Coupled Systems of Differential Algebraic and Partial Differential Equations in Circuit and Device Simulation

Modeling and Numerical Analysis

Habilitationsschrift
zur Erlangung der Lehrbefähigung
für das Fach Mathematik

vorgelegt dem Rat der Mathematisch-Naturwissenschaftlichen Fakultät II der Humboldt-Universität zu Berlin

von

Dr. Caren Tischendorf

geb. am 9. Juni 1969 in Berlin

Prof. J. Mlynek
Präsident
der Humboldt-Universität zu Berlin

Prof. E. Kuhlke
Dekan

Berlin, den 17.03.2003

Gutachter/Gutachterinnen:

1. Prof. Roswitha März (Humboldt-Universität zu Berlin)
2. Prof. Stephen L. Campbell (North Carolina State University, USA)
3. Prof. Michael Hanke (NADA Royal Institute of Technology, Schweden)

Preface

This work has been accompanied and supported by many people. It is a great pleasure for me to have now the opportunity to express my gratitude to all of them.

First, I gratefully want to thank my academic teacher Prof. R. März for her support and inspiring suggestions during all the time.

Extensive thanks to my colleagues from Humboldt-University of Berlin and from Lund University in Sweden for providing a pleasant working environment and numerous fruitful discussions. In particular, I wish to thank Prof. C. Führer (Lund Univ., Sweden), Prof. W. Römisch (Humboldt Univ. of Berlin) and Dr. R. Lamour (Humboldt Univ. of Berlin) for their generous support and encouragement.

Furthermore, I would like to express my gratitude to Prof. M. Günther (Univ. Wuppertal), Prof. G. Söderlind (Lund Univ., Sweden), Prof. G. Wachutka (TU Munich), Dr. D. Estévez Schwarz (Infineon Technologies, Munich), Dr. U. Feldmann (Infineon Technologies, Munich) and Dipl.-Math. A. Bartel (Univ. Karlsruhe) for many valuable conversations.

Finally, my special thanks to my parents who have always supported me.

Berlin, March 2003 Caren Tischendorf

Contents

Introduction

In various application areas like automotive industry, telecommunication or robotics, the technological progress is mainly driven by a terrific development in the field of optic and electronic components. The enormous increase of performance is based on a higher complexity of integrated systems as well as smaller and smaller dimensions of optic and electronic devices. Both trends are connected with new difficulties for the design of circuits.

Circuit design consists usually of several steps: process simulation, device simulation, compact modeling for semiconductor devices, extraction and generation of model parameters, circuit simulation. The way over a compact modeling of devices (macro-modeling) was advantageous up to now since one could simulate circuits without an expensive device simulation. Thus, a simulation of integrated circuits was enabled in general. Nowadays, however, the performance of high frequency devices depends not only on their geometrical dimensions. It is significantly influenced by the surrounding circuitry. This requires additional (time consuming) iterations during the circuit design for the extraction and generation of model parameters. Furthermore, high frequency parts of a circuit have to be modeled with a very high precision for a reliable evaluation of the circuit function. Consequently, for complex circuits with high frequency devices, it is recommended to combine circuit simulation directly with a device simulation of elements of the high frequency part. This results in coupled systems of differential-algebraic equations (DAEs) and partial differential equations (PDEs) of elliptic and parabolic type.

A coupling of DAEs and PDEs becomes more and more important also in other applications fields (see e.g. the simulation of blood flows in a human vascular system [QRV01] or the simulation of elastic multibody systems [Sim98, Sim00]). Theoretical investigations for such coupled systems are still at the initial stage [CM99b, FY99, LSEL99, Sim00, Gün01a]. But they have already shown that the existence and exploitation of certain structures is important for an efficient and successful coupled simulation.

Since structural information about circuit equation systems was missing for long time, for coupled circuit and device simulation, approaches have been established that either extend existing device-simulators by more complex boundary conditions (see e.g. [LSK$^+$96, Rot00]) or combine device-simulators with circuit-simulators in form of a "black box" (see e.g. [ESTZ96]). Both approaches are unsuited for complex circuits in high-frequency domain.

In this document, we follow an approach that includes the device model equations into the network equation system directly. Based on a detailed analysis of network and device modeling, we elaborate information about the structure of the coupled circuit and device systems. This enables us to investigate numerical properties of the systems as sensitivity with respect to perturbations. In contrast to the studies in [Gün01a, ABGT02] we deal with the case of instationary device equation systems here.

Furthermore, we present a treatment of the coupled systems as differential algebraic systems in infinite-dimensional Hilbert spaces. Such systems are also called singular or degenerate abstract differential equations. In contrary to applications, where a singularity is given at one time point only, our systems have an operator in front of the time derivative that is singular at all time points. Therefore, we call the systems abstract differential algebraic systems (ADAS). Finally, we develop a Galerkin approach for such abstract systems that regards the involved constraints properly.

This work is divided into 4 chapters. The first one is devoted to a detailed network analysis. First, network elements and topological properties of the network are discussed from a mathematical point of view. Then, the most common network approach, the modified nodal analysis (MNA) is presented in such a way that structures of the resulting differential-algebraic equation system become visible. This enables us to develop network topological criteria for the DAE-index. Furthermore, questions concerning unique solvability and the perturbation behavior of solutions can be answered. Finally, the BDF methods applied to the network equation systems are investigated with respect to feasibility, convergence and stability.

The second chapter deals with modeling of semiconductor devices. After a short discussion of different models on different levels, we consider the drift diffusion equations. Particular care is taken to the boundary conditions since they play an important role for the coupling between circuit and device equations. In the end, function and geometry of two device examples are explained.

In the third chapter, the coupling of the network and device model equations is discussed. Exploiting charge conservation of the drift diffusion model, the

occurrence of semiconductor currents in the network equations can be described by a matrix of the same structure as those for other basic network elements. After formulating the complete coupled system, we present a homogenization that yields to boundary conditions for the device equations independent of the network equations. This way, node potentials of the network appear in the device equations but explicit given function spaces can be chosen for the solution. In preparation of the fourth chapter, the coupled system is formulated as an abstract differential algebraic system. First, a classic description is given. Then, the system is described as variational equation (generalized form).

The fourth chapter is devoted to abstract differential algebraic equations in infinite-dimensional Hilbert spaces. First, we present an index concept that orientates on perturbations of the right hand side and bases on the tractability index for DAEs. Then, we develop network topological criteria for the index of the coupled network and device equations. Finally, a Galerkin approach for handling ADASs with monotone operators is proposed and discussed with respect to convergence and solution behavior of perturbed systems.

The purpose of Appendix A is to collect basic aspects of different fields that are relevant in the framework of this work. It is devoted to readers who are not familiar with one or the other field. Appendix B collects symbols, common notations and physical constants used in this thesis.

Chapter 1

Network Analysis

Electric networks are present in almost all parts of our live. Main application areas are telecommunications, home electronics, transport communication as well as controlling, optimization and automatization of production processes of any kind.

Modern integrated circuits are often distinguished by a very high complexity and a very high packing density. The development of electronic circuits has progressed to a state where millions of transistors can be put on a chip. The numerical simulation of such circuits requires modeling techniques that allow an automatic generation of network equations. Furthermore, the number of independent network variables describing the network should be as small as possible.

The modified nodal analysis (MNA) provides a system with a relatively small dimension (number of nodes plus the number voltage controlling elements). It is adopted successfully in numerous programs, e.g. in SPICE (originally developed at Berkeley in the 70s and early 80s; commercialized versions: HSPICE, PSPICE, et.al.) and in TITAN (developed by SIEMENS / Infineon Technologies). The resulting systems represent differential-algebraic equations (DAEs).

The aspect that one is usually confronted with highly nonlinear DAEs led to the popular opinion that DAEs arising from circuit simulation do not have a special structure and their index is generally unknown. But, our investigations in [MT97, Tis99, ET00, Est00, Tis01, EFM$^+$03], have shown that the network DAEs are well structured (depending on the network topology), the relevant subspaces may be described by suitable projections and the index

may be checked by topological criteria. Since some of these structural properties are important for the numerical simulation and for the further analysis of the coupled systems, we will present them in this chapter in detail.

1.1 Network Modeling

The numerical simulation of electric networks is closely related to the network modeling. Circuit models have to meet two contradicting demands. On the one hand, they have to describe the physical behavior of a circuit as correct as possible. On the other hand, the models should be simple enough to keep computing time reasonably small.

A well established approach meeting both demands to a certain extent is the description of the network by a graph with branches and nodes. Branch currents, branch voltages and node potentials are introduced as variables. The node potentials are defined as voltages with respect to one reference node, usually the mass node.

The physical behavior of each network element is modeled by a relation between its branch currents j and its branch voltages v. Network elements fully described by a relation between a single branch current and a single branch voltage are called one-port or two-terminal elements. One port reflects one branch, one terminal reflects one end of a branch. Correspondingly, the other elements are called multi-port or multi-terminal elements. The describing current-voltage relations are called characteristic equations. The characteristic equations of basic network elements are given later in this section.

In order to complete the network model, the topology of the elements has to be taken into account. Assuming the electrical connections between the circuit elements to be ideally conducting and the nodes to be ideal and concentrated, the topology can be described by Kirchhoff's laws as we will see in the next sections.

1.1.1 Network Elements

The analysis and design of circuits requires an approximation of real elements by appropriate models. The level of the models ranges from simple algebraic equations, over ordinary and partial differential equations to Boltzmann and

Schroedinger equations depending on the effects to be described. Due to the high number of network elements (up to millions of elements) belonging to one circuit one is restricted to relatively simple models. In order to describe the physics as good as possible, so called compact models represent the first choice in network simulation. Complex elements such as transistors are modeled by small circuits containing basic network elements described by algebraic and ordinary differential equations only.

The development of such replacement circuits forms its own research field and leads nowadays to transistor models with more than 500 parameters. As an example, a model circuit describing a MOSFET (metal oxide semiconductor field effect transistor) on a very low modeling level is depicted in Figure 1.1. The correct adjustment of the parameters becomes more and more a

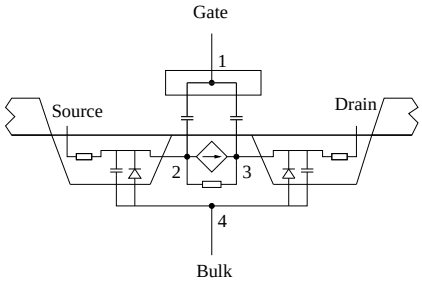

Figure 1.1: Circuit model for a MOSFET on a low modeling level

problem for the network design. Such a complexity motivates the idea to go back to a higher description level, that means to use distributed device models represented by a system of partial differential equations. This will be investigated in later chapters.

The basic network elements are ideal resistors, ideal condensers, ideal coils and ideal batteries. In order to express that we deal with ideal elements we use the terms resistor, capacitor, inductor, and voltage source, respectively. Furthermore, we shall use current sources that play an important role for the description of more complex elements such as transistors.

The characteristic equation of a resistor may be described as

$$v(t) = r(j(t), t) \qquad \text{or} \qquad j(t) = g(v(t), t) \tag{1.1}$$

where $j(t)$ reflects the current and $v(t)$ the voltage through the resistor at the time t. The functions r and g, respectively, can be linear and nonlinear. Typical examples are Ohms resistors and diodes. In the first case, the

functions r and g are linear and time-independent. In the second case, the function g reflects an exponential function with respect to the first argument.

Note, that j and v may form vector valued functions of time in case of multi-terminal elements. An element with n terminals is uniquely determined by the relation between $n-1$ branch voltages and $n-1$ branch currents. Therefore, one chooses often one reference terminal and considers the branches between the reference terminal and all other terminals (see Figure 1.2).

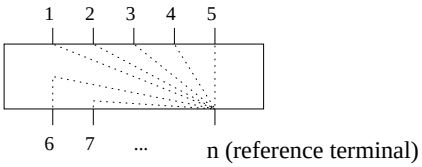

Figure 1.2: n-terminal circuit element

Capacitors store energy in their electric fields. Its charge $q(t)$ may be expressed by

$$q(t) = q_C(v(t), t).$$

The function q_C is usually monotone. Due to the definition of the current we get the voltage-current characteristics

$$j(t) = \frac{\mathrm{d}q_C(v(t), t)}{\mathrm{d}t}.$$

As described for resistors, the currents, charges and voltages may be vector valued functions in case of multi-terminal elements. Typical examples are the ideal condenser and the varactor diode. Transistor modeling requires often nonlinear multi-terminal capacitors.

Inductors store energies in their magnetic fields. The flux $\phi(t)$ is represented by

$$\phi(t) = \phi_L(j(t), t).$$

Regarding the definition of voltage, we arrive at

$$v(t) = \frac{\mathrm{d}\phi_L(j(t), t)}{\mathrm{d}t}.$$

Also in this case, the currents, fluxes and voltages may be vector valued functions. Most of coils have a nonlinear current-flux characteristics. One can approximate them by linear functions only in a small current range. In case of large currents, the flux growths sub-linearly.

Independent voltage and current sources are distinguished by the fact that the voltage and the current are given by

$$v(t) = v_s(t) \qquad \text{and} \qquad j(t) = i_s(t), \text{ respectively.}$$

If we have

$$v(t) = v_s(j(t)) \qquad \text{or} \qquad j(t) = i_s(\hat{j}(t))$$

then the sources are called current controlled. Again, v, j and $\hat{j}$ may be vector valued functions. Here, the branches corresponding to j are different from those corresponding to $\hat{j}$. In case of

$$v(t) = v_s(\hat{v}(t)) \qquad \text{or} \qquad j(t) = i_s(v(t)),$$

the sources are called voltage controlled. Here, the branches corresponding to v are different from those corresponding to $\hat{v}$.

In general, we have

$$v(t) = v_s(t, j(t), \hat{v}(t)) \qquad \text{for voltage sources}$$

and

$$j(t) = i_s(t, \hat{j}(t), v(t)) \qquad \text{for current sources.}$$

1.1.2 Network Topology

Assuming that quantum mechanical interactions between the network elements can be neglected, the electrical behavior of the network is described by Maxwell's equations which imply Kirchhoff's laws.

Considering one node with branch currents j_1, ..., j_n entering this node (see Figure 1.3) we may describe Kirchhoff's current law (KCL) as

$$\sum_{k=1}^{n} j_k = 0, \tag{1.2}$$

that means, the sum of all branch currents entering a node equals zero.

If we consider a loop with the branch voltages v_1, ..., v_n (see Figure 1.4), then we may formulate Kirchhoff's voltage law (KVL) as

$$\sum_{k=1}^{n} v_k = 0,$$

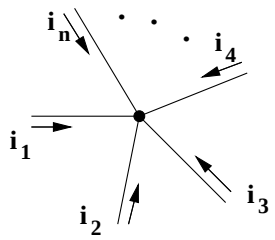

Figure 1.3: node with n conducting branches

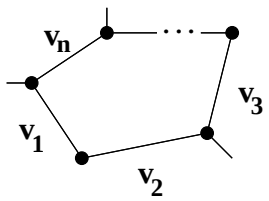

Figure 1.4: Loop with n conducting branches

that means, the sum of all branch voltages in a loop equals zero. Using Kirchhoff's laws, one can describe the network topology in an elegant way by the (reduced) incidence matrix $A = (a_{ij})$ that describes the relation between all nodes (except the mass node) and all branches of the network. It is defined as

$$a_{ij} = \begin{cases} 1 & \text{if the branch } j \text{ leaves the node } i, \\ -1 & \text{if the branch } j \text{ enters the node } i, \\ 0 & \text{else.} \end{cases}$$

In Appendix A.1, we have collected some useful properties of this matrix concerning the spaces spanned by certain rows or columns.

Let a connected network with n nodes and b branches be given. If $j = (j_1, j_2, ..., j_b)^{\mathrm{T}}$ is the vector of all branch currents of the circuit, then Kirchhoff's current law implies

$$Aj = 0. \tag{1.3}$$

As shown in Appendix A.1, the (reduced) incidence matrix has full column rank. Consequently, the maximal number of independent node equations describing the network is given by (1.3).

The incidence matrix allows, additionally, a simple description of the relation between node potentials and branch voltages of the network. If $v =$

$(v_1, v_2, ..., v_b)^{\mathrm{T}}$ is the vector of all branch voltages and $e = (e_1, e_2, ..., e_{n-1})^{\mathrm{T}}$ denotes the vector of all node potentials, then the relation

$$v = A^{\mathrm{T}} e \tag{1.4}$$

is satisfied. Each individual equation of (1.4) corresponds to one branch voltage. If we apply Kirchhoff's voltage law to a loop containing both nodes of the considered branch and the mass node, we get (1.4) directly.

1.1.3 The Modified Nodal Analysis

Let a connected electrical network be given. The nodal analysis is based on the network equations (1.3)

$$Aj = 0 \tag{1.5}$$

and (1.4)

$$v = A^{\mathrm{T}} e \tag{1.6}$$

as well as the characteristic equations of all network elements. As described in the section before, they may be written as

$$f(\frac{\mathrm{d}q_C(v,t)}{\mathrm{d}t}, \frac{\mathrm{d}\phi_L(j,t)}{\mathrm{d}t}, v, j, t) = 0. \tag{1.7}$$

The system (1.5)-(1.7) is a differential algebraic system, that means a coupled system of differential and algebraic equations in the network variables j, v und e. The dimension of this system equals $2b+n-1$. The approach leading to this system is called *sparse tableau analysis*.

The so called *modified nodal analysis* (MNA) requires a much smaller number of unknowns. In this case, one replaces the branch currents of all current defining elements in (1.5) by their characteristic equations, and all branch voltages by node voltages using (1.6). It is not difficult to see, that the resulting system represents a differential-algebraic equation.

In order to design an effective solution scheme, it is important to look at the structure of the equations. Other physical systems, like multibody systems, in mechanics, can be described by differential-algebraic equations, which are Euler equations of a variational principle (see e.g. [ESF98, RR99]). This gives them a structure, which is exploited by modern numerical schemes. In electrical network simulation, the structure is not so evident. Therefore, simulation methods could not be based on a particular structure until a few years ago. Our detailed investigations in [Tis99] lead to a structure based

description of the network equations. It has been extended for a more general class of networks in [ET00, EFM+03].

We will explain this structure here for networks containing resistors, capacitors, inductors and independent sources. Since we want to present the essential idea we will neglect the case of controlled sources. The treatment of controlled sources requires a lot of technical details which are presented in [ET00].

For the vector j_R of branch currents and the vector v_R of branch voltages of all resistors, we obtain

$$j_R = \tilde{g}(v_R, t).$$

Here, the function $\tilde{g}$ represents the composition of the functions g for each resistor. Since we are not interested in the individual functions g anymore, we will use g instead of $\tilde{g}$ in the following, that means $j_R = g(v_R, t)$. Analogously, for the branch currents j_C/j_L and the branch voltages v_C/v_L of all capacitors/inductors, we have

$$j_C = \frac{\mathrm{d}q_C(v_C, t)}{\mathrm{d}t}, \qquad v_L = \frac{\mathrm{d}\phi_L(j_L, t)}{\mathrm{d}t}.$$

Finally we get

$$v_V = v_s(t) \qquad \text{and} \qquad j_I = i_s(t)$$

for the branch voltages v_V of all voltage sources and the branch currents j_I of all current sources.

The essential idea for getting structure information is a numbering of the network branches in such a way that the incidence matrix forms a block matrix with blocks describing the different types of network elements. The blocks are then given as follows:

$$A = (A_R, A_C, A_L, A_V, A_I), \tag{1.8}$$

where the index stands for resistive, capacitive, inductive, voltage source and current source branches, respectively. Replacing the branch currents of all current defining elements in (1.5) by their characteristic equations, and all branch voltages by node voltages using (1.6), we obtain the system

$$\begin{aligned} A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}} e, t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V &= -A_I i_s(t), \\ \frac{\mathrm{d}\phi_L(j_L, t)}{\mathrm{d}t} - A_L^{\mathrm{T}} e &= 0, \\ A_V^{\mathrm{T}} e &= v_s(t) \end{aligned}$$

with the unknowns $e(t)$, $j_L(t)$, and $j_V(t)$. Consequently, the *classical* modified nodal approach results in a differential algebraic equation system of the form

$$\begin{aligned} A_C C(A_C^{\mathrm{T}} e, t) A_C^{\mathrm{T}} \frac{\mathrm{d}e}{\mathrm{d}t} + A_C \frac{\partial q_C}{\partial t}(A_C^{\mathrm{T}} e, t) & \\ + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V &= -A_I i_s(t), \qquad (1.9) \\ L(j_L, t)\frac{\mathrm{d}j_L}{\mathrm{d}t} + \frac{\partial \phi_L}{\partial t}(j_L, t) - A_L^{\mathrm{T}} e &= 0, \qquad (1.10) \\ A_V^{\mathrm{T}} e &= v_s(t). \qquad (1.11) \end{aligned}$$

if the functions $q_C(v,t)$ and $\phi_L(j,t)$ are sufficiently smooth and

$$C(v,t) := \frac{\partial q_C}{\partial v}(v,t), \qquad L(j,t) := \frac{\partial \phi_L}{\partial j}(j,t).$$

Denoting the number of nodes by n, the number of inductive branches by n_L and the number of voltage source branches by n_V, the dimension of the system is $n - 1 + n_L + n_V$.

In the *charge oriented* MNA approach, one introduces additionally charges q and fluxes ϕ as unknown variables. This implies the equivalent system

$$\begin{aligned} A_C \frac{\mathrm{d}q}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V &= -A_I i_s(t), \qquad (1.12) \\ \frac{\mathrm{d}\phi}{\mathrm{d}t} - A_L^{\mathrm{T}} e &= 0, \qquad (1.13) \\ A_V^{\mathrm{T}} e &= v_s(t), \qquad (1.14) \\ q &= q_C(A_C^{\mathrm{T}} e, t), \qquad (1.15) \\ \phi &= \phi_L(j_L, t). \qquad (1.16) \end{aligned}$$

At a first glance, the charge oriented system (1.12)-(1.16) seems to be disadvantageous since its dimension is significantly larger than the dimension of system (1.9)-(1.11). However it is, for several reasons, the main approach used in circuit simulators. For a detailed discussion of these reasons we refer to [GF99a, Gün01a]. We only want to mention a few aspects here. Replacement circuit models for semiconductor elements are often formulated by (1.15)-(1.16). This way charge and flux conservation is guaranteed automatically. Numerical methods applied to system (1.9)-(1.11) require the differentiation of the functions q_C and ϕ_L. Solving the resulting system of nonlinear equations requires the second derivatives of these functions, i.e., we need more smoothness. This plays a significant role for the numerical solution since models are usually not twice differentiable. Additionally, it is

computationally more expensive. Furthermore, charge and flux conservation is only fulfilled approximately.

Finally, the simple form of the equations (1.15) and (1.16) involves only function evaluations for the determination of q and ϕ. Consequently, from the computational point of view, the dimension of the charge oriented system equals the dimension of the classical system. In fact, one has to apply a numerical method to the system

$$\begin{aligned} A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}}e,t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}}e,t) + A_L j_L + A_V j_V &= -A_I i_s(t), \quad (1.17) \\ \frac{\mathrm{d}\phi_L(j_L,t)}{\mathrm{d}t} - A_L^{\mathrm{T}}e &= 0, \quad (1.18) \\ A_V^{\mathrm{T}}e &= v_s(t) \quad (1.19) \end{aligned}$$

directly (without the differentiation done in (1.9)-(1.11)). Note, that the system (1.17)-(1.19) represents a DAE with a proper stated leading term [Mär02a, Mär02b] if the matrices $C(v,t)$ and $L(j,t)$ are positive definite for all voltages v, currents j and time points t. These conditions seem to be natural from the physical point of view. For two-terminal capacitors, this means that a positive change of the voltage yields a current in forward direction. The current flows in reverse direction if the voltage change is negative.

1.2 Numerical Analysis

A straightforward approach based on BDF methods [Gea71] for the numerical solution of such DAEs has been used for numerous circuits and "is still among the best for this purpose" [VS94]. However, more and more complex circuit models can lead to DAEs of higher index (≥ 2). DAEs of higher index are ill-posed in the sense that small perturbations in the initial data may cause arbitrarily large changes in the solution data. See Appendix A.3 for more details about the index of a differential algebraic equation.

Furthermore, we are again confronted with basic questions. Do the circuit equations of the more complex models will have a unique solution? Does the numerical solution converge to the exact solution? Has the numerical solution the same qualitative behavior as the exact solution? Of special interests are answers to these problems based on criteria that can be checked during the model design.

Some answers, particularly for unique solvability, are known for certain kinds of networks [BM64, CR65, KR65, OW69, CL75, Mat76, Cam81, Sza82, HB84, NC84, Has86, Mat87, Fos92]. Typically, the results are based on the existence of specific trees or on the regularity of certain matrices. In [NC84] topological criteria for unique solutions of nonlinear resistive circuits containing linear controlled sources have been discussed for the first time. The authors apply a set of operations to the graph of the circuit and check a certain oriented structure, a cactus graph or a generalization thereof.

Beside investigations of special circuit systems, general differential-algebraic equations have widely been investigated during the last 20 years (see e.g. [Cam85, GM86, DHZ87, BCP89, HLR89, HW91, Kva90, Mär92, Rei90, RR91, LMM93, MS93, KM94, BC95, CG95b, CG95a, RR96, Lam97, BC98, KM98, KR99, HS02, RR02, HL03]). The results cover, among other things, unique solvability, feasibility of numerical methods as well as stability properties. However, most of the results suppose a certain structure (e.g. Hessenberg form), high smoothness and depend mainly on the index of the DAE.

Recently, the special structure and the index of the network equations have been investigated in several papers. In [GF95, GR96], it was shown that the index of a circuit may depend on the model level of transistors. A more general study of different circuit configurations was presented in [GF99a, GF99b]. It clarifies that the index may become arbitrarily high and may also depend on parameters. Furthermore, higher index configurations appear for certain charge and flux sources.

For a reliable and effective numerical simulation, the question how to avoid models described by higher index DAEs is of central importance. In [Tis97, Tis99] we have seen that networks consisting of passive resistors, capacitors and inductors and independent sources lead to a differential algebraic system with an index not higher than two. This result could be generalized to networks containing ideal transformers and gyrators [Rei98] and to networks containing controlled sources satisfying certain assumptions [ET00]. Since these results and the used approach are of importance for the analysis of the coupled systems considered in Chapter 3, we will briefly summarize the ideas and results in the next sections of this chapter.

1.2.1 Index of Network Equations

We consider electric networks consisting of resistors, capacitors, inductors and independent sources. As described in Section 1.1.1, these elements may

be multi-port elements. Furthermore, they may have nonlinear characteristics. We make the following assumptions

(A1) Smoothness. The functions $q_C(v,t)$, $\phi_L(j,t)$ and $g(v,t)$ are continuously differentiable for all $v \in \mathbb{R}^{n-1}$, $j \in \mathbb{R}^{n_L}$ and $t \in [t_0, T]$. The input signals $v_s(t)$ and $i_s(t)$ are continuous for all $t \in [t_0, T]$.

(A2) Local Passivity. The matrices

$$C(v,t) = \frac{\partial q_C}{\partial v}(v,t), \qquad L(j,t) = \frac{\partial \phi_L}{\partial j}(j,t), \qquad G(v,t) = \frac{\partial g}{\partial v}(v,t)$$

are positive definite for all $v \in \mathbb{R}^{n-1}$, $j \in \mathbb{R}^{n_L}$ and $t \in [t_0, T]$.

(A3) Consistency. The circuit contains no loops of voltage sources only and no cutsets of current sources only.

Assumption (A3) is necessary for a consistent model description. If it would not be satisfied, then the circuit equations would have either no solution or infinite many solutions due to Kirchhoff's laws.

Studying the network equations in more detail, it turns out that special loops and cutsets play an essential role as we will see in Theorem 1.5. The expressions *loop* and *cutset* are explained in Appendix A.1.

Definition 1.1
A loop consisting of capacitors and voltage sources only is called a *CV-loop* (see Figure 1.5). A cutset containing, inductors and current sources only, is called an *LI-cutset* (see Figure 1.5).

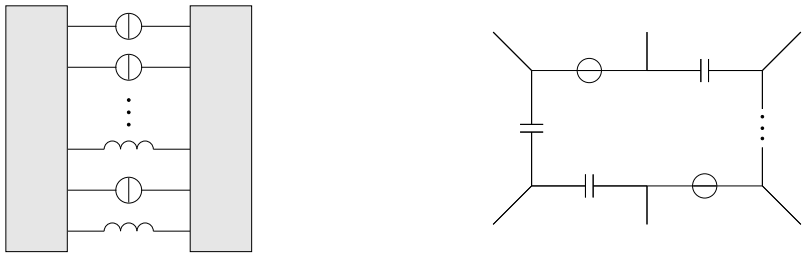

Figure 1.5: Example of an LI-cutset (left) and a CV-loop (right)

Using the splitting (1.8) of the incidence matrix, we can give a simple mathematical characterization of CV-loops and LI-cutsets.

Lemma 1.2 *The matrix (A_C, A_R, A_V) of a connected circuit has full row rank if and only if the circuit does not contain LI-cutsets.*

PROOF. Due to Theorem A.1, the matrix (A_C, A_R, A_V) has full row rank if and only if it has $n-1$ linear independent columns. By Theorem A.3, this is equivalent to the existence of a tree containing C-, R-, and V-branches only, which is the case for circuits without LI-cutsets.

□

Lemma 1.3 *Let Q_C be a projector onto* $\ker A_C$*. Then, the matrix $Q_C^T A_V$ has full column rank if and only if the circuit does not contain a CV-loop with a voltage source.*

PROOF. ($\Rightarrow$) We assume that the circuit has a CV-loop with a voltage source. Then, the columns of (A_V, A_C) are linear dependent (see Theorem A.2). Consequently, there is a non-zero vector $\binom{x}{y}$ such that

$$A_V x + A_C y = 0.$$

Consequently, $Q_C^{\mathrm{T}} A_V x = 0$. Furthermore, $x \neq 0$ since the loop contains a voltage source. Thus, $Q_C^{\mathrm{T}} A_V$ has not full column rank.

($\Leftarrow$) We assume that there is a non-zero vector x such that $Q_C^{\mathrm{T}} A_V x = 0$. Thus, $A_V x \in \ker Q_C^{\mathrm{T}} = \operatorname{im} A_C$. It exists a y such that $A_V x + A_C y = 0$. Since $x \neq 0$, we have a CV-loop with a voltage source.

□

As we will see later, the spaces $\ker (A_C\, A_R\, A_V)^{\mathrm{T}}$ and $\ker Q_C^{\mathrm{T}} A_V$ will play an essential role for the index determination. In order to be able to describe the elements of these spaces in a simple manner, we introduce two new projectors. Let Q_{CRV} be a projector onto $\ker (A_C\, A_R\, A_V)^{\mathrm{T}}$ and $\bar{Q}_{V-C}$ be a projector onto $\ker Q_C^{\mathrm{T}} A_V$. Additionally, let Q_{CRV} be chosen such that

$$\ker Q_{CRV} \supseteq \ker Q_C. \tag{1.20}$$

Such a choice is always possible since $\ker (A_C\, A_R\, A_V)^{\mathrm{T}} \cap \ker Q_C = \{0\}$.

The following lemma shows useful properties for a certain type of matrices, that is important for later discussion of the network equation systems.

Lemma 1.4 *For a positive definite matrix $M \in L(\mathbb{R}^m, \mathbb{R}^m)$ and any matrix $N \in L(\mathbb{R}^m, \mathbb{R}^k)$, it holds*

(i) $\ker N M N^{\mathrm{T}} = \ker N^{\mathrm{T}}$,

(ii) $\operatorname{im} NMN^{\mathrm{T}} = \operatorname{im} N$, *and*

(iii) $\operatorname{im} MN^{\mathrm{T}} \oplus \ker N = \mathbb{R}^m$.

PROOF: Ad(i) This is a simple consequence of the assumption that M is positive definite.

Ad(ii) The relation $\operatorname{im} NMN^{\mathrm{T}} \subseteq \operatorname{im} N$ is trivial. Considering (i), the equality follows from

$$\dim \operatorname{im} N = \dim \ker N^{\mathrm{T}}, \qquad \dim \operatorname{im} NMN^{\mathrm{T}} = \dim \ker NM^{\mathrm{T}}N^{\mathrm{T}}$$

and the fact that M^{T} is also positive definite.

Ad(iii) Part (i) implies immediately that

$$\operatorname{im} MN^{\mathrm{T}} \cap \ker N = \{0\}.$$

Let an arbitrary $y \in \mathbb{R}^k$ be given. Because of (ii), we find an $x \in \mathbb{R}^k$ such that

$$Ny = NMN^{\mathrm{T}}x.$$

Introducing $y_1 = MN^{\mathrm{T}}x$ and $y_2 = y - MN^{\mathrm{T}}x$, we get $y = y_1 + y_2$ with $y_1 \in \operatorname{im} MN^{\mathrm{T}}$ and $y_2 \in \ker N$.

□

With the collected facts in the lemmas before, we may proof the following index characterization for network DAEs.

Theorem 1.5 *Let the assumptions (A1)-(A3) be satisfied. Then, the DAE* (1.17)-(1.19) *has index*

(i) zero if and only if the network does not contain voltage sources and from each node exists a capacitive path to ground (mass node);

(ii) one if and only if the circuit has not index zero and, furthermore, it contains no LI-cutset and no CV-loop with at least one voltage source;

(iii) two in all other cases.

Remark 1.6 The theorem includes simple topological criteria for calculating the index of a network model. In contrast to the algorithm presented in [Rei98], this approach provides a very cheap and reliable index determination and has been successfully implemented as an index monitor in the industrial code TITAN (cf. [EFM+03]).

PROOF. We follow the matrix chain concept of the tractability index. For a short description of this approach see Appendix A.3. Recall the equation system (1.17)-(1.19):

$$\begin{aligned} A_C \frac{\mathrm{d}}{\mathrm{d}t} q_C(A_C^{\mathrm{T}} e, t) + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V &= -A_I i_s(t), \\ \frac{\mathrm{d}}{\mathrm{d}t} \phi_L(j_L, t) - A_L^{\mathrm{T}} e &= 0, \\ A_V^{\mathrm{T}} e &= v_s(t). \end{aligned}$$

Ad(i) We have to check, under which conditions, the matrix

$$G_0(A_C^{\mathrm{T}} e, j_L, t) = \begin{pmatrix} A_C C(A_C^{\mathrm{T}} e, t) A_C^{\mathrm{T}} & 0 & 0 \\ 0 & L(j_L, t) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is nonsingular. Since $C(\cdot)$ and $L(\cdot)$ are positive definite, this is obviously the case if and only if $\ker A_C^{\mathrm{T}} = \{0\}$ and the zero rows and columns in G_0 disappear (cf. Lemma 1.4). Regarding Theorem A.3, the null space is trivial if and only if the network has a tree containing capacitors only. But this is equivalent to the assertion that each node has a capacitive path to ground. The zero rows and columns in G_0 disappear if and only if the circuit does not contain voltage sources.

Ad(ii) The null space of $G_0(\cdot)$ is constant and has the form

$$\ker G_0(\cdot) = \ker A_C^{\mathrm{T}} \times \{0\} \times \mathbb{R}^{n_V},$$

where n_V denotes the number of V-branches in the network. For the matrix chain, we need a projector onto this null space. We are given one by

$$Q_0 := \begin{pmatrix} Q_C & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Regarding

$$B_0(A_C^{\mathrm{T}} e, j_L, t) = \begin{pmatrix} A_R G(A_R^{\mathrm{T}} e, t) A_R^{\mathrm{T}} & A_L & A_V \\ -A_L^{\mathrm{T}} & 0 & 0 \\ A_V^{\mathrm{T}} & 0 & 0 \end{pmatrix},$$

we arrive at

$$G_1(A_C^{\mathrm{T}} e, j_L, t) = \begin{pmatrix} A_C C(A_C^{\mathrm{T}} e, t) A_C^{\mathrm{T}} + A_R G(A_R^{\mathrm{T}} e, t) A_R^{\mathrm{T}} Q_C & 0 & A_V \\ -A_L^{\mathrm{T}} Q_C & L(j_L, t) & 0 \\ A_V^{\mathrm{T}} Q_C & 0 & 0 \end{pmatrix}. \tag{1.21}$$

Let $z = \begin{pmatrix} z_e \\ z_L \\ z_V \end{pmatrix}$ be an element of the null space of $G_1(\cdot)$. For brevity, we use only a dot for the arguments if their values are not important. The choice of z implies

$$\begin{pmatrix} (Q_C z_e)^{\mathrm{T}} & 0 & z_V^{\mathrm{T}} \end{pmatrix} G_1(\cdot) z = 0,$$

which yields to $z_e^{\mathrm{T}} Q_C^{\mathrm{T}} A_R G(\cdot) A_R^{\mathrm{T}} Q_C z_e = 0$. Applying Lemma 1.4, we get

$$A_R^{\mathrm{T}} Q_C z_e = 0, \tag{1.22}$$

since $G(\cdot)$ is positive definite. The last row in the equation $G_1(\cdot)z = 0$ reads as $A_V^{\mathrm{T}} Q_C z_e = 0$. Together with (1.22), we obtain

$$Q_C z_e = Q_{CRV} z_e.$$

The first row in $G_1(\cdot)z = 0$ implies

$$A_C C(\cdot) A_C^{\mathrm{T}} z_e + A_V z_V = 0,$$

if we regard again (1.22). Consequently, $Q_C^{\mathrm{T}} A_V z_V = 0$, that means

$$z_V = \bar{Q}_{V-C} z_V.$$

Now, it is not difficult to verify that

$$\begin{aligned} \ker G_1(\cdot) \;=\; \{z = \begin{pmatrix} z_e \\ z_L \\ z_V \end{pmatrix} :\; & Q_C z_e = Q_{CRV} z_e,\; z_V = \bar{Q}_{V-C} z_V, \\ & z_L = L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRV} z_e,\; P_C z_e = -H_C^{-1}(\cdot) A_V \bar{Q}_{V-C} z_V\} \end{aligned} \tag{1.23}$$

for $P_C := I - Q_C$ and $H_C(v,t) := A_C C(v,t) A_C^{\mathrm{T}} + Q_C^{\mathrm{T}} Q_C$. $H_C(\cdot)$ is positive definite, since $C(\cdot)$ is positive definite. Due to Assumption (A3), the matrices

$$(A_L, A_C, A_R, A_V)^{\mathrm{T}} \quad \text{and} \quad A_V$$

have full column rank (see Theorems A.1, A.2 and A.3). Consequently, we have

$$\ker A_L^{\mathrm{T}} Q_{CRV} = \ker Q_{CRV} \quad \text{and} \quad \ker A_V^{\mathrm{T}} \bar{Q}_{V-C} = \ker \bar{Q}_{V-C}.$$

Together with (1.23), we see that $G_1(\cdot)$ is nonsingular if and only if $Q_{CRV} = 0$ and $\bar{Q}_{V-C} = 0$. But this is equivalent to the assertion due to Lemma 1.2, Lemma 1.3, and the definition of these projectors.

