BSMOR: Block Structure-preserving Model Order Reduction

Hao Yu, Lei He
Electrical Engineering Dept., UCLA

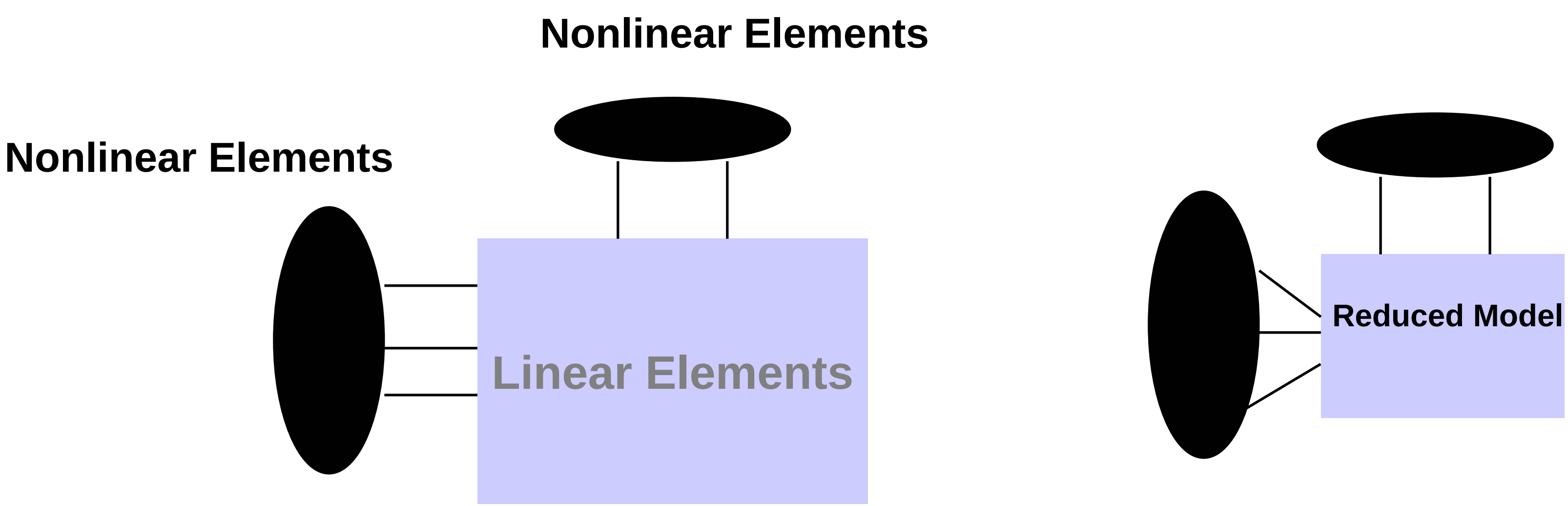
Sheldon S.D. Tan
Electrical Engineering Dept., UCR

http//:eda.ee.ucla.edu

This project is founded by SRC#1100, NSF, and Analog Devices

Motivation

- Deep submicron design needs to consider a large number of linear elements
  - Interconnect, Substrate, P/G grid, and Package
- Accurate extraction leads to the explosion of data storage and runtime
- Need efficient macro-model



Model Order Reduction

Outline

- Review of Model Order Reduction
  - Grimme's projection theorem
  - PRIMA
- Block Structure-preserving Model Order Reduction
  - BSMOR Method
  - Properties of BSMOR
  - Bordered-block-diagonal Decomposition
- Experiment Results
- Conclusions and Future Work

