# A Fast Block Structure Preserving Model Order Reduction for Inverse Inductance Circuits *

Hao Yu, Yiyu Shi, Lei He
EE. Dept., University of California
Los Angeles, CA 90095

{hy255,yshi,lhe}@ee.ucla.edu

David Smart
Analog Devices Inc.
Wilmington, MA 01867

david.smart@analog.com

## ABSTRACT

*Most existing $RCL^{-1}$ circuit reductions stamp inverse inductance $L^{-1}$ elements by a second-order nodal analysis (NA). The NA formulation uses nodal voltage variables and describes inductance by nodal susceptance. This leads to a singular matrix stamping in general. We introduce a new circuit stamping for $RCL^{-1}$ circuits using branch vector potentials. The new circuit stamping results in a first-order circuit matrix that is semi-positive definite and non-singular. We call this as vector-potential based nodal analysis (VNA). It enables an accurate and passive reduction. In addition, to preserve the structure of state matrices such as sparsity and hierarchy, we represent the flat VNA matrix in a bordered-block diagonal (BBD) form. This enables us to build and simulate the macromodel efficiently. In experiments performed on several test cases, our method achieves up to 15X faster modeling building time, up to 33X faster simulation time, and as much as 67X smaller waveform error compared to SAPOR, the best existing second order $RCL^{-1}$ reduction method.*

**Categories and Subject Descriptors:** B.7.2[Hardware]: Integrated circuits – Design aids

**General Terms:** Algorithms, Design

**Keywords:** Inductance and Interconnect Modeling, Model Order Reduction

## 1. INTRODUCTION

Inductance is important in analysis of high-performance interconnects. The small scale of high-speed signal nets in on-chip integration can be well handled using the existing model order reduction [1]. However, for system on a chip or on a package, there exist components with high complexity and strong magnetic coupling. To ensure signal and power integrity, complete $RLC$ models are required. A detailed 3D extraction using the partial-element equivalent circuit (PEEC) model [2] introduces densely coupled inductances. The extracted model is dense and large. Sparsification and model order reduction (MOR) are hence both needed to reduce a large scale $RLC$ circuit.

Compared to the dense partial inductance matrix ($L$), $L^{-1}$ matrix is easier to sparsify [3, 4], where $L^{-1}$ elements are related to the drop of the branch vector potential [4]. Existing model order reduction techniques [1], however, stamp $L$ in the MNA

(modified nodal analysis) [5] matrix with the first-order admittance. Because directly stamping $L^{-1}$ in the MNA matrix leads to an asymmetric circuit stamping, a reduction using PRIMA does not preserve passivity [6]. To passively reduce the linear circuit containing $L^{-1}$, ENOR [7] stamps the *nodal susceptance*: $\Gamma = E_l L^{-1} E_l^T$ ($E_l$ is the incident matrix for inductance). It results in a second-order NA (nodal analysis) matrix, which can be passively reduced. SAPOR further improves the accuracy of orthonormalization by using the second-order Arnoldi method [8].

With detailed discussion in Section 2, $\Gamma$ and $G$ matrices in NA are recognized to be singular in general. Thus the transfer function of NA cannot be expanded at $dc$ ($s = 0$). As a result, the reduced state matrices can not be directly stamped in the circuit matrix together with active devices for transient simulation. To avoid singularity, SuPreme [9] applies the double-inversion to stamp $L^{-1}$ in MNA. However, it introduces additional factorization costs due to the double inversion. In this paper, we introduce the *branch vector potential* as a new state variable and obtain a first-order admittance that contains $L^{-1}$ elements. The new stamping is called *vector-potential nodal analysis* (VNA). The state matrices in VNA are semi-positive definite and non-singular, which can be passively reduced by congruence transformation with the block Arnoldi orthonormalization. Compared to [9] with double inversions, our method is more efficient.

In addition, we propose a fast structure-preserving reduction to build and simulate the reduced macromodel efficiently. Different from SPRIM [10], a structure-preserving reduction that relates the structure of the high-order system to the first-order system, our method aims at preserving the structure of state matrices such as the sparsity and the hierarchy. In this paper, we extend a two-level bordered-block diagonal (BBD) decomposition developed in BSMOR [11], which can consider couplings between blocks. Note that no inductance was considered in [11], and its moment matching was not localized. In this paper, we show that the moments can be matched locally by using only the diagonal blocks in the BBD state matrices. The macromodel hence can be efficiently constructed and simulated. Experimental results show that compared to SAPOR, an existing second order $RCL^{-1}$ reduction method, our approach is up to 15X and 33X faster to build and analyze macromodels, respectively. In addition, our method has 67X smaller waveform error.

The rest of this paper is organized as follows. In Section 2, we detail the limitations of stamping $L^{-1}$ in NA. In Section 3, we show that circuits containing $L^{-1}$ elements can be written in a first-order form using VNA, which is non-singular and can be reduced with preserved passivity. We call it **VOR** (Vector-potential based model Order Reduction) method. In Section 4, we further introduce a BBD-structure preserving model order reduction (**BVOR**) for large sized VNA circuits. We present experiments in Section 5 and conclude in Section 6. Proofs of theorems and more details can be found in a technical report [12]

