Model Order Reduction of Coupled Circuit-Device Systems

Michael Hinze, Martin Kunkel, Andreas Steinbrecher, Tatjana Stykel
Model Order Reduction of Coupled Circuit-Device Systems

Michael Hinze∗, Martin Kunkel†, Andreas Steinbrecher‡, Tatjana Stykel‡

May 5, 2011

Abstract

We consider model order reduction of integrated circuits with semiconductor devices. Such circuits are modeled using modified nodal analysis by differential-algebraic equations coupled with the nonlinear drift-diffusion equations. A spatial discretization of these equations with a mixed finite element method yields a high dimensional nonlinear system of differential-algebraic equations. Balancing-related model reduction is used to reduce the dimension of the decoupled linear network equations, while the semidiscretized semiconductor model is reduced using proper orthogonal decomposition. Since the computational complexity of the reduced-order model through the nonlinearity of the drift-diffusion equations still depends on the number of variables of the full model, we apply the discrete empirical interpolation method to further reduce the computational complexity. We provide numerical comparisons which demonstrate the performance of the presented model reduction approach.

1 Introduction

Computer simulations play an significant role in design and production of very large integrated circuits or chips that have nowadays hundreds of millions of semiconductor devices placed on several layers and interconnected by wires. Caused by the decreasing physical size and increasing packing density and operating frequency, such devices cannot be modeled by lumped equivalent circuits any more. Therefore, the need for new models reflecting the complex continuous processes in semiconductors in more details is growing. An approach for modeling the semiconductor devices in the circuit relies on the drift-diffusion equations coupled to the network equations [1, 2]. A spatial discretization of the drift-diffusion equations leads to systems of very large state space dimension that makes the analysis and simulations unacceptably time consuming and expensive. In this context, model order reduction is of great importance. A general idea of model reduction is to approximate the large-scale system by a much smaller model that captures the input-output behavior of the original system to a required accuracy and also preserves essential physical properties. For circuit equations, passivity is the most important property to be preserved in the reduced-order model.

For linear dynamical systems, many different model reduction approaches have been developed over the last thirty years, see [3, 4] for recent collection books on this topic. Krylov subspace based methods such as PRIMA [5] and SPRIM [6, 7] are the most used passivity-preserving model reduction techniques in circuit simulation. A drawback of these methods is the ad hoc choice of interpolation points that strongly influence the approximation quality. Recently, an optimal point selection strategy based on tangential interpolation has been proposed in [8, 9] that provides an optimal H2-approximation.

An alternative approach for model reduction of linear systems is balanced truncation. In order to capture specific system properties, different balancing techniques have been developed for standard and generalized state space systems, see, e.g., [10, 11, 12, 13, 14]. In particular, passivity-preserving balanced truncation methods for electrical circuits (PABTEC) have been proposed in [15, 16, 17] that heavily exploit the topological structure of circuit equations. These methods are based on
balancing the solution of projected Lyapunov or Riccati equations and provide computable error bounds.

Model reduction of nonlinear equation systems may be performed by a trajectory piece-wise linear approach [18] based on linearization, or proper orthogonal decomposition (POD) (see, e.g., [19]), which relies on snapshot calculations and is successfully applied in many different engineering fields including computational fluid dynamics and electronics [19, 20, 21, 22, 23]. A connection of POD to balanced truncation was established in [24, 25].

A POD-based model reduction approach for the nonlinear drift-diffusion equations has been presented in [26], and then extended in [20] to parameterized electrical networks using the greedy sampling proposed in [27]. An advantage of the POD approach is its high accuracy with only few model parameters. However, for its application to the drift-diffusion equations it was observed that the reduction of the problem dimension not necessarily implies the reduction of the simulation time. Therefore, several adaption techniques such as missing point estimation [28] and discrete empirical interpolation method (DEIM) [29] have been developed to reduce the simulation cost for the reduced-order model.

In this paper, we consider model order reduction of coupled circuit-device systems that consist of the differential-algebraic equations modeling a circuit and the nonlinear drift-diffusion equations describing the semiconductor devices. Our approach is based on a combination of the PABTEC algorithm for the decoupled linear network equations and the DEIM-adopted POD method for the distributed device equations.

2 Model equations

In this section, we briefly introduce model equations for integrated circuits with semiconductor devices. For more details on network analysis and semiconductor physics, we refer to [2, 30, 31, 32, 33].

A general circuit can be represented as a directed graph with \( n_a + 1 \) nodes and \( n_b \) branches. The nodes of the graph correspond to the nodes of the circuit, while the branches of the graph correspond to the circuit elements like resistors, capacitors, inductors, voltage sources and current sources. The dynamical behavior of the circuit can then be described using modified nodal analysis (MNA), e.g., [34], by a nonlinear system of differential-algebraic equations (DAEs)

\[
\begin{align}
A_C \frac{d}{dt} q_C(A_C^T \eta) + A_R g(A_R^T \eta) + A_L \eta & = 0, \quad (1a) \\
\frac{d}{dt} \phi(\eta_L) - A_L^T \eta & = 0, \quad (1b) \\
A_V^T \eta & = 0, \quad (1c)
\end{align}
\]

where \( \eta \) denotes the vector of node potentials, \( \eta_L \), \( \eta_V \) and \( \eta_J \) are currents of inductive, voltage source and current source branches, respectively, while \( \eta_V \) and \( \eta_J \) are voltages of voltage sources and current sources, respectively. Furthermore, \( A_C \in \mathbb{R}^{n_a \times n_C}, A_L \in \mathbb{R}^{n_a \times n_L}, A_R \in \mathbb{R}^{n_b \times n_R}, A_V \in \mathbb{R}^{n_V \times n_V} \) and \( A_J \in \mathbb{R}^{n_R \times n_J} \) are the (reduced) incidence matrices describing the topology of the corresponding circuit elements, and the functions \( q_C : \mathbb{R}^{n_C} \to \mathbb{R}^{n_C}, g : \mathbb{R}^{n_R} \to \mathbb{R}^{n_R} \) and \( \phi : \mathbb{R}^{n_L} \to \mathbb{R}^{n_L} \) describe capacitor charges, resistor conductivities and electromagnetic fluxes in the inductors, respectively.

