Scalable Symbolic Model Order Reduction

Yiyu Shi*, Lei He* and C. J. Richard Shi+

*Electrical Engineering Department, UCLA

+Electrical Engineering Department, University of Washington, Seattle

This work is supported in part by NSF CAREER Award CCR-0306682, Defense Advanced Projects Agency under contract W31P4-05-O-R164 and in part by Defense Threat Reduction Agency under contract #FA9451-06-C-0384.

Motivation

- With the advance of design technology, especially when we have entered the nano regime. Plenty of algorithms exist in literature discussing how to analyze and simulate those symbolic circuits.
  - All those methods are practical only if the circuit has a moderate size.
  - However, the circuits from physical extraction usually contain millions of nodes.
- Numerous model order reduction (MOR) techniques have been successfully applied to the reduction of linear large scale circuits over the past decade.
- However, despite their wide application, unsolved problems do exist when directly extending them to symbolic circuits.
  - Symbolic model order reduction is proposed accordingly [Shi:tcad'06]

Prior Art

- The idea of symbolic model order reduction (SMOR) was first introduced in [Shi:tcad'06], which contains three different methods
  - Symbolic isolation
    - It first removes all the symbols from the circuits, and the nodes to which the symbols are connected are modeled as ports.
    - The time and space complexity for the reduced model increases cubically with the number of ports, i.e., the number of symbols
  - Nominal projection
    - It uses the nominal values of the symbols to compute the projection matrix.
    - It is accurate only when the symbol values slightly deviate from the nominal value.
  - First order expansion
    - It uses the first order expansion of the matrix inversion and multiplication to find the projection matrix, which is first order matrix polynomial w.r.t. all the symbols.
    - Again, no large change is allowed for the symbols in order for the method to be accurate.

Major Contribution of Our Work

- This paper presents a scalable SMOR algorithm, namely S^2MOR.
  - We first separate the original multi-port multi-symbol system into a set of single-port systems by superposition theorem, and then integrate them together to form a lower-bordered block diagonal (LBBD) structured system.
  - Each block is reduced independently, with a stochastic programming to distribute the given overall model order between blocks for best accuracy. The entire system is efficiently solved by low-rank update.
  - Compared with existing SMOR algorithms, given the same memory space, S^2MOR improves accuracy by up to 78% at a similar reduction time. In addition, the factorization and simulation of the reduced model by S^2MOR is up to 17X faster.

Outline

- Port Separation and Model Reduction
- Simulation and Update of the Reduced Model
- Min-max Programming based Projection Order Decision
- Experimental Results
- Conclusions

Port Separation and Model Reduction

Symbolic MNA equation

$$(\mathbf{G}+s\mathbf{C})x+\sum_{i=1}^{a} P_i s_i \circ (P_i^T x) = \mathbf{B}u$$

$$y = \mathbf{L}^T x,$$

symbol i

Incidence vector for symbol i

$$w_i = s_i \circ (P_i^T x)$$

Symbol-less MNA equation

$$(\mathbf{G}+s\mathbf{C})x = \sum_{i=1}^{p} B_i u_i - \sum_{i=1}^{a} P_i w_i,$$

the i[th] column of B matrix

superposition theorem

Symbol-less MNA equation set

$$(\mathbf{G}+s\mathbf{C})x^{(i)} = \begin{cases} B_i u_i, & 1 \le i \le p \\ P_{i-p} w_{i-p}, & p+1 \le i \le p+a \end{cases}$$

$$x = \sum_{i=0}^{p+a} x^{(i)},$$

Port Separation and Model Reduction

We can further show that the symbol-less MNA equation set

$$(\mathbf{G} + s\mathbf{C})x^{(i)} = \begin{cases} B_i u_i, & 1 \le i \le p \\ P_{i-p} w_{i-p}, & p+1 \le i \le p+a \end{cases}$$