Background

- **MNA Matrix**

State → $x(s)$, Input → $u(s)$

$$Gx(s) + sCx(s) = Bu(s)$$

Output → $y(s)$

$$y(s) = Lx(s)$$

- **Krylov subspace**

$$A = -(G + s_0 C)^{-1} C, R = (G + s_0 C)^{-1} B$$

$$\Rightarrow x \in \operatorname{span}\{R, AR, A^2 R, \cdots\}$$

The qth-order Krylov subspace

$$K(A, R; q) \equiv span\{R, AR, A^2 R, \ldots A^{q-1} R\}$$

$$q = ceil(n / n_p)$$

n: dimension of the spanned space

n_p: number of ports

Grimme's Projection Theorem

If $V = span\{v_1,\ldots,v_q\} \supseteq K(A,R;q)$

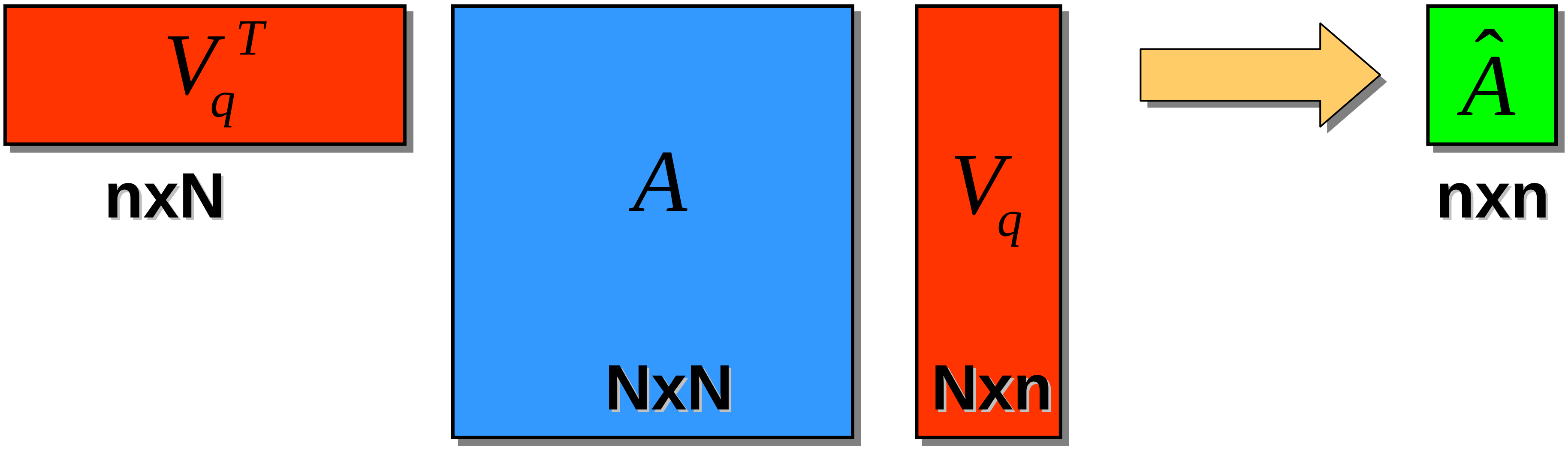
The projected $\hat{\mathrm{H}}(s)$ by V macthes the first q moments of H(s)

$H(s){:}L(I\text{-}sA)^{-1}R$

$\hat{H}(s){:}\hat{L}(I\text{-}s\hat{A})^{-1}\hat{R}$

$\hat{L} = V^T L, \hat{R} = VR, \hat{A} = V^T AV$

PRIMA



- **To improve matching accuracy**
  - **Apply Arnoldi orthnormalization to obtain independent basis**

$$V_q^T V_q = I$$

- **To preserve passivity**
  - **Project G and C respectively in form of congruence transformation**

$$\hat{L} = V_q L, \hat{B} = V_q B, \hat{G} = V_q^T G V_q, \hat{C} = V_q^T C V_q$$

