

Received 7 November 2023, accepted 17 January 2024, date of publication 30 January 2024, date of current version 7 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359853

RESEARCH ARTICLE

Structured Model Order Reduction of System-Level Power Delivery Networks

ANTONIO CARLUCCI¹, (Graduate Student Member, IEEE),
STEFANO GRIVET-TALOCIA¹, (Fellow, IEEE), SIDDHARTH KULASEKARAN², (Member, IEEE),
AND KALADHAR RADHAKRISHNAN², (Senior Member, IEEE)

¹Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy

²Intel Corporation, Chandler, AZ 85226, USA

Corresponding author: Antonio Carlucci (antonio.carlucci@polito.it)

The work of Antonio Carlucci and Stefano Grivet-Talocia was supported by Intel Corporation under the 2022-23 Intel Strategic Research Segment (SRS) Grant titled as “API-S: Accelerated system-Level Transient Power Integrity Solvers.”

ABSTRACT This paper proposes a comprehensive model order reduction framework to enable fast power integrity verification at the system level. This approach is developed to compress models of complete power delivery networks of high-end multiprocessor systems, where electromagnetic models of board and package are connected through banks of per-core Fully Integrated Voltage Regulators to chip models and loads in a closed-loop configuration. Due to complexity in both dynamical behavior and number of signals to be monitored, a direct transient simulation at the system level is very challenging. We show that a careful topological formulation of the circuit equations leads to a global model format that enables a structured projection framework for the elimination of the redundant states. Within this framework, we present and compare two alternative approaches based on approximate interpolation and empirical balancing, here adapted for the application at hand. In both cases, the resulting system is proven to be unconditionally stable both in open and in closed-loop configuration. Transient simulation of the reduced system provides a speedup exceeding $100\times$ with respect to SPICE.

INDEX TERMS Power integrity, integrated voltage regulators, model order reduction, moment matching, balanced truncation, macromodeling.

I. INTRODUCTION

Advancements in computational performance of modern high-end microprocessors is mainly driven by combining increasingly many processing cores to parallelize and distribute workload. This trend results from technological limits in power density and area, that have ultimately determined the slowdown in frequency scaling and single-core performance [1]. Consequently, microprocessors designed for pervasive applications such as high-performance computing (HPC) and artificial intelligence (AI) are necessarily many-cores platforms, with possibly hundreds of cores.

The associate editor coordinating the review of this manuscript and approving it for publication was Michel Nakhla.

Power delivery networks (PDNs) for many-cores systems take advantage of Fully-Integrated Voltage Regulators (FIVRs) as an innovative architectural solution where multiple voltage regulators are integrated on the package and die, effectively providing a second stage of regulation after the main platform Voltage Regulator (VR). This dual-stage topology has been introduced to maintain efficiency in spite of growing power levels and to enable fine-grain power management over a large number of independent power domains in different cores. FIVRs consist in an array of multi-phase integrated Buck switching regulators, whose purpose is to deliver a stably regulated and filtered voltage to individual cores. Per-core feedback controllers modulate the instantaneous duty cycle in the respective converter so as

to reject load voltage variations. Key components of FIVRs are in-package inductors and on-die capacitors used to realize the output low-pass filter, whose size is constrained to fit in a small volume. The distributed nature of this architecture makes design more difficult and inherently system-level, as it involves the board, package and die levels at the same time. Similarly, verification of such large-scale complex systems becomes a challenging task, because unintended electromagnetic interaction between distinct cores becomes inevitable as many components operate in close vicinity. Thus, performance verification aimed at certifying desired specifications in terms of voltage droop and power noise requires a system-level analysis in time-domain where all subsystems are simulated concurrently.

System-level simulations are set up by joining highly-accurate models of individual parts, which are usually extracted from commercial electromagnetic solvers and available as tabulated S-parameters data or lumped equivalent circuits (SPICE netlist) [2]. In case of S-parameters data, typical workflows rely on rational fitting to build macromodels that are instantiated in the larger system-level netlist. These individual models are accurate representations of the corresponding physical components, thus making the resulting numerical simulation computationally heavy and time-consuming, even if performed with state-of-the-art circuit solvers such as HSPICE. This represents an issue that is still to be addressed, as computational solutions for fast transient simulation tailored for multi-core power delivery networks have not been satisfactorily addressed in the literature yet. The present work contributes to fill this gap by leveraging model reduction (MOR) to simplify system complexity and ultimately bring transient analysis runtime to a fraction of what is currently required. The same issue has been addressed in [3] and [4], where reduced models are constructed through a compressed and parameterized black-box method, and [5], where structured projection is used instead.

This work presents an intrusive, projection-based reduction method whose key novelties are

- a structured approach where the reduced PDN model retains the FIVR switches as a separate and unaffected system block, as only the linear components of the PDN are compressed. This effectively enables adapting linear MOR to a system with mild nonlinearities (due to duty cycle modulation). In addition, this avoids neglecting the nonlinear effects of the switches in transient simulation. In this regard, it fundamentally differs from [4], where parameterization is used to construct a black-box surrogate to model duty cycle modulation.
- scalability and applicability to many-cores platforms thanks to a computationally cheap formulation that operates directly on the network equations as obtained via Modified Nodal Analysis (MNA). Building a reduced model only requires performing frequency-domain AC analyses of the full-order model. For this reason,

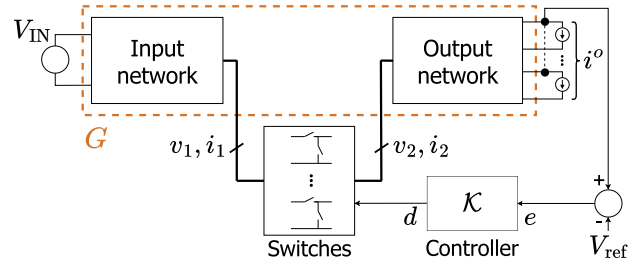


FIGURE 1. Topology of a generic power delivery network with an integrated switching regulation stage.

it represents an extension of [5], where the structured approach depends on a state-space representation, possibly difficult to obtain and handle. Moreover, the moment-matching method of [5] is here improved using balancing-based reduction to make the scheme more efficient and realization-independent, hence more robust. Furthermore, fundamental properties such as exact DC behavior and passivity are guaranteed by construction.

Compared to related projection-based MOR algorithms, the present work leverages the particular system topology to devise a method where a) reduction of a descriptor form is carried out without computing spectral projectors as in the theory of [6]; b) passivity preservation is achieved at a reduced computational cost with respect to [7], since solution of generic large-scale Lur'e equations is avoided; c) a randomized SVD scheme is employed to limit the memory footprint of projection matrices computation.

With these features, we can demonstrate a fast transient solver with a computational speedup of 50-200 \times compared to the full-order model and up to 230 \times with respect to HSPICE in realistic power integrity transient analyses, with error in the order of 5 – 10 mV or less, compatibly with the accuracy required in the application at hand. Results of transient analyses using the proposed reduced models yield essentially the same information as the full-order model in terms of maximum voltage droop, inter-core coupling noise, transient response and DC characteristic.

II. NOTATION

We denote scalar variables as x , vector variables as \mathbf{x} or \mathbf{X} , and matrices as \mathbf{X} or \mathcal{X} , with the latter notation reserved for block matrices related to small-signal linearized system representations. The all-zero and identity matrices are denoted as $\mathbf{0}$ and \mathbb{I} , respectively. We will adopt a MATLAB-consistent notation to denote horizontal stacking ($\mathbf{X}_1, \mathbf{X}_2$) and vertical stacking ($\mathbf{X}_1; \mathbf{X}_2$) of matrix blocks of compatible size. Vertical stacking will also be denoted through the operator $\text{col}\{\cdot\}$, whereas $\text{diag}\{\cdot\}$ and $\text{blkdiag}\{\cdot\}$ will denote (block) diagonal stacking. Also, we denote with $\mathbf{X}_{(r_1:r_2, c_1:c_2)}$ the submatrix of \mathbf{X} formed by its rows from r_1 to r_2 and its columns from c_1 to c_2 , with the operator: alone selecting all rows/columns. The standard Kronecker

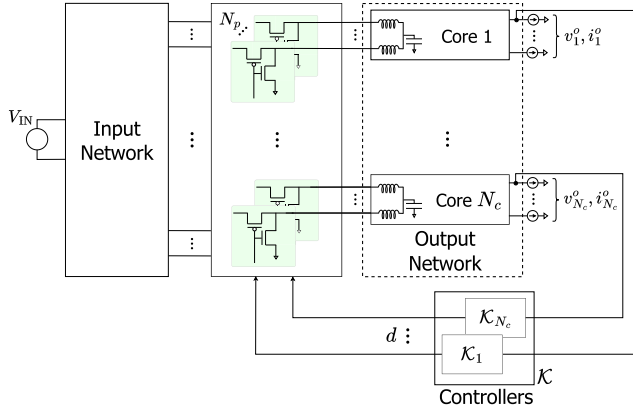


FIGURE 2. Detailed block schematic of FIVR-based multi-core power delivery network, with expanded per-core view of the output network, switches and controllers.

matrix product is \otimes and operator $\text{orth}\{\mathbf{X}\}$ produces a matrix whose orthogonal columns span the range $\mathcal{R}\{\mathbf{X}\}$ of matrix \mathbf{X} . Finally, for a generic descriptor system, we will use the alternative notation

$$\left(\begin{array}{c|c} \mathbf{E}, \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right) \Leftrightarrow \begin{cases} \mathbf{E}\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \end{cases} \quad (1)$$

with \mathbf{A} replacing \mathbf{E} , \mathbf{A} for regular state-space with $\mathbf{E} = \mathbb{I}$.

III. PROBLEM STATEMENT

This paper considers the system topology depicted in Fig. 1, representing a power delivery network (PDN) whose architecture includes a voltage regulation stage implemented through switching converters. The main supply to the power delivery network is provided by the ideal (constant) input voltage V_{IN} . On the output side, the PDN delivers power to multiple loads, represented by ideal time-dependent current sources. Load voltage regulation relies on a bank of switching regulators that are here assumed to be PWM-controlled Buck converters with fixed switching period and time-varying duty cycle. The duty cycles $\mathbf{d}(t)$ are synthesized by a feedback controller \mathcal{K} , based on the error signal $\mathbf{e}(t)$ between the actual load voltage and a predefined target value V_{ref} .

Besides the switching and control circuitry, a system-level PDN description includes suitable models of all passive components and parasitic elements, represented in Fig. 1 through the *input* and *output network* blocks. Unlike the converter switches, which introduce a nonlinear duty cycle modulation effect, both input and output networks are Linear and Time-Invariant (LTI). This motivates the system partitioning evidenced in Fig. 1, where all LTI components are collectively viewed as a large subsystem \mathbf{G} , while the converter switches constitute a separate block. This decomposition underpins the structured approach of this work, whereby methods are provided to build compact models of \mathbf{G} that are accurate and stable when operating in the closed-loop configuration of Fig. 1.

This topology captures the essential features of PDNs of many-core architectures with per-core Fully-Integrated Voltage Regulators (FIVRs), which is the application that motivates this work (see Fig. 2). The reduction methods outlined here can be directly employed to formulate and compress the system equations in order to enable a fast transient solver for system-level prediction of the PDN performance under realistic load current transients.

In order to set notation, we define with reference to Fig. 2

- N_c number of cores;
- N_o number of load ports (voltages to be monitored) for each core;
- $P = N_c N_o$ total number of output voltages to be monitored;
- N_p number of phases of each FIVR.

Each building block is now described in more detail.

A. INPUT NETWORK

The block denoted as *Input network* in Figs. 1–2 represents the electrical behavior of the board and the package up to the interface with the FIVR switches. This block is intended to model all parasitics and resonances of the combined board/package PDN portion, together with its on-board and on-package decoupling capacitors and a linear model of the platform VRM. This block is equipped with $1 + N_c N_p$ ports since connected with V_{IN} and with each FIVR phase of each core. The input model is thus a large-scale Linear Time-Invariant (LTI) system, which can be represented by a frequency-domain multi-input multi-output (MIMO) transfer function $\mathbf{G}_{\text{in}}(s)$ through

$$\begin{pmatrix} I_{\text{IN}} \\ \mathbf{V}_1 \end{pmatrix} = \mathbf{G}_{\text{in}}(s) \begin{pmatrix} V_{\text{IN}} \\ \mathbf{I}_1 \end{pmatrix} \quad (2)$$

where $\mathbf{V}_1, \mathbf{I}_1 \in \mathbb{C}^{N_c N_p}$.

The characterization of the input model is available in terms of its constitutive parts:

- an electromagnetic characterization of the coupled PDN interconnect (excluding VRM and decoupling capacitors), available through frequency-domain scattering matrix samples over the bandwidth of interest computed by a full-wave Maxwell equation solver. This description is converted to a state-space macromodel in a pre-processing phase, using standard passivity-constrained rational fitting algorithms [8].
- a set of models of decoupling capacitors, which can be available either as RLC subcircuits or as frequency response samples, in which case a macromodeling step is used to derive associated state-space descriptions.
- a linear VRM model, expressed as an RL subcircuit.

It can be easily shown (see Appendix A) that assembling the above building blocks using a standard Modified Nodal Analysis (MNA) stamping leads to the general descriptor form

$$\mathbf{G}_{\text{in}}(s) : \begin{cases} \mathbf{E}_{\text{in}} \dot{\mathbf{x}}_{\text{in}} = \mathbf{A}_{\text{in}} \mathbf{x}_{\text{in}} + \mathbf{B}_{\text{in}} \mathbf{u}_{\text{in}} \\ \mathbf{y}_{\text{in}} = \mathbf{C}_{\text{in}} \mathbf{x}_{\text{in}} + \mathbf{D}_{\text{in}} \mathbf{u}_{\text{in}} \end{cases} \quad (3)$$

where

$$\mathbf{u}_{\text{in}} \triangleq \begin{pmatrix} V_{\text{IN}} \\ \mathbf{i}_1 \end{pmatrix}, \quad \mathbf{y}_{\text{in}} \triangleq \begin{pmatrix} I_{\text{IN}} \\ \mathbf{v}_1 \end{pmatrix}. \quad (4)$$

Note that the descriptor matrix \mathbf{E}_{in} is generally singular, so that (3) is a Differential Algebraic system of equations (DAE).