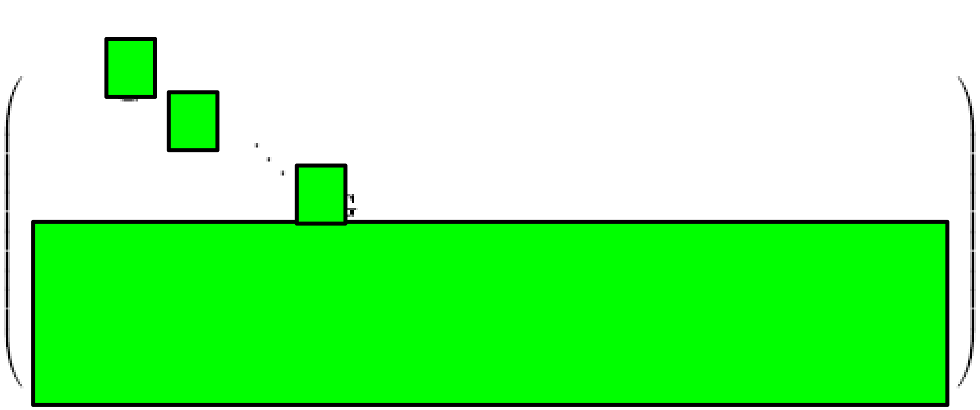
$$x = \sum_{i=0}^{p+a} x^{(i)},$$

can be expressed in the following compact form

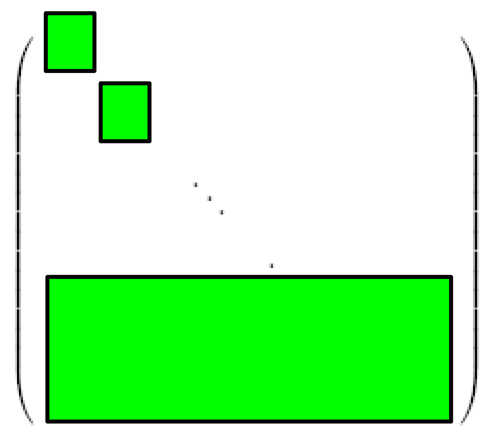
$$(\hat{\mathbf{G}} + s\hat{\mathbf{C}})z = \hat{\mathbf{B}}u$$

where

$\hat{\mathrm{G}} =$



$\hat{\mathrm{C}} =$



Note that G and C are lower bordered block diagonal matrices (LBBD matrices)

Port Separation and Model Reduction

- We can prove that if orthonormalized matrices V_i satisfies

$$\mathbf{V}_i \subseteq \begin{cases} \kappa_q\{\mathbf{G}, \mathbf{C}, B_i\} & 1 \leq i \leq p \\ \kappa_q\{\mathbf{G}, \mathbf{C}, P_{i-p}\} & p+1 \leq i \leq p+a \end{cases}$$

q-th order Krylov subspace

then with the block-diagonal projection matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & & & \\ & \mathbf{V}_2 & & \\ & & \ddots & \\ & & & \mathbf{V}_{p+a} \end{pmatrix},$$

the first q moments of the reduced system and the original systems are exactly matched. In addition, the reduced system still keeps the LBBD strucuture.

Outline

- Port Separation and Model Reduction
- Simulation and Update of the Reduced Model
- Min-max Programming based Projection Order Decision
- Experimental Results
- Conclusions

Simulation and Update of the Reduced Model

- We can fully utilize the LBBD structure of the reduced model. As an example, the Gr matrix can be expressed as

$$\hat{G}_r = \mathcal{D} + \mathcal{L}\mathcal{H}^T$$

where

$$\mathcal{D} = \begin{pmatrix} G_{r,1} & & \\ & \ddots & \\ & & G_{r,p+a} \end{pmatrix} \qquad \mathcal{L} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -P_{r,1}^{p+1} & 0 & \cdots & 0 \\ 0 & -P_{r,2}^{p+2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdot & 0 \\ 0 & 0 & \cdots & -P_{r,a}^{p+a} \end{pmatrix}$$

$$\mathcal{H} = \begin{pmatrix} s_1 \circ P_{r,1}^{1T} & s_2 \circ P_{r,2}^{1T} & \cdots & s_a \circ P_{r,a}^{1T} \\ s_1 \circ P_{r,1}^{2T} & s_2 \circ P_{r,2}^{2T} & \cdots & s_a \circ P_{r,a}^{2T} \\ \vdots & \vdots & \vdots & \vdots \\ s_1 \circ P_{r,1}^{(p+a)T} & s_2 \circ P_{r,2}^{(p+a)T} & \cdots & s_a \circ P_{r,a}^{(p+a)T} \end{pmatrix},$$

