

Parameterized Interconnect Order Reduction with Explicit-and-Implicit Multi-Parameter Moment Matching for Inter/Intra-Die Variations

Xin Li¹, Peng Li² and Lawrence T. Pileggi¹

¹Dept. of Electrical & Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{xinli, pileggi}@ece.cmu.edu

²Dept. of Electrical Engineering
Texas A&M University
College Station, TX 77843, USA
pli@neo.tamu.edu

Abstract

In this paper we propose a novel parameterized interconnect order reduction algorithm, CORE, to efficiently capture both inter-die and intra-die variations. CORE applies a two-step explicit-and-implicit scheme for multi-parameter moment matching. As such, CORE can match significantly more moments than other traditional techniques using the same model size. In addition, a recursive Arnoldi algorithm is proposed to quickly construct the Krylov subspace that is required for parameterized order reduction. Applying the recursive Arnoldi algorithm significantly reduces the computation cost for model generation. Several RC and RLC interconnect examples demonstrate that CORE can provide up to 10x better modeling accuracy than other traditional techniques, while achieving smaller model complexity (i.e. size). It follows that these interconnect models generated by CORE can provide more accurate simulation result with cheaper simulation cost, when they are utilized for gate-interconnect co-simulation.

1. Introduction

As IC technologies are scaled to deep submicron region, metal interconnect wires pose strong signal integrity problem and significantly impact the overall timing performance [1]. Meanwhile, it becomes increasingly difficult to control the relative process variations in nano-scale technologies [2]-[3]. The random fluctuations in manufacturing process introduce uncertainty in circuit behavior, thereby significantly impacting the circuit performance and product yield.

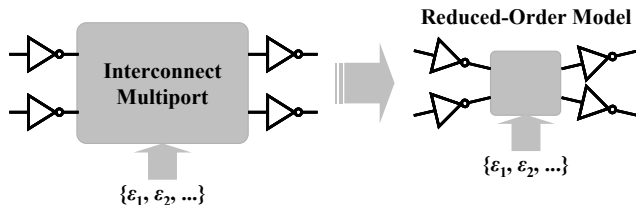


Fig. 1. Interconnect model with the variations $\{\epsilon_1, \epsilon_2, \dots\}$.

In order to model and analyze interconnect behavior with process variations, various techniques have been proposed for variational interconnect order reduction [4]-[5]. These approaches apply projection operator and generate reduced-order (small-size) interconnect models that can be further utilized to speedup the gate-interconnect co-simulation, as shown in Fig. 1. In addition, the projection subspace and/or the reduced-order system matrices are approximated as low-order polynomials of process parameters such that the process variation effects can be incorporated into the interconnect model. These process parameters, for example, can be the width and thickness of the interconnect metal wires.

Recently, intra-die variations become increasingly important in modern IC technologies [3]. These intra-die variations model the individual, but statistically correlated, local variations within

the same die. As will be discussed in Section 2, we find that the projection subspace and/or the reduced-order system matrices become strongly nonlinear in the presence of large-scale intra-die variations and, hence, cannot be accurately approximated by low-order polynomials. Therefore, although the techniques proposed in [4]-[5] have been successfully applied for inter-die variations, they become inefficient in modeling intra-die variation effects.

Recently-developed parameterized interconnect order reduction techniques propose to approximate the system transfer function, instead of the projection subspace and/or the reduced-order system matrices, by low-order polynomials of process parameters [6]-[7]. The system transfer function is weakly nonlinear even under intra-die variations and, therefore, is much easier to approximate with high accuracy. In addition, the authors in [7] propose the concept of multi-parameter moments which are utilized as a criterion for moment matching and generating reduced-order models. However, the algorithm proposed in [7] for multi-parameter moment matching can result in extremely large model size, especially when there exist many spatially correlated intra-die random variations.

In this paper we propose a novel Compact Orders Reduction algorithm for parameterized Extraction (CORE). The novelty of CORE lies in our unique two-step moment matching scheme that first *explicitly* matches the multi-parameter moments for the process parameters ϵ_i and then *implicitly* matches the moments for the frequency parameter s through projection. The main advantage of such an explicit-and-implicit moment matching is that an extremely compact (small-size) reduced-order model can be generated by CORE to simultaneously match a large number of multi-parameter moments. Since CORE dramatically increases the number of the matched multi-parameter moments, it can significantly reduce the modeling error (e.g. up to 10x in our tested examples) compared with other traditional techniques using the same model size.

Another important contribution of CORE is to formulate the parameterized system matrices into block lower triangular ones, based on a careful analysis of the underlying variational structure. Then, a *recursive Arnoldi* algorithm is proposed to quickly construct the Krylov subspace for projection. Compared with the *direct Arnoldi* algorithm used in PRIMA [8], the proposed recursive Arnoldi algorithm can achieve up to 10x speedup for parameterized interconnect order reduction in our tested examples.

The remainder of the paper is organized as follows. In Section 2 we review the background on variational interconnect order reduction and provide a comprehensive study on the advantages and disadvantages of the traditional techniques. Then we propose our CORE algorithm in Section 3. The efficacy of CORE is demonstrated by several interconnect examples in Section 4. Finally, we draw conclusions in Section 5.

