Model Order Reduction of Electrical Networks with Semiconductors modelled by Quantum-Drift-Diffusion Models

Michael Hinze and Ulrich Matthes

[1] Michael Hinze, Department of Mathematics, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany,
`michael.hinze@uni-hamburg.de`
[2] Ulrich Matthes, Department of Mathematics, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany,
`ulrich.matthes@math.uni-hamburg.de`

Abstract. In this paper we consider model order reduction (MOR) techniques for electrical networks with semiconductors modelled by the drift-diffusion and the quantum-drift-diffusion equations. We compare interpolation with parametric proper orthogonal decomposition (POD)-MOR.

Keywords: Model Order Reduction, Parametrized Dynamical Systems, Quantum Drift-Diffusion Equations, Integrated Circuits, POD-DEIM

AMS subject classifications: 93A30, 65B99, 65M60, 65M20

1 Introduction

Electrical networks with complex components like semiconductors can be reduced using MOR methods. In this paper we investigate the POD approach with the discrete empirical interpolation method (DEIM) for the reduction of the nonlinearities, see [7]. We emphasize that our MOR approach is not restricted to electrical networks with semiconductors but also extends to networks containing many simple components, and complex components modeled by PDE systems (see Figure 1) if the network allows modeling with modified nodal analysis (MNA).

For small length of the semiconductor quantum effects may become important. So we compare the drift-diffusion (DD) and quantum-drift-diffusion (QDD) equations for semiconductors. We find that for larger semiconductor length the difference of the DD and QDD model results is negligible, but the quantum term is more challenging for the POD-MOR approach, see Section 3. For smaller semiconductor length, where the DD and QDD model differ more, both are inaccurate by a large factor and energy transport in the semiconductor should be considered, see e.g. [11].
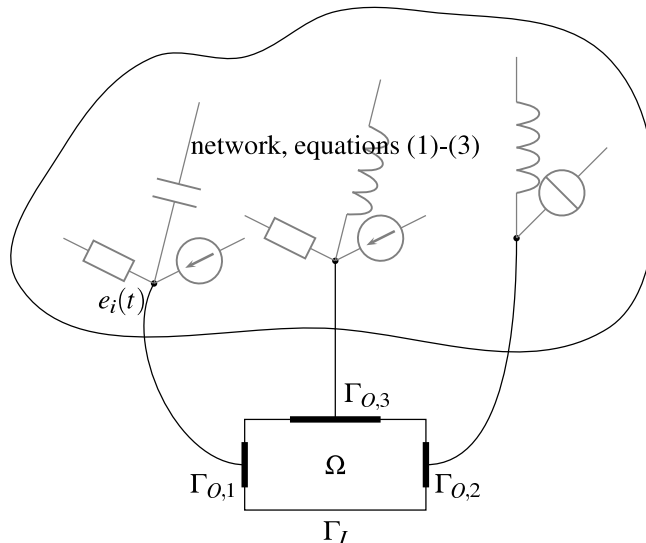
Fig. 1. Sketch of a network with many simple components and a complex component representing a semiconductor.

Second we investigate how complex components of networks can be reduced. It is necessary to use simulation based MOR approaches for parametric MOR? Or can the full model results used for the snapshots there simply interpolated over the parameter space, without the need of a reduced order model? In Section 6 we compare both methods for frequency as parameter and find that only if the solution does not depend much on the parameter, at low frequencies, the interpolation approach is comparative. In all other cases the POD-MOR approach outperforms the interpolation approach clearly in accuracy for two full simulations only, even when a very large number of expensive full simulations points in the (1-dim) parameter space are used.

In Section 2 we describe the mathematical model for the electrical network with complex components and the QDD-model for semiconductors, and in Section 3 we present some simulation results. In Section 4 we briefly summarize the implementation of the POD-DEIM method. In Section 5 we describe our interpolation method and in Section 6 compare it with parametric POD-DEIM for a simple test example. Finally we present some conclusions.