Ad(iii) It remains to show that

$$\bar{G}_2(A_C^{\mathrm{T}} e, j_L, t) = G_1(A_C^{\mathrm{T}} e, j_L, t) + B_0(A_C^{\mathrm{T}} e, j_L, t) P_0 Q_1(A_C^{\mathrm{T}} e, j_L, t)$$

is nonsingular, if $P_0 = I - Q_0$ and $Q_1(A_C^{\mathrm{T}} e, j_L, t)$ is a projector onto the null space of $G_1(A_C^{\mathrm{T}} e, j_L, t)$. Regarding (1.23), (1.20), and

$$P_C H_C^{-1}(\cdot) = H_C^{-1}(\cdot) P_C^{\mathrm{T}},$$

it is not difficult to verify that

$$Q_1(v,j,t) := \begin{pmatrix} Q_{CRV} & 0 & -H_C^{-1}(v,t)A_V\bar{Q}_{V-C} \\ L^{-1}(j,t)A_L^{\mathrm{T}}Q_{CRV} & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C} \end{pmatrix}$$

is a projector onto $\ker G_1(v,j,t)$. Using this projector, we obtain

$$B_0(\cdot)P_0Q_1(\cdot) = \begin{pmatrix} A_L L^{-1}(\cdot)A_L^{\mathrm{T}}Q_{CRV} & 0 & -A_R G(\cdot)A_R^{\mathrm{T}}P_C H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \\ 0 & 0 & -A_L^{\mathrm{T}}P_C H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \\ 0 & 0 & -A_V^{\mathrm{T}}H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \end{pmatrix}$$

and

$$\bar{G}_2(\cdot) = \begin{pmatrix} A_C C(\cdot)A_C^{\mathrm{T}} + A_R G(\cdot)A_R^{\mathrm{T}}Q_C + A_L L^{-1}(\cdot)A_L^{\mathrm{T}}Q_{CRV} & 0 & A_V - A_R G(\cdot)A_R^{\mathrm{T}}P_C H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \\ -A_L^{\mathrm{T}}Q_C & L(j_L,t) & -A_L^{\mathrm{T}}P_C H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \\ A_V^{\mathrm{T}}Q_C & 0 & -A_V^{\mathrm{T}}H_C^{-1}(\cdot)A_V\bar{Q}_{V-C} \end{pmatrix}.$$

Let $z = \begin{pmatrix} z_e \\ z_L \\ z_V \end{pmatrix}$ belong to the null space of $\bar{G}_2(\cdot)$. This yields to

$$0 = \begin{pmatrix} Q_{CRV}^{\mathrm{T}} & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C}^{\mathrm{T}} \end{pmatrix} \bar{G}_2(\cdot)z = \begin{pmatrix} Q_{CRV}^{\mathrm{T}}A_L L^{-1}(\cdot)A_L^{\mathrm{T}}Q_{CRV}z_e \\ \bar{Q}_{V-C}^{\mathrm{T}}A_V^{\mathrm{T}}H_C^{-1}(\cdot)A_V\bar{Q}_{V-C}z_V \end{pmatrix}.$$

Applying Lemma 1.4, we get

$$A_L^{\mathrm{T}}Q_{CRV}z_e = 0 \quad \text{and} \quad A_V\bar{Q}_{V-C}z_V = 0,$$

since $L^{-1}(\cdot)$ and $H_C^{-1}(\cdot)$ are positive definite. Note that the inverse of a positive definite matrix is also positive definite.

Recall that the matrices $(A_C, A_R, A_V, A_L)^{\mathrm{T}}$ and A_V have full column rank. Consequently,

$$Q_{CRV}z_e = 0 \quad \text{and} \quad \bar{Q}_{V-C}z_V = 0.$$

Furthermore, $0 = \bar{G}_2(\cdot)z = G_1(\cdot)z$ and, hence,

$$z = Q_1(\cdot)z = Q_1(\cdot)\begin{pmatrix} Q_{CRV}z_e \\ 0 \\ \bar{Q}_{V-C}z_V \end{pmatrix} = 0,$$

that means, $\bar{G}_2(\cdot)$ is nonsingular.

□

1.2.2 Unique Solvability

Solutions of DAEs have not only to follow the flow given by the inherent regular differential equation but also to satisfy constraints. It is well-known, for index-1 systems, that these constraints are given by the derivative-free part of the DAE. For higher index systems, certain hidden constraints have additionally to be satisfied. They arise through the differentiation of parts of the DAE. Using our structural information about the circuit systems, we may formulate all constraints as follows.

The derivative-free part of the system (1.17)-(1.19) is obviously described by

$$\begin{aligned} Q_C^{\mathrm{T}}[A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V + A_I i_s(t)] &= 0, & (1.24)\\ A_V^{\mathrm{T}} e - v_s(t) &= 0. & (1.25) \end{aligned}$$

We want to call these constraints index-1 constraints in the following. From the section before, we know that a higher index (= 2) is caused by CV-loops and LI-cutsets. Correspondingly, the hidden constraints are caused by the differentiations of the KVL equations of CV-loops and the KCL equations of LI-cutsets. Therefore, we need an additional smoothness assumption.

(A4) Let the charges $q_C(v,t)$ of capacitors of CV-loops and the fluxes $\phi_L(j_L,t)$ of inductors of LI-cutsets be twice continously differentiable. Furthermore, the input signals $i_s(t)$ and $v_s(t)$ are assumed to be continuously differentiable.

Then, the hidden constraints my be described by (see [ET00])

$$\begin{aligned} \bar{Q}_{V-C}^{\mathrm{T}} A_V^{\mathrm{T}} H_1^{-1}(\cdot) P_C^{\mathrm{T}} [A_C \frac{\partial q_C}{\partial t}(A_C^{\mathrm{T}} e, t) + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L & \\ + A_V j_V + A_I i(\cdot)] + \bar{Q}_{V-C}^{\mathrm{T}} v_s'(t) &= 0, & (1.26)\\ Q_{CRV}^{\mathrm{T}} [A_L L^{-1}(\cdot)(A_L^{\mathrm{T}} e - \frac{\partial \phi_L}{\partial t}(j_L, t)) + A_I i_s'(t)] &= 0. & (1.27) \end{aligned}$$

This becomes clear by the proof of the following theorem.

Theorem 1.7 *Let the assumptions (A1)-(A4) be satisfied. Then, the DAE* (1.17)-(1.19)*, completed by an initial value satisfying all constraints* (1.24)-(1.27)*, is locally uniquely solvable.*

Remark 1.8 The theorem shows, additionally, that an initial value is consistent for the system (1.17)-(1.19) if it satisfies the constraints (1.24)-(1.27). Recall that an initial value is called consistent, if it exists locally a solution through this initial point.

Remark 1.9 For the computation of consistent initial values, one only needs to calculate a DC-operation point[1] in order to fulfill the index-1 constraints, and solve, furthermore, a linear system to satisfy the hidden constraints (see [Est02, Est00]). More precisely, if x^{DC} is a DC-operation point, then the value

$$x_0 := \begin{pmatrix} P_{CRV} e^{DC} + Q_{CRV} \hat{e} \\ j_L^{DC} \\ \bar{P}_{V-C} j_V^{DC} + \bar{Q}_{V-C} \hat{j}_V \end{pmatrix}$$

is consistent if $Q_{CRV}\hat{e}$ is the unique solution of the linear system

$$\begin{aligned} &Q_{CRV}^{\mathrm{T}} A_L L^{-1}(j_L^{DC}, t_0) A_L^{\mathrm{T}} Q_{CRV} \hat{e} = \\ &\quad Q_{CRV}^{\mathrm{T}} \left[A_L L^{-1}(j_L^{DC}, t_0) \left(\frac{\partial \phi_L(j_L^{DC}, t_0)}{\partial t} - P_{CRV} e^{DC} \right) - A_I i_s'(t_0) \right] \end{aligned}$$

and $\bar{Q}_{V-C}\hat{j}_V$ is the unique solution of

$$\begin{aligned} &\bar{Q}_{V-C}^{\mathrm{T}} A_V^{\mathrm{T}} H_C^{-1}(A_C^{\mathrm{T}} e^{DC}, t_0) A_V \bar{Q}_{V-C} \hat{j}_V = \\ &\qquad \bar{Q}_{V-C}^{\mathrm{T}} \Big[A_V^{\mathrm{T}} H_C^{-1}(A_C^{\mathrm{T}} e^{DC}, t_0) P_C^{\mathrm{T}} \Big(A_C \frac{\partial q_C(A_C^{\mathrm{T}} e^{DC}, t_0)}{\partial t} \\ &\qquad\quad + A_R g(A_R^{\mathrm{T}} e^{DC}, t_0) + A_L j_L^{DC} + A_I i_s(t_0) \Big) - v_s'(t_0) \Big]. \end{aligned}$$

Indeed, a few technical computations show that x_0 constructed in such a way fulfills all constraints (1.24)-(1.27).

Remark 1.10 Applying well-known solvability results for index-1 DAEs (see e.g. [GM86]) we obtain unique solvability for circuit systems which contain neither LI-cutsets nor CV-loops. However, most solvability results for index-2 systems are not applicable since the considered DAE does not have a Hessenberg structure.

Proof: First, we rewrite the system (1.17)-(1.19) as the extended system

$$\begin{aligned} A_C \frac{\mathrm{d}q}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V &= -A_I i_s(t), && (1.28) \\ \frac{\mathrm{d}\phi}{\mathrm{d}t} - A_L^{\mathrm{T}} e &= 0, && (1.29) \\ A_V^{\mathrm{T}} e &= v_s(t), && (1.30) \\ q &= q_C(A_C^{\mathrm{T}} e, t), && (1.31) \\ \phi &= \phi_L(j_L, t). && (1.32) \end{aligned}$$

[1] A DC operation point is a point satisfying the stationary network equations.

Obviously, (e, j_L, j_V) is a solution of the system (1.17)-(1.19) if and only if (q, ϕ, e, j_L, j_V) with

$$q = q_C(A_C^{\mathrm{T}} e, t), \qquad \phi = \phi_L(j_L, t)$$

is a solution of the extended system (1.28)-(1.32). The advantage of the extended formulation is its special structure. It has a constant leading coefficient matrix and the higher index variables appear only linearly. More precisely, we may formulate (1.28)-(1.32) as [Est00]

$$\bar{A}x' + \bar{b}(Ux, t) + \bar{B}Tx = 0. \tag{1.33}$$

for $x = (q, \phi, e, j_L, j_V)$, where T and U are complementary projectors

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{CRV} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{Q}_{V-C} \end{pmatrix}, \qquad U = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & P_{CRV} & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & \bar{P}_{V-C} \end{pmatrix},$$

and the matrices $\bar{A}$ and B are given by

$$\bar{A} = \begin{pmatrix} A_C & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \bar{B}T = \begin{pmatrix} 0 & 0 & 0 & 0 & A_V \bar{Q}_{V-C} \\ 0 & 0 & -A_L^T Q_{CRV} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, we have

$$\bar{b}(P_{CRV}e, j_L, \bar{P}_{V-C}j_V, q, \phi) = \begin{pmatrix} A_R g(A_R^{\mathrm{T}} P_{CRV} e, t) + A_L j_L + A_V \bar{P}_{V-C} j_V + A_I i_s(t) \\ -A_L^{\mathrm{T}} P_{CRV} e \\ A_V^{\mathrm{T}} P_{CRV} e - v_s(t) \\ q - q_C(A_C^{\mathrm{T}} P_{CRV} e, t) \\ \phi - \phi_L(j_L, t) \end{pmatrix}.$$

Note, that the projector T represents a projector onto the space of the index-2 components $N \cap S(\cdot)$ with

$$N = \ker \bar{A}, \quad S(\cdot) = \{z : \ \bar{b}'_x(\cdot)Uz \in \operatorname{im} \bar{A}\}.$$

The Theorem 2.4.6 of [Est00] implies that x is a solution of (1.33) if and only if the initial value x_0 satisfies the constraints given by the derivative-free part, i.e.

$$W_0 \bar{b}(Ux_0, t_0) = 0, \tag{1.34}$$

for

$$W_0 = \begin{pmatrix} Q_{CRV}^T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C}^T & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix},$$

and x is a solution of the index-1 system

$$\begin{aligned} &\left(\bar{A} + W_1(Ux,t)(K_{W_1}W_0\bar{b})'_x(Ux,t)\right)x'(t) + \bar{b}(Ux,t) \\ &+ \bar{B}Tx - \hat{W}_1\bar{b}(Ux,t) + W_1(Ux,t)(K_{W_1}W_0\bar{b})'_t(Ux,t) \;=\; 0 \end{aligned} \tag{1.35}$$

for

$$W_1(\cdot) = \begin{pmatrix} Q_{CRV}^T & 0 & 0 & 0 & Q_{CRV}^T A_L L^{-1}(\cdot) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C}^T & \bar{Q}_{V-C}^T A_V^T H_C^{-1}(\cdot) A_C & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$K_{W_1}W_0 = \begin{pmatrix} Q_{CRV}^T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C}^T & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix}, \qquad \hat{W}_1 = \begin{pmatrix} Q_{CRV}^T & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{Q}_{V-C}^T & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Note, that W_0 is a projector along $\operatorname{im}\bar{A}$ and W_1 is a projector along

$$\operatorname{im}\left[\bar{A} + (\bar{b}'_x(Ux,t) + BT)\bar{Q}_0\right]$$

for any projector $\bar{Q}_0$ onto $\ker\bar{A}$. The condition (1.34) means exactly that x_0 fulfills the constraints (1.24)-(1.25). Consequently, it remains to show that the index-1 system (1.35) has a unique solution.

For brevity, we rewrite (1.35) as

$$\bar{\bar{A}}(x,t)x'(t) + \bar{\bar{b}}(x,t) = 0. \tag{1.36}$$

The leading coefficient matrix $\bar{\bar{A}}(x,t)$ has a constant nullspace

$$\ker\bar{\bar{A}}(x,t) = \ker\left(\bar{A} + W_1(Ux,t)(K_{W_1}W_0\bar{b})'_x(Ux,t)\right) = \ker\bar{A}$$

since $\bar{A}$ is constant. Furthermore, $\bar{\bar{A}}(x,t)$ and $\bar{\bar{b}}(x,t)$ are continuously differentiable with respect to x and continuous with respect to t. Consequently (cf. [GM86]), the index-1 system (1.36) has a unique solution if and only if x_0 is consistent for the system (1.36), i.e.

$$\bar{\bar{b}}(x_0,t_0) \in \operatorname{im}\bar{\bar{A}}(x_0,t_0).$$

Regarding all constraints (1.24)-(1.27), we obtain

$$\bar{\bar{b}}(x_0,t_0) = \begin{pmatrix} P_C^{\mathrm{T}} A_R G(A_R^{\mathrm{T}} e_0,t_0) + P_C^{\mathrm{T}} A_L j_{L0} + P_C^{\mathrm{T}} A_V j_{V0} + P_C^{\mathrm{T}} A_I i_s(t_0) - QCRV^{\mathrm{T}} A_L L^{-1}(\cdot) A_L^{\mathrm{T}} e_0 \\ -A_L^{\mathrm{T}} e_0 \\ \bar{Q}_{V-C}^{\mathrm{T}} A_V^{\mathrm{T}} H_C^{-1}(\cdot) P_C^{\mathrm{T}} (A_R g(A_R^{\mathrm{T}} e_0,t_0) + A_L j_{L0} + A_V j_{V0} + A_I i_s(t_0)) \\ 0 \\ 0 \end{pmatrix}$$

$$= \bar{\bar{A}}(x_0,t_0) y$$

if we choose y such that

$$y_1 = -A_L^{\mathrm{T}} e_0, \quad A_C y_2 = P_C^{\mathrm{T}} (A_R g(A_R^{\mathrm{T}} e_0,t_0) + A_L j_{L0} + A_V j_{V0} + A_I i_s(t_0)),$$
$$y_3 = 0, \quad y_4 = 0, \quad y_5 = 0.$$

Consequently, the system (1.36) and, finally, the system (1.17)-(1.19) have a unique solution. □

Remark 1.11 The proof has shown that $Q_{CRV} e$ and $\bar{Q}_{V-C} j_V$ represent the index-2 variables of the system. They are of importance when solving the network equation systems (1.17)-(1.19) numerically since they are caused by instability effects. We will present the details for that in Section 1.2.4. Network topologically spoken, $Q_{CRV} e$ and $\bar{Q}_{V-C} j_V$ represent the voltages of LI-cutsets and currents of CV-loops, respectively.

1.2.3 Perturbation Analysis

Our aim is to derive some statements about the feasibility, convergence and stability of numerical methods applied to the circuit equation systems. All problems are closely related to the question about the solution behavior of slightly perturbed network equations. The following theorem gives detailed information about the influence of small perturbations.

Theorem 1.12 *Let the assumptions of Theorem (A1)-(A4) be satisfied and a solution* $x_* := (e_*, j_{L*}, j_{V*})$ *of* (1.17)-(1.19) *be given.*

(i) Then, the perturbed initial value problem

$$A_C \frac{dq_C(A_C^T e,t)}{dt} + A_R r(A_R^T e) + A_L j_L + A_V j_V + A_I i(\cdot) = \delta_e, \quad (1.37)$$

$$\frac{d\phi_L(j_L,t)}{dt} - A_L^T e = \delta_L, \quad (1.38)$$

$$A_V^T e - v = \delta_V, \quad (1.39)$$

combined with a consistent initial value $x_0 = (e_0, j_{L0}, j_{V0})$ *is locally uniquely solvable if the initial error* $|x_0 - x_*(t_0)|$ *and the perturbations* $\|\delta\|_\infty$ *(for* $\delta := (\delta_e, \delta_L, \delta_V)$*),* $\|Q_{CRV}^T \delta_e'\|_\infty$*,* $\|\bar{Q}_{V-C}^T \delta_V'\|_\infty$ *are sufficiently small.*

(ii) For the solution $x = (e, j_L, j_V)$ *of the perturbed system* (1.37)-(1.39)*, the inequality*

$$\|x - x_*\|_\infty \le c(\|\delta\|_\infty + \|Q_{CRV}^T \delta_e'\|_\infty + \|\bar{Q}_{V-C}^T \delta_V'\|_\infty + |x_0 - x_*(t_0)|) \quad (1.40)$$

is true for a constant $c > 0$*.*

Remark 1.13 The theorem implies immediately that the perturbation index [HW91] of the network equation systems (1.17)-(1.19) does not exceed 2 and coincides with the differential as well as the tractability index.

Proof: We want to apply Theorem 3.12 of [Tis96] to the extended system

$$\begin{aligned} A_C \frac{\mathrm{d}q}{\mathrm{d}t} + A_R g(A_R^T e) + A_L j_L + A_V j_V + A_I i(\cdot) &= \delta_e, \\ \frac{\mathrm{d}\phi}{\mathrm{d}t} - A_L^T e &= \delta_L, \\ A_V^T e - v &= \delta_V, \\ q - q_C(A_C^T e, t) &= 0, \\ \phi - \phi_L(j_L, t) &= 0. \end{aligned}$$

The Theorem 3.12 of [Tis96] reads as follows. Let $x_* \in \{x \in C([t_0, T], \mathbb{R}^m) : Px \in C^1([t_0, T], \mathbb{R}^m)\}$ be a solution of the index-2 tractable DAE

$$A(t)\dot{x}(t) + g(x(t), t) = 0.$$

Further, let $UQ_0 x_* \in C^1([t_0, T], \mathbb{R}^m)$ be fulfilled and let Q_1 as well as $UQ_0 G_2^{-1} g_x'$ be of class C^1. If the structural condition

$$Q_1(t)(I + \hat{g}_x'(x, t) - \hat{g}_x'(x_*(t), t))^{-1} T(t) Q_0 = 0$$

with

$$\hat{g}(x, t) := [U(t)Q + P_0 Q_1(t)] G_2^{-1}(t) g(x, t)$$

is satisfied for all (x, t) with $|x - x_*(t)| \le \varrho$, the perturbation q is continuous, and its part

$$\hat{q}(t) := [U(t)Q_0 + P_0 Q_1(t)] G_2^{-1}(t) q(t)$$

is of class C^1, then the perturbed initial value problem

$$\begin{aligned} A(t)\dot{x}(t) + g(x(t),t) &= q(t) \\ P_0P_1(s)x(s) &= u_s \in \operatorname{im} P_0P_1(s), \qquad |u_s - P_0P_1(s)x_*(s)| \le \tau \\ \|q\|_\infty + \|\hat{q}'\|_\infty &\le \sigma \end{aligned}$$

is locally uniquely solvable for sufficiently small τ and σ. Moreover, the inequality

$$\begin{aligned} \|x - x_*\|_\infty + \|(P_0P_1(x - x_*))'\|_\infty \le \\ \text{const}\,(\|q\|_\infty + \|\hat{q}'\|_\infty + |u_s - P_0P_1(s)x_*(s)|) \end{aligned}$$

is satisfied. For the definition of Q_0, P_0, Q_1, P_1 and G_2 see Appendix A.3 and use $G_0 := A$.

Regarding Remark 3 on page 51 in [Tis96], the assumptions of the theorem are satisfied since the space $N \cap S(\cdot)$ of index-2 components is constant (see proof of Theorem 1.7). Considering that Tx appears only linearly in (1.33), it turns out that besides the perturbation δ itself the derivative of the part $P_0Q_1(\cdot)G_2^{-1}(\cdot)\delta$ has an influence on the solution of the perturbed system. For the system (1.33) we get

$$P_0 = \begin{pmatrix} \bar{P}_C & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Using the nonsingular matrix functions [ET00]

$$\begin{aligned} H_C(A_C^Te,t) &= A_CC(A_C^Te,t)A_C^{\mathrm{T}} + Q_C^{\mathrm{T}}Q_C \\ \bar{H}_C &= A_CA_C^{\mathrm{T}} + Q_C^{\mathrm{T}}Q_C \\ H_L(j_L,t) &= A_{CRV}^{\mathrm{T}}A_LL^{-1}(j_L,t)A_L^{\mathrm{T}}Q_{CRV} + P_{CRV}^{\mathrm{T}}P_{CRV} \\ H_{V-C}(A_C^{\mathrm{T}}e,t) &= \bar{Q}_{V-C}^{\mathrm{T}}A_V^{\mathrm{T}}H_C^{-1}(A_C^{\mathrm{T}}e,t)A_V\bar{Q}_{V-C} + \bar{P}_{V-C}^{\mathrm{T}}\bar{P}_{V-C} \end{aligned}$$

and

$$\begin{aligned} \hat{P}_C &= \bar{P}_CA_C^{\mathrm{T}}\bar{H}_C^{-1}(\cdot)A_V\bar{Q}_{V-C}H_{V-C}^{-1}(\cdot)\bar{Q}_{V-C}^{\mathrm{T}}A_V^{\mathrm{T}}H_C^{-1}(\cdot)A_C \\ \hat{P}_L &= A_L^{\mathrm{T}}Q_{CRV}H_L^{-1}(\cdot)Q_{CRV}^{\mathrm{T}}A_LL^{-1}(\cdot), \end{aligned}$$

we obtain

$$Q_1(\cdot) = \begin{pmatrix} \hat{P}_C & 0 & 0 & 0 & 0 \\ 0 & \hat{P}_L & 0 & 0 & 0 \\ 0 & Q_{CRV}H_L^{-1}(\cdot)Q_{CRV}^{\mathrm{T}}A_LL^{-1}(\cdot) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\bar{Q}_{V-C}H_{V-C}^{-1}\bar{Q}_{V-C}^{\mathrm{T}}A_V^{\mathrm{T}}H_C^{-1}(\cdot)A_C & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The nonsingular matrix G_2 is given by

$$G_2(\cdot) = \begin{pmatrix} A_C & 0 & A_R G(\cdot) A_R^{\mathrm{T}} & A_L & A_V \\ 0 & I & -A_L^{\mathrm{T}} & 0 & 0 \\ 0 & 0 & A_V^{\mathrm{T}} & 0 & 0 \\ \hat{P}_C + \bar{Q}_C & 0 & -C(\cdot) A_C^{\mathrm{T}} & 0 & 0 \\ 0 & \hat{P}_L & 0 & -L(\cdot) & 0 \end{pmatrix},$$

which leads to

$$PQ_1(\cdot)G_2^{-1}(\cdot) = \begin{pmatrix} 0 & 0 & \bar{P}_C A_C^{\mathrm{T}} \bar{H}_C^{-1}(\cdot) A_V \bar{Q}_{V-C} H_{V-C}^{-1}(\cdot) \bar{Q}_{V-C}^{\mathrm{T}} & \hat{P}_C & 0 \\ A_L^{\mathrm{T}} Q_{CRV} H_L^{-1}(\cdot) Q_{CRV}^{\mathrm{T}} & 0 & 0 & 0 & \hat{P}_L \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the theorem is proven. □

1.2.4 Feasibility, Convergence and (In)stability of the BDF Methods

Let again the assumptions (A1)-(A4) be fulfilled. We consider a partition π of a compact interval $[t_0, T]$ with the following properties.

$$\begin{aligned} &\pi : t_0 < t_1 < \cdots < T, \\ &h_{min} \le h_n := t_n - t_{n-1} \le h_{max}, \quad h_{min} > 0, \quad n \ge 1, \\ &\kappa_1 \le \frac{h_{n-1}}{h_n} \le \kappa_2, \quad n \ge 1, \end{aligned} \tag{1.41}$$

where κ_1 and κ_2 are suitable constants such that the BDF methods are stable for explicit ordinary differential equations (cf. [Gri83, GM86]).

Then, the BDF method applied to network equation systems (1.17)-(1.19) may be formulated as

$$\begin{aligned} A_C \frac{1}{h_n} \sum_{i=0}^{k} \alpha_{ni} q_C(A_C^{\mathrm{T}} e_n, t_n) + A_R g(A_R^{\mathrm{T}} e_n, t_n) & \\ + A_L j_{nL} + A_V j_{nV} + A_I i_s(t_n) &= \delta_{ne}, \qquad (1.42) \\ \frac{1}{h_n} \sum_{i=0}^{k} \alpha_{ni} \phi_L(j_{nL}, t_n) - A_L^{\mathrm{T}} e_n &= \delta_{nL}, \qquad (1.43) \\ A_V^{\mathrm{T}} e_n - v_s(t_n) &= \delta_{nV}. \qquad (1.44) \end{aligned}$$

Here, δ_n describes the perturbations on the n-th step for $n \ge k$, which is caused by numerical computations including errors arising from solving the

nonlinear equations (e.g. with a Newton-like method). Obviously, the system (1.42)-(1.44) is equivalent to

$$\begin{aligned} A_C \frac{1}{h_n} \sum_{i=0}^{k} \alpha_{ni} q_n + A_R g(A_R^{\mathrm{T}} e_n, t_n) + A_L j_{nL} & \\ + A_V j_{nV} + A_I i_s(t_n) &= \delta_{ne}, \qquad (1.45) \\ \frac{1}{h_n} \sum_{i=0}^{k} \alpha_{ni} \phi_n - A_L^{\mathrm{T}} e_n &= \delta_{nL}, \qquad (1.46) \\ A_V^{\mathrm{T}} e_n - v_s(t_n) &= \delta_{nV}, \qquad (1.47) \\ q_n - q_C(A_C^{\mathrm{T}} e_n, t_n) &= 0, \qquad (1.48) \\ \phi_n - \phi_L(j_{nL}, t_n) &= 0 \qquad (1.49) \end{aligned}$$

if

$$q_n = q_C(A_C^{\mathrm{T}} e_n, t_n) \quad \text{and} \quad \phi_n = \phi_L(j_{nL}, t_n).$$

But the system (1.45)-(1.49) represents nothing else than the BDF method applied to the extended system (1.28)-(1.32). Since the equations (1.48) and (1.49) require only function evaluations, we don't need to consider perturbations in these equations. We denote the exact solution by x_* and the local truncation error (lte) by τ_n. Then, we may formulate the following feasibility and stability result.

Theorem 1.14 *Let the assumptions (A1)-(A4) be fulfilled. Supposed there is a constant $c > 0$ such that the starting values satisfy the relation*

$$\|x_n - x_*(t_n)\| \le c h_n, \quad n < k, \qquad (1.50)$$

then it holds that:

(i) There are constants $c_1 > 0$ and $r > 0$ such that the BDF with

$$\|\delta_n\| \le c_1 \text{ for } n \ge k \quad \text{and} \quad \frac{\|Q_{CRV}^T \delta_{ne}\| + \|\bar{Q}_{V-C}^T \delta_{nV}\|}{h_n} \le c_1 \text{ for } n \ge 0,$$

is feasible for all partitions (1.41) with sufficiently small stepsizes, i.e., the nonlinear equations are solvable with $x_n \in B(x_*(t_n), r)$.

(ii) Supposed there is a constant $c_2 > 0$ with

$$\begin{aligned} \|\delta_n\| &\le c_2 h_n \text{ for } n \ge k, \\ \|Q_{CRV}^T \delta_{ne}\| + \|\bar{Q}_{V-C}^T \delta_{nV}\| &\le c_2 h_n^2 \text{ for } n \ge 0, \end{aligned}$$

we find a constant $c_3 > 0$ such that the following error estimation holds:

$$\max_{n\geq k}\|x_*(t_n)-x_n\| \leq c_3\Big[\max_{n<k}\|x_*(t_n)-x_n\| + \max_{n\geq k}\|\delta_n-\tau_n\| + \max_{n\geq 0}\frac{\|Q_{CRV}^T\delta_{ne}\|+\|\bar{Q}_{V-C}^T\delta_{nV}\|}{h_n}\Big].$$

PROOF: Since $N\cap S(\cdot)$ is constant and the errors $Q_{CRV}^{\mathrm{T}}\delta_e$ as well as $\bar{Q}_{V-C}^{\mathrm{T}}\delta_V$ reflect the errors in those constraints that have to be differentiated, we may apply Theorem 4.2 of [Tis96]. Regarding the expressions for $Q_1(\cdot)$ and $G_2(\cdot)$ in the proof of Theorem 1.12, we obtain the assertion immediately. □

Remark 1.15 Theorem 1.14 shows that the BDF methods applied to network equation systems are stable if the circuit does not contain L-I cutsets and C-V loops. In general, they are weekly unstable.

Corollary 1.16 *Suppose that the errors in the initial values are $O(h^k)$ and the errors in the Newton iteration satisfy $O(h^k)$ for all equations as well as $O(h^{k+1})$ for the equations*

$$Q_{CRV}^T(A_L j_L + A_I i(t)) = 0, \tag{1.51}$$
$$\bar{Q}_{V-C}^T(A_V^T e - v(t)) = 0. \tag{1.52}$$

Then, the k-step BDF method applied to network equation systems (1.17)-(1.19) *is convergent and globally accurate of $O(h^k)$.*

This is a simple conclusion from Theorem 1.14 if one regards that the errors $Q_{CRV}^{\mathrm{T}}\delta_e$ and $\bar{Q}_{V-C}^{\mathrm{T}}\delta_V$ correspond to the errors solving the equations (1.51) and (1.52), respectively.

Remark 1.17 Since, the equations (1.51) and (1.52) are linear, they are always solved quite accurately by Newton's method. This explains that the BDF method often works well in practice even for systems of index 2.

In the next chapter, we deal with modeling of semiconductor devices. Particular care is taken to the boundary conditions since they are important for the coupling of network and device equations.

Chapter 2

Semiconductor Device Modeling

Considering the literature, one finds an enormous amount of books and papers dealing with semiconductor device modeling. We refer here only to [Sel84, GG86, GG89, MRS90, Wac91]. These contributions provide a comprehensive overview of the topic and focus onto the mathematical background.

Semiconductor device models describe the electron transport in the the semiconductor. In consideration of the degree of simplification one distinguishes between quantum level transport and semi-classical transport completed with balance equations. The first one yields the Schrödinger equation and the second one leads to the Boltzmann equation. Simplifying the Boltzmann equation further by the method of moments [Sch90], one obtains the so called energy balance equations (considering four moments) or the drift diffusion equations (considering only two moments).

From the practical point of view, the interest in semiconductor device modeling is to replace as much laboratory testing as possible by numerical simulation in order to minimize the costs. Thus, mathematical models requiring expensive simulations are not preferable. For most semiconductor technologies, the drift diffusion equations seem to represent a reasonable compromise between computational efficiency and an accurate description of the underlying physics.

However with the increased miniaturization of semiconductor devices, one comes closer and closer to the limits of validity of the drift diffusion equations. The reason for this is, on one hand, that in ever smaller devices the free carriers can not longer be modeled as a continuum. On the other hand,

the drift diffusion equations are derived through a limiting process where the mean free path of a particle tends to zero. Through miniaturization this mean free path becomes larger and larger in comparison to the size of the device. In addition, quantum mechanical effects play a more and more important role in novel device structures.

Nevertheless, the drift diffusion equations remain an important tool since microscopic effects not described by them appear only locally. Thus, the most likely approach will be to use more sophisticated models only locally, and to use the drift diffusion equations in the parts of the device where they are sufficient to describe the physics (usually in the bulk of the semiconductor).

Therefore, we concentrate in Section 2.1 on the drift diffusion equations considered as an important model description for the device part in coupled network and device simulation.

Regarding the coupling between network and device simulation, the type of contacts between the semiconductor and the network has to be taken into account. Usually, they consist of layers of metal, insulator, or other semiconductors. This implies that the boundary conditions for the transport equations in the bulk of the semiconductor have to be formulated in such a way that the physical processes at the interfaces are described appropriately. We will discuss them in Section 2.2, but for a more detailed description we refer to [Sch90, Sze81].

Finally, we present the geometry and doping concentration for two typical examples of semiconductor devices in Section 2.3.

2.1 The Drift Diffusion Model for Semiconductor Devices

In this part we describe the drift diffusion model for semiconductor devices including a brief derivation. For a more detailed discussion see e.g. [Sel84, Sze81].

First, we introduce Ω to be a nonempty, open and bounded domain with a regular boundary $\Gamma = \partial\Omega$ in $\mathbb{R}^N$ with $1 \leq N \leq 3$ such that $\bar{\Omega}$ describes the range of the semiconductor inclusive its contacts.

2.1.1 Current-Density Equations

The conductivity of semiconductors is strongly connected to the number of its free charge carriers. As charge carriers, we have to consider not only electrons but also holes. If a semiconductor atom lacks one of its valence electrons, then it may attract an electron from another atom. This can be considered as a movement of a hole from one atom to another one. Correspondingly, holes are considered as positive charge carriers, whereas electrons are negative charge carriers.

As the name of the drift-diffusion model already expresses, the current in a semiconductor is mainly driven by drift and diffusion. The drift current is caused by an electric field E that is present due to the existence of free charge carriers. It is given by

$$q\mu_n n\mathrm{E} \qquad \text{and} \qquad q\mu_p p\mathrm{E}$$

for electrons and holes, respectively. Here, q represents the elementary charge. The variables n and p denote the concentrations of electrons and holes, respectively. The electron and hole mobilities, μ_n and μ_p are bounded, strictly positive functions depending on semiconductor material, doping, temperature and the electric field E.

The diffusion current is caused by a movement of charge carriers that aims to compensate inhomogeneous concentrations. The diffusion current is proportional to the gradient of the charge carrier concentration. More precisely, we have

$$qD_n \mathrm{grad}\, n \qquad \text{and} \qquad -qD_p \mathrm{grad}\, p.$$

D_n and D_p are called carrier diffusivities. In general, they are bounded, strictly positive functions depending on semiconductor material, doping and temperature.

In thermal equilibrium, the mobilities μ_p, μ_n and the diffusivities D_n, D_p related by

$$D_n = \frac{kT}{q}\mu_n \qquad \text{and} \qquad D_p = \frac{kT}{q}\mu_p$$

for non-degenerate semiconductors. Here, T denotes the temperature and k is the Boltzmann constant. The last equations are called Einstein relations.

Since the electric field E is related to the electrostatic potential V by

$$\mathrm{E} = -\mathrm{grad}\, V, \tag{2.1}$$

we obtain for the current densities of electrons and holes

$$J_n = -q\mu_n n \operatorname{grad} V + qD_n \operatorname{grad} n, \tag{2.2}$$
$$J_p = -q\mu_p p \operatorname{grad} V - qD_p \operatorname{grad} p. \tag{2.3}$$

Note that one has to consider an additional current, if a magnetic field is applied to the semiconductor. However, this is usually negligible for devices included in integrated circuits.