2. Background

Without loss of generality, an RLC network can be described by the following modified nodal analysis (MNA) equation:

$$\begin{cases} (G + sC) \cdot x = B \cdot u \\ y = L^T \cdot x \end{cases} \quad (1)$$

where $u \in R^{m \times 1}$ and $y \in R^{n \times 1}$ denote the inputs and outputs, $x \in R^{N \times 1}$ represents the state variables, and $G, C \in R^{N \times N}$, $B \in R^{N \times m}$ and $L \in R^{n \times m}$ are the system matrices. n is the input number, m is the output number and N is the size of the MNA equation.

Given the linear network in (1), the purpose of model order reduction is to find a reduced-order (small-size) system:

$$\begin{cases} (\tilde{G} + s\tilde{C}) \cdot \tilde{x} = \tilde{B} \cdot u \\ y = \tilde{L}^T \cdot \tilde{x} \end{cases} \quad (2)$$

such that the input-output relation is ‘‘almost’’ unchanged. The reduced-order model in (2) can be generated by using Krylov subspace projection [8] or truncated balanced realization [9].

When process variations are considered, the matrices G and C in (1) are functions of process parameters. There are several techniques previously proposed to generate parameterized reduced-order models. Most of them fall into one of the following two categories: quadratic polynomial fitting and multi-parameter moment matching.

2.1 Quadratic Polynomial Fitting

When process variations are considered, the most straightforward way for parameterized interconnect order reduction is to approximate the reduced-order system matrices \tilde{G} , \tilde{C} , \tilde{B} and \tilde{L} in (2) as low-order (e.g. quadratic) polynomials of process parameters. The authors in [4] demonstrate that, with inter-die variations, the projection subspace and/or the reduced-order system matrices are weakly nonlinear and, therefore, can be accurately approximated by quadratic polynomials.

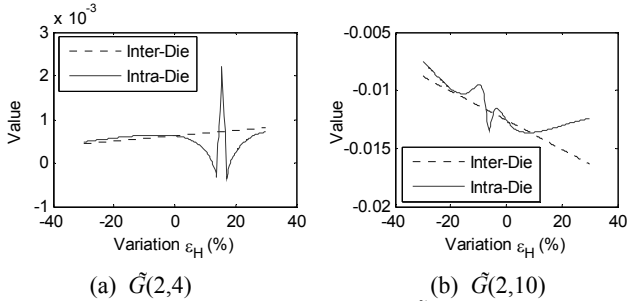


Fig. 2. The reduced-order system matrix \tilde{G} is strongly nonlinear with intra-die process variations.

However, when intra-die variations exist or when interconnect wires sit on multiple layers, the projection subspace and/or the reduced-order system matrices become strongly nonlinear and are difficult to fit. As a demonstration example, we extract the 10-th order model for an RC clock tree which consists of 1275 RC components. Fig. 2 shows two elements of the reduced-order system matrix \tilde{G} , i.e. $\tilde{G}(2,4)$ and $\tilde{G}(2,10)$, when the metal thickness is changed by $\epsilon_H = \pm 30\%$. In this example, results are compared when ϵ_H represents the inter-die variation (i.e. affect the entire interconnect wire) and the intra-die variation (i.e. only affect part of the interconnect wire) respectively. Note that the reduced-order system matrix \tilde{G} becomes strongly nonlinear in the presence of intra-die variations. Intuitively, the strong nonlinearity

in Fig. 2 appears because the reduced-order system matrices in (2) are the purely mathematical equations to represent the system. For a given linear system, its state-matrix representation is not unique. Therefore, a small perturbation on the physical RC elements can result in a dramatic change in these mathematic matrices.

2.2 Multi-Parameter Moment Matching

The authors in [6]-[7] propose to approximate the system transfer function H by low-order polynomials of process parameters. Unlike the reduced-order system matrices that are purely mathematical, the transfer function H is physical and uniquely determined for a given linear system. Therefore, a small perturbation on the physical interconnect parameters should only result in a small change in H . In other words, H is weakly nonlinear even if intra-die variations exist.

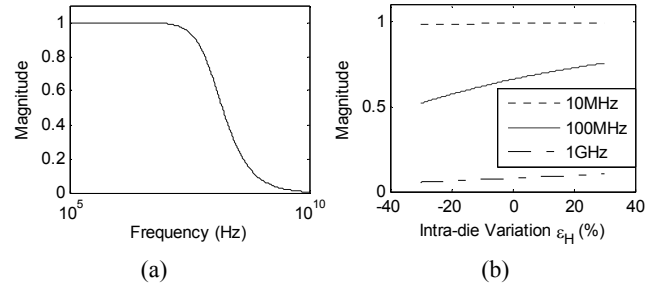


Fig. 3. (a) The nominal transfer function H ; (b) the transfer function H with intra-die variations.

Taking the same RC clock tree as an example, Fig. 3 shows its nominal transfer function and the transfer function with intra-die variations. It is shown in Fig. 3(b) that the transfer function remains smooth even if the intra-die metal thickness varies up to $\epsilon_H = \pm 30\%$.