Simulation and Update of the Reduced Model

- Then from matrix inversion lemma, the following algorithm to solve $(\hat{\mathbf{G}}_r + s_0\hat{\mathbf{C}}_r)x = \hat{\mathbf{B}}_r u$ can be easily derived.
  - First factorize $\mathcal{E} = \mathcal{D} + s_0\hat{C}_r$ which is block diagonal
  - Then factorize a small matrix

$$\mathcal{M} = \mathbf{I} + \mathcal{H}^T(\mathcal{D} + s_0\hat{C}_r)\mathcal{L}\ (\mathcal{M} \in R^{a \times a})$$

  - Then we solve $\mathcal{E}x' = \mathbf{B}_r u$
  - Next, solve $\mathcal{M}x'' = x'$.
  - Finally solve $\mathcal{E}x''' = \mathcal{L}x''$
  - And the solution of the original system can be obtained as

$$x = x' - x'''$$

- The main advantage of the above algorithm is that instead solving the full system, we turn to solve a set of much smaller systems, and thus obtain significant speedup.

Outline

- Port Separation and Model Reduction
- Simulation and Update of the Reduced Model
- Min-max Programming based Projection Order Decision
- Experimental Results
- Conclusions

Min-max Programming based Projection Order Decision

- From the structure of the block diagonal projection matrix we can see that each sub-projection matrix allows a different size.
- accordingly for a given overall size, we need to decide the size of each sub-matrix to achieve the best accuracy.
- The problem can be cast as

$$\min_{q_1,\ldots q_{p+a}} \max_{s_1^\circ,\ldots s_a^\circ} f(q_1,\ldots q_{p+a}; s_1^\circ,\ldots s_a^\circ)$$

$$\text{s.t.} \quad \sum_{i=1}^{p+a} q_i = d$$

$$q_i \in Z^+ \cup \{0\}, \quad 1 \le i \le p+a,$$

$$s_i^\circ \in \omega_i, \quad 1 \le i \le a$$

minimize the error

worst error for all possible symbol values

constraint on the total size

the sizes must be integer

permitted range of the symbol vales

Min-max Programming based Projection Order Decision

- This non-convex mixed-integer min-max programming is difficult to solve, so we propose to iteratively solve two sub-problems, each of which can be solved efficiently

$$\min_{q_1,\dots q_{p+a}} \quad \max_{s_1\circ,\dots s_a\circ} f(q_1,\dots q_{p+a}; s_1\circ,\dots s_a\circ)$$

$$s.t. \quad \sum_{i=1}^{p+a} q_i = d$$

$$q_i \in Z^+ \cup \{0\}, \quad 1 \le i \le p+a,$$

$$s_i\circ \in \omega_i, \quad 1 \le i \le a$$

$$(\mathbf{P1}:) \min_{q_1,\dots q_{p+a}} \quad f(q_1,\dots q_{p+a}; s_1\circ,\dots s_a$$

