Chapter 1
Model Order Reduction of Integrated Circuits in Electrical Networks

Michael Hinze, Martin Kunkel, Ulrich Matthes, and Morten Vierling

Abstract We consider integrated circuits with semiconductors modeled by modified nodal analysis and drift-diffusion equations. The drift-diffusion equations are discretized in space using mixed finite element method. This discretization yields a high-dimensional differential-algebraic equation. Balancing-related model reduction is used to reduce the dimension of the decoupled linear network equations, while the semidiscretized semiconductor models are reduced using proper orthogonal decomposition. We among other things show that this approach delivers reduced-order models which depend on the location of the semiconductor in the network. Since the computational complexity of the reduced-order models through the nonlinearity of the drift-diffusion equations still depend 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. We compare reduced and fine models and give numerical results for a basic network with one diode. Furthermore we discuss residual based sampling to construct POD models which are valid over certain parameter ranges.

1.1 Introduction

Computer simulations play a 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. Decreasing

M. Hinze (✉) • U. Matthes • M. Vierling
Department of Mathematics, University of Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
e-mail: michael.hinze@uni-hamburg.de; ulrich.matthes@math.uni-hamburg.de; morten.vierling@uni-hamburg.de

M. Kunkel
Fakultät für Luft- und Raumfahrttechnik, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: ich@martinkunkel.de

P. Benner (ed.), *System Reduction for Nanoscale IC Design*,
Mathematics in Industry 20, DOI 10.1007/978-3-319-07236-4_1

physical size, increasing packing density, and increasing operating frequencies necessitate the development of new models reflecting the complex continuous processes in semiconductors and the high-frequency electromagnetic coupling in more detail. Such models include complex coupled partial differential equation (PDE) systems where spatial discretization leads to high-dimensional ordinary differential equation (ODE) or differential-algebraic equation (DAE) systems which require unacceptably high simulation times. In this context model order reduction (MOR) is of great importance. In the present work we as a first step towards model order reduction of complex coupled systems consider electrical circuits with semiconductors modeled by drift-diffusion (DD) equations as proposed in e.g. [46, 52]. Our general idea of model reduction of this system consists in approximating this 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 30 years, see [6, 42] for recent collection books on this topic. Krylov subspace based methods such as PRIMA [32] and SPRIM [15, 16] 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 [3, 20] that provides an optimal $\mathbb{H}_2$-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., [19, 31, 35, 37, 49]. In particular, passivity-preserving balanced truncation methods for electrical circuits (PABTEC) have been proposed in [38, 39, 51] 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 [40] based on linearization, or proper orthogonal decomposition (POD) (see, e.g., [45]), which relies on snapshot calculations and is successfully applied in many different engineering fields including computational fluid dynamics and electronics [23, 29, 45, 48, 53]. A connection of POD to balanced truncation was established in [41, 54].

A POD-based model reduction approach for the nonlinear drift-diffusion equations has been presented in [25], and then extended in [23] to parameterized electrical networks using the greedy sampling proposed in [33]. 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 [4]

and discrete empirical interpolation method (DEIM) [10, 11] have been developed to reduce the simulation cost for the reduced-order model.

In this paper, we review results of [23–27] related to model order reduction of coupled circuit-device systems consisting of the differential-algebraic equations modeling an electrical circuit and the nonlinear drift-diffusion equations describing the semiconductor devices. In a first step we show how proper orthogonal decomposition (POD) can be used to reduce the dimension of the semiconductor models. It among other things turns out, that the reduced model for a semiconductor depends on the position of the semiconductor in the network. We present numerical investigations from [25] for the reduction of a 4-diode rectifier network, which clearly indicate this fact. Furthermore, we apply the Discrete Empirical Interpolation Method (DEIM) of [10] for a further reduction of the nonlinearity, yielding a further reduction of the overall computational complexity. Moreover, we adapt to the present situation the Greedy sampling approach of [33] to construct POD models which are valid over certain parameter ranges. In a next step we combine the passivity-preserving balanced truncation method for electrical circuits (PABTEC) [38, 51] to reduce the dimension of the decoupled linear network equations with POD MOR for the semiconductor model. Finally, we present several numerical examples which demonstrate the performance of our approach.

1.2 Basic Models

In this section we combine mathematical models for electrical networks with mathematical models for semiconductors. Electrical networks can be efficiently modeled by a differential-algebraic equation (DAE) which is obtained from modified nodal analysis (MNA). Denoting by e the node potentials and by j_L and j_V the currents of inductive and voltage source branches, the DAE reads (see [18, 28, 52])

$$A_C \frac{d}{dt} q_C(A_C^\top e, t) + A_R g(A_R^\top e, t) + A_L j_L + A_V j_V = -A_I i_s(t), \tag{1.1}$$

$$\frac{d}{dt} \phi_L(j_L, t) - A_L^\top e = 0, \tag{1.2}$$

$$A_V^\top e = v_s(t). \tag{1.3}$$

Here, the incidence matrix $A = [A_R, A_C, A_L, A_V, A_I] = (a_{ij})$ represents the network topology, e.g. at each non mass node i, $a_{ij} = 1$ if the branch j leaves node i and $a_{ij} = -1$ if the branch j enters node i and $a_{ij} = 0$ elsewhere. The indices R, C, L, V, I denote the capacitive, resistive, inductive, voltage source, and current source branches, respectively. The functions q_C, g and ϕ_L are continuously differentiable defining the voltage-current relations of the network components. The continuous functions v_s and i_s are the voltage and current sources.

Under the assumption that the Jacobians

$$D_C(e,t) := \frac{\partial q_C}{\partial e}(e,t), \quad D_G(e,t) := \frac{\partial g}{\partial e}(e,t), \quad D_L(j,t) := \frac{\partial \phi_L}{\partial j}(j,t)$$

are positive definite, analytical properties (e.g. the index) of DAE (1.1)–(1.3) are investigated in [14] and [13]. In linear networks, the matrices D_C, D_G and D_L are positive definite diagonal matrices with capacitances, conductivities and inductances on the diagonal.

Often semiconductors themselves are modeled by electrical networks. These models are stored in a library and are stamped into the surrounding network in order to create a complete model of the integrated circuit. Here we use a different approach which uses the transient drift-diffusion equations as a continuous model for semiconductors. Advantages are the higher accuracy of the model and fewer model parameters. On the other hand, numerical simulations are more expensive. For a comprehensive overview of the drift-diffusion equations we refer to [1, 2, 8, 30, 43]. Using the notation introduced there, we have the following system of partial differential equations for the electrostatic potential $\psi(t,x)$, the electron and hole concentrations $n(t,x)$ and $p(t,x)$ and the current densities $J_n(t,x)$ and $J_p(t,x)$:

$$\begin{aligned}
\operatorname{div}(\varepsilon \operatorname{grad} \psi) &= q(n - p - C), \\
-q\partial_t n + \operatorname{div} J_n &= qR(n,p,J_n,J_p), \\
q\partial_t p + \operatorname{div} J_p &= -qR(n,p,J_n,J_p), \\
J_n &= \mu_n q(U_T \operatorname{grad} n - n \operatorname{grad} \psi), \\
J_p &= \mu_p q(-U_T \operatorname{grad} p - p \operatorname{grad} \psi),
\end{aligned}$$

with $(t,x) \in [0,T] \times \Omega$ and $\Omega \subset \mathbb{R}^d (d = 1,\ldots,3)$. The nonlinear function R describes the rate of electron/hole recombination, q is the elementary charge, ε the dielectricity, μ_n and μ_p are the mobilities of electrons and holes. The temperature is assumed to be constant which leads to a constant thermal voltage U_T. The function C is the time independent doping profile. Note that we do not formulate into quasi-Fermi potentials since the additional non-linearities would imply higher simulation time for the reduced model. Further details are given in [23]. The analytical and numerical analysis of systems of this form is subject to current research, see [7, 17, 46, 52].

1.2.1 Coupling

In the present section we develop the complete coupled system for a network with n_s semiconductors. We will not specify an extra index for semiconductors, but we

keep in mind that all semiconductor equations and coupling conditions need to be introduced for each semiconductor.

For the sake of simplicity we assume that to a semiconductor m semiconductor interfaces $\Gamma_{O,k} \subseteq \Gamma \subset \partial\Omega$, $k = 1, \ldots, m$ are associated, which are all Ohmic contacts, compare Fig. 1.2. The dielectricity ε shall be constant over the whole domain Ω. We focus on the Shockley-Read-Hall recombination

$$R(n,p) := \frac{np - n_i^2}{\tau_p(n + n_i) + \tau_n(p + n_i)}$$

which does not depend on the current densities. Herein, τ_n and τ_p are the average lifetimes of electrons and holes, and n_i is the constant intrinsic concentration which satisfy $n_i^2 = np$ if the semiconductor is in thermal equilibrium.

The scaled complete coupled system is constructed as follows. (We neglect the tilde-sign over the scaled variables.) The current through the diodes must be considered in Kirchhoff's current law. Consequently, the term $A_S j_S$ is added to Eq. (1.1), e.g.

$$A_C \frac{d}{dt} q_C(A_C^\top e, t) + A_R g(A_R^\top e, t) + A_L j_L + A_V j_V + A_S j_S = -A_I i_s(t), \tag{1.4}$$

$$\frac{d}{dt} \phi_L(j_L, t) - A_L^\top e = 0, \tag{1.5}$$

$$A_V^\top e = v_s(t). \tag{1.6}$$

In particular the matrix A_S denotes the semiconductor incidence matrix. Here,

$$j_{S,k} = \int_{\Gamma_{O,k}} (J_n + J_p - \varepsilon \partial_t \nabla\psi) \cdot \nu \, d\sigma. \tag{1.7}$$

I.e. the current is the integral over the current density $J_n + J_p$ plus the displacement current in normal direction ν. Furthermore, the potentials of nodes which are connected to a semiconductor interface are introduced in the boundary conditions of the drift-diffusion equations (see also Fig. 1.2):

$$\psi(t,x) = \psi_{bi}(x) + (A_S^\top e(t))_k = U_T \log\left(\frac{\sqrt{C(x)^2 + 4n_i^2} + C(x)}{2n_i} \right) + (A_S^\top e(t))_k, \tag{1.8}$$

$$n(t,x) = \frac{1}{2}\left(\sqrt{C(x)^2 + 4n_i^2} + C(x) \right), \tag{1.9}$$

$$p(t,x) = \frac{1}{2}\left(\sqrt{C(x)^2 + 4n_i^2} - C(x) \right), \tag{1.10}$$

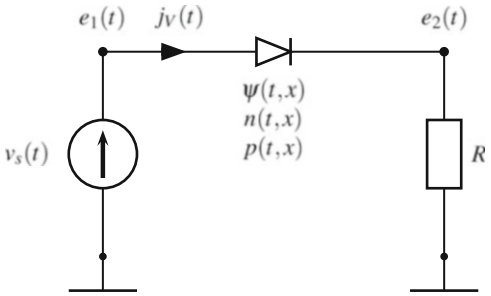

Fig. 1.1 Basic test circuit with one diode

for $(t,x) \in [0,T] \times \Gamma_{O,k}$. Here, $\psi_{bi}(x)$ denotes the build-in potential and n_i the constant intrinsic concentration. All other parts of the boundary are isolation boundaries $\Gamma_I := \Gamma \setminus \Gamma_O$, where $\nabla\psi \cdot \nu = 0$, $J_n \cdot \nu = 0$ and $J_p \cdot \nu = 0$ holds. For a basic example consider the network in Fig. 1.1 where the network is described by

$$A_V = \begin{pmatrix} 1, 0 \end{pmatrix}^\top, \; A_S = \begin{pmatrix} -1, 1 \end{pmatrix}^\top, \; A_R = \begin{pmatrix} 0, 1 \end{pmatrix}^\top, \text{ and } g(A_R^\top e, t) = \frac{1}{R} e_2(t).$$

The complete model forms a partial differential-algebraic equation (PDAE). The analytical and numerical analysis of such systems is subject to current research, see [7, 17, 46, 52]. The simulation of the complete coupled system is expensive and numerically difficult due to bad scaling of the drift-diffusion equations. The numerical issues can be significantly reduced by the unit scaling procedure discussed in [43]. That means we substitute

$$x = L\tilde{x}, \quad \psi = U_T\tilde{\psi}, \quad n = \|C\|_\infty \tilde{n}, \quad p = \|C\|_\infty \tilde{p}, \quad C = \|C\|_\infty \tilde{C},$$
$$J_n = \frac{qU_T\|C\|_\infty}{L}\mu_n \tilde{J}_n, \quad J_p = \frac{qU_T\|C\|_\infty}{L}\mu_p \tilde{J}_p, \quad n_i = \tilde{n}_i\|C\|_\infty,$$

where L denotes a specific length of the semiconductor (Fig. 1.2). The scaled drift-diffusion equations then read

$$\lambda \Delta\psi = n - p - C, \tag{1.11}$$
$$-\partial_t n + \nu_n \operatorname{div} J_n = \quad R(n,p), \tag{1.12}$$
$$\partial_t p + \nu_p \operatorname{div} J_p = -R(n,p), \tag{1.13}$$
$$J_n = \quad \nabla n - n\nabla\psi, \tag{1.14}$$
$$J_p = -\nabla p - p\nabla\psi, \tag{1.15}$$

where we omit the tilde for the scaled variables. The constants are given by $\lambda := \frac{\varepsilon U_T}{L^2 q\|C\|_\infty}$, $\nu_n := \frac{U_T\mu_n}{L^2}$ and $\nu_p := \frac{U_T\mu_p}{L^2}$, see e.g. [43].