We will assume that

(A1) the matrix \( A_{V'} \) has full column rank,

(A2) the matrix \( \begin{bmatrix} A_C & A_L & A_R & A_{V'} \end{bmatrix} \) has full row rank,

(A3) the functions \( q_C, g \) and \( \phi \) are continuously differentiable and their Jacobians

\[
\begin{align*}
\frac{\partial q_C(u_C)}{\partial u_C} & = C(u_C), & \frac{\partial g(u_R)}{\partial u_R} & = G(u_R), & \frac{\partial \phi(\eta_L)}{\partial \eta_L} & = L(\eta_L)
\end{align*}
\]

are positive definite for all admissible \( u_C = A_C^T \eta, u_R = A_R^T \eta \) and \( \eta_L \), respectively.
Assumptions (A1) and (A2) imply that the circuit contains neither loops of voltage sources (V-loops) nor cutsets of current sources (I-cutsets), respectively, while assumption (A3) means that all circuit elements are passive, i.e., they do not generate energy. Using (2), the MNA equations (1) can be written in the compact form

\[
\begin{align*}
E(x) \frac{d}{dt} x &= Ax + f(x) + Bu, \\
y &= B^T x,
\end{align*}
\]

where

\[
x = \begin{bmatrix} \eta \\ \eta_L \end{bmatrix}, \quad u = \begin{bmatrix} \eta \\ u_q \end{bmatrix}, \quad y = \begin{bmatrix} -u_j \\ -i_q \end{bmatrix}
\]

are the state, input and output vectors, respectively, and

\[
E(x) = \begin{bmatrix} A_C C(A_C^T \eta) A_C^T & 0 & 0 \\
0 & \mathcal{L}(i_L) & 0 \\
0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & -A_L & -A_q \\
A_L^T & 0 & 0 \\
A_q^T & 0 & 0 \end{bmatrix},
\]

\[
f(x) = \begin{bmatrix} -A_R g(A_R^T \eta) \\
0 \\
0 \\
0 \end{bmatrix}, \quad B = \begin{bmatrix} -A_j \\
0 \\
0 \\
0 \end{bmatrix}.
\]

In the following, we will distinguish between linear circuit elements like linear resistors, capacitors and inductors, and nonlinear circuit elements like nonlinear capacitors, inductors, diodes and transistors. A circuit element is called linear if the current-voltage relation for this element is linear. Otherwise, the circuit element is called nonlinear. Without loss of generality, we may assume that the circuit elements are ordered such that the incidence matrices can be partitioned as

\[
A_C = \begin{bmatrix} A_C & A_C \end{bmatrix}, \quad A_L = \begin{bmatrix} A_L & A_L \end{bmatrix}, \quad A_R = \begin{bmatrix} A_R & A_R \end{bmatrix},
\]

where the incidence matrices \(A_C, A_L\) and \(A_R\) correspond to the linear circuit components, and \(A_C^T, A_L^T\) and \(A_R^T\) are the incidence matrices for the nonlinear devices. We also assume that the linear and nonlinear elements are not mutually connected, i.e.,

\[
C(A_C^T \eta) = \begin{bmatrix} \tilde{C} & 0 \\
0 & \tilde{C}(A_C^T \eta) \end{bmatrix}, \quad \mathcal{L}(i_L) = \begin{bmatrix} \tilde{L} & 0 \\
0 & \mathcal{L}(i_L) \end{bmatrix}, \quad g(A_R^T \eta) = \begin{bmatrix} \tilde{G} A_R^T \eta \\
\tilde{g}(A_R^T \eta) \end{bmatrix}
\]

where \(\tilde{C} \in \mathbb{R}^{n_C,n_C}, \tilde{L} \in \mathbb{R}^{n_C,n_L}\) and \(\tilde{G} \in \mathbb{R}^{n_R,n_R}\) are the capacitance, inductance and conductance matrices for the corresponding linear elements, whereas \(\tilde{C}, \tilde{L} : \mathbb{R}^{n_C} \to \mathbb{R}^{n_C,n_C}\) and \(\tilde{G}, \tilde{g} : \mathbb{R}^{n_R} \to \mathbb{R}^{n_R}\) describe the corresponding nonlinear components, and \(i_L\) is the vector of currents through the nonlinear inductors. If the circuit contains some critical semiconductors that have to be modeled by distributed device equations, then we consider the further partitioning

\[
A_R = \begin{bmatrix} A_N & A_S \end{bmatrix}, \quad \tilde{g}(A_R^T \eta) = \begin{bmatrix} \tilde{g}_N(A_R^T \eta) \\
\tilde{g}_S(A_R^T \eta) \end{bmatrix}
\]

where the subscripts \(N\) and \(S\) stand for other nonlinear resistive elements with simple current-voltage relations and for semiconductors, respectively.