2.1.2 Continuity Equations

The continuity equations describe particle conservation and are given by

$$-q\partial_t n + \operatorname{div} J_n = qR, \tag{2.4}$$
$$q\partial_t p + \operatorname{div} J_p = -qR. \tag{2.5}$$

Here, R describes the generation/recombination rate. There are several physical mechanisms causing generation and recombination of electrons and holes. The main ones are phonon transitions, photon transitions, Auger (three particle) transitions and impact ionization. Corresponding to the different mechanisms, different models have been developed in order to describe the generation and recombination process. The mostly used models are *Shockley-Read-Hall recombination*

$$R_{\mathrm{RSH}} = \frac{np - n_i^2}{\tau_p(n + n_i) + \tau_n(p + n_i)},$$

Optic recombination

$$R_{\mathrm{OPT}} = C^{\mathrm{OPT}}(np - n_i^2),$$

Auger recombination

$$R_{\mathrm{AU}} = (C_n^{\mathrm{AU}} n + C_p^{\mathrm{AU}} p)(np - n_i^2),$$

and *impact ionization*

$$R_{\mathrm{II}} = -\alpha_n \frac{\|J_n\|}{q} - \alpha_p \frac{\|J_p\|}{q}.$$

Here, n_i represents the intrinsic charge density. If the semiconductor is in equilibrium, then np is constant and n_i is defined by

$$n_i^2 = np.$$

The factors τ_n and τ_p reflect the average lifetimes of electrons and holes, respectively. The constants C^{OPT}, C_n^{AU} and C_p^{AU} have to be determined by experiments. α_n and α_p are the ionization rates for electrons and holes, respectively. They may be approximated by

$$\begin{aligned} \alpha_n &= \alpha_n^{\infty} \cdot \exp\left(-\left(\frac{E_n^{\mathrm{crit}}}{E}\right)^{\beta_n}\right), \\ \alpha_p &= \alpha_p^{\infty} \cdot \exp\left(-\left(\frac{E_p^{\mathrm{crit}}}{E}\right)^{\beta_p}\right). \end{aligned}$$

with constants α_n, α_p, β_n, β_p, E_n^{crit} and E_p^{crit}.

If all effects are present, then one simply adds up all rates such that

$$R = R_{\mathrm{RSH}} + R_{\mathrm{OPT}} + R_{\mathrm{AU}} + R_{\mathrm{II}}.$$

2.1.3 Poisson Equation

The transport equations (2.2)-(2.5) constitute equations for the concentrations of electrons and holes (n and p) as well as the densities of electron and hole current (J_n and J_p). Additionally, the existence of these charge carriers causes an electrical field. In order to obtain a self-consistent formulation, the transport equations have to be completed by an equation that determines this electrical field. This is given by the third Maxwell equation, which relates the electric field to the electric charges. It reads

$$\operatorname{div} \mathrm{D} = \varrho \tag{2.6}$$

where D is the electric displacement and ϱ is the charge density. The electric charge is the source of the electric displacement. The electric field E is related to D by

$$\mathrm{D} = \varepsilon \mathrm{E} \tag{2.7}$$

with the permittivity constant ε of the medium if the medium is homogeneous. Inserting (2.7) and (2.1) into (2.6), we get

$$\operatorname{div}(-\varepsilon \operatorname{grad} V) = \varrho.$$

In a semiconductor, the local charge is composed of electrons, holes, donor atoms and acceptor atoms. Thus, the charge ϱ is given by

$$\varrho = q(p - n + N_D^+ - N_A^-),$$

where q is the elementary charge, N_D^+ the donor concentration and N_A^- the acceptor concentration. Finally, we arrive at the Poisson equation

$$\operatorname{div}(-\varepsilon \operatorname{grad} V) = q(p - n + N_D^+ - N_A^-). \tag{2.8}$$

The impurity atoms are assumed to be fixed in the semiconductor, i.e. N_D^+ and N_A^- are independent of time and, thus, given as functions of the position. This assumption is justified if the impurity concentrations are sufficiently small. But it holds no longer true in case of high power transistors and high power diodes.

2.1.4 Complete Drift-Diffusion Model

Summarizing the model equations (2.2), (2.3), (2.4), (2.5) and (2.8), we get the drift-diffusion model equations

$$\begin{aligned}
\operatorname{div}(\varepsilon \operatorname{grad} V) &= q(n - p - N), && (2.9)\\
-\partial_t n + \tfrac{1}{q} \operatorname{div} J_n &= R, && (2.10)\\
\partial_t p + \tfrac{1}{q} \operatorname{div} J_p &= -R, && (2.11)\\
J_n &= q(D_n \operatorname{grad} n - \mu_n n \operatorname{grad} V), && (2.12)\\
J_p &= q(-D_p \operatorname{grad} p - \mu_p p \operatorname{grad} V). && (2.13)
\end{aligned}$$

The unknowns are the electrostatic potential V, the electron and hole concentrations, n and p, as well as the current densities of electrons and holes, J_n and J_p. Note, that J_n and J_p are given by (2.12)-(2.13). Thus, inserting (2.12) and (2.13) into (2.10) and (2.11) yields to a system in the primary variables V, n and p only.

The doping concentration $N := N_D^+ - N_A^-$ represents a given function depending only on the position variable x. The sizes ε and q are constants. The mobilities μ_n and μ_p as well as the diffusivities $\mathcal{D}_n$ and D_p are bounded, strictly positive functions. They may depend on position x (due to dependency on doping) and on the gradient of the potential $\operatorname{grad} V$ (due to dependency on the electric field E. Finally, the generation/recombination rate R may depend on n, p, J_n, J_p and $\operatorname{grad} V$ corresponding to the applied model.

The system (2.9)-(2.13) represents a system of five coupled partial differential equations. The Poisson equation (2.9) is of elliptic type. Regarding the current density equations (2.12) and (2.13), the continuity equations (2.10) and (2.11) are of parabolic type.

Note that we assume a constant temperature. It is justified for applications with low performance devices. In case of high performance devices, one has to consider the temperature T as a variable. The drift diffusion equations have to be completed by an energy balance equation (see e.g. [Wac95, Jün01, AGH02]). It is a future task to combine such energy models with the network equations.

2.2 The Boundary Conditions

Semiconductors have essentially three different types of adjoining materials. The contacts between the network and the semiconductor are usually layers of metal. The second kind of bounding materials are insulators (e.g. oxide). Finally, they may be bounded by other semiconductors. Such semiconductor-semiconductor interfaces are called heterojunction. Here, we are interested in devices with one semi-conducting material only and we do not consider heterojunction.

The following sections are devoted to a brief explanation of the boundary conditions connected with different types of interfaces. For a more detailed description, we recommend the books [Sze81, Sel84, Sch90].

2.2.1 Metal-Semiconductor Contacts

Many semiconductor devices have low resistance or rectifying contacts. The corresponding models are called Ohmic and Schottky contacts, respectively. They have been established in many device simulation programs and we want to consider such contacts here. Note that, in [Sch90], a new model was presented for non-ideal contacts including tunneling effects (regarded in Ohmic contact models) as well as thermionic emissions (regarded in Schottky contact models) and generalizing both types of metal contact models. This results, in general, in mixed boundary conditions which are coupled in a highly nonlinear manner.

Ohmic Contacts

Ohmic contacts are characterized by a high doping of the semiconductor. This implies a large band bending and a very thin barrier at the metal-semiconductor interface. In this case, tunneling of electrons is the dominant

transport mechanism. It leads to high current densities at low voltage drops and, consequently, to a low resistance of the contact.

Since tunneling is not included in the drift diffusion equations describing electron transport in the semiconductor volume, one should place the actual boundary for the simulation domain at the end of the tunneling region. At high doping concentrations, the tunneling length comprises the total depletion region, and the boundary is placed at the depletion layer edge. Consequently, we have charge neutrality at the actual boundary that means

$$n - p - N = 0.$$

Furthermore, the electrostatic potential at the boundary is given by

$$V = V_{\text{ap}} + V_{\text{bi}}, \tag{2.14}$$

where V_{ap} is the applied voltage and V_{bi} is the so called built-in potential of the semiconductor. The built-in potential depends on semiconductor material, doping concentration, and temperature.

For very high doping (ideal ohmic contact), the resistance tends to zero which implies [Sch90]

$$np = n_i^2$$

with the intrinsic concentration n_i depending on material and temperature. This leads to Dirichlet boundary conditions for the electron and hole concentrations

$$n = \frac{1}{2}(\sqrt{N^2 + 4n_i^2} + N), \tag{2.15}$$

$$p = \frac{1}{2}(\sqrt{N^2 + 4n_i^2} - N). \tag{2.16}$$

Schottky Contacts

Schottky contacts have a rectifying behavior. This is caused by low semiconductor doping which leads to a slow weak band bending and a thick barrier at the interface. Therefore, thermionic emission of electrons is the dominant transport mechanism here.

In this case, the carriers crossing the interface have to overcome the barrier height ϕ_{B} arising from the band bending. This implies

$$V = V_{\text{ap}} + V_{\text{bi}} + V_{\text{B}}, \tag{2.17}$$

where V_{ap}, V_{bi} are as in (2.14) and

$$qV_{\mathrm{B}} = \phi_{\mathrm{B}}.$$

In zero order approximation, the barrier height ϕ_{B} is a constant, dependent on the combination of the materials.

Denoting ν as the unit outer normal vector on the contact, the boundary condition for the continuity equations read as

$$J_n \cdot \nu = -qv_n(n - n_0), \tag{2.18}$$

$$J_p \cdot \nu = qv_p(p - p_0), \tag{2.19}$$

where v_n and v_p are the recombination velocities depending on material and temperature. The quantities n_0 and p_0 are the *quasi-equilibrium* concentrations. They depend on barrier height and temperature.

Note that we have Dirichlet boundary conditions for the potential V and, consequently, mixed boundary conditions for the concentrations n and p.

2.2.2 Semiconductor-Insulator Interface

As a consequence of Maxwell's third law (2.6), we obtain as boundary condition for the Poisson equation

$$\varepsilon \frac{\partial V}{\partial \nu} - \varepsilon_i \frac{\partial V_i}{\partial \nu} = \sigma,$$

where ε and ε_i are the dielectric constants of the semiconductor and the insulator, respectively, and σ is the surface charge at the interface.

Regarding the existence of surface recombination, we obtain

$$J_n \cdot \nu = -qR_{\mathrm{surf}} \quad \text{and} \quad J_p \cdot \nu = qR_{\mathrm{surf}} \tag{2.20}$$

with

$$R_{\mathrm{surf}} = \frac{np - n_i^2}{\frac{1}{s_p}(n + n_1) + \frac{1}{s_n}(p + p_1)}$$

for the continuity equations. Here, n and p denote the electron and hole concentration at the contact, respectively. The recombination velocities s_n and s_p as well as the concentrations n_1, p_1, and the intrinsic carrier concentration n_i are parameters depending on the material of the semiconductor and the effective doping. By ν we denote the outer unit normal vector. If surface recombination can be neglected, then we arrive at

$$J_n \cdot \nu = 0, \qquad J_p \cdot \nu = 0. \tag{2.21}$$

2.3 Device Examples

This section is devoted to an illustration of geometry and doping for typical semiconductors. We start with a p-n diode model which belongs to the basic semiconductor devices. The second example describes a MOSFET (Metal Oxide Semiconductor Field Effect Transistor). It is one the most important semiconductor devices in chip industry since it can be used as a switch that needs almost no power.

2.3.1 Diode

A diode is a semiconductor element that has rectifying properties. The current flows mainly in one direction. In Figure 2.1, an example of a p-n diode is depicted. Here, Ω^+/Ω^- represent the positive(p)/negative(n) doping regions of the diode. Γ_1 and Γ_2 constitute metal contacts connecting the semiconductor with the network. Depending on the doping, they are of Ohmic or Schottky type. The rest of the boundary of the diode can be considered as electrically neutral.

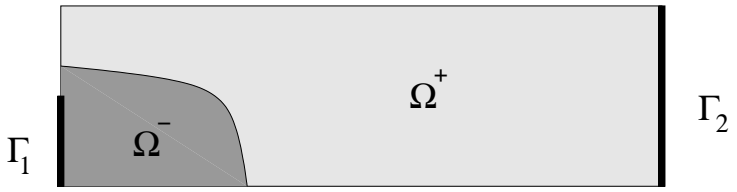

Figure 2.1: Two-dimensional cross section of a p-n diode

If the applied potential at Γ_1 is lower than the potential applied to Γ_2, then electrons from Ω^- and holes from Ω^+ accumulate at the boundary between Ω^- and Ω^+. Consequently, negative charge carriers from the n-region move to the p-region, positive charge carriers from the p-region move to the n-region and a significant current flows through the diode. It usually increases exponentially with an increasing voltage difference. However, if the applied potential at Γ_1 is lower than the potential applied to Γ_2, then the depletion zone at the Ω^-/Ω^+ boundary is enlarging. Only a few free charge carriers are situated there and the current flowing through the diode is almost zero.

2.3.2 MOSFET

A MOSFET (Metal Oxide Semiconductor Field Effect Transistor) is used as a switch for digital circuits. Figure 2.2 shows a two-dimensional cross section of an n-channel MOSFET. Ω_i^-/Ω^+ (i=1,2) represent again the negative/positive doping regions. The n-regions Ω_i^- are usually highly doped. The p-region Ω^+ has a lower doping concentration. Source and drain are always Ohmic metal contacts. Depending on technology, the bulk contact may be an insulator or a Schottky metal contact. The gate interface accomplished by the oxide layer constitutes a metal-insulator-semiconductor contact. The rest of the boundary can be considered as insulating boundaries.

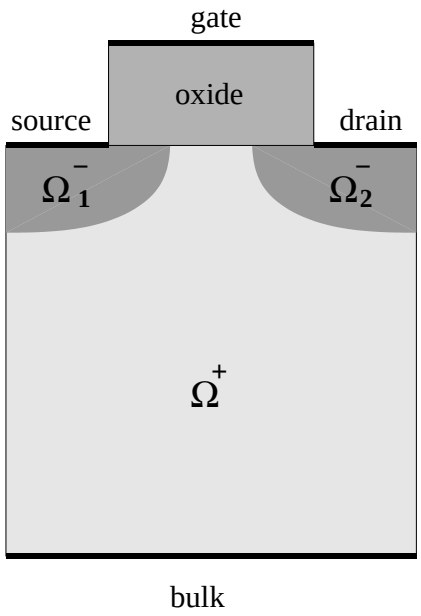

Figure 2.2: Two-dimensional cross section of a MOSFET

Application of a sufficiently high voltage at the gate contact creates a large electrical field in a channel normal to the oxide layer surface (see Figure 2.3). Consequently, electrons from Ω^+ are accumulated in a thin channel close to the oxide layer. Although Ω^+ has a positive doping, the electron density dominates the hole density in this channel. Therefore, it is called inversion layer. Applying a voltage between source and drain makes the charge carriers moving towards the drain along the channel and parallel to the oxide layer. Finally, a current moves from source to drain. Thus, the source-drain current can be switched on and off by applying different voltages to the gate.

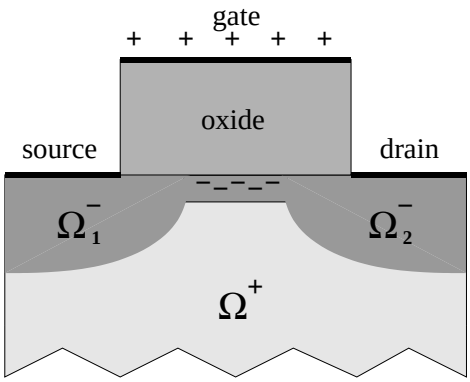

Figure 2.3: MOSFET channel for charge carrier transport if a high potential is applied to gate.

The next chapter is devoted to the coupling of the network and device model equations. On the one hand, the boundary conditions of the semiconductor equations depend on the node potentials of the network. On the other hand, the network currents flowing through a semiconductor depend on the current densities for holes and electrons as well as the gradient of the electrostatic potential of the semiconductor.

Chapter 3

Coupling of the Network and Device Model Equations

From the engineering point of view, the coupling of network and device simulation is not a new task (see e.g. [ELD82, MP92, LSK$^+$96, Sch96, ESTZ96, Rot00]).

In contrast, the mathematical analysis of coupled network and device model equations represents a very young research field. First results have already been obtained in [Grö87] and [GG89]. In [Grö87], a semiconductor connected to a resistor has been considered. In [GG89], the connection of a semiconductor with a simple circuit has been studied. Simple means here that the currents entering the semiconductor may be expressed by a function of the applied voltages. In both cases, the network is treated as a special boundary condition for the semiconductor. This approach fails if more than one device belongs to the network. This is, in particular, the case for integrated circuits. However, as we will see later in this chapter, the approach may be extended by a modification of the considered operator equation.

More recently, networks containing uniform lossy transmission lines have been investigated in [Gün01b] and [Gün01a]. The resulting equation systems represent also coupled systems of differential algebraic equations and partial differential equations. In contrast to the case here, the PDEs are of hyperbolic type.

To the authors knowledge, an existence analysis for integrated networks containing semiconductor device models has only been developed in [ABGT02]. There, the stationary drift diffusion model for diodes has been considered. Here, we are dealing for the first time with the instationary case.

For brevity, we will consider a network that contains exactly one semiconductor first. One can simply verify that the following arguments remain true if one considers a network with several semiconductors. At the end of Section 3.3, we will formulate the coupled system for networks with n_S semiconductors.

3.1 The Coupling Conditions

As we have seen in Chapter 1, the network equations for linear electric networks including capacitors, inductors, resistors, independent voltage and current sources are described by

$$\begin{aligned} A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}} e, t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V + A_I i_s &= 0, \quad (3.1) \\ \frac{\mathrm{d}\phi_L(j_L, t)}{\mathrm{d}t} - A_L^{\mathrm{T}} e &= 0, \quad (3.2) \\ A_V^{\mathrm{T}} e - v_s &= 0. \quad (3.3) \end{aligned}$$

The unknowns are the nodal potentials e, the currents of inductors j_L and the currents of voltage sources j_V. The matrices A_C, A_L, A_R, A_I and A_V are constant and span the incidence matrix. Thus, they are constant and have the entries -1, 0, and 1 only. Finally, i_s and v_s are input functions of time only. The small s shall indicate that the input functions describe wave forms of sources.

Equation (3.1) represents the KCL for each node except the mass node. Equation (3.2) and (3.3) describe the element characteristics of inductors and voltage sources, respectively.

The task is now to include the semiconductor device that is described by the drift-diffusion model. Consequently, the currents of the semiconductor device have to be added to the KCL equation (3.1). We denote the vector of all branch currents leaving the semiconductor device by $\hat{j}_S$. Let b_S be the number of terminals of the semiconductor. Again, n is chosen to be the number of all nodes of the network. Furthermore, the mass node has number n. Then, we introduce the matrix $\hat{A}_S \in L(\mathbb{R}^{b_S}, \mathbb{R}^{n-1})$ with the entries

$$\hat{a}_{ik} := \begin{cases} 1 & \text{if the current } j_{S_k} \text{ enters node } i, \\ 0 & \text{else.} \end{cases} \quad (3.4)$$

This way, the matrix $\hat{A}_S$ describes the incidence of currents of the semiconductor device. However, it differs from the other incidence matrices A_N

(N corresponds to one of the basic network elements) by the fact that each column contains only entries 1 but no -1. We arrive at the system

$$\begin{aligned} A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}}e,t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}},t)e + A_L j_L + A_V j_V + \hat{A}_S \hat{j}_S + A_I i_s &= 0, \quad (3.5) \\ \frac{\mathrm{d}\phi_L(j_L,t)}{\mathrm{d}t} - A_L^{\mathrm{T}} e &= 0, \quad (3.6) \\ A_V^{\mathrm{T}} e - v_s &= 0. \quad (3.7) \end{aligned}$$

In Chapter 2 we have seen that the semiconductor equations are given by

$$\begin{aligned} \mathrm{div}\,(\varepsilon \mathrm{grad}\, V) &= q(n-p-N), & (3.8) \\ -\partial_t n + \tfrac{1}{q}\,\mathrm{div}\, J_n &= R(n,p,J_n,J_p), & (3.9) \\ \partial_t p + \tfrac{1}{q}\,\mathrm{div}\, J_p &= -R(n,p,J_n,J_p), & (3.10) \\ J_n &= q(D_n \mathrm{grad}\, n - \mu_n n \mathrm{grad}\, V), & (3.11) \\ J_p &= q(-D_p \mathrm{grad}\, p - \mu_p p \mathrm{grad}\, V). & (3.12) \end{aligned}$$

Here, the unknowns are the electrostatic potential V, the electron concentration n, the hole concentration p, the electron current density J_n and the hole current density J_p.

How are the device equations (3.8)-(3.12) connected with the network equations (3.5)-(3.7)? This depends on the kind of contacts. We have to distinguish between metal–semiconductor contacts and metal–insulator–semiconductor contacts. For example, a MOSFET (see Section 2.3.2) has usually three metal–semiconductor contacts (source, drain, bulk) and one metal–insulator–semiconductor contact (gate).

We consider metal–semiconductor contacts first. Let Γ_k be a metal–semiconductor contact. The index k shall indicate that this contact represents terminal k of the semiconductor. Then, the current flowing through terminal k is given by

$$\hat{j}_{S_k} = \int_{\Gamma_k} J_{\mathrm{tot}} \cdot \nu \, \mathrm{d}\sigma.$$

J_{tot} represents the total current density and ν is the outward unity normal vector of Γ_k. In the semiconductor we meet three types of currents: the current of electrons, the current of holes, and the displacement current caused by the electrostatic potential. Consequently, the total current density is given by

$$J_{\mathrm{tot}} = J_n + J_p - \varepsilon\, \partial_t \,\mathrm{grad}\, V.$$

This implies

$$\hat{j}_{S_k} = \int_{\Gamma_k} (J_n + J_p - \varepsilon\, \partial_t \operatorname{grad} V) \cdot \nu \, \mathrm{d}\sigma. \tag{3.13}$$

Note that the displacement current disappears if one considers the stationary case only. But it can not be neglected in the non-stationary case in order to guarantee charge conservation. See Section 3.4 for more details about this.

The semiconductor current represents only one part of the coupling. The other part is given by the boundary condition for the electrostatic potential V. Regarding the results in Section 2.2, we find that

$$V(x,t) = V_{\mathrm{ap}}(t) + W(x) \qquad \text{on } \Gamma_k \tag{3.14}$$

where

$$W(x) = \begin{cases} V_{\mathrm{bi}}(x) & \text{for Ohmic contacts,} \\ V_{\mathrm{bi}}(x) + V_{\mathrm{B}} & \text{for Schottky contacts.} \end{cases}$$

Here, $V_{\mathrm{ap}}(t)$ is the applied potential at time t. It equals $e_j(t)$ if terminal k (corresponding to Γ_k) enters node number j. Regarding equation (3.4), we find that

$$V_{\mathrm{ap}}(t) = \begin{cases} 0 & \text{if terminal } k \text{ enters the mass node,} \\ \hat{a}_k^{\mathrm{T}} e(t) & \text{else} \end{cases}$$

for $\hat{a}_k^{\mathrm{T}} := (\hat{a}_{1k}, ..., \hat{a}_{n-1,k})$. Introducing the vector $c(x) = (c_1(x), ..., c_{b_S}(x))$ with the entries

$$c_k(x) = \begin{cases} 1 & \text{if } x \text{ belongs to } \Gamma_k, \\ 0 & \text{else,} \end{cases}$$

we may write equation (3.14) as

$$V(x,t) = c(x) \cdot \hat{A}_S^{\mathrm{T}} e(t) + W(x) \qquad \text{on } \Gamma. \tag{3.15}$$

We turn to metal-insulator-semiconductor interfaces (see Figure 3.1). Let Γ_k denote the boundary between the semiconductor and the insulator. Furthermore, we denote the boundary between the metal and the insulator by Γ_{MI}. Since the insulator is free of charge carriers, the particle current normal to the boundary equals zero, that means

$$(J_n + J_p) \cdot \nu = 0$$

on Γ_k and on Γ_{MI}. Consequently,

$$\hat{j}_{S_k} = -\int_{\Gamma_{\mathrm{MI}}} \varepsilon_{\mathrm{i}k}\, \partial_t \operatorname{grad} V_{\mathrm{i}k} \cdot \nu \, \mathrm{d}\sigma,$$

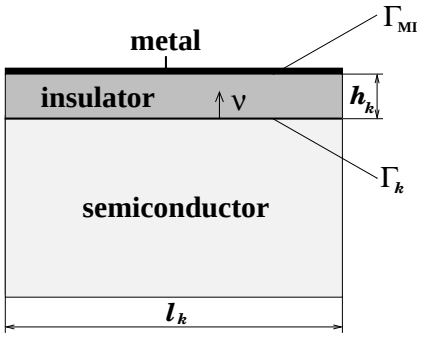

Figure 3.1: Metal-insulator-semiconductor interface

where V_{ik} represents the electrostatic potential V_i in the insulator. Due to the absence of charge carriers, we get

$$\Delta V_{ik} = \operatorname{div} \operatorname{grad} V_{ik} = 0. \tag{3.16}$$

If the thickness h_k of the insulator is sufficiently small with respect to the length (in $\mathbb{R}^2$) or diameter (in $\mathbb{R}^3$) l_k of the insulator, we may assume that V_{ik} varies in the direction of ν only. Consequently, we may solve the Poisson equation in one dimension in the insulator. It yields

$$\operatorname{grad} V_{ik}(x + h_k\nu, t) \cdot \nu = \operatorname{grad} V_{ik}(x, t) \cdot \nu$$

and

$$V_{ik}(x + h_k\nu, t) = V_{ik}(x, t) + h_k \operatorname{grad} V_{ik}(x, t) \cdot \nu \tag{3.17}$$

for all $x \in \Gamma_k$. We know from Section 2.2 that

$$\varepsilon \operatorname{grad} V \cdot \nu - \varepsilon_{ik} \operatorname{grad} V_{ik} \cdot \nu = \sigma_k \tag{3.18}$$

for all $x \in \Gamma_k$. Regarding that the surface charge σ_k on Γ_k depends only on the material (i.e., not on time), we get

$$\varepsilon_{ik}\, \partial_t \operatorname{grad} V_{ik}(x + h_k\nu, t) \cdot \nu = \varepsilon\, \partial_t \operatorname{grad} V(x, t) \cdot \nu$$

for all $x \in \Gamma_k$ and, finally,

$$\hat{j}_{S_k} = -\int_{\Gamma_k} \varepsilon\, \partial_t \operatorname{grad} V \cdot \nu \, \mathrm{d}\sigma. \tag{3.19}$$

It remains to determine the boundary conditions for the potential V. Using (3.17) and (3.18), we get

$$V_{ik}(x, t) = V_{ik}(x + h_k\nu, t) - \frac{h_k}{\varepsilon_{ik}}(\varepsilon \operatorname{grad} V(x, t) \cdot \nu - \sigma_k).$$

At the metal contact, the potential is given by

$$V_{\mathrm{i}k}(x + h_k\nu, t) = V_{\mathrm{ap}}(t) + V_{\mathrm{bi}}(x)$$

for all $x \in \Gamma_k$. Here, V_{bi} denotes again the built-in potential of the semiconductor which depends on the material, on doping and on temperature. The applied potential reads again as

$$V_{\mathrm{ap}}(t) = c(x) \cdot \hat{A}_S^{\mathrm{T}} e(t). \tag{3.20}$$

Using the continuity of the potential at Γ_k, that means

$$V_{\mathrm{i}k}(x, t) = V(x, t) \qquad \text{on} \quad \Gamma_k,$$

we get

$$V(x, t) = V_{\mathrm{ap}}(t) + V_{\mathrm{bi}}(x) - \frac{h_k}{\varepsilon_{\mathrm{i}k}}(\varepsilon \operatorname{grad} V(x, t) \cdot \nu - \sigma_k).$$

Equivalently, we have mixed boundary conditions

$$\varepsilon \operatorname{grad} V(x, t) \cdot \nu + \alpha(x) V(x, t) = \alpha(x) V_{\mathrm{ap}}(t) + \beta(x)$$

with

$$\alpha(x) := \frac{\varepsilon_{\mathrm{i}k}}{h_k} \quad \text{and} \quad \beta(x) := \sigma_k + \alpha V_{\mathrm{bi}} \qquad \forall\, x \in \Gamma_k.$$

Regarding (3.20), we get

$$\varepsilon \operatorname{grad} V(x, t) \cdot \nu + \alpha(x) V(x, t) = \alpha(x) c(x) \cdot \hat{A}_S^{\mathrm{T}} e(t) + \beta(x). \tag{3.21}$$

Note that for short channel transistors, the thickness of the insulator is not sufficiently small with respect to the length. In this case, one has solve the Poisson equation in the insulator numerically. More precisely, we would have to add equation (3.16) to the coupled system and to complete the model by corresponding boundary conditions. This would be also necessary if particle currents in the insulator can not be neglected (e.g. for MIS tunnel diodes).

Since we are more interested in the analysis of the device-network coupling, we shall concentrate on the case where (3.19) and (3.21) are satisfied. We do not expect fundamental differences for the case where the system is extended by the Poisson equation for the insulator.

3.2 Charge Conservation

As mentioned before, the presented device model is charge conserving. In fact, we have

$$\operatorname{div}(J_n + J_p) = q(\partial_t n - \partial_t p)$$

if we add the continuity equations (3.9) and (3.10). Additionally,

$$\operatorname{div}(\varepsilon\, \partial_t \operatorname{grad} V) = q(\partial_t n - \partial_t p),$$

if we differentiate the Poisson equation (3.8) with respect to time. This implies

$$\operatorname{div} J_{\text{tot}} = \operatorname{div}(J_n + J_p - \varepsilon\, \partial_t \operatorname{grad} V) = 0$$

and, by Gauss law,

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \oint_\Gamma J_{\text{tot}} \cdot \nu \, \mathrm{d}\sigma = \int_\Omega \operatorname{div} J_{\text{tot}} \, \mathrm{d}x = 0.$$

Furthermore, we get for the sum of all currents leaving the semiconductor

$$\begin{aligned}\sum_k \hat{j}_{S_k} \;=\; & \sum_{\substack{k \\ \Gamma_k \subseteq \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}}}} \int_{\Gamma_k} (J_n + J_p - \varepsilon\, \partial_t \operatorname{grad} V) \cdot \nu \, \mathrm{d}\sigma \\ & - \sum_{\substack{k \\ \Gamma_k \subseteq \Gamma_{\mathrm{MI}}}} \int_{\Gamma_k} \varepsilon\, \partial_t \operatorname{grad} V \cdot \nu \, \mathrm{d}\sigma = \oint_\Gamma J_{\text{tot}} \cdot \nu \, \mathrm{d}\sigma = 0.\end{aligned}$$

This means, that the current flowing through one terminal of the semiconductor may be described by the negative sum of the currents flowing through the other terminals. We choose one terminal (usual the bulk terminal) and call it the reference terminal. We delete the current of the reference terminal from $\hat{j}_S$ and denote the resulting vector by j_S. This implies

$$\hat{A}_S \hat{j}_S = A_S j_S$$

if the entries of the matrix A_S are defined as

$$a_{ik} := \begin{cases} 1 & \text{if the current } j_{S_k} \text{ enters node } i, \\ -1 & \text{if the reference terminal is connected to node } i, \\ 0 & \text{else} \end{cases}$$

for $i = 1, ..., n-1$ and $k = 1, ..., b_S - 1$. Recall that n is the number of nodes of the network and b_S is the number of terminals of the semiconductor.

This way, A_S has the same form as the other incidence matrices A_N for $N \in \{C, R, L, V, I\}$. Including the row for the mass node, each column of A_S has exactly one 1 and one -1. Finally, we have

$$A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}} e, t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}} e, t) + A_L j_L + A_V j_V + A_S j_S + A_I i_s = 0 \quad (3.22)$$

instead of equation (3.5). Regarding (3.13) and (3.19), we obtain

$$j_S = \int_\Gamma [(J_n + J_p) \cdot \nu \, \chi_1 - \varepsilon \, \partial_t \operatorname{grad} V \cdot \nu \, \chi_2] \, \mathrm{d}\sigma \quad (3.23)$$

if the vector valued functions χ_1 and χ_2 are defined as

$$\chi_{1k}(x) = \begin{cases} 1 & \text{if } x \in \Gamma_k \text{ and } \Gamma_k \subseteq (\Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}}), \\ 0 & \text{else,} \end{cases}$$

$$\chi_{2k} = \begin{cases} 1 & \text{if } x \in \Gamma_k \text{ and } \Gamma_k \subseteq (\Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \cup \Gamma_{\mathrm{MI}}), \\ 0 & \text{else,} \end{cases}$$

and

$$\int_\Gamma v(x) \, \mathrm{d}x := \begin{pmatrix} \int_\Gamma v_1(x) \, \mathrm{d}x \\ \vdots \\ \int_\Gamma v_m(x) \, \mathrm{d}x \end{pmatrix}$$

for any vector valued function v with $v(x) \in \mathbb{R}^m$.

Note that, in case of several semiconductors, one has to choose one reference terminal for each semiconductor and to follow the procedure for each semiconductor.

3.3 Complete Coupled System

At this point, we have all informations collected in order to formulate the complete coupled network-device equation system. We have seen that the coupling conditions depend mainly on the kind of contacts of the device. Therefore, we introduce Γ_{O}, Γ_{S} and Γ_{MI} as the unification of all Ohmic contacts, Schottky contacts and metal-insulator contacts, respectively. The remaining boundaries Γ_{I} do not have a connection to the network and are considered as ideal insulating materials. As usual, we denote the whole boundary of the semiconductor by Γ, that means

$$\Gamma = \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \cup \Gamma_{\mathrm{MI}} \cup \Gamma_{\mathrm{I}}.$$

We start with the network equations (3.22),(3.6)-(3.7)

$$A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}}e,t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}}e,t) + A_L j_L + A_V j_V + A_S j_S + A_I i_s = 0, \quad (3.24)$$

$$\frac{\mathrm{d}\phi_L(j_L,t)}{\mathrm{d}t} - A_L^{\mathrm{T}}e = 0, \quad (3.25)$$

$$A_V^{\mathrm{T}}e - v_s = 0. \quad (3.26)$$

The unknowns e, j_L and j_V depend on time t only. The next part is given by the semiconductor equations (3.8)-(3.12)

$$\begin{aligned}
\operatorname{div}(\varepsilon \operatorname{grad} V) &= q(n-p-N), && (3.27)\\
-\partial_t n + \tfrac{1}{q}\operatorname{div} J_n &= R(n,p,J_n,J_p), && (3.28)\\
\partial_t p + \tfrac{1}{q}\operatorname{div} J_p &= -R(n,p,J_n,J_p), && (3.29)\\
J_n &= q(D_n \operatorname{grad} n - \mu_n n \operatorname{grad} V), && (3.30)\\
J_p &= q(-D_p \operatorname{grad} p - \mu_p p \operatorname{grad} V). && (3.31)
\end{aligned}$$

Here, the variables V, n, p, J_n, and J_p depend on position x and on time t. The third part describes the semiconductor current as follows (see (3.23))

$$j_S = \int_\Gamma [(J_n + J_p)\cdot \nu\, \chi_1 - \varepsilon\, \partial_t \operatorname{grad} V \cdot \nu\, \chi_2]\, \mathrm{d}\sigma. \quad (3.32)$$

The fourth part consists of the boundary conditions for the Poisson equation (3.27), which depend on the network potentials e. Considering (3.15) and (3.21), we find

$$\begin{aligned}
V &= c\cdot \hat{A}_S^{\mathrm{T}} e + W && \text{on } \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}}, && (3.33)\\
\varepsilon \operatorname{grad} V \cdot \nu + \alpha V &= \alpha c \cdot \hat{A}_S^{\mathrm{T}} e + \beta && \text{on } \Gamma_{\mathrm{MI}}, && (3.34)\\
\operatorname{grad} V \cdot \nu &= 0 && \text{on } \Gamma_{\mathrm{I}}. && (3.35)
\end{aligned}$$

Recall that the entries of c satisfy

$$c_k(x) = \begin{cases} 1 & \text{if } x \text{ belongs to } \Gamma_k, \\ 0 & \text{else.} \end{cases}$$

Finally, the system is completed by the boundary conditions for the continuity equations (3.28)-(3.29) (see Section 2.2)

$$\begin{aligned}
n &= \tfrac{1}{2}(\sqrt{N^2+4n_i^2}+N), \quad p = \tfrac{1}{2}(\sqrt{N^2+4n_i^2}-N) && \text{on } \Gamma_{\mathrm{O}}, && (3.36)\\
J_n\cdot\nu &= -qv_n(n-n_0), \quad J_p\cdot\nu = qv_p(p-p_0) && \text{on } \Gamma_{\mathrm{S}}, && (3.37)\\
J_n\cdot\nu &= -qR_{\mathrm{surf}}(n,p), \quad J_p\cdot\nu = qR_{\mathrm{surf}}(n,p) && \text{on } \Gamma_{\mathrm{MI}}, && (3.38)\\
J_n\cdot\nu &= 0, \quad J_p\cdot\nu = 0 && \text{on } \Gamma_{\mathrm{I}}. && (3.39)
\end{aligned}$$

Obviously, they are not directly coupled to the network.

Remark 3.1 If the network contains $n_S > 1$ devices described by the drift-diffusion equations, then V, n, p, J_n, J_p are vector valued functions where V_l, n_l, p_l, J_{ln}, J_{lp} are the unknowns corresponding to the semiconductor with number l ($1 \leq l \leq n_S$). In this case, equations (3.27)-(3.39) are to be understood as n_S times given, i.e. for each set of unknowns V_l, n_l, p_l, J_{ln}, J_{lp} ($1 \leq l \leq n_S$).