2   Modeling of the electrical network with semiconductors and Quantum-Drift-Diffusion equations

We now describe the mathematical model for electrical networks with many simple components like resistors, capacitors, and inductors and complex components

like semiconductors modeled by DD or QDD equations. First the network containing only the simple components is modeled by a differential algebraic equation (DAE) system which is obtained by a modified nodal analysis (MNA) [10], including the Ohmic contacts $\Gamma_{O,k}, k = 1, \ldots, N_c$ of the semiconductors as network nodes. Denoting by e the node potentials and by j_L, j_V, and j_S the currents of inductive, voltage source, and semiconductor branches, the DAE reads (see, e.g. [6, 10, 18])

$$A_C \frac{d}{dt} q_C(A_C^\top e, t) + A_R g(A_R^\top e, t) + A_L j_L + A_V j_V + A_S j_S = -A_I i_s(t), \quad (2.1)$$

$$\frac{d}{dt} \phi_L(j_L, t) - A_L^\top e = 0, \quad (2.2)$$

$$A_V^\top e = v_s(t). \quad (2.3)$$

Here, the incidence matrix $A = [A_R, A_C, A_L, A_V, A_S, A_I] = (a_{ij})$ 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 $a_{ij} = 0$ else. The indices R, C, L, V, S, I denote the capacitive, resistive, inductive, voltage source, semiconductor, and current source branches, respectively. In particular the matrix A_S denotes the semiconductor incidence matrix. The vector valued functions q_C, g and ϕ_L are continuously differentiable defining the voltage-current relations of the network components. The continuous vector valued functions v_s and i_s are the voltage and current sources. For details we refer to [8].

In a second step the semiconductors are modeled by PDE systems, which are then coupled to the DAE of the network via the nodes related to the Ohmic contacts. Here we first use the transient drift-diffusion equations as a continuous model for semiconductors, see e.g. [1, 2] and the references cited there. We use the notation and scaling introduced there. For small semiconductors quantum effects have to be considered to improve the semiconductor model. This leads to the QDD-equations f.e. [11, Ch.12] or [14]. We use the quantum correction form of the DD-equations with holes.

So we obtain the following scaled system of PDEs 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 $J_p(t, x)$:

$$\lambda \Delta \psi = n - p - C, \quad (2.4)$$

$$-\partial_t n + \nu_n \operatorname{div} J_n = R(n, p), \quad (2.5)$$

$$\partial_t p + \nu_p \operatorname{div} J_p = -R(n, p), \quad (2.6)$$

$$J_n = \nabla n - n \nabla \psi - \epsilon^2 n \nabla \left(\frac{\Delta \sqrt{n}}{\sqrt{n}} \right), \quad (2.7)$$

$$J_p = -\nabla p - p \nabla \psi. \quad (2.8)$$

Here $(t, x) \in [0, T] \times \Omega$ and $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$. The nonlinear function R describes the rate of electron/hole recombination, where 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. Here, τ_n and τ_p are the average lifetimes of electrons and holes, and n_i is the constant intrinsic concentration which satisfies $n_i^2 = np$ if the semiconductor is in thermal equilibrium. The scalar $\lambda > 0$ is the scaled Debye length, and ν_n and ν_p are the scaled 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.

The coefficent

$$\epsilon^2 = \frac{\hbar^2}{2mk_B T_0 L^2}$$

is called squared Planck constant and grows in importance with shrinking length L of the semiconductor. The QDD model (2.4)-(2.8) only differs from the DD equations through the Bohm potential

$$-\epsilon^2 n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}} \right).$$

For a derivation of this quantum correction of the DD system we refer to [11, Ch.12] and [14].

This system is supplemented with the boundary conditions

$$\psi(t, x) = \psi_{bi}(x) + (A_S^\top e(t))_k = U_T \log\left(\frac{\sqrt{C(x)^2 + 4n_i^2} + C(x)}{2n_i} \right) + (A_S^\top e(t))_k, \quad (2.9)$$

$$n(t, x) = \frac{1}{2}\left(\sqrt{C(x)^2 + 4n_i^2} + C(x) \right), \quad p(t, x) = \frac{1}{2}\left(\sqrt{C(x)^2 + 4n_i^2} - C(x) \right), \quad (2.10)$$