For modeling of such critical semiconductors, we use the nonlinear drift-diffusion equations in
mixed formulation given by

\begin{align}
\varepsilon \div g \psi &= q(n - p - N), \\
q \partial_t n - \div J_n &= -q R(n, p), \\
q \partial_t p + \div J_p &= -q R(n, p), \\
g \psi &= \nabla \psi, \\
J_n &= q U_T \mu_n \nabla n - q \mu_n n g \psi, \\
J_p &= -q U_T \mu_p \nabla p - q \mu_p p g \psi,
\end{align}

(4a) – (4f)

on a space-time domain \( \Omega \times [0, T] \) with \( \Omega \subset \mathbb{R}^d, \, d = 1, 2, 3 \). Here, \( \psi = \psi(\xi, t) \) is the electric potential, \( n = n(\xi, t) \) and \( p = p(\xi, t) \) are the electron and hole densities, \( J_n = J_n(\xi, t) \) and \( J_p = J_p(\xi, t) \) are the current densities of electrons and holes, respectively, \( N = N(\xi) \) is the doping profile, and

\[ R(n, p) = \frac{np - n_0^2}{\tau_p (n + n_0) + \tau_n (p + n_0)} \]

is the Shockley-Read-Hall recombination-generation rate, where \( \tau_p \) and \( \tau_n \) are the electron and hole lifetimes, respectively, and \( n_0 \) is the intrinsic concentration. Furthermore, \( \mu_n \) and \( \mu_p \) are the mobilities of electrons and holes, respectively, \( \varepsilon \) is the dielectric permittivity and \( q \) is the unit charge. The temperature is assumed to be constant which leads to a constant thermal voltage \( U_T \). For a comprehensive overview of the drift-diffusion equations, we refer to [31, 32, 35]. The semiconductor model (4) is coupled to the MNA system (3) through the semiconductor current vector \( \mathbf{i}_S \) with the components

\[ (i_S)_k = \int_{\Gamma_{O,k}} (J_n + J_p - \varepsilon \partial_t g \psi) \cdot \nu \, d\sigma, \]

(5)

where \( \nu \) is the normal vector to the interface \( \Gamma_{O,k} \), and through the boundary conditions for the drift-diffusion equations (4) depending on the node potentials \( \eta \) of the network. On Ohmic contacts \( \Gamma_{O,k} \), they read

\begin{align}
\psi(\xi, t) &= \psi_{ap}(A^T_S \eta(t)) + \psi_{bi}(\xi), \quad (\xi, t) \in \Gamma_{O,k} \times [0, T] \\
n(\xi, t) &= \phi_n(\xi), \quad p(\xi, t) = \phi_p(\xi), \quad (\xi, t) \in \Gamma_{O,k} \times [0, T],
\end{align}

(6a) – (6b)

where \( A_S \) is the semiconductor incidence matrix, \( \psi_{ap} \) denotes the applied potential, and \( \psi_{bi}, \phi_n \) and \( \phi_p \) are given time-independent functions [2]. Furthermore, on isolated boundaries \( \Gamma_I \) the boundary conditions are given by

\[ g \psi \cdot \nu = 0, \quad J_n \cdot \nu = 0, \quad J_p \cdot \nu = 0. \]

(6c)

\[ \begin{array}{c}
\eta_1 \\
\eta_2 \\
C_1 \\
R_1 \\
n_{g'} \\
\eta_3 \\
R_3 \\
\eta_4 \\
\Gamma_{O,1} \\
\Gamma_I \\
\scriptsize{DD equations} \\
\Gamma_{O,2} \\
\Gamma_I \\
\end{array} \]

\[ \begin{array}{c}
\eta_k \\
\eta_{k+1} \\
R_k \\
C_{k-1} \\
C_k \\
\end{array} \]

Figure 1: RC chain with a diode.
In Figure 1, a coupled circuit-device system with one semiconductor diode is shown. In particular, the boundary conditions for the electric potential read \( \psi(\xi, t) = \eta_2(t) + \psi_{bi}(\xi) \) for all \((\xi, t) \in \Gamma_{O,1} \times [0, T]\) and \( \psi(\xi, t) = \eta_3(t) + \psi_{bi}(\xi) \) for all \((\xi, t) \in \Gamma_{O,2} \times [0, T]\). Equations (3), (4) and (5) represent a coupled system of partial differential-algebraic equations. The analytical properties and numerical methods for such a system have been investigated in [1, 20, 36]. Using the finite element method for a spatial discretization of the drift-diffusion model (7), other nonlinear equations can be reduced in a similar way.

3.1 Decoupling

Our goal is now to extract a linear subcircuit from a nonlinear circuit. For this purpose, we use a decoupling procedure from [37] that consists in the replacement of the nonlinear inductors with a decoupling procedure from [37] that consists in the replacement of the nonlinear inductors.