## 2. BACKGROUND

Considering branch resistive (**G**), capacitive (**C**), inductive (**L**) elements, the external excitation current $\mathbf{I}(s)$ as input, and the probing voltage $u(s)$ as output, the KCL and KVL equations in
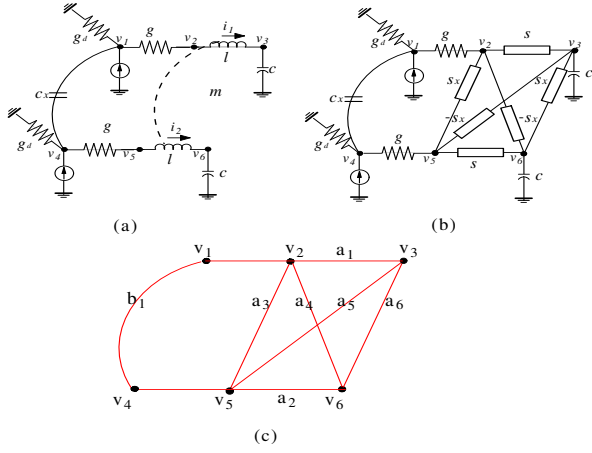
Figure 1: **Two coupled wires with (a) inductances and (b) susceptances ($S = L^{-1}$). In addition, (c) is a VNA-graph composed by nodal voltage and branch vector-potential for the two-wire example, where $b$-branch stands for capacitive branch, and $a$-branch stands for inductive branch.**



Figure 2: **Stamping the first-order admittance in MNA matrix, where (a), (b) and (c) represent $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ in Eqn. (3). $\mathcal{G}$ is non-singular.**



Figure 3: **Stamping the second-order admittance in NA matrix, where (a), (b) and (c) represent for $G$, $\Gamma$, $C$ and $B$ in Eqn. (7). $G$ and $\Gamma$ are both singular for $6 \times 6$ matrices. Note that $S = L^{-1}$ in the table.**

the $s$-domain are

$$\mathbf{E}\mathbf{i}_b^T = 0, \qquad \mathbf{E}^T\mathbf{v}_n = \mathbf{v}_b, \tag{1}$$

where $\mathbf{E} = [\mathbf{E}_g \quad \mathbf{E}_c \quad \mathbf{E}_l \quad \mathbf{E}_i]^T$ is the incident matrix, $\mathbf{i}_b = [\mathbf{i}_g \quad \mathbf{i}_c \quad \mathbf{i}_l \quad \mathbf{i}_i]^T$ and $\mathbf{v}_b = [\mathbf{v}_g \quad \mathbf{v}_c \quad \mathbf{v}_l \quad \mathbf{v}_i]^T$ are branch current and voltage, respectively, and $\mathbf{v}_n$ is nodal-voltage. The subscripts $(g, c, l, i)$ stand for the corresponding components for $(\mathbf{G}, \mathbf{C}, \mathbf{L}, \mathbf{I})$. For example, the branch currents are determined by

$$\mathbf{i}_g = \mathbf{G}\mathbf{v}_g, \ \mathbf{i}_c = s\mathbf{C}\mathbf{v}_c, \ \mathbf{v}_l = s\mathbf{L}\mathbf{i}_l, \ \mathbf{i}_i = -\mathbf{I}(s). \tag{2}$$

By defining nodal admittance matrices: $G = \mathbf{E}_g\mathbf{G}\mathbf{E}_g^T$ and $C = \mathbf{E}_c\mathbf{C}\mathbf{E}_c^T$, and only reserving the nodal-voltage $\mathbf{v}_n$ ($\in R^{n_v}$, $n_v$ is the number of nodal-voltage variables) and inductive branch-current $\mathbf{i}_l$ ($\in R^{n_i}$, $n_i$ is the number of inductive branch-current variables) as the state variables, (1) and (2) can be represented in the MNA matrix with the *first-order admittance*

$$(\mathcal{G} + s\mathcal{C})x(s) = \mathcal{B}\mathbf{I}(s) \quad u(s) = \mathcal{B}^T x(s). \tag{3}$$

where

$$x(s) = \begin{bmatrix} \mathbf{v}_n \\ \mathbf{i}_l \end{bmatrix}, \mathcal{B} = \begin{bmatrix} \mathbf{E}_i \\ 0 \end{bmatrix},$$

and

$$\mathcal{G} = \begin{bmatrix} G & \mathbf{E}_l \\ -\mathbf{E}_l^T & 0 \end{bmatrix}, \mathcal{C} = \begin{bmatrix} C & 0 \\ 0 & \mathbf{L} \end{bmatrix} \tag{4}$$

when stamping in $\mathbf{L}$. Note that the input and output ports are assumed to be identical. Fig. 2 presents the $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ for the circuit example in Fig. 1. Note that $\mathcal{G}$ and $\mathcal{C}$ become

$$\mathcal{G} = \begin{bmatrix} G & \mathbf{E}_l \\ -\mathbf{L}^{-1}\mathbf{E}_l^T & 0 \end{bmatrix}, \mathcal{C} = \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \tag{5}$$

when stamping in $\mathbf{L}^{-1}$. The difference between (4) and (5) is that the state matrix $\mathcal{G}$ in (5) does not satisfy: $\mathcal{G} + \mathcal{G}^T \succ 0$, i.e., it is not *symmetric* and *semi-positive definite*. Therefore, directly applying the congruence transformation by projecting (5) does not preserve passivity [6].