for $(t, x) \in [0, T] \times \Gamma_{O,k}$, where the potential of the nodes which are connected to a semiconductor interface enter in the boundary conditions for ψ. 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. The semiconductor model (2.4)-(2.8) is coupled to the network through the semiconductor current vector j_S with the components
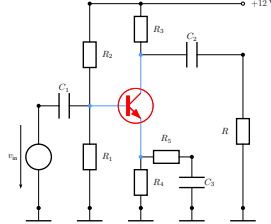
$$j_{S,k} = \int_{\Gamma_{O,k}} (J_n + J_p - \varepsilon\partial_t\nabla\psi) \cdot \nu \, d\sigma, \quad k = 1, \dots, N_c, \quad (2.11)$$

where ν denotes the unit outward normal to the interface $\Gamma_{O,k}$. Further details are given in [8]. Contributions to the analytical and numerical analysis of PDAE systems of the presented form can be found in [2, 5, 17, 18].

For the analytical treatment of the DD and QDD systems we refer the reader to [4, 11]. From the numerical point of view the denominator $\sqrt{n}$ in the Bohm

potential is critical, it requires positivity preserving schemes for the time integration of (2.4)-(2.8). The space discretization is done by mixed Raviart-Thomas finite elements for vector valued functions and piecewise constant elements for scalar valued functions. The time integration is performed with DASPK [3].

The discretized model reads

$$A_C \frac{\mathrm{d}q_C}{\mathrm{d}t}\left(A_C^\top e(t), t\right) + A_R g\left(A_R^\top e(t), t\right),$$

$$+ A_L j_L(t) + A_V j_V(t) + A_S j_S(t) = -A_I i_s(t),$$

$$\frac{\mathrm{d}\phi_L}{\mathrm{d}t}\left(j_L(t), t\right) - A_L^\top e(t) = 0,$$

$$A_V^\top e(t) = v_s(t),$$

$$j_S(t) - C_1 J_n(t) - C_2 J_p(t) - C_3 \dot{g}_\psi(t) = 0,$$

$$\begin{pmatrix} 0 \\ -M_L \dot{n}(t) \\ M_L \dot{p}(t) \\ 0 \\ 0 \\ 0 \end{pmatrix} + A_{FEM} \begin{pmatrix} \psi(t) \\ n(t) \\ p(t) \\ g_\psi(t) \\ J_n(t) \\ J_p(t) \end{pmatrix} + \mathcal{F}(n^h, p^h, g_\psi^h) - b(A_S^\top e(t)) = 0.$$

The POD model order reduction of the semiconductors in this model is now
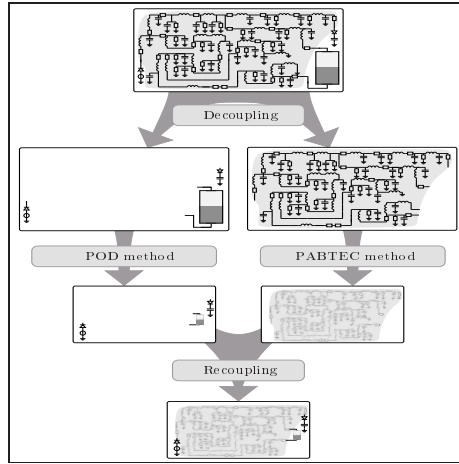

Fig. 2. PABTEC for MOR of linear networks combined with MOR for the complex components from [9].

done as in [7]. Let us recall that for MOR of the overall network one may combine the passitivity preserving balanced truncation method of electrical circuits

(PABTEC) [15, 16] with the MOR approach of e.q. [7] for the semiconductors, see the sketch in Fig. 2. For details we refer to [9].

3   Numerical Results
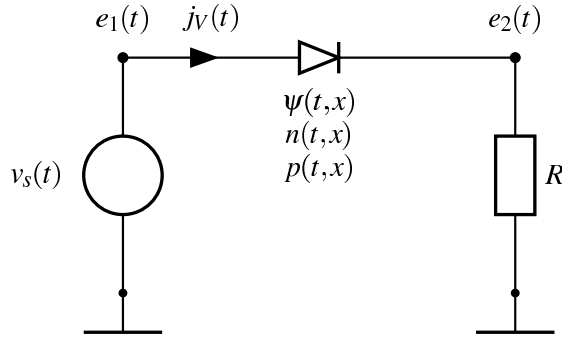
Fig. 3. Basic circuit with one diode.

In this section we present a comparison of simulation results for the DD and the QDD equations in a simple network, see Figure 3, as described in [8]. As a solver for the full and the reduced DAE systems we use DASPK, see [3, 13].