B. OUTPUT NETWORK

Each individual core is described by a corresponding output model, which represents the PDN portion between the FIVR switches and the loading current sources (see Fig. 2). The collection of all N_c core-wise output models forms the output network of Fig. 1. Each output model includes

- detailed electrical models of integrated inductors, which are part of the output filtering network of each FIVR. Such models are here available in terms of frequency scattering samples computed by a full-wave Maxwell equation solver, applied to the coupled coils related to all phases of any individual FIVR. Such samples are converted to passive rational macromodels and are therefore available in state-space form.
- on-die integrated passive component models, including MIM capacitors that complete the FIVR filtering network, as well as PDN die models. All such components are again described by R(L)C subcircuits.

The output model for the k -th core is therefore a passive LTI systems represented by a MIMO transfer function $\mathbf{G}_k(s)$ through

$$\begin{pmatrix} \mathbf{I}_{2,k} \\ \mathbf{V}_k^o \end{pmatrix} = \mathbf{G}_k(s) \begin{pmatrix} \mathbf{V}_{2,k} \\ \mathbf{I}_k^o \end{pmatrix} \quad (5)$$

where $\mathbf{V}_{2,k}, \mathbf{I}_{2,k} \in \mathbb{C}^{N_p}$ collect the interface variables to the switches and $\mathbf{V}_k^o, \mathbf{I}_k^o \in \mathbb{C}^{N_o}$ are the voltages and currents at the k -th core load ports. Following a similar MNA stamping procedure as for the input network, the output models are available in descriptor form

$$\mathbf{G}_k(s) : \begin{cases} \mathbf{E}_k \dot{\mathbf{x}}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k \\ \mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{D}_k \mathbf{u}_k \end{cases} \quad k = 1, \dots, N_c \quad (6)$$

where

$$\mathbf{u}_k \triangleq \begin{pmatrix} \mathbf{v}_{2,k} \\ \mathbf{i}_k^o \end{pmatrix}, \quad \mathbf{y}_k = \begin{pmatrix} \mathbf{i}_{2,k} \\ \mathbf{v}_k^o \end{pmatrix} \quad (7)$$

C. FIVR CONTROLLERS

The controllers are responsible for driving the switches in a feedback loop by sensing the output voltage(s) to be regulated. For the k -th core, an error signal quantifying the difference between the regulated load voltage and a target reference V_{ref} is defined as

$$e_k(t) = \mathbf{N}_k \mathbf{v}_k^o - V_{\text{ref}}, \quad (8)$$

where $\mathbf{N}_k \in \mathbb{R}^{1 \times N_o}$ is a selector vector that extracts a single port voltage. Such error signal is processed by a compensator, whose output is the duty cycle signal $d_k(t)$ of the FIVR

switches. The controllers are then described by the following state-space realizations

$$\begin{cases} \dot{\mathbf{x}}_{\mathcal{K},k} = \mathbf{A}_{\mathcal{K},k} \mathbf{x}_{\mathcal{K},k} + \mathbf{B}_{\mathcal{K},k} \mathbf{e}_k \\ d_k = \mathbf{C}_{\mathcal{K},k} \mathbf{x}_{\mathcal{K},k} \end{cases} \quad k = 1, \dots, N_c \quad (9)$$

whose coefficients are assumed to be known.

D. SWITCHES

Power transistor switches are key components of Buck regulators, which require a hierarchy of models for various stages of design and verification. Although efficiency investigations require detailed transistor-level descriptions, much simpler models are sufficient for system-level power integrity analysis and verification. In this context, the natural description of the switching bank of the multiphase FIVR associated to a single core k is

$$\mathbf{w}_k = \Delta_k \mathbf{z}_k, \quad \mathbf{w}_k \triangleq \begin{pmatrix} \mathbf{i}_{1,k} \\ \mathbf{v}_{2,k} \end{pmatrix}, \quad \mathbf{z}_k \triangleq \begin{pmatrix} \mathbf{v}_{1,k} \\ \mathbf{i}_{2,k} \end{pmatrix} \quad (10)$$

where Δ_k is a $2N_p$ -port hybrid matrix representation. In case an accurate description of the voltage ripple induced by the PWM switching pattern is desired, a time-dependent, piecewise constant model for Δ_k can be used. Instead, if only a system-level evaluation of the voltage droop and long-term transient response is required as induced by load current transients, a low-frequency averaged switch model is sufficient. This is the model herein adopted, which basically amounts to representing each pair of switches through an ideal transformer having its turn ratio equal to the instantaneous duty cycle. Under this assumption, we have

$$\Delta_k = \begin{pmatrix} \mathbf{0} & -d_k \mathbb{I}_{N_p} \\ d_k \mathbb{I}_{N_p} & \mathbf{0} \end{pmatrix} \quad (11)$$

E. FULL SYSTEM DESCRIPTION

We now assemble all individual submodels representing the input network, the output models associated to all cores with the corresponding averaged switches models and the controllers. We first collect all relevant circuit variables in the following global vectors

$$\begin{aligned} \mathbf{i}^o &= \text{col}\{\mathbf{i}_k^o\}_{k=1}^{N_c} & \mathbf{v}^o &= \text{col}\{\mathbf{v}_k^o\}_{k=1}^{N_c} \\ \mathbf{i}_1 &= \text{col}\{\mathbf{i}_{1,k}\}_{k=1}^{N_c} & \mathbf{v}_1 &= \text{col}\{\mathbf{v}_{1,k}\}_{k=1}^{N_c} \\ \mathbf{i}_2 &= \text{col}\{\mathbf{i}_{2,k}\}_{k=1}^{N_c} & \mathbf{v}_2 &= \text{col}\{\mathbf{v}_{2,k}\}_{k=1}^{N_c} \end{aligned} \quad (12)$$

Combining (3) with (6)-(9) and (11) for $k = 1, \dots, N_c$ leads to the complete system description

$$\mathbf{E} \dot{\mathbf{x}} = \mathbf{A} \mathbf{x} + \mathbf{B}_w \mathbf{w} + \mathbf{B}_u \mathbf{u} \quad (13a)$$

$$\mathbf{z} = \mathbf{C}_z \mathbf{x} + \mathbf{D}_{zw} \mathbf{w} + \mathbf{D}_{zu} \mathbf{u} \quad (13b)$$

$$\mathbf{y} = \mathbf{C}_y \mathbf{x} + \mathbf{D}_{yw} \mathbf{w} + \mathbf{D}_{yu} \mathbf{u} \quad (13c)$$

$$\mathbf{w} = \Delta(\mathbf{d}) \mathbf{z} \quad (13d)$$

where

$$\mathbf{w} \triangleq \begin{pmatrix} \mathbf{i}_1 \\ \mathbf{v}_2 \end{pmatrix}, \quad \mathbf{z} \triangleq \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{i}_2 \end{pmatrix} \quad (14)$$

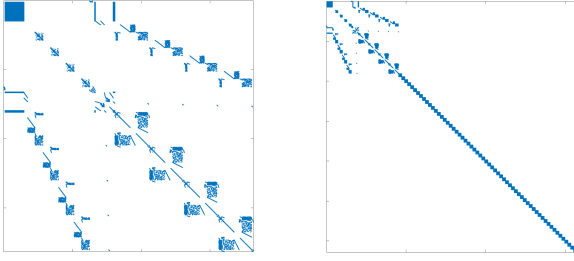


FIGURE 3. Left panel: sparsity pattern of the input network \mathbf{A}_{in} matrix in (3) for a 60-core enterprise server system (see Sec. VIII-B). Only 0.6% of all entries are nonzero. Right panel: sparsity pattern of the \mathbf{A} matrix in (13) for the same testcase.

collect the “internal” variables at the switch interface,

$$\mathbf{u} \triangleq \begin{pmatrix} V_{\text{IN}} \\ \mathbf{i}^o \end{pmatrix}, \quad \mathbf{y} \triangleq \begin{pmatrix} I_{\text{IN}} \\ \mathbf{v}^o \end{pmatrix} \quad (15)$$

define input and output variables, respectively, and

$$\Delta(\mathbf{d}) = \begin{pmatrix} \mathbf{0} & -\Delta'(\mathbf{d}) \\ \Delta'(\mathbf{d}) & \mathbf{0} \end{pmatrix} \quad (16)$$

where

$$\mathbf{d} = \text{col}\{\mathbf{d}_k\}_{k=1}^{N_c}, \quad \Delta'(\mathbf{d}) = \text{diag}\{\mathbf{d}\} \otimes \mathbb{I}_{N_p}. \quad (17)$$

The differential-algebraic system (13) can be interpreted as an open-loop description of the regulated PDN where the duty cycle signals collected in \mathbf{d} play the role of external inputs. This system must be complemented with the feedback control, represented collectively for all cores as

$$\dot{\mathbf{x}}_{\mathcal{K}} = \mathbf{A}_{\mathcal{K}} \mathbf{x}_{\mathcal{K}} + \mathbf{B}_{\mathcal{K}} \mathbf{e} \quad (18a)$$

$$\mathbf{e} = \mathbf{N} \mathbf{v}^o - \mathbf{V}_{\text{ref}} \quad (18b)$$

$$\mathbf{d} = \mathbf{C}_{\mathcal{K}} \mathbf{x}_{\mathcal{K}} \quad (18c)$$

where

$$\mathbf{N} = \text{col}\{\mathbf{N}_k\}_{k=1}^{N_c} \quad (19)$$

and \mathbf{V}_{ref} replicates the reference voltage V_{ref} into a column vector of size N_c . The cumulative state vector $\mathbf{x}_{\mathcal{K}}$ in (18) is the column stacking of the per-core vectors $\mathbf{x}_{\mathcal{K},k}$ used in (9).

F. DISCUSSION

It is easy to show that the descriptor matrix $\mathbf{E} \in \mathbb{R}^{n \times n}$ in (13) is obtained by stacking as diagonal blocks the individual submodel matrices \mathbf{E}_{in} and \mathbf{E}_k for $k = 1, \dots, N_c$. The same applies to \mathbf{A} and to the various other matrices in (13) and (18). The procedure of Appendix A can be further applied to split dynamic and algebraic states as $\mathbf{x} = (\mathbf{x}_d; \mathbf{x}_a)$, with $\mathbf{x}_d \in \mathbb{R}^{n_d}$, $\mathbf{x}_a \in \mathbb{R}^{n-n_d}$, obtaining the following canonical partition [9]

$$\begin{aligned} \mathbf{E} &= \begin{pmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \\ \mathbf{B} &= \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}, \quad \mathbf{C} = (\mathbf{C}_1 \quad \mathbf{C}_2). \end{aligned} \quad (20)$$

This decomposition is obtained with cheap algebraic operations (see Appendix A) and leads to a structured descriptor system where matrix $\mathbf{E}_{11} \in \mathbb{R}^{n_d \times n_d}$ is nonsingular and symmetric, and all structured descriptor matrices are sparse (see Fig. 3 for an example). For example, using the canonical partition (20) to represent the system (13) leads to a matrix \mathbf{B} with $(N_c N_o + 1) + 2N_c N_p$ columns, because it corresponds to horizontal concatenation of the $2N_c N_p$ columns of \mathbf{B}_w and the $(N_c N_o + 1)$ columns of \mathbf{B}_u . Similarly, \mathbf{C} would have the same number of rows as the columns of \mathbf{B} .

An additional assumption made throughout this work whenever the partitioning in (20) is used, is that the corresponding system is impulse-free. In particular, we assume that either \mathbf{A}_{22} is an invertible matrix (in this case (\mathbf{E}, \mathbf{A}) is said to have index 1), or any infinite dynamic mode of (\mathbf{E}, \mathbf{A}) is uncontrollable and unobservable [10]. This technical condition is practically verified for all RLC circuits that do not contain cuts of inductors and current sources, or loops of capacitors and voltage sources [11]. As detailed in Appendix A-B, the same property carries over to all systems discussed in this work, since they arise from the combination of MNA descriptions of RLC networks and passive macromodels.

One of the objectives of this work is to apply and compare various MOR approaches to reduce model complexity and simulation runtime, while avoiding any operation that requires dense linear algebra operations at the large scale. This is essential to obtain efficient reduction schemes. For this reason, although elimination of algebraic states is possible (see Appendix B) to obtain a large-scale standard state-space description, this operation will be never performed here since the obtained system matrices would be dense. In this respect, this work differentiates from and generalizes [5], where regular state-space descriptions were assumed. System (13) in canonical form (20) combined with (18) provides the starting point of our MOR formulation, and will be assumed in all subsequent derivations. The dynamic states in this system are \mathbf{x}_d and $\mathbf{x}_{\mathcal{K}}$, whereas the algebraic states include \mathbf{x}_a , \mathbf{z} , and \mathbf{w} .

IV. MODEL REDUCTION STRATEGY AND LINEARIZATION

The adoption of averaged switch models simplifies both the system description and the consequent numerical simulation, since the time-varying and possibly discontinuous behavior of the switching operator $\Delta(\mathbf{d})$ is eliminated. However, the coupled equations (13)-(18) are nonlinear, due to the product $\Delta(\mathbf{d})\mathbf{z}$ between the duty cycle variables \mathbf{d} and the electrical quantities in \mathbf{z} . The overall system is therefore a nonlinear differential-algebraic set of equations, where the nonlinearity is quadratic (bilinear). Linear reduction approaches are therefore not directly applicable, and more general nonlinear MOR schemes would be required.