With $\mathcal{G}$ and $\mathcal{C}$ matrices defined in (5), (3) can be written as

$$(G + sC)\mathbf{v}_n + \mathbf{E}_l\mathbf{i}_l = \mathbf{E}_i\mathbf{I}(s), \quad s\mathbf{i}_l = \mathbf{L}^{-1}\mathbf{E}_l^T\mathbf{v}_n. \tag{6}$$

It is in fact the KCL's law. Because in most applications only nodal-voltage $\mathbf{v}_n$ is of interest, the inductive branch-current $\mathbf{i}_l$ can be eliminated as an intermediate variable. Therefore, (6) can be further written in the NA matrix with the *second-order admittance*

$$(sC + G + \Gamma/s)x(s) = \mathbf{E}_i\mathbf{I}(s) \quad u(s) = \mathbf{E}_i^T x(s), \tag{7}$$

where

$$x(s) = \mathbf{v}_n, \quad \Gamma = \mathbf{E}_l\mathbf{L}^{-1}\mathbf{E}_l^T.$$

$\Gamma$ is the *nodal susceptance* similar to $G$ and $C$. In addition, the time-domain response by NA is obtained by

$$(G + \frac{C}{h} + \Gamma \cdot h)\mathbf{v}_n(t+h) = \frac{C}{h}\mathbf{v}_n(t) - \mathbf{E}_l\mathbf{i}_l(t) - \mathbf{E}_i\mathbf{I}(t+h) \tag{8}$$

at time instant $t$ with step $h$.

Fig.3 shows the stamping of $\Gamma$ for the example of two wires in Fig. 1. In this form, $G$, $C$ and $\Gamma$ are all symmetric and positive definite. SAPOR [8] can passively reduce (7) by constructing orthonormalized columns to span a second-order Krylov subspace.

Yet, there exist limitations related to the second-order NA stamping for $L^{-1}$ elements. It is primarily due to the singularity. Here the singularity means: (*i*) the circuit matrix is indefinite at some frequency point or time instant; and (*ii*) the circuit matrix is rank-deficient, i.e., there exists linear dependence between columns or rows. Correspondingly, we have the following observations:

OBSERVATION 1. *Nodal admittance is indefinite at dc.*

Due to the stamping of $\Gamma$, it is obvious that at *dc* ($s \to 0$, or $h \to \infty$), the nodal admittance in (7) and (8) becomes indefinite (approaching to infinity).

OBSERVATION 2. *A PEEC-based extraction results in an interconnect model with π-topology, where R and L are serially connected [2] for one piece of interconnect. Assuming that there exists dc-path for the interconnect network, then stamping R and L separately into G and Γ by NA results in a singular stamping.*

Observation 2 can be understood as follows. If there are $N_R$ resistors and $N_L$ inductors serially connected, then the number of nodal-voltage variables is $N_R + N_L + 1$. In other words, $G$ matrix has only $N_L$ stamped resistors but has total $N_R + N_L + 1$ nodal-voltage variables. Therefore, $G$ is rank-deficient, i.e., singular. The same reason can explain why Γ is singular. These singularities are clearly shown in Fig.3. Here we assume that there exists dc-path for interconnect network. It can be always satisfied when connecting interconnect network with input sources or active devices (See source resistors in Fig.1 (a)). But the singularity in NA remains even when connected externally [12].

A non-singular circuit matrix is needed by an accurate model order reduction and a SPICE-like time-domain simulation. Because Γ is singular, the second-order admittance $K = s_0^2 C + s_0 G + Γ$ becomes indefinite (approaching to infinity) at dc ($s_0 = 0$). As a result, the moment generation matrix in SAPOR [8] has large numerical error around dc as well, and hence the reduced model expanded at dc is inaccurate (See Fig.6 later on). Therefore, SAPOR needs to choose a non-zero expansion point $s_0$ to ensure that $A$ is not singular. $s_0$ often uses a large value due to the magnitude difference between $G$, $C$ and Γ matrices. This leads to an inaccurate matching at the low-frequency region (See Fig.6 later on). More importantly, a reduction with nonzero $s_0$ expansion results in circuit matrices $\widetilde{G}$, $\widetilde{C}$ and $\widetilde{Γ}$ that do not have real values. As a result, it is inconvenient to stamp them in a circuit matrix together with active devices for time-domain simulation.

In contrast, MNA [5] uses inductive-branch current $\mathbf{i}_l$ to describe inductance (See (4)). It combines the singular $G$ together with $\mathbf{E}_l$ (describing branch inductors). The resulting $\mathcal{G}$ is non-singular. For example, if we assume there is an inductive-branch current $\mathbf{i}_k$ between voltage node $\mathbf{v}_i$ and $\mathbf{v}_j$, the combined matrix $\mathcal{G}$ below has nonzero entries at node $\mathbf{v}_i$ and $\mathbf{v}_j$, i.e.,

$$
\mathcal{G} = \begin{bmatrix}
 & \mathbf{v}_i & & \mathbf{v}_j & & \mathbf{i}_k & \\
 & \vdots & & \vdots & & \vdots & \\
\cdots & 0 & \cdots & & \cdots & 1 & \cdots \\
 & \vdots & & \vdots & & \vdots & \\
\cdots & & \cdots & 0 & \cdots & -1 & \cdots \\
 & \vdots & & \vdots & & \vdots &
\end{bmatrix}. \tag{9}
$$

Therefore, the combined $\mathcal{G}$ is non-singular even when there is no conductance stamped at $\mathbf{v}_i$ and $\mathbf{v}_j$. Ensuring the non-singularity of $\mathcal{G}$ is critical for simulation. Because in both model reduction and time-domain simulation, $\mathcal{G}$ needs to be factorized accurately.