Limitation of PRIMA

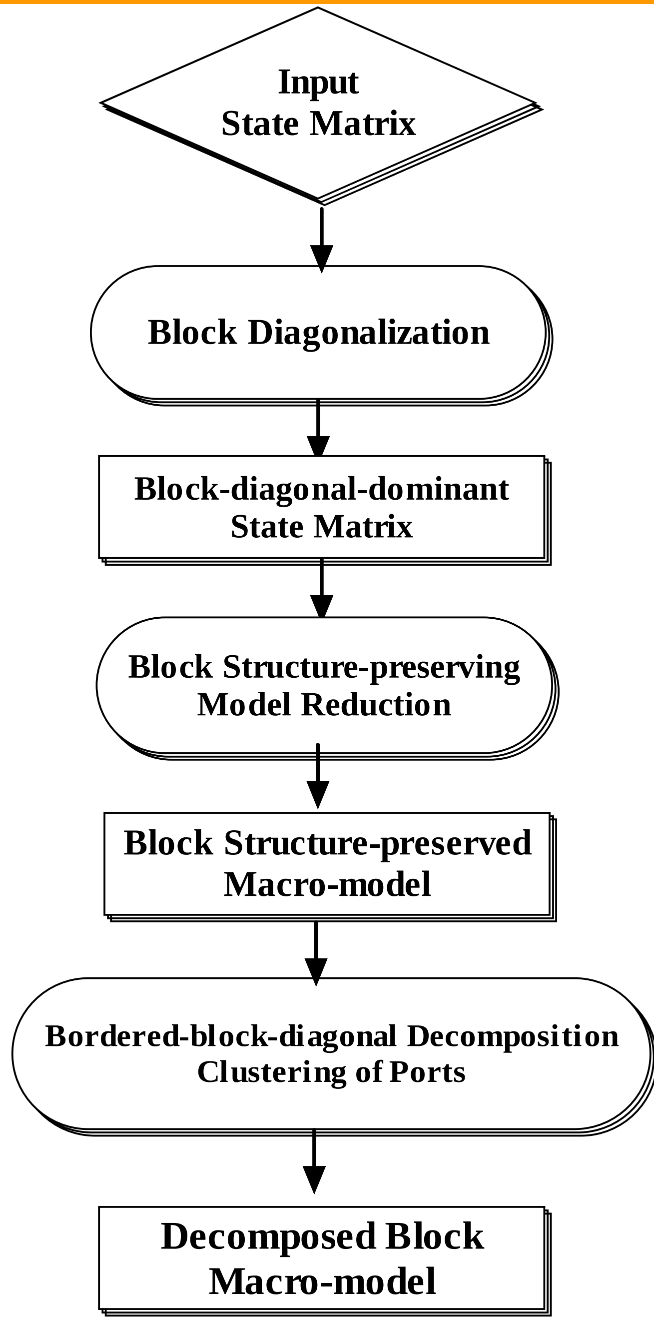
- **Flat-projection loses the substructure information of the original state matrices**
  - **Original state matrices are sparse, but reduced state matrices are dense**
  - **It becomes inefficient to match poles for structured state matrices**
- **It can not handle large number of ports efficiently**
  - **Accuracy degrades as port number increases**
  - **Reduced macro-model in form of flat port matrix is too large and dense to analyze**

Our Contribution

- Basic Idea
  - **Explore the substructure information by partitioning the state matrices**
  - **Partition the projection matrix accordingly and construct a new block-structured projection matrix**

- Properties
  - Reduced model matches *mq* poles for the block diagonal state matrices
  - Reduced model matches q dominant poles exactly and (m-1)q poles approximately for general state matrices with an additional block diagonalization procedure
  - Reduced state matrices are sparse
  - Reduced model can be further decomposed into blocks, each with a small number of ports

BSMOR Flow


Input
State Matrix
Block Diagonalization
Block-diagonal-dominant
State Matrix
Block Structure-preserving
Model Reduction
Block Structure-preserved
Macro-model
Bordered-block-diagonal Decomposition
Clustering of Ports
Decomposed Block
Macro-model

Outline

- **Review of Model Order Reduction**
  - **Grimme's projection theorem**
  - **PRIMA**
- **Block Structure-preserving Model Order Reduction**
  - **BSMOR Method**
  - **Properties of BSMOR**
  - **Bordered-block-diagonal Decomposition**
- **Experiment Results**
- **Conclusions and Future Work**

BSMOR Method

- **Given m blocks within G, C and B matrices (block-diagonal-dominant)**

$$G = \begin{bmatrix} G_{11} & \cdots & G_{1m} \\ \vdots & \ddots & \vdots \\ G_{m1} & \cdots & G_{mm} \end{bmatrix}, C = \begin{bmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mm} \end{bmatrix}, B = \begin{bmatrix} B_1 \\ \vdots \\ B_m \end{bmatrix}$$