Given a parameterized transfer function:

$$H(\epsilon_1, \dots, \epsilon_K, s) = \sum_{i=0}^{+\infty} \dots \sum_{iK=0}^{+\infty} H_{i1, \dots, iK, l} \cdot \epsilon_1^{i1} \dots \epsilon_K^{iK} \cdot s^l \quad (3)$$

where ϵ_i is the i -th process parameter and K is the total number of the variational process parameters, the coefficients $H_{i1, \dots, iK, l}$ are called the *multi-parameter moments* [7]. The authors in [7] propose an algorithm to compute the projection subspace and generate the reduced-order model such that the first several multi-parameter moments are matched.

It should be noted that if the Taylor expansion in (3) is matched up to M -th order for all ϵ_i and s , the number of the matched multi-parameter moments is much larger than M . Using the algorithm proposed in [7], the resulting model order is equal to n times the total number of the matched multi-parameter moments, where n is the number of the system inputs. Therefore, the model order generated by the algorithm in [7] can be extremely large in some practical applications. A large model order will result in expensive computation cost when using the model for gate-interconnect co-simulation.

3. CORE Algorithm

The main disadvantage of the algorithm in [7] is that the multi-parameter moment matching is completely achieved through projection. In order to match one additional multi-parameter moment, the projection subspace must be increased by n dimensions (n is the number of the system inputs), thereby increasing the final model order by n . Our proposed CORE

algorithm, however, utilizes a completely different moment matching scheme in order to increase the number of the matched multi-parameter moments for a given model order.

When process variations are considered, the state variables x in (1) can be approximated by the Taylor expansion:

$$x(\varepsilon_1, \dots, \varepsilon_K, s) = \sum_{i_1=0}^{+\infty} \dots \sum_{i_K=0}^{+\infty} x_{i_1, \dots, i_K}(s) \cdot \varepsilon_1^{i_1} \dots \varepsilon_K^{i_K} \quad (4)$$

Fig. 4 intuitively illustrates the equivalent system structure for this Taylor expansion. The input signal u passes various linear systems $x_{i_1, \dots, i_K}(s)$ and then all signals are added together with weight $\varepsilon_1^{i_1} \dots \varepsilon_K^{i_K}$ to generate the output signal y .

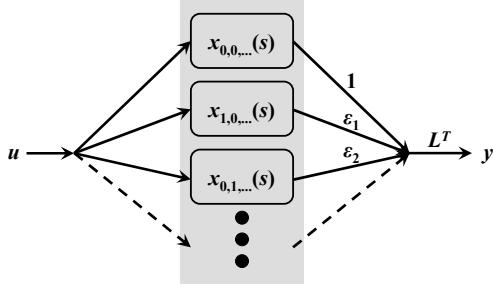


Fig. 4. Equivalent system structure for the Taylor expansion (4).

The main idea of CORE is that since all $x_{i_1, \dots, i_K}(s)$ in Fig. 4 share the same input, they can be conceptually considered as the different outputs of a large linear system, as shown by the grey area in Fig. 4. As such, the moments of all $x_{i_1, \dots, i_K}(s)$ can be *simultaneously* matched by using Krylov subspace projection.

CORE matches the multi-parameter moments through two steps. Firstly, the multi-parameter moments are matched for the process parameters ε_i , resulting in a set of governing equations to describe the large linear system $x_{i_1, \dots, i_K}(s)$ in Fig. 4. Next, the moments of all $x_{i_1, \dots, i_K}(s)$ are implicitly matched for the frequency parameter s through Krylov subspace projection.

3.1 Explicit Moment Matching for ε

Similar to (4), we expand the system matrices G and C in (1) by Taylor series:

$$G(\varepsilon_1, \dots, \varepsilon_K) = \sum_{i_1=0}^{+\infty} \dots \sum_{i_K=0}^{+\infty} G_{i_1, \dots, i_K} \cdot \varepsilon_1^{i_1} \dots \varepsilon_K^{i_K} \quad (5)$$

$$C(\varepsilon_1, \dots, \varepsilon_K) = \sum_{i_1=0}^{+\infty} \dots \sum_{i_K=0}^{+\infty} C_{i_1, \dots, i_K} \cdot \varepsilon_1^{i_1} \dots \varepsilon_K^{i_K}$$

The matrices B and L in (1) represent topological connections and, therefore, are independent on process variations.

Substituting (4)-(5) into (1) and *explicitly* matching the coefficients (i.e. the moments) for all cross-product terms $\varepsilon_1^{i_1} \dots \varepsilon_K^{i_K}$, we have:

$$\begin{aligned} (G_{0, \dots, 0} + sC_{0, \dots, 0}) \cdot x_{0, \dots, 0} &= B \cdot u \\ \varepsilon_1 \cdot (G_{1, \dots, 0} + sC_{1, \dots, 0}) \cdot x_{0, \dots, 0} + \varepsilon_1 \cdot (G_{0, \dots, 0} + sC_{0, \dots, 0}) \cdot x_{1, \dots, 0} &= 0 \\ &\vdots \\ \varepsilon_K \cdot (G_{0, \dots, 1} + sC_{0, \dots, 1}) \cdot x_{0, \dots, 0} + \varepsilon_K \cdot (G_{0, \dots, 0} + sC_{0, \dots, 0}) \cdot x_{0, \dots, 1} &= 0 \\ &\vdots \end{aligned} \quad (6)$$

$$y = L^T \cdot x = \sum_{i_1=0}^{+\infty} \dots \sum_{i_K=0}^{+\infty} L^T \cdot x_{i_1, \dots, i_K} \cdot \varepsilon_1^{i_1} \dots \varepsilon_K^{i_K} \quad (7)$$