Moreover, $\mathbf{L}$ is stamped together with $C$ in MNA, and the resulting $s\mathcal{C}$ becomes a zero matrix $\mathbf{0}$ at dc ($s = 0$). Hence, the MNA is definite at dc, different from NA which is indefinite (Observation 1). In summary, the singularity is only related to how we construct the circuit matrix. Only MNA can correctly describe inductances without a singular stamping. However, MNA is not symmetric and positive definite when stamping inverse-inductance. This will be solved below.

## 3. VNA MATRIX

### 3.1 Stamping $L^{-1}$ by Vector Potential Variable

We start with the differential Maxwell equations in terms of the vector potential $\mathbf{A}$ [4]:

$$
\nabla^2 \mathbf{A} = -\mu \mathbf{J}, \quad \frac{\partial \mathbf{A}}{\partial t} = -\mathbf{E} \tag{10}
$$

where $\mathbf{J}$ is current density, $\mathbf{E}$ is electrical field, and $\mu$ is permeability constant. For each filament (a constant branch current),



Figure 4: Stamp first-order admittance in VNA matrix, where (a), (b) and (c) represent $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ in Eqn.(14). $\mathcal{G}$ is non-singular.



Figure 5: Stamping the first-oder admittance in a bordered-block diagonal VNA matrix, where (a), (b) and (c) represent $\mathcal{G}$, $\mathcal{C}$ and $B$ in Eqn.(19). $\mathcal{G}$ is non-singular.

by defining a *branch vector potential* $\mathbf{A}_l$ as the *average* volume integral of $\mathbf{A}$ within the volume $\tau$:

$$
\mathbf{A}_l = \frac{1}{\tau} \int_\tau \mathbf{A} d\tau,
$$

(10) can be rewritten in the following circuit equations:

$$
\mathbf{L}^{-1} \mathbf{A}_l = \mathbf{i}_l, \quad \frac{\partial \mathbf{A}_l}{\partial t} = \mathbf{E}_l \mathbf{v}_n \tag{11}
$$

with $\mathbf{E}_l$ and $\mathbf{v}_n$ defined in (1). The detailed derivation can be found in [4]. However, [4] only presented a SPICE-compatible equivalent circuit, but did not discuss how to stamp the VPEC in the circuit matrix. In this paper, we show that by using branch vector potential $\mathbf{A}_l$ as a new state variable to replace the $\mathbf{i}_l$, $L^{-1}$ element can be stamped into a circuit matrix that is non-singular and can be passively reduced.

Note that (11) leads to

$$
(\mathbf{L}^{-1} \mathbf{E}_l^T) \mathbf{v}_n = s \mathbf{L}^{-1} \mathbf{A}_l. \tag{12}
$$

In addition, according to KCL law, we have

$$
(G + sC)\mathbf{v_n} + \mathbf{E}_l(\mathbf{L}^{-1}\mathbf{A}_l) = \mathbf{E}_i \mathbf{I}(s). \tag{13}
$$

As a result, by introducing a new state variable

$$
x = \begin{bmatrix} \mathbf{v}_n \\ \mathbf{A}_l \end{bmatrix},
$$

a first-order admittance can be obtained

$$\mathcal{G}x(s) + s\mathcal{C}x(s) = \mathcal{B}\mathbf{I}(s) \quad u(s) = \mathcal{B}^T x(s), \qquad (14)$$

where

$$\mathcal{G} = \begin{bmatrix} G & (\mathbf{E}_l\mathbf{L}^{-1}) \\ -(\mathbf{L}^{-1}\mathbf{E}_l^T) & 0 \end{bmatrix} \quad \mathcal{C} = \begin{bmatrix} C & 0 \\ 0 & \mathbf{L}^{-1} \end{bmatrix}$$

$$\mathcal{B} = \begin{bmatrix} \mathbf{E}_i \\ 0 \end{bmatrix}. \qquad (15)$$

Because

$$\mathcal{G} + \mathcal{G}^T = \begin{bmatrix} 2G & 0 \\ 0 & 0 \end{bmatrix}, \ \mathcal{C} + \mathcal{C}^T = \begin{bmatrix} 2C & 0 \\ 0 & 2\mathbf{L}^{-1} \end{bmatrix}, \qquad (16)$$

$\mathcal{G} + \mathcal{G}^T$ and $\mathcal{C} + \mathcal{C}^T$ are symmetric and semi-positive definite in VNA.

In addition, the $\mathcal{G}$ for interconnect is non-singular in VNA. Let $\mathbf{L}_{kk}^{-1}$ at branch $\mathbf{i}_k$ be an element between node $\mathbf{v}_i$ and $\mathbf{v}_j$. The product of $\mathbf{E}_l\mathbf{L}^{-1}$ is

$$\begin{bmatrix} & & \mathbf{i}_k & & \\ & & \vdots & & \\ \mathbf{v}_i & \cdots & 1 & \cdots & \\ & & \vdots & & \\ \mathbf{v}_j & \cdots & -1 & \cdots & \\ & & \vdots & & \end{bmatrix} \times \begin{bmatrix} & & \mathbf{i}_k & & \\ & & \vdots & & \\ \mathbf{i}_k & \cdots & \mathbf{L}_{kk}^{-1} & \cdots & \\ & & \vdots & & \end{bmatrix} =$$

$$\begin{bmatrix} & & \mathbf{i}_k & & \\ & & \vdots & & \\ \mathbf{v}_i & \cdots & \mathbf{L}_{kk}^{-1} & \cdots & \\ & & \vdots & & \\ \mathbf{v}_j & \cdots & -\mathbf{L}_{kk}^{-1} & \cdots & \\ & & \vdots & & \end{bmatrix}, \qquad (17)$$

which still has nonzero entries at $\mathbf{v}_i$ and $\mathbf{v}_j$. Therefore, stamping $\mathbf{E}_l\mathbf{L}^{-1}$ and its transpose together with $G$ results in a non-singular $\mathcal{G}$ similarly as MNA does, and Fig.4 presents $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ in VNA for the two-wire example. In contrast, both [7] and [8] use the NA form, which contains singular $G$ and $\Gamma$.