We start with the simulation of the full model at the frequency $\omega := 10^{10} \ Hz$. The number of POD basis functions s in the examples is chosen such that the lack of information content

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

ranges between 10^{-4} and 10^{-8}.

In Figure 4 we show the evolution of the currents through a 1000 nm diode ($\epsilon^2 = 9e - 5$) modeled by the QDD- and DD-equation respectively. As one can see the differences are negligible.

The results for a 100 nm diode ($\epsilon^2 = 9e - 3$) are depicted in Figure 5. In this case the difference in the currents is still small, see Figure 5(right).

We observe that the influence of the quantum term only increases slightly with decreasing structure size in the considered range from 1000 to 100 nm. To include quantum effects for these sizes more sophisticated semiconductor models should be considered.

4   Parametric MOR

To obtain reduced order models which are valid over a certain parameter range we apply parametric MOR (PMOR) with greedy sampling, see [12]. The param-
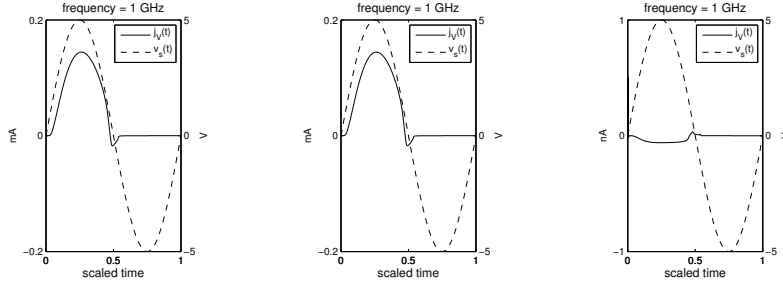
Fig. 4. Current and voltage: QDD (left) versus DD (middle), and current difference QDD - DD (in nA)(right), diode length 1000 nm, $\epsilon^2 = 9e - 5$.
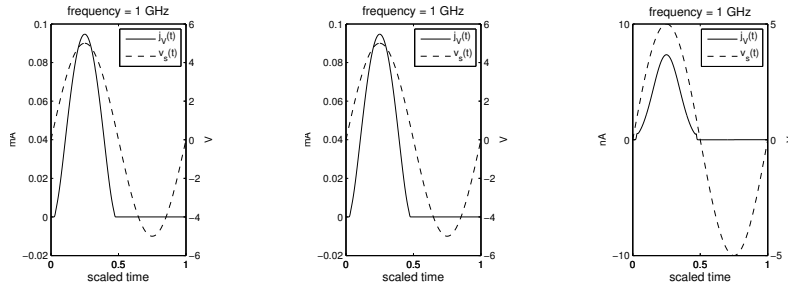
Fig. 5. Current and voltage: QDD (left) versus DD (middle), and current difference QDD - DD (in nA)(right), diode length 100 nm, $\epsilon^2 = 9e - 3$.

eter in our application is given by the frequecy which ranges is the parameter interval $\mathcal{P} := [f_l, f_u]$. The sampling procedure is summarized in Algorithm 1. For details we refer to e.g. [8].

Algorithm 1 (Sampling)

1. *Select $\omega_1 \in \mathcal{P}$, $P_{test} \subset \mathcal{P}$, tol > 0, and set $k := 1$, $P_1 := \{\omega_1\}$. Simulate the unreduced model at ω_1 and calculate the reduced model with POD basis functions U^1.*
2. *Calculate the residual $\|\mathcal{R}(z^{POD}(\omega, P_k))\|$ for all $\omega \in P_{test}$.*
3. *Check termination conditions, e.g.*
   - *$\max_{\omega \in P_{test}} \|\mathcal{R}(z^{POD}(\omega, P_k))\| < tol$, or*
   - *no further reduction of residual, then STOP.*
4. *Calculate $\omega_{k+1} := \arg\max_{\omega \in P_{test}} \|D(\omega)\mathcal{R}(z^{POD}(\omega, P_k))\|$.*
5. *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 $\omega_1, \ldots, \omega_k$.*
6. *Set $P_{k+1} := P_k \cup \{\omega_{k+1}\}$, $k := k + 1$ and goto 3.*

Here, $D(\omega)$ denotes a scaling matrix tailored to the dimension of the components appearing in the residual vector $\mathcal{R}$.