3 Model reduction approach

In this section, we present a model reduction approach for the coupled nonlinear DAE system (3), (7) and (9) based on decoupling this system into linear and nonlinear subsystems. Then the linear part is approximated by a reduced-order linear model of lower dimension using the PABTEC algorithm [15, 17], while the decoupled nonlinear equations are reduced using the POD method as described in [20]. Combining these reduced-order linear and nonlinear models, we obtain a nonlinear reduced-order model that approximates the coupled system (3), (7) and (9), see Figure 2. We now describe this model reduction procedure in more detail. For simplicity, in the model reduction of the nonlinear part, we restrict ourself to the semidiscretized drift-diffusion model (7). Other nonlinear equations can be reduced in a similar way.
and nonlinear capacitors by controlled current sources and controlled voltage sources, respectively. The nonlinear resistors are replaced by an equivalent circuit consisting of two serial linear resistors and one controlled current source connected parallel to one of the resistors. Such replacements introduce additional nodes and state variables, but neither additional CV-loops nor LI-cutsets occur in the decoupled linear subcircuit meaning that its index coincides with the index of the original circuit, see [38] for the index analysis of the circuit equations. An advantage of the suggested replacement strategy is exemplary demonstrated in the following example.

**Example 3.1** Consider again a circuit with a semiconductor diode as in Figure 1. We suggest to replace the diode by an equivalent circuit shown in Figure 3. If we would replace the diode by a current source, then a decoupled linear circuit would have I-cutset and, hence, lack well-posedness. Moreover, if we would replace the diode by a voltage source, then the resulting linear circuit would have CV-loop, i.e., it would be of index two, although the original circuit is of index one. Note that model reduction of index two problems is more involved than that of index one [39].

After the replacements described above, the extracted linear subcircuit can be modeled by the linear DAE system in the MNA form

\[
\begin{align*}
E \frac{4}{dt} x_t &= Ax_t + Bu_t, \\
y_t &= B^T x_t,
\end{align*}
\] (10a) (10b)
Figure 3: Decoupled linear RC chain with a replacement circuit.

with \( x^T = [\eta^T \eta_z^T \eta_z^T \eta_z^T \eta_z^T \eta_z^T \eta_z^T \eta_z^T] \) and

\[
E = \begin{bmatrix}
A_C C A_C^T & 0 & 0 \\
0 & L & 0 \\
0 & 0 & 0
\end{bmatrix},
A = \begin{bmatrix}
-A_R G A_R^T & -A_L & -A_V \\
A_L^T & 0 & 0 \\
A_V^T & 0 & 0
\end{bmatrix},
B = \begin{bmatrix}
-A_I & 0 \\
0 & 0 \\
0 & -I
\end{bmatrix},
\]

where the incidence and element matrices are given by

\[
A_C = \begin{bmatrix}
A_C^T & 0 \\
0 & L
\end{bmatrix},
A_R = \begin{bmatrix}
A_R^1 A_R^2 \\
0 & -I & I
\end{bmatrix},
A_L = \begin{bmatrix}
A_L^T & 0 \\
0 & 0
\end{bmatrix},
\]

(10c)

\[
A_V = \begin{bmatrix}
A_V & A_V \\
0 & 0
\end{bmatrix},
A_I = \begin{bmatrix}
A_I & A_L^T & A_L^T & A_L^T \\
0 & I & 0
\end{bmatrix},
\]

(10d)

\[
G = \begin{bmatrix}
G & 0 & 0 & 0 \\
0 & G_1 & 0 & 0 \\
0 & 0 & G_2 & 0
\end{bmatrix},
C = \tilde{C},
L = \tilde{L}.
\]

(10e)

Here, the matrices \( A_R^1 \) and \( A_R^2 \) have entries in \( \{0,1\} \) and \( \{-1,0\} \), respectively, and satisfy \( A_R^1 + A_R^2 = A_R \). Moreover, \( \eta_z \) is the potential of the introduced nodes, and the matrices \( G_1 \) and \( G_2 \) are diagonal with conductances of the introduced linear resistors in the replacement circuits on the diagonal, and the new input variables \( u_{\tilde{C}} \) and \( \eta_z \) are given by

\[
u_{\tilde{C}} = A_{\tilde{C}}^T \eta,
\]

(11)

\[
\eta_z = (G_1 + G_2) \eta - G_1 A_{\tilde{C}}^T \eta - G_2 A_{\tilde{C}}^T \eta
\]

(12)

One can show that the linear system (10) together with the decoupled nonlinear equations (7), (9) and the equations for the nonlinear inductors

\[
\tilde{L} A^T \tilde{I} - A_{\tilde{I}}^T \eta = 0
\]

is state equivalent to the coupled system (3), (7) and (9) together with the relations

\[
u_{\tilde{C}} = \tilde{C}(u_{\tilde{C}}) A_{\tilde{C}}^T u_{\tilde{C}},
\]

\[
\eta_z = (G_1 + G_2)^{-1} \left( G_1 (A_{\tilde{C}}^T \eta - G_2 A_{\tilde{C}}^T \eta) - \eta_z \right)
\]

in the sense that these both systems have the same state vectors up to a permutation, see [37] for detail.
3.2 Model reduction of the linear subcircuit using the PABTEC method

Once we have the decoupled linear DAE system (10) with \( E, A \in \mathbb{R}^{n_e,n_e} \) and \( B \in \mathbb{R}^{n_e,m_e} \), we can approximate this system by a reduced-order model

\[
\frac{d}{dt} \hat{x}_f = \hat{A} \hat{x}_f + \hat{B} u,
\]

\[
\hat{y}_f = \hat{C} \hat{x}_f,
\]

with \( \hat{E}, \hat{A} \in \mathbb{R}^{r_e,r_e}, \hat{B} \in \mathbb{R}^{r_e,m_e}, \hat{C} \in \mathbb{R}^{m_e,r_e} \) and \( r_e < n_e \). If the matrices \( G, C \) and \( L \) in (10f) are symmetric and positive definite, then system (10) is passive and reciprocal. The latter means that the transfer function of (10) given by \( G(s) = B^T(sE - A)^{-1}B \) satisfies \( G^T(s) = S_{\text{ext}}G(s)S_{\text{ext}} \) with the signature matrix

\[
S_{\text{ext}} = \text{diag}(I_{n_y+n_z+n_x}, -I_{n_y+n_c}),(14)
\]

Of course, these properties should be preserved in the reduced-order model (13). This would allow us to synthesize this model as a circuit with a small number of elements compared to the original circuit [40, 41].