## 3.2 Reducing $L^{-1}$ in First-Order Form

After stamping, we can perform model order reduction similar to PRIMA. Solving (14) results in a transfer function $H(s) = \mathcal{B}^T(\mathcal{G} + s\mathcal{C})^{-1}\mathcal{B}$. By expanding of $H(s)$ at $s_0$ ($s = s_0 + \sigma$), it can be easily verified that the result is contained in the block Krylov subspace: $\{\mathcal{A}, \mathcal{A}\mathcal{R}, \cdots \mathcal{A}^{q-1}\mathcal{R}, ...\}$. with two moment generation matrices: $\mathcal{A} = (\mathcal{G} + s_0\mathcal{C})^{-1}\mathcal{C}$ and $\mathcal{R} = (\mathcal{G} + s_0\mathcal{C})^{-1}\mathcal{B}$. A projection matrix $Q$ that spans the $q$th block Krylov subspace

$$\mathcal{K}(\mathcal{A}, \mathcal{R}, q) = \{\mathcal{A}, \mathcal{A}\mathcal{R}, \cdots \mathcal{A}^{q-1}\mathcal{R}\} \subseteq Q$$

can be constructed from the block Arnoldi method [1]. Using $Q$ to project $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ matrices respectively, we obtain

$$\hat{\mathcal{G}} = Q^T\mathcal{G}Q, \quad \hat{\mathcal{C}} = Q^T\mathcal{C}Q, \quad \hat{\mathcal{B}} = Q^T\mathcal{B}. \qquad (18)$$

The resulting reduced transfer function is $\hat{H}(s) = \hat{\mathcal{B}}^T(\hat{\mathcal{G}} + s\hat{\mathcal{C}})^{-1}\hat{\mathcal{B}}$. We call this reduction as **VOR** (nodal Vector-potential based model Order Reduction) method. The reduced model $\hat{H}(s)$ has the following properties [12].

THEOREM 1. *The first $q$ block moments expanded at $s_0$ are identical for $\hat{H}(s)$ and $H(s)$.*

THEOREM 2. *When the input and output are symmetric, $\hat{H}(s)$ projected by $Q$ is passive.*

However, the reduction by **VOR** is constructed and projected in a flat fashion, and hence the according macromodel is inefficient to construct and simulate. To improve efficiency, we further introduce a block structured reduction below.

# 4. BBD STRUCTURED REDUCTION

## 4.1 Bordered-Block Diagonal Representation

The first step in BBD decomposition is to partition the VNA circuit into several blocks by branch-tearing [11]. Note that branch-tearing can consider couplings between different blocks.

A *min-cut* multilevel algorithm [13] is used to partition a *VNA graph* into user-specified $m$ blocks, each block with size $n_{b_i}$ and $n_{p_i}$ external ports. Since the number of inductive branches is much larger than that of capacitive branches, we define the graph composed by inductive branches as *VNA graph*. The resulting $m$ blocks with no coupling in between are at the bottom level of BBD. The top-level is one global interconnection block that connects all blocks by corresponding cut matrices. The interconnection block has size $n_0$ and contains all coupling branches between any pair of blocks at the bottom level.

Precisely, the resulting BBD system equation is

$$Yx = Bu \qquad (19)$$

where

$$Y = \begin{bmatrix} Y_1 & 0 & \dots & 0 & X_{10} \\ 0 & Y_2 & \dots & 0 & X_{20} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & Y_m & X_{m0} \\ -(X_{10})^T & -(X_{20})^T & \dots & -(X_{m0})^T & \mathbf{Z}_0 \end{bmatrix}$$

$$B = \begin{bmatrix} B_1 & 0 & \dots & 0 & 0 \\ 0 & B_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & B_m & 0 \\ 0 & 0 & \dots & 0 & \mathbf{0} \end{bmatrix}$$

and

$$x = [x_1, x_2, \ldots, x_m, i_0]^T, u = [u_1, u_2, \ldots, u_m, 0]^T.$$

For each block $Y_i$, $x_i$ is the state variable, the first-order VNA admittance is $\mathcal{G}_{ii} + s\mathcal{C}_{ii}$ ($\in R^{n_{b_i} \times n_{b_i}}$), $B_i$ is the excitation-port incidence matrix ($\in R^{n_{b_i} \times n_{p_i}}$) for external current sources, and $X_{i0}$ is the connection-port incidence matrix (cut matrix) ($\in R^{n_{b_i} \times n_0}$).

For the global interconnection block $\mathbf{Z}_0$ at the bottom, $i_0$ is the state variable, $(\mathbf{Z}_g)_0 + s(\mathbf{Z}_c)_0$ ($\in R^{n_0 \times n_0}$) is its branch impedance matrix. It is diagonal and contains all coupling branch impedances described by the set of cut matrices of $X_{i0}$. Because of the branch-tearing, all external sources are counted at each block by $B_i$, and there are no external sources in $Z_0$. Note that the off-diagonal block in the original flat VNA matrix can be recovered by

$$\mathcal{G}_{ij} + s\mathcal{C}_{ij} = -(X_{i0})^T(\mathbf{Z}_0)^{-1}(X_{j0}). \qquad (20)$$

As $\mathbf{Z}_0$ is a diagonal matrix, inverting $\mathbf{Z}_0$ is efficient. Fig.1 (c) presents a VNA-graph for the two-wire example. BBD representation introduces a sparse representation, but the overall system size is increased by $\mathbf{Z}_0$. Below, we show that the reduction and factorization in BBD form can be performed at the block level with reduced computational cost and memory requirement.