5   Interpolation

A further possibility for obtaining a parametric reduced order model is interpolation over the frequency range.

In the following the interpolation is applied componentwise on the variables $\psi, n, p, g_\psi, J_n, J_p$ in the semiconductor, the current j, and the electric field e of the network. Here $g_\psi = \nabla \psi$. The interpolation is applied on snapshots taken at the same phase $\phi_i = f_k t_{i,k}$ for different parameters.

The interpolation is performed on the logarithm of the parameter frequency. The logarithm of the frequency is used since relative changes in the frequency are a better measure for the change of the variables $\psi, n, p, g_\psi, J_n, J_p, j, e$ with respect to the frequency than the absolute frequency difference. The time points are multiplied by the frequency to ensure that the interpolation is performed at the same phase.

The interpolation is now described in detail for the variable n. The other parameters can be treated analoguosly.

Let the parameters f_k be the frequencies with $k = 1, \ldots, m$ the index of the interpolation points. $n_{\log f_k}(t_{i,k})$ denotes the i-th snapshot of the electron density n for the parameter $\log f_k$ at time $t_{i,k}$ with phase $\phi_i = f_k t_{i,k}$.

Now we interpolate the electron density according to

$$n_{\log f}^{\text{inter}}(\phi_i) = \text{spline}_{\log f_k, k=1,\ldots,m}\{n_{\log f_1}(\phi_i), \ldots, n_{\log f_m}(\phi_i)\}$$

with a piecewise cubic or linear spline at each time point $t_i = \phi_i/f$.

Interpolation for the other variables $\psi, p, g_\psi, J_n, J_p, j, e$ is performed similarly.

The total relative error of the interpolation is

$$\text{err}^2 = \eta_\psi^2 + \eta_n^2 + \eta_p^2 + \eta_{g_\psi}^2 + \eta_{J_n}^2 + \eta_{J_p}^2 + \eta_j^2 + \eta_e^2, \qquad (5.12)$$

where e.g.

$$\eta_n = \frac{\|n_{\log f}^{\text{inter}}(\phi_i) - n_{\log f}(\phi_i)\|_{L^2}}{\|n_{\log f}(\phi_i)\|_{L^2}}.$$

The error of the reduced model is calculated in the same way.

6   Interpolation vs. POD-MOR

In this section we compare the results of the interpolation approach with these obtained by MOR. We use an one-dimensional parameter space. The parameter is the logarithm of the frequency. As a solver for the full and reduced DAE we use DASPK, see [3, 13].

As an example we again take the circuit with one diode depicted in Figure 3.

We choose the frequency of the input voltage v_s as model parameter with parameter space $\mathcal{P} := [10^8, \ 10^{12}] \ Hz$. We initialize PMOR with a reduced model constructed from the simulation of the full model at the reference frequency $\omega_1 := 10^{10} \ Hz$. In the case of interpolation we start with functions that are

constant with respect to the parameter. The number of POD basis functions s is chosen such that the lack of information content $\Delta(s)$ is approximately 10^{-7}. The relative error and the residual for the PMOR approach are plotted in Figure 8. We observe that the residual admits a structure similar to that of the approximation error. Using Algorithm 1 the next chosen frequency is $\omega_2 := 10^8 \; Hz$ since it maximizes the residual. For the selection of the interpolation points we can proceed similarly.

frequency	average error interpolation				POD-DEIM		
Hz	5	10	20	40	1	2	3
	-	-	-	-	5	10	15
1e8-1e9	0.2285	0.0676	0.0324	0.0163	0.2640	0.0159	0.0109
1e9-1e10	0.4028	0.2288	0.1210	0.0717	0.0980	0.0095	0.0058
1e10-1e11	0.6058	0.2632	0.1178	0.0516	0.0158	0.0066	0.0085
1e11-1e12	0.9667	0.3225	0.1213	0.0354	0.0183	0.0066	0.0057
1e8-1e12	0.5617	0.2167	0.0961	0.0425	0.0990	0.0097	0.0077

Table 1. Errors: Interpolation with 5, 10, 20, and 40 (log-equidistant) interpolation nodes (each obtained by a full simulation) vs. POD-DEIM with 1, 2, or 3 full simulations and 5, 10, or 15 reduced simulations