$$s.t. \quad \sum_{i=1}^{p+a} q_i = d$$

$$q_i \in Z^+ \cup \{0\}, \quad 1 \le i \le p+a,$$

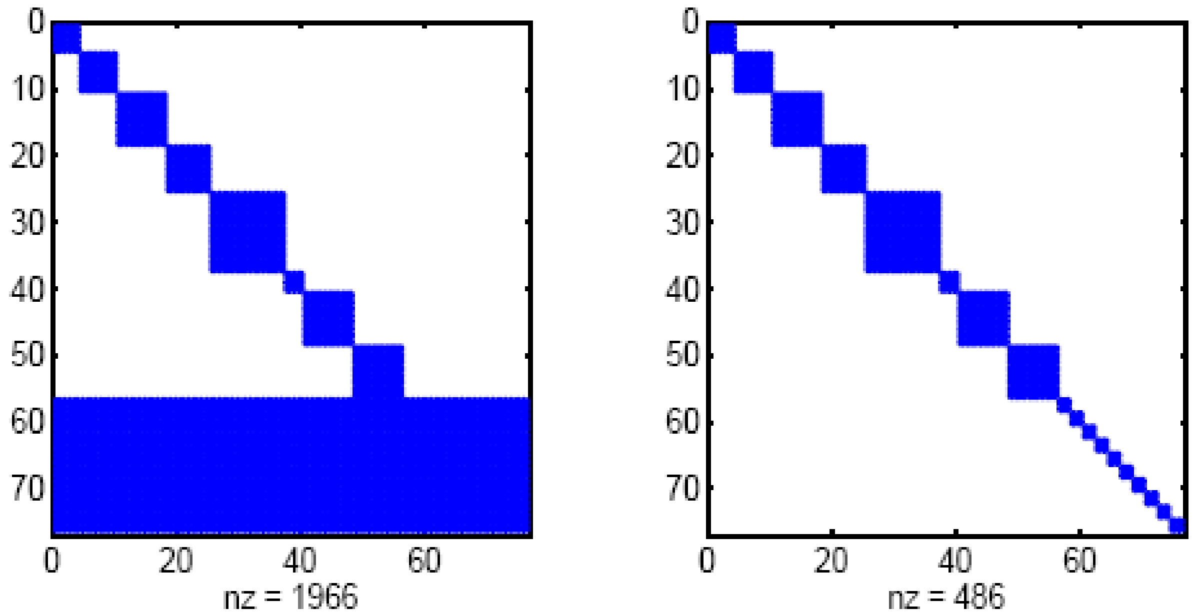
$$(\mathbf{P2}:) \max_{s_1\circ,\dots,s_a\circ} \quad f(q_1,\dots,q_{p+a}; s_1\circ,\dots s_a\circ)$$

$$s.t. \quad s_i\circ \in \omega_i,$$

Outline

- Port Separation and Model Reduction
- Simulation and Update of the Reduced Model
- Min-max Programming based Projection Order Decision
- Experimental Results
- Conclusions

