: In this setting, both partitions are applied and overall convergence holds when both inner and outer iterations converge. The main advantage of this setting is the major reduction in the solution of the individual subblocks: there are many such subblocks but each of them is very small and requires a much reduced runtime for its transient analysis.

IV. NUMERICAL RESULTS AND DISCUSSION

Numerical results are presented for an Intel-based enterprise server model with $N_c = 60$ cores, $N_p = 3$ FIVR phases per core, and $N_o = 57$ output ports per core. The resulting DAE system (1) includes $N = 45726$ states and $P = 3420$ total inputs/outputs.

Figure 2 demonstrates convergence properties of the three WR scenarios, confirming that very few iterations are indeed required for all schemes. The left panel reports the correction between successive iterations, whereas the right panel reports the error with respect to a reference solution. The three WR schemes basically gain one digit of precision per iterations. Adopting a condition to stop iterations when the practical engineering accuracy 10^{-4} is attained allows to limit the overall number of iterations to 4-5.

The runtime from a prototype multithread C implementation of WR-LP and WR-LPTP (based on $\nu_{\max} = 2$) is documented in Fig. 3 (at the time of writing the parallel implementation of the WR-TP scheme was not complete and its documentation is left to a future report). Figure 3 reports the total execution time required by a ramp-up (all 60 cores are switched on simultaneously) simulation requiring 11000 time steps. Simulations are performed with up to 30 parallel computing threads. The plot demonstrates an excellent parallel efficiency of both implementations, when compared to the reference ideal 100% efficiency (dashed line). For this particular structure, the WR-LP scheme results more efficient than the WR-LPTP, probably due to a very aggressive voltage stabilization as required by the hardware platform, which makes feedback between input
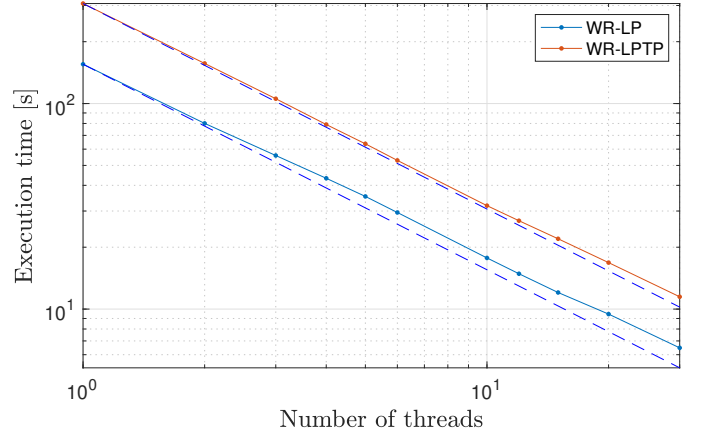
and output networks very weak. Considering that the reference runtime for a serial implementation of the solver directly applied to (1) is 26 seconds, these results also confirm a significant overhead required by the iterative nature of all WR approaches. Nonetheless, the excellent parallel efficiency of proposed WR implementations can further reduce this cost by running the solver on a massive multithread hardware.

V. CONCLUSIONS

This paper demonstrated the high potential of parallelized Waveform Relaxation schemes applied to fast transient Power Integrity verification of multicore architectures at the system level. Our future investigation will couple the proposed WR solvers with reduced-order modeling strategies, in order to further reduce simulation times.

REFERENCES

[1] A. Carlucci, S. Grivet-Talocia, S. Mongrain, S. Kulasekaran, and K. Radhakrishnan, "A structured Krylov subspace projection framework for fast power integrity verification," in *2023 IEEE 27th Workshop on Signal and Power Integrity (SPI)*, 2023, pp. 1–4.

[2] A. Carlucci, T. Bradde, S. Grivet-Talocia, S. Mongrain, S. Kulasekaran, and K. Radhakrishnan, "A compressed multivariate macromodeling framework for fast transient verification of system-level power delivery networks," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 13, no. 10, pp. 1553–1566, 2023.

[3] M. J. Gander, M. Al-Khaleel, and A. E. Ruehli, "Corrections to optimized waveform relaxation methods for longitudinal partitioning of transmission lines [aug 09 1732-1743]," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 1, pp. 312–312, 2010.

[4] J. K. White and A. L. Sangiovanni-Vincentelli, *Relaxation techniques for the simulation of VLSI circuits*. Springer New York, NY, 2012.

[5] E. Lelarasmee, A. Ruehli, and A. Sangiovanni-Vincentelli, "The waveform relaxation method for time-domain analysis of large scale integrated circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 1, no. 3, pp. 131–145, 1982.

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

[7] V. Loggia, S. Grivet-Talocia, and H. Hu, "Transient simulation of complex high-speed channels via waveform relaxation," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, no. 11, pp. 1823–1838, 2011.

[8] "IdEM R2018, Dassault Systèmes." [Online]. Available: www.3ds.com/products-services/simulia/products/idem/