In Table 1 we show the average relative errors of the interpolation for different frequency ranges from 1e8-1e9 to 1e11-1e12. This is done for interpolation of 5, 10, 20, and 40 nodes over the whole parameter range $\mathcal{P}$. In the last columns we show the errors of the reduced model from PMOR for comparison. In the case of interpolation the errors are small at low frequencies $\leq 10^9 \; Hz$, in particular for a large number of interpolation points. For high frequencies the error in the interpolated variables are large, even for a large number of interpolation points. The PMOR approach produces only small errors (especially for medium and high frequencies) even for a low number of full simulations. This can be explained with the good approximation properties of the reduced POD model. If the reduced POD model can be solved sufficently fast, then the two full simulations plus two PODs to generate two reduced models plus 10 (or even 100) reduced simulations (second last column in table 1) are faster than 20 or even 10 full simulations (forth and third column).

In Figure 6 (upper-left) the relative error of the interpolation is shown as calculated by formula (5.12). The errors are larger as for MOR for the same reference frequencies which are shown as thin lines. The weighted residuals (upper-right) are large. In the lower part we see the errors and residuals after one and two sampling steps.
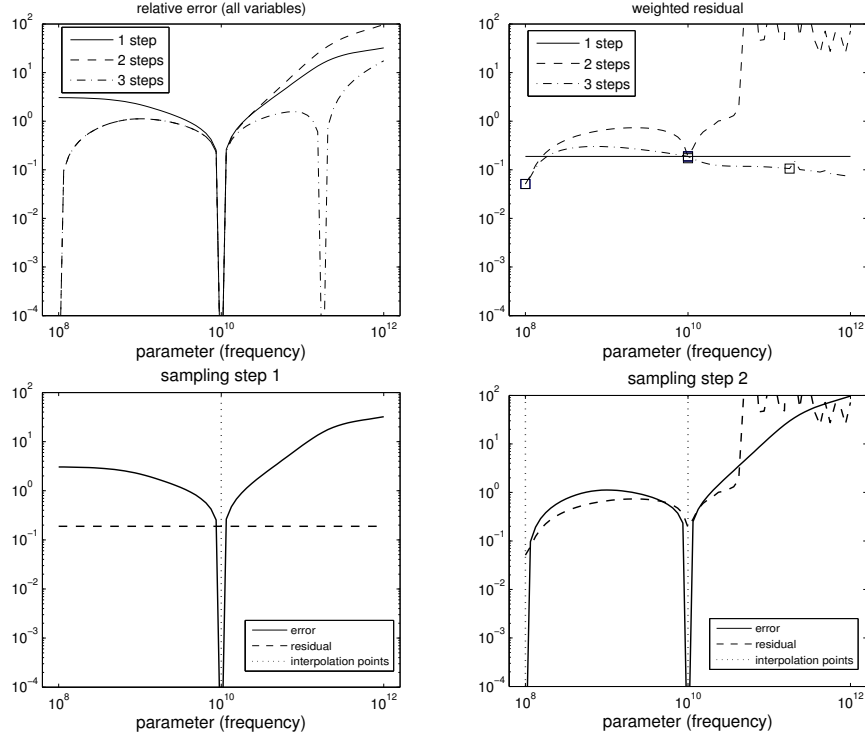
Fig. 6. Interpolation, residual based sampling, greedy approach: relative error (upper-left) weighted residuals (upper-right) for 1 (solid), 2 (dashed), or 3 (dash-doted) sampling steps and errors and residuals after one and two sampling steps (lower part).

In Figure 7 we show the results for log-equidistant interpolation at 5, 10, 20, or 40 interpolation points compared to PMOR with one full simulation at $\omega = 1e10$ Hz. Even for a number of 40 interpolation points the interpolation error is larger for medium and high frequencies.

In Figure 8 (upper-left) the relative error of the PMOR is shown as calculated by (5.12). The errors and residuals are substantial smaller for POD-MOR than for the interpolation approach, compare Figure 6. In the lower part we show the errors and residuals for PMOR after one and two sampling steps.