Sparsity of the reduced model

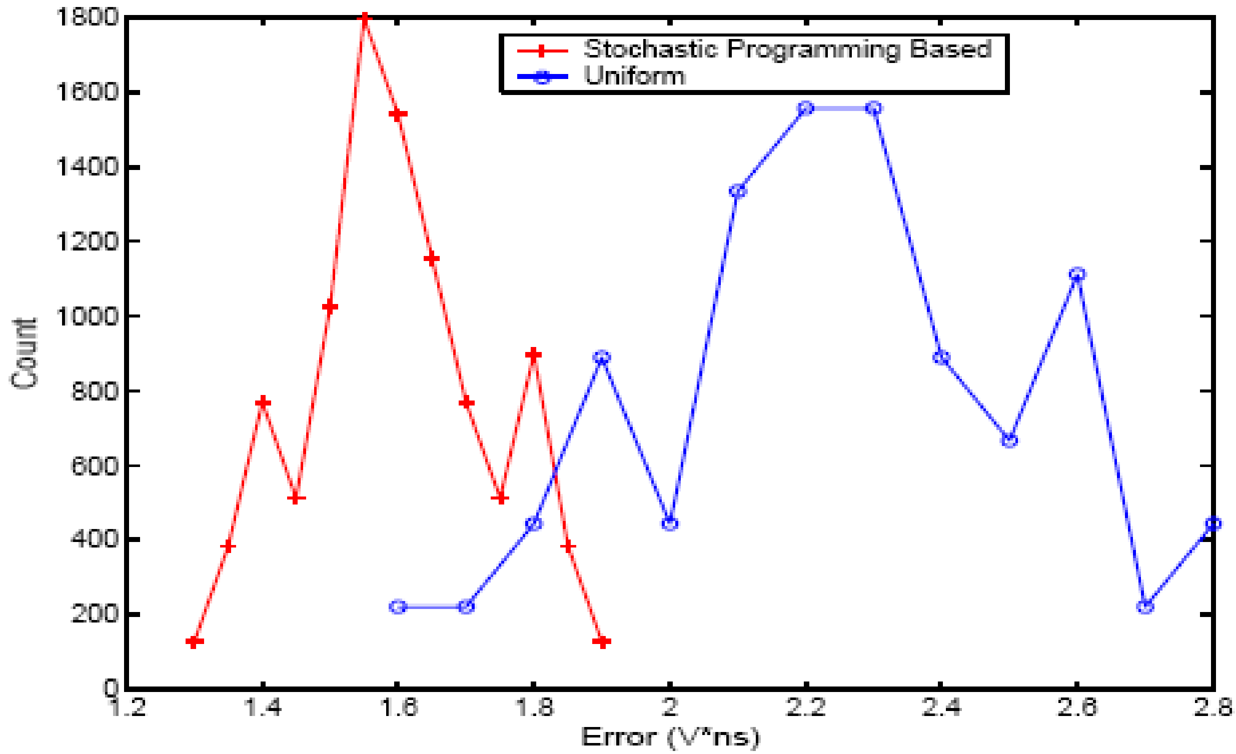


(a) Gr

(b) Cr

- Reduction for a low-noise amplier (LNA) design with parasitics, which contains 4920 nodes, 8 ports 10 symbols. The circuit is reduced to order 76 by the S2NMOR method

Effectiveness of the Stochastic Programming Based Projection Order Decision



- Accuracy comparison between uniform projection order and our stochastic programming based approach based on the 10k Monte Carlo simulation on the symbol values. Our method reduces the mean error by 30% and 3-sigma error by 50%.

Accuracy comparison

ckt name	node #	port #	symbol #
LNA1	1392	14	6
LNA2	5573	33	14
LNA3	11380	79	137
LNA4	49965	147	661

ckt name	var	S.I.	N.P.	F.E	S^2MOR
LNA1	10%	1.2	0.7	**0.9**	0.6 (-24%)
	30%	**1.2**	8.4	6.8	0.6 (-50%)
LNA2	10%	3.7	**1.1**	1.9	0.8 (-27%)
	30%	**3.6**	11.6	17.8	0.8 (-78%)
LNA3	10%	4.2	**3.7**	3.9	0.9 (-76%)
	30%	**4.2**	13.7	19.8	1.0 (-76%)
LNA4	10%	**5.2**	6.7	N.A.	1.6 (-69%)
	30%	**5.2**	28.4	N.A.	1.6 (-69%)

- Accuracy comparison between symbol isolation (S.I.), nominal projection (N.P.), first order expansion (F.E) and the S^2MOR with different variation amount (var) of symbol values. All the errors are in the unit of V*ns.

Runtime Comparison

ckt name	method	size	reduce	factor	update
LNA1	S.I.	300	427	43.7	13.6
	N.P.	300	421	43.7	43.6
	F.E.	300	374	43.7	43.7
	S^2MOR	930	484	7.6	1.5
LNA2	S.I.	420	835	86.4	38.4
	N.P.	420	816	86.5	86.9
	F.E.	420	741	86.4	86.5
	S^2MOR	1340	975	11.3	2.2
LNA3	S.I.	480	1124	91.5	47.6
	N.P.	480	1190	91.4	91.2
	F.E.	480	1011	91.6	91.5
	S^2MOR	1440	1238	12.3	2.9
LNA4	S.I.	500	2977	123.6	61.2
	N.P.	500	2918	123.6	123.5
	F.E.	500	2715	123.6	123.6
	S^2MOR	1610	3020	13.1	3.6

- Runtime comparison between symbol isolation (S.I.), nominal projection (N.P.), first order expansion (F.E.) and the S^2MOR method. The reduced sizes are also reported (size). All units are in seconds. The factorization and simulation time for the S^2MOR model from is up to 17X faster.

Major Contribution of our work

- This paper presents a scalable SMOR algorithm, namely S^2MOR.
  - We first separate the original multi-port multi-symbol system into a set of single-port systems by superposition theorem, and then integrate them together to form a lower-bordered block diagonal (LBBD) structured system.
  - Each block is reduced independently, with a stochastic programming to distribute the given overall model order between blocks for best accuracy. The entire system is efficiently solved by low-rank update.
  - Compared with existing SMOR algorithms, given the same memory space, S^2MOR improves accuracy by up to 78% at a similar reduction time. In addition, the factorization and simulation of the reduced model by S^2MOR is up to 17X faster.

Thank you!