In this work, we circumvent the above difficulties by

- 1) performing a structured order reduction, applied only to the linear components of (13) and (18). The averaged

switch equations (13d) are preserved by the reduction scheme;

- 2) the reduction of the linear components will consider the constraints induced by the interconnection of all blocks, including the averaged switch characteristics. This procedure is justified by the fact that the closed-loop system dynamics are severely restricted with respect to the open-loop dynamics, described by the individual blocks acting alone. The loading effects of output networks on the input network through the control loops can be shown to constrain the large-scale state-space onto a significantly lower-dimensional manifold. We attempt to identify such manifold through a projection framework.
- 3) The projection subspaces used in the proposed linear structured MOR are constructed based on a linearization process around a nominal operating point. This corresponds to embedding the reduced-order dynamics into a linear subspace \mathcal{L} .
- 4) In the transient simulation phase, the complete (non-linear) reduced-order system is solved, so that the only approximation step is the projection of the large-scale linear subsystem onto the identified linear subspace \mathcal{L} .

The following sections V-VII will provide a detailed formulation of above steps 1,2,4. The linearization process in step 3 is instead addressed below.

A. LINEARIZATION

Linearization to infer approximate dynamics is here justified based on physical considerations. In fact, any voltage regulator, including switching converters, is designed to stabilize the output voltage. By design, such structures are developed to make all signals have small fluctuations around some nominal operating point. In such conditions, small-signal linearized models are appropriate. In fact, small-signal analysis is routinely employed for analysis and controller design [12].

System equations are linearized by replacing the last line in (13) with its Taylor expansion around an operating point identified by the constant steady-state quantities $(\bar{\mathbf{d}}, \bar{\mathbf{z}})$

$$\mathbf{w} \approx \Delta(\bar{\mathbf{d}})\bar{\mathbf{z}} + \Delta(\bar{\mathbf{d}})(\mathbf{z} - \bar{\mathbf{z}}) + \Delta(\bar{\mathbf{d}} - \bar{\mathbf{d}})\bar{\mathbf{z}}. \quad (21)$$

The nominal operating point is computed by assuming a nominal loading condition $\mathbf{i}^o = \mathbf{I}_{\text{ref}}^o$ and solving the DC steady-state problem associated to (13)-(18).

In the linearization process, any variable ζ is decomposed as

$$\zeta = \bar{\zeta} + \tilde{\zeta} \quad (22)$$

where $\tilde{\zeta}$ is a small-signal component around the DC bias $\bar{\zeta}$. Using these definitions and (21) in (13)-(18), while subtracting the DC bias terms, leads to the linearized small-signal model, expressed in the following descriptor

form

$$\begin{cases} \mathcal{E} \dot{\tilde{\chi}} = \mathcal{A} \tilde{\chi} + \mathcal{B} \tilde{\mathbf{u}} \\ \tilde{\mathbf{v}}^o = \mathcal{C} \tilde{\chi} + \mathcal{D} \tilde{\mathbf{u}} \end{cases}, \quad \tilde{\chi} \triangleq \begin{pmatrix} \tilde{\mathbf{x}}_d \\ \tilde{\mathbf{x}}_{\mathcal{K}} \\ \tilde{\mathbf{x}}_a \\ \tilde{\mathbf{z}} \end{pmatrix} \quad (23)$$

where the small-signal state vector $\tilde{\chi}$ collects all states and variables $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{d}}$ are eliminated through (21) and (18). Note that (23) retains the structure (20), in particular $\mathcal{E} = \text{blkdiag}\{\mathbf{E}_{11}, \mathbb{I}_{n_{\mathcal{K}}}, \mathbf{0}_{n-n_d}, \mathbf{0}_{n_z}\}$. The small-signal transfer function corresponding to (23) is denoted as $\mathcal{H}(s) = \mathcal{D} + \mathcal{C}(s\mathcal{E} - \mathcal{A})^{-1}\mathcal{B}$. Note that only the small-signal output voltages $\tilde{\mathbf{v}}^o$ are retained, as the only quantities of interest.

A remark on notation is in order: we denote with calligraphic fonts (\mathcal{E}, \mathcal{A}) quantities related to the small-signal linearized system (23) and with bold upright fonts (\mathbf{E}, \mathbf{A}) quantities related to the original system (13)-(18). Note that the size of the dynamic states in these two cases is different, since the linearized system (23) includes also the small-signal components of $\tilde{\mathbf{x}}_a$ and $\tilde{\mathbf{z}}$.

V. MODEL REDUCTION

This section develops the proposed model reduction approach. The linear-nonlinear decomposition described in Sec. III shows that the dynamic complexity of the system (13) is entirely due to the linear subnetwork denoted as \mathbf{G} in Fig. 1 and represented by (13a)-(13c). Hence, a straightforward approach would be to replace \mathbf{G} with a reduced-order system $\hat{\mathbf{G}}$ using any of the available MOR algorithms. This, however, would require treating \mathbf{G} as an isolated linear block, thus neglecting that the signals at its ports are constrained by (13d). Our proposed approach constructs $\hat{\mathbf{G}}$ by taking this feedback into consideration, in order to optimize model accuracy in the closed-loop configuration.

We propose and compare two alternative MOR schemes, based on moment-matching [13] and balanced truncation [14], [15], respectively. Both approaches fit in the classical projection framework [16], so that $\hat{\mathbf{G}}$ is obtained by a Petrov-Galerkin projection of (13). A pair of matrices $\mathbf{W}, \mathbf{V} \in \mathbb{R}^{n \times q}$ with $q \ll n$ defines the reduced model via the following transformations

$$\begin{aligned} \hat{\mathbf{E}} &= \mathbf{W}^T \mathbf{E} \mathbf{V}, & \hat{\mathbf{A}} &= \mathbf{W}^T \mathbf{A} \mathbf{V}, \\ \hat{\mathbf{B}}_{w,u} &= \mathbf{W}^T \mathbf{B}_{w,u}, & \hat{\mathbf{C}}_{z,y} &= \mathbf{C}_{z,y} \mathbf{V}, \end{aligned} \quad (24)$$

with direct coupling blocks $\hat{\mathbf{D}} = \mathbf{D}$ left unchanged. In this projection, the dynamic-algebraic decomposition in (20) is preserved by constructing \mathbf{W}, \mathbf{V} to have the following block-diagonal structure

$$\mathbf{V} = \text{blkdiag}\{\mathbf{V}_d, \mathbb{I}_{n-n_d}\}, \quad \mathbf{W} = \text{blkdiag}\{\mathbf{W}_d, \mathbb{I}_{n-n_d}\} \quad (25)$$

with $\mathbf{V}_d, \mathbf{W}_d \in \mathbb{R}^{n_d \times q}$. Consequently, the canonical form (20) continues to hold for the reduced system, whose dynamic part is reduced to dimension q , and the algebraic part is left unaffected. The choice of reducing only the dynamic states is non-standard, as most well-known projection-based

methods do not assume any particular structure as (25) for the projection matrices. This choice is motivated by the following considerations:

- the reduced model must match exactly the DC response of the original model. This constraint can be enforced easily with a standard (unstructured) projection, but with a significantly higher computational cost and a much less efficient reduction. This point will be discussed in detail in Sec. V-C;
- balanced truncation of singular descriptor systems as (13) requires some technical assumptions and operations, which may be numerically challenging. The adopted structured projection avoids such difficulties by construction. This is discussed in Sec. V-B.

We follow the usual practice of constructing the right projection matrix \mathbf{V} solely based on model accuracy considerations, and then selecting the left projection matrix \mathbf{W} to enforce closed-loop stability (see Sec. V-D). The construction of \mathbf{V} is discussed in a moment matching (interpolation) framework in Sec. V-A and in an empirical balanced truncation framework in Sec. V-B. In both cases, we search for a dominant subspace \mathcal{L} to embed the reduced-order dynamics. Such subspace is constructed based on the linearized model (23), which is assumed as a proxy for the true system (13) under small-signal operating conditions around the nominal operating point.

A. APPROXIMATE INTERPOLATION

This section proposes a reduction approach that identifies a matrix \mathcal{V} whose range embeds the dominant dynamics of the linearized system (23). This is achieved through a moment-matching approach where only the first moment is enforced, resulting in an interpolation condition. Only the components of \mathcal{V} along the dynamic states of (13) are retained to define the right projection matrix \mathbf{V} .

Let us consider a set of K points $s_1, \dots, s_K \in \mathbb{C}$. It is well known [13] that, if $\mathcal{R}\{(s_k \mathcal{E} - \mathcal{A})^{-1} \mathcal{B}\} \subset \mathcal{R}\{\mathcal{V}\}$, then the reduced system response $\hat{\mathcal{H}}(s)$ defined as

$$\hat{\mathcal{H}}(s) = \left(\frac{\mathcal{W}^T \mathcal{E} \mathcal{V}, \mathcal{W}^T \mathcal{A} \mathcal{V}}{\mathcal{C} \mathcal{V}} \middle| \frac{\mathcal{W}^T \mathcal{B}}{\mathcal{D}} \right) \quad (26)$$

interpolates the original system response as $\hat{\mathcal{H}}(s_k) = \mathcal{H}(s_k)$. Let us consider a matrix \mathcal{M} collecting frequency-domain snapshots $\mathcal{X}_k = \mathcal{X}(j\omega_k) \triangleq (j\omega_k \mathcal{E} - \mathcal{A})^{-1} \mathcal{B}$ evaluated at K frequencies $\{\omega_k\}_{k=1}^K$,

$$\mathcal{M} \triangleq (\mathcal{X}_1 \quad \dots \quad \mathcal{X}_K). \quad (27)$$

This matrix is partitioned conformably with the structure of $\tilde{\mathcal{X}}$ in (23),

$$\mathcal{M} = \begin{pmatrix} \mathbf{M}_d \\ \mathbf{M}_{\mathcal{K}} \\ \mathbf{M}_a \\ \mathbf{M}_{\mathbf{z}} \end{pmatrix}. \quad (28)$$

The first block \mathbf{M}_d of n_d rows deserves particular attention, because it corresponds to the dynamic part of the linear

subsystem \mathbf{G} whose dominant subspace is sought for. We define

$$\mathbf{V}_d = \text{orth}\{\mathbf{M}_d\} \quad (29)$$

and we construct the right projection matrix \mathbf{V} as in (25). Finally, we define

$$\mathcal{V} = \text{blkdiag}\{\mathbf{V}_d, \mathbb{I}_{n_{\mathcal{K}}}, \mathbb{I}_{n-n_d}, \mathbb{I}_{n_z}\}, \quad (30)$$

$$\mathcal{W} = \text{blkdiag}\{\mathbf{W}_d, \mathbb{I}_{n_{\mathcal{K}}}, \mathbb{I}_{n-n_d}, \mathbb{I}_{n_z}\} \quad (31)$$

for any left projection matrix \mathbf{W} of compatible size. The interpolation condition $\hat{\mathcal{H}}(s_k) = \mathcal{H}(s_k)$ follows from the observation that the inclusion

$$\mathcal{R}\{\mathcal{V}\} \supset \mathcal{R}\{\mathcal{M}\} \quad (32)$$

holds by construction.

Fulfillment of this exact interpolation condition might come at the price of large dimension of the subspace $\mathcal{R}\{\mathbf{V}_d\}$, which is generically $N_o K$. This would lead to an unnecessarily large order of the reduced model. Therefore, we pursue only an approximate interpolation condition by taking $\mathcal{R}\{\mathbf{V}_d\}$ to be the subspace spanned by the leading singular vectors of \mathbf{M}_d . Using the Singular Value Decomposition (SVD) of \mathbf{M}_d ,

$$\mathbf{M}_d = (\mathbf{U}_1 \ \mathbf{U}_2) \text{blkdiag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\} \mathbf{Z}^T \quad (33)$$

we choose $\mathbf{V}_d = \mathbf{U}_1$, with \mathbf{U}_1 containing the first q left singular vectors. The order q is selected to enforce $\sigma_{q+1} \leq \epsilon \sigma_1$ where σ_i are the singular values in decreasing order and ϵ is the desired relative accuracy threshold. Given the large size of \mathbf{M}_d , the actual computations are here performed through a randomized SVD, see Sec. VI.

B. EMPIRICAL BALANCING

Model reduction via balanced truncation is known to be superior to interpolation-based reduction in terms of accuracy [17] and optimality [18]. In fact, explicit a priori error bounds are easily derived for balanced truncation, whereas only a-posteriori or heuristic error control is possible for interpolation-based reduction methods. Balanced truncation methods aim at finding a special set of state-space coordinates, in which the state variables that are simultaneously poorly controllable and observable (hence with a negligible contribution to the input-output response) can be safely removed. Such coordinate systems are determined based on the system Gramians, whose computation requires dense linear algebra operations. A standard balanced truncation for large-scale systems is therefore impractical. A further complication for our case of study is that we need to process a singular descriptor system (13). For such systems, balanced truncation requires the identification of spectral projectors, whose computation may be numerically problematic [19], [20]. The reader is referred to [16] for a complete overview of the state of the art.

In this work, we avoid dense large-scale linear algebra operations through an empirical (snapshot-based) balancing,

which directly computes a low-rank factorization of the Gramians using fast and cheap operations (thanks to the sparse nature of our system). Further, spectral projections are avoided by construction thanks to the adopted structured formulation and projection framework.