After some mathematical manipulations, equations (6) and (7) can be re-formulated as:

$$\begin{cases} (G_{AP} + sC_{AP}) \cdot x_{AP} = B_{AP} \cdot u \\ y = L_{AP}^T \cdot x_{AP} \end{cases} \quad (8)$$

where

$$G_{AP} = \begin{bmatrix} G_{0, \dots, 0} & 0 & \dots & 0 & \dots \\ G_{1, \dots, 0} & G_{0, \dots, 0} & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \dots \\ G_{0, \dots, 1} & 0 & \dots & G_{0, \dots, 0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (9)$$

$$C_{AP} = \begin{bmatrix} C_{0, \dots, 0} & 0 & \dots & 0 & \dots \\ C_{1, \dots, 0} & C_{0, \dots, 0} & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \dots \\ C_{0, \dots, 1} & 0 & \dots & C_{0, \dots, 0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (10)$$

$$x_{AP} = \begin{bmatrix} x_{0, \dots, 0} \\ x_{1, \dots, 0} \\ \vdots \\ x_{0, \dots, 1} \\ \vdots \end{bmatrix} \quad \text{and} \quad B_{AP} = \begin{bmatrix} B \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} \quad \text{and} \quad L_{AP} = \begin{bmatrix} L \\ \varepsilon_1 L \\ \vdots \\ \varepsilon_K L \\ \vdots \end{bmatrix} \quad (11)$$

The system specified in (8) is called the *augmented parameterized system* or simply *parameterized system* in this paper. The parameterized system in (8) is much larger than the original system in (1); however, the following theorem proves that the system matrices G_{AP} and C_{AP} in (8) are block lower triangular. Such a special property makes it possible to handle the large-size parameterized system in practical applications, which will be further discussed in Section 3.2.

Theorem 1: The parameterized system matrices G_{AP} and C_{AP} in (9)-(10) are block lower triangular with the constant diagonal blocks $G_{0, \dots, 0}$ and $C_{0, \dots, 0}$ respectively.

Proof: When we match the Taylor expansion coefficients for one of the cross-product terms $\varepsilon_1^{i_1} \dots \varepsilon_K^{i_K}$ in (6), the left-hand side of (6) is:

$$\varepsilon_1^{i_1} \dots \varepsilon_K^{i_K} \cdot \sum_{j_1=0}^{i_1} \dots \sum_{j_K=0}^{i_K} (G_{j_1, \dots, j_K} + sC_{j_1, \dots, j_K}) \cdot x_{i_1-j_1, \dots, i_K-j_K} \quad (12)$$

Equation (12) proves Theorem 1. ■

In addition, the following theorem further proves that the stability is preserved when converting the original system to the parameterized one. Preserving stability is a unique property due to the explicit moment matching scheme. It is one of the key advantages of the proposed CORE algorithm. Note that the stability cannot be guaranteed for the parameterized systems derived by other approaches, e.g., [6].

Theorem 2: The parameterized system in (8) is stable if and only if the original system in (1) is stable.

Proof: The poles of the parameterized system in (8) are the roots of the following determinant:

$$\left| G_{AP} + sC_{AP} \right| = \begin{vmatrix} G_{0, \dots, 0} + sC_{0, \dots, 0} & & & 0 \\ G_{1, \dots, 0} + sC_{1, \dots, 0} & \ddots & & \\ \vdots & & \ddots & \\ & & & G_{0, \dots, 0} + sC_{0, \dots, 0} \end{vmatrix} \quad (13)$$

According to Theorem 1, both G_{AP} and C_{AP} are block lower

triangular with constant diagonal blocks. Therefore, the matrix $G_{AP} + sC_{AP}$ is also block lower triangular with constant diagonal block. The determinant in (13) is equal to the product of the determinants of the diagonal blocks:

$$|G_{AP} + sC_{AP}| = |G_{0,\dots,0} + sC_{0,\dots,0}| \cdots |G_{0,\dots,0} + sC_{0,\dots,0}| \quad (14)$$

Note that $G_{0,\dots,0}$ and $C_{0,\dots,0}$ in (14) are the nominal system matrices. Therefore, equation (14) implies that the parameterized system contains repeated poles. However, the values of all these poles are exactly equal to those of the original nominal system. This conclusion proves Theorem 2, since the stability of a linear system is uniquely determined by its poles. ■

By matching the Taylor expansion coefficients for ε_i , the parameterized system in (8) provides the governing equations that $x_{i_1,\dots,i_k}(s)$ should satisfy. If we further expand $x_{i_1,\dots,i_k}(s)$ by Taylor series in s and match the expansion coefficients (i.e. the moments), the multi-parameter moments of the transfer function H (defined in (3)) can be matched. In the following subsection, we show how to simultaneously match the moments for all $x_{i_1,\dots,i_k}(s)$ using Krylov subspace projection.