The passive and reciprocal reduced-order model (13) can be computed via the PABTEC method [15] based on balanced truncation. First, we define the controllability and observability Gramians of system (10) as unique stabilizing solutions of the projected Riccati equations

\[
EXE^T + FXE^T + EXB^TB^TX + P_rB^TB^TP_r^T = 0, \quad X = P_rX^r_r, (15)
\]

\[
E^TYF + F^TYE + E^TYB^TB_f^TY + P_r^TB_f^TB_fP_r = 0, \quad Y = P_f^YP_f, (16)
\]

where

\[
F = A - B B^T - 2P_rB(I - M_0^TM_0)^{-1}M_0^TB^TP_r,
\]

\[
B_o = \sqrt{2} B J_o^{-1}, \quad B_c = \sqrt{2} J_c^{-1}B^T,
\]

\[
J_o^TJ_o = I - M_0^TM_0, \quad J_c^TJ_c = I - M_0M_0^T,
\]

\[
M_0 = I - 2 \lim_{s \to \infty} B^T(sE - A + B B^T)^{-1}B,
\]

and \( P_r \) and \( P_f \) are the spectral projectors onto the right and left deflating subspaces of the pencil \( \lambda E - (A - B B^T) \) corresponding to the finite eigenvalues. The balanced truncation approach is based on the transformation of system (10) into a balanced form whose Gramians are both equal to a diagonal matrix. Then the reduced-order model (13) is determined by the truncation of the states corresponding to small diagonal elements of the balanced Gramians. In practice, we do not need to balance system (10) explicitly. Instead, we can use the following algorithm developed in [15].

**Algorithm 3.2 (Passivity-preserving balanced truncation for electrical circuits (PABTEC))**

Given \( E, A, B, B^T \) for the linear model equations (10), compute \( (\hat{E}, \hat{A}, \hat{B}, \hat{C}) \) for a reduced-order model (13).

1. Compute the Cholesky factor \( R_X \) of the stabilizing solution \( X = R_X R_X^T \) of the projected Riccati equation (15).

2. Compute the eigenvalue decomposition

\[
R_X^T S_{\text{int}} E R_X = [U_1, U_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} [U_1, U_2]^T,
\]

where \( S_{\text{int}} = \text{diag}(I_{n_x+n_y}, -I_{n_z}, -I_{n_r+n_c}) \), \( [U_1, U_2] \) is orthogonal, \( \Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_r) \) and \( \Lambda_2 = \text{diag}(\lambda_{r+1}, \ldots, \lambda_{m_c}) \).

3. Compute the eigenvalue decomposition \( (I - M_0)S_{\text{ext}} = U_0 \Lambda_0 U_0^T \), where \( S_{\text{ext}} \) is as in (14), \( U_0 \) is orthogonal and \( \Lambda_0 = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{m_c}) \).
4. Compute the reduced-order model (13) with

\[ \hat{E} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{A} = \frac{1}{2} \begin{bmatrix} 2W^TAV & \sqrt{2}W^TC_{\infty} \\ -\sqrt{2}B_{\infty}B^TV & 2I - B_{\infty}C_{\infty} \end{bmatrix}, \]

\[ \hat{B} = \begin{bmatrix} W^TB \\ -B_{\infty}/\sqrt{2} \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} B^TV, & C_{\infty}/\sqrt{2} \end{bmatrix}, \]

where

\[ B_{\infty} = S_0|A_0|^{1/2}U_0^TS_\text{ext}, \quad C_{\infty} = U_0|A_0|^{1/2}, \]

\[ W = R_XU_1|A_1|^{-1/2}, \quad V = S_\text{int}R_XU_1|A_1|^{-1/2}, \]

\[ S_0 = \text{diag} (\text{sign}(\hat{\lambda}_1), \ldots, \text{sign}(\hat{\lambda}_m)), \quad |A_0| = \text{diag} (|\hat{\lambda}_1|, \ldots, |\hat{\lambda}_m|), \]

\[ S_1 = \text{diag} (\text{sign}(\lambda_1), \ldots, \text{sign}(\lambda_r)), \quad |A_1| = \text{diag} (|\lambda_1|, \ldots, |\lambda_r|). \]