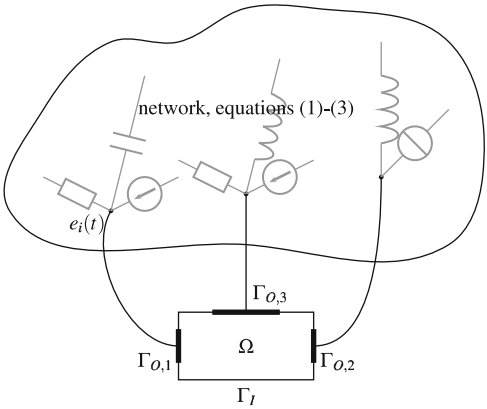

Fig. 1.2 Sketch of a coupled system with one semiconductor. Here $\psi(t,x) = e_i(t) + \psi_{bi}(x)$, for all $(t,x) \in [0,T] \times \Gamma_{O,1}$

1.3 Simulation of the Full System

Classical approaches for the simulation of drift-diffusion equations (e.g. Gummel iterations [21]) approximate J_n and J_p by piecewise constant functions and then solve Eqs. (1.12) and (1.13) with respect to n and p explicitly. This helps reducing the computational effort and increases the numerical stability. For the model order reduction approach proposed in the present work this method has the disadvantage of introducing additional non-linearities, arising from the exponential structure of the Slotboom variables, see [46]. Subsequently we propose two finite element discretizations for the drift-diffusion system which with regard to coping with nonlinearities are advantageous from the MOR reduction point of view, and which together with the equations for the electrical network finally lead to large-scale nonlinear DAE model for the fully coupled system.

1.3.1 Standard Galerkin Finite Element Approach

Let $\mathscr{T}$ denote a regular triangulation of the domain Ω with gridwidth h, whose simplexes are denoted by T. In the classical Galerkin finite element method the functions ψ, n and p are approximated by piecewise linear and globally continuous functions, while J_n and J_p are approximated by patchwise-piecewise constant functions, e.g.

$$\psi(t,x) := \sum_{i=1}^{N} \psi_i(t)\phi_i(x), \quad n(t,x) := \sum_{i=1}^{N} n_i(t)\phi_i(x), \quad p(t,x) := \sum_{i=1}^{N} p_i(t)\phi_i(x),$$

$$J_n(t,x) := \sum_{i=1}^{N} J_{n,i}(t)\varphi_i(x), \quad J_p(t,x) := \sum_{i=1}^{N} J_{p,i}(t)\varphi_i(x),$$

where the functions $\{\phi_i\}$ and $\{\varphi_i\}$ are the corresponding ansatz functions, and N denotes the number of degrees of freedom. For ψ, n and p the coefficients corresponding to the boundary elements are prescribed using the Dirichlet boundary conditions. Note that the time is not discretized at this point which refers to the so-called method of lines. The finite element method leads to the following DAE for the unknown vector-valued functions of time ψ, n, p, J_n, J_p for each semiconductor:

$$\begin{aligned} 0 &= \lambda S\psi(t) + Mn(t) - Mp(t) - C_h + b_\psi(A_S^T e(t)), \\ -M\dot{n}(t) &= -\nu_n D^\top J_n(t) + hR(n(t), p(t)), \\ M\dot{p}(t) &= -\nu_p D^\top J_p(t) - hR(n(t), p(t)), \\ 0 &= hJ_n(t) + Dn(t) - \operatorname{diag}\left(Bn(t) + \tilde{b}_n\right) D\psi(t) + b_n, \\ 0 &= hJ_p(t) - Dp(t) - \operatorname{diag}\left(Bp(t) + \tilde{b}_p\right) D\psi(t) + b_p, \end{aligned} \tag{1.16}$$

where S, M and D, B are assembled finite element matrices. The matrix $\operatorname{diag}(v)$ is diagonal with vector v forming the diagonal. The vectors $b_\psi(A_S^T e(t))$, b_n, $\tilde{b}_n$, b_p and $\tilde{b}_p$ implement the boundary conditions imposed on ψ, n and p through (1.8)–(1.10).

Discretization of the coupling condition for the current (1.7) completes the discretized system. In one spatial dimension we use

$$j_{S,k}(t) = \frac{aqU_T\|C\|_\infty}{L}\left(\mu_n J_{n,N}(t) + \mu_p J_{p,N}(t)\right) - \frac{a\varepsilon U_T}{Lh}\left(\dot{\psi}_N(t) - \dot{\psi}_{N-1}(t)\right),$$

1.3.2 Mixed Finite Element Approach

Since the electrical field represented by the (negative) gradient of the electrical potential ψ plays a dominant role in (1.11)–(1.15) and is present also in the coupling condition (1.7), we provide for it the additional variable $g_\psi = \nabla\psi$ leading to the following mixed formulation of the DD equations:

$$\lambda \operatorname{div} g_\psi = n - p - C, \tag{1.17}$$

$$-\partial_t n + \nu_n \operatorname{div} J_n = R(n, p), \tag{1.18}$$

$$\partial_t p + \nu_p \operatorname{div} J_p = -R(n, p), \tag{1.19}$$

$$g_\psi = \nabla\psi, \tag{1.20}$$

$$J_n = \nabla n - n g_\psi, \tag{1.21}$$

$$J_p = -\nabla p - p g_\psi. \tag{1.22}$$

The weak formulation of (1.17)–(1.22) then reads: Find $\psi, n, p \in [0,T] \times L^2(\Omega)$ and $g_\psi, J_n, J_p \in [0,T] \times H_{0,N}(\mathrm{div}, \Omega)$ such that

$$\lambda \int_\Omega \mathrm{div}\, g_\psi \; \varphi = \int_\Omega (n-p)\, \varphi - \int_\Omega C\, \varphi, \tag{1.23}$$

$$-\int_\Omega \partial_t n\, \varphi + \nu_n \int_\Omega \mathrm{div}\, J_n\, \varphi = \int_\Omega R(n,p)\, \varphi, \tag{1.24}$$

$$\int_\Omega \partial_t p\, \varphi + \nu_p \int_\Omega \mathrm{div}\, J_p\, \varphi = -\int_\Omega R(n,p)\, \varphi, \tag{1.25}$$

$$\int_\Omega g_\psi \cdot \phi = -\int_\Omega \psi\ \mathrm{div}\, \phi + \int_\Gamma \psi\ \phi \cdot \nu, \tag{1.26}$$

$$\int_\Omega J_n \cdot \phi = -\int_\Omega n\ \mathrm{div}\, \phi + \int_\Gamma n\, \phi \cdot \nu - \int_\Omega n\, g_\psi \cdot \phi, \tag{1.27}$$

$$\int_\Omega J_p \cdot \phi = \int_\Omega p\ \mathrm{div}\, \phi - \int_\Gamma p\, \phi \cdot \nu - \int_\Omega p\, g_\psi \cdot \phi, \tag{1.28}$$

are satisfied for all $\varphi \in L^2(\Omega)$ and $\phi \in H_{0,N}(\mathrm{div}, \Omega)$ where the space $H_{0,N}(\mathrm{div}, \Omega)$ is defined by

$$H(\mathrm{div}, \Omega) := \{y \in L^2(\Omega)^d :\ \mathrm{div}\, y \in L^2(\Omega)\},$$
$$H_{0,N}(\mathrm{div}, \Omega) := \{y \in H(\mathrm{div}, \Omega) :\ y \cdot \nu = 0 \text{ on } \Gamma_I\}.$$

Consequently, the boundary integrals on the right hand sides in Eqs. (1.26)–(1.28) reduce to integrals over the interfaces $\Gamma_{O,k}$, where the values of ψ, n and p are determined by the Dirichlet boundary conditions (1.8)–(1.10). We note that, in contrast to the standard weak form associated with (1.11)–(1.15), the Dirichlet boundary values are naturally included in the weak formulation (1.23)–(1.28) and the Neumann boundary conditions have to be included in the space definitions. This is advantageous in the context of POD model order reduction since the non-homogeneous boundary conditions (1.8)–(1.10) are not present in the space definitions.

Here, Eqs. (1.23)–(1.28) are discretized in space with Raviart-Thomas finite elements of degree 0 (RT_0), alternative discretization schemes for the mixed problem are presented in [8]. To describe the RT_0-approach for $d = 2$ spatial dimensions, let $\mathscr{T}$ be a triangulation of Ω and let $\mathscr{E}$ be the set of all edges. Let $\mathscr{E}_I := \{E \in \mathscr{E} :\ E \subset \bar{\Gamma}_I\}$ be the set of edges at the isolation (Neumann) boundaries. The potential and the concentrations are approximated in space by piecewise constant functions

$$\psi^h(t), n^h(t), p^h(t) \in L_h := \{y \in L^2(\Omega) :\ y|_T(x) = c_T,\ \forall T \in \mathscr{T}\},$$

with ansatz functions $\{\varphi_i\}_{i=1,\dots,N}$ and the discrete fluxes $g_\psi^h(t)$, $J_n^h(t)$ and $J_p^h(t)$ are elements of the space

$$RT_0 := \{y : \Omega \to \mathbb{R}^d : \ y|_T(x) = a_T + b_T x,\ a_T \in \mathbb{R}^d,\ b_T \in \mathbb{R}, [y]_E \cdot \nu_E = 0, \\ \text{for all inner edges } E\}.$$

Here, $[y]_E$ denotes the jump $y|_{T_+} - y|_{T_-}$ over a shared edge E of the elements T_+ and T_-. The continuity assumption yields $\mathrm{RT}_0 \subset H(\mathrm{div}, \Omega)$. We set

$$H_{h,0,N}(\mathrm{div}, \Omega) := (\mathrm{RT}_0 \cap H_{0,N}(\mathrm{div}, \Omega)) \subset H_{0,N}(\mathrm{div}, \Omega).$$

Then it can be shown, that $H_{h,0,N}$ posses an edge-oriented basis $\{\phi_j\}_{j=1,\dots,M}$. We use the following finite element ansatz in (1.23)–(1.28)

$$\left.\begin{aligned} \psi^h(t,x) &= \sum_{i=1}^N \psi_i(t)\varphi_i(x), & g_\psi^h(t,x) &= \sum_{j=1}^M g_{\psi,j}(t)\phi_j(x), \\ n^h(t,x) &= \sum_{i=1}^N n_i(t)\varphi_i(x), & J_n^h(t,x) &= \sum_{j=1}^M J_{n,j}(t)\phi_j(x), \\ p^h(t,x) &= \sum_{i=1}^N p_i(t)\varphi_i(x), & J_p^h(t,x) &= \sum_{j=1}^M J_{p,j}(t)\phi_j(x), \end{aligned}\right\} \tag{1.29}$$

where $N := |\mathscr{T}|$, i.e. the number of elements of $\mathscr{T}$, and $M := |\mathscr{E}| - |\mathscr{E}_N|$, i.e. the number of inner and Dirichlet boundary edges.