We have not specified initial conditions yet. Due to the constraints in the network equation system, we may not specify initial conditions for all variables. Since the constraints are in implicit form and since they involve not only network variables but also device variables, it is not a trivial task to set up correct initial conditions in advance. However, we will see later in this Chapter that the coupled system may be written as an abstract differential algebraic system (ADAS). This will allow us to apply the index concept for ADAS (see [LMT01]), which yields a formulation of suitable initial conditions for the coupled system.

3.4 Homogenization

Considering existence results and Galerkin approaches for partial differential equations, Dirichlet boundary conditions are usually treated by a suitable choice of a function space as solution space. In our case, the boundary conditions (3.33) and (3.34) for the Poisson equation depend on the node potentials that are described implicitly by the network equations. But an implicitly given function space is not very handy. Therefore, we shall homogenize the conditions (3.33) as follows.

If the semiconductor has b_S terminals, then let the terminal b_S be the reference terminal. We choose smooth functions $f_1(x)$, ..., $f_{b_S-1}(x)$ defined on whole Ω such that

$$f_k(x) = \begin{cases} 1 & \text{if } x \text{ belongs to } \Gamma_k \subseteq (\Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \cup \Gamma_{\mathrm{MI}}) \\ 0 & \text{if } x \text{ belongs to } (\Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \cup \Gamma_{\mathrm{MI}}) \backslash \Gamma_k \end{cases} \tag{3.40}$$

and, additionally,

$$\operatorname{grad} f_k \cdot \nu = 0 \qquad \text{on} \quad \Gamma \tag{3.41}$$

for $k = 1, ..., b_S - 1$. For common geometries, one should always find such smooth functions. If $f(x) = (f_1(x), ..., f_{b_S-1}(x))^{\mathrm{T}}$ and the reference terminal Γ_{b_S} enters node j, then we get

$$c \cdot \hat{A}_S^{\mathrm{T}} e = e_j + f \cdot A_S^{\mathrm{T}} e \qquad \text{on} \quad \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \cup \Gamma_{\mathrm{MI}}.$$

Furthermore, let g be a smooth function on whole Ω such that

$$g = W \qquad \text{on } \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}} \tag{3.42}$$

and

$$\operatorname{grad} g \cdot \nu = 0 \qquad \text{on } \Gamma_{\mathrm{MI}} \cup \Gamma_{\mathrm{I}}. \tag{3.43}$$

Introducing

$$\tilde{V}(x,t) := V(x,t) - e_j(t) - f(x) \cdot A_S^{\mathrm{T}} e(t) - g(x), \tag{3.44}$$

we get, for the boundary conditions (3.33)-(3.35),

$$\begin{aligned} \tilde{V} &= 0 && \text{on } \Gamma_{\mathrm{O}} \cup \Gamma_{\mathrm{S}}, && (3.45)\\ \varepsilon \operatorname{grad} \tilde{V} \cdot \nu + \alpha \tilde{V} &= \tilde{\beta} && \text{on } \Gamma_{\mathrm{MI}}, && (3.46)\\ \operatorname{grad} \tilde{V} \cdot \nu &= 0 && \text{on } \Gamma_{\mathrm{I}}, && (3.47) \end{aligned}$$

with $\tilde{\beta} := \beta - \alpha g - \varepsilon \operatorname{grad} g \cdot \nu$. Furthermore, the node potentials e appear in the Poisson equation (3.27) and in the current-density equations (3.30)-(3.31). More precisely, we have

$$\operatorname{div}(\varepsilon \operatorname{grad} \tilde{V}) = q(n - p - N) - \operatorname{div}(\varepsilon \operatorname{grad}(f \cdot A_S^{\mathrm{T}} e + g)) \tag{3.48}$$

as well as

$$\begin{aligned} J_n &= q(D_n \operatorname{grad} n - \mu_n n \operatorname{grad}(\tilde{V} + f \cdot A_S^{\mathrm{T}} e + g)), && (3.49)\\ J_p &= q(-D_p \operatorname{grad} p - \mu_p p \operatorname{grad}(\tilde{V} + f \cdot A_S^{\mathrm{T}} e + g)), && (3.50) \end{aligned}$$

As mentioned in Remark 3.1 V, n, p, J_n, J_p are vector valued functions if the network contains $n_S > 1$ semiconductors. In the following, we will use the notation V_l, n_l, p_l, J_{ln}, J_{lp} for the unknowns of the semiconductor with number l ($1 \le l \le n_S$). All equations in variables with the index l are to be understood as given for each $l = 1, ..., n_S$.

Summarizing all equations we arrive at the homogenized coupled system

$$A_C \frac{\mathrm{d}q_C(A_C^{\mathrm{T}}e,t)}{\mathrm{d}t} + A_R g(A_R^{\mathrm{T}}e,t) + A_L j_L + A_V j_V + A_S j_S + A_I i_s = 0, \quad (3.51)$$

$$\frac{\mathrm{d}\phi_L(j_L,t)}{\mathrm{d}t} - A_L^{\mathrm{T}}e = 0, \quad (3.52)$$

$$A_V^{\mathrm{T}}e - v_s = 0, \quad (3.53)$$

$$\begin{aligned}
\operatorname{div}(\varepsilon_l \operatorname{grad}\tilde{V}_l) &= q(n_l - p_l - N_l) - \operatorname{div}(\varepsilon_l \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}}e + g_l)), && (3.54)\\
-\partial_t n_l + \tfrac{1}{q}\operatorname{div} J_{ln} &= R_l(n_l, p_l, J_{ln}, J_{lp}), && (3.55)\\
\partial_t p_l + \tfrac{1}{q}\operatorname{div} J_{lp} &= -R_l(n_l, p_l, J_{ln}, J_{lp}), && (3.56)
\end{aligned}$$

$$\begin{aligned}
J_{ln} &= q(D_{nl}\operatorname{grad} n_l - \mu_{nl} n_l \operatorname{grad}(\tilde{V}_l + f_l \cdot A_S^{\mathrm{T}}e + g_l)), && (3.57)\\
J_{lp} &= q(-D_{pl}\operatorname{grad} p_l - \mu_{pl} p_l \operatorname{grad}(\tilde{V}_l + f_l \cdot A_S^{\mathrm{T}}e + g_l)), && (3.58)
\end{aligned}$$

$$j_{Sl} = \int_{\Gamma_l} [(J_{ln} + J_{lp}) \cdot \nu\, \chi_{1l} - \varepsilon_l\, \partial_t \operatorname{grad}\tilde{V}_l \cdot \nu\, \chi_{2l}]\, \mathrm{d}\sigma, \quad (3.59)$$

$$\begin{aligned}
\tilde{V}_l &= 0 && \text{on } \Gamma_{l\mathrm{O}} \cup \Gamma_{l\mathrm{S}}, && (3.60)\\
\varepsilon_l \operatorname{grad}\tilde{V}_l \cdot \nu + \alpha_l \tilde{V}_l &= \tilde{\beta}_l && \text{on } \Gamma_{l\mathrm{MI}}, && (3.61)\\
\operatorname{grad}\tilde{V}_l \cdot \nu &= 0 && \text{on } \Gamma_{l\mathrm{I}}, && (3.62)
\end{aligned}$$

$$\begin{aligned}
n_l &= \tfrac{1}{2}(\sqrt{N_l^2 + 4n_{il}^2} + N_l), \quad p_l = \tfrac{1}{2}(\sqrt{N_l^2 + 4n_{il}^2} - N_l) && \text{on } \Gamma_{\mathrm{O}}, && (3.63)\\
J_{ln} \cdot \nu &= -qv_{nl}(n_l - n_{0l}), \quad J_{lp} \cdot \nu = qv_{pl}(p_l - p_{0l}) && \text{on } \Gamma_{\mathrm{S}}, && (3.64)\\
J_{ln} \cdot \nu &= -qR_{l\mathrm{surf}}(n_l, p_l), \quad J_{lp} \cdot \nu = qR_{l\mathrm{surf}}(n_l, p_l) && \text{on } \Gamma_{l\mathrm{MI}}, && (3.65)\\
J_{ln} \cdot \nu &= 0, \quad J_{lp} \cdot \nu = 0 && \text{on } \Gamma_{l\mathrm{I}}. && (3.66)
\end{aligned}$$

In the following chapter we want to study the solution behavior of equation systems like the coupled system described above. Furthermore, we are interested in numerical solutions. For a systematic treatment, we prefer to formulate the coupled system as an abstract differential equation system in Hilbert spaces. Due to the constraints of the network, we will call it as abstract differential algebraic system (ADAS) or abstract DAE. Obviously, there exist several formulations as ADAS. In the next two sections we will develop two of them which we will investigate in Chapter 4 in detail. The first approach follows the standard/classical way. The second one formulates the system as variational equation.

3.5 The Coupled Problem as Abstract DAE

The intention of this section is to formulate the homogenized system (3.51)-(3.66) as an operator equation of the form

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t)+\mathcal{B}(u(t),t)=0 \qquad \text{for all } t\in[t_0,T] \tag{3.67}$$

where $\mathcal{A}$, $\mathcal{B}(\cdot,t)$ and $\mathcal{D}(\cdot,t)$ are operators acting in Hilbert spaces X, Y and Z as follows:

$$\mathcal{A}:\ Z\to Y, \qquad \mathcal{B}(\cdot,t):X\to Y, \qquad \mathcal{D}(\cdot,t):X\to Z. \tag{3.68}$$

The interesting case is $\mathcal{A}$ not bijective on the whole interval $[t_0,T]$. Therefore, systems of the form (3.67) are called as abstract differential algebraic systems. We shall discuss these systems in Chapter 4.

Regarding the homogeneous Dirichlet boundary conditions (3.60) for the potential $\tilde{V}$ on $\Gamma_{\mathrm{O}}\cup\Gamma_{\mathrm{S}}$, we are looking for solutions

$$u(t)=(e(t),j_L(t),j_V(t),j_S(t),\tilde{V}(\cdot,t),n(\cdot,t),p(\cdot,t),J_n(\cdot,t),J_p(\cdot,t))$$

for the coupled system (3.51)-(3.59) that belong to $X=\overset{9}{\underset{i=1}{\mathrm{X}}}X_i$ with

$$\begin{aligned}
&X_1=\mathbb{R}^{n-1}, \quad X_2=\mathbb{R}^{n_L}, \quad X_3=\mathbb{R}^{n_V}, \quad X_4=\overset{n_s}{\underset{l=1}{\mathrm{X}}}\mathbb{R}^{k_l-1}\\
&X_5=\{v\in\overset{n_s}{\underset{l=1}{\mathrm{X}}}H^2(\Omega_l):\ v_l=0 \text{ on } \Gamma_{l\mathrm{O}}\cup\Gamma_{l\mathrm{S}}\},\\
&X_6=X_7=\overset{n_s}{\underset{l=1}{\mathrm{X}}}H^1(\Omega_l), \quad X_8=X_9=\overset{n_s}{\underset{l=1}{\mathrm{X}}}H(\mathrm{div}\,;\Omega_l).
\end{aligned}$$

Here, n, n_L, n_V and n_S denote the number of nodes, inductors, voltage sources and semiconductors, respectively. Furthermore, k_l and Ω_l are the number of terminals and the space region of the semiconductor with number l. We need higher regularity for V since we have to evaluate the gradient of V at the boundary for the determination of the semiconductor current j_S. Furthermore, we choose

$$\begin{aligned}
Y &:= X_1\times X_2\times X_3\times(\overset{n_s}{\underset{l=1}{\mathrm{X}}}L_2(\Omega_l))^5\times X_4,\\
Z &:= \mathbb{R}^{n_C}\times X_2\times X_4\times(\overset{n_s}{\underset{l=1}{\mathrm{X}}}H^1(\Omega_l))^2,
\end{aligned}$$

where n_C denotes the number of capacitors in the network.

We can write the coupled system (3.51)-(3.59) as abstract differential algebraic system of the form (3.68) with

$$\mathcal{A} = \begin{pmatrix} A_C & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \end{pmatrix}, \quad \mathcal{D}(u,t) = \begin{pmatrix} q_C(A_C^{\mathrm{T}} u_1, t) \\ \phi_L(u_2, t) \\ -\mathfrak{r}_{1l}\, u_{5l} \\ u_{6l} \\ u_{7l} \end{pmatrix},$$

and

$$\mathcal{B}(u,t) = \begin{pmatrix} A_R g(A_R^{\mathrm{T}} u_1, t) + A_L u_2 + A_V u_3 + A_S u_4 + A_I i_s(t) \\ -A_L^{\mathrm{T}} u_1 \\ A_V^{\mathrm{T}} u_1 - v_s(t) \\ \mathrm{div}\,(\varepsilon_l \mathrm{grad}\, u_{5l}) - q(u_{6l} - u_{7l} - N_l) + \mathrm{div}\,(\varepsilon_l \mathrm{grad}\,(f_l \cdot A_S^{\mathrm{T}} u_1 + g_l)) \\ -\frac{1}{q}\,\mathrm{div}\, u_{8l} + R_l(u_{6l}, u_{7l}, u_{8l}, u_{9l}) \\ \frac{1}{q}\,\mathrm{div}\, u_{9l} + R_l(u_{6l}, u_{7l}, u_{8l}, u_{9l}) \\ u_{8l} - q(D_{nl} \mathrm{grad}\, u_{6l} - \mu_{nl} u_{7l} \mathrm{grad}\,(u_{5l} + f_l \cdot A_S^{\mathrm{T}} u_1 + g_l)) \\ u_{9l} - q(-D_{pl} \mathrm{grad}\, u_{7l} - \mu_{pl} u_{7l} \mathrm{grad}\,(u_{5l} + f_l \cdot A_S^{\mathrm{T}} u_1 + g_l)) \\ u_{4l} - \mathfrak{r}_{2l}\,(u_{8l} + u_{9l}) \end{pmatrix}$$

where

$$\begin{aligned} \mathfrak{r}_{1l} v &:= \int_{\Gamma_l} \varepsilon_l \,\mathrm{grad}\, v \cdot \nu \,\chi_{l2}\, \mathrm{d}\sigma, \\ \mathfrak{r}_{2l} v &:= \int_{\Gamma_l} v \cdot \nu \,\chi_{l1}\, \mathrm{d}\sigma. \end{aligned}$$

In the matrix notation of $\mathcal{D}(u,t)$ and $\mathcal{B}(u,t)$, all lines containing variables with the index l are to be understood as given for each semiconductor $l = 1, ..., n_S$. Correspondingly, the dimension of the identity operators in $\mathcal{A}$ is chosen such that it suits to $\mathcal{D}(u,t)$. In the next section, we will formulate the generalized problem corresponding to the coupled system. For the generalized formulation, we can apply the Galerkin method which is investigated in the next chapter.

3.6 The Generalized Problem of the Coupled System

The intention of this section is to formulate the homogenized system (3.51)-(3.66) as an operator equation of the form

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t) + \mathcal{B}(u(t),t) = 0 \qquad \text{for all } t \in [t_0, T] \tag{3.69}$$

where $\mathcal{A}(t)$, $\mathcal{B}(\cdot,t)$ and $\mathcal{D}(\cdot,t)$ are bounded operators for all $t \in [t_0, T]$.

Let Ω_l again be the domain of the semiconductor with number l, a bounded region in $\mathbb{R}^N$ with a regular boundary. Following the idea of generalized problems for partial differential equations (see e.g. [Zei90b]), we have to do the following steps.

(i) For each partial differential equation, define a suitable space of test functions.

(ii) Multiply the partial differential equations by a test function, integrate over the domain Ω and use the integration by parts formula.

(iii) Multiply each equation of the network system by a test real value.

(iv) Build the sum over all resulting equations.

We denote again by u all unknowns, i.e.

$$u_1 = e, \quad u_2 = j_L, \quad u_3 = j_V, \quad u_{4l} = j_{Sl}, \quad u_{5l} = \tilde{V}_l, \quad u_{6l} = n_l, \quad u_{7l} = p_l$$

for all $l = 1, ..., n_S$. J_{ln} and J_{lp} are considered as functions of n_l, p_l, $\operatorname{grad} n_l$, $\operatorname{grad} p_l$ and $\operatorname{grad} V_l$ in the following.

Ad(i) Due to the boundary condition (3.60) for u_5, a suitable space of test functions for the Poisson equation is given by

$$V_5 = \{v \in \overset{n_S}{\underset{l=1}{\mathrm{X}}} H^1(\Omega_l) : \ v_l = 0 \text{ on } \Gamma_{l\mathrm{O}} \cup \Gamma_{l\mathrm{S}}\}.$$

Although we have used the letter V already for the potential, we want to use it also for the space of test functions since it is very common in literature. Due to the context, it should always be clear whether we mean the potential or the space.

Considering (3.63)-(3.66) for the continuity equations, we choose

$$V_6 = V_7 = \{v \in \overset{n_S}{\underset{l=1}{\mathrm{X}}} H^1(\Omega_l) : \ v_l = 0 \text{ on } \Gamma_{l\mathrm{O}}\}.$$

Ad(ii) Multiplying the Poisson equation (3.54) by $v_5 \in V_5$, integrating over Ω, using the integration by parts formula, regarding the boundary conditions

(3.61)-(3.62) and the choice of f, g (see (3.40), (3.43)), we obtain

$$\begin{aligned}\int_{\Omega_l} \varepsilon_l \mathrm{grad}\, u_{5l} \cdot \mathrm{grad}\, v_{5l}\ \mathrm{d}x + \int_{\Gamma_{l\mathrm{MI}}} \alpha_l u_{5l} v_{5l}\ \mathrm{d}\sigma \\ + \int_{\Omega_l} q(u_{6l} - u_{7l}) v_{5l}\ \mathrm{d}x + \int_{\Omega_l} \varepsilon_l \mathrm{grad}\,(f_l \cdot A_S^{\mathrm{T}} u_{1l}) \cdot \mathrm{grad}\, v_{5l}\ \mathrm{d}x \\ = \int_{\Omega_l} q N_l v_{5l}\ \mathrm{d}x - \int_{\Gamma_{l\mathrm{MI}}} \tilde{\beta}_l v_{5l}\ \mathrm{d}\sigma - \int_{\Omega_l} \varepsilon_l \mathrm{grad}\, g_l \cdot \mathrm{grad}\, v_{5l}\ \mathrm{d}x \end{aligned} \tag{3.70}$$

for all $v_5 \in V_5$ and $l = 1, ..., n_S$. Regarding the current density equations (3.57)-(3.58) as well as the boundary conditions (3.64)-(3.66) for the continuity equations, we get analogously, for all $v_6 \in V_6$, $v_7 \in V_7$ and $l = 1, ..., n_S$,

$$\begin{aligned}\int_{\Omega_l} \partial_t u_{6l} v_{6l}\ \mathrm{d}x + \tfrac{1}{q} \int_{\Omega_l} J_{ln} \cdot \mathrm{grad}\, v_{6l}\ \mathrm{d}x + \int_{\Gamma_{l\mathrm{S}}} v_{ln}(u_{6l} - n_{l0}) v_{6l}\ \mathrm{d}\sigma \\ = -\int_{\Omega_l} R_l(u_{6l}, u_{7l}, J_{ln}, J_{lp}) v_{6l}\ \mathrm{d}x - \int_{\Gamma_{l\mathrm{MI}}} R_{l\mathrm{surf}}(u_{6l}, u_{7l}) v_{6l}\ \mathrm{d}\sigma \end{aligned} \tag{3.71}$$

and

$$\begin{aligned}\int_{\Omega_l} \partial_t u_{7l} v_{7l}\ \mathrm{d}x - \tfrac{1}{q} \int_{\Omega_l} J_{lp} \cdot \mathrm{grad}\, v_{7l}\ \mathrm{d}x + \int_{\Gamma_{l\mathrm{S}}} v_{lp}(u_{7l} - p_{l0}) v_{7l}\ \mathrm{d}\sigma \\ = -\int_{\Omega_l} R_l(u_{6l}, u_{7l}, J_{ln}, J_{lp}) v_{7l}\ \mathrm{d}x - \int_{\Gamma_{l\mathrm{MI}}} R_{l\mathrm{surf}}(u_{6l}, u_{7l}) v_{7l}\ \mathrm{d}\sigma \end{aligned} \tag{3.72}$$

with

$$J_{ln} = q(D_{ln} \mathrm{grad}\, u_{6l} - \mu_{ln} u_{6l}\, \mathrm{grad}\,(u_{5l} + f_l \cdot A_S^{\mathrm{T}} u_1 + g_l)), \tag{3.73}$$

$$J_{lp} = q(-D_{lp} \mathrm{grad}\, u_{7l} - \mu_{lp} u_{7l}\, \mathrm{grad}\,(u_{5l} + f_l \cdot A_S^{\mathrm{T}} u_1 + g_l)). \tag{3.74}$$

Ad(iii) The network equations (3.51)-(3.53) yield to

$$\begin{aligned} v_1^{\mathrm{T}} A_C \frac{\mathrm{d} q_C(A_C^{\mathrm{T}} u_1, t)}{\mathrm{d}t} + v_1^{\mathrm{T}} A_R g(A_R^{\mathrm{T}} u_1, t) \\ + v_1^{\mathrm{T}} A_L u_2 + v_1^{\mathrm{T}} A_V u_3 + v_1^{\mathrm{T}} A_S u_4 &= -v_1^{\mathrm{T}} A_I i_s(t), \end{aligned} \tag{3.75}$$

$$v_2^{\mathrm{T}} \frac{\mathrm{d}\phi_L(u_2, t)}{\mathrm{d}t} = v_2^{\mathrm{T}} A_L^{\mathrm{T}} u_1, \tag{3.76}$$

$$v_3^{\mathrm{T}} A_V^{\mathrm{T}} u_1 = v_3^{\mathrm{T}} v_s(t), \tag{3.77}$$

for all $v_1 \in V_1 := \mathbb{R}^{n-1}$, $v_2 \in V_2 := \mathbb{R}^{n_L}$, and $v_3 \in V_3 := \mathbb{R}^{n_V}$. Recall that i_s and v_s are input functions depending on time only. Finally, equation (3.59) implies

$$v_{4l}^{\mathrm{T}} u_{4l} = v_{4l}^{\mathrm{T}} \int_{\Gamma_l} [(J_{ln} + J_{lp}) \cdot \nu \, \chi_{l1} - \varepsilon_l \, \partial_t \operatorname{grad} u_{5l} \cdot \nu \, \chi_{l2}] \, \mathrm{d}\sigma \tag{3.78}$$

for all $v_4 \in V_4 := \overset{n_S}{\underset{l=1}{\mathrm{X}}} \mathbb{R}^{k_l - 1}$ and $l = 1, ..., n_S$.

Ad(iv) We introduce the function space

$$V := V_1 \times V_2 \times V_3 \times V_4 \times V_5 \times V_6 \times V_7.$$

Considering the boundary conditions (3.63) we are looking for solutions that belong to the affine space $u \in V + \tilde{u}$ where

$$\tilde{u} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \tilde{n} \\ \tilde{p} \end{pmatrix}$$

and $\tilde{n}, \tilde{p} \in \overset{n_S}{\underset{l=1}{\mathrm{X}}} H^1(\Omega_l)$ with

$$\tilde{n}_l = \tfrac{1}{2}(\sqrt{N_l^2 + 4n_{li}^2} + N_l), \qquad \tilde{p}_l = \tfrac{1}{2}(\sqrt{N_l^2 + 4n_{li}^2} - N_l)$$

on $\Gamma_{l\mathrm{O}}$ for all $l = 1, ..., n_S$. Here N_l and n_{li} denote the doping concentration and the intrinsic charge density of the semiconductor with the number l. Additionally, we introduce

$$\begin{aligned} Z &:= \mathbb{R}^{n_C} \times \mathbb{R}^{n_L} \times V_4 \times \hat{V}_6 \times \hat{V}_7, \\ H &:= \mathbb{R}^{n_C} \times \mathbb{R}^{n_L} \times V_4 \times \overset{n_S}{\underset{l=1}{\mathrm{X}}} L_2(\Omega_l) \times \overset{n_S}{\underset{l=1}{\mathrm{X}}} L_2(\Omega_l). \end{aligned}$$

Then, Z equipped with the norm

$$\|z\|_Z^2 := \|z_1\|^2 + \|z_2\|^2 + \|z_3\|^2 + \sum_{l=1}^{n_S} \int_{\Omega_l} (z_{4l}^2 + |\operatorname{grad} z_{4l}|^2 + z_{5l}^2 + |\operatorname{grad} z_{5l}|^2) \, \mathrm{d}x$$

forms a real, separable, and reflexive Banach space. Here, $\|\cdot\|$ denotes the Euclidean norm. Furthermore, H equipped with the scalar product

$$(\bar{z}|z)_H := \bar{z}_1^{\mathrm{T}} z_1 + \bar{z}_2^{\mathrm{T}} z_2 + \bar{z}_3^{\mathrm{T}} z_3 + \sum_{l=1}^{n_S} \int_{\Omega_l} (\bar{z}_{4l} z_{4l} + \bar{z}_{5l} z_{5l}) \, \mathrm{d}x \qquad \forall \bar{z}, z \in H$$

forms a real, separable Hilbert space. Additionally, $Z \subseteq H \subseteq Z^*$ represents an evolution triple. Note, that the notation $\bar{z}$ is used for an other element of H only. It does not mean the conjugate complex of z. We are dealing with real spaces here.

Building the sum over the equations (3.70)-(3.78), we arrive at the operator equation

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t) + \mathcal{B}(u(t),t) = 0 \qquad \text{for } t \in [t_0, T] \tag{3.79}$$

with the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$ defined as follows. The operator

$$\mathcal{D}: \ (V + \tilde{u}) \times [t_0, T] \to Z$$

is chosen as

$$\mathcal{D}(u,t) = \begin{pmatrix} q_C(A_C^{\mathrm{T}} u_1, t) \\ \phi_L(u_2, t) \\ \mathfrak{r}_1 u_5 \\ u_6 \\ u_7 \end{pmatrix} \tag{3.80}$$

with

$$\mathfrak{r}_1 u_5 = \begin{pmatrix} \vdots \\ \int_{\Gamma_l} \varepsilon_l \operatorname{grad} u_{5l} \cdot \nu \, \chi_{l2} \, \mathrm{d}\sigma \\ \vdots \end{pmatrix}$$

for all $u_5 \in V_5$. The derivative $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u,t)$ is to be understood as a Z^*-valued distribution satisfying

$$\langle \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t), z\rangle_Z = \frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{D}(u(t),t)|z)_H$$

for all $z \in Z$ and $u(t) \in V + \tilde{u}$. The operator $\mathcal{A}: \ Z^* \to V^*$ is given by

$$\langle \mathcal{A}\bar{z}, v\rangle_V := \langle \bar{z}, \mathcal{A}^* v\rangle_Z, \qquad \mathcal{A}^* v := \begin{pmatrix} A_C^{\mathrm{T}} v_1 \\ v_2 \\ A_S^{\mathrm{T}} v_4 \\ v_6 \\ v_7 \end{pmatrix} \tag{3.81}$$

for all $\bar{z} \in Z^*$ and $v \in V$. Finally, the operator $\mathcal{B}: \ (V + \tilde{u}) \times [t_0, T] \to V^*$

reads as

$$\begin{aligned}
\langle \mathcal{B}(u,t), v\rangle_V \;=\; & v_1^{\mathrm{T}}[A_R g(A_R^{\mathrm{T}}u_1,t) + A_L u_2 + A_V u_3 + A_S u_4 + A_I i_s(t)] \\
& - [v_2^{\mathrm{T}}A_L^{\mathrm{T}} + v_3^{\mathrm{T}}A_V^{\mathrm{T}}]u_1 + v_3^{\mathrm{T}}v_s(t) + v_4^{\mathrm{T}}u_4 - v_4^{\mathrm{T}}\mathfrak{r}_2(J_n + J_p) \\
+ \sum_{l=1}^{n_S} \Bigg[& \int_{\Omega_l} \varepsilon \mathrm{grad}\,(u_{l5} + f_l \cdot A_S^{\mathrm{T}}u_1 + g_l) \cdot \mathrm{grad}\, v_{5l}\ \mathrm{d}x \\
& + \int_{\Omega_l} q(u_{6l} - u_{7l} - N_l)v_{5l}\ \mathrm{d}x + \tfrac{1}{q}\int_{\Omega_l} (J_{ln} \cdot \mathrm{grad}\, v_{6l} - J_{lp} \cdot \mathrm{grad}\, v_{7l})\ \mathrm{d}x \\
& + \int_{\Omega_l} R_l(u_{6l}, u_{7l}, J_{ln}, J_{lp})(v_{6l} + v_{7l})\ \mathrm{d}x \\
& + \int_{\Gamma_{l\mathrm{MI}}} (\alpha_l u_{5l} - \tilde{\beta}_l)v_{5l}\ \mathrm{d}\sigma + \int_{\Gamma_{l\mathrm{MI}}} R_{l\mathrm{surf}}(u_{6l}, u_{7l})(v_{6l} + v_{7l})\ \mathrm{d}\sigma \\
& + \int_{\Gamma_{l\mathrm{S}}} [v_{ln}(u_{6l} - n_{l0})v_{6l} + v_{lp}(u_{7l} - p_{l0})v_{7l}]\ \mathrm{d}\sigma \Bigg]
\end{aligned} \tag{3.82}$$

for all $u \in V + \tilde{u}$ and $v \in V$, where

$$\mathfrak{r}_2(J_n + J_p) = \begin{pmatrix} \vdots \\ \int_{\Gamma_l} (J_{ln} + J_{lp}) \cdot \nu\, \chi_{l1}\ \mathrm{d}\sigma \\ \vdots \end{pmatrix}$$

and J_{ln}, J_{lp} given by (3.73), (3.74).

Remark 3.2 By means of standard arguments, it follows that a sufficiently smooth function u is a solution of (3.79) if and only if

$$u = (e, j_L, j_V, j_S, \tilde{V}, n, p)$$

with J_{ln}, J_{lp} given by (3.73), (3.74) satisfies the coupled system (3.51)-(3.66).

In the next chapter, we deal with general abstract differential algebraic systems. First, we introduce an index concept for such systems. Then, we present network topological criteria for the coupled circuit and device equations using the formulation given in Section 3.5. Finally, we investigate abstract systems as given in Section 3.6. A Galerkin approach providing unique solutions for linear systems with monotone operators is proposed as a first step for a treatment of general ADASs.

Chapter 4

Abstract Differential Algebraic Systems

In order to treat coupled systems of partial differential equations (PDEs) and differential-algebraic equations (DAEs) in a systematic way we shall study abstract differential algebraic systems (ADASs) of the following form

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t)+\mathcal{B}(u(t),t)=0 \qquad \text{for } t\in[t_0,T]. \tag{4.1}$$

This equation is to be understood as an operator equation with operators $\mathcal{A}$, $\mathcal{D}(\cdot,t)$ and $\mathcal{B}(\cdot,t)$ acting in real Hilbert spaces. More precisely, let X, Y, $Z\subseteq\tilde{Z}$ be Hilbert spaces and

$$\mathcal{A}\ :\ \tilde{Z}\to Y,\quad \mathcal{D}(\cdot,t)\ :\ X\to Z,\quad \mathcal{B}(\cdot,t)\ :\ X\to Y. \tag{4.2}$$

In [Mär02a, LMT01], the case $Z=\tilde{Z}$ has been considered. This is enough for the classical formulation of the coupled system (see Section 3.5). However, for the generalized formulation, we need the more general case $Z\subseteq\tilde{Z}$ (see Section 3.6).

Such systems with $\mathcal{A}$ and $\mathcal{D}$ being invertible have already been studied in [GGZ74]. However, the classical formulation as well as the generalized formulation of the coupled system lead to abstract differential algebraic systems with operators $\mathcal{A}$ or $\mathcal{D}(\cdot,t)$ that are not invertible on the whole time interval. Such systems are also called singular (see e.g. [Cam80, Cam82] for the finite dimensional case) or degenerate differential equations (see e.g. [Kur93, FY99]).

In [FY99], systems of the form (4.1) with linear operators have been studied. More precisely, initial value systems of the form

$$\begin{aligned} \mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}u + \mathcal{B}u &= \mathcal{A}g(t) \qquad \text{for } t \in (0,T], && (4.3)\\ \mathcal{D}u(0) &= v_0 \in \operatorname{im}\mathcal{D}, && (4.4) \end{aligned}$$

with $X = Y = Z$, $\mathcal{D} = M$ and $A = \mathcal{I}$ are treated by a semigroup approach. $\mathcal{I}$ denotes the identity operator. The system (4.3)-(4.4) is reduced to multi-valued differential equations of the form

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t} &\in Av + f(t) \qquad \text{for } t \in (0,T], && (4.5)\\ v(0) &= v_0, && (4.6) \end{aligned}$$

where $A := -\mathcal{B}\mathcal{D}^{-1}$. Here, the operator $\mathcal{D}^{-1}$ is defined as (multi-valued) function satisfying

$$\mathcal{D}^{-1}v = \{u \in D_{\mathcal{D}} :\ \mathcal{D}u = v\} \quad \text{for all } v \in \operatorname{im}\mathcal{D}.$$

with $D_{\mathcal{D}}$ being the definition domain of $\mathcal{D}$. The existence and uniqueness of classical solutions of (4.3)-(4.4) satisfying

$$\mathcal{D}u \in C^1([0,T];X) \quad \text{and} \quad \mathcal{B}u \in C([0,T];X)$$

is shown provided that $g \in C^1([0,T];X)$, $\mathcal{B}u(0) \in \operatorname{im}\mathcal{A}$,

$$\operatorname{Re}(-\mathcal{B}u|\mathcal{D}u)_X \le \beta\|\mathcal{D}u\|_X^2, \quad \text{for all } u \in D_{\mathcal{B}} \subseteq D_{\mathcal{D}} \tag{4.7}$$

as well as the operator

$$\lambda_0\mathcal{A}\mathcal{D} + \mathcal{B} :\ D_{\mathcal{B}} \to X \quad \text{is bijective for some } \lambda_0 > \beta. \tag{4.8}$$

A similar result is obtained for systems of the form (4.3)-(4.4) with $\mathcal{A} = D^*$ or $\mathcal{D} = \mathcal{I}$.

This approach via multi-valued differential equations concentrates on the dynamic part of the system. It is limited to systems with special constraints. So, for instance, condition (4.7) implies

$$\operatorname{Re}(\mathcal{B}w, \mathcal{D}v) = 0 \qquad \text{for all } w \in \ker\mathcal{D} \cap D_{\mathcal{B}},\ v \in D_{\mathcal{D}}.$$

Together with condition (4.8), we obtain, for the finite dimensional case, a DAE of index 1 with the constraint

$$(\mathcal{B}Q)^*\mathcal{B}u = (\mathcal{B}Q)^*g(t)$$

where Q denotes a projector onto $\ker \mathcal{D}$. This is obvious since $(\mathcal{B}Q)^*\mathcal{D} = 0$ if we use the Euclidean norm.

Furthermore, the semigroup approach has been extended to linear systems with time dependent operators in [FY99] supposed that $\mathcal{A} = I$ or $\mathcal{D} = I$. There, it is assumed that the operator $\mathcal{D}(t)(\lambda\mathcal{D}(t) + \mathcal{B}(t))^{-1}$ or $(\lambda\mathcal{D}(t) + \mathcal{B}(t))^{-1}\mathcal{D}(t)$, respectively, is bounded in a certain way for λ from a specific region in $\mathbb{C}$. Unfortunately, it is usually quite difficult to verify this condition for coupled systems in practice. Additionally, having DAEs in mind, we think that we should not concentrate on the pencil operators $\lambda\mathcal{D}(t) + \mathcal{B}(t)$ in the non-stationary case. Even in the finite-dimensional case, it is not reasonable to demand non-singularity of $\lambda\mathcal{D}(t) + \mathcal{B}(t)$ for existence and uniqueness of solutions (see e.g. [BCP89, GM86]). Regarding the trivial example

$$\left(\begin{pmatrix} -t & -t^2 \\ 1 & t \end{pmatrix} u\right)' + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u = f(t),$$

we see that $\lambda\mathcal{D}(t) + \mathcal{B}(t)$ is nonsingular for all λ but a solution exists only if $f_1(t) + t\, f_2(t) = 0$. On the other hand, considering the example

$$\left(\begin{pmatrix} t & 0 \\ 1 & 0 \end{pmatrix} u\right)' + \begin{pmatrix} 0 & t \\ 0 & 1 \end{pmatrix} u = \begin{pmatrix} -t \\ 0 \end{pmatrix},$$

the matrix pencil $\lambda\mathcal{D}(t) + \mathcal{B}(t)$ is singular of all λ but there is a unique solution satisfying

$$u_1(t) = -t, \qquad u_2(t) = 1.$$

Hence, investigating time dependent differential algebraic equations, one has turned away from matrix pencils about 15 years ago.

Using an operational method developed in [PG78], more general linear systems of the form (4.3)-(4.4) with either $\mathcal{A} = I$ or $\mathcal{D} = I$ are treated in [FY99]. But it is restricted to systems with constant injective operators $\mathcal{B}$ having a bounded inverse.