## 4.2 BBD-Structured Reduction

We first construct a block diagonal projection matrix,

$$\mathcal{Q} = diag[Q_1, Q_2, ..., Q_0] \qquad (21)$$

where $Q_i \in R^{n_{b_i} \times q}$ ($1 \le i \le m$).

Each projection block $Q_i$ is constructed from each diagonal block in (19) independently, where

$$\mathcal{K}(\mathcal{A}_{ii}, \mathcal{R}_{ii}, q) \subseteq Q_i. \qquad (22)$$

Note that $Q_0$ is an identity matrix $I_0$ ($\in R^{n_0 \times n_0}$), because there are no excitation sources for block $\mathbf{Z}_0$ and $\mathbf{Z}_0$ contains only diagonal entries.
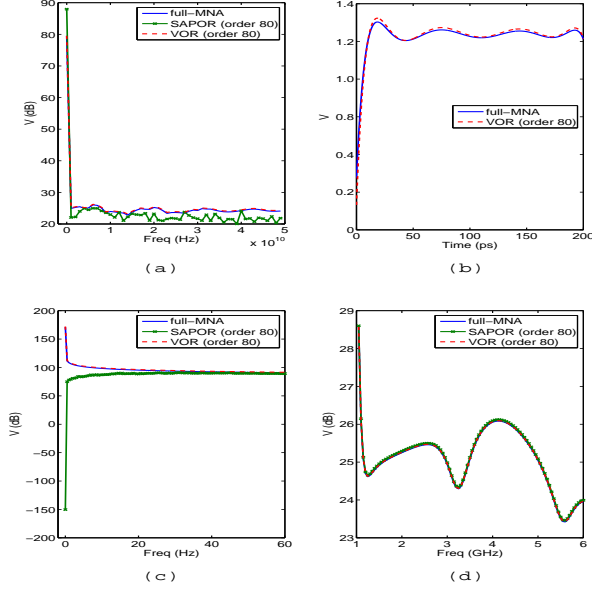
Figure 6: **Frequency and time domain waveform comparison of full-MNA, SAPOR (order=80) and VOR (order=80). The reduced models are expanded close to** $dc$ ($s_0 = 10Hz$) **in (a) and (b).VOR and original are visually identical. The reduced model by SAPOR can not converge in time-domain simulation.In addition, the reduced models are expanded at** $s_0 = 1GHz$ **in (c) and (d). VOR is identical to the exact MNA in both ranges,but SAPOR has large error at low-frequency range.**
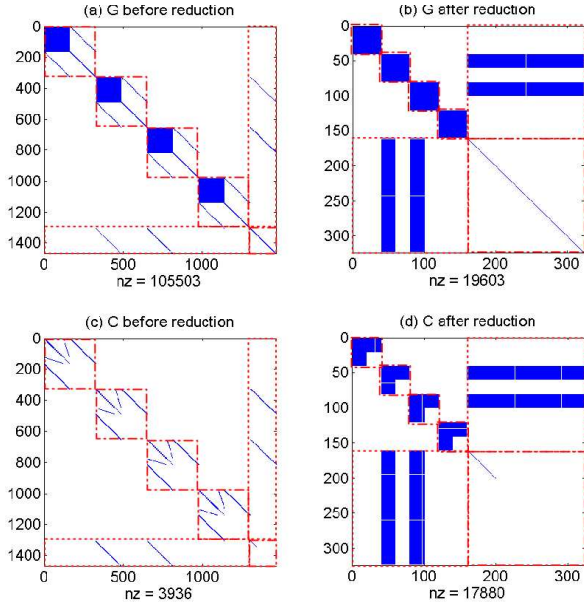


Figure 7: **BBD structure-preserving of state matrices** $\mathcal{G}$ **and** $\mathcal{C}$ **before and after reduction, where NZ is the number of non-zero entries.**

Because $\mathcal{Q}^T\mathcal{Q} = I$, i.e., each column of $\mathcal{Q}$ is still linearly independent and the total column-rank of $\mathcal{Q}$ (rank=$2mq$) is increased by a factor of $m$ compared to the column-rank of $Q$ (rank=$q$). The $m$-times increased column-rank means $m$-times more poles can be approximated after projection.

Decompose the BBD admittance into the diagonal and off-diagonal parts:

$$Y = (Y_g + Y_x) + s(Y_c + Y_x)$$

where the diagonal parts are

$$Y_g = diag[\mathcal{G}_1, ..., \mathcal{G}_m, (\mathbf{Z}_g)_0], \quad Y_c = diag[\mathcal{C}_1, ..., \mathcal{C}_m, (\mathbf{Z}_c)_0]$$

and the off-diagonal part is

$$Y_x = \begin{bmatrix} 0 & 0 & \dots & X_{10} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & X_{m0} \\ -X_{10}^T & \dots & -X_{m0}^T & 0 \end{bmatrix}.$$

We can obtain the order reduced state matrices by projecting $\mathcal{Q}$:

$$\widetilde{Y}_g + \widetilde{Y}_x = \mathcal{Q}^T(Y_g + Y_x)\mathcal{Q}, \;\; \widetilde{Y}_c + \widetilde{Y}_x = \mathcal{Q}^T(Y_c + Y_x)\mathcal{Q}, \;\; \widetilde{B} = \mathcal{Q}^T B$$

and hence define the transfer function

$$\widetilde{H}(s) = \widetilde{B}^T[(\widetilde{Y}_g + \widetilde{Y}_x) + s(\widetilde{Y}_c + \widetilde{Y}_x)]^{-1}\widetilde{B}. \qquad (23)$$

We call this reduction as **BVOR** (BBD based VOR) method.