3.2 Implicit Moment Matching for s

Unlike the aforementioned explicit moment matching for the process parameters ε_i , the moments of the frequency s of all $x_{i_1,\dots,i_k}(s)$ are *implicitly* matched by Krylov subspace projection. The Krylov subspace can be generated using the Arnoldi algorithm [8]. Fig. 5 summarizes a simplified implementation of the Arnoldi algorithm. Note that the computations in Fig. 5 are completely numerical, since they only involve operations on G_{AP} , C_{AP} and B_{AP} in (8). These three matrices are numerical, while the matrix L_{AP} in (8) contains the symbols ε_i for modeling process variations.

1. Start from G_{AP}, C_{AP}, B_{AP} in (8).
2. Solve $G_{AP}X_0 = B_{AP}$ for X_0 .
3. Orthogonalize and normalize each column in X_0 .
For $k = 1, 2, \dots$
4. $V = C_{AP}X_{k-1}$.
5. Solve $G_{AP}X_k = V$ for X_k .
6. Orthogonalize X_k to all X_i ($i = 0, 1, 2, \dots, k-1$).
7. Orthogonalize and normalize each column in X_k .
End For
8. $X = [X_0 X_1 X_2 \dots]$.

Fig. 5. Simplified Arnoldi algorithm.

After the Krylov subspace X is constructed in Fig. 5, the parameterized reduced-order interconnect model is achieved by:

$$\begin{aligned} \tilde{G}_{AP} &= X^T G_{AP} X & \tilde{C}_{AP} &= X^T C_{AP} X \\ \tilde{B}_{AP} &= X^T B_{AP} & \tilde{L}_{AP} &= X^T L_{AP} \end{aligned} \quad (15)$$

Theorem 3: An M -th order model created by (15) matches $M\varepsilon \lfloor M/n \rfloor$ multi-parameter moments, where $M\varepsilon$ is the total number of the Taylor expansion terms in (4), n is the number of the system inputs and the operator $\lfloor \bullet \rfloor$ stands for the truncation to the nearest integer toward zero.

Proof: The authors in [8] prove that an M -th order model can match the first $\lfloor M/n \rfloor$ moments for all state variables when using the Arnoldi algorithm in Fig. 5 to generate the Krylov subspace. Therefore, an M -th order model in (15) matches the first $\lfloor M/n \rfloor$ moments for all $x_{i_1,\dots,i_k}(s)$ in (8)-(11). On the other hand, each

$x_{i_1,\dots,i_k}(s)$ is one of the Taylor expansion coefficients for $x(\varepsilon_1, \dots, \varepsilon_k, s)$, as shown in (4). If the total number of the Taylor expansion terms in (4) is $M\varepsilon$, the total number of the matched multi-parameter moments is $M\varepsilon \lfloor M/n \rfloor$. ■

According to Theorem 3, the final model size M is independent on $M\varepsilon$ using the proposed CORE algorithm. However, increasing $M\varepsilon$ will consequently increase the parameterized system size in (8), thereby making the Krylov subspace computation more expensive. In many practical applications, a great number of the symbols ε_i might be required to model both inter-die and intra-die variations. As such, even a quadratic expansion in (4) can result in an extremely large system in (8). Next, we propose a *recursive Arnoldi* algorithm to address this computation cost problem.

For large systems, the most expensive operation during Krylov subspace generation is to solve the linear equations in Step 2 and Step 5 of Fig. 5. Remember that the system matrix G_{AP} is block lower triangular with constant diagonal block (Theorem 1). Given a linear equation whose coefficient matrix has such a special property:

$$\begin{bmatrix} E_{11} & 0 & 0 & \cdots \\ E_{21} & E_{11} & 0 & \cdots \\ E_{31} & E_{32} & E_{11} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \end{bmatrix} \quad (16)$$

the above linear equation can be efficiently solved using the recursive algorithm in Fig. 6.

1. Start from (16).
2. Compute the LU factorization $E_{11} = LU$.
For $k = 1, 2, \dots$
3. $H_k = H_k - \sum_{i=1}^{k-1} E_{ki} F_i$.
4. Solve $LV_k = H_k$ for V_k .
5. Solve $UF_k = V_k$ for F_k .
End For

Fig. 6. Recursive algorithm for solving the linear equation (16).

The recursive algorithm in Fig. 6 can significantly reduce the computation cost for two reasons. Firstly, instead of directly solving the original large linear equation, the recursive algorithm repeatedly solves a number of small equations with the coefficient matrix E_{11} . The size of E_{11} is much smaller than the original equation size in (16). Secondly, since the coefficient matrix in (16) has the constant diagonal block E_{11} , the LU factorization is only required once.

The recursive algorithm in Fig. 6 can be applied to solve the linear equations in Step 2 and Step 5 of Fig. 5, which are the most expensive operations for generating the Krylov subspace. In this paper, such an Arnoldi algorithm with recursive linear equation solver is called the *recursive Arnoldi* algorithm, while the traditional Arnoldi algorithm with direct linear equation solver in [8] is called the *direct Arnoldi* algorithm. As will be demonstrated by the numerical examples in Section 4, the proposed recursive Arnoldi algorithm can achieve up to 10x speedup in parameterized interconnect order reduction, compared with the direct Arnoldi algorithm used in PRIMA [8].