This in (1.23)–(1.28) yields

$$\lambda \sum_{j=1}^M g_{\psi,j}(t) \int_\Omega \mathrm{div}\,\phi_j\,\varphi_k - \sum_{i=1}^N (n_i(t) - p_i(t)) \int_\Omega \varphi_i\,\varphi_k = -\int_\Omega C\,\varphi_k,$$

$$-\sum_{i=1}^N \dot{n}_i(t) \int_\Omega \varphi_i\,\varphi_k + \nu_n \sum_{j=1}^M J_{n,j}(t) \int_\Omega \mathrm{div}\,\phi_j\,\varphi_k - \int_\Omega R(n^h, p^h)\,\varphi_k = 0,$$

$$\sum_{i=1}^N \dot{p}_i(t) \int_\Omega \varphi_i\,\varphi_k + \nu_p \sum_{j=1}^M J_{p,j}(t) \int_\Omega \mathrm{div}\,\phi_j\,\varphi_k + \int_\Omega R(n^h, p^h)\,\varphi_k = 0,$$

$$\sum_{j=1}^M g_{\psi,j}(t) \int_\Omega \phi_j \cdot \phi_l + \sum_{i=1}^N \psi_i(t) \int_\Omega \varphi_i\ \mathrm{div}\,\phi_l = \int_\Gamma \psi^h\,\phi_l \cdot \nu,$$

$$\sum_{j=1}^{M} J_{n,j}(t) \int_\Omega \phi_j \cdot \phi_l + \sum_{i=1}^{N} n_i(t) \int_\Omega \varphi_i \ \mathrm{div}\, \phi_l + \int_\Omega n^h g_\psi^h \cdot \phi_l = \int_\Gamma n^h \ \phi_l \cdot \nu,$$

$$\sum_{j=1}^{M} J_{p,j}(t) \int_\Omega \phi_j \cdot \phi_l - \sum_{i=1}^{N} p_i(t) \int_\Omega \varphi_i \ \mathrm{div}\, \phi_l + \int_\Omega p^h g_\psi^h \cdot \phi_l = -\int_\Gamma p^h \ \phi_l \cdot \nu,$$

which represents a nonlinear, large and sparse DAE for the approximation of the functions ψ, n, p, g_ψ, J_n, and J_p. In matrix notation it reads

$$\begin{pmatrix} 0 \\ -M_L \dot{n}(t) \\ M_L \dot{p}(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + \underbrace{\begin{pmatrix} & -M_L & M_L & \lambda D & & \\ & & & & \nu_n D & \\ & & & & & \nu_p D \\ D^\top & & & M_H & & \\ & D^\top & & & M_H & \\ & & -D^\top & & & M_H \end{pmatrix}}_{A_{FEM}} \begin{pmatrix} \psi(t) \\ n(t) \\ p(t) \\ g_\psi(t) \\ J_n(t) \\ J_p(t) \end{pmatrix}$$

$$+ \mathscr{F}(n^h, p^h, g_\psi^h) = b(A_S^T e(t)),$$

with

$$\mathscr{F}(n^h, p^h, g_\psi^h) := \begin{pmatrix} 0 \\ -\int_\Omega R(n^h, p^h)\ \varphi \\ \int_\Omega R(n^h, p^h)\ \varphi \\ 0 \\ \int_\Omega n^h g_\psi^h \cdot \phi \\ \int_\Omega p^h g_\psi^h \cdot \phi \end{pmatrix}, \quad b := \begin{pmatrix} -\int_\Omega C\ \varphi \\ 0 \\ 0 \\ \int_\Gamma \psi^h (A_S^T e(t))\ \phi \cdot \nu \\ \int_\Gamma n^h\ \phi \cdot \nu \\ -\int_\Gamma p^h\ \phi \cdot \nu \end{pmatrix}, \tag{1.30}$$

and

$$\int_\Omega R(n^h, p^h)\varphi := \begin{pmatrix} \int_\Omega R(n^h, p^h)\varphi_1 \\ \vdots \\ \int_\Omega R(n^h, p^h)\varphi_N \end{pmatrix}.$$

All other integrals in $\mathscr{F}$ and b are defined analogously. The matrices $M_L \in \mathbb{R}^{N\times N}$ and $M_H \in \mathbb{R}^{M\times M}$ are mass matrices in the spaces L_h and $H_{h,0,N}$, respectively, and $D \in \mathbb{R}^{N\times M}$. The final DAE for the mixed finite element discretization now takes the form

Problem 1.3.1 (Full Model)

$$A_C \frac{d}{dt} q_C(A_C^\top e(t), t) + A_R g(A_R^\top e(t), t) + A_L j_L(t) + A_V j_V(t)$$
$$+ A_S j_S(t) + A_I i_s(t) = 0, \tag{1.31}$$
$$\frac{d}{dt} \phi_L(j_L(t), t) - A_L^\top e(t) = 0, \tag{1.32}$$
$$A_V^\top e(t) - v_s(t) = 0, \tag{1.33}$$
$$j_S(t) - C_1 J_n(t) - C_2 J_p(t) - C_3 \dot{g}_\psi(t) = 0, \tag{1.34}$$
$$\begin{pmatrix} 0 \\ -M_L \dot{n}(t) \\ M_L \dot{p}(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{FEM} \begin{pmatrix} \psi(t) \\ n(t) \\ p(t) \\ g_\psi(t) \\ J_n(t) \\ J_p(t) \end{pmatrix} + \mathscr{F}(n^h, p^h, g_\psi^h) - b(A_S^T e(t)) = 0, \tag{1.35}$$

where (1.34) represents the discretized linear coupling condition (1.7).

We present numerical computations for the basic test circuit with one diode depicted in Fig. 1.1, where the model parameters are presented in Table 1.1. The input $v_s(t)$ is chosen to be sinusoidal with amplitude 5 V. The numerical results in Fig. 1.3 show the capacitive effect of the diode for high input frequencies. Similar results are obtained in [44] using the simulator MECS.

The discretized equations are implemented in MATLAB, and the DASPK software package [34] is used to integrate the high-dimensional DAE. Initial values are stationary states obtained by setting all time derivatives to 0. In order to solve the Newton systems which arise from the BDF method efficiently, one may reorder the variables of the sparse system with respect to minimal bandwidth. Then, one can use the internal DASPK routines for the solution of the linear systems. Alternatively one can implement the preconditioning subroutine of DASPK using a direct sparse solver. Note that for both strategies we only need to calculate the reordering matrices once, since the sparsity structure remains constant.

Table 1.1 Diode model parameters

Parameter	Value	Parameter	Value
L	10^{-4} cm	ε	$1.03545 \cdot 10^{-12}$ F/cm
U_T	0.0259 V	n_i	$1.4 \cdot 10^{10}$ 1/cm^3
μ_n	1350 cm^2/(V s)	τ_n	$330 \cdot 10^{-9}$ s
μ_p	480 cm^2/(V s)	τ_p	$33 \cdot 10^{-9}$ s
a	10^{-5} cm^2	$C(x),\ x < L/2$	$-9.94 \cdot 10^{15}$ 1/cm^3
		$C(x),\ x \geq L/2$	$4.06 \cdot 10^{18}$ 1/cm^3

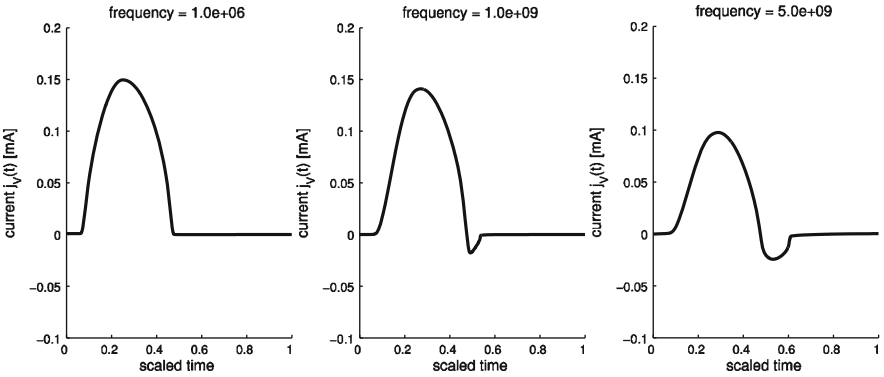

Fig. 1.3 Current j_V through the basic network for input frequencies 1 MHz, 1 GHz and 5 GHz. The capacitive effect is clearly demonstrated

1.4 Model Order Reduction Using POD

We now use proper orthogonal decomposition (POD) to construct low-dimensional surrogate models for the drift-diffusion equations. The idea consists in replacing the large number of local model-independent ansatz and test functions $\{\phi_i\}, \{\varphi_j\}$ in the finite element approximation of the drift-diffusion systems by only a few nonlocal model-dependent ansatz functions for the respective variables.

The snapshot variant of POD introduced in [45] works as follows. We run a simulation of the unreduced system and collect l snapshots $\psi^h(t_k,\cdot)$, $n^h(t_k,\cdot)$, $p^h(t_k,\cdot)$, $g^h_\psi(t_k,\cdot)$, $J^h_n(t_k,\cdot)$, $J^h_p(t_k,\cdot)$ at time instances $t_k \in \{t_1,\dots,t_l\} \subset [0,T]$. The optimal selection of the time instances is not considered here. We use the time instances delivered by the DAE integrator.

Since every component of the state vector $y := (\psi, n, p, g_\psi, J_n, J_p)$ has its own physical meaning we apply POD MOR to each component separately. Among other things this approach has the advantage of yielding a block-dense model and the approximation quality of each component is adapted individually.

Let X denote a Hilbert space and let $y^h : [0,T] \times X \to \mathbb{R}^r$ with some $r \in \mathbb{N}$. The Galerkin formulation (1.29) yields $y^h(t,\cdot) \in X_h := \operatorname{span}\{\phi^X_1,\dots,\phi^X_n\}$, where $\{\phi^X_j\}_{1\le j\le n}$ denote n linearly independent elements of X. The idea of POD consists in finding a basis $\{u^1,\dots,u^m\}$ of the span of the snapshots

$$\operatorname{span}\left\{ y^h(t_k,\cdot) = \sum_{i=1}^{n} y_i^{h,k}\phi_i^X(\cdot),\ \text{with } k = 1,\dots,l \right\}$$

satisfying

$$\{u^1,\dots,u^s\} = \operatorname*{arg\,min}_{\{v^1,\dots,v^s\}\subset X} \sum_{k=1}^{l} \left\| y^h(t_k,\cdot) - \sum_{i=1}^{s} \langle y^h(t_k,\cdot), v^i(\cdot)\rangle_X v^i(\cdot) \right\|_X^2 ,$$

for $1 \le s \le m$, where $1 \le m \le l$. The functions $\{u^i\}_{1\le i\le s}$ are orthonormal in X and can be obtained with the help of the singular value decomposition (SVD) as follows.

Let the matrix $Y := (y^{h,1}, \ldots, y^{h,l}) \in \mathbb{R}^{n\times l}$ contain as columns the coefficient vectors of the snapshots. Furthermore, let $M := (\langle \phi_i^X, \phi_j^X\rangle_X)_{1\le i,j\le n}$ be the positive definite mass matrix with its Cholesky factorization $M = LL^\top$. Let $(\tilde U, \Sigma, \tilde V)$ denote the SVD of $\tilde Y := L^\top Y$, i.e. $\tilde Y = \tilde U \Sigma \tilde V^\top$ with $\tilde U \in \mathbb{R}^{n\times n}$, $\tilde V \in \mathbb{R}^{l\times l}$, and a matrix $\Sigma \in \mathbb{R}^{n\times l}$ containing the singular values $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_m > \sigma_{m+1} = \ldots = \sigma_l = 0$. We set $U := L^{-\top}\tilde U_{(:,1:s)}$. Then, the s-dimensional POD basis is given by

$$\operatorname{span}\left\{ u^i(\cdot) = \sum_{j=1}^{n} U_{ji}\phi_j^X(\cdot),\ i = 1, \ldots, s \right\}.$$

The information content of $\{u^1, \ldots, u^s\}$ with respect to the scalar product $\langle\cdot,\cdot\rangle_X$ with

$$0 \le \Delta(s) = \sqrt{\frac{\sum_{i=s+1}^m \sigma_i^2}{\sum_{i=1}^m \sigma_i^2}} \le 1, \tag{1.36}$$

is given by $1 - \Delta(s)$. Here $\Delta(s)$ measures the lack of information of $\{u^1, \ldots, u^s\}$ with respect to $\operatorname{span}\{y^h(t_1,\cdot), \ldots, y^h(t_l,\cdot)\}$. An extended introduction to POD can be found in [36].

The POD basis functions are now used as trial and test functions in the Galerkin method.

If the snapshots satisfy inhomogeneous Dirichlet boundary conditions, as in (1.16), POD is performed for

$$\tilde\psi(t) = \psi(t) - \psi_r(t), \quad \tilde n(t) = n(t) - n_r(t), \quad \tilde p(t) = p(t) - p_r(t),$$

with ψ_r, n_r, p_r denoting reference functions satisfying the Dirichlet boundary conditions required for ψ, n and p. This guarantees that the POD basis admits homogeneous boundary conditions on the Dirichlet boundary.

In the case of the mixed finite element approach the introduction of a reference state is not necessary, since the boundary values are included more naturally through the variational formulation. The time-snapshot POD procedure then delivers Galerkin ansatz spaces for ψ, n, p, g_ψ, J_n and J_p. This leads to the ansatz

$$\left.\begin{aligned} \psi^{POD}(t) &= U_\psi \gamma_\psi(t), & g_\psi^{POD}(t) &= U_{g_\psi}\gamma_{g_\psi}(t),\\ n^{POD}(t) &= U_n\gamma_n(t), & J_n^{POD}(t) &= U_{J_n}\gamma_{J_n}(t),\\ p^{POD}(t) &= U_p\gamma_p(t), & J_p^{POD}(t) &= U_{J_p}\gamma_{J_p}(t). \end{aligned}\right\} \tag{1.37}$$

The injection matrices

$$U_\psi \in \mathbb{R}^{N\times s_\psi}, \qquad U_n \in \mathbb{R}^{N\times s_n}, \qquad U_p \in \mathbb{R}^{N\times s_p},$$
$$U_{g_\psi} \in \mathbb{R}^{M\times s_{g_\psi}}, \qquad U_{J_n} \in \mathbb{R}^{M\times s_{J_n}}, \qquad U_{J_p} \in \mathbb{R}^{M\times s_{J_p}},$$

contain the (time independent) POD basis functions, the vectors $\gamma_{(\cdot)}$ the corresponding time-variant coefficients. The numbers $s_{(\cdot)}$ are the respective number of POD basis functions included. Assembling the POD system yields the DAE

$$\begin{pmatrix} 0 \\ -\dot\gamma_n(t) \\ \dot\gamma_p(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{POD} \begin{pmatrix} \gamma_\psi(t) \\ \gamma_n(t) \\ \gamma_p(t) \\ \gamma_{g_\psi}(t) \\ \gamma_{J_n}(t) \\ \gamma_{J_p}(t) \end{pmatrix} + U^\top \mathscr{F}(n^{POD}, p^{POD}, g_\psi^{POD}) = U^\top b(A_S^T e(t)),$$

with

$$A_{POD} = U^\top A_{FEM} U$$
$$= \begin{pmatrix} & -U_\psi^\top M_L U_n & U_\psi^\top M_L U_p & \lambda U_\psi^\top D U_{g_\psi} & & \\ & & & & \nu_n U_n^\top D U_{J_n} & \\ & & & & & \nu_p U_p^\top D U_{J_p} \\ U_{g_\psi}^\top D^\top U_\psi & & & I & & \\ & U_{J_n}^\top D^\top U_n & & & I & \\ & & -U_{J_p}^\top D^\top U_p & & & I \end{pmatrix}$$

and $U = \mathrm{diag}(U_\psi, U_n, U_p, U_{g_\psi}, U_{J_n}, U_{J_p})$. Note that we exploit the orthogonality of the POD basis functions, e.g. $U_n^\top M_L U_n = U_p^\top M_L U_p = I_{N\times N}$ and $U_{g_\psi}^\top M_H U_{g_\psi} = U_{J_n}^\top M_H U_{J_n} = U_{J_p}^\top M_H U_{J_p} = I_{M\times M}$. The arguments of the nonlinear functional have to be interpreted as functions in space.

