# Large-Scale Differential-Algebraic Equations Arising in VLSI Interconnect Analysis

Roland W. Freund

Department of Mathematics University of California, Davis, USA http://www.math.ucdavis.edu/~freund/

Supported in part by NSF

## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points
- Open problems

## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points
- Open problems

## **VLSI** circuit simulation

- Circuit simulation uses computational methods to simulate and analyze the behavior of electronic circuits
- A circuit can be viewed as a network of electronic devices: transistors, resistors, capacitors, inductors, ...
- Today's VLSI circuits can have  $\mathcal{O}(10^9)$  transistors

#### Are we really just solving DAEs?

- Network topology is described by a graph:
  - Kirchhoff's current laws
  - Kirchhoff's voltage laws
- Equations that characterize the circuit devices:

$$f(i,v) = 0, \quad g\left(i, \frac{d}{dt}v\right) = 0, \quad \dots$$

• All these equations can be summarized as a system of DAEs:

$$\mathbf{F}\left(\mathbf{x}, \frac{d}{dt}\mathbf{x}, t\right) = \mathbf{0}$$

#### The catch

- $\mathbf{F}$ :  $\mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R} \mapsto \mathbb{R}^N$  where N is of the order of the number of devices in the circuit
- For a state-of-the-art circuit:  $N = \mathcal{O}(10^9)$
- No way!
- We are always using today's computers to design tomorrow's largest and more complicated machines

## The VLSI circuit design process



## A small piece of a chip cross section



## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points
- Open problems

#### State-of-the-art VLSI circuits

- 45 nm feature size
- O(10<sup>9</sup>) transistors
- O(10) km of 'wires' (the interconnect)
- Up to 15 layers



## **VLSI** interconnect parasitics

- Wires are not ideal:
  - Resistance
  - Capacitance
  - Inductance
- Consequences:
  - Timing behavior
  - Noise
  - Energy consumption
  - Power distribution



## **Interconnect parasitics extraction**



• Replace 'pieces' of the interconnect by RCL networks:



# Need for model order reduction



## **RCL** networks as directed graphs



• Network topology  $\iff$  Graph incidence matrix  $\mathcal{A}$ 

## **RCL** network equations

- Kirchhoff's current laws:  $\mathcal{A} i_{\mathcal{E}} = 0$
- Kirchhoff's voltage laws:  $\mathcal{A}^{\mathsf{T}}v = v_{\mathcal{E}}$
- Equations for R's, C's, and L's:



#### **RCL** networks as descriptor systems

• System of linear time-invariant DAEs of the form

$$C\frac{d}{dt}x(t) + Gx(t) = Bu(t)$$
$$y(t) = B^{T}x(t)$$
where C,  $G \in \mathbb{R}^{N \times N}$  and  $B \in \mathbb{R}^{N \times m}$ 

- $\mathbf{x}(t) \in \mathbb{R}^N$  is the unknown vector of state variables
- Large state-space dimension N
- *m* inputs, *m* outputs

#### **Reduced-order models**

• System of DAEs of the same form:

$$\mathbf{C}_n \frac{d}{dt} \mathbf{z}(t) + \mathbf{G}_n \mathbf{z}(t) = \mathbf{B}_n \mathbf{u}(t)$$
$$\tilde{\mathbf{y}}(t) = \mathbf{B}_n^{\mathsf{T}} \mathbf{z}(t)$$

But now:

 $\mathbf{C}_n, \ \mathbf{G}_n \in \mathbb{R}^{n \times n}$  and  $\mathbf{B}_n \in \mathbb{R}^{n \times m}$ 

where  $n \ll N$ 

#### **Transfer functions**

• Original descriptor system:

$$\mathbf{H}(s) = \mathbf{B}^{\mathsf{T}} \left( s \, \mathbf{C} + \mathbf{G} \right)^{-1} \mathbf{B}$$

• Reduced-order model:

$$\mathbf{H}_n(s) = \mathbf{B}_n^{\mathsf{T}} \left( s \, \mathbf{C}_n + \mathbf{G}_n \right)^{-1} \mathbf{B}_n$$