In addition, we define the moment generation matrices for the original BBD as

$$\begin{aligned} \mathcal{A} &= [(Y_g + Y_x) + s_0(Y_c + Y_x)]^{-1}(Y_c + Y_x) \\ \mathcal{R} &= [(Y_g + Y_x) + s_0(Y_c + Y_x)]^{-1}B, \end{aligned}$$

and those for its diagonal blocks as

$$\mathcal{A}_0 = (Y_g + s_0 Y_c)^{-1} Y_c, \quad \mathcal{R}_0 = (Y_g + s_0 Y_c)^{-1} B.$$

Accordingly, we have the following properties for reduced $\widetilde{H}(s)$ [12].

THEOREM 3. *The qth-block Krylov space of BBD system is contained by the qth-block Krylov space spanned by the diagonal block of the BBD system, i.e.,* $\mathcal{K}(\mathcal{A}, \mathcal{R}, q) \subseteq \mathcal{K}(\mathcal{A}_0, \mathcal{R}_0, q)$. *Therefore, the first q expanded block moments at* $s_0$ *are identical for* $\widetilde{H}(s)$ *and* $H(s)$.

THEOREM 4. *When the input and output are symmetric,* $\widetilde{H}(s)$ *projected by* $\mathcal{Q}$ *is passive.*

Moreover, because the structured projection preservers the BBD structure, it enables a two-level domain-decomposition proposed in [11] to solve each reduced block independently. Each reduced block matrix is first solved individually with LU factorization and substitution. The results from each reduced block are then further used to solve the coupling block. The final $x_i$ of each reduced block is updated with the result from the coupling current $i_0$. As shown by experiments in Section 5, because building and solving the sparse BBD system are both performed at block level, the overall runtime is reduced although the system size becomes larger than the original MNA. However, there is a trade-off of sparsity and partition-time. When increasing the block number, the resulting BBD form becomes sparser but the partition time becomes larger to construct the cut matrix.

Table 1: Runtime comparison of SAPOR, VOR and BVOR with similar accuracy.

| ckt | size (before:after) | block # | Original(MNA) sim(s) | SAPOR(NA) build(s) | SAPOR(NA) sim(s) | VOR(VNA) build(s) | VOR(VNA) sim(s) | BVOR(BBD-VNA) build(s) | BVOR(BBD-VNA) sim(s) |
|---|---|---|---|---|---|---|---|---|---|
| bus8x20 | 325:40 | 4 | 99.3 | 2.02 | 11.90 | 1.41 | 9.32 | 0.36 | 4.32 |
| bus32x20 | 1200:200 | 16 | 1.96e3 | 16.9 | 102.3 | 14.3 | 89.2 | 2.12 | 12.5 |
| bus128x20 | 6000:500 | 64 | 8.73e4 | 92.7 | 1.05e3 | 59.6 | 905.2 | 6.01 | 31.6 |
| bus256x20 | 12000:800 | 64 | NA | NA | NA | NA | NA | 67.5 | 158.8 |
| clktree1 | 1215:200 | 4 | 2.09e3 | 18.7 | 104.1 | 15.7 | 91.2 | 1.88 | 11.2 |
| clktree2 | 2100:300 | 8 | 6.02e3 | 23.4 | 553.3 | 14.2 | 429.2 | 3.06 | 22.7 |
| clktree3 | 4500:400 | 16 | 1.83e4 | 101.9 | 640.0 | 58.2 | 569.2 | 6.15 | 25.2 |
| clktree4 | 36000:800 | 16 | NA | NA | NA | NA | NA | 653.2 | 2.38e3 |
| mesh1 | 800:80 | 8 | 524.1 | 10.0 | 31.2 | 5.12 | 16.76 | 1.12 | 7.48 |
| mesh2 | 2000:200 | 16 | 5.42e3 | 19.1 | 105.5 | 15.7 | 55.8 | 2.01 | 12.2 |
| mesh3 | 6000:400 | 32 | 1.08e5 | 96.3 | 649.9 | 51.2 | 575.3 | 7.29 | 25.2 |
| mesh4 | 48000:1200 | 32 | NA | NA | NA | NA | NA | 1.07e3 | 16.9e3 |

## 5. NUMERICAL EXPERIMENTS

Numerical experiments are presented to demonstrate the accuracy and efficiency of the proposed $VOR$ and $BVOR$ methods. They are compared with the exact MNA and SAPOR [8], the best existing second-order NA based reduction. All methods are implemented in MATLAB and C. Experiments are run on a Linux workstation with Intel Pentium IV 2.66G CPU and 2G RAM. The $RCL^{-1}$ circuits are extracted for buses, clock-trees and meshes. Model order reduction is first performed to obtain the reduced state matrices that are stamped back for the frequency and time-domain simulations. The Backward-Euler method is used to obtain the time-domain transient waveform.