Nonlinear abstract systems of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{D}u) + \mathcal{B}u = F(t, \mathcal{K}u), \quad t \in [0, T]$$

where $\mathcal{D}$, $\mathcal{B}$, $\mathcal{K}$ are linear closed operators from a complex Banach space X into a Banach space Y, have been investigated in [FP86, FP89, BF98, FR99]. The theory developed there bases mainly on properties of the operator

$$T := \mathcal{D}(\lambda\mathcal{D} + \mathcal{B})^{-1},$$

λ being a regular point of the operator pencil $\lambda\mathcal{D} + \mathcal{B}$. Most results are presented for problems with the resolvent operator $(T-\xi I)^{-1}$ having a simple pole at $\xi = 0$. In the finite dimensional case, such problems are DAEs of index 1. In [FR99] also problems are investigated where $(T-\xi I)^{-1}$ has a pole of multiple order at $\xi = 0$. Considering again the finite dimensional case, these are problems of higher index. Existence and uniqueness of solutions of such problems are obtained by a study of the transformed problem

$$\frac{\mathrm{d}}{\mathrm{d}t}(Tv) + v = f(t, Nv), \quad t \in [0, T]$$

with $N = \mathcal{K}(\lambda\mathcal{D} + \mathcal{B})^{-1}$, $f(t, w) = e^{-\lambda t}F(t, e^{\lambda t}w)$ and

$$v(t) = e^{-\lambda t}(\lambda\mathcal{D} + \mathcal{B})u(t).$$

Beside certain smoothness conditions and consistent initial conditions, the nonlinear function f has to fulfill a structural condition of the form

$$\pi_k f(t, Nv) = \pi_k f(t, N\sum_{j=k}^{m-1} \Pi_j v), \quad k = m-1, ..., 1$$

for certain projectors Π_j satisfying $\Pi_k v = \pi_k(Pv)\varphi_k$ if m is the order of the pole of $(T-\xi I)^{-1}$ in $\xi = 0$ and $\ker T^m$ is spanned by $\{\varphi_k = T^{m-1-k}\varphi_n\}_{k=0}^{m-1}$. Although the assumptions are shown to be satisfied for a sample circuit with an LI-cutset in [FR99], we don't know network topological conditions (as presented in Section 1.2.1) for general networks that guarantee all assumptions. In particular, the determination of the order of the pole of the operator $(T-\xi I)^{-1}$ in $\xi = 0$ becomes a problem for coupled systems described in Section 3.5 and 3.6.

Nevertheless, the order of the pole plays a significant role for the characterization of the systems. Indeed, in the linear, finite dimensional case, the pole order equals to the index of the DAE.

Since we are interested in network topological conditions characterizing the behavior of solutions of our coupled PDE-DAE system, we will present an index concept for abstract differential algebraic systems in Section 4.1 that is based on the tractability index for DAEs used in Chapter 1. It orients towards the sensitivity of solutions due to perturbations of the right hand side. The main reason for studying this kind of index is our interest in the time domain behavior of the circuit. One can compare it with the time index introduced in [LSEL99] and the modal index introduced in [CM99a] for linear

PDAEs with constant coefficients. For a comparison of the different index concepts in certain case studies, we refer to [LMT01].

In Section 4.2, we will show that the index of coupled systems described in Section 3.5 does not exceed 2. Furthermore, a network topological criterion for index-1 systems will be given.

It remains the question when solutions for general coupled network-device systems described in Section 3.5 exist and are unique. In order to answer this question, techniques of the theory of DAEs for networks and the theory of PDEs for semiconductor devices should be combined. Furthermore, we are also interested in a numerical solution of the coupled systems. Therefore, we prefer approaches for the treatment of semiconductor devices that are based on a Galerkin approach (see e.g. [GG86, Grö87, GG89, Gaj93]). For this, the generalized drift diffusion equations are treated as variational problem.

We aim to combine this approach with projector techniques from DAE theory in order to treat the coupled network-device systems. Therefore, we have to consider the generalized formulation presented in Section 3.6. In Section 4.5 we present a Galerkin method for linear abstract differential algebraic systems. In the further sections, it will be shown that the linear system has a unique solution under certain monotonicity conditions and that the Galerkin method converges to the unique solution of the system. In contrast to Galerkin methods for parabolic differential equations, the choice of the basis of the function space is not arbitrary anymore in order to guarantee convergence. This is not really surprising, since the solution has to satisfy certain constraints as we know from the theory of DAEs.

In Section 4.7 we will see that the assumed monotonicity conditions imply the abstract system to have index 1 and the unique solution to depend continuously on time-dependent perturbations of the right hand side. Furthermore, the solution will be shown to depend continuously on perturbations of the initial data. Additionally, we will see in Section 4.5 that the DAEs resulting from the Galerkin approach have at most index 1, i.e., they suit to the abstract system from the numerical point of view.

4.1 Index concept for ADA Systems

From the finite dimensional case we know that the sensitivity of solutions of DAEs with respect to perturbations depends on its index. Since we are interested in the transient behavior of solutions of the coupled system, we

want to follow the concept in [LMT01] extending the tractability index for DAEs (see Appendix A.3) to abstract differential algebraic systems.

Having the classical formulation of the coupled system (3.67) in mind, we consider systems of the form

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t),t) + \mathcal{B}(u(t),t) = 0 \qquad \text{for } t \in [t_0, T] \tag{4.9}$$

with operators $\mathcal{A}$, $\mathcal{B}(\cdot,t)$ and $\mathcal{D}(\cdot,t)$ acting in Hilbert spaces X, Y and Z as follows:

$$\mathcal{A}:\ Z \to Y, \qquad \mathcal{B}(\cdot,t):\ X \to Y, \qquad \mathcal{D}(\cdot,t):\ X \to Z.$$

We assume the existence of the Fréchet derivatives $\mathcal{B}_0$ and $\mathcal{D}_0$ of the operators $\mathcal{B}(\cdot,t)$ and $\mathcal{D}(\cdot,t)$. More precisely, we assume the existence of linear, continuous operators $\mathcal{B}_0(u,t)$ and $\mathcal{D}_0(u,t)$ satisfying

$$\begin{aligned} \mathcal{B}(u+h,t) - \mathcal{B}(u,t) - \mathcal{B}_0(u,t)h &= o(\|h\|), \quad h \to 0, \\ \mathcal{D}(u+h,t) - \mathcal{D}(u,t) - \mathcal{D}_0(u,t)h &= o(\|h\|), \quad h \to 0, \end{aligned}$$

for all h in some neighborhood of zero in X, all $u \in X$ and all $t \in [t_0, T]$. Furthermore, we assume that $\operatorname{im}\mathcal{D}_0(u,t)$ and $\ker\mathcal{D}_0(u,t)$ do not depend on u and t. Finally, $\mathcal{A}$ and $\mathcal{D}$ are assumed to be well matched in the sense that

$$\ker\mathcal{A} \oplus \operatorname{im}\mathcal{D}_0(u,t) = Z \tag{4.10}$$

forms a topological direct sum for all $u \in X$ and $t \in [t_0, T]$. We will see in the next section that these assumptions are satisfied for our coupled system provided that R_l is sufficiently smooth and the functions μ_{ln}, μ_{lp}, D_{ln}, D_{lp}, $\operatorname{grad} f_l$, $\operatorname{div}(\varepsilon_l \operatorname{grad} f_l)$ and $\operatorname{grad} g_l$ are bounded on Ω_l.

We introduce $\mathcal{G}_0(u,t) := \mathcal{A}\mathcal{D}_0(u,t)$ for all $u \in X$ and $t \in [t_0, T]$. Since we are interested in abstract differential algebraic systems containing equations without time derivatives, we assume that $\operatorname{codim}(\operatorname{cl}(\operatorname{im}(G_0(u,t))) > 0$ for all $u \in X$ and $t \in [t_0, T]$.

Remark 4.1 Since $\mathcal{A}$ and $\mathcal{D}$ are assumed to be well matched, the relations

$$\operatorname{im}\mathcal{G}_0(u,t) = \operatorname{im}\mathcal{A} \quad \text{and} \quad \ker\mathcal{G}_0(u,t) = \ker\mathcal{D}_0(u,t) \tag{4.11}$$

are fulfilled for all $u \in X$ and $t \in [t_0, T]$. Indeed, there is a constant projection operator $\mathcal{R}:\ Z \to Z$ satisfying

$$\operatorname{im}\mathcal{R} = \operatorname{im}\mathcal{D}_0(u,t) \quad \text{and} \quad \ker\mathcal{R} = \ker\mathcal{A}$$

for all $u \in X$ and $t \in [t_0, T]$ because $\ker \mathcal{A}$ and $\operatorname{im} \mathcal{D}_0(u,t)$ form a topological direct sum. Consequently,

$$\operatorname{im} \mathcal{G}_0(u,t) = \operatorname{im} \mathcal{A}\mathcal{R} = \operatorname{im} \mathcal{A}$$

and

$$\ker \mathcal{G}_0(u,t) = \ker \mathcal{R}\mathcal{D}_0(u,t) = \ker \mathcal{D}_0(u,t).$$

Remark 4.2 Considering the formulation of (4.9), the natural solution space is given by

$$C^1_{\mathcal{D}}([t_0,T],X) := \{u \in C([t_0,T],X) : \mathcal{D}(u(\cdot),\cdot) \in C^1([t_0,T],Z)\}.$$

For a linearization

$$\mathcal{A}\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{D}_0(u_*(t),t)u) + \mathcal{B}_0(u_*(t),t)u = q$$

of (4.9) at $u_* \in C^1_{\mathcal{D}}([t_0,T],X)$, the natural solution space reads as

$$C^1_{\mathcal{D}_0}([t_0,T],X) := \{u \in C([t_0,T],X) : \mathcal{D}_0(u_*(\cdot),\cdot)u(\cdot) \in C^1([t_0,T],Z)\}.$$

The next proposition shows that both solution spaces coincide.

Proposition 4.3 *Let $\mathcal{D}_0(u,t)$ depend continuously differentiable on u, t. Additionally, assume the partial derivative $\mathcal{D}_t'(u,t)$ to exist and to be continuous. Furthermore, suppose that the constant space $\ker \mathcal{G}_0(u,t)$ splits X, i.e., there is a linear space $W \subseteq X$ such that*

$$\ker \mathcal{G}_0(u,t) \oplus W = X$$

forms a topological direct sum. Then, $C^1_{\mathcal{D}}([t_0,T],X) = C^1_{\mathcal{D}_0}([t_0,T],X)$.

Proof: Since $\ker \mathcal{G}_0(u,t)$ splits X, we find a projection operator $\mathcal{P}_0 : X \to X$ with

$$\operatorname{im} \mathcal{P}_0 = W \quad \text{and} \quad \ker \mathcal{P}_0 = \ker \mathcal{G}_0(u,t).$$

Using the mean value theorem in Banach spaces, we get

$$\mathcal{D}(u,t) - \mathcal{D}(\mathcal{P}_0 u,t) = \int_0^1 \mathcal{D}_0(su + (1-s)\mathcal{P}_0 u,t)(\mathcal{I} - \mathcal{P}_0)u \,\mathrm{d}s = 0 \quad (4.12)$$

for all $u \in X$ and $t \in [t_0,T]$. Additionally, $\mathcal{D}_0(u,t)$ acts bijectively from W to $\operatorname{im} \mathcal{R}$ for all $u \in X$ and $t \in [t_0,T]$. Applying the implicit function theorem to

$$F : Z \times W \times [t_0,T] \to \operatorname{im} \mathcal{R} \quad \text{with} \quad F(v,w,t) := \mathcal{D}(w,t) - \mathcal{R}v$$

at $v = \mathcal{D}(u(t), t)$ and $w = \mathcal{P}_0 u(t)$ for $u \in C^1_{\mathcal{D}}([t_0, T], X)$, we obtain a continuously differentiable function $g(\cdot, t) : Z \to W$ satisfying $F(v, g(v, t), t) = 0$ for v in a neighborhood of $\mathcal{D}(u(t), t)$. In particular, we have

$$\mathcal{P}_0 u(t) = g(\mathcal{D}(u(t), t), t).$$

Due to the smoothness assumptions, g is also continuously differentiable with respect to t. Consequently,

$$\mathcal{D}(u(\cdot), \cdot) \in C^1([t_0, T], Z) \quad \Leftrightarrow \quad \mathcal{P}_0 u(\cdot) \in C^1([t_0, T], W) \tag{4.13}$$

if we regard (4.12). Analogously to (4.12), we obtain

$$\mathcal{D}_0(u, t) = \mathcal{D}_0(\mathcal{P}_0 u, t)$$

for all $u \in X$ and $t \in [t_0, T]$. This implies $\mathcal{D}_0(u_*(\cdot), \cdot) \in C^1([t_0, T], L(X, Z))$ for $u_* \in C^1_{\mathcal{D}}([t_0, T], X)$. Thus,

$$\mathcal{D}_0(u_*(\cdot), \cdot)u(\cdot) \in C^1([t_0, T], Z) \quad \Leftrightarrow \quad \mathcal{P}_0 u(\cdot) \in C^1([t_0, T], W)$$

since $\ker \mathcal{D}_0(u_*(\cdot), \cdot) = \ker \mathcal{P}_0$ and $\mathcal{D}_0(u_*(\cdot), \cdot)$ acts bijectively from $\operatorname{im} \mathcal{P}_0$ to $\operatorname{im} \mathcal{R}$. Regarding (4.13), the proposition is proven.

□

As the tractability index for finite-dimensional differential algebraic systems, the following index concept for ADASs bases on linearizations.

Definition 4.4
The abstract differential algebraic system (4.9) has *index 1* if there is a projection operator $\mathcal{Q}_0 : X \to X$ onto the constant space $\ker \mathcal{G}_0(u, t)$ such that the operator

$$\mathcal{G}_1(u, t) := \mathcal{G}_0(u, t) + \mathcal{B}_0(u, t)\mathcal{Q}_0$$

is injective and $\operatorname{cl}(\operatorname{im} \mathcal{G}_1(u, t)) = Y$ for all $u \in X$ and $t \in [t_0, T]$.

Remark 4.5 The definition is independent of the choice of the projection operator $\mathcal{Q}_0$. If $\tilde{\mathcal{Q}}_0$ is an other projection operator onto $\ker \mathcal{G}_0$, then

$$\tilde{\mathcal{G}}_1 = \mathcal{G}_0 + \mathcal{B}_0\tilde{\mathcal{Q}}_0 = \mathcal{G}_1(I + \mathcal{Q}_0\tilde{\mathcal{Q}}_0\mathcal{P}_0)$$

holds for $\mathcal{P}_0 := \mathcal{I} - \mathcal{Q}_0$ since

$$\mathcal{Q}_0 = \tilde{\mathcal{Q}}_0\mathcal{Q}_0 \quad \text{and} \quad \tilde{\mathcal{Q}}_0 = \mathcal{Q}_0\tilde{\mathcal{Q}}_0.$$

For brevity, we have omitted the arguments u and t. The operator $I + \mathcal{Q}_0\tilde{\mathcal{Q}}_0\mathcal{P}_0$ is continuous and injective. Its inverse operator is given by $I - \mathcal{Q}_0\tilde{\mathcal{Q}}_0\mathcal{P}_0$ and, thus, also continuous. Consequently,

$$\ker \tilde{\mathcal{G}}_1 = \ker \mathcal{G}_1 \quad \text{and } \operatorname{im} \tilde{\mathcal{G}}_1 = \operatorname{im} \mathcal{G}_1.$$

Remark 4.6 This index definition shall characterize the behavior of abstract differential algebraic systems with respect to perturbations of the right hand side. It should not be confused with the Fredholm index of operators.

Definition 4.7
The abstract differential algebraic system (4.9) has *index 2* if and only if there are projection operators $\mathcal{Q}_0 : X \to X$ onto $\ker G_0(u,t)$ and $\mathcal{Q}_1(u,t) : X \to X$ onto $\ker \mathcal{G}_1(u,t)$ such that $\operatorname{codim}(\operatorname{cl}(\operatorname{im} \mathcal{G}_1(u,t))) > 0$ and the operator

$$\mathcal{G}_2(u,t) := \mathcal{G}_1(u,t) + \mathcal{B}_0(u,t)\mathcal{P}_0\mathcal{Q}_1(u,t)$$

is injective as well as $\operatorname{cl}(\operatorname{im} \mathcal{G}_2(u,t)) = Y$ for all $u \in X$ and $t \in [t_0, T]$.

Remark 4.8 The definitions require implicitly $\ker \mathcal{G}_0$ and $\ker \mathcal{G}_1$ to be closed in X. Due to the choice of the spaces X and Y, this condition is satisfied for the coupled system as abstract DAE in the classical form (see Section 3.5).

4.2 Network Topological Index Criteria

In Chapter 1, we have seen that the network DAEs have an index ≤ 1 if the network neither contains LI-cutsets nor CV-loops with at least one voltage source. The question arises whether one can find such network topological criteria also for the index of the coupled system (3.51)-(3.66) formulated as abstract DAE (3.67) (see Section 3.5).

This question can be answered positively for coupled systems with one-dimensional semiconductor models having two metal-semiconductor contacts. It turns out that LI-cutsets are also relevant for the coupled systems. However CV-loops have to be replaced by CVS-loops, i.e. loops consisting of capacitors, voltage sources and semiconductors.

Lemma 4.9 *The network does not contain LI-cutsets if and only if the matrix*

$$(A_C, A_R, A_V, A_S) \quad \textit{has full row rank.} \tag{4.14}$$

The network does not contain CVS-loops with at least one voltage source or one semiconductor if and only if the matrix

$$(Q_C^T A_V, Q_C^T A_S) \quad \textit{has full column rank} \tag{4.15}$$

for any projector Q_C onto $\ker A_C^T$.

This lemma is a consequence of the Theorems A.1-A.3. We may apply the same arguments as in the proof of Lemma 1.2 and Lemma 1.3.

Recall that the capacitance, resistance and inductance matrices are given by

$$C(v,t) = \frac{\mathrm{d}q_C(v,t)}{\mathrm{d}v}, \quad G(v,t) = \frac{\mathrm{d}g(v,t)}{\mathrm{d}v}, \quad \text{and} \quad L(j,t) = \frac{\mathrm{d}\phi_L(j,t)}{\mathrm{d}j}.$$

Theorem 4.10 *Let $C(v,t)$, $G(v,t)$ and $L(j,t)$ be positive definite for all voltages v, currents j and $t \in [t_0,T]$. Additionally, let the network be consistent, i.e., it contains neither loops of voltage sources only nor cut-sets of current sources only. Furthermore, let all semiconductor models be one-dimensional with two metal-semiconductor contacts. We assume further R_l to be continuously differentiable and the functions μ_{ln}, μ_{lp}, D_{ln}, D_{lp}, $\operatorname{grad} f_l$, $\operatorname{div}(\varepsilon_l \operatorname{grad} f_l)$, $\operatorname{grad} g_l$ to be bounded on Ω_l for all semiconductors ($l = 1,...,n_S$). Finally, $\varepsilon_l > 0$ for all $l = 1,...,n_S$. Then, the ADAS (3.68) has index 1 if and only if the network contains neither LI-cutsets nor CVS-loops with at least one voltage source or one semiconductor.*

Proof: Recall that

$$\mathcal{A} = \begin{pmatrix} A_C & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \end{pmatrix}, \quad \mathcal{D}(u,t) = \begin{pmatrix} q_C(A_C^{\mathrm{T}}u_1,t) \\ \phi_L(u_2,t) \\ -\mathfrak{r}_{1l}\, u_{5l} \\ u_{6l} \\ u_{7l} \end{pmatrix},$$

and

$$\mathcal{B}(u,t) = \begin{pmatrix} A_R g(A_R^{\mathrm{T}}u_1,t) + A_L u_2 + A_V u_3 + A_S u_4 + A_I i_s(t) \\ -A_L^{\mathrm{T}}u_1 \\ A_V^{\mathrm{T}}u_1 - v_s(t) \\ \operatorname{div}(\varepsilon_l \operatorname{grad} \tilde{u}_{5l}) - q(u_{6l} - u_{7l} - N_l) + \operatorname{div}(\varepsilon_l \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}}u_1 + g_l)) \\ -\frac{1}{q}\operatorname{div} u_{8l} + R_l(u_{6l},u_{7l},u_{8l},u_{9l}) \\ \frac{1}{q}\operatorname{div} u_{9l} + R_l(u_{6l},u_{7l},u_{8l},u_{9l}) \\ u_{8l} - q(D_{nl}\operatorname{grad} u_{6l} - \mu_{nl}u_{7l}\operatorname{grad}(u_{5l} + f_l \cdot A_S^{\mathrm{T}}u_1 + g_l)) \\ u_{9l} + q(-D_{pl}\operatorname{grad} u_{7l} - \mu_{pl}u_{7l}\operatorname{grad}(u_{5l} + f_l \cdot A_S^{\mathrm{T}}u_1 + g_l)) \\ u_{4l} - \mathfrak{r}_{2l}(u_{8l} + u_{9l}) \end{pmatrix}$$

where

$$\mathfrak{r}_{1l}v := \int_{\Gamma_l} \varepsilon_l \operatorname{grad} v \cdot \nu\, \chi_{l2}\, \mathrm{d}\sigma, \tag{4.16}$$

$$\mathfrak{r}_{2l}v := \int_{\Gamma_l} v \cdot \nu\, \chi_{l1}\, \mathrm{d}\sigma. \tag{4.17}$$

Let x_{lL} and x_{lR} be the left and the right contact of the semiconductor with number l. Furthermore, we choose, without loss of generality, the right contact as the reference terminal. Then, (4.16), (4.17) read as

$$\mathfrak{r}_{1l}v = -[\varepsilon_l \operatorname{grad} v](x_{lL}), \qquad \mathfrak{r}_{2l}v = -v(x_{lL}).$$

Under the assumptions, the Fréchet derivatives of the operators $\mathcal{D}(\cdot,t)$ and $\mathcal{B}(\cdot,t)$ exist and are given as follows:

$$\mathcal{D}_0(u,t) = \begin{pmatrix} C(A_C^{\mathrm{T}}u_1,t)A_C^{\mathrm{T}}u_1 & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & L(u_2,t)u_2 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & \mathfrak{r}_{1l} & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & I\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,I\,0\,0 \end{pmatrix}$$

and

$$\mathcal{B}_0(u) = \begin{pmatrix} A_R G(A_R^{\mathrm{T}}e,t)A_R^{\mathrm{T}} & A_L & A_V & A_S & 0 \\ -A_L^{\mathrm{T}} & 0 & 0 & 0 & 0 \\ A_V^{\mathrm{T}} & 0 & 0 & 0 & 0 \\ \operatorname{div}(\varepsilon_l \operatorname{grad} f_l \cdot A_S^{\mathrm{T}}) & 0 & 0 & 0 & \operatorname{div}(\varepsilon \operatorname{grad}) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ q\mu_n u_{6l} \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}}) & 0 & 0 & 0 & q\mu_n u_{6l} \operatorname{grad} \\ q\mu_p u_{7l} \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}}) & 0 & 0 & 0 & q\mu_p u_{7l} \operatorname{grad} \\ 0 & 0 & 0 & I & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -q & q & 0 & 0 \\ R'_{ln}(u) & R'_{lp}(u) & R'_{lJ_n}(u) - \frac{1}{q}\operatorname{div} & R'_{lJ_p}(u) \\ R'_{ln}(u) & R'_{lp}(u) & R'_{lJ_n}(u) & R'_{lJ_p}(u) + \frac{1}{q}\operatorname{div} \\ \mathfrak{j}_{ln}(u) & 0 & I & 0 \\ 0 & \mathfrak{j}_{lp}(u) & 0 & I \\ 0 & 0 & -\mathfrak{r}_{2l} & -\mathfrak{r}_{2l} \end{pmatrix}$$

For brevity, we have introduced

$$\begin{aligned} \mathfrak{j}_{ln}(u) &:= q\mu_{ln} \operatorname{grad}(u_{5l} + f_l \cdot A_S^{\mathrm{T}}u_1 + g_l) - qD_{ln} \operatorname{grad}, \\ \mathfrak{j}_{lp}(u) &:= q\mu_{lp} \operatorname{grad}(u_{5l} + f_l \cdot A_S^{\mathrm{T}}u_1 + g_l) + qD_{lp} \operatorname{grad}. \end{aligned}$$

Then, the image space of

$$\mathcal{G}_0(u,t) = \begin{pmatrix} A_C C(A_C^{\mathrm{T}}u_1,t)A_C^{\mathrm{T}} & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & L(u_2,t) & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & I\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,I\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0 & 0\,0 & \mathfrak{r}_{1l} & 0\,0\,0\,0 \end{pmatrix}$$

is given by

$$\operatorname{im} A_C^{\mathrm{T}} \times X_2 \times 0 \times 0 \times X_6 \times X_7 \times 0 \times 0 \times X_4.$$

Obviously, $\operatorname{codim}(\operatorname{cl}(\operatorname{im}\mathcal{G}_0(u,t))) > 0$ for all $u \in X$ and $t \in [t_0, T]$. Furthermore, $\mathcal{G}_0(u)$ has the nontrivial nullspace

$$\ker A_C^{\mathrm{T}} \times 0 \times X_3 \times X_4 \times \mathcal{N}_{\mathfrak{r}_1} \times 0 \times 0 \times X_8 \times X_9$$

where

$$\mathcal{N}_{\mathfrak{r}_1} = \{v \in X_5 : \ \mathfrak{r}_{1l} v_l = 0, \ l = 1, ..., n_S\}.$$

This follows from Lemma 1.4 and the assumption that $C(A_C^{\mathrm{T}} u_1, t)$ and $L(u_2, t)$ are positive definite. We choose

$$\begin{array}{rccc} \mathcal{Q}_{\mathfrak{r}_1} : & X_5 & \to & X_5 \\ & u & \mapsto & v \end{array}$$

with

$$v_l(x) = u_l(x) - h_l(x) \cdot \mathfrak{r}_{1l} u_l \qquad \forall x \in \Omega_l, \qquad l = 1, ..., n_S,$$

as projection operator onto $\mathcal{N}_{\mathfrak{r}_1}$. Here, h_l is chosen as a smooth function satisfying

$$h_l(x_{lL}) = h_l(x_{lR}) = 0 \quad \text{and} \quad \mathfrak{r}_{1l} h_l = 1.$$

Then,

$$\mathcal{Q}_0 = \begin{pmatrix} Q_C & 0\,0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0\,0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0\,I\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0\,0\,I & 0 & 0\,0\,0\,0 \\ 0 & 0\,0\,0 & \mathcal{Q}_{\mathfrak{r}_1} & 0\,0\,0\,0 \\ 0 & 0\,0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0\,0\,0 & 0 & 0\,0\,0\,0 \\ 0 & 0\,0\,0 & 0 & 0\,0\,I\,0 \\ 0 & 0\,0\,0 & 0 & 0\,0\,0\,I \end{pmatrix}$$

is a projection operator onto $\ker \mathcal{G}_0$. The continuity of $\mathcal{Q}_0$ follows from the continuous embedding of $H^1(\Omega_l)$ into $L_2(\partial\Omega_l)$ for all $l = 1, ..., n_S$ and the choice of the space X_5. Furthermore,

$$\mathcal{G}_1(u,t) = \begin{pmatrix} A_C C(A_C^{\mathrm{T}} u_1, t) A_C^{\mathrm{T}} + A_R G(A_R^{\mathrm{T}} u_1, t) A_R^{\mathrm{T}} Q_C & 0 & A_V & A_S \\ -A_L^{\mathrm{T}} Q_C & L(u_2,t) & 0 & 0 \\ A_V^{\mathrm{T}} Q_C & 0 & 0 & 0 \\ \operatorname{div}(\varepsilon_l \operatorname{grad} f_l \cdot A_S^{\mathrm{T}} Q_C) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ q\mu_n u_{6l} \operatorname{grad} f_l \cdot A_S^{\mathrm{T}} Q_C & 0 & 0 & 0 \\ q\mu_p u_{7l} \operatorname{grad} f_l \cdot A_S^{\mathrm{T}} Q_C & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0\,0 & 0 & 0 \\ 0 & 0\,0 & 0 & 0 \\ 0 & 0\,0 & 0 & 0 \\ \operatorname{div}(\varepsilon_l \operatorname{grad} \mathcal{Q}_{\mathfrak{r}_{1l}}) & 0\,0 & 0 & 0 \\ 0 & I\,0 & R'_{lJ_n}(u) - \frac{1}{q}\operatorname{div} & R'_{lJ_p}(u) \\ 0 & 0\,I & R'_{lJ_n}(u) & R'_{lJ_p}(u) + \frac{1}{q}\operatorname{div} \\ q\mu_{ln} u_{6l} \operatorname{grad} \mathcal{Q}_{\mathfrak{r}_{1l}} & 0\,0 & I & 0 \\ q\mu_{lp} u_{7l} \operatorname{grad} \mathcal{Q}_{\mathfrak{r}_{1l}} & 0\,0 & 0 & I \\ \mathfrak{r}_{1l} & 0\,0 & -\mathfrak{r}_{2l} & -\mathfrak{r}_{2l} \end{pmatrix}$$

(i) We show that $\mathcal{G}_1(u,t)$ is not injective if the network contains LI-cutsets or CVS-loops.
If the network contains an LI-cutset, then we find a nontrivial $w \in X_1$ with

$$A_C^{\mathrm{T}} w = 0,\ A_R^{\mathrm{T}} w = 0,\ A_V^{\mathrm{T}} w = 0,\ A_S^{\mathrm{T}} w = 0.$$

It implies $w = Q_C w$ and, consequently,

$$v = (w, L^{-1}(u_2,t) A_L^{\mathrm{T}} w, 0, 0, 0, 0, 0, 0, 0)$$

belongs to the nullspace of $\mathcal{G}_1(u,t)$.
If the network contains a CVS-loop, then we find a nontrivial $(w_1, w_2, w_3) \in \mathbb{R}^{n_C} \times X_3 \times X_4$ with

$$A_C w_1 + A_V w_2 + A_S w_3 = 0.$$

Applying Lemma 1.4, we find a $w_4 \in \operatorname{im} P_C$ such that

$$A_C w_1 = A_C C(A_C^{\mathrm{T}} u_1, t) A_C^{\mathrm{T}} w_4.$$

Choosing $w_5 \in X_5$ with

$$w_5(x) = -h(x) w_3$$

and $h(x) = (h_1(x), ..., h_{n_S}(x))$, the vector function

$$v = (w_4, 0, w_2, w_3, w_5, 0, 0, 0, 0)$$

belongs to the nullspace of $\mathcal{G}_1(u,t)$. The existence of at least one voltage source or one semiconductor in the CVS-loop ensures that the constructed v is nontrivial because at least w_2 or w_3 is nontrivial.

(ii) Now we show that $\mathcal{G}_1(u,t)$ is injective if the network contains neither LI-cutsets nor CVS-loops with at least one voltage source or one semiconductor. We assume that

$$\mathcal{G}_1(u,t) w = 0. \tag{4.18}$$

Integrating the 4th line of (4.18)

$$\operatorname{div}(\varepsilon_l \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}} Q_C w_1)) + \operatorname{div}(\varepsilon_l \operatorname{grad}(\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l})) = 0,$$

we obtain

$$\begin{aligned} &\varepsilon_l \operatorname{grad}(f_l \cdot A_S^{\mathrm{T}} Q_C w_1) + \varepsilon_l \operatorname{grad}(\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l}) \\ &\qquad = \ \mathfrak{r}_{1l}(f_l \cdot A_S^{\mathrm{T}} Q_C w_1) + \mathfrak{r}_{1l}(\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l}) \ = \ 0, \end{aligned}$$

since $\mathfrak{r}_{1l} f_l = 0$ for the chosen function f_l (see (3.41)). Integrating once again, we obtain

$$\begin{aligned} f_l \cdot A_S^{\mathrm{T}} Q_C w_1 + \mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l} \\ = \; f_l(x_{lL}) \cdot A_S^{\mathrm{T}} Q_C w_1 + [\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l}](x_{lL}) \; = \; A_S^{\mathrm{T}} Q_C w_1, \end{aligned}$$

because $f_l(x_{lL}) = 1$ (see (3.40)) and $\operatorname{im} \mathcal{Q}_{\mathfrak{r}_1}$ belongs to X_5. Evaluating this equation at x_{lR}, we arrive at

$$A_S^{\mathrm{T}} Q_C w_1 = 0 \quad \text{and, consequently,} \quad \mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l} = 0. \tag{4.19}$$

From the 3rd line of (4.18), we know that $A_V^{\mathrm{T}} Q_C w_1 = 0$. The 1st line of (4.18) implies

$$Q_C^{\mathrm{T}} A_R G(A_R^{\mathrm{T}} u_1, t) A_R^{\mathrm{T}} Q_C w_1 + Q_C^{\mathrm{T}} A_V w_3 + Q_C^{\mathrm{T}} A_S w_4 = 0.$$

Regarding the two relations for w_1 above and the fact that $G(A_R^{\mathrm{T}} u_1, t)$ is positive definite, we obtain

$$(A_C A_R A_V A_S)^{\mathrm{T}} Q_C w_1 = 0, \quad \text{i.e.} \quad Q_C w_1 = 0, \tag{4.20}$$

since the network does not contain LI-cutsets. Furthermore, using the 2nd and the 5th-8th line of (4.18) as well as the fact that $L(u,t)$ is positive definite, we get

$$w_2 = 0, \quad w_8 = w_9 = 0, \quad w_6 = w_7 = 0. \tag{4.21}$$

The 1st line of (4.18) reads as

$$A_C C(A_C^{\mathrm{T}} u_1, t) A_C^{\mathrm{T}} w_1 + A_V w_3 + A_S w_4 = 0$$

now and implies

$$A_C^{\mathrm{T}} w_1 = 0, \quad w_3 = 0, \quad w_4 = 0,$$

since $C(A_C^{\mathrm{T}} u_1, t)$ is positive definite and the network does not contain CVS-loops with at least one voltage source or one semiconductor. Together with (4.20) and the 9th line of (4.18), this yields to

$$w_1 = 0, \quad \mathfrak{r}_{1l} w_{5l} = 0.$$

Regarding (4.19), we get $w_{5l} = 0$. That means that $G_1(u,t)$ is injective.

(iii) It remains to show that $\overline{\operatorname{im}\mathcal{G}_1(u)} = Y$ if the network contains neither LI-cutsets nor CVS-loops with at least one voltage source or one semiconductor. Since $C^\infty(\Omega_l)$ is dense in $L_2(\Omega_l)$, it is enough to show that

$$\mathbb{R}^{n-1} \times \mathbb{R}^{n_L} \times \mathbb{R}^{n_V} \times (\overset{n_S}{\underset{l=1}{\mathrm{X}}} C^\infty(\Omega_l))^5 \times \mathbb{R}^{n_S} \subseteq \operatorname{im}\mathcal{G}_1(u,t).$$

Let y belong to $\mathbb{R}^{n-1}\times\mathbb{R}^{n_L}\times\mathbb{R}^{n_V}\times(\mathrm{X}_{l=1}^{n_S} C^\infty(\Omega_l))^5\times\mathbb{R}^{n_S}$. With similar arguments like above, one can show that the matrix

$$\begin{pmatrix} A_C C(A_C^{\mathrm{T}}u_1,t)A_C^{\mathrm{T}} + A_R G(A_R^{\mathrm{T}}u_1,t)A_R^{\mathrm{T}}Q_C & A_V & A_S \\ A_V^{\mathrm{T}}Q_C & 0 & 0 \\ A_S^{\mathrm{T}}Q_C & 0 & 0 \end{pmatrix}$$

is nonsingular because the network contains neither LI-cutsets nor CVS-loops with at least one voltage source or one semiconductor. Consequently, we find w_1, w_3 and w_4 such that

$$\begin{aligned} A_C C(A_C^{\mathrm{T}}u_1,t)A_C^{\mathrm{T}}w_1 + A_R G(A_R^{\mathrm{T}}u_1,t)A_R^{\mathrm{T}}Q_C w_1 & \\ + A_V w_3 + A_S w_4 &= y_1, \\ A_V^{\mathrm{T}}Q_C w_1 &= y_3, \\ A_S^{\mathrm{T}}Q_C w_1 &= -\int_{x_{lL}}^{x_{lR}} \tfrac{1}{\varepsilon_l(\tau)} \int_{x_{lL}}^{\tau} y_{4l}(s)\,\mathrm{d}s\,\mathrm{d}\tau. \end{aligned} \tag{4.22}$$

Defining

$$\begin{aligned} w_2 &= L^{-1}(u_2,t)y_2 + L^{-1}(u_2,t)A_L^{\mathrm{T}}Q_C w_1, \\ w_{5l}(x) &= \int_{x_{lL}}^{x} \tfrac{1}{\varepsilon_l(\tau)} \int_{x_{lL}}^{\tau} y_{4l}(s)\,\mathrm{d}s\,\mathrm{d}\tau + (1-f_l(x))A_S^{\mathrm{T}}Q_C w_1 \\ &\quad - h_l(x)(w_{4l} - y_{9l} - \mathfrak{r}_{2l}y_{7l} - \mathfrak{r}_{2l}y_{8l}), \\ w_{8l} &= y_{7l} - \tfrac{q}{\varepsilon_l}\mu_{ln}u_{6l}\int_{x_{lL}}^{x} y_{4l}(s)\,\mathrm{d}s, \\ w_{9l} &= y_{8l} - \tfrac{q}{\varepsilon_l}\mu_{lp}u_{7l}\int_{x_{lL}}^{x} y_{4l}(s)\,\mathrm{d}s, \\ w_{6l} &= y_{5l} - R'_{lJ_n}(u)w_{8l} + \tfrac{1}{q}\operatorname{div} w_{8l} - R'_{lJ_p}(u)w_{9l}, \\ w_{7l} &= y_{6l} - R'_{lJ_n}(u)w_{8l} - R'_{lJ_p}(u)w_{9l} - \tfrac{1}{q}\operatorname{div} w_{9l} \end{aligned}$$

successively, we obtain $w \in X$ and $\mathcal{G}_1(u,t)w = y$. Note that (4.22) ensures $w_{5l}(x_{lR}) = 0$ in order to guarantee $w_5 \in X_5$. □

Theorem 4.11 *Let the assumptions of theorem 4.10 be satisfied. Additionally, we assume that ε_l is constant and greater than zero on Ω_l for all $l = 1, ..., n_S$. Let, furthermore, the network have an LI-cutset or a CVS-loop. Then, the ADAS* (3.68) *has the index 2.*

Remark 4.12 Theorem 4.10 and 4.11 correspond exactly to Theorem 25 and Theorem 26 in [Sch02]. There, a simplified semiconductor model, described by the Poisson equation only, has been studied instead of the complete system of drift-diffusion equations. Furthermore, it has been assumed there that the network has index 1 if all semiconductors are removed. We don't need this assumption here.