We define the following input-state and dual (output-state) responses

$$\mathcal{X}^c(j\omega) = (j\omega\mathcal{E} - \mathcal{A})^{-1}\mathcal{B} \quad (34)$$

$$\mathcal{X}^o(j\omega) = (j\omega\mathcal{E}^\top - \mathcal{A}^\top)^{-1}\mathcal{C}^\top \quad (35)$$

by initially assuming that \mathcal{E} is full rank. In such case, the controllability and observability Gramians can be computed [15] through the following frequency domain integrals

$$\mathcal{W}^c = \text{Re} \left\{ \frac{1}{\pi} \int_0^{+\infty} \mathcal{X}^c(j\omega) \mathcal{X}^{c\top}(-j\omega) d\omega \right\} \quad (36)$$

$$\mathcal{W}^o = \text{Re} \left\{ \frac{1}{\pi} \int_0^{+\infty} \mathcal{X}^o(j\omega) \mathcal{X}^{o\top}(-j\omega) d\omega \right\}. \quad (37)$$

The balanced coordinates where a simple reduction by truncation can be applied is provided by the eigenvectors of the product of the two Gramians [18]. When \mathcal{E} is singular, the above Gramians must be replaced by the so-called *proper* Gramians, obtained by spectral projections onto the deflating subspaces associated to the finite eigenvalues of the pencil $\lambda\mathcal{E} - \mathcal{A}$. The evaluation of these projectors is here not necessary, since we start from a block-structured representation of the system (13) in form (20), where dynamic state variables are explicit. Spectral projection simply amounts to downselecting from (34)-(35) the first $n_d + n_K$ components. However, since the controller states \mathbf{x}_K are not involved in the proposed structured projection, we further restrict the evaluation to the first n_d states, resulting in

$$\mathbf{X}_d^v(j\omega) = [\mathcal{X}^v(j\omega)]_{(1:n_d,:)}, \quad (38)$$

where $v \in \{c, o\}$ is used as a placeholder for both controllability and observability. Replacing $\mathcal{X}^v(j\omega)$ in the integrals (36)-(37) with its upper block row $\mathbf{X}_d^v(j\omega)$ leads to

$$\mathbf{W}_d^v = \text{Re} \left\{ \frac{1}{\pi} \int_0^{+\infty} \mathbf{X}_d^v(j\omega) \mathbf{X}_d^{v\top}(-j\omega) d\omega \right\} \quad (39)$$

where \mathbf{W}_d^v matches the upper $n_d \times n_d$ block of \mathcal{W}^v . The above result holds under technical assumptions that are here verified by construction. See Appendix C for details.

The evaluation of the dynamic components $\mathbf{X}_d^v(j\omega)$ at any desired frequency ω is numerically cheap due to the sparse nature of the system matrices. However, integration still needs to be performed, which is a possibly expensive operation. Following the strategy of [14], we approximate the integrals (39) through a quadrature rule

$$\mathbf{W}_d^v \approx \text{Re} \left\{ \sum_{k=1}^K \alpha_k \mathbf{X}_d^v(j\omega_k) \mathbf{X}_d^{v\top}(-j\omega_k) \right\} \quad (40)$$

based on a set of nodes $\{\omega_k, k = 1, \dots, K\}$ spread over the frequency band of interest, with weights α_k that depend on the

discretization scheme (e.g. trapezoidal rule). Note that only contributions up to the maximum frequency $\Omega = \max\{\omega_k\}$ are considered, thus effectively leading to an approximation of *bandlimited* Gramians [21]. This is a desired feature, since accuracy in model reduction is only required over the frequency band of interest $\omega \in [0, \Omega]$.

Further splitting of real and imaginary parts as $\mathbf{X}_d^v(j\omega) = \mathbf{X}_d^{v,r}(j\omega) + j\mathbf{X}_d^{v,i}(j\omega)$ leads to the following representation of the Gramians

$$\mathbf{W}_d^v \approx \mathbf{R}^v \mathbf{R}^{v\top} \quad (41)$$

through real snapshot matrices

$$\mathbf{R}^v \triangleq \left(\dots, \sqrt{\alpha_k} \mathbf{X}_d^{v,r}(j\omega_k), \sqrt{\alpha_k} \mathbf{X}_d^{v,i}(j\omega_k), \dots \right). \quad (42)$$

Thanks to (41), the factors \mathbf{R}^c and \mathbf{R}^o are used to obtain the desired right and left projection matrices \mathbf{V}_d , \mathbf{W}_d through the Square Root Balancing method [6], [22]. This procedure is standard and requires to first compute the SVD decomposition

$$\mathbf{R}^{c\top} \mathbf{E}_{11} \mathbf{R}^o = (\mathbf{U}_1 \ \mathbf{U}_2) \text{blkdiag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\} (\mathbf{Z}_1 \ \mathbf{Z}_2)^\top \quad (43)$$

where $\boldsymbol{\Sigma}_1$ contains the largest singular values up to a predefined relative accuracy threshold. The projection matrices are then computed as

$$\mathbf{V}_d = \mathbf{R}^c \mathbf{U}_1 \boldsymbol{\Sigma}_1^{-1/2}, \quad \mathbf{W}_d = \mathbf{R}^o \mathbf{Z}_1 \boldsymbol{\Sigma}_1^{-1/2} \quad (44)$$

from which we retain \mathbf{V}_d only.

Thanks to this representation, we see that the explicit calculation of the Gramians is not required, and all computations leading to the projection matrix \mathbf{V}_d can be performed through large-scale (but sparse) evaluations of the snapshots, followed by dense but low-rank decompositions. In addition, all SVD operations are here performed through a randomized SVD scheme (Sec. VI) to further limit both memory and CPU cost in the construction of \mathbf{V}_d .

C. DC ENFORCEMENT

This section shows how the reduced model can be constrained to preserve the steady-state behavior of the original system under constant excitation. For this purpose, it is sufficient to match the DC response as $\hat{\mathbf{G}}(0) = \mathbf{G}(0)$. We illustrate two alternative ways to achieve this condition in the proposed workflow, with reference to (13).

1) BASIS AUGMENTATION

Building on a well-known fact from classical moment matching algorithms, the basis \mathbf{V} can be augmented as

$$\mathbf{V} \leftarrow \text{orth} \left\{ (\mathbf{V} \quad \mathbf{A}^{-1} \mathbf{B}_w \quad \mathbf{A}^{-1} \mathbf{B}_u) \right\} \quad (45)$$

With this modification, $\mathcal{R}\{\mathbf{V}\}$ includes the subspace spanned by $\mathbf{A}^{-1}(\mathbf{B}_w, \mathbf{B}_u)$, which is sufficient to achieve interpolation at $\omega = 0$ as desired. A major downside of this approach is that the basis size (columns of \mathbf{V}) hence the order of the reduced system may become exceedingly large in case the number of system inputs/outputs is large. This is indeed the case for the power distribution networks under consideration.

2) RECIPROCAL TRANSFORMATION

This technique is described in [23] and is based on the reciprocal frequency transformation $s \leftarrow 1/s$, which maps DC to infinite frequency and vice-versa. Starting from $\mathbf{G}(s)$ defined by (13a)-(13c), a descriptor realization of the corresponding reciprocal system $\mathbf{G}_r(s) = \mathbf{G}(1/s)$ is $\mathbf{G}_r = (\mathbf{E}_r, \mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r, \mathbf{D}_r)$, whose expressions are given in Appendix D.

This reciprocally-transformed system is such that

$$\lim_{s \rightarrow \infty} \mathbf{G}_r(s) = \mathbf{D}_r = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} = \mathbf{G}(0). \quad (46)$$

Projection of the state-space matrices in (75) does not affect the feedthrough term \mathbf{D}_r . Hence, a reduced system $\hat{\mathbf{G}}_r = (\hat{\mathbf{E}}_r, \hat{\mathbf{A}}_r, \hat{\mathbf{B}}_r, \hat{\mathbf{C}}_r, \hat{\mathbf{D}}_r)$ obtained from (75) using $\mathbf{W}_d, \mathbf{V}_d$ as in (24) is such that $\hat{\mathbf{D}}_r = \mathbf{D}_r$. Finally, an additional reciprocal transformation on $\hat{\mathbf{G}}_r(s)$ yields a reduced model $\hat{\mathbf{G}}(s) = \hat{\mathbf{G}}_r(1/s)$ back in the original frequency variable. This model satisfies

$$\hat{\mathbf{G}}(0) = \hat{\mathbf{D}}_r = \mathbf{D}_r = \mathbf{G}(0), \quad (47)$$

as desired.

It is important to remark that, just as reported in [23] and [24] for the case of nonsingular \mathbf{E} , Gramian matrices are invariant under the above reciprocal frequency transformation (see Appendix D). This implies that \mathbf{V}_d can still be constructed in the original s -domain following the procedure in either Sec. V-A or Sec. V-B. The resulting basis can be used to reduce the intermediate reciprocally-transformed system \mathbf{G}_r , followed by a final reciprocal transformation on the reduced system.

D. ENFORCING CLOSED-LOOP STABILITY

Up to this point, only the determination of the right projection matrix \mathbf{V} has been addressed, with emphasis on the accurate characterization of the subspace \mathcal{L} where dominant dynamics take place. We now focus on the left projection matrix \mathbf{W} , which is here entirely driven by the necessity of guaranteeing closed-loop input-output stability of the reduced-order interconnected system.

We start by noting that the full system (13) can be decomposed in two parts: i) an LTI subsystem (\mathbf{G} in Fig. 1) formed by (13a)-(13c) with external ports (inputs \mathbf{u} , outputs \mathbf{y}) and internal interface ports (inputs \mathbf{w} , outputs \mathbf{z}); ii) the FIVR switches represented by (13d), which are connected to the internal interface ports of \mathbf{G} . The input-output representation of both subsystems is immittance (hybrid, since some ports are voltage-controlled and some other ports are current-controlled for both subsystems).

The LTI subsystem is physically passive and is represented as a passive descriptor system, thanks to the formulation in Appendix A. Therefore, the descriptor form of the Positive Real Lemma (PRL) [25] holds, which states that there exists a matrix \mathbf{P} such that $\mathbf{E}^T \mathbf{P} = \mathbf{P}^T \mathbf{E} \geq 0$ and

$$\begin{pmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P}^T \mathbf{A} & \mathbf{P}^T \mathbf{B} - \mathbf{C}^T \\ (\mathbf{P}^T \mathbf{B} - \mathbf{C}^T)^T & -(\mathbf{D} + \mathbf{D}^T) \end{pmatrix} \leq 0. \quad (48)$$

Additionally, the averaged switch model (13d) is lossless, since the hybrid matrix $\Delta(\mathbf{d})$ in (16) is skew-symmetric and $\Delta(\mathbf{d}) + \Delta(\mathbf{d})^T = \mathbf{0}$. This remains true even when the duty cycle signals $\mathbf{d}(t)$ are time-varying, as obtained by the feedback controllers (18), since for any trajectory of $\mathbf{d}(t)$, the signals \mathbf{w}, \mathbf{z} are related instantaneously by $\mathbf{w}(t) = \Delta(\mathbf{d}(t)) \cdot \mathbf{z}(t)$, thus satisfying $\mathbf{w}(t)^T \mathbf{z}(t) = 0$ at all times. The latter condition implies no power dissipation nor generation. Now, since the interconnection of passive subsystems is also passive [26], and since any passive system is also stable, we conclude that if we can enforce the reduction process to preserve passivity, then the reduced-order closed-loop system will also be passive hence stable.

We now proceed to find a left projection matrix \mathbf{W} that preserves passivity. This is in fact straightforward: starting from any solution \mathbf{P} of (48) and setting $\mathbf{W} = \mathbf{P}\mathbf{V}$ guarantees that also the reduced system matrices defined in (24) satisfy the PRL condition (48), with a solution matrix $\hat{\mathbf{P}} = \mathbb{I}$, as can be verified by direct substitution. Note that the MNA formulation of an RLC(M) network is a particular case of the above with $\mathbf{P} = \mathbb{I}$ and $\mathbf{W} = \mathbf{V}$. In such case, the projection is also a congruence transform, a fact exploited in well-known passivity-preserving projection algorithms such as PRIMA [27].

The explicit solution of the PRL equations (48) is very expensive for large-scale systems. Fortunately, this operation is not necessary for the proposed block-structured system. As discussed in Appendix A, the PRL solution \mathbf{P} of a system obtained by assembling arbitrary RLC(M) elements and passive macromodels can be obtained by the block-diagonal assembly of the individual PRL matrices

$$\mathbf{P} = \text{blkdiag}_{\ell} \{\mathbf{P}_{\ell}\}. \quad (49)$$

If the ℓ -th block is a state-space macromodel, the corresponding \mathbf{P}_{ℓ} needs to be determined by explicit solution of the corresponding PRL equations. A computational method for finding \mathbf{P}_{ℓ} is through the associated algebraic Riccati equation, see e.g. [16, chap. 2 and 5] and references therein. If instead the block is made of RLC(M) elements, the corresponding $\mathbf{P}_{\ell} = \mathbb{I}$. We see that this solution provides a significantly less expensive approach with respect to other known passive projection methods, such as [7] and [28].