All matrix-matrix multiplications are calculated in an offline phase. The nonlinear functional $\mathscr{F}$ has to be evaluated online. The reduced model for the network now reads

Problem 1.4.1 (POD MOR Surrogate)

$$A_C \frac{d}{dt} q_C(A_C^\top e(t), t) + A_R g(A_R^\top e(t), t) + A_L j_L(t) + A_V j_V(t) + A_S j_S(t) + A_I i_s(t) = 0, \tag{1.38}$$

$$\frac{d}{dt}\phi_L(j_L(t),t) - A_L^\top e(t) = 0, \tag{1.39}$$

$$A_V^\top e(t) - v_s(t) = 0, \tag{1.40}$$

$$j_S(t) - C_1 U_{J_n} \gamma_{J_n}(t) - C_2 U_{J_p} \gamma_{J_p}(t) - C_3 U_{g_\psi} \dot{\gamma}_{g_\psi}(t) = 0, \tag{1.41}$$

$$\begin{pmatrix} 0 \\ -\dot{\gamma}_n(t) \\ \dot{\gamma}_p(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{POD} \begin{pmatrix} \gamma_\psi(t) \\ \gamma_n(t) \\ \gamma_p(t) \\ \gamma_{g_\psi}(t) \\ \gamma_{J_n}(t) \\ \gamma_{J_p}(t) \end{pmatrix} + U^\top \mathscr{F}(n^{POD}, p^{POD}, g_\psi^{POD}) - U^\top b(A_S^T e(t)) = 0. \tag{1.42}$$

1.4.1 Numerical Investigation

We now present numerical examples for POD MOR of the basic test circuit in Fig. 1.1 and validate the reduced model at a fixed reference frequency of 10^{10} Hz. Figure 1.4 (left) shows the development of the error between the reduced and the unreduced numerical solutions, plotted over the neglected information Δ, see (1.36), which is measured by the relative error between the non-reduced states ψ, n, p, J_n, J_p and their projections onto the respective reduced state space. The number of POD basis functions for each variable is chosen such that the indicated approximation quality is reached, i.e. $\Delta := \Delta_\psi \simeq \Delta_n \simeq \Delta_p \simeq \Delta_{g_\psi} \simeq \Delta_{J_n} \simeq \Delta_{J_p}$. Since we compute all POD basis functions anyway, this procedure does not involve any additional costs.

In Fig. 1.4 (right) the simulation times are plotted versus the neglected information Δ. As one also can see, the simulation based on standard finite elements takes twice as long as if based on RT elements. However, this difference is not observed for the simulation of the corresponding reduced models.

Figure 1.5 shows the total number of singular vectors $k = k_\psi + k_n + k_p + k_{J_n} + k_{J_p}$ required in the POD model to guarantee a given state space cut-off error Δ. While the number of singular vectors included increases only linearly, the cut-off error tends to zero exponentially.

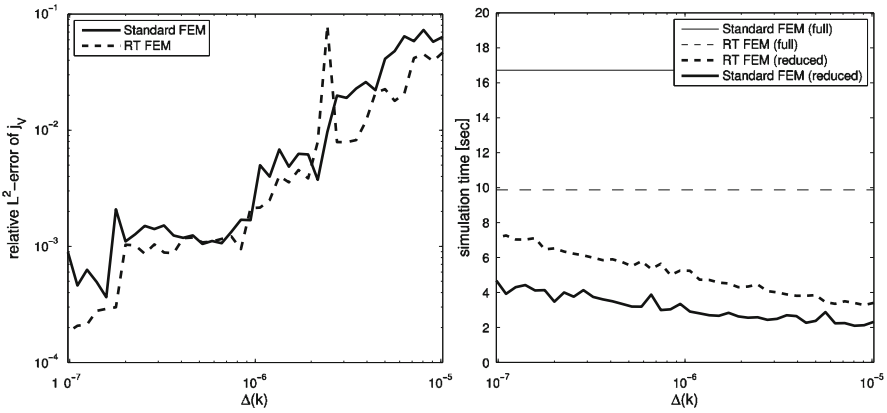

Fig. 1.4 *Left*: L_2 error of j_V between reduced and unreduced problem, both for standard and Raviart-Thomas FEM. *Right*: Time consumption for simulation runs for *left figure*. The *fine lines* indicate the time consumption for the simulation of the original full system

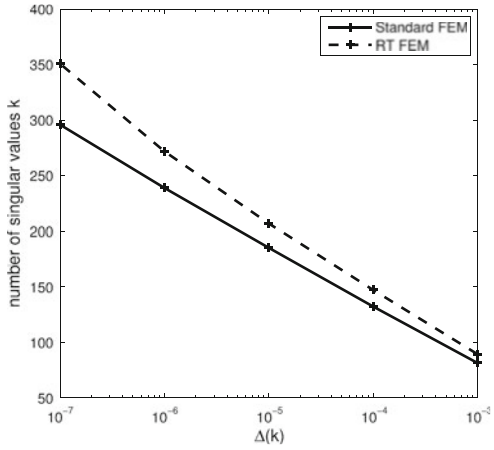

Fig. 1.5 The number of required singular values grows only logarithmically with the requested accuracy

1.4.2 Numerical Investigation, Position of the Semiconductor in the Network

Finally we note that the presented reduction method accounts for the position of the semiconductors in a given network in that it provides reduced-order models which for identical semiconductors may be different depending on the location of the semiconductors in the network. The POD basis functions of two identical semiconductors may be different due to their different operating states. To demonstrate this fact, we consider the rectifier network in Fig. 1.6 (left). Simulation results are

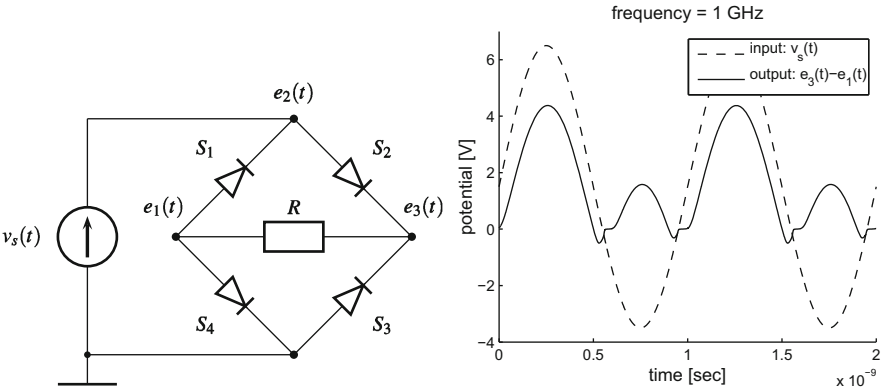

Fig. 1.6 *Left*: Rectifier network. *Right*: Simulation results for the rectifier network. The input v_s is sinusoidal with frequency 1 GHz and offset $+1.5$ V

Table 1.2 Distances between reduced models in the rectifier network

Δ	$d(U^1, U^2)$	$d(U^1, U^3)$
10^{-4}	0.61288	$5.373 \cdot 10^{-8}$
10^{-5}	0.50766	$4.712 \cdot 10^{-8}$
10^{-6}	0.45492	$2.767 \cdot 10^{-7}$
10^{-7}	0.54834	$1.211 \cdot 10^{-6}$

plotted in Fig. 1.6 (right). The distance between the spaces U^1 and U^2 which are spanned, e.g., by the POD-functions U^1_ψ of the diode S_1 and U^2_ψ of the diode S_2 respectively, is measured by

$$d(U^1, U^2) := \max_{\substack{u \in U^1 \\ \|u\|_2 = 1}} \min_{\substack{v \in U^2 \\ \|v\|_2 = 1}} \|u - v\|_2.$$

Exploiting the orthonormality of the bases U^1_ψ and U^2_ψ and using a Lagrange framework, we find

$$d(U^1, U^2) = \sqrt{2 - 2\sqrt{\lambda}},$$

where λ is the smallest eigenvalue of the positive definite matrix $SS^\top$ with $S_{ij} = \langle u^1_{\psi,i}, u^2_{\psi,j} \rangle_2$. The distances for the rectifier network are given in Table 1.2. While the reduced model for the diodes S_1 and S_3 are almost equal, the models for the diodes S_1 and S_2 are significantly different. Similar results are obtained for the reduction of n, p, etc.

1.4.3 MOR for the Nonlinearity with DEIM

The nonlinear function $\mathscr{F}$ in (1.42) has to be evaluated online which means that the computational complexity of the reduced-order model still depends on the number of unknowns of the unreduced model. A reduction method for the nonlinearity is given by Discrete Empirical Interpolation (DEIM) [10]. This method is motivated by the following observation. The nonlinearity in (1.42), see also (1.30), is given by

$$U^\top \mathscr{F}(U\gamma(t)) = \begin{pmatrix} 0 \\ U_n^\top F_n(U_n\gamma_n(t), U_p\gamma_p(t)) \\ U_p^\top F_p(U_n\gamma_n(t), U_p\gamma_p(t)) \\ 0 \\ U_{J_n}^\top F_{J_n}(U_n\gamma_n(t), U_{g_\psi}\gamma_{g_\psi}(t)) \\ U_{J_p}^\top F_{J_p}(U_n\gamma_p(t), U_{g_\psi}\gamma_{g_\psi}(t)) \end{pmatrix},$$

see e.g. [23]. The subsequent considerations apply for each block component of $\mathscr{F}$. For the sake of presentation we only consider the second block

$$\underbrace{U_n^\top}_{\text{size } s_n\times\mathbf{N}} \quad \underbrace{F_n}_{\mathbf{N} \text{ evaluations}} \quad \Big(\underbrace{U_n}_{\text{size } \mathbf{N}\times s_n} \gamma_n(t), \underbrace{U_p}_{\text{size } \mathbf{N}\times s_p} \gamma_p(t) \Big), \tag{1.43}$$

and its derivative with respect to γ_p,

$$\underbrace{U_n^\top}_{\text{size } s_n\times\mathbf{N}} \underbrace{\frac{\partial F_n}{\partial p}(U_n\gamma_n(t), U_p\gamma_p(t))}_{\text{size } \mathbf{N}\times\mathbf{N}, \text{ sparse}} \underbrace{U_p}_{\text{size } \mathbf{N}\times s_p}.$$

Here, the matrices $U_{(\cdot)}$ are dense and the Jacobian of F_n is sparse. The evaluation of (1.43) is of computational complexity $O(N)$. Furthermore, we need to multiply large dense matrices in the evaluation of the Jacobian. Thus, the POD model order reduction may become inefficient.

To overcome this problem, we apply Discrete Empirical Interpolation Method (DEIM) proposed in [10], which we now describe briefly. The snapshots $\psi^h(t_k,\cdot)$, $n^h(t_k,\cdot)$, $p^h(t_k,\cdot)$, $g_\psi^h(t_k,\cdot)$, $J_n^h(t_k,\cdot)$, $J_p^h(t_k,\cdot)$ are collected at time instances $t_k \in \{t_1,\dots,t_l\} \subset [0,T]$ as before. Additionally, we collect snapshots $\{F_n(n(t_k),p(t_k))\}$ of the nonlinearity. DEIM approximates the projected function (1.43) such that

$$U_n^\top F_n(U_n\gamma_n(t), U_p\gamma_p(t)) \approx (U_n^\top V_n (P_n^\top V_n)^{-1}) P_n^\top F_n(U_n\gamma_n(t), U_p\gamma_p(t)),$$

where $V_n \in \mathbb{R}^{N\times\tau_n}$ contains the first τ_n POD basis functions of the space spanned by the snapshots $\{F_n(n(t_k),p(t_k))\}$ associated with the largest singular values. The selection matrix $P_n = \left(e_{\rho_1},\dots,e_{\rho_{\tau_n}}\right) \in \mathbb{R}^{N\times\tau_n}$ selects the rows of F_n corresponding

to the so-called DEIM indices $\rho_1, \ldots, \rho_{\tau_n}$ which are chosen such that the growth of a global error bound is limited and $P_n^\top V_n$ is regular, see [10] for details.