PROOF: Regarding the proof of Theorem 4.10, it remains to show that there is a bounded projection operator $\mathcal{Q}_1(u,t)$ onto $\ker \mathcal{G}_1(u,t)$ which provides an injective operator $\mathcal{G}_2(u,t)$ with

$$\mathrm{cl}\,(\mathrm{im}\,\mathcal{G}_2(u,t)) = Y.$$

(i) Existence of $\mathcal{Q}_1$. First, we introduce two new projectors. Let Q_{C-VS} be a projector onto $\ker Q_C^{\mathrm{T}}(A_V, A_S)$ and Q_{CRVS} be a projector onto $\ker\,(A_C, A_R, A_V, A_S)^{\mathrm{T}}$ with $\ker Q_C \subseteq \ker Q_{CRVS}$. Then,

$$\mathcal{Q}_1(u,t) := \begin{pmatrix} Q_{CRVS} & 0 & -H_C^{-1}(A_C^{\mathrm{T}}u_1,t)(A_V,A_S)Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\ L^{-1}(u_2,t)A_L^{\mathrm{T}}Q_{CRVS} & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & (0,-h_l(x))Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\ 0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \end{pmatrix}$$

is a bounded projection operator in X for the nonsingular matrix

$$H_C(A_C^{\mathrm{T}}u_1, t) := A_C C(A_C^{\mathrm{T}}u_1, t)A_C^{\mathrm{T}} + Q_C^{\mathrm{T}}Q_C.$$

Furthermore, a few computations show that $\mathrm{im}\,\mathcal{Q}_1(u,t) \subseteq \ker \mathcal{G}_1(u,t)$. Assume that w belongs to $\ker \mathcal{G}_1(u,t)$. Following step (ii) in the proof of Theorem 4.10, we obtain that (see (4.19) and (4.20))

$$\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l} = 0 \text{ and } (A_C, A_R, A_V, A_S)^{\mathrm{T}} Q_C w_1 = 0, \text{ i.e., } Q_C w_1 = Q_{CRVS} w_1.$$

Considering the 2nd and 5th-8th line of $\mathcal{G}_1(u,t)w = 0$, we obtain

$$w_2 = L_2^{-1}(u_2,t)A_L^{\mathrm{T}}Q_{CRVS}w_1, \quad w_{8l} = w_{9l} = 0, \quad w_{6l} = w_{7l} = 0.$$

The first line of $\mathcal{G}_1(u,t)w = 0$ implies

$$Q_C^{\mathrm{T}}(A_V, A_S)\begin{pmatrix} w_3 \\ w_4 \end{pmatrix} = 0, \text{ i.e., } \begin{pmatrix} w_3 \\ w_4 \end{pmatrix} = Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix}$$

and

$$P_C w_1 = -H_C^{-1}(A_C^{\mathrm{T}} u_1, t)(A_V, A_S)\begin{pmatrix} w_3 \\ w_4 \end{pmatrix}.$$

Finally, by construction of $\mathcal{Q}_{\mathfrak{r}_{1l}}$ and the 9-th line of $\mathcal{G}_1(u,t)w = 0$, we get

$$w_{5l} = \mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l} + h_l(x)\mathfrak{r}_{1l} w_{5l} = h_l(x)\mathfrak{r}_{1l} w_{5l} = -h_l(x) w_4.$$

Summarizing all equations, we obtain $\ker Q_C = \ker Q_{CRVS}$.

(ii) Injectivity of $\mathcal{G}_2$. By definition of $\mathcal{G}_2$, we get

$$\mathcal{G}_2(u,t) = \mathcal{G}_1(u,t) + \mathcal{B}_0(u,t)(I - \mathcal{Q}_0)\mathcal{Q}_1(u,t) =$$

$$\mathcal{G}_1(u,t) + \begin{pmatrix}
A_L L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRVS} & 0 & -A_R G(\cdot) A_R^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & A_L^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & -A_V^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) Q_{C-VS} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & \begin{array}{c}-\mathrm{div}\,(\varepsilon_l \mathrm{grad}\,[f_l \cdot A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) \\ +(0,\ h_l)] Q_{C-VS}\,)\end{array} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & \begin{array}{c}-q\mu_{ln} u_{6l} \mathrm{grad}\,[f_l \cdot A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) \\ +(0,\ h_l)] Q_{C-VS}\end{array} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & \begin{array}{c}-q\mu_{lp} u_{7l} \mathrm{grad}\,[f_l \cdot A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) \\ +(0,\ h_l)] Q_{C-VS}\end{array} & 0\ 0\ 0\ 0\ 0 \\
0 & 0 & (0,\ 0) & 0\ 0\ 0\ 0\ 0
\end{pmatrix}.$$

For brevity, we have used a dot instead of the arguments u and t. We assume that w belongs to $\ker \mathcal{G}_2(u,t)$. Then, the 4-th line of $\mathcal{G}_2(u,t)w = 0$ reads as

$$\begin{aligned}\mathrm{div}\,(\varepsilon_l \mathrm{grad}\,[f_l \cdot A_S^{\mathrm{T}} Q_C w_1 - f_l \cdot A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix} \\ -(0,\ h_l) Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix}]) &= 0.\end{aligned}$$

Integrating this equation once yields to

$$\begin{aligned}\mathrm{grad}\,[f_l \cdot A_S^{\mathrm{T}} Q_C w_1 - f_l \cdot A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V,A_S) Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix} \\ -(0,\ h_l) Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix} + Q_{\mathfrak{r}_{1l}} w_{5l}] &= -(0,\ \tfrac{1}{\varepsilon_l}) Q_{C-VS}\begin{pmatrix} w_3 \\ w_4 \end{pmatrix},\end{aligned}$$

if we regard that $(\operatorname{grad} f_l)(x_{lL}) = 0$ and $(\varepsilon_l \operatorname{grad} h_l)(x_{lL}) = 1$. Integrating once again and evaluating the resulting function at x_{lR}, we obtain

$$\begin{aligned} A_S^{\mathrm{T}} Q_C w_1 - A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) \\ = \quad (0, \ \tfrac{x_{lR}-x_{lL}}{\varepsilon_l}) Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right), \end{aligned}$$

since $f_l(x_{lL}) = 1$, $f_l(x_{lR}) = h_l(x_{lL}) = h_l(x_{lR}) = 0$ and $\mathcal{Q}_{\mathfrak{r}_{1l}} w_{5l} \in H_0^1(\Omega_l)$. Together with the 3-rd line of $\mathcal{G}_2(u,t)w = 0$ this reads as

$$\begin{aligned} (A_V, A_S)^{\mathrm{T}} Q_C w_1 - (A_V, A_S)^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) \\ = \quad \begin{pmatrix} 0 & 0 \\ 0 & \frac{x_{lR}-x_{lL}}{\varepsilon_l} \end{pmatrix} Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right). \end{aligned}$$

Multiplying this equation by Q_{C-VS}^{T}, we arrive at

$$Q_{C-VS}^{\mathrm{T}} \left((A_V, A_S)^{\mathrm{T}} H_C^{-1}(\cdot)(A_V, A_S) + \begin{pmatrix} 0 & 0 \\ 0 & \frac{x_{lR}-x_{lL}}{\varepsilon_l} \end{pmatrix} \right) Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) = 0.$$

Since

$$(A_V, A_S)^{\mathrm{T}} H_C^{-1}(\cdot)(A_V, A_S) \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 0 & \frac{x_{lR}-x_{lL}}{\varepsilon_l} \end{pmatrix}$$

are positive semidefinite, we get

$$Q_{C-VS}^{\mathrm{T}} (A_V, A_S)^{\mathrm{T}} H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) = 0$$

and

$$Q_{C-VS}^{\mathrm{T}} \begin{pmatrix} 0 & 0 \\ 0 & \frac{x_{lR}-x_{lL}}{\varepsilon_l} \end{pmatrix} Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) = 0.$$

We introduce

$$\left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right) := Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right).$$

Thus,

$$A_V v_1 + A_S v_2 = 0 \quad \text{and} \quad v_2 = 0.$$

since $H_C^{-1}(\cdot)$ is positive definite and $\frac{x_{lR}-x_{lL}}{\varepsilon_l}$ is positive. Since the network does not contain loops of voltage sources only, the matrix A_V has full column rank (see Theorem A.2). Hence, we have also $v_1 = 0$, i.e.,

$$Q_{C-VS} \left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) = 0. \tag{4.23}$$

This implies

$$\mathcal{G}_2(u,t)w = \mathcal{G}_1(u,t)w + \begin{pmatrix} A_L L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRVS} w_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Multiplying the first line of this equation system by Q_{CRVS}^{T} yields to

$$Q_{CRVS}^{\mathrm{T}} A_L L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRVS} w_1 = 0, \text{ i.e., } \quad A_L^{\mathrm{T}} Q_{CRVS} w_1 = 0.$$

since $L^{-1}(\cdot)$ is positive definite. Consequently,

$$(A_C, A_R, A_L, A_V)^{\mathrm{T}} Q_{CRVS} w_1 = 0$$

Since the network does not contain cutsets of current sources only, we find always a tree containing capacitive, resistive, inductive and voltage source branches only. Thus, the matrix (A_C, A_R, A_L, A_V) has full row rank (see Theorem A.3) and

$$Q_{CRVS} w_1 = 0. \tag{4.24}$$

Consequently, $\mathcal{G}_1(u,t)w = 0$, i.e., $w = \mathcal{Q}_1(u,t)w$. On the other hand, (4.23) and (4.24) imply $\mathcal{Q}_1(u,t)w = 0$, i.e., $w = 0$.

(iii) Dense solvability of $\mathcal{G}_2$. Since $C^\infty(\Omega_l)$ is dense in $L_2(\Omega_l)$, it is enough to show that

$$\mathbb{R}^{n-1} \times \mathbb{R}^{n_L} \times \mathbb{R}^{n_V} \times (\overset{n_S}{\underset{l=1}{\mathrm{X}}} C^\infty(\Omega_l))^5 \times \mathbb{R}^{n_S} \subseteq \operatorname{im} \mathcal{G}_2(u,t).$$

Let y belong to $\mathbb{R}^{n-1} \times \mathbb{R}^{n_L} \times \mathbb{R}^{n_V} \times (\mathrm{X}_{l=1}^{n_S} C^\infty(\Omega_l))^5 \times \mathbb{R}^{n_S}$. With similar arguments like above, one can show that the matrix

$$\begin{pmatrix} \begin{matrix} A_C C(\cdot) A_C^{\mathrm{T}} + A_R G(\cdot) A_R^{\mathrm{T}} Q_C \\ + A_L L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRVS} \end{matrix} & (A_V, A_S) - A_R G(\cdot) A_R^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \\ (A_V, A_S)^{\mathrm{T}} Q_C & -\left[(A_V, A_S)^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) + \begin{pmatrix} 0 & 0 \\ 0 & \frac{x_{lR} - x_{lL}}{\varepsilon_l} \end{pmatrix} \right] Q_{C-VS} \end{pmatrix}$$

is nonsingular. Consequently, we find w_1, w_3 and w_4 such that

$$\begin{aligned} A_C C(\cdot) A_C^{\mathrm{T}} w_1 + A_R G(\cdot) A_R^{\mathrm{T}} Q_C w_1 & \\ + A_L L^{-1}(\cdot) A_L^{\mathrm{T}} Q_{CRVS} w_1 + A_V w_3 + A_S w_4 & \\ - A_R G(\cdot) A_R^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \begin{pmatrix} w_3 \\ w_4 \end{pmatrix} &= y_1, \\ A_V^{\mathrm{T}} Q_C w_1 - A_V^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) Q_{C-VS} \begin{pmatrix} w_3 \\ w_4 \end{pmatrix} &= y_3 \end{aligned}$$

and

$$\begin{aligned} A_S^{\mathrm{T}} Q_C w_1 - [A_S^{\mathrm{T}} P_C H_C^{-1}(\cdot)(A_V, A_S) + (0, \ \tfrac{x_{lR} - x_{lL}}{\varepsilon_l})] Q_{C-VS} \begin{pmatrix} w_3 \\ w_4 \end{pmatrix} & \\ = \quad - \int_{x_{lL}}^{x_{lR}} \tfrac{1}{\varepsilon_l} \int_{x_{lL}}^{\tau} y_{4l}(s) \, \mathrm{d}s \, \mathrm{d}\tau. & \end{aligned} \tag{4.25}$$

Defining

$$\begin{aligned}
w_2 &= L^{-1}(\cdot)y_2 + L^{-1}(\cdot)A_L^{\mathrm{T}}Q_C w_1 \\
&\quad - L^{-1}(\cdot)A_L^{\mathrm{T}}P_C H_C^{-1}(\cdot)(A_V, A_S)Q_{C-VS}\left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right), \\
w_{8l} &= y_{7l} - \tfrac{q}{\varepsilon_l}\mu_{ln}u_{6l}[\int_{x_{lL}}^{x} y_{4l}(s)\,\mathrm{d}s - (0,I)Q_{C-VS}\left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right)], \\
w_{9l} &= y_{8l} - \tfrac{q}{\varepsilon_l}\mu_{lp}u_{7l}[\int_{x_{lL}}^{x} y_{4l}(s)\,\mathrm{d}s - (0,I)Q_{C-VS}\left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right)], \\
w_{6l} &= y_{5l} - R'_{lJ_n}(\cdot)w_{8l} + \tfrac{1}{q}\mathrm{div}\, w_{8l} - R'_{lJ_p}(\cdot)w_{9l}, \\
w_{7l} &= y_{6l} - R'_{lJ_n}(\cdot)w_{8l} - R'_{lJ_p}(\cdot)w_{9l} - \tfrac{1}{q}\mathrm{div}\, w_{9l}, \\
w_{5l}(x) &= \int_{x_{lL}}^{x} \tfrac{1}{\varepsilon_l}\int_{x_{lL}}^{\tau} y_{4l}(s)\,\mathrm{d}s\,\mathrm{d}\tau + (1-f_l(x))A_S^{\mathrm{T}}Q_C w_1 \\
&\quad - (1-f_l(x))A_S^{\mathrm{T}}P_C H_C^{-1}(\cdot)(A_V, A_S)Q_{C-VS}\left(\begin{smallmatrix} w_3 \\ w_4 \end{smallmatrix}\right) \\
&\quad - h_l(x)(w_{4l} - y_{9l} - \mathfrak{r}_{2l}w_{8l} - \mathfrak{r}_{2l}w_{9l})
\end{aligned}$$

successively, we obtain $w \in X$ and $\mathcal{G}_2(u,t)w = y$. Note that (4.25) ensures $w_{5l}(x_{lR}) = 0$ in order to guarantee $w_5 \in X_5$.

□

Remark 4.13 As in case of DAEs, we do not have a free choice for initial conditions of all components. Theorem 4.10 and 4.11 provide us an information about those components we may choose an initial condition for.

- If the coupled system has index 1, then we may require an initial condition for $\mathcal{P}_0 u(t_0)$ in $\mathrm{im}\,\mathcal{P}_0$, i.e., for $P_C e$ in $\mathrm{im}\,P_C$, j_L, $\mathfrak{r}_{1l}\tilde{V}_l$, n_l and p_l for all $l = 1, ..., n_S$. They correspond to the voltages of all capacitive branches, the currents of all inductive branches, the gradient of the electrostatic potential at the left terminal of each semiconductor as well as the electron and hole density on each semiconductor. Note, that we could require an initial condition for the gradient of the electrostatic potential at the right terminal of a semiconductor, if we would choose the left one as reference terminal.

- In case of an index-2 system, we may require an initial condition for

$\mathcal{P}_0\mathcal{P}_1(u(t_0),t_0)u(t_0)$ in

$$\operatorname{im}\mathcal{P}_0\mathcal{P}_1(u(t_0),t_0) = \operatorname{im}\begin{pmatrix} P_C & 0 & P_C H_C^{-1}(A_C^{\mathrm{T}} e(t_0),t_0)(A_V,A_S)Q_{C-VS} & 0 & 0\,0\,0\,0 \\ -L^{-1}(j_L(t_0),t_0)A_L^{\mathrm{T}}Q_{CRVS} & I & (0,0) & 0 & 0\,0\,0\,0 \\ 0 & 0 & (0,0) & 0 & 0\,0\,0\,0 \\ 0 & 0 & (0,0) & 0 & 0\,0\,0\,0 \\ 0 & 0 & (0,h_l)Q_{C-VS} & h_l\mathfrak{r}_{1l} & 0\,0\,0\,0 \\ 0 & 0 & (0,0) & 0 & I\,0\,0\,0 \\ 0 & 0 & (0,0) & 0 & 0\,I\,0\,0 \\ 0 & 0 & (0,0) & 0 & 0\,0\,0\,0 \\ 0 & 0 & (0,0) & 0 & 0\,0\,0\,0 \end{pmatrix}.$$

Obviously, we have a free choice for the electron and hole density on each semiconductor. Additionally, a few computations show that we have a free choice for all inductors that do not belong to S_I if the set of inductive branches S_I is chosen such that each LI-cutset contains exactly one inductor of S_I. Furthermore, all voltages of capacitive branches that do not belong to $S_{C,S}$ can be chosen freely if the set of capacitive branches $S_{C,S}$ is chosen such that each CVS-loop contains either one capacitor or one semiconductor of $S_{C,S}$. Finally, the gradient of the electrostatic potential at the right terminal of a semiconductor may be chosen freely if the semiconductor does not belong to $S_{C,S}$.

It remains the interesting question, how one can solve the coupled system numerically and how sensitive the solution depends on perturbations. The index characterization above gives a hint for the sensitivity with respect to perturbations of the right hand side. But first of all we need the existence and uniqueness of solutions.

In the next sections, we will show that for linear systems arising from a generalized formulation of coupled systems (cf. Section 3.6). For this, we formulate a Galerkin approach, which requires the solution of DAEs. Then, we will show that the Galerkin method converges, under certain assumptions, to a solution of the abstract differential algebraic system. Finally, we will see that the solution is unique and depends continuously on the data. Naturally, the combination of the Galerkin approach with a numerical method for DAEs (e.g. BDF, cf. Section 1.2.4) provides a numerical method for the solution of abstract differential algebraic systems.

4.3 Linear ADA Systems

The goal is to obtain existence and uniqueness of solutions for abstract differential algebraic equations of the form

$$\mathcal{A}(\mathcal{D}u(t))' + \mathcal{B}(t)u(t) = q(t) \qquad \text{for almost all} \quad t \in (t_0, T) \tag{4.26}$$

with linear operators $\mathcal{A}$, $\mathcal{D}$ and $\mathcal{B}(t)$ acting in special Hilbert spaces which naturally arise by a generalized formulation of coupled systems of differential algebraic equations and partial differential equations of elliptic and parabolic type (cf. Section 3.6). Here, $(\mathcal{D}u)'$ denotes the generalized derivative of $\mathcal{D}u$ in the sense of distributions (see Appendix A.4).

Obviously, if $\mathcal{A}$ and $\mathcal{D}$ represent identity mappings and $\mathcal{B}$ is an elliptic operator, then the system (4.26) represents a system of parabolic differential equations. Usual approaches to obtain existence results in the theory of parabolic differential equations are the theory of semigroups and the Galerkin method (see e.g. [Zei90a]). From the numerical point of view, the Galerkin method is preferable. Firstly, it provides canonically a numerical method for solving the ADA system. Secondly, the theory of semigroups treats the abstract system as an evolution on a manifold. However, this manifold is unknown at the outset and must be calculated when solving the system numerically. But, already in the finite dimensional case, a calculation of the manifold (for a description see e.g. [BM02, RR02]) would be connected with a significant computational expense, in particular for systems of higher index. Furthermore, it would be necessary to investigate the influence of perturbations of the manifold onto solutions, since we can not expect to calculate the manifold exactly.

Although the coupled network-device systems are nonlinear, we are interested in linear systems here. We consider the following new approach as a starting point for a general treatment of such coupled systems. Existence proofs for nonlinear differential equations are often based on the construction of suitable fixed point mappings using the existence of solutions for linear differential equation systems. Therefore, we consider unique solvability statements for linear ADA systems as a substantial basis for a general approach treating abstract differential algebraic systems.

We make the following assumptions.

(A1) The spaces V, Z and H are real Hilbert spaces. $Z \subseteq H \subseteq Z^*$ is an evolution triple (see Appendix A.4).

(A2) The mapping

$$\mathcal{D} : V \to Z$$

is linear, continuous and surjective. The mapping

$$\mathcal{A} : Z^* \to V^*$$

represents the dual mapping of $\mathcal{D}$, that means,

$$\langle \mathcal{A}f, v\rangle_V = \langle f, \mathcal{D}v\rangle_Z \qquad \text{for all} \quad v \in V.$$

The mapping

$$\mathcal{B}(t) : V \to V^*$$

is linear, uniformly bounded and uniformly strongly monotone for all $t \in [t_0,T]$. More precisely, there are constants $c_1, c_2 > 0$ such that

$$\langle \mathcal{B}(t)u, v\rangle_V \le c_1\|u\|\,\|v\|, \quad \langle \mathcal{B}(t)u, u\rangle_V \ge c_2\|u\|^2$$

for all $u, v \in V$ and $t \in [t_0,T]$.

4.4 The solution space $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$

Let $t_0 < T < \infty$. For evolution equations, the natural solution space is given by the Sobolev space $W^1_2(t_0,T;V,H)$ (see Appendix A.4). For linear ADAS of the form (4.26), we have to modify it, since we need the generalized derivative of $(\mathcal{D}u)(t)$ which belongs to Z and not to V. Consequently, we define

$$W^1_{2,\mathcal{D}}(t_0,T;V,Z,H) := \{u \in L_2(t_0,T;V) : \ (\mathcal{D}u)' \in L_2(t_0,T;Z^*)\}$$

where $(\mathcal{D}u)'$ denotes the generalized derivative of $\mathcal{D}u$, that means

$$\int_{t_0}^{T} \varphi'(t)\mathcal{D}u(t)\mathrm{d}t = -\int_{t_0}^{T} \varphi(t)(\mathcal{D}u)'(t)\mathrm{d}t \quad \text{for all } \varphi(t) \in C_0^\infty(t_0,T).$$

Here, $\mathcal{D}u : (t_0,T) \to Z$ is defined by $(\mathcal{D}u)(t) = \mathcal{D}u(t)$ for all $t \in (t_0,T)$. Note that the space $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$ represents also an extension of the natural solution space

$$C^1_D = \{u \in C([t_0,T],\mathbb{R}^n) : \ Du \in C^1([t_0,T],\mathbb{R}^m)\}$$

for differential algebraic equations of the form

$$A[Du(t)]' + Bu(t) = q(t)$$

with matrices A, B and D (see Appendix A.3).

Proposition 4.14 *The space $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$ forms a real Banach space with the norm*

$$\|u\|_{W^1_{2,\mathcal{D}}} := \|u\|_{L_2(t_0,T;V)} + \|(\mathcal{D}u)'\|_{L_2(t_0,T;Z^*)}.$$

PROOF. Obviously, $\|\cdot\|_{W^1_{2,\mathcal{D}}}$ is a norm since $\|\cdot\|_{L_2(t_0,T;V)}$ and $\|\cdot\|_{L_2(t_0,T;Z^*)}$ are norms. It remains to show that $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$ is complete. Let (u_n) be a Cauchy sequence in $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$. Since $L_2(t_0,T;V)$ and $L_2(t_0,T;Z^*)$ are Banach spaces, we find $u \in L_2(t_0,T;V)$ and $v \in L_2(t_0,T;Z^*)$ with

$$u_n \to u \qquad \textit{and} \qquad (\mathcal{D}u_n)' \to v.$$

Since $\mathcal{D}:\ V \to Z$ is continuous, we have

$$\mathcal{D}u_n \to \mathcal{D}u \qquad \text{in} \quad L_2(t_0,T;Z).$$

The continuous embedding $Z \subseteq Z^*$ yields the continuous embedding $L_2(t_0,T;Z) \subseteq L_2(t_0,T;Z^*)$. Consequently,

$$\mathcal{D}u_n \to \mathcal{D}u \qquad \text{in} \quad L_2(t_0,T;Z^*).$$

Since $L_2(t_0,T,Z^*) \subseteq L_1(t_0,T,Z^*)$, we get

$$\mathcal{D}u_n \to \mathcal{D}u \qquad \text{and} \qquad (\mathcal{D}u_n)' \to v \qquad \text{in} \quad L_1(t_0,T;Z^*). \tag{4.27}$$

For $\varphi \in C_0^\infty(t_0,T)$, we have

$$\int_{t_0}^{T} \varphi' \mathcal{D}u_n \,\mathrm{d}t = -\int_{t_0}^{T} \varphi(\mathcal{D}u_n)' \,\mathrm{d}t.$$

(4.27) allows us to apply the limit $n \to \infty$ which yields to

$$\int_{t_0}^{T} \varphi' \mathcal{D}u \,\mathrm{d}t = -\int_{t_0}^{T} \varphi v \,\mathrm{d}t.$$

But this means $v = (\mathcal{D}u)'$ and hence $u \in W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$.

□

Proposition 4.15 *If $u \in W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$, then $\mathcal{D}u$ belongs to the classical Sobolev space*

$$W^1_2(t_0,T;Z,H) = \{v \in L_2(t_0,T;Z):\ v' \in L_2(t_0,T;Z^*)\}.$$

PROOF: Let u belong to $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$. This implies $u \in L_2(t_0,T;V)$. Since $\mathcal{D}: V \to Z$ is continuous, $\mathcal{D}u$ belongs to $L_2(t_0,T;Z)$ and the proposition is proven.

□

The last proposition implies immediately two important properties of the function $\mathcal{D}u$ if $u \in W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$. The first one is a simple conclusion of the continuous embedding

$$W^1_2(t_0,T;Z,H) \subseteq C([t_0,T],H).$$

Corollary 4.16 *If $u \in W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$, then there exists a uniquely determined continuous function $z:[t_0,T] \to H$ which coincides almost everywhere on $[t_0,T]$ with the function $\mathcal{D}u$. Furthermore,*

$$\max_{t_0 \le t \le T} \|z(t)\|_H \le \text{const} \|\mathcal{D}u\|_{W^1_2}.$$

As a consequence of the generalized integration by parts formula, we obtain the next corollary.

Corollary 4.17 *For all u, $v \in W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$ and arbitrary s, t with $t_0 \le s \le t \le T$, the following integration by parts formula holds:*

$$\big(\mathcal{D}u(t)|\mathcal{D}v(t)\big)_H - \big(\mathcal{D}u(s)|\mathcal{D}v(s)\big)_H = \int_s^t \langle(\mathcal{D}u)'(\tau),\mathcal{D}v(\tau)\rangle_Z + \langle(\mathcal{D}v)'(\tau),\mathcal{D}u(\tau)\rangle_Z \, d\tau. \quad (4.28)$$

Here, the values of $\mathcal{D}u$ and $\mathcal{D}v$ are the values of the continuous functions z_u, $z_v:[t_0,T] \to H$ in the sense of Corollary 4.16.

Remark 4.18 All statements and arguments of this section remain true if $\mathcal{D}$ depends on t provided that $\mathcal{D}(\cdot,t): V \to Z$ is uniformly Lipschitz continuous for almost all $t \in (t_0,T)$, i.e.

$$\|\mathcal{D}(u,t) - \mathcal{D}(v,t)\|_Z \le c\|u-v\|_V \quad \forall u,v \in V, \quad \text{for almost all } t \in (t_0,T)$$

with a constant $c > 0$ being independent of t.

4.5 The Galerkin Method for ADA Systems

As explained in the section before, the natural solution space for ADA Systems of the form (4.26) is given by $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$. Therefore, we con-

sider the initial value problem

$$\begin{aligned} \mathcal{A}[\mathcal{D}u(t)]' + \mathcal{B}(t)u(t) &= q(t) \qquad \text{for almost all} \quad t \in (t_0, T), \quad (4.29) \\ \mathcal{D}u(t_0) &= z_0 \in Z \qquad (4.30) \end{aligned}$$

with $u \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$. We will need, additionally, the following assumptions.

(A3) Let $z_0 \in Z$ and $q \in L_2(t_0, T; Z^*)$ be given.

(A4) $\{w_1, w_2, ...\}$ is a basis in V and $\{z_1, z_2, ...\}$ is a basis in Z such that, for all $n \in \mathbb{N}$,

$$\exists m_n \in \mathbb{N}: \quad \{\mathcal{D}w_1, ..., \mathcal{D}w_n\} \subseteq \operatorname{span}\{z_1, ..., z_{m_n}\}.$$

Furthermore, (z_{n0}) is a sequence from Z with

$$z_{n0} \to z_0 \quad \text{in } Z \quad \text{as } n \to \infty,$$

where

$$z_{n0} \in \operatorname{span}\{\mathcal{D}w_1, ..., \mathcal{D}w_n\} \quad \text{for all } n.$$

In order to formulate the Galerkin method, we set

$$u_n(t) = \sum_{i=1}^{n} c_{in}(t) w_i.$$

Additionally, we use the formulation

$$\langle \mathcal{A}[\mathcal{D}u(t)]', v\rangle_V + \langle \mathcal{B}(t)u(t), v\rangle_V = \langle q(t), v\rangle_V \qquad \forall\, v \in V \qquad (4.31)$$

which is equivalent to (4.29) for almost all $t \in (t_0, T)$. Then, we obtain the Galerkin equations if we replace u by u_n and v by w_i:

$$\begin{aligned} \langle \mathcal{A}[\mathcal{D}u_n(t)]', w_i\rangle_V + \langle \mathcal{B}(t)u_n(t), w_i\rangle_V &= \langle q(t), w_i\rangle_V, \qquad (4.32) \\ \mathcal{D}u_n(t_0) &= z_{n0}, \qquad (4.33) \end{aligned}$$

for all $i = 1, ..., n$. By assumption (A2) equation (4.32) reads as

$$\langle [\mathcal{D}u_n(t)]', \mathcal{D}w_i\rangle_Z + \langle \mathcal{B}(t)u_n(t), w_i\rangle_V = \langle q(t), w_i\rangle_V. \qquad (4.34)$$

Regarding the continuous embedding $H \subseteq V^*$, we may also write

$$([\mathcal{D}u_n(t)]' | \mathcal{D}w_i)_H + \langle \mathcal{B}(t)u_n(t), w_i\rangle_V = \langle q(t), w_i\rangle_V.$$

Consequently, the Galerkin equations are given by

$$\left(\sum_{j=1}^{n}[c_{jn}(t)\mathcal{D}w_j]'|\mathcal{D}w_i\right)_H + \sum_{j=1}^{n}\langle\mathcal{B}(t)w_j,w_i\rangle_V c_{jn}(t) = \langle q(t),w_i\rangle_V, \quad (4.35)$$

$$\mathcal{D}u_n(t_0) = z_{n0}, \quad (4.36)$$

for all $i = 1,...,n$. If we take into account assumption (A4), then we find coefficients a_{ik} with $i = 1,...,n$ and $k = 1,...,m_n$ such that

$$\mathcal{D}w_i = \sum_{k=1}^{m_n} a_{ik}z_k \quad \forall\, i = 1,...,n.$$

Note that the coefficients are simply given by $a_{ik} = (\mathcal{D}w_i|z_k)_H$ if the basis $\{z_1, z_2, ...\}$ is an orthonormal basis in Z.

Consequently, equation (4.35) is equivalent to

$$\sum_{k=1}^{m_n}\left(\sum_{j=1}^{n}[c_{jn}(t)\mathcal{D}w_j]'|a_{ik}z_k\right)_H + \sum_{j=1}^{n}\langle\mathcal{B}(t)w_j,w_i\rangle_V c_{jn}(t) = \langle q(t),w_i\rangle_V$$

for all $i = 1,...,n$. This can be rewritten as

$$\sum_{k=1}^{m_n} a_{ik}\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{j=1}^{n}(\mathcal{D}w_j|z_k)_H c_{jn}(t)\right) + \sum_{j=1}^{n}\langle\mathcal{B}(t)w_j,w_i\rangle_V c_{jn}(t) = \langle q(t),w_i\rangle_V.$$

Furthermore, equation (4.36) is equivalent to

$$\sum_{j=1}^{n} c_{jn}(t_0)(\mathcal{D}w_j|z_k)_H = \sum_{j=1}^{n}\alpha_{jn}(\mathcal{D}w_j|z_k)_H \qquad \forall\, k = 1,...,m_n,$$

where $z_{n0} = \sum_{j=1}^{n}\alpha_{jn}\mathcal{D}w_j$. The existence of the coefficients α_{jn} and the second equivalence are ensured by Assumption (A4). Hence, the Galerkin equations represent an initial value differential algebraic equation

$$A(Dc_n(t))' + B(t)c_n(t) = r(t) \quad (4.37)$$

$$Dc_n(t_0) = D\alpha_n \quad (4.38)$$

for the coefficients $c_{nj}(t)$ if we introduce the vector function $c_n(\cdot)$ as

$$c_n(t) = (c_{nj}(t))_{j=1,...,n} \qquad \text{for all} \quad t \in [t_0, T]$$

and the vector $\alpha_n = (\alpha_{jn}(t))_{j=1,\dots,n}$. The matrices A, D and $B(t)$ are given by

$$A = (a_{ik})_{\substack{i=1,\dots,n \\ k=1,\dots,m_n}}, \quad D = (d_{kj})_{\substack{k=1,\dots,m_n \\ j=1,\dots,n}}, \quad B(t) = (b_{ij}(t))_{\substack{i=1,\dots,n \\ j=1,\dots,n}}$$

with

$$d_{kj} = (\mathcal{D}w_j|z_k)_H \qquad \text{and} \qquad b_{ij}(t) = \langle \mathcal{B}(t)w_j, w_i \rangle_V$$

for $i, j = 1, ..., n$ and $k = 1, ..., m_n$. Finally, the vector function for the right hand side reads as

$$r(t) = (r_j(t))_{j=1,\dots,n}$$

with $r_j(t) = \langle q(t), w_j \rangle_V$ for $j = 1, ..., n$.

Proposition 4.19 *The differential algebraic equation* (4.37) *arising from the Galerkin approach for the IVP* (4.29)-(4.30) *has a proper stated leading term, that means*

$$\ker A \oplus \operatorname{im} D = \mathbb{R}^{m_n}.$$

Proof. We shall show first that the intersection of the spaces $\ker A$ and $\operatorname{im} D$ is trivial. We assume $y \in \mathbb{R}^{m_n}$ to belong to this intersection, that means

$$\sum_{k=1}^{m_n} a_{ik} y_k = 0 \quad \text{for all} \quad i = 1, ..., n.$$

Furthermore there exist $p_j \in \mathbb{R}^n$ for $j = 1, ..., n$ such that

$$y_k = \sum_{j=1}^{n} (\mathcal{D}w_j|z_k)_H p_j \quad \text{for all} \quad k = 1, ..., m_n.$$

Multiplying the last equations by a_{ik} and summing over $k = 1, ..., m_n$ we get

$$\sum_{j=1}^{n} (\mathcal{D}w_j| \sum_{k=1}^{m_n} a_{ik} z_k)_H p_j = 0 \quad \text{for all } i = 1, ..., n.$$

This implies

$$(\sum_{j=1}^{n} p_j \mathcal{D}w_j|\mathcal{D}w_i)_H = 0 \quad \text{for all} \quad i = 1, ..., n.$$

because of the definition of the coefficients a_{ik}. Multiplying this by p_i and building the sum over $i = 1, ..., n$ we obtain

$$\| \sum_{j=1}^{n} p_j \mathcal{D}w_j \|_H^2 = 0, \text{ which implies } \sum_{j=1}^{n} p_j \mathcal{D}w_j = 0.$$

Consequently,

$$y_k = \sum_{j=1}^{n} (p_j \mathcal{D} w_j | z_k)_H = 0 \quad \text{for all} \quad k = 1, ..., m_n.$$

It remains to show that the sum of the spaces $\ker A$ and $\operatorname{im} D$ spans the whole space $\mathbb{R}^{m_n}$. For this, it is enough to verify that

$$\dim \ker A \geq m_n - \dim \operatorname{im} D.$$

Let d be the dimension of $\operatorname{im} D$. If $d = m_n$ we are done. If $d < m_n$, then let, without loss of generality, the first d rows of D be linearly independent. Then, we find for all k with $d < k \leq m_n$ real values $\lambda_{k1}, ..., \lambda_{kd}$ such that

$$d_{ki} = \sum_{j=1}^{k} \lambda_{kj} d_{ji} \qquad \forall\, i = 1, ..., m_n.$$

Regarding the definition of d_{ki} we may rewrite this equation as

$$(\mathcal{D} w_i | z_k - \sum_{j=1}^{k} \lambda_{kj} z_j)_H = 0 \qquad \forall\, i = 1, ..., m_n.$$

Using the definition of A we find

$$\sum_{l=1}^{m_n} a_{il} (z_l | z_k - \sum_{j=1}^{k} \lambda_{kj} z_j)_H = 0 \qquad \forall\, i = 1, ..., m_n.$$

This implies $A y_k = 0$ for $y_k = (y_{kl})_{l=1,...,m_n}$ with

$$y_{kl} = (z_l | z_k - \sum_{j=1}^{k} \lambda_{kj} z_j)_Z \quad \forall\, l = 1, ..., m_n.$$

Consequently, $\dim \ker A \geq m_n - d$ holds if $\{y_k;\, k = d+1, ..., m_n\}$ is linearly independent. We assume linear dependency, that means, we find $\mu_{d+1}, ..., \mu_{m_n}$ such that

$$\sum_{k=d+1}^{m_n} \mu_k y_k = 0.$$

This implies

$$\sum_{k=d+1}^{m_n} \mu_k (z_l | z_k - \sum_{j=1}^{k} \lambda_{kj} z_j)_H = 0 \quad \forall\, l = 1, ..., m_n.$$

Defining

$$\xi_{kj} := \begin{cases} -\lambda_{kj} & \text{if } 1 \le j \le d \\ 1 & \text{if } j = k \\ 0 & else \end{cases}$$

for all $k = d+1, ..., m_n$, the last equation reads as

$$(z_l| \sum_{k=d+1}^{m_n} \mu_k \sum_{j=1}^{m_n} \xi_{kj} z_j)_H = 0 \quad \forall\, l = 1, ..., m_n.$$

This yields

$$(\sum_{l=1}^{m_n} \sum_{k=d+1}^{m_n} \mu_k \xi_{kl} z_l| \sum_{j=1}^{m_n} \sum_{k=d+1}^{m_n} \mu_k \xi_{kj} z_j)_H = 0$$

and hence

$$\sum_{j=1}^{m_n} \sum_{k=d+1}^{m_n} \mu_k \xi_{kj} z_j = 0$$

Since $\{z_1, ..., z_{m_n}\}$ is linearly independent, we get

$$\sum_{k=d+1}^{m_n} \mu_k \xi_{kj} = 0 \quad \forall\, j = 1, ..., m_n.$$

Considering the definition of ξ_{kj}, we obtain that $\mu_j = 0$ for $j = d+1, ..., m_n$, that means $\{y_k;\, k = d+1, ..., m_n\}$ is linearly independent.