3.3 Summary

Fig. 7 summarizes the overall implementation of CORE. Compared with other traditional approaches, CORE offers several

advantages in three aspects.

1. Start from the MNA equation (1) where x , G and C are expanded by Taylor series in (4)-(5).
2. Apply explicit moment matching on the process parameters ε_i and formulate the parameterized system (8)-(11).
3. Apply the recursive Arnoldi algorithm, i.e. the Arnoldi algorithm in Fig. 5 with the recursive linear equation solver in Fig. 6, to generate the Krylov subspace X .
4. Apply the projection in (15) to generate the reduced-order model.

Fig. 7. Summary of the CORE implementation.

- *Flexibility.* While the algorithm proposed in [7] must use the same expansion order for both the process parameters ε_i and the frequency s , CORE matches the moments for ε_i and s separately, thereby providing the flexibility of selecting different expansion orders for ε_i and s . Note that, in many practical applications, the process parameters ε_i only vary about $\pm 30\%$, while the frequency s must cover a very wide range, e.g. from 0 Hz up to 10 GHz. Therefore, it is important to distinguish ε_i and s and apply different expansion orders for each of them.
- *Compactness.* In order to match the same number of the multi-parameter moments, CORE can generate a compact model whose size (i.e. order) is much smaller than that created in [7]. As proved in Theorem 3, an M -th order model generated by CORE can match $M\varepsilon \lfloor M/n \rfloor$ multi-parameter moments, while using the same model size only matches $\lfloor M/n \rfloor$ moments in [7]. Note that $M\varepsilon$ (i.e. the total number of the Taylor expansion terms in (4)) can be very large in many practical applications, which makes CORE superior to the traditional approach in [7]. Table 1 shows a comparison on model sizes, assuming that the system is single-input and there are 10 variational process parameters ε_i . In Table 1, we use the same Taylor expansion order for all ε_i and s , since the algorithm in [7] cannot apply different expansion orders for each of them.

Table 1. Model order for a single-input system with 10 variational process parameters

Expansion Order	1	3	5	7	9	
# of Matched Moments	12	364	4368	32K	168K	
Model Order	[7]	12	364	4368	32K	168K
Order	CORE	1	3	5	7	9

- *Scalability.* Since the proposed explicit moment matching scheme results in the block lower triangular matrix G_{AP} , the recursive Arnoldi algorithm can be applied to quickly construct the Krylov subspace. This makes CORE scalable to large-size problems. The total computation cost for CORE is roughly equal to $M\varepsilon$ times the nominal order reduction cost for the same circuit, where $M\varepsilon$ is the total number of the Taylor expansion terms in (4). The parameterized systems derived by other approaches (e.g., [6]) do not have the block lower triangular structure and, therefore, the recursive Arnoldi algorithm cannot be applied to reduce the computation cost.

However, the passivity of the CORE model is difficult to prove. The main difficulty is that the resulting reduced-order model is a parameterized one (see (15) and Fig. 4), including the variational process parameters ε_i . It is difficult to achieve passivity

with variational ε_i . Fortunately, passivity is not always required in many simulation algorithms. For example, the algorithm in [10] does not require passive interconnect models by using the successive chord method for nonlinear iteration.

4. Numerical Examples

In this section we demonstrate the efficacy of CORE using several circuit examples. In all examples, the modeling accuracy is compared using the 50% delay errors that are measured by 10^4 Monte Carlo simulations with step input. All experiments are run on a SUN — 1 GHz server.

4.1 RC Clock Tree

An RC clock tree, including 801 RC elements, is synthesized using the IBM CMOS 90 nm process and is distributed on three metal layers: M5, M6 and M7. In this example, three symbols $\varepsilon_1 \sim \varepsilon_3$ are respectively utilized to model the metal width variations of M5-M7 up to $\pm 30\%$.

Table 2 compares the reduced-order model size and accuracy for three different methods: the quadratic fitting approach [4], the traditional multi-parameter moment matching approach [7] and CORE. Compared with [7], our CORE model has smaller size while matching much more multi-parameter moments. The frequency response of the RC clock tree at one Monte Carlo sampling point is plotted in Fig. 8.

Table 2. Reduced-order model size and accuracy for the RC clock tree

		[4]	[7]	CORE
Model Order		20	36	20
# of Matched Moments		—	36	200
Delay Error	Min	0.02%	1.33%	0.96%
	Avg	61.65%	2.15%	1.45%
	Max	635.05%	3.55%	2.57%

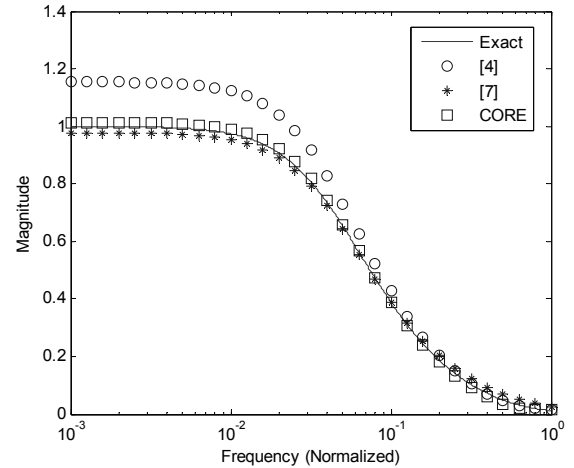


Fig. 8. Frequency response of the RC clock tree at one Monte Carlo sampling point ($\varepsilon_1 = -0.3\%$, $\varepsilon_2 = 9.9\%$, $\varepsilon_3 = -6.0\%$).