- **Construct a new projection matrix V-tilde with the block structure accordingly based on V from PRIMA**

$$V = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \rightarrow \tilde{V} = \begin{bmatrix} v_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_m \end{bmatrix}$$

- **Block Structure-preserved Projection**

$$\tilde{G} = \tilde{V} G \tilde{V}, \tilde{C} = \tilde{V} C \tilde{V}, \tilde{B} = \tilde{V} B$$

Properties (I)

- **q-moment matching**

$\tilde{H}_q(s)$ matches the first q moments of $H(s)$:

$$K(A,R;q) \subseteq V_q \subseteq \tilde{V}_q$$

- **Passivity preservation**

$$\tilde{V}_q^T \tilde{V}_q = I$$

$$\tilde{V}_q^T (G+G^T)\tilde{V}_q \succ 0, \quad \tilde{V}_q^T (C+C^T)\tilde{V}_q \succ 0$$

Properties (II)

- **Block structure-preserving**
  - **Results in a sparse reduced matrices (but not PRIMA)**
  - **Enable further block-ports decomposition (but not PRIMA)**

$$G = \begin{bmatrix} G_{11} & \cdots & G_{1m} \\ \vdots & \ddots & \vdots \\ G_{m1} & \cdots & G_{mm} \end{bmatrix}$$

$$BSMOR: \tilde{G}_q = \begin{bmatrix} v_1^T G_{11} v_1^T & \dots & v_1^T G_{1m} v_m^T \\ \vdots & \ddots & \vdots \\ v_1^T G_{m1} v_1^T & \dots & v_1^T G_{m1} v_m^T \end{bmatrix}, \quad PRIMA: \hat{G}_q = \sum_{i=1}^{m} \sum_{j=1}^{m} v_i^T G_{ij} v_j^T$$

Properties (III)

Theorem: the reduced model matches mq poles, if G and C matrices are block diagonal

Proof: the resulted Heisenberg matrix A-tilde is block diagonal

$$\tilde{A} = \tilde{V}_q^T A \tilde{V}_q$$

$$= diag[\tilde{A}_{11}, ..., \tilde{A}_{mm}]$$

$$= \begin{bmatrix} v_1^T(G_{11}^{-1}C_{11})v_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_m^T(G_{mm}^{-1}C_{mm})v_m \end{bmatrix}$$

$$\text{If } Null = eigen(\tilde{A}_{11}) \cup \cdots eigen(\tilde{A}_{mm})$$

$$\tilde{A} \text{ matches } mq \text{ poles of } A$$

BBDC Analysis

- **Resulted MIMO macro-model has preserved block structure but has dense couplings between blocks**
  - **Each block is now represented by a subset of ports**

$$Y(s)=\begin{bmatrix} Y_{11} & \cdots & Y_{1m} \\ \vdots & \ddots & \vdots \\ Y_{m1} & \cdots & Y_{mm} \end{bmatrix}$$

- **Represent it into bordered-block-diagonal form with a global coupling block (with branch addmaince Y_{00})**

$$\begin{bmatrix} Y_{11} & \cdots & 0 & C_{10} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & Y_{mm} & C_{m0} \\ C_{10}^T & \cdots & C_{m0}^T & -Y_{00}^{-1} \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_m \\ I_0 \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ I_m \\ 0 \end{bmatrix}$$