7   Conclusions

For the considered size of the semiconductors (in the range of 100 to 1000nm), the difference in simulation and MOR for the DD and QDD equations is negligible. We compare a PMOR approach related to frequency as parameter based on residual based greedy sampling with snapshot interpolation. Interpolation works well for lower frequencies, but is clearly outperformed by PMOR in the approximation related to hight frequencies.
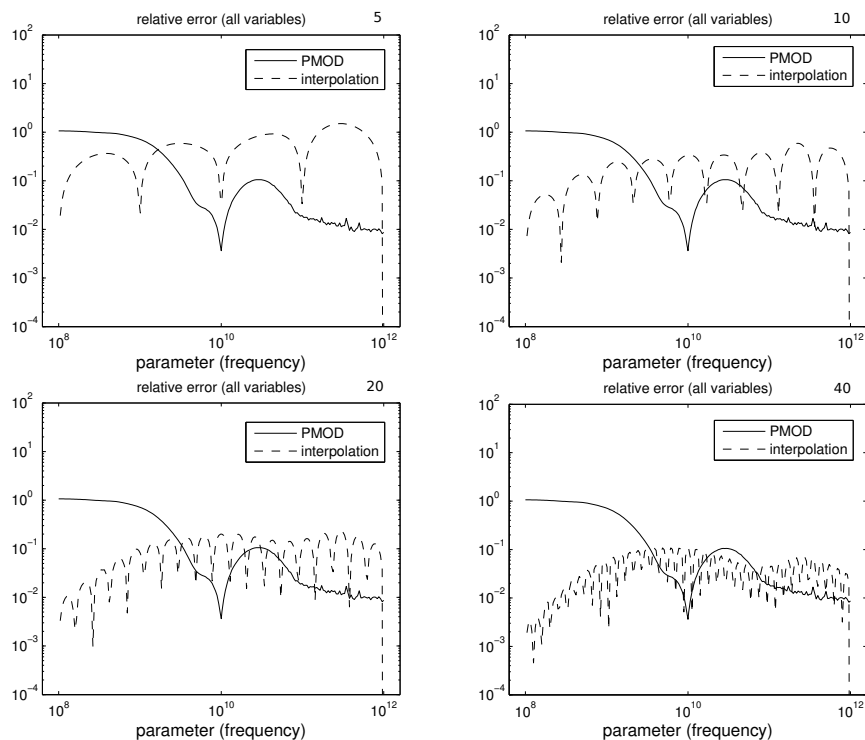
Fig. 7. Interpolation log-equidistant, interpolation error with 5, 10, 20, or 40 interpolation points (dashed); PMOR with one full simulation at $\omega = 1e10$ Hz for comparison (solid).

References

1. Anile, A., Mascali, G., Romano, V.: Mathematical problems in semiconductor physics. Lectures given at the C. I. M. E. summer school, Cetraro, Italy, July 15–22, 1998. Lecture Notes in Mathematics. Berlin: Springer (2003)
2. Bodestedt, M., Tischendorf, C.: PDAE models of integrated circuits and index analysis. Math. Comput. Model. Dyn. Syst. 13:(1), 1–17 (2007)
3. Brown, P., Hindmarsh, A., Petzold, A.: A description of DASPK: A solver for large-scale differential-algebraic systems. Tech. rep., *Lawrence Livermore National Report* UCRL (1992)
4. Gajewski, H.: On existence, uniqueness and asymptotic behavior of solutions of the basic equitions for carrier transport in semiconductors, ZAMM 65:2, 101-108 (1985)
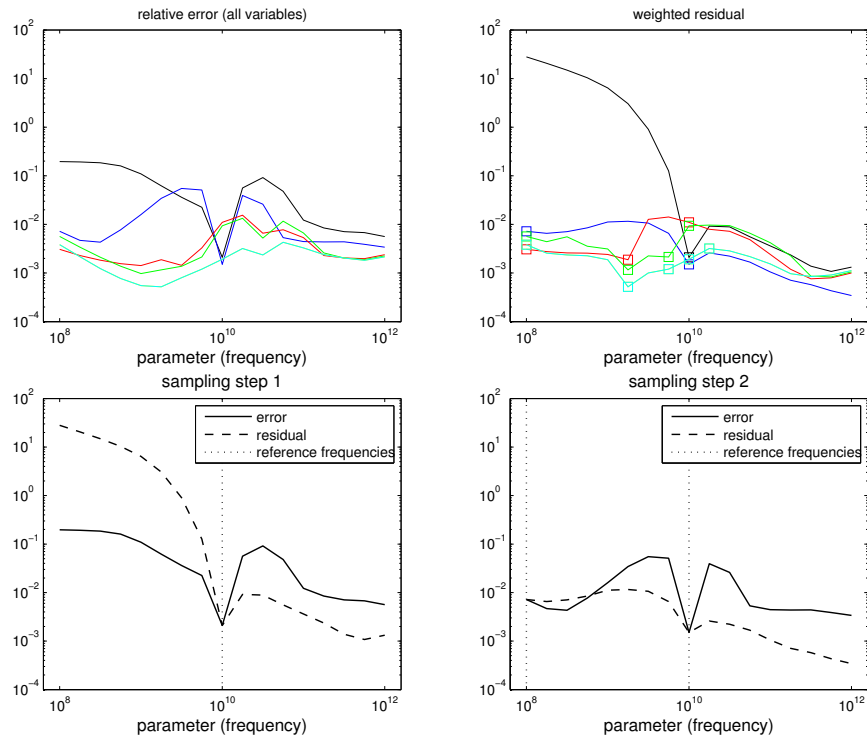
Fig. 8. MOR, residual based sampling, greedy approach: relative error (upper-left) weighted residuals (upper-right) and error and residuals after one and two sampling steps (lower part) (black, blue, red, green, and cyan corresponding to the sampling steps 1 to 5).