VI. EXPLOITING RANDOMIZED SVD

In this section, we discuss how the randomized SVD algorithm can be used to enable fast computation of the empirical Gramians in a large-scale setting. Notice that the block entries in the \mathbf{R}^v matrix in (42) are not necessarily sparse, even if the matrices \mathcal{E} and \mathcal{A} are. Therefore, explicitly storing these matrices to perform SVD in a second step might become impractical due to large memory requirements. Since only their principal components are ultimately needed, it is possible to leverage the randomized SVD algorithm from [29] for computing a low-rank representation of \mathbf{R}^v in an online fashion, meaning that the snapshots do not need

Algorithm 1 Online Randomized SVD for Gramian Computation**Require:** $\{\omega_k\}_{k=1}^K, (\mathcal{E}, \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}), \rho$

```

1: Initialize  $\mathbf{R}_\Omega^v \leftarrow \mathbf{0}$ 
2: for  $k = 1, \dots, K$  do
3:    $\Omega_k', \Omega_k'' \leftarrow$  random matrices in  $\mathbb{R}^{M \times p}$ 
4:    $\mathbf{X}_{d,k}^v \leftarrow [(\mathbf{j}\omega_k \mathcal{E} - \mathcal{A})^{-1} \mathcal{B}]_{(1:n_d,:)}$ 
5:    $\mathbf{R}_\Omega^v \leftarrow \mathbf{R}_\Omega^v + \sqrt{\alpha_k} [\mathbf{X}_{d,k}^{v,r} \Omega_k' + \mathbf{X}_{d,k}^{v,i} \Omega_k'']$ 
6: end for
7: Compute  $\mathbf{Q}$  from a QR decomposition of  $\mathbf{R}_\Omega^v$ 
8:  $\mathbf{Q}_1 \leftarrow \mathbf{Q}_{(:,1:\rho)}$ 
9: for  $k = 1, \dots, K$  do
10:   $\mathbf{Z}_k \leftarrow \mathbf{Q}_1^\top [\sqrt{\alpha_k} (\mathbf{j}\omega_k \mathcal{E} - \mathcal{A})^{-1} \mathcal{B}]_{(1:n_d,:)}$ 
11:   $\mathbf{Z}_k^r \leftarrow \text{Re}\{\mathbf{Z}_k\}, \mathbf{Z}_k^i \leftarrow \text{Im}\{\mathbf{Z}_k\}$ 
12: end for
13:  $\mathbf{Z} \leftarrow (\mathbf{Z}_1^r \mathbf{Z}_1^i \dots \mathbf{Z}_K^r \mathbf{Z}_K^i)$ 
14: Perform SVD decomposition of  $\mathbf{Z} = \mathbf{U}_z \mathbf{\Sigma} \mathbf{V}^\top$ 
15:  $\mathbf{U} \leftarrow \mathbf{Q}_1 \mathbf{U}_z$ 
16: return  $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}$  such that  $\mathbf{R}^v \approx \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ 

```

to be all stored in memory simultaneously. This procedure is particularized for current application in Algorithm 1.

Two passes over the K data matrices $\mathbf{X}_d(\mathbf{j}\omega_k)$ are performed. In the first pass (Steps 2–6), the product between $\mathbf{R}^v \in \mathbb{C}^{n_d \times M}$ and p random vectors is computed, where p is larger than the desired rank ρ in the low-rank decomposition. According to [29], it is typically sufficient to set $p = \rho + 10$ when the singular values of \mathbf{R}^v decay rapidly. By regarding these random vectors as the columns of a matrix $\Omega \in \mathbb{R}^{M \times p}$, a block-row partitioning compatible with the block-columns of \mathbf{R}^v

$$\Omega = \text{col}\{\dots, \Omega_k', \Omega_k'', \dots\} \quad (50)$$

allows us to decompose the product $\mathbf{R}^v \Omega$ as

$$\mathbf{R}^v \Omega = \sum_{k=1}^K \sqrt{\alpha_k} \left(\mathbf{X}_d^{v,r}(\mathbf{j}\omega_k) \Omega_k' + \mathbf{X}_d^{v,i}(\mathbf{j}\omega_k) \Omega_k'' \right) \quad (51)$$

which shows how $\mathbf{R}^v \Omega$ can be computed by summing contributions for each k without ever forming \mathbf{R}^v . The remaining algorithmic steps are standard and not specific to present application, so we refer the reader to [29] for a complete discussion.

VII. NUMERICAL INTEGRATION

Transient simulation consists in integrating the system of differential equations (13), coupled with the controller (18). In this section, we outline the time-stepping scheme used to provide the results reported in the foregoing Sec. VIII. The first step in our derivation is the discretization of (13) using the backward Euler scheme with a fixed timestep δ ,

$$\mathbf{x}_{h+1} = (\mathbf{E} - \delta \mathbf{A})^{-1} (\mathbf{E} \mathbf{x}_h + \delta \mathbf{B}_w \mathbf{w}_{h+1} + \delta \mathbf{B}_u \mathbf{u}_{h+1}) \quad (52a)$$

$$\mathbf{z}_{h+1} = \mathbf{C}_z \mathbf{x}_{h+1} + \mathbf{D}_{zw} \mathbf{w}_{h+1} + \mathbf{D}_{zu} \mathbf{u}_{h+1} \quad (52b)$$

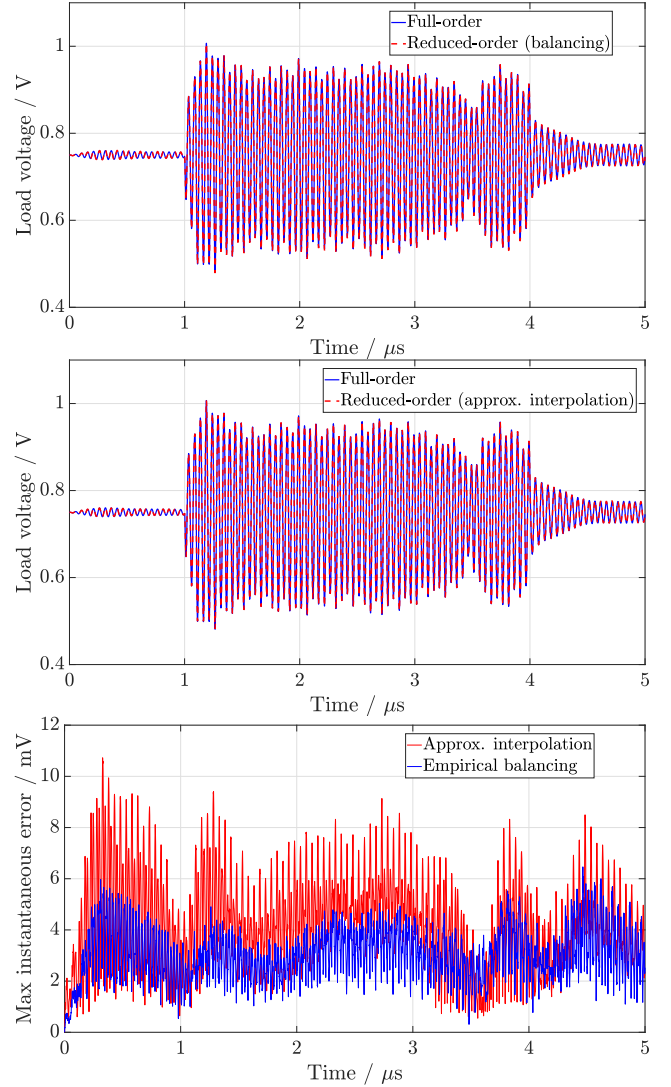


FIGURE 4. For the mobile benchmark of Sec. VIII-A, top and middle panels report the load voltage transient response (at the third port of the first core) to a sequence of current steps, used to compare two reduced models of order $q = 470$. Bottom panel: maximum instantaneous error between full-order and reduced models across all load voltages in the same simulation.

$$\mathbf{y}_{h+1} = \mathbf{C}_y \mathbf{x}_{h+1} + \mathbf{D}_{yw} \mathbf{w}_{h+1} + \mathbf{D}_{yu} \mathbf{u}_{h+1} \quad (52c)$$

where ζ_h denotes a first-order approximation of the corresponding variable $\zeta(h\delta)$. The controller equations (18) can be discretized analogously. The remaining equation (13d) is treated differently and discretized semi-explicitly as

$$\mathbf{w}_{h+1} = \mathbf{\Delta}(\mathbf{d}_h) \mathbf{z}_{h+1} \quad (53)$$

In this way, the solution at timestep $h + 1$ is obtained by solving (52) and (53) simultaneously in a linear system, where the value of \mathbf{d}_h is known and available from the previous timestep. Since the format of (52) is invariant upon proposed model order reduction schemes, the above time discretization holds for and will be applied to both original and reduced systems. We remark that the matrix $\mathbf{E} - \delta \mathbf{A}$

in (52a) is constant and can be efficiently pre-factorized for a fast inversion at each time step.

VIII. NUMERICAL EXPERIMENTS

A. A MOBILE PLATFORM

This section validates the proposed reduction approaches on a relatively small-scale test case, namely a mobile computing system equipped with a 4-cores Intel® Core™ microprocessor. The corresponding PDN includes four FIVRs with $N_p = 4$ phases each, $N_o = 36$ output ports per core and 144 output ports overall. The initial state dimension of the full-order system is $n = 2673$.

Application of the balanced truncation procedure described in Sec. V-B gives a reduced model of order $q = 470$, corresponding to $\sigma_{q+1}/\sigma_1 < 10^{-5}$. Another model with the same q has been constructed with the approximate interpolation approach of Sec. V-A. The approximation error is evaluated numerically in a realistic scenario where both reduced models are used in a transient simulation, in which the load ports are excited with current signals that model chip activity. In this experiment, individual cores transition between an off-state, where no current is drawn from the corresponding output ports, and an on-state, where the load current \mathbf{i}_k^o is a random piecewise linear waveform with peak per-core current 10 A (uniformly distributed among all $N_o = 356$ core ports) and average current 0.11 A. The power spectrum of this signal is contrived to have a peak at 20 MHz, close to the resonant peak of the PDN output impedance. Results of this transient analysis are reported in Fig. 4. The load voltage measured at the third load port of the first core is shown in the upper panels, and a more comprehensive picture of the approximation error is given in the bottom panel, which reports the maximum error across all 144 load ports for each time instant. It is seen that both models essentially capture the correct voltage response, with the truncated balancing approach providing a generally more accurate result for the same reduced model order q . The larger oscillations starting at $t = 1 \mu\text{s}$ correspond to the first core being turned on, whereas the smaller ringing between 0 and $1 \mu\text{s}$ is the effect of drawing current from the second core (core coupling through the input network through the FIVRs switches). This simulation runs in 17 seconds, that is $56\times$ faster than the full-order model using the same solver (955 s) with the same fixed timestep $\delta = 0.1 \text{ ns}$. The same simulation in HSPICE runs in 972 seconds (with maximum timestep $\delta = 0.1 \text{ ns}$).

B. AN ENTERPRISE SERVER PLATFORM

This section describes the application of the proposed model reduction framework to the PDN of an enterprise server based on an Intel® Xeon® microprocessor with $N_c = 60$ modeled cores and $N_p = 3$ FIVR phases. There are $N_o = 57$ load ports for each core, and 3420 output ports overall where voltage needs to be stabilized and monitored. The MNA algorithm could be used to obtain a descriptor realization of the linear subnetwork, whose state dimension is in excess of $5 \cdot 10^4$.

TABLE 1. Modeling times for the enterprise server benchmark of Sec. VIII-B.

Operation	Time
Computation of \mathbf{P}	30 s
Computation of \mathbf{R}^c	44 min
Computation of \mathbf{R}^o	41 min
Reciprocal transformation	150 s
Projection	20 s
Approx.interpolation	$\approx 47 \text{ min}$
Empirical balancing	$\approx 88 \text{ min}$

Using $K = 25$ evaluations of the frequency-domain snapshots as described in Sec. V, reduced models could be built with different accuracies depending on the reduced order q . These were tested in time-domain using a benchmark simulation where the excitation current profile consists of step signals at each individual load port having amplitude 20/57 A and rise time 3 ns. In this $1.1\text{-}\mu\text{s}$ long simulation, groups of 15 cores are turned on and off simultaneously, and the total current drawn from each core in the on-state is 20 A.

A reduced model with $q = 380$, built with the empirical balancing approach is seen to achieve a maximum error of 3.3 mV in this test case, as reported in Fig. 6. This model is also compared with another one with the same q built using the approximate interpolation alternative. The errors of both models in the benchmark time-domain simulation, defined as the maximum instantaneous error across all load ports, are compared in the bottom panel of Fig. 6. The time required to simulate these reduced models is 65 s, more than ten times faster than the 902 s taken by the full-order model using the same transient solver.

Results concerning numerical experiments with different reduced models and methodologies are summarized in Table 2. This shows that both proposed methodologies yield acceptable time-domain peak errors, which can be decreased by increasing the reduced order q . The empirical balancing approach shows lower error than approximate interpolation when comparing different models with the same q or with the same relative truncation threshold σ_{q+1}/σ_1 . Although the balancing approach is more accurate and robust, building a model takes longer, as summarized in Table 1, where the modeling times required by each algorithmic step of Sec. V are detailed. Note that the majority of time is spent in the phase of snapshot computation, which is basically an AC analysis where a large sparse linear system is solved for each frequency point, and is perhaps one of the computationally cheapest ways to gather information on the system to be reduced. Time spent for this task could be further reduced by means of more sophisticated approaches for AC analysis, by parallelizing the computation for different frequencies and exploiting iterative methods available for large and sparse linear systems. Additional computational steps specific to the model reduction task take negligible time with respect to AC response computation. Note that this computational cost is paid only once to construct the model, which can then be used for power integrity verification through multiple repeated time-domain simulations with different load current waveforms or compensator models.

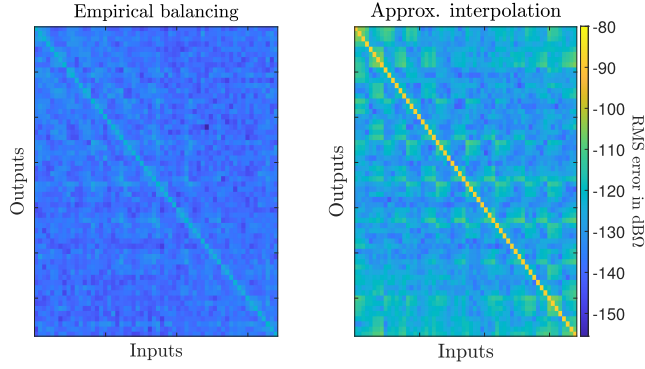


FIGURE 5. RMS error over frequency for each response of the reduced small-signal output impedance submatrix [part of $\hat{\mathcal{H}}(j\omega)$] over the bandwidth $[0, 1]$ GHz, relative to the enterprise server benchmark of Sec. VIII-B. Models were built with a fixed relative threshold $\sigma_{q+1}/\sigma_1 \leq 10^{-4}$.

TABLE 2. Enterprise server benchmark (60 cores). Comparison between reduction methods in terms of worst-case error and transient simulation time.