- **Enable multi-level partioned solution by branch-tearing [Wu:TCAS'76]**

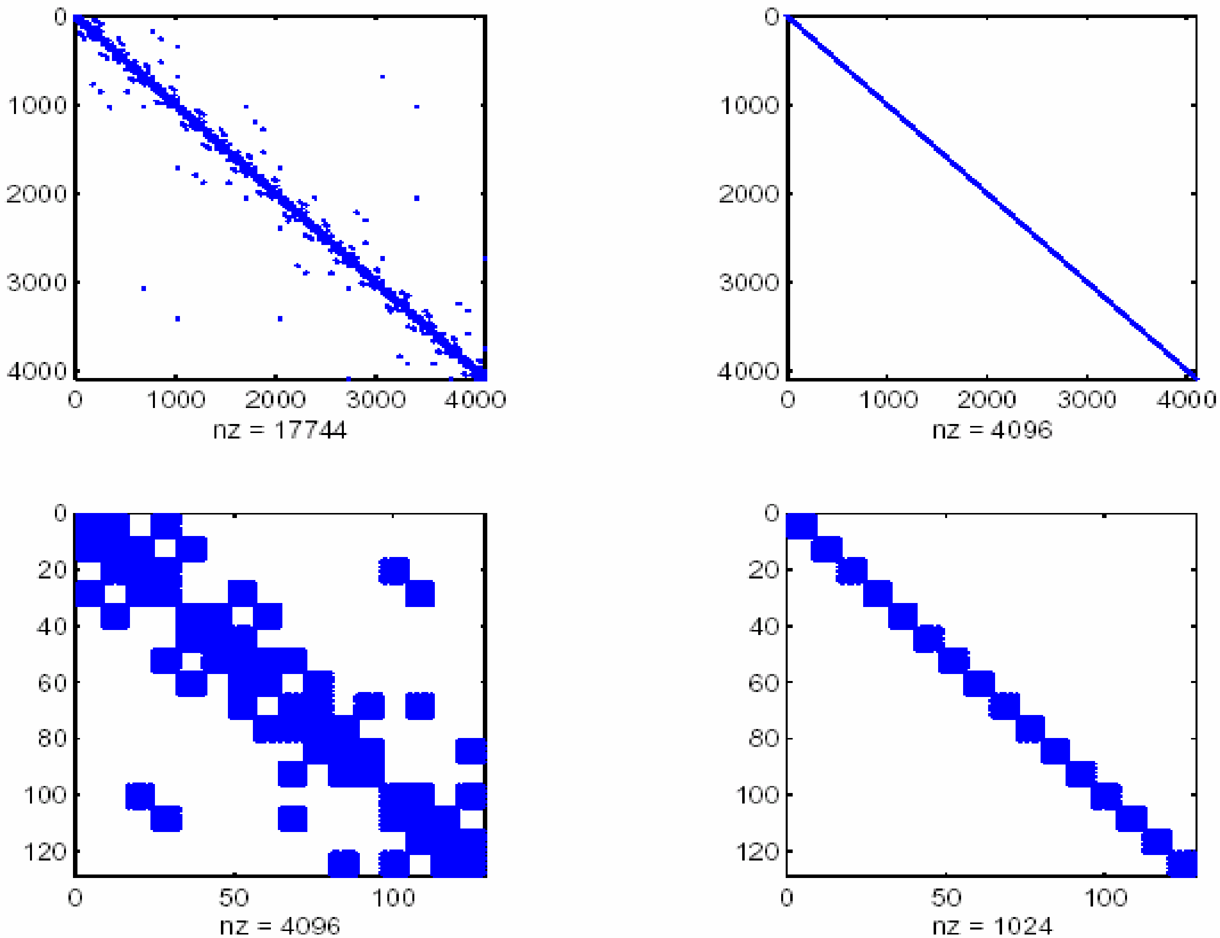
Outline

- Review of Model Order Reduction
  - Grimme's projection theorem
  - PRIMA
- Block Structure-preserving Model Order Reduction
  - Block Identification
  - Structured Projection
  - Bordered-block-diagonal Analysis
- Experiment Results
- Conclusions and Future Work