One can show that the reduced-order system (13), (17) is passive and reciprocal, and we have the following a priori \( L_2 \)-norm error bound

\[ \| y_t - y^r \|_2 \leq 2\| I + \hat{G} \|_{H_\infty}(|\lambda_{r+1}| + \ldots + |\lambda_q|)\| u_t \|_2, \]

provided \( 2\| I + \hat{G} \|_{H_\infty}(|\lambda_{r+1}| + \ldots + |\lambda_q|) < 1 \), see [13, 15]. Here, the \( H_\infty \)-norm is defined as

\[ \| I + \hat{G} \|_{H_\infty} = \sup_{s \in \mathbb{R}} \| (I + \hat{G}(s)) \|, \]

where \( \| \cdot \| \) denotes the spectral matrix norm. Furthermore, if we choose \( r \) in the PABTEC algorithm such that \( 2\| I + \hat{G} \|_{H_\infty}(|\lambda_{r+1}| + \ldots + |\lambda_q|) < 1 \), where \( \hat{G}(s) = \hat{C}(s\hat{E} - \hat{A})^{-1}\hat{B} \) is the transfer function of (13), then we obtain the a posteriori error bound

\[ \| y_t - y^r \|_2 \leq 2\| I + \hat{G} \|_{H_\infty}(|\lambda_{r+1}| + \ldots + |\lambda_q|)\| u_t \|_2 \]

that is inexpensive to compute.

Note that the projectors \( P_l, P_r \) and the matrix \( M_0 \) required in Algorithm 3.2 can be constructed in explicit form using the topological structure of the MNA equations (10), see [15, 17]. Moreover, for RC and RL circuits, the PABTEC algorithm can be simplified in such a way that a projected Lyapunov equation has to be solved instead of the projected Riccati equation, that reduces the computational complexity considerably [16].

3.3 Model reduction of the nonlinear semiconductor model using the POD method

For the approximation of the nonlinear semiconductor model (7) by a reduced-order model, we use the POD method [19] combined with the DEIM approach [29] for efficient evaluation of nonlinearities. For this purpose, we first run a simulation of the coupled system system (3), (7) and (9) and collect \( k \) snapshots \( \psi(t_j), n^h(t_j), p^h(t_j), g^h_{\psi}(t_j), J_n^h(t_j), J_p^h(t_j) \) at time instances \( t_j \in [0, T], j = 1, \ldots, k \). Note that already at this stage we can replace the decoupled linear subsystem (10) of (3) with the reduced-order model (13), (17) in order to reduce the simulation time. The snapshot variant of POD introduced in [19] finds a best approximation of the space spanned by the snapshots with respect to the considered scalar product. Since every component \( \psi^h, n^h, p^h, g^h_{\psi}, J_n^h, J_p^h \) of the state vector of (7) has its own physical meaning, we approximate these components separately by the vectors

\[ \psi^{POD}(t) = U_{\psi}\gamma_{\psi}(t), \quad n^{POD}(t) = U_n\gamma_n(t), \quad p^{POD}(t) = U_p\gamma_p(t), \quad g_{\psi}^{POD}(t) = U_{g_{\psi}}\gamma_{g_{\psi}}(t), \quad J_n^{POD}(t) = U_{J_n}\gamma_{J_n}(t), \quad J_p^{POD}(t) = U_{J_p}\gamma_{J_p}(t), \]

respectively, where the projection matrices \( U_* \in \mathbb{R}^{n_s \times s_*} \) with \( * \in \{ \psi, n, p, g_{\psi}, J_n, J_p \} \) contain the (time-independent) POD basis vectors, the functions \( \gamma_* \) are the corresponding time-variant coefficients, and the numbers \( s_* \) denote the respective numbers of the POD basis functions included.

The projection matrices \( U_* \) are determined from the singular value decomposition of the matrices composed of the corresponding snapshots. Let, for example, \( \Psi = \tilde{U}_{\psi}S_{\psi}\tilde{V}_{\psi}^T \) be the singular
The Galerkin projection of system (7) yields the reduced-order model

\[ \begin{bmatrix} 0 \\ -\frac{d}{dt}\gamma_p \\ \frac{d}{dt}\gamma_p \end{bmatrix} = -A_{POD} \begin{bmatrix} \gamma_p \\ \gamma_n \\ \gamma_{J_n} \end{bmatrix} - U^T F(U_n\gamma_n, U_p\gamma_p, U_{g_0}\gamma_{g_0}) + U^T b(A_F^T \eta), \] (19)

where \( A_{POD} = U^T A_{FEM} U \) and \( U = \text{diag}(U_\psi, U_n, U_p, U_{g_0}, U_{J_n}, U_{J_p}) \). The coupling relation (9) can then be approximated by

\[ \dot{i}_S^h = C_1 U_{J_n}\gamma_{J_n} + C_2 U_{J_p}\gamma_{J_p} + C_3 U_{g_0} \frac{d}{dt}\gamma_{g_0}. \] (20)

As for the original system (7) and (9), we denote the relation between \( A_F^T \eta \) and \( \dot{i}_S^h \) by

\[ \dot{i}_S^h = g_S(A_F^T \eta). \] (21)

All matrix-matrix multiplications are calculated in an off-line phase, whereas the nonlinear function \( F \) has to be evaluated on-line. For the reduction of the evaluation time, we use DEIM proposed in [29].