Method	q	σ_{q+1}/σ_1	Error	Time
Interpolation (Sec. V-A)	380	$1.7 \cdot 10^{-4}$	5.1 mV	65 s
Interpolation (Sec. V-A)	446	$9.2 \cdot 10^{-5}$	3.4 mV	69 s
Balancing (Sec. V-B)	380	$2 \cdot 10^{-4}$	3.3 mV	65 s
Balancing (Sec. V-B)	483	$9.4 \cdot 10^{-5}$	3.0 mV	70 s

Besides time-domain analysis, we also evaluate the frequency response error of the linearized transfer function $\mathcal{H}(j\omega)$ versus the reduced model $\hat{\mathcal{H}}(j\omega)$. In particular, we consider the root-mean-square (RMS) error for each individual matrix entry of the small-signal output impedance in a bandwidth $[0, 1]$ GHz. Fig. 5 reports the RMS errors of two models corresponding to the second and fourth rows in Tab. 2, which are built with the same relative truncation threshold $\sigma_{q+1}/\sigma_1 \leq 10^{-4}$. Each point in this picture indicates the RMS error over frequency of the corresponding impedance matrix entry. A further numerical demonstration of the accuracy and correctness of the proposed framework is provided in Fig. 7, showing the magnitude of two entries $Z_{1,1}$ and $Z_{2,400}$ of the small-signal output impedance matrix. The original model is compared with the result obtained from the empirical balancing approach with reduced order $q = 600$, corresponding to $\sigma_{q+1}/\sigma_1 = 2 \cdot 10^{-5}$. Here, the reciprocal transformation is used to enforce the DC point during model reduction, so that reduced and full models are identical at $\omega = 0$. Accuracy at all other frequencies is enforced during projection by an appropriate selection of the reduced order model. In this example, such order has been selected so that the worst-case absolute deviation in the output impedance at any frequency is $\approx 10^{-7} \Omega$. This accuracy is far more than required for the considered application.

C. SCALABILITY ANALYSIS

Starting from the 60-cores benchmark of Sec. VIII-B, here we analyze how runtime scales with the model complexity by including only part of all available cores in the model, in particular $N_c \in \{16, 32, 48, 60\}$. Reduced-order models

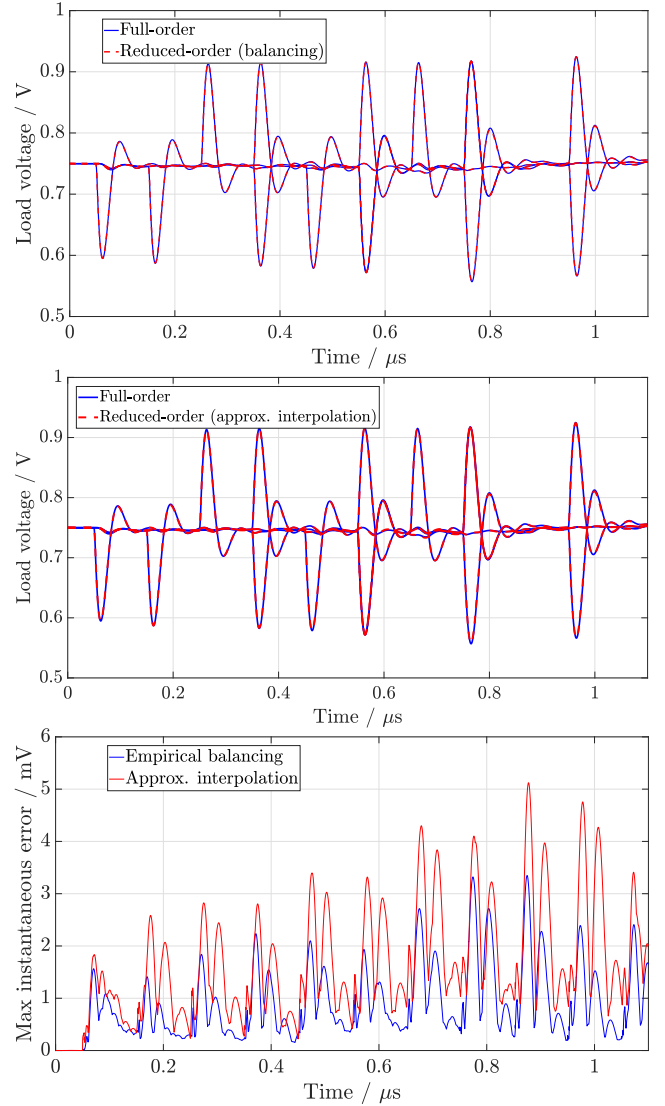


FIGURE 6. Top panel: load voltage response for the enterprise server benchmark in Sec. VIII-B. Top and middle panels compare the responses of the reduced models based on empirical balancing and interpolation, respectively, to the reference full-order responses. Bottom panel reports the corresponding maximum error across all load ports. Both reduced models have the same order $q = 380$.

are built for each case with a fixed relative threshold $\sigma_{q+1}/\sigma_1 = 10^{-4}$ and solved in time-domain using a MATLAB implementation of the integration scheme of Sec. VII. The runtime is compared with the time taken by the same solver to simulate the full-order model with the same N_c , and the runtime of the reference commercial solver HSPICE. All simulations are performed on the same workstation based on Core i9-7900X CPU running at 3.3 GHz with 64 GB of RAM.

We remark that, once the descriptor system (13) is available by parsing the netlist and filling in the MNA matrices, there are two main options to carry out the transient simulation of the full-order system:

- if no further processing is done and the descriptor system is solved directly, the matrix $(\mathbf{E} - \delta\mathbf{A})$ is large and sparse,

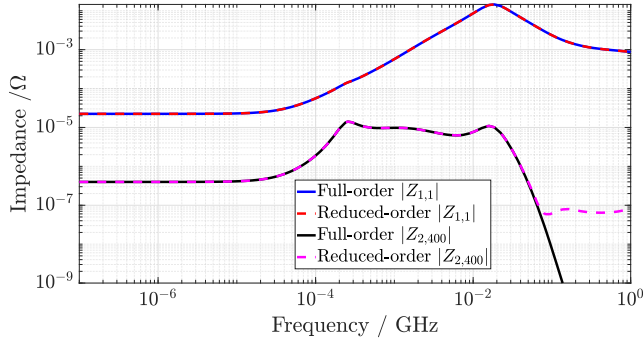


FIGURE 7. Enterprise server benchmark: selected responses from the output impedance matrix of the full-order $\mathcal{H}(j\omega)$ and reduced-order $\mathcal{H}(j\omega)$ obtained through empirical balancing, with $q = 615$ and $\sigma_{q+1}/\sigma_1 = 1 \cdot 10^{-5}$.

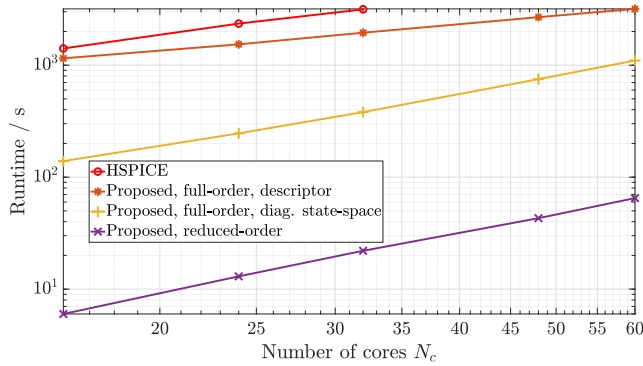


FIGURE 8. Scalability analysis reporting the transient simulation runtime required by various combinations of models/solvers for the enterprise server benchmark. Runtime is plotted as a function of the N_c modeled cores.

TABLE 3. Runtime as a function of N_c as in Fig. 8. Speedup refers to reduced vs. full-order models solved in MATLAB.

N_c	HSPICE	MATLAB (full)	MATLAB (reduced)	Speedup
16	1410 s	1151 s	6 s	200×
24	2351 s	1535 s	13 s	120×
32	3152 s	1950 s	22 s	90×
48	—	2683 s	43 s	60×
60	—	3181 s	64 s	50×

and the repeated application of its inverse in (52a) at each timestep can be done efficiently by pre-computing its LU factorization.

- if the system size is not too large, one might choose to first convert it to a standard state space using the procedure in Appendix B. This procedure eliminates some states but produces dense and still large matrices. A standard diagonalization can be further applied to make the resulting \mathbf{A} diagonal or quasi-diagonal. The latter step can become computationally infeasible as the size of these matrices becomes large, due to storage and CPU requirements (exact diagonalization takes $O(n^3)$ operations).

In the present case, it was possible to apply both approaches. The corresponding results are depicted in Fig. 8. These issues do not arise with the reduced model, so that a standard state-space description in diagonal form (modal realization)

can be obtained at negligible cost. The corresponding results correspond to the blue line in Fig. 8.

Figure 8 shows that the direct descriptor solver takes almost as long as HSPICE to simulate the full-order model, compatibly with the fact that no simplification is applied to the MNA descriptor system. Note that HSPICE failed to converge when more than $N_c = 32$ cores were included. Solving a full-order diagonal state-space realization is much more efficient, despite the overhead of diagonalization that in this case is 63 s for \mathbf{A}_{in} and 0.27 s for the output model. After model reduction applied, the runtime decreases by a factor ranging from 50× to 200× depending on the system configuration, as reported in Table 3. Speedup is reduced for larger N_c because runtime scales differently in the full and reduced models (notice the slopes in Fig. 8). In fact, the full-order runtime is dominated by the cost of applying the operator $(\mathbf{E} - \delta\mathbf{A})^{-1}$, an $n \times n$ matrix in LU-factored form [see Sec. VII, in particular (52a)]. As this is reduced to a quasi-diagonal $q \times q$ matrix, the computational bottleneck becomes the repeated solution of the linear system (52) combined with (53), which has the cost of solving for the $N_c N_p$ unknowns in \mathbf{w}_{h+1} .

APPENDIX A FROM CIRCUITS AND MACROMODELS TO DESCRIPTOR FORM

Here, we show how a standard MNA formulation can be used to assemble a descriptor system in form (3) or (6) for a multiport structure including an arbitrary number of RLCM components and state-space macromodels in immittance form. We demonstrate the procedure for a single macromodel \mathbf{G}_M in admittance form, and we target an impedance representation of the multiport. Generalizations are straightforward.

A. FORMULATION

We define the reduced and block-partitioned incidence matrix $(\mathbf{A}_R, \mathbf{A}_L, \mathbf{A}_C, \mathbf{A}_X, \mathbf{A}_P)$ of the structure assuming a global reference node. The various column blocks are indexed by a subscript referring to the subgraphs of resistors, inductors, capacitors, macromodel ports and external interface ports. The KCL and KVL equations read

$$\sum_v \mathbf{A}_v \mathbf{i}_v = 0, \quad \mathbf{v}_v = \mathbf{A}_v^T \mathbf{e}, \quad v \in \{R, L, C, X, P\} \quad (54)$$

where \mathbf{e} collects the nodal voltages and $\mathbf{v}_v, \mathbf{i}_v$ collect voltages and currents of the v -th subgraph, defined using normal reference directions. The characteristics of all circuit blocks are

$$\mathbf{i}_R = \mathbf{G}\mathbf{v}_R, \quad \mathbf{v}_L = \mathbf{L}\dot{\mathbf{i}}_L, \quad \mathbf{i}_C = \mathbf{C}\dot{\mathbf{v}}_C \quad (55)$$

where matrices \mathbf{G}, \mathbf{L} and \mathbf{C} collect all conductances, (possibly coupled) inductances and capacitances of the various circuit elements. The macromodel is known through its state-space

equations

$$\begin{cases} \dot{\mathbf{x}}_X = \mathbf{A}_M \mathbf{x}_X + \mathbf{B}_M \mathbf{v}_X \\ \mathbf{i}_X = \mathbf{C}_M \mathbf{x}_X + \mathbf{D}_M \mathbf{v}_X \end{cases} \quad (56)$$

where input and output variables are voltages \mathbf{v}_X and currents \mathbf{i}_X . Finally, port excitations are ideal current sources such that $\mathbf{i}_P = -\mathbf{I}$. Assembling all above equations leads to the following representation of the multiport impedance relating port voltages \mathbf{V} and currents \mathbf{I}

$$\mathbf{Z}(s) = \mathbf{B}^T (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} : \begin{cases} \mathbf{E}\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{I} \\ \mathbf{V} = \mathbf{B}^T \mathbf{x} \end{cases} \quad (57)$$

where $\mathbf{B}^T = (\mathbf{A}_P^T \quad \mathbf{0} \quad \mathbf{0} \quad \mathbf{0})$,

$$\mathbf{x} = \begin{pmatrix} \mathbf{e} \\ \mathbf{i}_L \\ \mathbf{x}_X \\ \mathbf{i}_X \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \mathbf{A}_C \mathbf{C}_A^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (58)$$

and

$$\mathbf{A} = - \begin{pmatrix} \mathbf{A}_R \mathbf{G}_A^T & \mathbf{A}_L & \mathbf{0} & \mathbf{A}_X \\ -\mathbf{A}_L^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}_M \mathbf{A}_X^T & \mathbf{0} & -\mathbf{A}_M & \mathbf{0} \\ -\mathbf{D}_M \mathbf{A}_X^T & \mathbf{0} & -\mathbf{C}_M & \mathbb{I} \end{pmatrix} \quad (59)$$

B. PASSIVITY CHARACTERIZATION

The structure of the system (57)-(59) can be exploited to infer and characterize its passivity properties. Let us assume that the macromodel (56) is passive, so that its state-space realization $(\mathbf{A}_M, \mathbf{B}_M, \mathbf{C}_M, \mathbf{D}_M)$ satisfies the Positive Real Lemma (PRL) inequality (48) with a matrix $\mathbf{P}_M = \mathbf{P}_M^T > 0$ (being a regular state-space, the PRL applies with $\mathbf{E}_M = \mathbb{I}$). We know that the interconnection of passive systems is also passive, therefore also (57)-(59) must fulfil the PRL conditions (48) with a suitable matrix \mathbf{P} . A straightforward algebraic verification shows that we can define this matrix as the block-diagonal collection $\mathbf{P} = \text{blkdiag}\{\mathbb{I}, \mathbb{I}, \mathbf{P}_M, \mathbb{I}\}$, where the presence of the identity matrices is a consequence of the internal passivity of the RLCM blocks. In this particular example, the PRL (48) holds with the particular conditions $\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} \leq 0$, $\mathbf{P} \mathbf{B} - \mathbf{C}^T = \mathbf{0}$, and $\mathbf{D} = \mathbf{0}$.