• 'Good' reduced-order model

 $\iff$  'Good' approximation  $\mathbf{H}_n \approx \mathbf{H}$ 

#### • Original dimension $N \approx 10^{4-6}$



• Reduced dimension  $n \ll N$   $(n \approx 10^{0-2})$ 

$$\mathbf{H}_{n}(s) = \mathbf{B}_{n}^{\mathsf{T}} \left( s \mathbf{C}_{n} + \mathbf{G}_{n} \right)^{-1} \mathbf{B}_{n}$$

#### **Moment matching**

• Choose a suitable expansion point  $s_0 \in \mathbb{C}$  and expand H(s) about that point

$$H(s) = M_0 + M_1 (s - s_0) + \dots + M_i (s - s_0)^i + \dots$$

• Determine reduced-order transfer function  $H_n(s)$  such that  $H_n(s) = M_0 + M_1 (s - s_0) + \dots + M_{q-1} (s - s_0)^{q-1}$   $+ \widetilde{M}_q (s - s_0)^q + \widetilde{M}_{q+1} (s - s_0)^{q+1} + \dots$   $= H(s) + \mathcal{O} ((s - s_0)^q)$ for some q = q(n)

#### Padé and Padé-type approximation

• Padé approximation:  $C_n, G_n \in \mathbb{R}^{n \times n}, B_n \in \mathbb{R}^{n \times m}$  such that  $H_n(s) = H(s) + \mathcal{O}\left((s - s_0)^{q(n)}\right)$ 

and q(n) is maximal

• 
$$q(n) \ge 2 \left\lfloor \frac{n}{m} \right\rfloor$$
 with equality in the 'generic' case

Padé-type approximation.

$$\mathbf{H}_n(s) = \mathbf{H}(s) + \mathcal{O}\left((s - s_0)^{\tilde{q}(n)}\right)$$

where  $\tilde{q}(n)$  is no longer maximal, e.g.,  $\tilde{q}(n) = \left\lfloor \frac{n}{m} \right\rfloor$ 

## Some history

- AVVE (Pillage and Rohrer, '90): Explicit computation and matching of moments
- PVL, MPVL (Feldmann and F., '94 and '95): Avoids numerical issues of AWE by computing Padé reducedorder models via the Lanczos process
- Arnoldi-based reduction (Silveira et al, '96):
  Padé-type reduced-order models via the Arnoldi process

## An RCL network



Exact and Padé model

# Padé may produce unstable poles



## Some more history

- PRIMA (Odabasioglu, Celik, and Pileggi, '97): Passive reduced-order models via explicit projection onto Krylov subspaces
- SPRIM (F., '04, '09, and '11) Structure-Preserving Reduced Interconnect Macromodeling

#### **Projection-based order reduction**

• Choose an  $N \times n$  matrix

with Rank  $V_n = n$ 

and explicitly project the data matrices of

 $\mathbf{V}_n =$ 

 $C \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t)$  $\mathbf{y}(t) = \mathbf{B}^{\mathsf{T}} \mathbf{x}(t)$ 

onto the subspace spanned by the columns of  $\mathbf{V}_n$ 

#### **Projection-based order reduction**

• Resulting reduced-order model:

$$C_n \frac{d}{dt} \mathbf{z}(t) + G_n \mathbf{z}(t) = \mathbf{B}_n \mathbf{u}(t)$$
$$\tilde{\mathbf{y}}(t) = \mathbf{B}_n^{\mathsf{T}} \mathbf{z}(t)$$

where

$$\mathbf{C}_n := \mathbf{V}_n^{\mathsf{T}} \mathbf{C} \mathbf{V}_n, \quad \mathbf{G}_n := \mathbf{V}_n^{\mathsf{T}} \mathbf{G} \mathbf{V}_n, \quad \mathbf{B}_n := \mathbf{V}_n^{\mathsf{T}} \mathbf{B}$$

• Preserves passivity:

 $\mathbf{C} \succeq \mathbf{0}, \quad \mathbf{G} + \mathbf{G}^{\mathsf{T}} \succeq \mathbf{0} \quad \Rightarrow \quad \mathbf{C}_n \succeq \mathbf{0}, \quad \mathbf{G}_n + \mathbf{G}_n^{\mathsf{T}} \succeq \mathbf{0}$ 