The matrix $W_n := (U_n^\top V_n (P_n^\top V_n)^{-1}) \in \mathbb{R}^{s_n \times \tau_n}$ as well as the whole interpolation method is calculated in an offline phase. In the simulation of the reduced-order model we instead of (1.43) evaluate:

$$\underbrace{W_n}_{\text{size } s_n \times \tau_n} \quad \underbrace{P_n^\top F_n}_{\boldsymbol{\tau_n} \text{ evaluations}} \quad \Big(\underbrace{U_n}_{\text{size } \mathbf{N} \times s_n} \gamma_n(t), \underbrace{U_p}_{\text{size } \mathbf{N} \times s_p} \gamma_p(t) \Big), \tag{1.44}$$

with derivative

$$\underbrace{W_n^\top}_{\text{size } s_n \times \tau_n} \quad \underbrace{\frac{\partial P_n^\top F_n}{\partial p}(U_n \gamma_n(t), U_p \gamma_p(t))}_{\text{size } \boldsymbol{\tau_n} \times \mathbf{N}, \text{ sparse}} \underbrace{U_p}_{\text{size } \mathbf{N} \times s_p}.$$

In the applied finite element method a single functional component of F_n only depends on a small constant number $c \in \mathbb{N}$ components of $U_n \gamma_n(t)$. Thus, the matrix-matrix multiplication in the derivative does not really depend on N since the number of entries per row in the Jacobian is at most c.

But there is still a dependence on N, namely the calculation of $U_n \gamma_n(t)$. To overcome this dependency we identify the required components of the vector $U_n \gamma_n(t)$ for the evaluation of $P_n^\top F_n$. This is done by defining selection matrices $Q_{n,n} \in \mathbb{R}^{c\tau_n \times s_n}$, $Q_{n,p} \in \mathbb{R}^{c\tau_p \times s_p}$ such that

$$P_n^\top F_n(U_n \gamma_n(t), U_p \gamma_p(t)) = \hat{F}_n(Q_{n,n} U_n \gamma_n(t), Q_{n,p} U_p \gamma_p(t)),$$

where $\hat{F}_n$ denotes the functional components of F_n selected by P_n restricted to the arguments selected by $Q_{n,n}$ and $Q_{n,p}$.

Supposed that $\tau_n \approx s_n \ll N$ we obtain a reduced-order model which does not depend on N any more.

1.4.4 Numerical Implementation and Results with DEIM

We again use the basic test circuit with a single 1-dimensional diode depicted in Fig. 1.1. The parameters of the diode are summarized in [23]. The input $v_s(t)$ is chosen to be sinusoidal with amplitude 5 V. In the sequel the frequency of the voltage source will be considered as a model parameter.

We first validate the reduced model at a fixed reference frequency of $5 \cdot 10^9$ Hz. Figure 1.7 shows the development of the relative error between the POD reduced, the POD-DEIM reduced and the unreduced numerical solutions, plotted over the lack of information Δ of the POD basis functions with respect to the space spanned by the snapshots. The figure shows that the approximation quality of the POD-DEIM

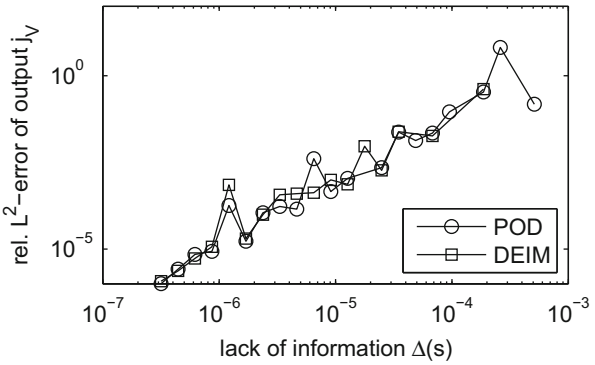

Fig. 1.7 Relative error between DEIM-reduced and unreduced nonlinearity at the fixed frequency $5 \cdot 10^9$ Hz

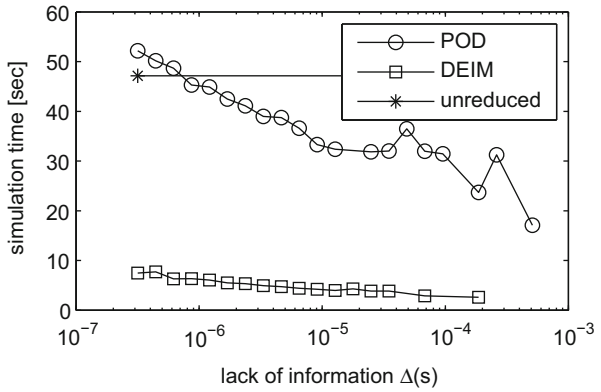

Fig. 1.8 Time consumption for simulation runs of Fig. 1.7. The *horizontal line* indicates the time consumption for the simulation of the original full system

reduced model is comparable with the more expensive POD reduced model. The number of POD basis functions $s_{(\cdot)}$ for each variable is chosen such that the indicated approximation quality is reached, i.e. $\Delta := \Delta_\psi \simeq \Delta_n \simeq \Delta_p \simeq \Delta_{g_\psi} \simeq \Delta_{J_n} \simeq \Delta_{J_p}$. The numbers $\tau_{(\cdot)}$ of POD-DEIM basis functions are chosen likewise.

In Fig. 1.8 the simulation times are plotted versus the lack of information Δ. The POD reduced-order model does not reduce the simulation times significantly for the chosen parameters. The reason for this is its dependency on the number of variables of the unreduced system. Here, the unreduced system contains 1000 finite elements which yields 12,012 unknowns. The POD-DEIM reduced-order model behaves very well and leads to a reduction in simulation time of about 90% without reducing the accuracy of the reduced model. However, we have to report a minor drawback; not all tested reduced models converge for large $\Delta(s) \geq 3 \cdot 10^{-5}$. This is indicated in the figures by missing squares. This effect is even more pronounced for

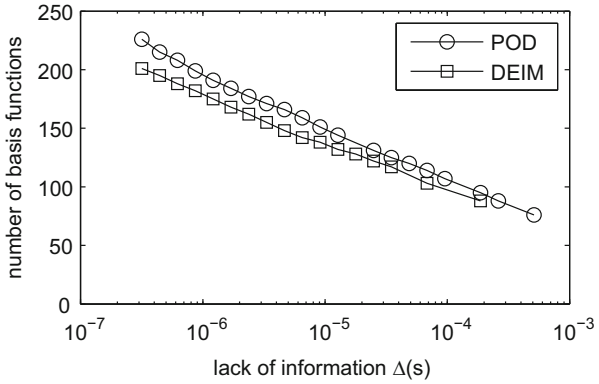

Fig. 1.9 The number of required POD basis function and DEIM interpolation indices grows only logarithmically with the requested information content

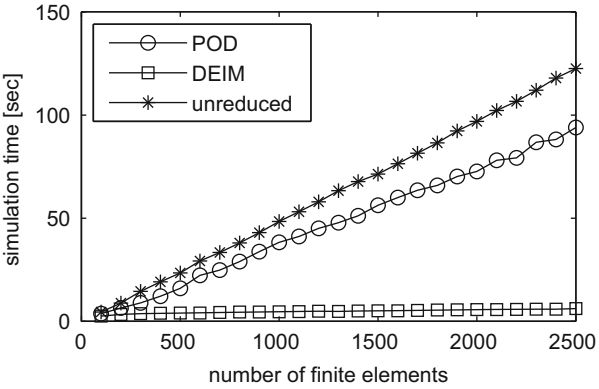

Fig. 1.10 Computation times of the unreduced and the reduced-order models plotted versus the number of finite elements

spatially two-dimensional semiconductors. It seems to be caused by the fact, that only a sufficiently large POD basis captured the physics of the semiconductors well enough.

In Fig. 1.9 we plot the corresponding total number of required POD basis functions. It can be seen that with the number of POD basis functions increasing linearly, the lack of information tends to zero exponentially. Furthermore, the number of DEIM interpolation indices behaves in the same way.

In Fig. 1.10 we investigate the dependence of the reduced models on the number of finite elements N. One sees that the simulation times of the unreduced model depends linearly on N. The POD reduced-order model still depends on N linearly with a smaller constant. The dependence on N of our POD-DEIM implementation is negligible.

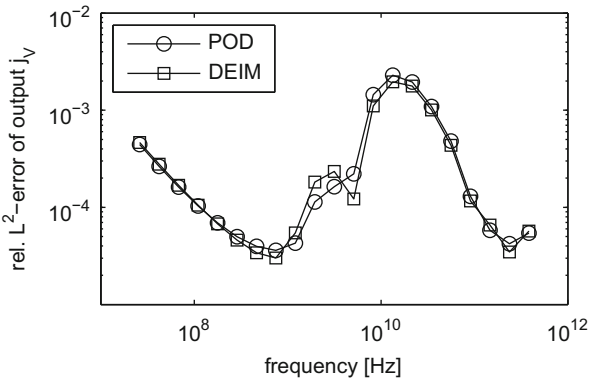

Fig. 1.11 The reduced models are compared with the unreduced model at various input frequencies

Finally, we analyze in Fig. 1.11 the behaviour of the models with respect to parameter changes. We consider the frequency of the sinusoidal input voltage as model parameter. The reduced-order models are created based on snapshots gathered in a full simulation at a frequency of $5 \cdot 10^9$ Hz. We see that the POD model and the POD-DEIM model behave very similarly. The adaptive enlargement of the POD basis using the residual greedy approach of [33] is discussed in the next section based on the results presented in [23].

Summarizing all numerical results we conclude that the significantly faster POD-DEIM reduction method yields a reduced-order model with the same qualitative behaviour as the reduced model obtained by classical POD MOR.

1.5 Residual-Based Sampling

Although POD model order reduction often works well, one has to keep in mind that the reduced system depends on the specific inputs and parameters used to generate the snapshots. A possible remedy consists in performing simulations over a certain input and/or parameter sample and then to collect all simulations in a global snapshot matrix $Y := [Y^1, Y^2, \ldots]$. Here, each Y^i represents the snapshots taken for a certain input resp. parameter.

In this section we propose a strategy to choose inputs/parameters in order to obtain a reduced model, which is valid over the whole input/parameter range. Possible parameters are physical constants of the semiconductors (e.g. length, permeability, doping) and parameters of the network elements (e.g. frequency of sinusoidal voltage sources, value of resistances). We do not distinguish between inputs and parameters of the model.

Let there be $r \in \mathbb{N}$ parameters and let the space of considered parameters be given as a bounded set $\mathscr{P} \subset \mathbb{R}^r$. We construct the reduced model based on snapshots from a simulation at a reference parameter $\omega_1 \in \mathscr{P}$. One expects that the reduced model approximates the unreduced model well in a small neighborhood of ω_1, but one cannot expect that the reduced model is valid over the complete parameter set $\mathscr{P}$. In order to create a suitable reduced-order model we consider additional snapshots which are obtained from simulations at parameters $\omega_2, \omega_3, \ldots \in \mathscr{P}$. The iterative selection of ω_{k+1} at a step k is called parameter sampling. Let P_k denote the set of selected reference parameters, $P_k := \{\omega_1, \omega_2, \ldots, \omega_k\} \subset \mathscr{P}$.

We neglect the discretization error of the finite element method and its influence on the coupled network and define the error of the reduced model as

$$\mathscr{E}(\omega; P) := z^h(\omega) - z^{POD}(\omega; P), \tag{1.45}$$

where $z^h(\omega) := (e^h(\omega), j_V^h(\omega), j_L^h(\omega), y^h(\omega))^\top$ is the solution of Problem 1.3.1 at the parameter ω with discretized semiconductor variables $y^h := (\psi^h, n^h, p^h, g_\psi^h, J_n^h, J_p^h)^\top$. $z^{POD}(\omega; P)$ denotes the solution of the coupled system in Problem 1.4.1 with reduced semiconductors, where the reduced model is created based on simulations at the reference parameters $P \subset \mathscr{P}$. The error is considered in the space X with norm

$$\begin{aligned}\|z\|_X := \Big\| \Big(& \|e\|_2, \|j_V\|_2, \|j_L\|_2, \\ & \|\psi\|_{L^2([0,T],L^2(\Omega))}, \|n\|_{L^2([0,T],L^2(\Omega))}, \|p\|_{L^2([0,T],L^2(\Omega))}, \\ & \|g_\psi\|_{L^2([0,T],H_{0,N}(\mathrm{div},\Omega))}, \\ & \|J_n\|_{L^2([0,T],H_{0,N}(\mathrm{div},\Omega))}, \|J_p\|_{L^2([0,T],H_{0,N}(\mathrm{div},\Omega))} \Big) \Big\|.\end{aligned}$$

Obvious extensions apply when there is more than one semiconductor present.