5. Günther, M.: Partielle differential-algebraische Systeme in der numerischen Zeit-bereichsanalyse elektrischer Schaltungen. VDI Fortschritts-Berichte, Reihe 20, Rechnerunterstützte Verfahren, Nr. 343 (2001)
6. Günther, M., Feldmann, U., ter Maten, J.: Modelling and discretization of circuit problems. Schilders, W. H. A. (ed.) et al., Handbook of numerical analysis. Vol XIII. Special volume: Numerical methods in electromagnetics. Amsterdam: Elsevier/North Holland. Handbook of Numerical Analysis 13, 523-629 (2005)
7. Hinze, M., Kunkel, M.: Discrete empirical interpolation in pod model order reduction of drift-diffusion equations in electrical networks. SCEE Proceedings 2010, Toulouse (2010)
8. Hinze, M., Kunkel, M.: Residual based sampling in POD model order reduction of drift-diffusion equations in parametrized electrical networks. ZAMM 92:91-104 (2011)
9. Hinze, M., Kunkel, M., Steinbrecher, A., Stykel, T.: Model order reduction of coupled circuit-device systems. Int. J. Numer. Model. 25:362-377 (2012)
10. Ho, C., Ruehli, A., Brennan, P.: The modified nodal approach to network analysis. IEEE Trans. Circuits Syst. 22, 504–509 (1975)

11. Jüngel, A.: Leture Notes in Physics: Transport Equations for Semiconductors, Springer (2009)
12. Patera, A., Rozza, G.: Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations. Version 1.0. Copyright MIT 2006–2007, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering. (2007)
13. Petzold, L.R.: A description of DASSL: A differential/algebraic system solver. IMACS Trans. Scientific Computing 1, 65–68 (1993)
14. Pinnau, R.: The Linearized Transient Quantum Drift Diffusion Model, ZAMM 80:327-344 (2000)
15. Reis T., Stykel T.: PABTEC: Passivity-Preserving Balanced Truncation for Electrical Circuits, IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems , vol. 29, no. 9, pp. 1354-1367, (2010)
16. Salih, H., Steinbrecher, A., Stykel, T.: MATLAB Toolbox PABTEC - A users guide. *Technical Report.* Tech. rep., Institut für Mathematik, Technische Universität Berlin, Germany (2011)
17. Selva Soto, M., Tischendorf, C.: Numerical analysis of DAEs from coupled circuit and semiconductor simulation. Appl. Numer. Math. 53:(2-4), 471–488 (2005)
18. Tischendorf, C.: Coupled Systems of Differential Algebraic and Partial Differential Equations in Circuit and Device Simulation. Habilitation thesis, Humboldt-University of Berlin (2003)