Sparsity

Before reduction

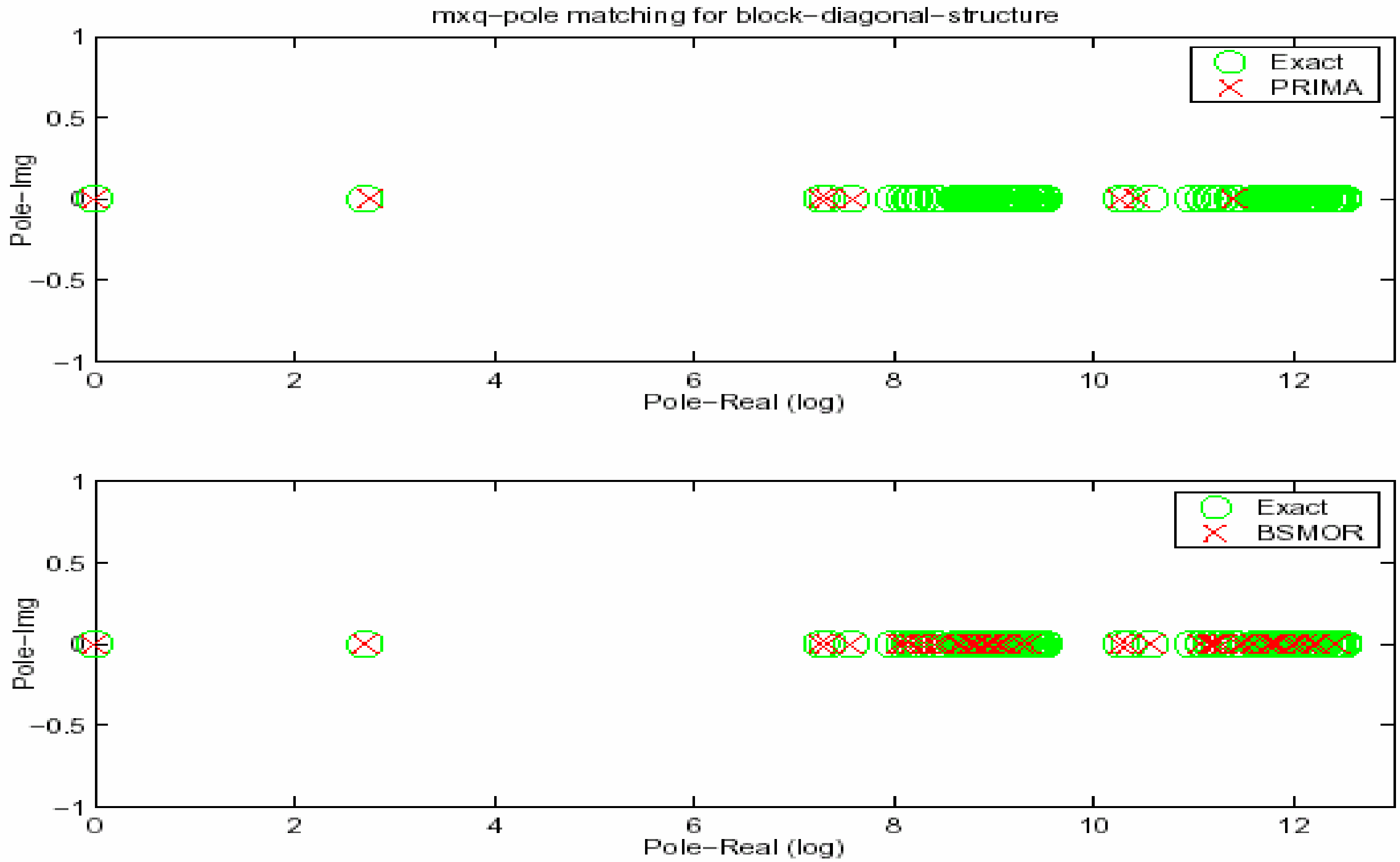
After reduction



Non-zero patterns for G, C matrices of a uniform RC-mesh (256x256) before and after a 16×16 BSMOR reduction with 8 iterations, where NZ is the number of non-zero.