#### Choice of projection matrix

• Choose expansion point  $s_0 \in \mathbb{C}$  for transfer function and rewrite:  $H(s) = B^{\mathsf{T}} (s \operatorname{C} + \operatorname{G})^{-1} B = B^{\mathsf{T}} (I - (s - s_0) \operatorname{A})^{-1} R$ where

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

•  $\hat{n}$ -th block Krylov subspace:

$$\mathcal{K}_{\widehat{n}}(\mathbf{A},\mathbf{R}) := \operatorname{colspan}_{\widehat{n}} \begin{bmatrix} \mathbf{R} & \mathbf{A}\mathbf{R} & \mathbf{A}^{2}\mathbf{R} & \cdots \end{bmatrix}$$

# Krylov + Projection = Padé-type

•  $\hat{n}$ -th block Krylov subspace:

$$\mathcal{K}_{\hat{n}}(\mathbf{A},\mathbf{R}) := \operatorname{colspan}_{\hat{n}} | \mathbf{R} | \mathbf{A}\mathbf{R} | \mathbf{A}^{2}\mathbf{R} | \cdots$$

• Choose the projection matrix  $\mathbf{V}_n$  such that  $\mathcal{K}_{\widehat{n}}(\mathbf{A},\mathbf{R})\subseteq \mathsf{Range}\,\mathbf{V}_n$ 

• Krylov subspace + Projection = Padé-type approximant:  $H_n(s) = H(s) + O\left((s - s_0)^{\tilde{q}}\right), \text{ where } \tilde{q} \ge \lfloor \hat{n}/m \rfloor$ 

PRIMA and SPRIM are methods of this type

## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points
- Open problems

# **Preservation of RCL structure**



## **General RCL network equations**

• System of linear time-invariant DAEs of the form

$$C \frac{d}{dt} \mathbf{x}(t) + \mathbf{G} \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{B}^{\mathsf{T}} \mathbf{x}(t)$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 & \mathbf{G}_3 \\ -\mathbf{G}_2^{\mathsf{T}} & \mathbf{0} & \mathbf{0} \\ -\mathbf{G}_3^{\mathsf{T}} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}$$

• Passivity:

 $\mathbf{C} \succeq \mathbf{0}$  and  $\mathbf{G} + \mathbf{G}^{\top} \succeq \mathbf{0}$ 

#### **PRIMA** does not preserve structure

• PRIMA = projection onto n-th block Krylov subspace:

Range  $\mathbf{V}_n = \mathcal{K}_n(\mathbf{A}, \mathbf{R})$ 

• Block structure of the data matrices:

$$C = \begin{bmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & G_2 & G_3 \\ -G_2 & 0 & 0 \\ -G_3 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 & 0 \\ 0 & 0 \\ 0 & B_2 \end{bmatrix}$$

• PRIMA reduced-order matrices:

$$\mathbf{C}_n = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}, \quad \mathbf{G}_n = \begin{bmatrix} & & \\ & & \\ & & \\ & & \end{bmatrix}, \quad \mathbf{B}_n = \begin{bmatrix} & & \\ &$$

#### **SPRIM** does preserve block structure

• Structure of SPRIM reduced-order matrices:

$$\mathbf{C}_{n} = \begin{bmatrix} \tilde{\mathbf{C}}_{1} & 0 & 0 \\ 0 & \tilde{\mathbf{C}}_{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \mathbf{G}_{n} = \begin{bmatrix} \tilde{\mathbf{G}}_{1} & \tilde{\mathbf{G}}_{2} & \tilde{\mathbf{G}}_{3} \\ -\tilde{\mathbf{G}}_{2}^{\mathsf{T}} & 0 & 0 \\ -\tilde{\mathbf{G}}_{3}^{\mathsf{T}} & 0 & 0 \end{bmatrix}, \ \mathbf{B}_{n} = \begin{bmatrix} \tilde{\mathbf{B}}_{1} & 0 \\ 0 & 0 \\ 0 & \tilde{\mathbf{B}}_{2} \end{bmatrix}$$