4.2 Scaling with Problem Size

The second circuit example is a larger RC network that includes 1275 RC elements.

A. Modeling Inter-Die Variations

Table 3. Reduced-order model size and accuracy for the RC network (inter-die only)

		[4]	[7]	CORE
Model Order		10	21	10
# of Matched Moments		–	21	60
Delay Error	Min	0.00%	0.06%	0.00%
	Avg	0.00%	11.19%	0.10%
	Max	0.00%	41.19%	1.49%

Firstly, we only consider the inter-die variations on metal width and thickness up to $\pm 30\%$. Table 3 compares the reduced-order model size and accuracy for different approaches. Note that the quadratic fitting approach [4] is extremely accurate in this example when only inter-die variations are considered.

B. Impact of Intra-Die Variations

Next, we partition the entire interconnect network into 10 individual regions based on the physical locations. The intra-die metal width and thickness variations in each region are represented by individual symbols. In this example, we have two symbols for modeling the inter-die width and thickness variations and 20 additional symbols (two symbols in each region) for modeling the intra-die width and thickness variations.

Table 4 compares the reduced-order model size and accuracy when different inter-die and intra-die variation scales are selected. As we would expect, the quadratic fitting approach [4] has large modeling errors in the presence of intra-die variations. As shown in Table 4, CORE creates the most compact (small-size) and accurate parameterized interconnect models. After the intra-die variations reach $\pm 15\%$, the average modeling error of CORE is 10x smaller than those of the traditional approaches.

Table 4. Reduced-order model size and accuracy for the RC network (inter-die & intra-die)

		[4]	[7]	CORE	
Model Order		10	46	10	
# of Matched Moments		–	46	2760	
Delay Error	25% Inter & 5% Intra	Min	0.00%	0.00%	0.00%
		Avg	0.27%	1.46%	0.13%
		Max	6.77%	6.02%	1.56%
	15% Inter & 15% Intra	Min	0.00%	0.00%	0.00%
		Avg	3.81%	1.36%	0.11%
		Max	106.9%	5.43%	0.43%
	5% Inter & 25% Intra	Min	0.00%	0.00%	0.00%
		Avg	4.44%	1.72%	0.11%
		Max	261.8%	8.21%	0.50%

C. Computation Cost

Table 5 shows the model generation cost for three different approaches. In this example, the quadratic fitting approach [4] is the most expensive one, since it fits the quadratic functions for all reduced-order system matrices $\tilde{G}, \tilde{C} \in R^{10 \times 10}$ and $\tilde{B}, \tilde{L} \in R^{10 \times 1}$. Each quadratic fitting requires solving an over-determined linear equation. On the other hand, CORE is more expensive than the traditional multi-parameter moment matching approach [7]. This implies that more computation effort is required in order to match more multi-parameter moments by using CORE.

However, it should also be noted that the model simulation cost is typically more important than the model generation cost in

many interconnect analysis problems, since the parameterized interconnect model is only generated once and it can be repeatedly applied many times (e.g. for Monte Carlo simulation). Table 6 shows the Monte Carlo simulation time (10^4 samples) when using the reduced-order models to compute the step responses. In this example, using the parameterized models achieves up to 10x speedup, compared with the approach of directly applying PRIMA for interconnect order reduction at each Monte Carlo sampling point. In addition, due to the difference in model order, using the parameterized model extracted by [7] (Order = 46) is 2x slower than using the CORE model (Order = 10). This implies that the CORE model yields cheaper simulation cost, although it is more expensive to build.

Table 5. Reduced-order model generation cost (Sec.) for the RC network

[4]	[7]	CORE
751.71	0.99	12.07

Table 6. Monte Carlo simulation cost (Sec.) with 10^4 samples for the RC network

PRIMA	[4]	[7]	CORE
2042	147	346	147

It is also worth noting the significant runtime speedup by applying the proposed recursive Arnoldi algorithm for CORE. We implemented CORE with both the direct Arnoldi algorithm used in PRIMA [8] and the recursive Arnoldi algorithm proposed in Section 3.2. In this RC network example, the nominal MNA matrix size (equation (1)) is 767 and the augmented parameterized system size (equation (8)) is 2.12×10^6 (276x larger). Since these system matrices are typically sparse, the iterative GMRES algorithm [11] is utilized in the direct Arnoldi implementation for solving linear equations. The GMRES algorithm can significantly reduce the computation cost by taking advantage of the matrix sparsity. Even with the GMRES solver, the direct Arnoldi algorithm takes 168.27 seconds to generate the reduced-order model, which is 10x larger than that of the recursive Arnoldi algorithm (12.07 seconds) shown in Table 5.