Furthermore we define the residual $\mathscr{R}$ by evaluation of the unreduced model (1.31)–(1.35) at the solution of the reduced model $z^{POD}(\omega; P)$, i.e.

$$\begin{aligned}\mathscr{R}(z^{POD}(\omega; P)) := & \begin{pmatrix} 0 \\ -M_L \dot{n}^{POD}(t) \\ M_L \dot{p}^{POD}(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{FEM} \begin{pmatrix} \psi^{POD}(t) \\ n^{POD}(t) \\ p^{POD}(t) \\ g_\psi^{POD}(t) \\ J_n^{POD}(t) \\ J_p^{POD}(t) \end{pmatrix} \\ & + \mathscr{F}(n^{POD}, p^{POD}, g_\psi^{POD}) - b(A_S^T e^{POD}(t)).\end{aligned} \tag{1.46}$$

Note that the residual of Eqs. (1.31)–(1.34) vanishes.

We note that the same definitions are used in [22] for linear descriptor systems. In [22] an error estimate is obtained by deriving a linear ODE for the error and exploiting explicit solution formulas. Here we have a nonlinear DAE and at the present state we are not able to provide an upper bound for the error $\|\mathscr{E}(\omega;P)\|_X$ which would yield a rigorous sampling method using for example the Greedy algorithm of [33].

We propose to consider the residual as an estimate for the error. The evaluation of the residual is cheap since it only requires the solution of the reduced system and its evaluation in the unreduced DAE. It is therefore possible to evaluate the residual at a large set of test parameters $P_{test} \subset \mathscr{P}$. Similar to the Greedy algorithm of [33], we add to the set of reference parameters the parameter where the residual becomes maximal.

The magnitude of the components in error and residual may be large and a proper scaling should be applied. For the error we consider the component-wise relative error, i.e.

$$\frac{\|\psi^h(\omega)-\psi^{POD}(\omega;P)\|_{L^2([0,T],L^2(\Omega))}}{\|\psi^h(\omega)\|_{L^2([0,T],L^2(\Omega))}},\ \frac{\|n^h(\omega)-n^{POD}(\omega;P)\|_{L^2([0,T],L^2(\Omega))}}{\|n^h(\omega)\|_{L^2([0,T],L^2(\Omega))}},\ \ldots,$$

and the residual is scaled by a block-diagonal matrix containing the weights

$$D(\omega)\mathscr{R}(z^{POD}(\omega;P)) = \begin{pmatrix} d_\psi(\omega)I & & & & & & \\ & d_n(\omega)I & & & & & \\ & & d_p(\omega)I & & & & \\ & & & d_{g_\psi}(\omega)I & & & \\ & & & & d_{J_n}(\omega)I & & \\ & & & & & d_{J_p}(\omega)I & \end{pmatrix} \times \mathscr{R}(z^{POD}(\omega;P)).$$

The weights $d_{(\cdot)}(\omega) > 0$ may be parameter-dependent. These weights are chosen in a way that the norm of the residual and the relative error are component-wise equal at the reference frequencies ω_k where we know $z^h(\omega_k)$ from simulation of the unreduced model, i.e.

$$d_\psi(\omega_k) := \frac{\|\psi^h(\omega_k)-\psi^{POD}(\omega_k;P)\|_{L^2([0,T],L^2(\Omega))}}{\|\psi^h(\omega_k)\|_{L^2([0,T],L^2(\Omega))}\cdot\|\mathscr{R}_1(z^{POD}(\omega_k;P))\|_{L^2([0,T],L^2(\Omega))}}, \tag{1.47}$$

and similarly for the other components. If $\|\mathscr{R}_1(z^{POD}(\omega_k;P))\|_{L^2([0,T],L^2(\Omega))}=0$ holds we chose $d_\psi(\omega_k):=1$.

In one dimensional parameter sampling with $\mathscr{P} := [\underline{p}, \overline{p}]$, we approximate $d_{(\cdot)}(\omega)$ by piecewise linear interpolation of the weights $d_{(\cdot)}(\omega_1)$, …, $d_{(\cdot)}(\omega_k)$. Extrapolation is done by nearest-neighbor interpolation to ensure the positivity of the weights.

Algorithm 1.1 Sampling

1. Select $\omega_1 \in \mathscr{P}$, $P_{test} \subset \mathscr{P}$, $tol > 0$, and set $k := 1$, $P_1 := \{\omega_1\}$.
2. Simulate the unreduced model at ω_1 and calculate the reduced model with POD basis functions U_1.
3. Calculate weight functions $d_{(\cdot)}(\omega) > 0$ according to (1.47) for all $\omega_k \in P_k$.
4. Calculate the scaled residual $\|D(\omega)\mathscr{R}(z^{POD}(\omega, P_k))\|$ for all $\omega \in P_{test}$.
5. Check termination conditions, e.g.
   - $\max_{\omega \in P_{test}} \|D(\omega)\mathscr{R}(z^{POD}(\omega, P_k))\| < tol$,
   - no progress in weighted residual.
6. Calculate $\omega_{k+1} := \arg\max_{\omega \in P_{test}} \|D(\omega)\mathscr{R}(z^{POD}(\omega, P_k))\|$.
7. Simulate the unreduced model at ω_{k+1} and create a new reduced model with POD basis U_{k+1} using also the already available information at $\omega_1, \ldots, \omega_k$.
8. Set $P_{k+1} := P_k \cup \{\omega_{k+1}\}$, $k := k + 1$ and goto 3.

We summarize our ideas in the sampling Algorithm 1.1. The step 7 in this algorithm can be executed in different ways. If offline time and offline memory requirements are not critical one may combine snapshots from all simulations of the full model and redo the model order reduction on the large snapshot ensemble. Otherwise we can create a new reduced model at reference frequency ω_{k+1} with POD-basis $\bar{U}$ and then perform an additional POD step on $(U_k, \bar{U})$.

1.5.1 Numerical Investigation for Residual Based Sampling

We now apply Algorithm 1.1 to provide a reduced-order model of the basic circuit and we choose the frequency of the input voltage v_s as model parameter. As parameter space we chose the interval $\mathscr{P} := [10^8,\ 10^{12}]$ Hz. We start the investigation with a reduced model which is created from the simulation of the full model at the reference frequency $\omega_1 := 10^{10}$ Hz. The number of POD basis functions s is chosen such that the lack of information $\Delta(s)$ is approximately 10^{-7}. The relative error and the weighted residual are plotted in Fig. 1.12 (left). We observe that the weighted residual is a rough estimate for the relative approximation error. Using Algorithm 1.1 the next additional reference frequency is $\omega_2 := 10^8$ Hz since it maximizes the weighted residual. The second reduced model is constructed on the same lack of information $\Delta := 10^{-7}$. Here we note that in practical applications, the error is not known over the whole parameter space.

The next two iterations of the sampling algorithm are also depicted in Fig. 1.12. Based on the residual in step 2, one selects $\omega_3 := 1.0608 \cdot 10^9$ Hz as the next reference frequency. Since no further progress of the weighted residual is achieved in step 3, the algorithm terminates. The maximal errors and residuals are given in Table 1.3.

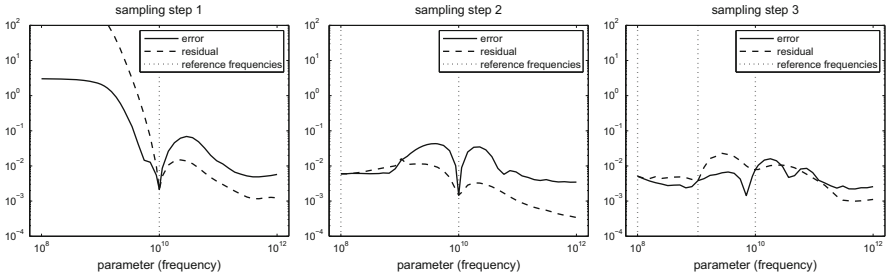

Fig. 1.12 *Left*: Relative reduction error (*solid line*) and weighted residual (*dashed line*) plotted over the frequency parameter space. The reduced model is created based on simulations at the reference frequency $\omega_1 := 10^{10}$ Hz, which is marked by *vertical dotted line*. *Middle*: Relative reduction error (*solid line*) and weighted residual (*dashed line*) plotted over the frequency parameter space. The reduced model is created based on simulations at the reference frequencies $\omega_1 := 10^{10}$ Hz and $\omega_2 := 10^8$ Hz. The reference frequencies are marked by *vertical dotted lines*. *Right*: Relative reduction error (*solid line*) and weighted residual (*dashed line*) plotted over the frequency parameter space. The reduced model is created based on simulations at the reference frequency $\omega_1 := 10^{10}$ Hz, $\omega_2 := 10^8$ Hz, and $\omega_3 := 1.0608 \cdot 10^9$ Hz. The reference frequencies are marked by *vertical dotted lines*

Table 1.3 Progress of refinement method

Step k	Reference parameters P_k	Max. scaled residual (at frequency)	Max. relative error (at frequency)
1	$\{1.0000 \cdot 10^{10}\}$	$9.9864 \cdot 10^2$ ($1.0000 \cdot 10^8$)	$3.2189 \cdot 10^0$ ($1.0000 \cdot 10^8$)
2	$\{1.0000 \cdot 10^8$, $1.0000 \cdot 10^{10}\}$	$1.5982 \cdot 10^{-2}$ ($1.0608 \cdot 10^9$)	$4.3567 \cdot 10^{-2}$ ($3.4551 \cdot 10^9$)
3	$\{1.0000 \cdot 10^8$, $1.0608 \cdot 10^9$, $1.0000 \cdot 10^{10}\}$	$2.2829 \cdot 10^{-2}$ ($2.7283 \cdot 10^9$)	$1.6225 \cdot 10^{-2}$ ($1.8047 \cdot 10^{10}$)

1.6 PABTEC Combined with POD MOR

In the current section, we combine the PABTEC approach of Chap. 2 and simulation based POD model order reduction techniques to determine reduced-order models for coupled circuit-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. Details of the approach are given in [27].

Now we return to the network equations (1.31)–(1.35). The coupling relation (1.34) can shortly be written as $j_S(t) = \vartheta(x_S(t))$, where $x_S(t) = \left[\psi^T(t), n^T(t), p^T(t), g_\psi^T(t), J_n^T(t), J_p^T(t)\right]^T$ is the state vector of the semidiscretized drift-diffusion equations (1.35). Determining the state $x_S(t)$ from Eq. (1.35) for a given voltage $A_S^T e(t)$, say $x_S(t) = \chi(A_S^T e(t))$, and substituting it into (1.34), we

obtain the relationship

$$j_S(t) = g(A_S^T e(t)), \tag{1.48}$$

where $g(A_S^T e(t)) := \vartheta(\chi(A_S^T e(t)))$ describes the voltage-current relation for the semidiscretized semiconductors. This relation can be considered as an input-to-output map, where the input is the voltage vector $A_S^T e(t)$ at the contacts of the semiconductors and the output is the approximate semiconductor current $j_S(t)$.

Electrical networks usually contains very large linear subnetworks modeling interconnects. In POD MOR we need to simulate the coupled DAE system (1.31)–(1.35) in order to determine the snapshots. To reduce the simulation time, we can first to separate the linear subsystem and approximate it by a reduced-order linear model of lower dimension using the PABTEC algorithm [38, 51], see also Chap. 2 in this book. The decoupled device equations are then reduced using the POD method presented in Sect. 1.4. Combining these reduced-order linear and nonlinear models, we obtain a nonlinear reduced-order model that approximates the coupled system (1.31)–(1.35).

1.6.1 Decoupling

For the extraction of a linear subcircuit, we use a decoupling procedure from [47] that consists in the replacement of the nonlinear inductors and nonlinear capacitors by controlled current sources and controlled voltage sources, respectively. The nonlinear resistors and semiconductor devices 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 loops consisting of capacitors and voltage sources (CV-loops) nor cutsets consisting of inductors and current sources (LI-cutsets) occur in the decoupled linear subcircuit meaning that its index coincides with the index of the original circuit, see [13] for the index analysis of the circuit equations. An advantage of the suggested replacement strategy is demonstrated in the following example.

Example 1.6.1 Consider a circuit with a semiconductor diode as in Fig. 1.13. We suggest to replace the diode by an equivalent circuit shown in Fig. 1.14. 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 of index one problems [50].