### 3.4 Recoupling

After model reduction of the linear DAE system (10) using the PABTEC method, we obtain the reduced-order model (13), (17). In particular, this model has the form

\[ \dot{E} \frac{d}{dt} \hat{x}_t = \hat{A} \hat{x}_t + \begin{bmatrix} \hat{B}_1 \hat{B}_2 \hat{B}_3 \hat{B}_4 \hat{B}_5 \end{bmatrix} \begin{bmatrix} \dot{v}_j \\ \dot{z}_j \\ \dot{\gamma}_C \\ \dot{u}_C \end{bmatrix}, \quad \begin{bmatrix} \dot{y}_t \\ \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \\ \dot{y}_5 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{bmatrix} \dot{x}_t, \]

where \( \dot{y}_{tj} = \dot{C}_j \dot{x}_t, j = 1, \ldots, 5 \), approximate the corresponding components of the output \( y_t \) in (10b). Combining this system with the unchanged nonlinear circuit equations and the reduced semiconductor model (19), (21) as described in [37], we get the reduced-order nonlinear model

\[ \dot{E} \frac{d}{dt} \hat{x} = \hat{A} \hat{x} + \hat{f}(\hat{x}) + \hat{B}u, \] (22a)

\[ \dot{y} = \hat{C} \hat{x}, \] (22b)
First, we describe the diode by the voltage-current relation \( f \) with the frequency \( \pi f \). Consider an RC circuit with one diode as shown in Figure 1. The input is given by

\[
\begin{align*}
\dot{x} &= E L (\vec{u}_L) u_L, \\
\dot{\vec{u}}_C &= -C \vec{u}_L \\
\dot{u}_R &= -G \vec{u}_L \\
\end{align*}
\]

where \( \dot{x}^T = [\dot{x}_L^T \ \ u_L^T \ \ \vec{u}_L^T \ \ u_R^T] \) and \( u^T = [\dot{x}_L^T \ \ u_L^T] \) and

\[
\dot{\vec{x}}(\hat{x}) = \begin{bmatrix} \hat{E} & 0 & 0 & 0 \\
0 & \hat{\vec{E}} (\vec{u}_L) & 0 & 0 \\
0 & 0 & \hat{\vec{C}} (\vec{u}_L) & 0 \\
0 & 0 & 0 & \hat{u}_R \end{bmatrix}, \quad \hat{f}(\hat{x}) = \begin{bmatrix} 0 \\
0 \\
0 \\
\hat{\vec{g}} (\vec{u}_L) \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} \hat{B}_1 & \hat{B}_4 \\
0 & 0 \\
0 & 0 \\
0 & 0 \end{bmatrix}, \quad (22c)
\]

\[
\hat{A} = \begin{bmatrix} \hat{A} + \hat{B}_2 (G_1 + G_2) \hat{C}_2 & \hat{B}_3 & \hat{B}_5 & \hat{B}_2 G_1 \\
-\hat{C}_3 & 0 & 0 & 0 \\
-\hat{C}_5 & 0 & 0 & 0 \\
-G_1 \hat{C}_2 & 0 & 0 & -G_1 \end{bmatrix}, \quad \hat{\vec{C}} = \begin{bmatrix} \hat{\vec{C}}_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \end{bmatrix}, \quad (22d)
\]

The coupled system (22), (19) and (21) represents then an approximation to the nonlinear DAE system (3), (7) and (9), where both the linear subcircuit as well as the semiconductor model are reduced. Note that both model reduction approaches presented in Sections 3.2 and 3.3 for the decoupled linear subcircuit and nonlinear drift-diffusion equations can be executed independently. In this case, the model equations (22), (7) and (9) with unreduced \( \hat{g} \) instead of \( \tilde{g} \) in (22c) approximate the original coupled nonlinear DAE system (3), (7) and (9), where only the linear subsystem is reduced and the semiconductor model remains unchanged. Finally, the model equations (3), (19) and (21), where the approximate function \( \tilde{g} \) instead of \( \hat{g} \) is used in (3f) represent an approximation to the original nonlinear DAE system (3), (7) and (9), where only the distributed semiconductor model is reduced and the circuit equations remain unchanged.

### 4 Numerical experiments

In this section, we present some results of numerical experiments to demonstrate the applicability of the presented model reduction approaches for coupled circuit-device systems. For model reduction of linear circuit equations, we use the MATLAB Toolbox PABTEC [42]. The POD method is implemented in C++ based on the FEM library deal.II [43] for discretizing the drift-diffusion equations. The obtained large and sparse nonlinear DAE system (3), (7), (9) as well as the small and dense reduced-order model (19), (21), (22) are integrated using the DASPK software package [44] based on a BDF method, where the nonlinear equations are solved using Newton’s method. Furthermore, the direct sparse solver SuperLU [45] is employed for solving linear systems.

Consider an RC circuit with one diode as shown in Figure 1. The input is given by

\[
u(t) = u_q(t) = 10 \sin(2\pi f_0 t)^4
\]

with the frequency \( f_0 = 10^4 \) Hz, see Figure 4. The output of the system is \( y(t) = -v_q(t) \).

We simulate the models over the fixed time horizon \([0, \frac{\tau_0}{\pi}]\). The linear resistors have the same resistance \( R = 2 \, k\Omega \) and the linear capacitors have the same capacitance \( C = 0.02 \, \mu F \).