The first example is an 8-bit bus with 20 segments per bit, similar to the example used by SAPOR [8]. We first expand first-order admittance in VNA and the second-order admittance in NA around $dc$. All reductions have an order of $q = 80$. The near-end of the first bit is excited with a unit current impulse, and the response at far-end is observed. Fig.6 (a) and (b) show the frequency and time domain responses of the exact MNA, reduced models by SAPOR and by VOR. Because state matrices in NA are close to singular, SAPOR shows large numerical errors beyond 1GHz. In contrast, VOR has a similar accuracy as the exact MNA for up to 50GHz. The reduced model by VOR is also as accurate as MNA in time-domain transient simulation. However, the one by SAPOR does not converge. In Fig.6 (c) and (d), we further expand VNA and NA at a non-zero frequency point $s_0 = 1GHz$. SAPOR converges at high frequency range as VOR does, but it is still inaccurate in low frequency range when compared to VOR. The reduced model by BVOR has up to 67X smaller error than that by SAPOR.

The second example is a 32-bit bus with 20 segments per bit, where shielding is inserted for every 8 bits. The partitioning results in 4 blocks, each with 180 resistors, 250 capacitors, and around 25600 nodal susceptors. The near end of the first bit in each block is excited with a unit current impulse. BVOR is applied to reduce each block independently, but VOR, SPRIM and SAPOR are applied to the entire circuit, all with order $q = 20$. Fig.7 (a) and (b) show the non-zero BBD pattern of the $\mathcal{G}$ matrix in VNA before and after the reduction, and Fig.7 (c) and (d) show those for the $\mathcal{C}$ matrix in VNA. Clearly, the reduced state matrices preserve the BBD structure and are sparse. $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{C}}$ have a 16.1% and 14.6% sparsification ratio, respectively.

Table 1 compares runtime for different typed $RCL^{-1}$ circuits with different size (number of state variables), where 10% of nodes in each cirucit have unit impulse current sources. The runtime here includes both the macromodel building and simulation time. The reduction and simulation are preformed in the frequency-domain. All reduced models have the same size and similar accuracy. For a circuit with $6000^2$ elements (bus128x20), BVOR reduces the orthonormalization cost and hence is 10X (6.01s versus 59.6s) faster than VOR to build macromodels. BVOR is also 15X (6.01s versus 92.7s) faster to build than SAPOR for the same circuit. The additional speedup is due to the fact that the first-order VNA matrix used in VOR is sparser than the NA matrix used in SAPOR. Moreover, for the same circuit (bus128x20) as the BBD structure is preserved after the reduction, the two-level

analysis in BVOR further reduces the simulation by 33X and 29X when compared to SAPOR and VOR, respectively.

## 6. CONCLUSIONS

Existing NA based second-order model reduction has singular matrix stamping for $RCL^{-1}$ circuits. This paper presented a vector-potential based nodal analysis (VNA) to accurately stamp $L^{-1}$ element and to enable a passive and simpler first-order reduction. By further representing the flat VNA matrix into a bordered-block diagonal (BBD) form, moment matching can be achieved locally by only using diagonal blocks in the BBD matrix. As a result, each block can be reduced independently and hence orthonormalization cost is minimized. Moreover, the reduced model is sparse and can be efficiently analyzed in a two-level fashion. In experiments performed on several test cases, our method achieved up to 15X faster modeling building time, up to 33X faster simulation time, and as much as 67X smaller waveform error compared to SAPOR.

## 7. REFERENCES

[1] A. Odabasioglu, M. Celik, and L. Pileggi, "PRIMA: Passive reduced-order interconnect macro-modeling algorithm," *IEEE Trans. on CAD*, pp. 645–654, 1998.

[2] A. E. Ruehli, "Equivalent circuits models for three dimensional multiconductor systems," *IEEE Trans. on MTT*, pp. 216–220, 1974.

[3] A. Devgan, H. Ji, and W. Dai, "How to efficiently capture on-chip inductance effects: introducing a new circuit element K," in *Proc. ICCAD*, 2000.

[4] H. Yu and L. He, "A provably passive and cost efficient model for inductive interconnects," *IEEE Trans. on CAD*, pp. 1283–1294, 2005.

[5] C. W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. on CAS*, pp. 504–509, 1975.

[6] H. Zheng and L. T. Pileggi, "Robust and passive model order reduction for circuits containing susceptance elements," in *Proc. ICCAD*, 2002.

[7] B. N. Sheehan, "ENOR: Model order reduction of RLC circuits using nodal equations for efficient factorization," in *Proc. DAC*, 1999.

[8] Y. Su, J. Wang, and et. al., "SAPOR: Second-order arnoldi method for passive order reduction of RCS circuits," in *Proc. ICCAD*, 2004.

[9] T. Chen, C. Luk, H. Kim, and C. Chen, "SuPREME: Substrate and power-delivery reluctance-enhanced macromodel evaluation," in *Proc. ICCAD*, 2003.

[10] R. W. Freund, "SPRIM: Structure-preserving reduced-order interconnect macro-modeling," in *Proc. ICCAD*, 2004.

[11] H. Yu, L. He, and S.-D. Tan, "Block structure preserving model reduction," in *IEEE BMAS-workshop*, 2005.

[12] H. Yu, Y. Shi, D. Smart, L. He, and S.-D. Tan, "An efficient first order block structure preserving model order reduction for $rcl^{-1}$ circuits," *UCLA Engr. 06-262., http://eda.ee.ucla.edu*, 2006.

[13] G. Karypis, R. Aggarwal, and V. K. S. Shekhar, "Multilevel hypergraph partitioning: application in VLSI domain," *IEEE Trans. on VLSI*, pp. 69–79, 1999.