□

Proposition 4.20 *For the leading term matrix functions of the differential algebraic equation* (4.37) *holds*

(i) $(\operatorname{im} A)^{\perp} = \ker D$,

(ii) AD *is positive semidefinite,*

(iii) $B(t)$ *is positive definite for all* $t \in [t_0, T]$.

Proof. Ad(i) Any vector $y \in \mathbb{R}^n$ belongs to $\ker D$ if and only if

$$\sum_{j=1}^{n} (\mathcal{D} w_j | z_l)_H y_j = 0 \qquad \forall\, l = 1, ..., m_n$$

or, equivalently,

$$\sum_{j=1}^{n}(\sum_{k=1}^{m_n} a_{jk}z_k|z_l)_H y_j = 0 \qquad \forall\, l = 1, ..., m_n.$$

This is equivalent to

$$\sum_{k=1}^{m_n}\sum_{j=1}^{n} y_j a_{jk} z_k = 0,$$

that means, $\sum_{j=1}^{n} y_j a_{jk} = 0$ for all $k = 1, ..., m_n$, since $\{z_k; k = 1, ..., m_n\}$ is linearly independent. But this means nothing else than $y \in (\ker A)^{\perp}$.

Ad(ii) For any $y \in \mathbb{R}^n$ we have

$$\begin{aligned} y^{\mathrm{T}}ADy &= \sum_{k=1}^{m_n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ik}d_{kj}y_i y_j = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathcal{D}w_j|\mathcal{D}w_i)_H y_i y_j \\ &= (\sum_{j=1}^{n} y_j\mathcal{D}w_j|\sum_{i=1}^{n} y_i\mathcal{D}w_i)_H = \|\sum_{j=1}^{n} y_j\mathcal{D}w_j\|_H^2 \geq 0. \end{aligned}$$

Ad(iii) Let $y \neq 0$ be any vector in $\mathbb{R}^n$. Then we get

$$y^{\mathrm{T}}B(t)y = \sum_{i=1}^{n}\sum_{j=1}^{n} y_i\langle\mathcal{B}(t)w_j, w_i\rangle_V y_j = \langle\mathcal{B}(t)(\sum_{j=1}^{n} y_j w_j), \sum_{i=1}^{n} y_i w_i\rangle_V > 0$$

since $\mathcal{B}(t)$ is strongly monotone and $w_1, ..., w_n$ are linearly independent.

□

Proposition 4.21 *The differential algebraic equation* (4.37) *arising from the Galerkin approach for the IVP* (4.29)-(4.30) *has at most index 1.*

Proof. If AD is singular, then let Q be any projector onto $\ker AD$. We shall show that the matrix

$$G_1 := AD + B(t)Q$$

is nonsingular. We assume y to belong to the null space of G_1, i.e.

$$ADy + B(t)Qy = 0 \tag{4.39}$$

Multiplying the equation by $(Qy)^{\mathrm{T}}$ we get

$$(Qy)^{\mathrm{T}}B(t)Qy = 0$$

since $\ker Q^T = \operatorname{im} A$ (see Proposition 4.20). The positive definiteness of B yields $Qy = 0$ and, regarding (4.39), we have $ADy = 0$. But the latter equation means nothing else than $y = Qy$ and, finally, $y = 0$.

□

4.6 Unique Solvability

In order to obtain unique solutions via the Galerkin method we shall need, additionally, the following assumption.

(A5) The null space of $\mathcal{D}$ splits the space V. This implies the existence of a projection operator $\mathcal{Q}: V \to V$ with $\operatorname{im}\mathcal{Q} = \ker\mathcal{D}$. The operator $\mathcal{P}: V \to V$ is the complementary projection operator, that means, $\mathcal{P} = I - \mathcal{Q}$ where I is the identity operator on V. Furthermore, the basis $\{w_1, w_2, ...\}$ is chosen such that

$$w_i \in \operatorname{im}\mathcal{P} \quad \text{for odd } i, \text{ and } w_i \in \operatorname{im}\mathcal{Q} \quad \text{for even } i.$$

This assumption shall guarantee that the dynamic part of the solution will be approximated by linear combinations of the basis functions with odd i. Correspondingly, the non-dynamic part of the solution will be approximated by linear combinations of the basis functions with even i. In applications, it should not be a problem to fulfill assumption (A5).

In the proof of the existence and uniqueness of solutions, we will need some properties of the adjoint operator of the projection operator Q. Therefore, we summarize them in the following lemma.

Lemma 4.22 *Let V and Z be Banach spaces. Furthermore, let $\mathcal{D}$ be a linear, continuous, and surjective operator $\mathcal{D}: V \to Z$.*
If $\mathcal{Q}: V \to V$ is a projection operator onto $\ker\mathcal{D}$, then the adjoint operator $\mathcal{Q}^*: V^* \to V^*$ defined by

$$\langle \mathcal{Q}^*\bar{v}, v\rangle_V = \langle w, \mathcal{Q}v\rangle_V \qquad \forall\ \bar{v} \in V^*,\ v \in V$$

is a projection operator along $\operatorname{im}\mathcal{D}^*$ for the adjoint operator $\mathcal{D}^*: Z^* \to V^*$ defined by

$$\langle \mathcal{D}^*\bar{z}, v\rangle_V = \langle \bar{z}, \mathcal{D}v\rangle_Z \qquad \forall\ \bar{z} \in Z^*,\ v \in V.$$

Proof: (i) $\mathcal{Q}^{*2} = \mathcal{Q}$. For all $\bar{v} \in V^*$ and $v \in V$, we have

$$\langle \mathcal{Q}^{*2}\bar{v}, v\rangle_V = \langle \mathcal{Q}^*\bar{v}, \mathcal{Q}v\rangle_V = \langle \bar{v}, \mathcal{Q}^2v\rangle_V = \langle \bar{v}, \mathcal{Q}v\rangle_V = \langle \mathcal{Q}^*\bar{v}, v\rangle_V.$$

(ii) Continuity. Let $v \neq 0$ belong to V and $\bar{v} \in V^*$. This implies

$$\left|\left\langle \mathcal{Q}^*\bar{v}, \frac{v}{\|v\|}\right\rangle\right| = \left|\left\langle \bar{v}, \frac{\mathcal{Q}v}{\|v\|}\right\rangle\right| = \left|\left\langle \bar{v}, \frac{\mathcal{Q}v}{\|\mathcal{Q}v\|}\right\rangle\right| \frac{\|\mathcal{Q}v\|}{\|v\|} \leq \text{const}\|\bar{v}\|_{V^*}$$

since Q is continuous. But this means $\|\mathcal{Q}^*\bar{v}\|_{V^*} \leq \text{const}\|\bar{v}\|_{V^*}$.

(iii) $\operatorname{im}\mathcal{D}^* \subseteq \ker\mathcal{Q}^*$. For all $\bar{z} \in Z^*$, we get

$$\langle \mathcal{Q}^*(\mathcal{D}^*\bar{z}), v\rangle_V = \langle \mathcal{D}^*\bar{z}, \mathcal{Q}v\rangle_V = \langle \bar{z}, \mathcal{D}(\mathcal{Q}v)\rangle_Z = 0.$$

(iv) $\ker\mathcal{Q}^* \subseteq \operatorname{im}\mathcal{D}^*$. Let $\bar{v} \in \ker\mathcal{Q}^*$, i.e.,

$$0 = \langle \mathcal{Q}^*\bar{v}, v\rangle_V = \langle \bar{v}, \mathcal{Q}v\rangle_V \tag{4.40}$$

for all $v \in V$. Since $\mathcal{D}$ is surjective, we find, for all $z \in Z$ a $v \in V$ such that $z = \mathcal{D}v$. This allows us to define a functional $\bar{z} \in Z^*$ by

$$\langle \bar{z}, z\rangle_Z := \langle \bar{v}, v\rangle_V$$

for any $v \in V$ with $z = \mathcal{D}v$. The functional $\bar{z}$ is well defined since, for any $v_1,\ v_2 \in V$ with

$$\mathcal{D}v_1 = z = \mathcal{D}v_2,$$

it follows $v_1 - v_2 \in \ker\mathcal{D} = \operatorname{im}\mathcal{Q}$ and, consequently,

$$\langle \bar{v}, v_1\rangle_V = \langle \bar{v}, v_2\rangle_V$$

if we regard (4.40). Finally, for all $v \in V$,

$$\langle \mathcal{D}^*\bar{z}, v\rangle_V = \langle \bar{z}, \mathcal{D}v\rangle_Z = \langle \bar{v}, v\rangle_V,$$

which yields to $\bar{v} = \mathcal{D}^*\bar{z} \in \operatorname{im}\mathcal{D}^*$.

□

Note that the surjectivity of $\mathcal{D}$ was needed for the relation $\ker\mathcal{Q}^* \subseteq \operatorname{im}\mathcal{D}^*$ only.

Theorem 4.23 *Let the assumptions (A1)-(A5) be satisfied. Then, the ADA system* (4.29)-(4.30) *has exactly one solution* $u \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$.

The following proof orients towards the existence proof for first-order linear evolution equations presented in [Zei90a]. The main differences are the following.

1. We are looking for solutions $u \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$ instead of $u \in W^1_2(t_0, T; V, H)$.
2. The Galerkin equations represent a differential algebraic equation instead of an explicit ordinary differential equation.

3. Appropriate initial conditions are given only for $\mathcal{D}u(t_0)$ instead of whole $u(t_0)$.

4. Assumption (A5) is needed to ensure the existence of the generalized derivative $(\mathcal{D}u)'$.

PROOF. For brevity we set $W = W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$.

Step 1: Uniqueness. We suppose u_1 and u_2 to be two solutions of the system (4.29)-(4.30). Then, the difference $u = u_1 - u_2$ satisfies the initial value problem

$$\begin{aligned} \mathcal{A}[\mathcal{D}u(t)]' + \mathcal{B}(t)u(t) &= 0 \qquad \text{for almost all} \quad t \in (t_0, T), \\ \mathcal{D}u(t_0) &= 0 \end{aligned}$$

with $u \in W$. This yields

$$\int_{t_0}^{T} \left\langle \mathcal{A}[\mathcal{D}u(t)]', u(t) \right\rangle_V \mathrm{d}t + \int_{t_0}^{T} \left\langle \mathcal{B}(t)u(t), u(t) \right\rangle_V \mathrm{d}t = 0.$$

Regarding Assumption (A1) and (A2), we have

$$\left\langle \mathcal{A}(\mathcal{D}u)'(t), u(t) \right\rangle_V = \left((\mathcal{D}u)'(t) | \mathcal{D}u(t) \right)_H.$$

By the integration by parts formula (4.28) we get

$$\tfrac{1}{2}\|\mathcal{D}u(T)\|_Z^2 - \tfrac{1}{2}\|\mathcal{D}u(t_0)\|_H^2 = -\int_{t_0}^{T} \left\langle \mathcal{B}u(t), u(t) \right\rangle_V \mathrm{d}t.$$

Since $\mathcal{B}(t)$ is uniformly strongly monotone, there is a constant $c > 0$ such that

$$\tfrac{1}{2}\|\mathcal{D}u(T)\|_Z^2 - \tfrac{1}{2}\|\mathcal{D}u(t_0)\|_Z^2 \leq -c \int_{t_0}^{T} \|u(t)\|_V^2 \, \mathrm{d}t.$$

The initial condition $\mathcal{D}u(t_0) = 0$ implies

$$\tfrac{1}{2}\|\mathcal{D}u(T)\|_Z^2 + c \int_{t_0}^{T} \|u(t)\|_V^2 \, \mathrm{d}t \leq 0,$$

and, consequently, $u(t) = 0$ for almost all $t \in (t_0, T)$.

Step 2: Existence proof via the Galerkin method.

(I) Solution of the Galerkin equations. The Galerkin equations (4.37)-(4.38) represent an initial value differential algebraic equation with index 1 (see Proposition 4.21). Since $q \in L^2(t_0, T; V^*)$, the right hand side r of the Galerkin equations belongs to $L^2(t_0, T; \mathbb{R}^n)$. Applying Theorem A.9, the Galerkin equations have a unique solution in

$$L^2_D(t_0, T; \mathbb{R}^n) = \{u \in L^2(t_0, T; \mathbb{R}^n) : \ Du \in C([t_0, T], \mathbb{R}^{m_n})\}.$$

(II) A priori estimates for the Galerkin solution. Multiplying the Galerkin equations (4.35) by $c_{nj}(t)$ and summing over $j = 1, ..., n$, we obtain

$$\big((\mathcal{D}u_n)'(t)|\mathcal{D}u_n(t)\big)_H + \big\langle \mathcal{B}(t)u_n(t), u_n(t)\big\rangle_V = \big\langle q(t), u_n(t)\big\rangle_V$$

Due to the product formula for real valued functions we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathcal{D}u_n(t)|\mathcal{D}u_n(t)\big)_H = 2\,\big((\mathcal{D}u_n)'(t)|\mathcal{D}u_n(t)\big)_H.$$

This implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathcal{D}u_n(t)\|^2_H + 2\,\big\langle \mathcal{B}(t)u_n(t), u_n(t)\big\rangle_V = 2\,\big\langle q(t), u_n(t)\big\rangle_V$$

Integration over t yields

$$\begin{aligned} &\|\mathcal{D}u_n(T)\|^2_H - \|\mathcal{D}u_n(t_0)\|^2_H \\ &+ 2\int_{t_0}^{T} \big\langle \mathcal{B}(t)u_n(t), u_n(t)\big\rangle_V \,\mathrm{d}t \ = \ 2\int_{t_0}^{T} \big\langle q(t), u_n(t)\big\rangle_V \,\mathrm{d}t. \end{aligned}$$

Since $\mathcal{B}(t)$ is strongly monotone with a constant C_0 independent of t, we get

$$\begin{aligned} \|\mathcal{D}u_n(T)\|^2_H + 2C_0\int_{t_0}^{T} \|u_n(t)\|^2_V \,\mathrm{d}t \ \le \\ \|\mathcal{D}u_n(t_0)\|^2_H + 2\int_{t_0}^{T} \big\langle q(t), u_n(t)\big\rangle_V \,\mathrm{d}t. \end{aligned}$$

Using the classical inequality

$$2|xy| \le C_0^{-1}x^2 + C_0y^2,$$

and the assumption that q belongs to $L_2(t_0, T; V^*)$, we find

$$\begin{aligned} \|\mathcal{D}u_n(T)\|^2_H + 2C_0\int_{t_0}^{T} \|u_n(t)\|^2_V \,\mathrm{d}t \le \\ \|\mathcal{D}u_n(t_0)\|^2_H + C_0^{-1}\int_{t_0}^{T} \|q\|^2_{V^*} \,\mathrm{d}t + C_0\int_{t_0}^{T} \|u_n(t)\|^2_V \,\mathrm{d}t. \end{aligned}$$

Consequently, there is a constant C such that

$$\int_{t_0}^{T} \|u_n(t)\|_V^2 \, \mathrm{d}t \le C \left(\|\mathcal{D}u_n(t_0)\|_H^2 + \int_{t_0}^{T} \|q\|_{V^*}^2 \, \mathrm{d}t \right). \tag{4.41}$$

(III) Weak convergence of the Galerkin method in $L_2(t_0, T; V)$. Because of $\mathcal{D}u_n(t_0) = z_{n0} \to z_0$ in Z as $n \to \infty$, the a priori estimate (4.41) yields the boundedness of the sequence (u_n) in the Hilbert space $L_2(t_0, T; V)$. Therefore there is a weakly convergent subsequence $(u_{n'})$ with

$$u_{n'} \rightharpoonup u \quad \text{in } L_2(t_0, T; V) \qquad \text{as} \quad n \to \infty. \tag{4.42}$$

The goal is now to show that u belongs to W and that u is a solution of the original equation (4.29)-(4.30). If this done, then we know because of uniqueness (see Step 1) that all weakly convergent subsequences $(u_{n'})$ have the same limit u and thus

$$u_n \rightharpoonup u \quad \text{in } L_2(t_0, T; V) \qquad \text{as} \quad n \to \infty.$$

(III-1) We shall show the key equation

$$\begin{aligned} -\big(z_0|\mathcal{D}v\big)_H \varphi(t_0) \; &- \int_{t_0}^{T} \big(\mathcal{D}u(t)|\mathcal{D}v\big)_H \varphi'(t) \, \mathrm{d}t \\ &+ \int_{t_0}^{T} \big\langle \mathcal{B}(t)u(t), v \big\rangle_V \varphi(t) \, \mathrm{d}t = \int_{t_0}^{T} \big\langle q(t), v \big\rangle_V \varphi(t) \, \mathrm{d}t \end{aligned} \tag{4.43}$$

for all $v \in V$ and real functions

$$\varphi \in C^1[t_0, T] \qquad \text{with} \quad \varphi(T) = 0. \tag{4.44}$$

Let φ be as in (4.44). We multiply the Galerkin equations (4.35) by φ and use the integration by parts formula (4.28) in order to get

$$\begin{aligned} -\big(z_0|\mathcal{D}w_i\big)_H \varphi(t_0) \; &- \int_{t_0}^{T} \big(\mathcal{D}u_n(t)|\mathcal{D}w_i\big)_H \varphi'(t) \, \mathrm{d}t \\ &+ \int_{t_0}^{T} \big\langle \mathcal{B}(t)u_n(t), w_i \big\rangle_V \varphi(t) \, \mathrm{d}t = \int_{t_0}^{T} \big\langle q(t), w_i \big\rangle_V \varphi(t) \, \mathrm{d}t \end{aligned} \tag{4.45}$$

for all $i = 1, ..., n$. In order to be able to apply the weak limit, we shall show that the integral terms on the left-hand side are linear continuous

functionals on the space $L_2(t_0,T;V)$. Using the Hölder inequality and the continuity of $\mathcal{D}$, we get

$$\left|\int_{t_0}^{T}\left(\mathcal{D}u_n(t)|\mathcal{D}w_i\right)_H\varphi'(t)\,\mathrm{d}t\right| \le \int_{t_0}^{T}\|\mathcal{D}u_n(t)\|_H\|\mathcal{D}w_i\|_H|\varphi'(t)|\,\mathrm{d}t$$
$$\le C_1\|w_i\|_V\left(\int_{t_0}^{T}\|u_n(t)\|_V^2\,\mathrm{d}t\right)^{\frac{1}{2}} = C_1\|w_i\|_V\|u_n\|_{L_2(t_0,T;V)} \quad (4.46)$$

for all $i=1,...,n$. Since $\mathcal{B}(t)$ is bounded with a constant independent of t, we find

$$\left|\int_{t_0}^{T}\left\langle\mathcal{B}(t)u_n(t),w_i\right\rangle_V\varphi(t)\,\mathrm{d}t\right| \le C_2\int_{t_0}^{T}\|u_n(t)\|_V\|w_i\|_V|\varphi(t)|\,\mathrm{d}t$$
$$\le C_3\|w_i\|_V\|u_n\|_{L_2(t_0,T;V)} \quad (4.47)$$

for all $i=1,...,n$. Applying now the weak limit (4.42) to equation (4.45), we obtain

$$-\left(z_0|\mathcal{D}w_i\right)_H\varphi(t_0) - \int_{t_0}^{T}\left(\mathcal{D}u(t)|\mathcal{D}w_i\right)_H\varphi'(t)\,\mathrm{d}t$$
$$+\int_{t_0}^{T}\left\langle\mathcal{B}(t)u(t),w_i\right\rangle_V\varphi(t)\,\mathrm{d}t = \int_{t_0}^{T}\left\langle q(t),w_i\right\rangle_V\varphi(t)\,\mathrm{d}t \quad (4.48)$$

for all $i=1,...,n$. By assumption (A4), there exists a sequence (v_n) with

$$v_n \to v \quad \text{in } V \qquad \text{as} \quad n\to\infty,$$

where each v_n is a finite linear combination of certain basis elements w_i. Regarding the continuity of $\mathcal{D}$, the inequalities (4.46), (4.47), and $q\in L_2(t_0,T;V^*)$, we obtain that equation (4.48) is also satisfied if we replace w_i by v, that means the key equation (4.43) is satisfied.

III-2 Proof that u belongs to $W^1_{2,\mathcal{D}}(t_0,T;V,Z,H)$. Considering Assumption (A5), the Galerkin equations (4.32) with basis elements with even i imply that

$$\left\langle q(t)-\mathcal{B}(t)u_n(t),w_i\right\rangle_V = \left\langle\mathcal{A}[\mathcal{D}u_n(t)]',w_i\right\rangle_V$$
$$= \left\langle[\mathcal{D}u_n(t)]',\mathcal{D}w_i\right\rangle_Z = 0 \quad (4.49)$$

for all $n=1,2,...$ and almost all $t\in(t_0,T)$. Recall that $w_i\in\operatorname{im}\mathcal{Q}=\ker\mathcal{D}$ for even i. Due to (4.47), we may apply again the weak limit, that means

$$\int_{t_0}^{T}\langle q(t),w_i\rangle_V\varphi(t)\,\mathrm{d}t - \int_{t_0}^{T}\langle\mathcal{B}(t)u(t),w_i\rangle_V\varphi(t)\,\mathrm{d}t = 0$$

for all $\varphi \in C_0^\infty(t_0, T)$. Applying the variational lemma, we get

$$\langle q(t), w_i \rangle_V - \langle \mathcal{B}(t)u(t), w_i \rangle_V = 0$$

for almost all $t \in (t_0, T)$. For any $v \in \operatorname{im} \mathcal{Q}$, we find a sequence (v_n) with

$$v_n \to v \qquad \text{in} \quad V \qquad \text{as} \quad n \to \infty$$

where each v_n is a linear combination of the basis elements w_i with even i. Consequently,

$$\langle q(t), v \rangle_V - \langle \mathcal{B}(t)u(t), v \rangle_V = 0 \tag{4.50}$$

for all $v \in \operatorname{im} Q$. This allows us to define a functional $\bar{z}(t) \in Z^*$ such that

$$\langle \bar{z}(t), z \rangle_Z = \langle q(t) - \mathcal{B}(t)u(t), v \rangle_V \tag{4.51}$$

for almost all $t \in (t_0, T)$ and for all $z \in Z$ with $z = \mathcal{D}v$ for some $v \in V$. Since $\mathcal{D}$ is surjective, the functional is defined for all $z \in Z$. Furthermore, $\bar{z}(t)$ is well defined since

$$\langle \bar{z}(t), \mathcal{D}v_1 \rangle_Z = \langle \bar{z}(t), \mathcal{D}v_2 \rangle_Z$$

for any v_1, $v_2 \in V$ with $\mathcal{D}v_1 = Dv_2$. This is a conclusion from the fact that $v_1 - v_2$ belongs to $\operatorname{im} \mathcal{Q}$ and (4.50). We shall show that $\bar{z}$ belongs to $L_2(t_0, T; Z^*)$. We have

$$\begin{aligned} \|\bar{z}(t)\|_{Z^*} = \sup_{\|z\|_Z \le 1} \left| \langle \bar{z}(t), z \rangle_Z \right| &= \sup_{\|\mathcal{D}v\|_Z \le 1,\ v \in \ker \mathcal{Q}} \left| \langle q(t) - \mathcal{B}(t)u(t), v \rangle_V \right| \\ &\le \sup_{\|\mathcal{D}v\|_Z \le 1,\ v \in \ker \mathcal{Q}} \|q(t) - \mathcal{B}(t)u(t)\|_{V^*} \|v\|_V. \end{aligned}$$

Note that $\mathcal{D}$ is bijective from $\ker \mathcal{Q}$ to Z. Using the open mapping theorem and regarding that D is also linear and continuous, we find a constant $C \ge 0$ such that

$$\|v\|_V \le C \|\mathcal{D}v\|_Z \qquad \text{for all} \quad v \in \ker \mathcal{Q}.$$

This implies

$$\|\bar{z}(t)\|_{Z^*} \le C \|q(t) - \mathcal{B}(t)u(t)\|_{V^*} \tag{4.52}$$

for almost all $t \in (t_0, T)$. Since $q \in L_2(t_0, T; V^*)$, $u \in L_2(t_0, T; V)$, and $B(t)$ is uniformly bounded, we obtain $\bar{z} \in L_2(t_0, T; Z^*)$. Using the key equation (4.43) and (4.51), we arrive at

$$-\int_{t_0}^{T} \left(\mathcal{D}u(t) | \mathcal{D}v \right)_H \varphi'(t) = \int_{t_0}^{T} \left\langle \bar{z}(t), \mathcal{D}v \right\rangle_V \varphi(t) \, \mathrm{d}t$$

for all $v \in V$ and $\varphi \in C_0^\infty(t_0, T)$. Since $\mathcal{D}$ is surjective, we have

$$-\int_{t_0}^{T} \langle \mathcal{D}u(t), z\rangle_Z \varphi'(t) = \int_{t_0}^{T} \langle \bar{z}(t), z\rangle_Z \varphi(t)\ \mathrm{d}t$$

for all $z \in Z$. This is equivalent to

$$\left\langle -\int_{t_0}^{T} \mathcal{D}u(t)\varphi'(t) - \int_{t_0}^{T} \bar{z}(t)\varphi(t)\ \mathrm{d}t, z \right\rangle_Z = 0 \qquad \forall z \in Z.$$

since $\varphi'\mathcal{D}u$ and φz belong to $L_2(t_0, T; Z^*)$ for all $\varphi \in C_0^\infty(t_0, T)$. But this means that

$$-\int_{t_0}^{T} \mathcal{D}u(t)\varphi'(t) = \int_{t_0}^{T} \bar{z}(t)\varphi(t)\ \mathrm{d}t,$$

and, finally, $\mathcal{D}u$ has the generalized derivative $\bar{z} \in L_2(t_0, T; Z^*)$. Hence, u belongs to $W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$.

III-3 Proof that u fulfills (4.29). Since $\bar{z} = (Du)'$, we have, by (4.51),

$$\langle (Du)'(t), \mathcal{D}v\rangle_Z = \langle q(t) - \mathcal{B}(t)u(t), v\rangle_V \tag{4.53}$$

for all $v \in V$ and almost all $t \in (t_0, T)$. By Assumption (A2), we get

$$\langle \mathcal{A}(Du)'(t), v\rangle_V = \langle q(t) - \mathcal{B}(t)u(t), v\rangle_V.$$

But this means that (4.29) is satisfied.

III-4 Proof that u fulfills (4.30). Since $u \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$, we can apply the integration by parts formula (4.28). This yields

$$(\mathcal{D}u(T), \varphi(T)\mathcal{D}v)_H - (\mathcal{D}u(t_0), \varphi(t_0)\mathcal{D}v)_H = \int_{t_0}^{T} \langle (\mathcal{D}u)'(t), \varphi(t)\mathcal{D}v\rangle_Z + (\mathcal{D}u(t)|\varphi'(t)\mathcal{D}v)_H\ \mathrm{d}t$$

for all $\varphi \in C^1[t_0, T]$ and $v \in V$. In particular, if $\varphi(t_0) = 1$ and $\varphi(T) = 0$, then equation (4.43) along with (4.53) yields

$$(\mathcal{D}u(t_0) - z_0|\mathcal{D}v)_H = 0 \qquad \text{for all} \quad v \in V.$$

Since $\mathcal{D}$ is surjective and Z is dense in H, we get $\mathcal{D}u(t_0) = z_0$.

□

4.7 Continuous Dependence on the Data

Theorem 4.24 *Let the assumptions (A1)-(A5) be satisfied. Then, the ADA system* (4.29)-(4.30) *has at most index 1, that means, the map* $\mathcal{G}_1(t) : V \to V^*$ *defined as*

$$\langle \mathcal{G}_1(t)u, v\rangle_V = \langle \mathcal{A}(\mathcal{D}u), v\rangle_V + \langle \mathcal{B}(t)\mathcal{Q}u, v\rangle_V \qquad \text{for all} \quad v \in V$$

is injective and densely solvable for all $u \in V$*. The system has index 0 if and only if* $\mathcal{D}$ *is injective.*

Remark 4.25 Here, $\mathcal{D}u$ is considered as the unique element of Z^* satisfying

$$\langle \mathcal{D}u, z\rangle_Z = (\mathcal{D}u|z)_H \qquad \forall z \in Z.$$

Proof. *Step 1.* Injectivity of $\mathcal{G}_1(t)$. Let u belong to the null space of $\mathcal{G}_1(t)$. This implies

$$\langle \mathcal{A}(\mathcal{D}u), v\rangle_V + \langle \mathcal{B}(t)\mathcal{Q}u, v\rangle_V = 0 \qquad \text{for all} \quad v \in V.$$

Due to Assumption (A2) we have

$$\langle \mathcal{D}u, \mathcal{D}v\rangle_Z + \langle \mathcal{B}(t)\mathcal{Q}u, v\rangle_V = 0 \qquad \text{for all} \quad v \in V. \tag{4.54}$$

In particular, for $v = \mathcal{Q}u$, we get

$$\langle \mathcal{B}(t)\mathcal{Q}u, \mathcal{Q}u\rangle_V = 0.$$

Since $\mathcal{B}(t)$ is strictly monotone, it implies $\mathcal{Q}u = 0$. This yields $\mathcal{D}u = 0$ if we use $v := u$ in (4.54). Consequently, $u = \mathcal{Q}u = 0$.

Step 2. $\operatorname{im}\mathcal{G}_1(t) = V^*$. Since $\mathcal{G}_1(t)$ is linear, it is sufficient to show that $\operatorname{im}\mathcal{G}_1(t)$ is closed and

$$(\operatorname{im}\mathcal{G}_1(t))^\perp = \{v \in V : \ \langle \mathcal{G}_1(t)u, v\rangle_V = 0 \quad \forall u \in V\} = \{0\}.$$

(I) $\operatorname{im}\mathcal{G}_1(t)$ is closed. Let (v_n) be a Cauchy sequence in $\operatorname{im}\mathcal{G}_1(t) \subseteq V^*$. Since $\mathcal{A}$, $\mathcal{D}$, $\mathcal{B}(t)$, and $\mathcal{Q}$ are linear and continuous, the map $\mathcal{G}_1(t)$ is also linear and continuous. From step 1 we know that $\mathcal{G}_1(t)$ is injective. Hence, $\mathcal{G}_1(t)$ is bijective from $V \to \operatorname{im}\mathcal{G}_1(t)$ and the inverse $\mathcal{G}_1^{-1}(t) : \operatorname{im}\mathcal{G}_1(t) \to V$ is continuous. This implies that $(\mathcal{G}_1^{-1}(t)v_n)$ is a Cauchy sequence in the Banach space V. Thus, $(\mathcal{G}_1^{-1}(t)v_n)$ converges in V. We denote its limit by u. Using the continuity of $\mathcal{G}_1(t)$, we get

$$v_n \to \mathcal{G}_1(t)u \in \operatorname{im}\mathcal{G}_1(t),$$

that means, $\operatorname{im}\mathcal{G}_1(t)$ is closed.

(II) $(\operatorname{im}\mathcal{G}_1(t))^\perp = \{0\}$. Let v belong to $(\operatorname{im}\mathcal{G}_1(t))^\perp$. This yields

$$\langle \mathcal{A}(\mathcal{D}u) + \mathcal{B}(t)\mathcal{Q}u, v\rangle_V = 0 \qquad \forall u \in V.$$

It implies

$$\langle \mathcal{A}(\mathcal{D}u), v\rangle_V = 0 \qquad \forall u \in \operatorname{im}\mathcal{P}.$$

Since $\mathcal{D}$ is surjective, we get

$$\langle \mathcal{A}z, v\rangle_V = 0 \qquad \forall z \in Z.$$

Due to Assumption (A1) and (A2), we have

$$(z|\mathcal{D}v)_H = 0 \qquad \forall z \in Z.$$

It implies $\mathcal{D}v = 0$ if we choose $z = \mathcal{D}v$. Consequently, $v \in \operatorname{im}\mathcal{Q}$ and

$$0 = \langle \mathcal{B}(t)\mathcal{Q}u, v\rangle_V = \langle \mathcal{B}(t)\mathcal{Q}u, \mathcal{Q}v\rangle_V \qquad \forall u \in V.$$

Choosing $u = v$, we arrive at

$$0 = \langle \mathcal{B}(t)\mathcal{Q}v, \mathcal{Q}v\rangle_V \geq C\|\mathcal{Q}v\|_V^2,$$

since $\mathcal{B}(t)$ is strongly monotone. Finally, we have

$$v = \mathcal{Q}v = 0.$$

Step 3. If $\mathcal{D}$ is injective, then $\mathcal{Q}$ is the zero mapping and $\mathcal{G} = \mathcal{A}\mathcal{D}$ is bijective due to step 1 and step 2. Conversely, injectivity of $\mathcal{A}\mathcal{D}$ implies injectivity of $\mathcal{D}$. Consequently, (4.29)-(4.30) has index 0 if and only if $\mathcal{D}$ is injective.

□

From the theory of DAEs we know that index-1 systems have solutions which depend continuously on the data. The following theorem shows that this is also the case for the solution of the index-1 ADA system (4.29)-(4.30).

Theorem 4.26 *Let the assumptions (A1)-(A5) be satisfied. Furthermore, let $u \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$ be the unique solution of the ADA system* (4.29)-(4.30)*. Then, the map*

$$(z_0, q) \mapsto u$$

is linear and continuous from $Z \times L_2(t_0, T; V^*)$ to $W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$, i.e., there is a constant $C > 0$ such that

$$\|u\|_{W^1_{2,\mathcal{D}}} \leq C(\|z_0\|_H + \|q\|_{L_2(t_0,T;V^*)}),$$

for all $z_0 \in Z$ and $b \in L_2(t_0, T; V^*)$.

PROOF. In the proof of Theorem 4.23 we have seen that

$$u_n \rightharpoonup u \qquad \text{in} \quad L_2(t_0, T; V) \qquad \text{as} \quad n \to \infty.$$

From the Banach-Steinhaus theorem it follows that

$$\|u\|_{L_2(t_0,T;V)} \leq \varliminf_{n\to\infty} \|u_n\|_{L_2(t_0,T;V)}.$$

Using the a priori estimate (4.41), the continuity of $\mathcal{D}$ and Assumption (A4), we find a constant $C_1 \geq 0$ such that

$$\|u\|_{L_2(t_0,T;V)} \leq C_1 \left(\|z_0\|_H + \|q\|_{L_2(t_0,T;V_*)} \right) \tag{4.55}$$

Using inequality (4.52), we obtain

$$\|(\mathcal{D}u)'\|_{L_2(t_0,T;Z^*)} \leq C_2 \left(\|u\|_{L_2(t_0,T;V)} + \|q\|_{L_2(t_0,T;V^*)} \right)$$

if we regard that $\bar{z} = (\mathcal{D}u)'$ in the proof of theorem 4.23. Together with (4.55), it implies the assertion with the constant $C = C_1C_2 + C_1 + C_2 > 0$. □

4.8 Strong Convergence of the Galerkin Method

Theorem 4.27 *Let the assumptions (A1)-(A5) be satisfied. Then, for all $n = 1, 2, ...$, the Galerkin equations* (4.37)-(4.38) *have exactly one solution*

$$u_n \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H).$$

The sequence (u_n) converges as $n \to \infty$ to the solution u of (4.29)-(4.30) *in the following sense:*

$$u_n \to u \quad \text{in } L_2(t_0, T; V) \qquad \text{and} \qquad \max_{t_0\leq t\leq T} \|\mathcal{D}u_n(t) - \mathcal{D}u(t)\|_H \to 0.$$

PROOF. In the proof of Theorem 4.23, the Galerkin equations (4.37)-(4.38) are shown to have a unique solution $u_n \in W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$. Furthermore, it was shown that the unique solution u of (4.29)-(4.30) belongs to $W^1_{2,\mathcal{D}}(t_0, T; V, Z, H)$.