For simplicity, we assume that the circuit does not contain nonlinear devices other than semiconductors. Then after the replacements described above, the extracted

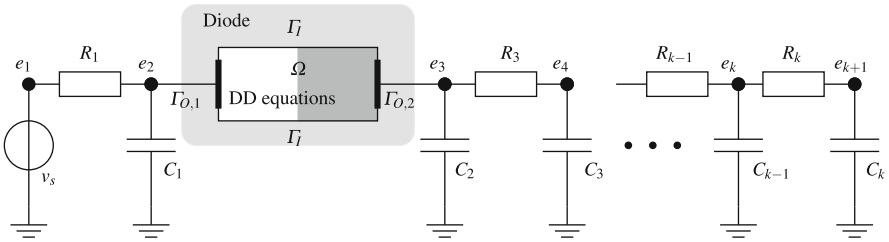

Fig. 1.13 RC chain with a diode

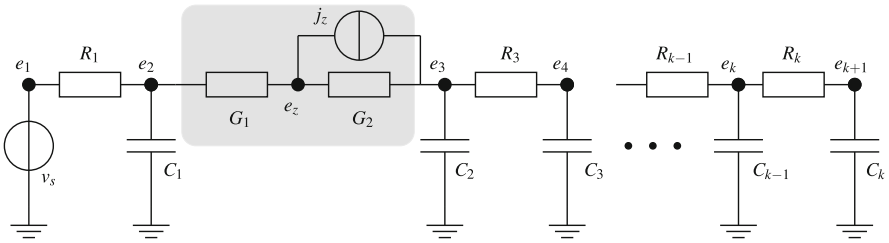

Fig. 1.14 Decoupled linear RC chain with a circuit replacing the diode

linear subcircuit can be modeled by the linear DAE system in the MNA form

$$E\dot{x}(t) = Ax(t) + Bu_l(t), \tag{1.49a}$$

$$y_l(t) = B^T x(t), \tag{1.49b}$$

with $x(t) = \left[\, e^T(t)\ \ e_z^T(t) \,\middle|\, j_L^T(t) \,\middle|\, j_V^T(t) \,\right]$, $u_l^T(t) = \left[\, i_s^T(t)\ j_z^T(t) \,\middle|\, v_s^T(t) \,\right]$ and

$$E = \begin{bmatrix} A_{C,l}CA_{C,l}^T & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} -A_{R,l}G_lA_{R,l}^T & -A_{L,l} & -A_{V,l} \\ A_{L,l}^T & 0 & 0 \\ A_{V,l}^T & 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} -A_{I,l} & 0 \\ 0 & 0 \\ 0 & -I \end{bmatrix}, \tag{1.49c}$$

where the incidence and element matrices are given by

$$A_{C,l} = \begin{bmatrix} A_C \\ 0 \end{bmatrix}, \quad A_{L,l} = \begin{bmatrix} A_L \\ 0 \end{bmatrix}, \quad A_{V,l} = \begin{bmatrix} A_V \\ 0 \end{bmatrix}, \quad A_{I,l} = \begin{bmatrix} A_I & A_S^2 \\ 0 & I \end{bmatrix}, \tag{1.49d}$$

$$A_{R,l} = \begin{bmatrix} A_R & A_S^1 & A_S^2 \\ 0 & -I & I \end{bmatrix}, \quad G_l = \begin{bmatrix} G & 0 & 0 \\ 0 & G_1 & 0 \\ 0 & 0 & G_2 \end{bmatrix}. \tag{1.49e}$$

Here, C, L and G are the capacitance, inductance and conductance matrices, A_S^1 and A_S^2 have entries in $\{0, 1\}$ and $\{-1, 0\}$, respectively, and satisfy $A_S^1 + A_S^2 = A_S$. Moreover, $e_z(t)$ is the potential of the introduced nodes, and the new input variable $j_z(t)$ is given by

$$j_z(t) = (G_1 + G_2)G_1^{-1} g(A_S^T e(t)) - G_2 A_S^T e(t), \tag{1.50}$$

where the matrices G_1 and G_2 are diagonal with conductances of the introduced linear resistors in the replacement circuits on the diagonal. One can show that the linear system (1.49) together with the decoupled nonlinear equations (1.35), (1.48) is state equivalent to the coupled system (1.31)–(1.35) together with the equation

$$e_z(t) = (G_1 + G_2)^{-1}\big(G_1(A_{\widetilde{\mathcal{R}}}^1)^T e(t) - G_2(A_{\widetilde{\mathcal{R}}}^2)^T e(t) - j_z(t)\big) \tag{1.51}$$

in the sense that these both systems have the same state vectors up to a permutation, see [47] for detail.

1.6.2 Model Reduction Approach

Applying the PABTEC method to the linear DAE system (1.49), we obtain a reduced-order model

$$\hat{E}\frac{d}{dt}\hat{x}(t) = \hat{A}\hat{x}(t) + \begin{bmatrix} \hat{B}_1 & \hat{B}_2 & \hat{B}_3 \end{bmatrix} \begin{bmatrix} i_s(t) \\ j_z(t) \\ v_s(t) \end{bmatrix}, \qquad \begin{bmatrix} \hat{y}_{l,1}(t) \\ \hat{y}_{l,2}(t) \\ \hat{y}_{l,3}(t) \end{bmatrix} = \begin{bmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \end{bmatrix} \hat{x}(t), \tag{1.52}$$

where $\hat{y}_{l,j} = \hat{C}_j\hat{x}(t)$, $j = 1, 2, 3$, approximate the corresponding components of the output y_l in (1.49b). Combining this reduced model with the semidiscretized drift-diffusion equations (1.35) via (1.48), we can determine the approximate snapshots which can then be used to compute the POD-reduced model as in (1.42). The coupling relation (1.41) can then be approximated by

$$\hat{j}_S(t) = C_1 U_{J_n} \gamma_{J_n}(t) + C_2 U_{J_p} \gamma_{J_p}(t) + C_3 U_{g_\psi} \dot{\gamma}_{g_\psi}(t). \tag{1.53}$$

As for the original system (1.34) and (1.35), we denote the relation between $A_S^T e(t)$ and $\hat{j}_S(t)$ by

$$\hat{j}_S(t) = \hat{g}(A_S^T e(t)). \tag{1.54}$$

Using (1.50) and (1.51), we have $-(A_S^2)^T e(t) - e_z(t) = -A_S^T e(t) + G_1 g(A_S^T e(t))$. Then it follows from $-(A_S^2)^T e(t) - e_z(t) \approx \hat{C}_2\hat{x}(t)$ that the semiconductor voltage

vector $u_S(t) = A_S^T e(t)$ can be approximated by $\hat{u}_S(t)$ satisfying $-G_1\hat{C}_2\hat{x}(t) - G_1\hat{u}_S(t) + \hat{g}(\hat{u}_S(t)) = 0$. Thus, combining the reduced linear system (1.52) with the reduced semiconductor model (1.42), we obtain a reduced-order coupled DAE system

$$\hat{E}\frac{d}{dt}\hat{x}(t) - (\hat{A} + \hat{B}_2(G_1 + G_2)\hat{C}_2)\hat{x}(t) - \hat{B}_2 G_1\hat{u}_S(t) - \hat{B}_1 i_s(t) - \hat{B}_3 v_s(t) = 0, \tag{1.55}$$

$$-G_1\hat{C}_2\hat{x}(t) - G_1\hat{u}_S(t) + \hat{g}(\hat{u}_S(t)) = 0, \tag{1.56}$$

$$\hat{j}_S(t) - C_1 U_{J_n}\gamma_{J_n}(t) - C_2 U_{J_p}\gamma_{J_p}(t) - C_3 U_{g_\psi}\dot{\gamma}_{g_\psi}(t) = 0, \tag{1.57}$$

$$\begin{pmatrix} 0 \\ -\dot{\gamma}_n(t) \\ \dot{\gamma}_p(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{POD} \begin{pmatrix} \gamma_\psi(t) \\ \gamma_n(t) \\ \gamma_p(t) \\ \gamma_{g_\psi}(t) \\ \gamma_{J_n}(t) \\ \gamma_{J_p}(t) \end{pmatrix} + U^\top \mathscr{F}(n^{POD}, p^{POD}, g_\psi^{POD}) - U^\top b(\hat{u}_S(t)) = 0. \tag{1.58}$$

Note that model reduction of the linear subsystem and the semiconductor model can be executed independently.

1.6.3 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, see Chap. 2. The POD method is implemented in C++ based on the FEM library deal.II [5] for discretizing the drift-diffusion equations. The obtained large and sparse nonlinear DAE system (1.31)–(1.35) as well as the small and dense reduced-order model (1.55)–(1.58) are integrated using the DASPK software package [9] based on a BDF method, where the nonlinear equations are solved using Newton's method. Furthermore, the direct sparse solver SuperLU [12] is employed for solving linear systems.

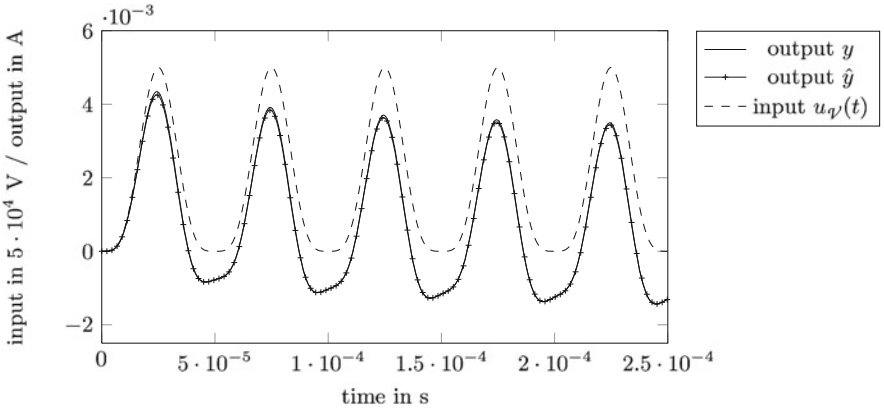

Fig. 1.15 Input voltage and output currents for the basic diode with the voltage-current relation (1.59)

Consider again an RC circuit with one diode as shown in Fig. 1.13. The input is given by

$$v_s(t) = u_V(t) = 10 \sin(2\pi f_0 t)^4$$

with the frequency $f_0 = 10^4$ Hz, see Fig. 1.15. The output of the system is $y(t) = -j_V(t)$. We simulate the models over the fixed time horizon $[0, \frac{2.5}{f_0}]$. The linear resistors have the same resistance $R = 2\,\mathrm{k}\Omega$ and the linear capacitors have the same capacitance $C = 0.02\,\mu\mathrm{F}$.

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

$$g(u_S) = 10^{-14}\,(\exp(40u_S) - 1)\,, \tag{1.59}$$

and apply only the PABTEC method to the decoupled linear system (1.49) that models the linear circuit given in Fig. 1.14. System (1.49) with $n_l = 1503$ variables was approximated by a reduced model (1.52) of dimension 24. The outputs y and $\hat{y}$ of the original nonlinear system (1.31)–(1.33), (1.48), (1.59) and the reduced-order nonlinear model (1.55), (1.56) with $\hat{g}$ replaced by g are plotted in Fig. 1.15. Simulation time and the absolute and relative L_2-norm errors in the output are presented in Table 1.4. 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 (1.17)–(1.22) for the diode. The parameters of the diode are summarized in Table 1.5. 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 1.6, 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 (1.31)–

Table 1.4 Simulation time and approximation errors for the nonlinear RC circuit with the basic diode described by the voltage-current relation (1.59)

System	Dimension	Simulation time (s)	Absolute error $\|y-\hat{y}\|_{L_2}$	Relative error $\|y-\hat{y}\|_{L_2}/\|y\|_{L_2}$
Unreduced	1503	0.584		
Reduced	24	0.054	$5.441 \cdot 10^{-7}$	$1.760 \cdot 10^{-2}$

Table 1.5 Diode model parameters

Parameter	Value
ε	$1.03545 \cdot 10^{-12}$ F/cm
U_T	0.0259 V
n_0	$1.4 \cdot 10^{10}$ 1/cm^3
μ_n	1350 cm^2/(V s)
τ_n	$330 \cdot 10^{-9}$ s
μ_p	480 cm^2/(V s)
τ_p	$33 \cdot 10^{-9}$ s
Ω	$[0, l_1] \times [0, l_2] \times [0, l_3]$
l_1 (length)	10^{-4} cm
l_2 (width)	10^{-5} cm
l_3 (depth)	10^{-5} cm
$N(\xi),\ \xi_1 < l_1/2$	$-9.94 \cdot 10^{15}$ 1/cm^3
$N(\xi),\ \xi_1 \geq l_1/2$	$4.06 \cdot 10^{18}$ 1/cm^3
FEM-mesh	500 elements, refined at $\xi_1 = l_1/2$

Table 1.6 Statistics for model reduction of the coupled circuit-device system

Network (MNA equations)	Diode (DD equations)	Dim.	Simul. time (s)	Jacobian evaluations	Absolute error $\|y-\hat{y}\|_{L_2}$	Relative error $\|y-\hat{y}\|_{L_2}/\|y\|_{L_2}$
Unreduced	Unreduced	7510	23.37	20		
Reduced	Unreduced	6031	16.90	17	$2.165 \cdot 10^{-8}$	$7.335 \cdot 10^{-4}$
Unreduced	Reduced	1609	1.51	16	$2.952 \cdot 10^{-6}$	$1.000 \cdot 10^{-1}$
Reduced	Reduced	130	1.19	11	$2.954 \cdot 10^{-6}$	$1.000 \cdot 10^{-1}$

(1.35) 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 (1.42) for the diode. The number of columns s_* of the projection matrices U_* is determined from the condition $\Delta_* \leq tol_{\text{POD}}$ with Δ_* defined in (1.36) and a tolerance $tol_{\text{POD}} = 10^{-6}$ 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 (1.42) for the diode is

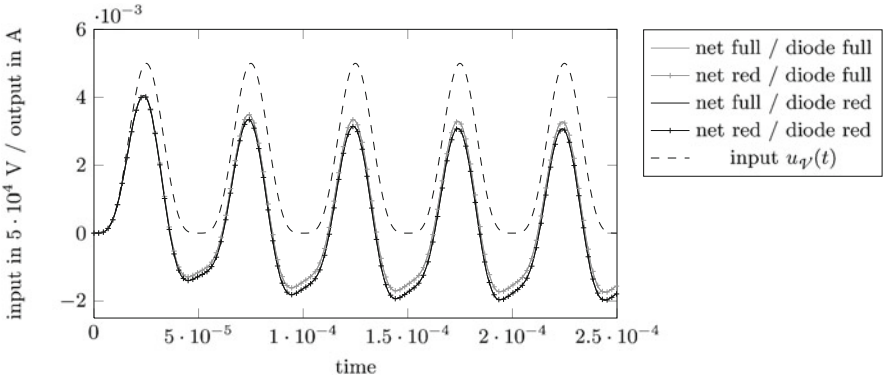

Fig. 1.16 Input voltage and output currents for the four model reduction setups

a dense DAE of dimension 105 while the original model (1.35) has dimension 6006, for the diode only. Coupling it with the unreduced and reduced linear networks, we obtain the results in Table 1.6 (last two rows). The simulation results for the different model reduction setups are also illustrated in Fig. 1.16.