C. REDUCTION TO CANONICAL FORM

A closer look at (57)-(59) reveals the following facts:

- the upper-left block in matrix \mathbf{E} is not generally full-rank. However, elimination of selected nodal voltages in favor of the capacitor branch voltages along locally connected capacitor trees (assuming no capacitor loops) leads to a change of variable that reduces this upper block to $\mathbf{A}_C \mathbf{C}_A^T \rightarrow \text{blkdiag}\{\hat{\mathbf{C}}, \mathbf{0}\}$ where $\hat{\mathbf{C}}$ is full-rank. This operation can be performed by basic row/column sums.
- the inductance matrix \mathbf{L} can be rank deficient when perfect inductor couplings are present. In such case, a simple eigendecomposition reveals the vanishing

eigenvalues so that the change of variable induced by the (orthogonal) eigenvectors of \mathbf{L} reduces this block to $\text{blkdiag}\{\hat{\mathbf{L}}, \mathbf{0}\}$ where $\hat{\mathbf{L}}$ is full-rank.

We conclude that a standard MNA stamping process, followed by the above-described low-cost algebraic operations and a reordering of the states, leads to representation of any multiport structure including interconnected state-space macromodels with arbitrary RLCM elements as a descriptor form, with matrices $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ in the canonical form (20). This form is instrumental to the derivations in Appendix B.

APPENDIX B FROM DESCRIPTOR TO STATE-SPACE FORM

In this section, we report a procedure to obtain an equivalent standard state-space representation of a descriptor system arranged in the form (20). This operation could be formally accomplished through the Weierstrass form to separate finite and infinite modes [6]. This procedure is known to be numerically unreliable. Furthermore, general methods are available to remove non-dynamic modes in a descriptor system, such as the one described in [30] (based on the SVD-coordinate form [9]), or to recast it as a standard state-space system altogether (such as the shuffle algorithm [31]). The above general methods are fortunately unnecessary here. In fact, such decompositions are required for so-called *high-index* systems, which are characterized by the presence of impulsive modes, or equivalently by a direct dependence of the output on high-order derivatives of the input, hence by an asymptotic behavior of the frequency response $\mathbf{Z}(s) \sim O(s^m)$ for $s \rightarrow \infty$ as a high-order polynomial. For the considered application, however,

- the LTI system is passive: it can be shown [26] that the associated descriptor is at most index-two (i.e., at most polynomial order $m = 1$ for $s \rightarrow \infty$);
- further, the specific configuration of the considered PDN system can be shown to satisfy an even stronger (and beneficial) condition, with $m = 0$ and $\mathbf{Z}(s)$ bounded for $s \rightarrow \infty$.

The various situations that may occur are analyzed in detail below, with reference to (20).

A. CASE 1: \mathbf{A}_{22} INVERTIBLE

In the simplest case, the index of the regular pencil (\mathbf{E}, \mathbf{A}) is at most one, the frequency response $\mathbf{Z}(s)$ is bounded for $s \rightarrow \infty$ and the block \mathbf{A}_{22} is invertible [20, Thm 2.1]. In this case, all dynamic modes are finite and \mathbf{x}_d can be eliminated to recast (20) as an equivalent descriptor system with reduced-size matrices $(\check{\mathbf{E}}, \check{\mathbf{A}}, \check{\mathbf{B}}, \check{\mathbf{C}}, \check{\mathbf{D}})$ with nonsingular $\check{\mathbf{E}} = \mathbf{E}_{11}$ and

$$\begin{aligned} \check{\mathbf{A}} &= \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}, & \check{\mathbf{B}} &= \mathbf{B}_1 - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{B}_2, \\ \check{\mathbf{C}} &= \mathbf{C}_1 - \mathbf{C}_2 \mathbf{A}_{22}^{-1} \mathbf{A}_{21}, & \check{\mathbf{D}} &= \mathbf{D} - \mathbf{C}_2 \mathbf{A}_{22}^{-1} \mathbf{B}_2 \end{aligned} \quad (60)$$

A regular state-space realization is obtained by inverting $\check{\mathbf{E}}$ and redefining $\check{\mathbf{A}} \leftarrow \check{\mathbf{E}}^{-1} \check{\mathbf{A}}$ and $\check{\mathbf{B}} \leftarrow \check{\mathbf{E}}^{-1} \check{\mathbf{B}}$.

B. CASE 2: \mathbf{A}_{22} NON INVERTIBLE, NO IMPULSIVE MODES

In the more general case where \mathbf{A}_{22} is singular, expressions (60) become ill-defined. The index of the pencil (\mathbf{E}, \mathbf{A}) is two and there are impulsive modes. In this subcase, we assume that these impulsive modes are not visible in the input-output response, meaning that they cannot be excited by the inputs (uncontrollable) or are not detected on output (unobservable). In the present context, this situation holds when the circuit does not contain cutsets of inductors and current sources or loops of capacitors and voltage sources. Such cases would practically require ideal inductances directly connected in series to current-controlled ports (on-chip excitation and \mathbf{G}_{in} ports connected to FIVRs), or ideal capacitances directly connected in parallel to voltage-controlled ports (V_{IN} or \mathbf{G}_k ports connected to FIVRs). No PDN of practical use falls in this situation. These topological conditions can be stated algebraically (see [11]) as

$$\forall \mathbf{v} \in \mathbb{R}^{n-n_d} : \mathbf{A}_{22}\mathbf{v} = \mathbf{0} \implies \mathbf{C}_2\mathbf{v} = \mathbf{0} \quad (61a)$$

$$\forall \mathbf{p} \in \mathbb{R}^{n-n_d} : \mathbf{p}^T \mathbf{A}_{22} = \mathbf{0} \implies \mathbf{p}^T \mathbf{B}_2 = \mathbf{0} \quad (61b)$$

Conditions (61) enable an algebraic procedure to eliminate \mathbf{x}_a . We first introduce a change of coordinate and a partitioning on \mathbf{x}_a to cast (20) as

$$\mathbf{E}_{11}\dot{\mathbf{x}}_d = \mathbf{A}_{11}\mathbf{x}_d + \mathbf{A}'_{12}\mathbf{x}'_a + \mathbf{A}''_{12}\mathbf{x}''_a + \mathbf{B}_1\mathbf{u} \quad (62a)$$

$$\mathbf{0} = \mathbf{A}'_{21}\mathbf{x}_d + \mathbf{A}'_{22}\mathbf{x}'_a + \mathbf{B}'_2\mathbf{u} \quad (62b)$$

$$\mathbf{0} = \mathbf{A}''_{21}\mathbf{x}_d + \mathbf{B}''_2\mathbf{u} \quad (62c)$$

$$\mathbf{y} = \mathbf{C}_1\mathbf{x}_d + \mathbf{C}'_2\mathbf{x}'_a + \mathbf{C}''_2\mathbf{x}''_a + \mathbf{D}\mathbf{u} \quad (62d)$$

where \mathbf{A}'_{22} is nonsingular. This is always possible, e.g., through a rank-revealing QR factorization or an SVD. The consequence of conditions (61) is that, in this transformed representation, $\mathbf{B}''_2 = \mathbf{0}$ and $\mathbf{C}''_2 = \mathbf{0}$. The variable \mathbf{x}'_a can be solved for in (62b) and eliminated, leading to

$$\mathbf{E}_{11}\dot{\mathbf{x}}_d = \bar{\mathbf{A}}_{11}\mathbf{x}_d + \mathbf{A}'_{13}\mathbf{x}''_a + \bar{\mathbf{B}}_1\mathbf{u} \quad (63a)$$

$$\mathbf{0} = \mathbf{A}'_{21}\mathbf{x}_d \quad (63b)$$

$$\mathbf{y} = \bar{\mathbf{C}}_1\mathbf{x}_d + \bar{\mathbf{D}}\mathbf{u} \quad (63c)$$

with $\bar{\mathbf{A}}_{11} = \mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}'_{22}{}^{-1}\mathbf{A}'_{21}$, $\bar{\mathbf{B}}_1 = \mathbf{B}_1 - \mathbf{A}'_{12}\mathbf{A}'_{22}{}^{-1}\mathbf{B}'_2$, $\bar{\mathbf{C}}_1 = \mathbf{C}_1 - \mathbf{C}'_2\mathbf{A}'_{22}{}^{-1}\mathbf{A}'_{21}$ and $\bar{\mathbf{D}} = \mathbf{D} - \mathbf{C}'_2\mathbf{A}'_{22}{}^{-1}\mathbf{B}'_2$. This new canonical form has been studied in detail in [32] and [33], and basically shows that the states \mathbf{x}_d are constrained by (63b) to $\ker\{\mathbf{A}'_{21}\}$. Therefore, we can write $\mathbf{x}_d = \mathbf{K}\xi_d$ where the columns of $\mathbf{K} \in \mathbb{R}^{n_d \times \mu_d}$ with $\mu_d < n_d$ are an orthogonal basis for $\ker\{\mathbf{A}'_{21}\}$. Using this representation, simple algebraic manipulations lead to a state-space realization with reduced-size state vector $\xi_d \in \mathbb{R}^{\mu_d}$, with nonsingular $\check{\mathbf{E}} = \mathbb{I}$ and

$$\check{\mathbf{A}} = \mathbf{K}^T \mathbf{E}_{11}^{-1} \Pi \bar{\mathbf{A}}_{11} \mathbf{K}, \quad \check{\mathbf{B}} = \mathbf{K}^T \mathbf{E}_{11}^{-1} \Pi \bar{\mathbf{B}}_1, \quad (64a)$$

$$\check{\mathbf{C}} = \bar{\mathbf{C}}_1 \Pi \mathbf{K}, \quad \check{\mathbf{D}} = \bar{\mathbf{D}}, \quad (64b)$$

where $\Pi = \mathbb{I} - \mathbf{A}''_{12}(\mathbf{A}'_{21}\mathbf{E}_{11}^{-1}\mathbf{A}'_{12})^{-1}\mathbf{A}'_{21}\mathbf{E}_{11}^{-1}$. We conclude that also for index-two systems with uncontrollable and

unobservable impulsive modes, a regular state-space representation holds.

C. CASE 3: \mathbf{A}_{22} NON INVERTIBLE, IMPULSIVE MODES

When inductor-current source cutsets or capacitor-voltage source loops are present, so that (61) do not hold, it is still possible to derive an equivalent state-space representation, but the latter includes also one term in the output equation that is proportional to the derivative of the inputs. Since this case is of no interest in this work, we omit such details and refer the Reader to [32] and [33], where a technical discussion and a complete derivation can be found.

APPENDIX C SYSTEM GRAMIANS

In this section, we show that balanced truncation on the realizations $(\check{\mathbf{E}}, \check{\mathbf{A}}, \check{\mathbf{B}}, \check{\mathbf{C}}, \check{\mathbf{D}})$ with $\check{\mathbf{E}}$ nonsingular derived in Appendix B and valid for Case 1 (60) or Case 2 (64) can be equivalently carried out by working with the original semi-explicit descriptor $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ in (20) with singular \mathbf{E} , without performing explicitly the conversion to state-space.

We define the state responses of the regular realizations as

$$\check{\mathbf{X}}^c(s) = (s\check{\mathbf{E}} - \check{\mathbf{A}})^{-1}\check{\mathbf{B}}, \quad \check{\mathbf{X}}^o(s) = (s\check{\mathbf{E}}^T - \check{\mathbf{A}}^T)^{-1}\check{\mathbf{C}}^T \quad (65)$$

and the associated Gramians

$$\check{\mathbf{W}}^v = \frac{1}{2\pi} \int_{-\infty}^{\infty} \check{\mathbf{X}}^v(j\omega) \check{\mathbf{X}}^v(-j\omega)^T d\omega, \quad v = \{c, o\}. \quad (66)$$

Similarly, for the original descriptor (20) we define

$$\mathbf{X}^c(s) = (s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}, \quad \mathbf{X}^o(s) = (s\mathbf{E}^T - \mathbf{A}^T)^{-1}\mathbf{C}^T \quad (67)$$

from which we extract the first block-rows corresponding to dynamic equations

$$\mathbf{X}_d^v(s) \triangleq [\mathbf{X}^v(s)]_{(1:n_d,:)}, \quad v = \{c, o\} \quad (68)$$

to define the Gramians

$$\mathbf{W}_d^v = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{X}_d^v(j\omega) \mathbf{X}_d^v(-j\omega)^T d\omega, \quad v = \{c, o\} \quad (69)$$

Note that, based on the block structure of (20), we have

$$\mathbf{X}^c(s) = \begin{pmatrix} s\mathbf{E}_{11} - \mathbf{A}_{11} & -\mathbf{A}_{12} \\ -\mathbf{A}_{21} & -\mathbf{A}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} \quad (70)$$

and similarly for the response $\mathbf{X}^o(s)$.

We consider separately the cases 1-2 of Appendix B.