First, we describe the diode by the voltage-current relation

\[
\tilde{g}(u_R) = 10^{-14} \left( \exp(40 u_R) - 1 \right), \quad (23)
\]

and apply only the PABTEC method to the decoupled linear system (10) that models the linear circuit given in Figure 3. System (10) with \( n_L = 1503 \) variables was approximated by a reduced model (13) of dimension \( n_L = 24 \). This dimension was determined as \( n_L = \text{rank}(I - M_0) \) and \( r \) satisfies the condition \(|\lambda_{r+1}| + \cdots + |\lambda_0| < \text{tol}_{\Pi} \) with a prescribed tolerance \( \text{tol}_{\Pi} = 10^{-7} \). The outputs \( y \) and \( \tilde{y} \) of the original nonlinear system (3) and the reduced-order nonlinear model (22), respectively, are plotted in Figure 4. Simulation time and the absolute and relative \( L_2 \)-norm errors in the output are presented in Table 1. One can see that the simulation time is reduced by a factor of 10, while the relative error is below 2%.
As the next step, we introduce the drift-diffusion model (4) for the diode. The parameters of the diode are summarized in Table 2. Note that we do not expect to obtain the same output $y$ as in the previous experiment. To achieve this, one would need to perform a parameter identification for the drift-diffusion model which is not done in this paper. In Table 3, we collect the numerical results for different model reduction strategies. The outputs of the systems with the reduced network and/or POD-reduced diode are compared to the full semidiscretized model (3), (7) and (9) with 7510 variables. First, we reduce the extracted linear network and do not modify the diode. This reduces the number of variables by about 20%, and the simulation time is reduced by 27%. It should also be noted that the reduced network is not only smaller but it is also easier to integrate for the DAE solver. An indicator for the computational complexity is the number of Jacobian evaluations or, equivalently, the number of LU decompositions required during integration.

Finally, we create a POD-reduced model (19) and (21) for the diode. The number of columns $s_*$ of the projection matrices $U_*$ is determined from the condition $\Delta_* \leq \text{tol}_{\text{POD}}$ with $\Delta_*$ defined in (18) and a tolerance $\text{tol}_{\text{POD}} = 10^{-6}$ for each component. We also apply the DEIM method for the reduction of nonlinearity evaluations in the drift-diffusion model. The resulting reduced-order model (19) for the diode is a dense DAE of dimension 105 while the original model (7) has dimension 6006. Coupling it with the unreduced and reduced linear networks, we obtain the results in Table 3 (last two rows). The simulation results for the different model reduction setups are also illustrated in Figure 5.

## 5 Summary

In the present paper, we developed a framework to combine the PABTEC and simulation-based POD model order reduction techniques to determine reduced-order models for coupled circuit-
Table 2: Diode model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>$1.03545 \cdot 10^{-12}$ F/cm</td>
</tr>
<tr>
<td>$U_T$</td>
<td>0.0259 V</td>
</tr>
<tr>
<td>$n_0$</td>
<td>$1.4 \cdot 10^{10}$ 1/cm$^3$</td>
</tr>
<tr>
<td>$\mu_n$</td>
<td>1350 cm$^2$/V sec</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>$330 \cdot 10^{-9}$ sec</td>
</tr>
<tr>
<td>$\mu_p$</td>
<td>480 cm$^2$/V sec</td>
</tr>
<tr>
<td>$\tau_p$</td>
<td>$33 \cdot 10^{-9}$ sec</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>$[0,l_1] \times [0,l_2] \times [0,l_3]$</td>
</tr>
<tr>
<td>$l_1$ (length)</td>
<td>$10^{-4}$ cm</td>
</tr>
<tr>
<td>$l_2$ (width)</td>
<td>$10^{-5}$ cm</td>
</tr>
<tr>
<td>$l_3$ (depth)</td>
<td>$10^{-5}$ cm</td>
</tr>
<tr>
<td>$N(\xi), \xi_1 &lt; l_1/2$</td>
<td>$-9.94 \cdot 10^{15}$ 1/cm$^3$</td>
</tr>
<tr>
<td>$N(\xi), \xi_1 \geq l_1/2$</td>
<td>$4.06 \cdot 10^{18}$ 1/cm$^3$</td>
</tr>
<tr>
<td>FEM-mesh</td>
<td>500 elements, refined at $\xi_1 = l_1/2$</td>
</tr>
</tbody>
</table>

Table 3: Statistics for model reduction of the coupled circuit-device system.

<table>
<thead>
<tr>
<th>network (MNA equations)</th>
<th>diode (drift-diffusion equations)</th>
<th>dim. simul.</th>
<th>Jacobian absolute error</th>
<th>relative error $|y - \hat{y}|<em>{L_2}/|y|</em>{L_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>unreduced</td>
<td>unreduced</td>
<td>7510</td>
<td>23.37s</td>
<td>20</td>
</tr>
<tr>
<td>reduced</td>
<td>unreduced</td>
<td>6031</td>
<td>16.90s</td>
<td>17</td>
</tr>
<tr>
<td>unreduced</td>
<td>reduced</td>
<td>1609</td>
<td>1.51s</td>
<td>16</td>
</tr>
<tr>
<td>reduced</td>
<td>reduced</td>
<td>130</td>
<td>1.19s</td>
<td>11</td>
</tr>
</tbody>
</table>

device systems. While the PABTEC method preserves the passivity and reciprocity in the reduced linear circuit model, the POD approach delivers high-fidelity reduced-order models for the semiconductor devices. Numerical examples demonstrate that the recoupling of the respective reduced-order models delivers an overall reduced-order model for the circuit-device system which allows significantly faster simulations (speedup-factor is about 20) while keeping the relative errors below 10%.

Finally, we note that the model reduction concept developed in this paper is not restricted to the reduction of electrical networks containing semiconductor devices. It can also be extended to the reduction of networks modeling e.g. nonlinear multibody systems containing many simple mass-spring-damper components and only a few high-fidelity components described by PDE systems.

References


Figure 5: Input voltage and output currents for the four model reduction setups.