The presented numerical results 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 section 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.

Acknowledgements The work reported in this paper was supported by the German Federal Ministry of Education and Research (BMBF), grant no. 03HIPAE5. Responsibility for the contents of this publication rests with the authors.

References

1. Anile, A.M, Mascali, G., Romano, V.: Mathematical Problems in Semiconductor Physics. Lecture Notes in Mathematics. Springer, Berlin (2003). Lectures given at the C. I. M. E. Summer School, Cetraro, Italy, 15–22 July 1998
2. Anile, A., Nikiforakis, N., Romano, V., Russo, G.: Discretization of semiconductor device problems. II. In: Schilders, W.H.A., et al. (eds.) Special Volume: Numerical Methods in Electromagnetics. Handbook of Numerical Analysis, vol. 13. Elsevier/North Holland, Amsterdam; Handb. Numer. Anal. **13**, 443–522 (2005)
3. Antoulas, A., Beattie, C., Gugercin, S.: Interpolatory model reduction of large-scale dynamical systems. In: Mohammadpour, J., Grigoriadis, K. (eds.) Efficient Modeling and Control of Large-Scale Systems, pp. 3–58. Springer, New York (2010)

4. Astrid, P., Weiland, S., Willcox, K., Backx, T.: Missing point estimation in models described by proper orthogonal decomposition. IEEE Trans. Autom. Control **53**(10), 2237–2251 (2008)
5. Bangerth, W., Hartmann, R., Kanschat, G.: deal.II – a general-purpose object-oriented finite element library. ACM Trans. Math. Softw. **33**(4), Article No. 24 (2007)
6. Benner, P., Hinze, M., ter Maten, E. (eds.): Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74. Springer, Berlin/Heidelberg (2011)
7. Bodestedt, M., Tischendorf, C.: PDAE models of integrated circuits and index analysis. Math. Comput. Model. Dyn. Syst. **13**(1), 1–17 (2007)
8. Brezzi, F., Marini, L., Micheletti, S., Pietra, P., Sacco, R., Wang, S.: Discretization of semiconductor device problems. I. In: Schilders, W., ter Maten, E. (eds.) Special Volume: Numerical Methods in Electromagnetics. Handbook of Numerical Analysis, vol. 13, pp. 317–441. Elsevier, Amsterdam (2005)
9. Brown, P., Hindmarsh, A., Petzold, A.: A description of DASPK: a solver for large-scale differential-algebraic systems. Lawrence Livermore National Report, UCRL (1992)
10. Chaturantabut, S., Sorensen, D.C.: Discrete empirical interpolation for nonlinear model reduction. Tech. Rep. 09-05, Department of Computational and Applied Mathematics, Rice University (2009)
11. Chaturantabut, S., Sorensen, D.: Nonlinear model reduction via discrete empirical interpolation. SIAM J. Sci. Comput. **32**(5), 2737–2764 (2010)
12. Demmel, J., Eisenstat, S., Gilbert, J., Li, X., Liu, J.: A supernodal approach to sparse partial pivoting. SIAM J. Matrix Anal. Appl. **20**(3), 720–755 (1999)
13. Estévez, S.D., Tischendorf, C.: Structural analysis for electric circuits and consequences for MNA. Int. J. Circuit Theory Appl. **28**, 131–162 (2000)
14. Estévez, S.D., Feldmann, U., März, R., Sturtzel, S., Tischendorf, C.: Finding beneficial DAE structures in circuit simulation. In: Jäger, W., et al. (eds.) Mathematics - Key Technology for the Future. Joint Projects Between Universities and Industry, pp. 413–428. Springer, Berlin (2003)
15. Freund, R.: SPRIM: structure-preserving reduced-order interconnect macromodeling. In: Technical Digest of the 2004 IEEE/ACM International Conference on Computer-Aided Design, pp. 80–87. IEEE Computer Society, Los Alamos, CA (2004)
16. Freund, R.: Structure-preserving model order reduction of RCL circuit equations. In: Schilders, W., van der Vorst, H.A., Rommes, J. (eds.) Model Order Reduction: Theory, Research Aspects and Applications. Mathematics in Industry, vol. 13, pp. 49–73. Springer, Berlin/Heidelberg (2008)
17. Günther, M.: Partielle differential-algebraische Systeme in der numerischen Zeitbereichsanalyse elektrischer Schaltungen. VDI Fortschritts-Berichte, Reihe 20, Rechnerunterstützte Verfahren, Nr. 343 (2001)
18. Günther, M., Feldmann, U., ter Maten, J.: Modelling and discretization of circuit problems. In: Schilders, W.H.A., et al. (eds.) Handbook of Numerical Analysis, vol. 13. Elsevier/North Holland, Amsterdam; Handb. Numer. Anal. **13**, 523–629 (2005)
19. Gugercin, S., Antoulas, A.: A survey of model reduction by balanced truncation and some new results. Int. J. Control **77**(8), 748–766 (2004)
20. Gugercin, S., Antoulas, A., Beattie, C.: $\mathscr{H}_2$ model reduction for large-scale linear dynamical systems. SIAM J. Matrix Anal. Appl. **30**(2), 609–638 (2008)
21. Gummel, H.: A self-consistent iterative scheme for one-dimensional steady state transistor calcula- tions. IEEE Trans. Electron Devices **11**(10), 455–465 (1964)
22. Haasdonk, B., Ohlberger, M.: Efficient reduced models and a-posteriori error estimation for parametrized dynamical systems by offline/online decomposition. SimTech Preprint 2009-23 (2009)
23. Hinze, M., Kunkel, M.: Residual based sampling in POD model order reduction of drift-diffusion equations in parametrized electrical networks. Z. Angew. Math. Mech. **92**, 91–104 (2012)
24. Hinze, M., Kunkel, M.: Discrete empirical interpolation in POD model order reduction of drift-diffusion equations in electrical networks. In: Scientific Computing in Electrical Engineering

SCEE 2010. Mathematics in Industry, vol. 16, Part 5, pp. 423–431. Springer, Berlin/Heidelberg (2012). doi:10.1007/978-3-642-22453-945
25. Hinze, M., Kunkel, M., Vierling, M.: POD model order reduction of drift-diffusion equations in electrical networks. In: Benner, P., Hinze, M., ter Maten, E. (eds.) Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, pp. 171–186. Springer, Berlin/Heidelberg (2011)
26. Hinze, M., Kunkel, M., Matthes, U.: POD model order reduction of electrical networks with semiconductors modeled by the transient drift–diffusion equations. In: Günther, M., Bartel, A., Brunk, M., Schöps, S., Striebel, M. (eds.) Progress in Industrial Mathematics at ECMI 2010. Mathematics in Industry, vol. 17, pp. 163–169. Springer, Berlin/Heidelberg (2012)
27. Hinze, M., Kunkel, M., Steinbrecher, A., Stykel, T.: Model order reduction of coupled circuit-device systems. Int. J. Numer. Model. **25**, 362–377 (2012)
28. Ho, C., Ruehli, A., Brennan, P.: The modified nodal approach to network analysis. IEEE Trans. Circuits Syst. **22**, 504–509 (1975)
29. Holmes, P., Lumley, J., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics. Cambridge University Press, Cambridge (1996)
30. Markowich, P.: The Stationary Semiconductor Device Equations. Computational Microelectronics. Springer, Wien/New York (1986)
31. Moore, B.: Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans. Autom. Control **26**(1), 17–32 (1981)
32. Odabasioglu, A., Celik, M., Pileggi, L.: PRIMA: passive reduced-order interconnect macromodeling algorithm. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **17**(8), 645–654 (1998)
33. Patera, A., Rozza, G.: Reduced Basis Approximation and a Posteriori Error Estimation for Parametrized Partial Differential Equations. MIT Pappalardo Graduate Monographs in Mechanical Engineering. Massachusetts Institute of Technology, Cambridge (2007)
34. Petzold, L.R.: A description of DASSL: a differential/algebraic system solver. IMACS Trans. Sci. Comput. **1**, 65–68 (1993)
35. Phillips, J., Daniel, L., Silveira, L.: Guaranteed passive balancing transformations for model order reduction. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **22**(8), 1027–1041 (2003)
36. Pinnau, R.: Model reduction via proper orthogonal decomposition. In: Schilders, W.H.A., et al. (eds.) Model Order Reduction: Theory, Research Aspects and Applications. Selected Papers Based on the Presentations at the Workshop 'Model Order Reduction, Coupled Problems and Optimization', Leiden, The Netherlands, 19–23 September 2005. Mathematics in Industry, vol. 13, pp. 95–109. Springer, Berlin (2008)
37. Reis, T., Stykel, T.: Positive real and bounded real balancing for model reduction of descriptor systems. Int. J. Control **83**(1), 74–88 (2010)
38. Reis, T., Stykel, T.: PABTEC: passivity-preserving balanced truncation for electrical circuits. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **29**(9), 1354–1367 (2010)
39. Reis, T., Stykel, T.: Lyapunov balancing for passivity-preserving model reduction of RC circuits. SIAM J. Appl. Dyn. Syst. **10**(1), 1–34 (2011)
40. Rewieński, M.: A Trajectory Piecewise-Linear Approach to Model Order Reduction of Nonlinear Dynamical Systems. Ph.D. thesis, Massachusetts Institute of Technology (2003)
41. Rowley, C.: Model reduction for fluids, using balanced proper orthogonal decomposition. Int. J. Bifur. Chaos Appl. Sci. Eng. **15**(3), 997–1013 (2005)
42. Schilders, W., van der Vorst, H., Rommes, J. (eds.): Model Order Reduction: Theory, Research Aspects and Applications. Mathematics in Industry, vol. 13. Springer, Berlin/Heidelberg (2008)
43. Selberherr, S.: Analysis and Simulation of Semiconductor Devices. Springer, Wien/New York (1984)
44. Selva, S.M.: A coupled system for electrical circuits. Numerical simulations. PAMM **6**(1), 51–54 (2006)

45. Sirovich, L.: Turbulence and the dynamics of coherent structures. I: coherent structures. II: symmetries and transformations. III: dynamics and scaling. Q. Appl. Math. **45**, 561–590 (1987)
46. Soto, M.S., Tischendorf, C.: Numerical analysis of DAEs from coupled circuit and semiconductor simulation. Appl. Numer. Math. **53**(2–4), 471–88 (2005)
47. Steinbrecher, A., Stykel, T.: Model order reduction of nonlinear circuit equations. Int. J. Circuit Theory Appl. **41**, 1226–1247 (2013). doi:10.1002/cta.1821
48. Striebel, M., Rommes, J.: Model order reduction of nonlinear systems in circuit simulation: status and applications. In: Benner, P., Hinze, M., ter Maten, E. (eds.) Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, pp. 279–292. Springer, Berlin/Heidelberg (2011)
49. Stykel, T.: Gramian-based model reduction for descriptor systems. Math. Control Signals Syst. **16**, 297–319 (2004)
50. Stykel, T.: Balancing-related model reduction of circuit equations using topological structure. In: Benner, P., Hinze, M., ter Maten, E. (eds.) Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, pp. 53–80. Springer, Berlin/Heidelberg (2011)
51. Stykel, T., Reis, T.: The PABTEC algorithm for passivity-preserving model reduction of circuit equations. In: Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems, MTNS 2010, Budapest, 5–9 July 2010, paper 363. ELTE, Budapest (2010)
52. Tischendorf, C.: Coupled Systems of Differential Algebraic and Partial Differential Equations in Circuit and Device Simulation. Habilitation thesis, Humboldt-Universität Berlin (2004)
53. Verhoeven, A., ter Maten, E., Striebel, M., Mattheij, R.: Model order reduction for nonlinear IC models. In: Korytowski, A., Malanowski, K., Mitkowski, W., Szymkat, M. (eds.) System Modeling and Optimization. IFIP Advances in Information and Communication Technology, vol. 312, pp. 476–491. Springer, Berlin/Heidelberg (2009)
54. Willcox, K., Peraire, J.: Balanced model reduction via the proper orthogonal decomposition. AIAA J. **40**(11), 2323–2330 (2002)