A. CASE 1: \mathbf{A}_{22} INVERTIBLE

If \mathbf{A}_{22} is invertible, we can show that

$$\check{\mathbf{W}}^v = \mathbf{W}_d^v, \quad v = \{c, o\} \quad (71)$$

In fact, block inversion formulas applied to (70) show that

$$\mathbf{X}_d^c(s) = (s\check{\mathbf{E}} - \check{\mathbf{A}})^{-1}\check{\mathbf{B}}, \quad \mathbf{X}_d^o(s) = (s\check{\mathbf{E}}^T - \check{\mathbf{A}}^T)^{-1}\check{\mathbf{C}}^T \quad (72)$$

with $\check{\mathbf{E}}, \check{\mathbf{A}}$ and $\check{\mathbf{B}}$ defined as in (60). Therefore,

$$\check{\mathbf{X}}^c(s) = \mathbf{X}_d^c(s), \quad \check{\mathbf{X}}^o(s) = \mathbf{X}_d^o(s) \quad (73)$$

and (71) holds trivially.

B. CASE 2: \mathbf{A}_{22} NON INVERTIBLE, NO IMPULSIVE MODES

In the case where \mathbf{A}_{22} is singular, under the same assumptions made in Appendix B and adopting the same notation, one can show that

$$\mathbf{X}_d^c(s) = \mathbf{K}\check{\mathbf{X}}^c(s), \quad (74a)$$

$$\mathbf{X}_d^o(s) = \mathbf{\Pi}^\top \mathbf{E}_{11}^{-\top} \mathbf{K}\check{\mathbf{X}}^o(s), \quad (74b)$$

$$\mathbf{W}_d^c = \mathbf{K}\check{\mathbf{W}}^c \mathbf{K}^\top, \quad (74c)$$

$$\mathbf{W}_d^o = (\mathbf{\Pi}^\top \mathbf{E}_{11}^{-\top} \mathbf{K})\check{\mathbf{W}}^o (\mathbf{\Pi}^\top \mathbf{E}_{11}^{-\top} \mathbf{K})^\top \quad (74d)$$

Detailed derivations are here omitted for the sake of brevity, the Reader is referred to [32] and [33] for details. From (74c)-(74d) we see that \mathbf{W}_d^c and \mathbf{W}_d^o are obtained by embedding in a space of dimension n_d the gramians $\check{\mathbf{W}}^c$, $\check{\mathbf{W}}^o$, which in turn have numerical rank μ_d . Therefore, the low-rank factorization (41) applied to \mathbf{W}_d^c and \mathbf{W}_d^o will produce the same projection bases (44) as if applied to $\check{\mathbf{W}}^c$ and $\check{\mathbf{W}}^o$. This implies that balanced truncation of the singular descriptor form (20) and the regular state-space (64) provide indeed the same result.

APPENDIX D RECIPROCAL TRANSFORMATION

Starting from a descriptor system $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ in semi-explicit form (20), a realization of the transfer function $\mathbf{G}_r(s) \triangleq \mathbf{G}(1/s)$ is given by $(\mathbf{E}_r, \mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r, \mathbf{D}_r)$, with

$$\begin{aligned} \mathbf{E}_r &= \mathbf{E}_{11}, & \mathbf{A}_r &= \mathbf{E}_{11} \mathbf{A}_{r,11}, \\ \mathbf{B}_r &= \mathbf{E}_{11} \mathbf{B}_{r,1}, & \mathbf{C}_r &= \mathbf{C}_{r,1} \end{aligned} \quad (75)$$

where all matrices are defined by the partitioning

$$\left(\begin{array}{c|c} \mathbf{A}^{-1}\mathbf{E} & \mathbf{A}^{-1}\mathbf{B} \\ \hline -\mathbf{C}\mathbf{A}^{-1}\mathbf{E} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{array} \right) = \left(\begin{array}{cc|c} \mathbf{A}_{r,11} & \mathbf{A}_{r,12} & \mathbf{B}_{r,1} \\ \mathbf{A}_{r,21} & \mathbf{A}_{r,22} & \mathbf{B}_{r,2} \\ \hline \mathbf{C}_{r,1} & \mathbf{C}_{r,2} & \mathbf{D}_r \end{array} \right) \quad (76)$$

with $\mathbf{A}_{r,11} \in \mathbb{R}^{n_d \times n_d}$.

Proof: By manipulation of the transfer function,

$$\begin{aligned} \mathbf{G}_r(s) &= \mathbf{D} + \mathbf{C}(s^{-1}\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} \\ &= \mathbf{D} + \mathbf{C}(s\mathbf{I} - \mathbf{A}^{-1}\mathbf{E} + \mathbf{A}^{-1}\mathbf{E})(\mathbf{A}^{-1}\mathbf{E} - s\mathbf{I})^{-1}\mathbf{B} \\ &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} - \mathbf{C}\mathbf{A}^{-1}\mathbf{E}(s\mathbf{I} - \mathbf{A}^{-1}\mathbf{E})^{-1}\mathbf{A}^{-1}\mathbf{B} \end{aligned}$$

whose realization is (76). Assuming the original system in the canonical form in (20), $\mathbf{A}_{r,12}$, $\mathbf{A}_{r,22}$, $\mathbf{C}_{r,2}$ result identically vanishing, thus corresponding to a block of unobservable states that can be removed. Finally, left-multiplication by \mathbf{E}_{11} yields (75). \square

REFERENCES

- [1] K. Radhakrishnan, M. Swaminathan, and B. K. Bhattacharyya, "Power delivery for high-performance microprocessors—Challenges, solutions, and future trends," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 4, pp. 655–671, Apr. 2021.
- [2] S. Govindan, K. Bharath, S. Venkataraman, and D. Gope, "A state-space-based method to model Vccin feedthrough noise in microprocessors with fully integrated voltage regulators," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 9, pp. 1391–1401, Sep. 2021.
- [3] A. Carlucci, S. Grivet-Talocia, S. Mongrain, S. Kulasekaran, and K. Radhakrishnan, "Towards accelerated transient solvers for full system power integrity verification," in *Proc. IEEE 31st Conf. Electr. Perform. Electron. Packag. Syst. (EPEPS)*, San Jose, CA, USA, Oct. 2022, pp. 1–3.
- [4] 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 Trans. Compon., Packag., Manuf. Technol.*, vol. 13, no. 10, pp. 1553–1566, Oct. 2023.
- [5] A. Carlucci, S. Grivet-Talocia, S. Mongrain, S. Kulasekaran, and K. Radhakrishnan, "A structured Krylov subspace projection framework for fast power integrity verification," in *Proc. IEEE 27th Workshop Signal Power Integrity (SPI)*, May 2023, pp. 1–4.
- [6] T. Stykel, "Gramian-based model reduction for descriptor systems," *Math. Control, Signals, Syst. (MCS)*, vol. 16, no. 4, pp. 297–319, Mar. 2004.
- [7] Z. Zhang, Q. Wang, N. Wong, and L. Daniel, "A moment-matching scheme for the passivity-preserving model order reduction of indefinite descriptor systems with possible polynomial parts," in *Proc. 16th Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2011, pp. 49–54.
- [8] *IdEM, Dassault Systèmes*. Accessed: Jan. 31, 2024. [Online]. Available: www.3ds.com/products-services/simulia/products/idem/
- [9] D. Bender and A. Laub, "The linear-quadratic optimal regulator for descriptor systems," *IEEE Trans. Autom. Control*, vol. AC-32, no. 8, pp. 672–688, Aug. 1987.
- [10] G. Verghese, B. Levy, and T. Kailath, "A generalized state-space for singular systems," *IEEE Trans. Autom. Control*, vol. AC-26, no. 4, pp. 811–831, Aug. 1981.
- [11] T. Reis and F. Glazov, "Systems theoretic properties of linear RLC circuits," *IFAC-PapersOnLine*, vol. 54, no. 9, pp. 38–45, 2021.
- [12] R. Erickson and D. Maksimović, *Fundamentals of Power Electronics*, 3rd ed. Cham, Switzerland: Springer, 2020.
- [13] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. thesis, Graduate College Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 1997.
- [14] J. R. Phillips and L. M. Silveira, "Poor man's TBR: A simple model reduction scheme," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 1, pp. 43–55, Jan. 2005.
- [15] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," *AIAA J.*, vol. 40, pp. 2323–2330, Jan. 2002.
- [16] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, Eds., *System- and Data-Driven Methods and Algorithms Part of the Multi-Volume Work Model Order Reduction*, vol. 1. Berlin, Germany: De Gruyter, 2021, doi: [10.1515/9783110498967](https://doi.org/10.1515/9783110498967).
- [17] D. F. Enns, "Model reduction with balanced realizations: An error bound and a frequency weighted generalization," in *Proc. 23rd IEEE Conf. Decis. Control*, Dec. 1984, pp. 127–132.
- [18] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
- [19] Z. Zhang and N. Wong, "An efficient projector-based passivity test for descriptor systems," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 29, no. 8, pp. 1203–1214, Aug. 2010.
- [20] T. Stykel, "Analysis and numerical solution of generalized Lyapunov equations," Ph.D. dissertation, Fakultät II—Mathematik und Naturwissenschaften der Technischen Universität, Berlin, Germany, 2002.
- [21] W. Gawronski and J.-N. Juang, "Model reduction in limited time and frequency intervals," *Int. J. Syst. Sci.*, vol. 21, no. 2, pp. 349–376, Feb. 1990.
- [22] M. G. Safonov and R. Y. Chiang, "A Schur method for balanced-truncation model reduction," *IEEE Trans. Autom. Control*, vol. 34, no. 7, pp. 729–733, Jul. 1989.
- [23] M. Green and D. Limebeer, *Linear Robust Control* (Dover Books on Electrical Engineering). New York, NY, USA: Dover, 2012.
- [24] B. Liljegren-Sailer and I. V. Gosea, "Data-driven and low-rank implementations of balanced singular perturbation approximation," 2023, *arXiv:2303.05361*.
- [25] R. W. Freund and F. Jarre, "An extension of the positive real lemma to descriptor systems," *Optim. Methods Softw.*, vol. 19, no. 1, pp. 69–87, Feb. 2004.
- [26] S. Grivet-Talocia and B. Gustavsen, *Passive Macromodeling: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2015.
- [27] A. Odabasioglu, M. Celik, and L. T. Pileggi, "PRIMA: Passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 8, pp. 645–654, Aug. 1998.

- [28] T. Reis and T. Stykel, "PABTEC: Passivity-preserving balanced truncation for electrical circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 29, no. 9, pp. 1354–1367, Sep. 2010.
- [29] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, Jan. 2011.
- [30] A. Varga. (2018). *Descriptor System Tools (DSTOOLS) User's Guide*. Accessed: Jan. 31, 2024. [Online]. Available: <https://arxiv.org/pdf/1707.07140.pdf>
- [31] D. G. Luenberger, "Time-invariant descriptor systems," *Automatica*, vol. 14, no. 5, pp. 473–480, Sep. 1978.
- [32] M. Heinkenschloss, D. C. Sorensen, and K. Sun, "Balanced truncation model reduction for a class of descriptor systems with application to the oseen equations," *SIAM J. Sci. Comput.*, vol. 30, no. 2, pp. 1038–1063, Jan. 2008.
- [33] S. Gugercin, T. Stykel, and S. Wyatt, "Model reduction of descriptor systems by interpolatory projection methods," *SIAM J. Scientific Comput.*, vol. 35, no. 5, pp. B1010–B1033, Jan. 2013.



ANTONIO CARLUCCI (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electronic engineering from the Politecnico di Torino, Turin, Italy, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the EMC Group. His research interests include large-scale simulation of electronic systems using macromodeling methods. He received the Best Student Paper Award of SPI 2023, the 27th IEEE Workshop on Signal and Power Integrity.



STEFANO GRIVET-TALOCIA (Fellow, IEEE) received the Laurea and Ph.D. degrees in electronic engineering from Politecnico di Torino, Turin, Italy, in 1994 and 1998, respectively. From 1994 to 1996, he was with the NASA/Goddard Space Flight Center, Greenbelt, MD, USA. He is currently a Full Professor of electrical engineering with Politecnico di Torino. He co-founded the academic spinoff company IdemWorks, Turin, in 2007, serving as the President until its acquisition by CST, in 2016. He has authored about 200 journal articles and conference papers. His current research interests include passive macro modeling of lumped and distributed interconnect structures, model-order reduction, modeling and simulation of fields, circuits, and their interaction, wavelets, time-frequency transforms, and their applications. He was a co-recipient of the 2007 Best Paper Award of IEEE TRANSACTIONS ON ADVANCED PACKAGING. He received the IBM Shared University Research Award, in 2007, 2008, and 2009. He was an Associate Editor of IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY from 1999 to 2001. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY. He was the General Chair of the 20th and 21st IEEE Workshops on Signal and Power Integrity (SPI2016 and SPI2017).



SIDDHARTH KULASEKARAN (Member, IEEE) received the B.Tech. degree in electrical engineering from NIT, Trichy, India, in 2010, and the M.Sc. and Ph.D. degrees in electrical, electronic and communications engineering from Arizona State University, Tempe, AZ, USA, in 2012 and 2017, respectively. Since 2017, he has been a Senior Analog Engineer with the Power Delivery Core Competency Team, Intel Corporation, Chandler, AZ, USA. At Intel, he focuses on developing modeling methodologies and measurement metrologies for characterizing power delivery networks. His work involves closing the gap between modeling and measurements through accurate 3D EM extraction and loss estimation. His research interests include integrated voltage regulators, advanced packaging, and passive technologies.



KALADHAR RADHAKRISHNAN (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA. In 2000, he joined Intel. He is an Intel Fellow and a Power Delivery Architect with the Technology Development Group, Intel. He has played a significant role in shaping and driving power delivery technologies for Intel microprocessors. He has authored four book chapters and more than 50 technical papers in peer-reviewed journals, and has been awarded 40 U.S. patents. His research interests include integrated voltage regulators, advanced packaging, and passive technologies. He was a two-time recipient of the Intel Achievement Award.

...