4.3 RLC Network

Table 7. Reduced-order model size and accuracy for the RLC network (inter-die & intra-die)

		[4]	[7]	CORE	
Model Order		35	38	35	
# of Matched Moments		–	38	46550	
Delay Error	25% Inter & 5% Intra	Min	0.00%	13.17%	0.00%
		Avg	2.60%	20.29%	1.39%
		Max	99.94%	26.70%	4.67%
	15% Inter & 15% Intra	Min	0.00%	13.57%	0.00%
		Avg	9.11%	20.36%	1.24%
		Max	99.61%	25.31%	3.31%
	5% Inter & 25% Intra	Min	0.01%	12.94%	0.00%
		Avg	14.77%	20.31%	1.20%
		Max	99.62%	27.04%	3.78%

The final circuit example is an RLC network that consists of 123 RLC elements. We partition the entire interconnect network into 5 individual regions based on the physical locations. Three individual symbols are utilized to represent the intra-die metal

width, thickness and inductance variations in each region. In this example, we have 18 symbols in total to model both inter-die and intra-die variations.

A. Modeling Accuracy

Table 7 compares the reduced-order model size and accuracy for three different approaches. As shown in Table 7, the average modeling error of CORE is 7x (and 12x) smaller than those of the traditional approaches, when the intra-die variations reach $\pm 15\%$ (and $\pm 25\%$).

B. Simulation Cost

Table 8 shows the Monte Carlo simulation time (10^4 samples) when using the reduced-order models to compute the step responses. In this example, using the parameterized models achieves up to 7.5x runtime speedup, compared with the approach of directly applying PRIMA for interconnect order reduction at each Monte Carlo sampling point.

Table 8. Monte Carlo simulation cost (Sec.) with 10^4 samples for the RLC network

PRIMA	[4]	[7]	CORE
1757	234	285	232

5. Conclusions

In this paper we propose a novel parameterized interconnect order reduction algorithm, CORE, that can efficiently handle both inter-die and intra-die variations. CORE maximizes the number of the matched multi-parameter moments by using a novel two-step explicit-and-implicit moment matching scheme. As such, CORE can significantly reduce the modeling error without increasing the model complexity (i.e. size). As is demonstrated by the numerical examples, the modeling accuracy is improved by up to 10x compared with other traditional techniques. In addition, a recursive Arnoldi algorithm is proposed to reduce the computation cost of the model generation. Compared with the direct Arnoldi algorithm utilized in PRIMA [8], the proposed recursive Arnoldi algorithm can achieve up to 10x speedup for parameterized interconnect order reduction in our tested examples. We also demonstrate that using parameterized models can achieve up to 10x runtime speedup in Monte Carlo simulation with 10^4 samples, compared with the approach of directly applying PRIMA for interconnect order reduction at each Monte Carlo sampling point.

6. Acknowledgements

The authors wish to thank Dr. Frank Liu, IBM Austin Research Laboratories, for providing the RC clock tree example. This work has been supported by the MARCO Focus Center for Circuit & System Solutions (C2S2, www.c2s2.org) under contract 2003-CT-888 and the Semiconductor Research Cooperation (SRC, www.src.org) under contract 2005-TJ-1298.

7. References

- [1] M. Celik, L. Pileggi and A. Odabasioglu, *IC Interconnect Analysis*, Kluwer Academic Publishers, 2002.
- [2] S. Nassif, "Modeling and analysis of manufacturing variations," *IEEE CICC*, pp. 223-228, 2001.
- [3] M. Orshansky; L. Milor and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction," *IEEE Trans. Semiconductor Manufacturing*, vol. 17, no. 1, pp. 2-11, Feb. 2004.
- [4] Y. Liu, L. Pileggi and A. Strojwas, "Model order-reduction of RC(L) interconnect including variational analysis," *IEEE/ACM DAC*, pp. 201-206, 1999.
- [5] P. Heydari and M. Pedram, "Model reduction of variable-geometry interconnects using variational spectrally-weighted balanced truncation," *IEEE/ACM ICCAD*, pp. 586-591, 2001.
- [6] J. Wang, P. Ghanta and S. Vrudhula, "Stochastic analysis of interconnect performance in the presence of process variations," *IEEE/ACM ICCAD*, pp. 880-886, 2004.
- [7] L. Daniel, O. Siong, L. Chay, K. Lee and J. White, "A multi-parameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models," *IEEE Trans. CAD*, vol. 23, no. 5, pp. 678-693, May. 2004.
- [8] A. Odabasioglu, M. Celik and L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. CAD*, vol. 17, no. 8, pp. 645-654, Aug. 1998.
- [9] J. Phillips, L. Daniel and L. Silveira, "Guaranteed passive balancing transforms for model order reduction," *IEEE Trans. CAD*, vol. 22, no. 8, pp. 1027-1041, Aug. 2003.
- [10] E. Acar, F. Dartu and L. Pileggi, "TETA: transistor-level waveform evaluation for timing analysis," *IEEE Trans. CAD*, vol. 21, no. 5, pp. 605-616, May. 2002.
- [11] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, 1996.