- 16x16-BSMOR shows 72% and 93% sparsity for G and C matrices of a 256x256 RC-mesh
- Matrices reduced by PRIMA are fully dense

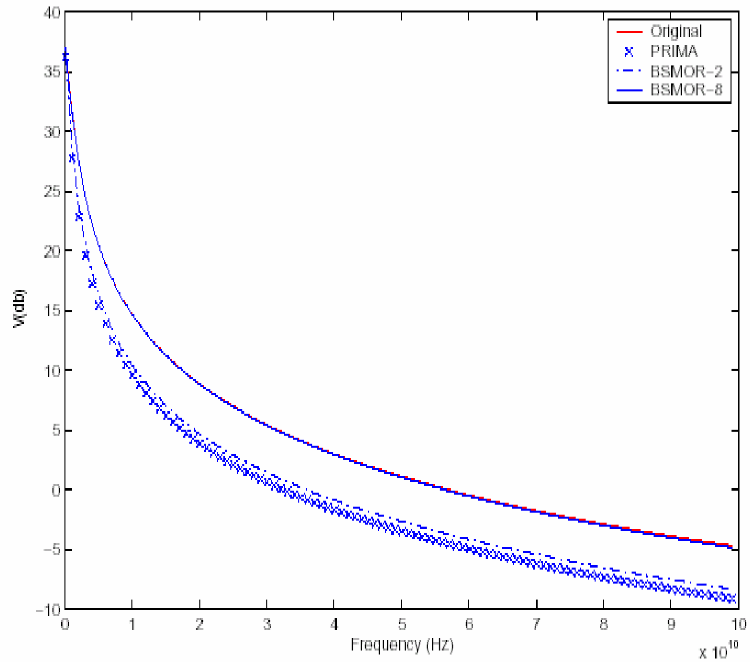
Pole Matching



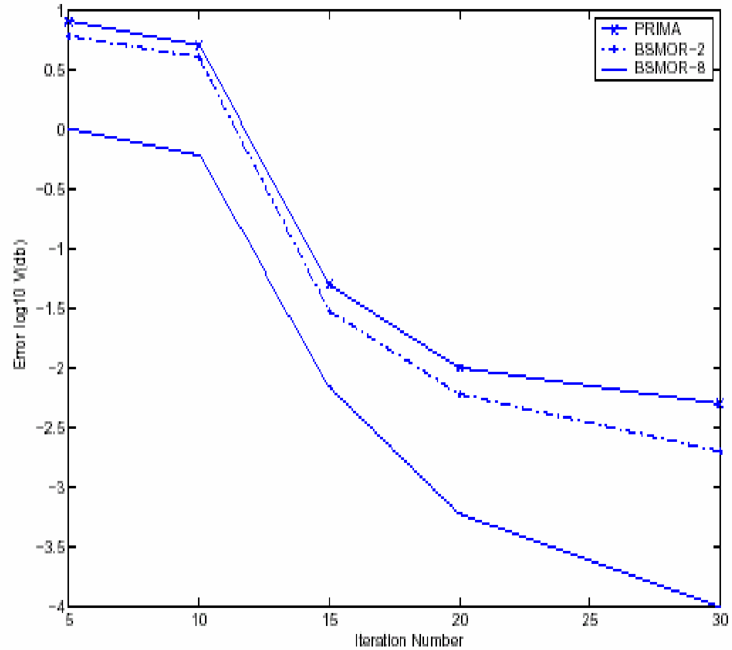
mq poles matching comparison of PRIMA and BSMOR.

- For a non-uniform mesh composed by 32 sub-meshes
- 8x8-BSMOR exactly matches 8 poles and closely matches additional 56 poles
- PRIMA only matches 8 poles

Frequency Response



Frequency responses of the BSMOR, PRIMA, and original model at one port of a uniform mesh (256x256) after 10 iterations.



n errors of the frequency response of the BSMOR and PRIMA for increasing order models of a uniform mesh (256x256) up to 20GHz.