 Projection onto Krylov subspaces guarantees a Padé-type property:

$$\mathbf{H}_n(s) = \mathbf{H}(s) + \mathcal{O}\left((s - s_0)^{\tilde{q}}\right)$$

with  $\tilde{q}$  the same integer as for PRIMA

• For SPRIM, we actually observe higher accuracy

# An RCL network with mostly C's and L's



Exact and models corresponding to block Krylov subspace of dimension  $\hat{n} = 120$ 

## **SPRIM**

• Let  $\widehat{\mathbf{V}}_{\widehat{n}}$  be any matrix such that Range  $\widehat{\mathbf{V}}_{\widehat{n}} = \mathcal{K}_{\widehat{n}}(\mathbf{A},\mathbf{R})$ 

• Recall:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 & \mathbf{G}_3 \\ -\mathbf{G}_2^\mathsf{T} & \mathbf{0} & \mathbf{0} \\ -\mathbf{G}_3^\mathsf{T} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}$$

#### **SPRIM**, continued

• Partition  $\hat{\mathbf{V}}_{\hat{n}}$  accordingly:

$$\widehat{\mathrm{V}}_{\widehat{n}} = egin{bmatrix} \mathrm{V}_{\widehat{n}}^{(1)} \ \mathrm{V}_{\widehat{n}}^{(2)} \ \mathrm{V}_{\widehat{n}}^{(3)} \ \mathrm{V}_{\widehat{n}}^{(3)} \end{bmatrix}$$

• For l = 1, 2, 3: If Rank  $V_{\hat{n}}^{(i)} < \hat{n}$ , replace  $V_{\hat{n}}^{(i)}$  by matrix of full column rank

#### **SPRIM**, continued

• Set

$$\mathrm{V}_n = egin{bmatrix} \mathrm{V}_n^{(1)} & 0 & 0 \ 0 & \mathrm{V}_n^{(2)} & 0 \ 0 & 0 & \mathrm{V}_n^{(3)} \end{bmatrix}$$

• Block structure is preserved:

$$\mathbf{C}_{n} = \begin{bmatrix} \tilde{\mathbf{C}}_{1} & 0 & 0\\ 0 & \tilde{\mathbf{C}}_{2} & 0\\ 0 & 0 & 0 \end{bmatrix}, \ \mathbf{G}_{n} = \begin{bmatrix} \tilde{\mathbf{G}}_{1} & \tilde{\mathbf{G}}_{2} & \tilde{\mathbf{G}}_{3}\\ -\tilde{\mathbf{G}}_{2}^{\mathsf{T}} & 0 & 0\\ -\tilde{\mathbf{G}}_{3}^{\mathsf{T}} & 0 & 0 \end{bmatrix}, \ \mathbf{B}_{n} = \begin{bmatrix} \tilde{\mathbf{B}}_{1} & 0\\ 0 & 0\\ 0 & \tilde{\mathbf{B}}_{2} \end{bmatrix}$$

•  $\mathcal{K}_{\widehat{n}}(\mathbf{A}, \mathbf{R}) = \operatorname{Range} \mathbf{V}_{\widehat{n}} \subseteq \operatorname{Range} \mathbf{V}_{n} \Rightarrow$  Padé-type property!

# An RCL network with mostly C's and L's



Exact and models corresponding to  $\hat{n}=90$ 

# An RCL network with mostly C's and L's



Exact and models corresponding to  $\hat{n} = 90$ 

#### A package example



Exact and models corresponding to  $\hat{n} = 128$ 

# A package example



Exact and models corresponding to  $\hat{n} = 128$ 

#### Padé-type property of SPRIM

• General theory of projection onto block Krylov subspaces: PRIMA and SPRIM produce Padé-type models with

 $\mathrm{H}_n(s) = \mathrm{H}(s) + \mathcal{O}\left((s - s_0)^{\tilde{q}}\right), \quad \text{where} \quad \tilde{q} \ge \lfloor \hat{n}/m \rfloor$ 

• Theorem (F., '08)

The n-th SPRIM model satisfies

 $\mathrm{H}_n(s) = \mathrm{H}(s) + \mathcal{O}\left((s - s_0)^{\widetilde{q}}\right), \quad \text{where} \quad \widetilde{q} \ge 2\lfloor \widehat{n}/m \rfloor$ 

- Twice as accurate as PRIMA!
- This is a consequence of structure preservation!

## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points (with Efrem B. Rensi)
- Open problems

#### **Need for restarts**

- To obtain a Padé-type property, projection matrix  $\mathbf{V}_n$  with  $\mathcal{K}_{\widehat{n}}(\mathbf{A},\mathbf{R})\subseteq \mathsf{Range}\,\mathbf{V}_n$
- Need to first generate  $\widehat{V}_{\widehat{n}}$  such that  $\mathcal{K}_{\widehat{n}}(\mathbf{A}, \mathbf{R}) = \operatorname{Range} \widehat{V}_{\widehat{n}}$
- Use suitable (band) variant of the Arnoldi process
- But: prohibitive for large  $\widehat{n}$
- Remedy: (thick) restarts

#### **Using restarts**

- Motivated by recent work by Eiermann et al.
- Restart after each cycle of *r* Arnoldi steps
- Extract 'good' eigenvector information Y from the last batch of r Arnoldi vectors
- Use the columns of Y as the first vectors in the next cycle
- Repeat
- Project with

$$\mathbf{V}_n = \begin{bmatrix} \mathbf{V}^{(1)} & \mathbf{V}^{(2)} & \cdots & \mathbf{V}^{(l)} \end{bmatrix}$$

#### 'Good' eigenvector information

• Recall:

$$\mathbf{H}(s) = \mathbf{B}^{\mathsf{T}} \Big( \mathbf{I} - (s - s_0) \mathbf{A} \Big)^{-1} \mathbf{R}$$

Poles of H are of the form

$$s = s_0 + \frac{1}{\lambda}, \quad \lambda \in \sigma(\mathbf{A})$$

 'Good' eigenvector information:
 Good approximate eigenvectors corresponding to poles close to the frequency range of interest

#### Without restarts



Uniform convergence throughout frequency range

# **Obtaining 'good' eigenvector information**



Single point  $s_0 = 1 \times 10^8$ 

## **Obtaining 'good' eigenvector information**



Single point  $s_0 = 250 \times 10^8$ 

## **Obtaining 'good' eigenvector information**



Single point  $s_0 = (1 + 250 i) \times 10^8$ 

#### Changing expansion points

- Extract 'good' eigenvector information Y from the last batch of r Arnoldi vectors
- At each restart allow for changing expansion point:

 $\mathbf{A}(s_0) = -(s_0 \mathbf{C} + \mathbf{G})^{-1} \mathbf{C} \implies \mathbf{A}(\tilde{s}_0) = -(\tilde{s}_0 \mathbf{C} + \mathbf{G})^{-1} \mathbf{C}$ 

'Converged' eigenvectors v do not change:

 $\lambda$ 

where

$$A(s_0) \mathbf{v} = \lambda \mathbf{v} \quad \Longleftrightarrow \quad A(\tilde{s}_0) \mathbf{v} = \tilde{\lambda} \mathbf{v}$$
$$\frac{1}{2} - \frac{1}{2} = \tilde{s}_0 - s_0$$

#### Multiple expansion points

• Due to changing expansion points

 $s_0^{(1)}, s_0^{(2)}, \ldots, s_0^{(l)},$ 

the resulting reduced-order model is characterized by a multi-point Padé-type property:

$$H_n(s) = H(s) + O\left(\left(s - s_0^{(j)}\right)^{q_j}\right), \quad j = 1, 2, ..., l$$

• Except for  $s_0^{(1)}$ , the other expansion points are complex

## Single vs. multiple expansion points



Single point — no restarts Three points — thick restarts

## Single vs. multiple expansion points



Single point — no restarts Three points — thick restarts

n = 80

n = 42

## Outline

- Numerical simulation of electronic circuits
- DAEs arising in VLSI interconnect analysis
- Structure-preserving model order reduction
- Thick restarts and multiple expansion points
- Open problems

## **Open problems**

- SPRIM preserves the block structures of RCL networks
- Preservation of the fine structure of the blocks?
- Optimal structure-preserving Padé-type reduction?
- Automated selection of changing expansion points to make thick restarts practical?
- We still cannot handle RCL descriptor systems as large as we would need to
- Meaningful reduced-order models for very inaccurate system data?