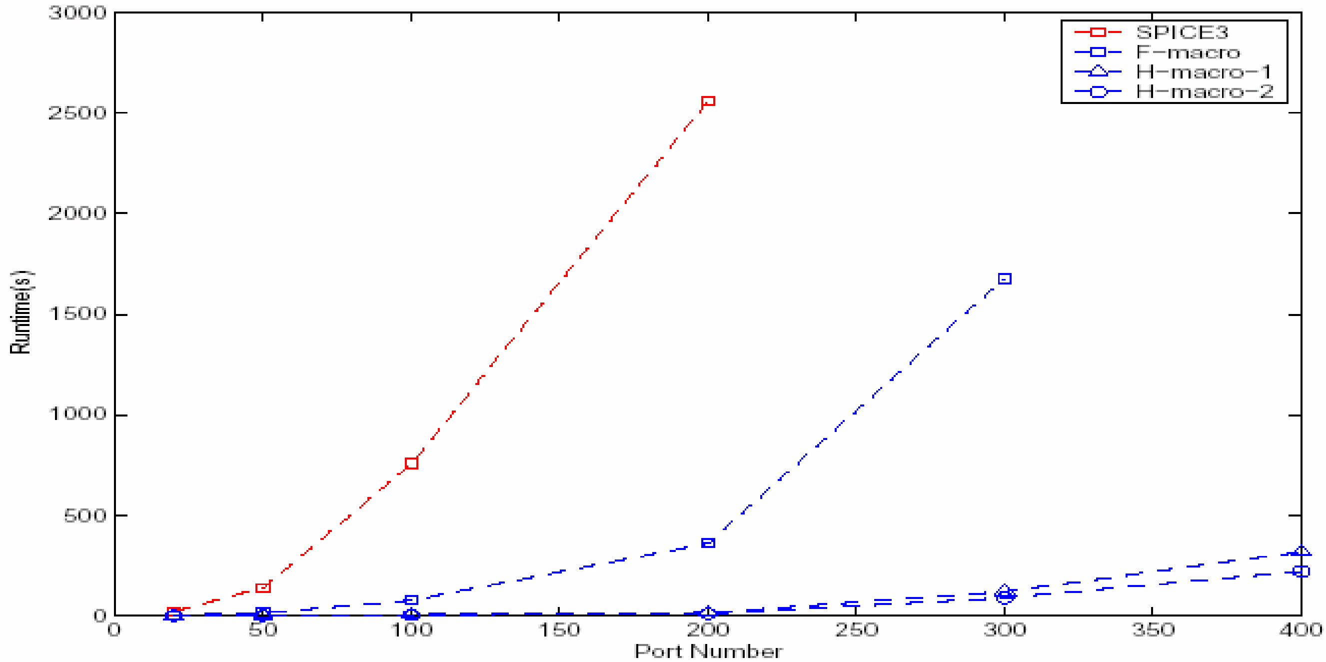
- Comparison of 2x2-BSMOR, 8x8-BSMOR and PRIMA
  - 8x8-BSMOR has best accuracy in all iterations of block Arnoldi procedures
  - Increasing block number leads to more matched poles and hence improved accuracy

Reduction Time

Ckt	elements	err-bound	BSMOR			PRIMA	
			block#	iter#	time	iter#	time
mesh1	1K	1e-8	2x2	4	0.03s	10	0.09s
mesh2	10K	1e-8	8x8	6	0.07s	20	0.28s
mesh3	80K	1e-6	16x16	6	0.42s	30	3.82s
mesh4	160K	1e-6	16x16	6	5.14s	40	46.98s
mesh5	320K	1e-4	32x32	6	10.27s	60	104.62s
mesh6	1M	1e-4	64x64	8	240.22s	80	4982.76s

- **Under the same error bound, BSMOR has 20X smaller reduction time than PRIMA**
  - **Fewer iterations are needed by BSMOR**

Simulation Time



The scalability trend of simulation time for the original model, flat macro-model, partitioned models with different hierarchical levels.

- **The dense macro-model by PRIMA leads to a similar runtime growth as the original model**
- **Level-(1,2) BBDC has a much slower growth (up to 30X simulation time reduction)**

Conclusions and Future Work

- **BSMOR achieves higher model reduction efficiency and accuracy by leveraging and preserving the structure information of the input state matrices**
- **Reduced block can be further hierarchically analyzed that further boosts the efficiency**

- **How to find the best way to do the block diagonalization**
- **How to apply BSMOR to system that has strong inductive couplings**

- **Updates available at http://eda.ee.ucla.edu**