Step 1. We shall show that $\max_{t_0\leq t\leq T} \|\mathcal{D}u_n(t) - \mathcal{D}u(t)\|_H \to 0$. Since

$$W^1_2(t_0, T; V, H) \subseteq C([t_0, T], H),$$

we get $\mathcal{D}u \in C([t_0,T],H)$ and $\mathcal{D}u_n \in C([t_0,T],H)$. We introduce

$$H_n = \operatorname{span}\{\mathcal{D}w_1,...,\mathcal{D}w_n\} \subseteq Z \subseteq H.$$

Due to Assumption (A5),

$$H_n = \operatorname{span}\{\mathcal{D}w_i,\ 1 \le i \le n,\ i \text{ odd}\}.$$

Since the set of all polynomials $p:\ [t_0,T] \to Z$ is dense in $W_2^1(t_0,T;Z,H)$ we find, for $u \in W_{2,\mathcal{D}}^1(t_0,T;V,Z,H)$ and each $\varepsilon > 0$, a polynomial

$$p(t) = \sum_i t^i a_i$$

with coefficients $a_i \in Z$ and

$$\begin{aligned}\|\mathcal{D}u - p\|_{W_2^1(t_0,T;Z,H)} &= \Big(\int_{t_0}^T \|\mathcal{D}u(t) - p(t)\|_Z^2\ \mathrm{d}t\Big)^{\frac{1}{2}} \\ &\quad + \Big(\int_{t_0}^T \|(\mathcal{D}u)'(t) - p'(t)\|_{Z^*}^2\ \mathrm{d}t\Big)^{\frac{1}{2}} \le \varepsilon.\end{aligned}$$

Due to the fact that $\underset{n}{\cup} H_n$ is dense in Z, we obtain that the set of all polynomials with coefficients in $\underset{n}{\cup} H_n$ is dense in $W_2^1(t_0,T;Z,H)$. Thus there exists a sequence (p_n) of polynomials

$$p_n:\ [t_0,T] \to H_n$$

with

$$p_n \to \mathcal{D}u \qquad \text{in} \quad W_2^1(t_0,T;Z,H) \qquad \text{as} \quad n \to \infty.$$

The continuity of the embedding $W_2^1(t_0,T;Z,H) \subseteq C([t_0,T]),H)$ implies

$$\max_{t_0 \le t \le T} \|\mathcal{D}u(t) - p_n(t)\|_H \le \text{const}\, \|\mathcal{D}u - p_n\|_{W_2^1(t_0,T;Z,H)} \to 0. \tag{4.56}$$

Since

$$\max_{t_0 \le t \le T} \|\mathcal{D}u_n(t) - \mathcal{D}u(t)\|_H \le \max_{t_0 \le t \le T} \|\mathcal{D}u_n(t) - p_n(t)\|_H + \max_{t_0 \le t \le T} \|\mathcal{D}u(t) - p_n(t)\|_H$$

it is enough to show that

$$\max_{t_0 \le t \le T} \|\mathcal{D}u_n(t) - p_n(t)\|_H \to 0 \qquad \text{as} \quad n \to \infty. \tag{4.57}$$

Regarding $\mathcal{D}u_n(t_0) = z_{n0} \to z_0$ in Z as $n \to \infty$, the inequality (4.56) yields

$$\begin{aligned} \|\mathcal{D}u_n(t_0) - p_n(t_0)\|_H &\leq \|\mathcal{D}u_n(t_0) - \mathcal{D}u(t_0)\|_H + \|\mathcal{D}u(t_0) - p_n(t_0)\|_H \\ &= \|z_{n0} - z_0\|_H + \|\mathcal{D}u(t_0) - p_n(t_0)\|_H \\ &\to 0 \qquad \text{as} \quad n \to \infty. \end{aligned} \tag{4.58}$$

Here we have used the continuity of the embedding $Z \subseteq H$. Since $p_n(t)$ belongs to H_n for all $t \in [t_0, T]$, we find coefficients $\beta_{in}(t)$ such that

$$p_n(t) = \sum_{i=1}^{n} \beta_{in}(t) \mathcal{D}w_i.$$

Introducing

$$v_n(t) = \sum_{i=1}^{n} \beta_{in}(t) w_i.$$

we have $p_n = \mathcal{D}v_n$. From the Galerkin equations (4.34) we get

$$\langle (\mathcal{D}u_n)', \mathcal{D}u_n - p_n \rangle_Z = \langle q - \mathcal{B}(t)u_n, u_n - v_n \rangle_V.$$

In the proof of Theorem 4.23 we have seen that the generalized derivative $(\mathcal{D}u)'$ satisfies (4.51) with $\bar{z} = (\mathcal{D}u)'$. This implies

$$\langle (\mathcal{D}u_n)', \mathcal{D}u_n - p_n \rangle_Z = \langle (\mathcal{D}u)', \mathcal{D}u_n - p_n \rangle_Z + \langle \mathcal{B}(t)u - \mathcal{B}(t)u_n, u_n - v_n \rangle_V.$$

Since $\mathcal{B}(t)$ is strongly monotone, we have

$$\langle \mathcal{B}(t)u - \mathcal{B}(t)u_n, u - u_n \rangle_V \geq 0$$

and, consequently,

$$\langle (\mathcal{D}u_n)', \mathcal{D}u_n - p_n \rangle_Z \leq \langle (\mathcal{D}u)', \mathcal{D}u_n - p_n \rangle_Z + \langle \mathcal{B}(t)u - \mathcal{B}(t)u_n, u - v_n \rangle_V. \tag{4.59}$$

The equation (4.49) yields

$$\langle q - \mathcal{B}(t)u_n, v \rangle_V = 0 \qquad \forall\ v \in \operatorname{im} \mathcal{Q}.$$

Together with (4.50) it implies

$$\langle \mathcal{B}(t)u - \mathcal{B}(t)u_n, v \rangle_V = 0 \qquad \forall\ v \in \operatorname{im} \mathcal{Q}.$$

Using this equation in (4.59), we get

$$\langle (\mathcal{D}u_n)', \mathcal{D}u_n - p_n \rangle_Z \leq \langle (\mathcal{D}u)', \mathcal{D}u_n - p_n \rangle_Z + \langle \mathcal{B}(t)u - \mathcal{B}(t)u_n, \mathcal{P}(u - v_n) \rangle_V. \tag{4.60}$$

If we apply the integration by parts formula (4.28), we obtain

$$\begin{aligned}
&\tfrac{1}{2}\|\mathcal{D}u_n(t) - p_n(t)\|_H^2 - \tfrac{1}{2}\|\mathcal{D}u_n(t_0) - p_n(t_0)\|_H^2 \\
&\quad = \int_{t_0}^t \langle (\mathcal{D}u_n)'(s) - p_n'(s), \mathcal{D}u_n(s) - p_n(s)\rangle_Z \,\mathrm{d}s \\
&\quad \le \int_{t_0}^t \langle (\mathcal{D}u)'(s) - p_n'(s), \mathcal{D}u_n(s) - p_n(s)\rangle_Z \,\mathrm{d}s \\
&\qquad + \int_{t_0}^t \langle \mathcal{B}(s)u(s) - \mathcal{B}(s)u_n(s), \mathcal{P}(u(s) - v_n(s))\rangle_V \,\mathrm{d}s \\
&\quad \le \|(\mathcal{D}u)' - p_n'\|_{L_2(t_0,T;Z^*)}\|\mathcal{D}u_n - p_n\|_{L_2(t_0,T;Z)} \\
&\qquad + \|\mathcal{B}u - \mathcal{B}u_n\|_{L_2(t_0,T;V^*)}\|\mathcal{P}(u - v_n)\|_{L_2(t_0,T;V)} \qquad (4.61)
\end{aligned}$$

Since $\mathcal{D}|_{\operatorname{im} P} : \operatorname{im}\mathcal{P} \to Z$ is linear, bijective and continuous, its inverse $\mathcal{D}|_{\operatorname{im} P}^{-1} : Z \to \operatorname{im}\mathcal{P}$ is continuous. Hence, we find a constant $C > 0$ such that

$$\|\mathcal{P}(u - v_n)\|_{L_2(t_0,T;V)} \le C\,\|\mathcal{D}u - p_n\|_{L_2(t_0,T;Z)}. \qquad (4.62)$$

The a priori estimate (4.41) shows that $(\mathcal{D}u_n)$ is bounded in $L_2(t_0,T;Z)$. From Theorem 4.23 we know that $u \in L_2(t_0,T;V)$. Since $\mathcal{D}$ is continuous, we have $\mathcal{D}u \in L_2(t_0,T;Z)$. Therefore, (p_n) converges in $L_2(t_0,T;Z)$. In particular, (p_n) is bounded in $L_2(t_0,T;Z)$. The uniform continuity of $\mathcal{B}(t)$ implies $\mathcal{B}u \in L_2(t_0,T;V^*)$ and the boundedness of $(\mathcal{B}u_n)$ in $L_2(t_0,T;V^*)$. Thus, we obtain from (4.61) and (4.62) that

$$\begin{aligned}
&\tfrac{1}{2}\|\mathcal{D}u_n(t) - p_n(t)\|_H^2 - \tfrac{1}{2}\|\mathcal{D}u_n(t_0) - p_n(t_0)\|_H^2 \\
&\quad \le C_1\|\mathcal{D}u - p_n\|_{W_2^1(t_0,T;Z,V)} \;\to\; 0 \qquad \text{as} \quad n \to \infty.
\end{aligned}$$

Regarding (4.58), the estimation (4.57) is proven.

Step 2. Strong convergence of (u_n) in $L_2(t_0,T;V)$. From Theorem 4.23 we know that

$$u_n \rightharpoonup u \qquad \text{in} \quad L_2(t_0,T;V) \qquad \text{as} \quad n \to \infty.$$

Since $\mathcal{B}$ is linear and uniformly continuous, we get

$$\mathcal{B}u_n \rightharpoonup \mathcal{B}u \qquad \text{in} \quad L_2(t_0,T;V^*) \qquad \text{as} \quad n \to \infty$$

and, hence,

$$\int_{t_0}^T \langle \mathcal{B}u_n, u\rangle_V \to \int_{t_0}^T \langle \mathcal{B}u, u\rangle_V \qquad \text{as} \quad n \to \infty. \qquad (4.63)$$

Furthermore, the assumption $q \in L_2(t_0, T, V^*)$ yields

$$\int_{t_0}^{T} \langle q, u_n \rangle_V \to \int_{t_0}^{T} \langle q, u \rangle_V \qquad \text{as} \quad n \to \infty. \tag{4.64}$$

From the integration by parts formula (4.28) , we have

$$\tfrac{1}{2}\|(\mathcal{D}u(T) - \mathcal{D}u_n(T)\|_H^2 - \tfrac{1}{2}\|(\mathcal{D}u(t_0) - \mathcal{D}u_n(t_0)\|_H^2 = \int_{t_0}^{T} \langle (\mathcal{D}u)'(t) - (\mathcal{D}u_n)'(t), \mathcal{D}u(t) - \mathcal{D}u_n(t) \rangle_Z \, \mathrm{d}t$$

and

$$\big(\mathcal{D}u_n(T)|\mathcal{D}u(T)\big)_H - \big(\mathcal{D}u_n(t_0)|\mathcal{D}u(t_0)\big)_H = \int_{t_0}^{T} \langle (\mathcal{D}u_n)'(t), \mathcal{D}u(t) \rangle_Z + \langle (\mathcal{D}u)'(t), \mathcal{D}u_n(t) \rangle_Z \, \mathrm{d}t.$$

Using again (4.51) for the generalized derivative $(\mathcal{D}u)' = \bar{z}$, we obtain

$$\tfrac{1}{2}\|(\mathcal{D}u(T) - \mathcal{D}u_n(T)\|_H^2 - \tfrac{1}{2}\|(\mathcal{D}u(t_0) - \mathcal{D}u_n(t_0)\|_H^2 = \int_{t_0}^{T} \langle q(t) - \mathcal{B}(t)u(t), u(t) - u_n(t) \rangle_V - \langle (\mathcal{D}u_n)'(t), \mathcal{D}u(t) - \mathcal{D}u_n(t) \rangle_Z \, \mathrm{d}t \tag{4.65}$$

as well as

$$\big(\mathcal{D}u_n(T)|\mathcal{D}u(T)\big)_H - \big(\mathcal{D}u_n(t_0)|\mathcal{D}u(t_0)\big)_H = \int_{t_0}^{T} \langle (\mathcal{D}u_n)'(t), \mathcal{D}u(t) \rangle_Z + \langle (q(t) - \mathcal{B}(t)u(t), u_n(t) \rangle_V \, \mathrm{d}t. \tag{4.66}$$

From the Galerkin equations (4.34) we get

$$\langle (\mathcal{D}u_n)'(t), \mathcal{D}u_n(t) \rangle_Z + \langle \mathcal{B}(t)u_n(t), u_n(t) \rangle_V = \langle q(t), u_n(t) \rangle_V. \tag{4.67}$$

The strong monotonicity of $\mathcal{B}(t)$ implies

$$C \, \|u - u_n\|^2_{L_2(t_0,T;V)} \leq \int_{t_0}^{T} \langle \mathcal{B}(t)u(t) - \mathcal{B}(t)u_n(t), u(t) - u_n(t) \rangle_V \mathrm{d}t,$$

where C is a positive constant. Applying (4.65), we obtain

$$C \, \|u - u_n\|^2_{L_2(t_0,T;V)} \leq \int_{t_0}^{T} \langle q(t) - \mathcal{B}(t)u_n(t), u(t) - u_n(t) \rangle_V \mathrm{d}t - \int_{t_0}^{T} \langle (\mathcal{D}u_n)'(t), \mathcal{D}u(t) - \mathcal{D}u_n(t) \rangle_Z \mathrm{d}t + \tfrac{1}{2}\|\mathcal{D}u(t_0) - \mathcal{D}u_n(t_0)\|_H^2.$$

Regarding (4.67), it yields

$$C\,\|u-u_n\|^2_{L_2(t_0,T;V)} \le \int_{t_0}^{T} \langle q(t)-\mathcal{B}(t)u_n(t), u(t)-u_n(t)\rangle_V \mathrm{d}t$$
$$-\int_{t_0}^{T} \langle (\mathcal{D}u_n)'(t), \mathcal{D}u(t)\rangle_Z \mathrm{d}t + \tfrac{1}{2}\|\mathcal{D}u(t_0)-\mathcal{D}u_n(t_0)\|^2_H.$$

Using (4.66), we get

$$\begin{aligned} C\,\|u-u_n\|^2_{L_2(t_0,T;V)} \;&\le \int_{t_0}^{T} \langle q(t)-\mathcal{B}(t)u_n(t), u(t)\rangle_V \mathrm{d}t \\ &+ \int_{t_0}^{T} \langle q(t)-\mathcal{B}(t)u(t), u_n(t)\rangle_V \mathrm{d}t - (\mathcal{D}u_n(T), \mathcal{D}u(T))_H \\ &+ (\mathcal{D}u_n(t_0), \mathcal{D}u(t_0))_H + \tfrac{1}{2}\|\mathcal{D}u(t_0)-\mathcal{D}u_n(t_0)\|^2_H. \end{aligned} \tag{4.68}$$

If we apply (4.63) and (4.64), we see that the right hand side of inequality (4.68) converges to

$$2\int_{t_0}^{T} \langle q(t)-\mathcal{B}(t)u(t), u(t)\rangle_V \mathrm{d}t - (\mathcal{D}u(T), \mathcal{D}u(T))_H + (\mathcal{D}u(t_0), \mathcal{D}u(t_0))_H. \tag{4.69}$$

as $n \to \infty$. Note that we have already proven in step 1 that

$$\|\mathcal{D}u(t_0)-\mathcal{D}u_n(t_0)\|_H \to 0 \qquad \text{as} \quad n \to \infty.$$

Applying once more the integration by parts formula (4.28) and regarding (4.51) for the generalized derivative $(\mathcal{D}u)' = \bar{z}$, we get

$$\begin{aligned} &(\mathcal{D}u(T), \mathcal{D}u(T))_H - (\mathcal{D}u(t_0), \mathcal{D}u(t_0))_H \;= \\ &\quad 2\int_{t_0}^{T} \langle (\mathcal{D}u)'(t), \mathcal{D}u(t)\rangle_Z \mathrm{d}t \;=\; 2\int_{t_0}^{T} \langle q(t)-\mathcal{B}(t)u(t), u(t)\rangle_V \mathrm{d}t. \end{aligned} \tag{4.70}$$

Summarizing (4.68)-(4.70), we obtain

$$\|u-u_n\|^2_{L_2(t_0,T;V)} \to 0 \qquad \text{as} \quad n \to \infty$$

which implies the assertion.

□

Summary and perspectives

The further miniaturization of optic and electronic components demands a refined network analysis describing certain semiconductor elements by contributed models. Using the instationary drift-diffusion model, the device equations represent a system of elliptic and parabolic differential equations. The network is described by a differential-algebraic system. Both systems are mutually coupled via boundary conditions and integral relations.

The coupled system can be analyzed as an abstract differential algebraic system in infinite-dimensional Hilbert spaces. For the one-dimensional case (with respect to space), network topological criteria for the index of the coupled system are described. Furthermore it is shown that the index does not exceed 2. This corresponds exactly to the results for networks with compact models instead of distributed models. It is still an open question whether the results remain true for a higher-dimensional case.

Finally, a Galerkin approach for handling linear ADASs with monotone operators is proposed. It is shown to provide solutions that converge to the unique solution of the abstract differential-algebraic system. Furthermore, the solution is proven to depend continuously on the data. The most interesting point of the Galerkin approach for ADASs is the choice of basis functions. Choosing the basis functions properly (in certain subspaces), the Galerkin solution is enforced to satisfy the constraints of the system.

It remains to answer the problem whether the nonlinear coupled system has a unique solution and how to solve it numerically. We consider the presented Galerkin approach for linear ADASs as a starting point for an exciting research about the numerical treatment of linear and nonlinear abstract differential-algebraic equations in general. Thus, we expect interesting results not only for coupled circuit and device systems but also for coupled systems of DAEs and PDEs in other application fields.

Appendix A

Basics

In this chapter, we collect basic definitions and results from different mathematical fields used in the Chapters before. It is devoted to readers who are not familiar with one or another field.

A.1 Graph Theory

We start with basic definitions.

1. Graph. A *graph* is a set of branches. If all branches are equipped with an orientation, then the graph is called a *directed* graph. The ends of the branches are called nodes.

2. Path. A set of branches $\{b_1, b_2, ..., b_n\}$ of a graph G is called a *path* between two nodes j and k, if the branches have the following properties.

   (i) Successive branches b_i and b_{i+1} have always one common node.

   (ii) Each node belongs to maximal two branches of the set.

   (iii) The nodes j and k belong to exactly one (usually not the same) branch of the set.

   EXAMPLE. In Figure A.1, the set $\{d, h, i, f, b\}$ is a path. The sets $\{e, f, j, g\}$ and $\{e, f, g, h, i, c\}$ do not represent paths.

3. Connected graph. A graph is called a *connected graph*, if there exists at least one path between any two nodes of the graph.

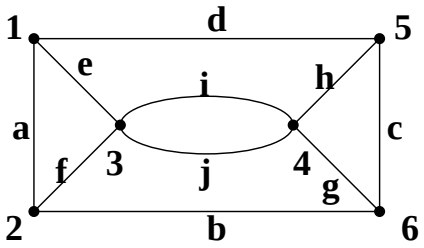

Figure A.1: Example circuit with 6 nodes and 10 branches

4. Loop. A subgraph G_l of a connected graph G is called a *loop*, if G_l is connected and each node of G_l connects exactly two branches of G_l.

   EXAMPLE. In Figure A.1, the set $\{a, b, c, d\}$ is a loop. However, the set $\{a, e, f, g, j\}$ does not represent a loop.

5. Tree. A subgraph G_t of a connected graph G is called a *tree*, if

   (i) G_t is connected,

   (ii) G_t contains all nodes of G,

   (iii) G_t has no loops.

   EXAMPLE. In Figure A.1, the set $\{a, e, d, g, i\}$ is a tree. The sets $\{a, e, h, c\}$ and $\{a, b, c, d, g, e\}$ are not trees.

   Note that, for each connected graph, one can construct a tree (see e.g. [CDK87]). Furthermore, each tree of a connected graph consists of exactly $n - 1$ branches for n being the number of nodes.

6. Cutset. A set of branches G_c of a connected graph is called a *cutset* if

   (i) removing all branches of G_c from G leads to an unconnected graph,

   (ii) adding, afterwards, any branch of G_c leads again to a connected graph.

   EXAMPLE. In Figure A.1, the set $\{a, e, d\}$ is a cutset. The sets $\{e, h, g, f, b\}$ and $\{c, d\}$ are not cutsets.

7. Incidence matrix. Let a directed graph G with n nodes and b branches be given. The *incidence matrix* $A_a \in \mathbb{R}^{n \times b}$ is defined as $A_a = (a_{ij})$ with

$$a_{ij} = \begin{cases} 1 & \text{if the branch } j \text{ leaves the node } i, \\ -1 & \text{if the branch } j \text{ enters the node } i, \\ 0 & \text{else.} \end{cases}$$

The rows of the incidence matrix A_a of a connected graph are linear dependent. More precisely, the sum of all rows of this matrix equals zero. This is caused by the fact that each column contains exactly one 1 and one -1. All other entries of A_a are zero. This becomes clear if one regards that each column corresponds to exactly one branch, and each branch leaves and enters exactly one node, respectively. Consequently, one row of the incidence matrix is needless for the description of the network.

If one erases one row of A_a, then we obtain the so called reduced incidence matrix A. For electrical networks, one usually neglects the row corresponding to the mass node.

In literature, the reduced incidence matrix A is often called only incidence matrix A.

Theorem A.1 *The (reduced) incidence matrix A of a connected graph with n nodes has $n-1$ linear independent rows and, hence, full row rank.*

Proof. Let, without loss of generality, the mass node be the node with number n. Furthermore, let a_k denote the k-th row of the (reduced) incidence matrix A for $k = 1, ..., n-1$. We assume that the rows a_k are linear dependent. The, we find coefficients λ_1, λ_2, ..., λ_{n-1} such that

$$\sum_{k=1}^{n-1} \lambda_k a_k = 0,$$

We numerate the nodes such that $\lambda_k \neq 0$ for all $1 \leq k \leq n_0$ and $\lambda_k = 0$ for all $n_0 + 1 \leq k \leq n-1$. Because of the linear dependency, we have $n_0 \geq 1$. This yields to

$$\sum_{k=1}^{n_0} \lambda_k a_k = 0 \quad \text{und} \quad A = \begin{pmatrix} A_u \\ \hline A_l \end{pmatrix} = \left(\begin{array}{c} a_1 \\ \dots \\ a_{n_0} \\ \hline a_{n_0+1} \\ \dots \\ a_{n-1} \end{array}\right).$$

A_u has at least one column with not only zero entries. Otherwise the nodes $\{1, ..., n_0\}$ and the nodes $\{n_0 + 1, ..., n\}$ would not be connected. Now we permute the columns of A_u such that the first b_0 ($b_0 \geq 1$) columns contain at least one 1 or one -1 and all other columns have zero entries only. This

implies

$$A_u = \begin{pmatrix} A_{ul} & | & A_{ur} \end{pmatrix} = \begin{pmatrix} \underbrace{*}_{b_0} & | & \underbrace{0}_{b-b_0} \end{pmatrix},$$

where b is the number of branches of the network and the number of columns of A, respectively. Since each column of A_{ul} contains maximal one $+1$ and maximal one -1, and

$$\sum_{k=1}^{k_0} \lambda_k a_{kj} = 0$$

holds for all $j = 1, ..., b$, each column of A_{ul} contains exactly one $+1$ and exactly one -1. This yields to

$$A = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \begin{matrix} n_0 \\ n-1-n_0 \end{matrix},$$

i.e., the nodes $\{1, ..., n_0\}$ and $\{n_0 + 1, ..., n\}$ are not connected. This contradicts the assumption of a connected graph.

□

Theorem A.2 *A subset G_l of a connected graph G with l branches has loops if and only if the columns of the incidence matrix A corresponding to these l branches are linear dependent.*

Proof.

(⇒) Let, without loss of generality, b_1, ..., b_r be branches of G_l that form a loop. Let a_1, ..., a_r be the columns of A corresponding to these branches. We introduce

$$c_j = \begin{cases} 1 & \text{if the branch } j \text{ has the same orientation as the loop} \\ -1 & \text{if the branch } j \text{ has the opposite orientation of the loop} \end{cases}$$

for all $j = 1, ..., r$. Since each node of G_l connects exactly two nodes of G_l, each row of the matrix

$$\begin{pmatrix} c_1 a_1 & c_2 a_2 & ... & c_r a_r \end{pmatrix}$$

contains exactly one 1 and exactly one -1. This implies

$$\sum_{i=1}^{r} c_i a_i = 0,$$

that means, the columns a_1, ..., a_r are linear dependent.

($\Leftarrow$) Let the subgraph G_l has l branches. Let, without loss of generality, the first l columns of A correspond to these branches. Due to our assumption, these columns are linear dependent. If G_l is not connected itself, then G_l consists of finite many connected, disjunct subgraphs G_{li} $(i = 1, ..., r)$. The subgraphs may be sorted such that

$$A_a = \begin{pmatrix} A_1 & 0 & \dots & 0 & * \\ 0 & A_2 & \dots & 0 & * \\ \vdots & \vdots & \ddots & \vdots & * \\ 0 & 0 & \dots & A_r & * \end{pmatrix}.$$
$$G_{l1} \quad G_{l2} \quad \dots \quad G_{lr}$$

Then we find at least one $i_0 \in \{1, ..., r\}$ such that the columns of A_{li_0} are linear dependent. Let k be the number of nodes of G_{li_0}. If G_{li_0} would not contain a loop, then G_{li_0} would be a tree with $k-1$ branches. Since A_{li_0} represents an incidence matrix for G_{li_0}, $k-1$ rows of A_{li_0} would be linear independent. But this would mean that also all $k-1$ columns of A_{li_0} are linear independent. This contradicts the assumption. It follows that G_{li_0} contains at least one loop and, hence, G_l contains a loop.

□

Theorem A.3 *Let A be the (reduced) incidence matrix of a connected graph with n nodes. Then, $n-1$ columns of A are linear independent if and only if the branches of these columns form a tree.*

Corollary A.4 *If A will be partitioned into $A := (A_t, A_r)$ such that the columns of A_t correspond to the rows of the tree, then A_t is nonsingular.*

Proof.

($\Rightarrow$) Let $A_s \in \mathbb{R}^{(n-1)\times(n-1)}$ be a nonsingular quadratic sub-matrix of A. Then, A_s belongs to a subgraph G_s of G consisting of n nodes and $n-1$ branches. Since the columns of A_s are linear independent, Theorem A.2 implies that G_s does not contain loops. Consequently, G_s is also connected and, finally, a tree.

($\Leftarrow$) Let the matrix of all branches belonging to the tree be denoted by A_t. Then, A_t is the reduced incidence matrix of a connected subgraph with $n-1$ branches. From Theorem A.1, it follows that A_t is nonsingular.

□

For more information about network graphs, we refer to [CDK87, CL75].

A.2 Properties of Projectors

Here, we shall summarize important properties of projectors. We begin with some definitions [Zei86]. Let X be a Banach space and V, W be linear subspaces of X.

1. A linear continuous operator $P : X \to X$ is called a *projector* or *projection operator* if $P^2 = P$.
2. A projector P projects *onto* a linear subspace V of X if $\operatorname{im} P = V$.
3. A projector P projects *along* a linear subspace V of X if $\ker Q = V$.
4. The sum $X = V \oplus W$ is called a *direct sum* if, for all $x \in X$, there exist exactly one $v \in V$ and exactly one $w \in W$ such that $x = v + w$.
5. The direct sum $X = V \oplus W$ is called a *topological direct sum* if the operators P and Q defined as

   $$Px = v \quad \text{and} \quad Qx = w \qquad \text{for } x = v + w, \quad v \in V,\, w \in W$$

   are continuous, that is, projection operators.

Thus, we have the following properties.

1. Let $P : X \to X$ be a projector and $I : X \to X$ be the identity map. Then, $Q := I - P$ is also a projector. If P projects onto V, then Q projects along V. If P projects along V, then Q projects onto V. This is a simple conclusion of the fact that

   $$\operatorname{im} P = \ker Q \qquad \text{and} \qquad \ker P = \operatorname{im} Q.$$

2. If V and W are linear subspaces of X such that

   $$V \oplus W = X,$$

   is a topological direct sum then there is a projector $Q : X \to X$ such that

   $$\operatorname{im} Q = V \qquad \text{and} \qquad \ker Q = W.$$

3. The sum $X = V \oplus W$ is a topological direct sum if and only if it is a direct sum and V, W are closed in X. Thus, each projector has a closed nullspace and a closed image space.

4. As a consequence of 2. and 3., we have for the finite-dimensional space $X = \mathbb{R}^m$ that there is always a projector $Q : \mathbb{R}^m \to \mathbb{R}^m$ such that

$$\operatorname{im} Q = V \qquad \text{and} \qquad \ker Q = W.$$

if $\mathbb{R}^m = V \oplus W$.

A.3 Index and Solvability of DAEs

A.3.1 Linear DAEs

In correspondence to the type of equations arising from circuit and device simulation, we shall consider DAEs with a so called properly stated leading term (see [Mär02a, Mär02b, Mär02c])

$$A(t)(D(t)x(t))' + B(t)x(t) = q(t), \tag{A.1}$$

where

$$\begin{aligned} A &\in C([t_0,T], L(\mathbb{R}^m,\mathbb{R}^n)), \quad D \in C([t_0,T], L(\mathbb{R}^n,\mathbb{R}^m)), \\ B &\in C([t_0,T], L(\mathbb{R}^n,\mathbb{R}^n)), \quad q(t) \in \mathbb{R}^n \end{aligned}$$

and $t \in [t_0, T]$ with $t_0 < T$ in $\mathbb{R}$.

Definition A.5
The leading term of (A.1) is stated properly if the coefficients $A(t)$ and $D(t)$ are well matched in the sense that

$$\ker A(t) \oplus \operatorname{im} D(t) = \mathbb{R}^m \qquad \forall\ t \in [t_0, T]$$

and there is a continuously differentiable projector function

$$R :\ [t_0, T] \to L(\mathbb{R}^m, \mathbb{R}^m)$$

such that $\operatorname{im} R(t) = \operatorname{im} D(t)$ and $\ker R(t) = \ker A(t)$ for all $t \in [t_0, T]$.

Remark A.6 If a DAE has a properly stated leading term, then we have

$$\operatorname{im} A(t)D(t) = \operatorname{im} A(t)R(t) = \operatorname{im} A(t)$$

and

$$\ker A(t)D(t) = \ker R(t)D(t) = \ker D(t).$$

Formulating DAEs in such a form provides advantages for the numerical integration. In this case, the discretization and the decoupling into the dynamic and non-dynamic parts commute if $\operatorname{im} D(t)$ and certain subspaces of it (depending on the index of the DAE) do not vary with t (see [HMTa, HMTb]). This implies that methods for ordinary differential equations remain there stability properties if applied to DAEs with properly stated leading term. Effects that implicit methods behave like explicit ones if applied to DAEs (cf. [AP91]) can be avoided.

The following index definition (see [Mär02c]) for DAEs of the form (A.1) generalizes the so-called global index proposed in [GP84] as well as the tractability index (see e.g. [Mär92]). Regarding the Galerkin equations from Chapter 4, we need solutions with only week smoothness assumptions. Consequently, other index notions related to derivative array systems and reduction techniques (e.g. [Cam87, KM94, RR96]) are not well suited for our considerations.

Definition A.7
An equation (A.1) with properly stated leading term has the index μ, if there is a continuous matrix function sequence G_i and a continuous projector function sequence Q_i such that

(i) $Q_i(t)$ is a projector onto $\ker G_i(t)$ for all $t \in [t_0, T]$,

(ii) $G_i(t)$ has constant rank $r_i > 0$ on $[t_0, T]$,

(iii) $r_{\mu-1} < r_\mu = n$,

(iv) $Q_i(t)Q_j(t) = 0$ for $j = 0, ..., i-1, \quad i > 0, \quad t \in [t_0, T]$, and

(v) $DP_0...P_iD^- \in C^1([t_0, T], L(\mathbb{R}^m, \mathbb{R}^m))$

where

$$\begin{aligned} G_{i+1} &= G_i + B_iQ_i, \quad G_0 = AD, \\ B_{i+1} &= B_iP_i - G_{i+1}D^-(DP_0...P_iD^-)'DP_0...P_i, \quad B_0 = B, \\ P_i &= I - Q_i, \end{aligned}$$

and D^- denotes the reflexive generalized inverse of D such that

$$D^-DD^- = D^-, \quad DD^-D = D, \quad DD^- = R, \quad D^-D = P_0.$$

Obviously, the natural solution space for (A.1) is given by

$$C_D^1([t_0,T],\mathbb{R}^n) := \{x \in C([t_0,T],\mathbb{R}^n) : \ Dx \in C^1([t_0,T],\mathbb{R}^m)\}.$$

Indeed the following theorem holds for initial value problems of index 1 (see [Mär02c])

$$\begin{aligned} A(t)(D(t)x(t))' + B(t)x(t) &= q(t), && \text{(A.2)}\\ D(t_0)x_0 &= z_0 \in \operatorname{im} D(t_0). && \text{(A.3)} \end{aligned}$$

Theorem A.8 *If $q \in C([t_0,T],\mathbb{R}^n)$, then the index-1 IVP* (A.2)-(A.3) *has a unique solution $x \in C_D^1([t_0,T],\mathbb{R}^n)$. Furthermore, the solution depends continuously on the data. More precisely, there is a constant $C > 0$ such that*

$$\max_{t_0\le t\le T} \|x(t)\| + \max_{t_0\le t\le T} \|(Dx)'(t)\| \le C\left(\|z_0\| + \max_{t_0\le t\le T} \|q(t)\|\right).$$

However, the Galerkin equations obtained in Chapter 4 have right hand sides q that belong to $L^2(t_0,T;\mathbb{R}^n)$ only. Therefore, we need to generalize Theorem A.8 as follows

Theorem A.9 *If $q \in L^2(t_0,T;\mathbb{R}^n)$, then the index-1 IVP* (A.2)-(A.3) *has a unique solution x in*

$$L_D^2(t_0,T;\mathbb{R}^n) := \{x \in L^2(t_0,T;\mathbb{R}^n) : \ Dx \in C([t_0,T],\mathbb{R}^m)\}.$$

The equation (A.2) *holds for almost all $t \in [t_0,T]$. Furthermore, Dx is differentiable for almost all $t \in [t_0,T]$ and there is a constant $C > 0$ such that*

$$\begin{aligned} \|x\|_{L^2(t_0,T;\mathbb{R}^n)} + \|Dx\|_{C([t_0,T],\mathbb{R}^m)} + \|(Dx)'\|_{L^2(t_0,T;\mathbb{R}^m)} & \\ \le\ C\left(\|z_0\| + \|q\|_{L^2(t_0,T;\mathbb{R}^n}\right). & \end{aligned}$$

For continuous solutions, the right hand side belonging to the non-dynamical part has to be continuous. The next theorem describes this more precisely.

Theorem A.10 *If $q \in L^2(t_0,T;\mathbb{R}^n)$ and $Q_0G_1^{-1}q \in C([t_0,T];\mathbb{R}^n)$, then the solution x of the index-1 IVP* (A.2)-(A.3) *belongs to $C([t_0,T];\mathbb{R}^n)$ and we find a constant $C > 0$ such that*

$$\begin{aligned} \|x\|_{C([t_0,T],\mathbb{R}^n)} + \|(Dx)'\|_{L^2(t_0,T;\mathbb{R}^m)} & \\ \le\ C\left(\|z_0\| + \|q\|_{L^2(t_0,T;\mathbb{R}^n)} + \|Q_0G_1^{-1}q\|_{C([t_0,T],\mathbb{R}^n)}\right). & \end{aligned}$$

Proof of Theorems A.9 and A.10: The proof is straightforward. We simply have to combine standard techniques from DAE and Volterra operator theory. Due to the index-1 assumption, the matrix

$$G_1(t) = A(t)D(t) + B(t)Q_0(t)$$

is nonsingular for all $t \in [t_0, T]$. Recall that $Q_0(t)$ is a projector onto $\ker A(t)D(t)$. Multiplying (A.2) by $D(t)G^{-1}(t)$ and $Q_0(t)G^{-1}(t)$, respectively, we obtain the system

$$\begin{aligned} (Dx)'(t) - R'(t)(Dx)(t) + (DG_1^{-1}BD^-)(t)(Dx)(t) &= (DG_1^{-1}r)(t), \quad \text{(A.4)} \\ (Q_0x)(t) + (Q_0G_1^{-1}BD^-)(t)(Dx)(t) &= (Q_0G_1^{-1}r)(t), \quad \text{(A.5)} \end{aligned}$$

which is equivalent to (A.2). Here, we have used the properties

$$(DG_1^{-1}A)(t) = R(t), \qquad (G_1^{-1}BQ_0)(t) = Q_0(t)$$

for all $t \in [t_0, T]$. Recall that $R(t) = D(t)D^-(t)$ is a continuously differentiable projector onto $\operatorname{im} D(t)$ along $\ker A(t)$ and $D^-(t)$ is a generalized inverse that satisfies $D^-(t)D(t) = P_0(t)$.
For $z := Dx$, equation (A.4) together with (A.3) represents an ordinary initial value problem of the form

$$z'(t) = \hat{A}(t)z(t) + b(t), \qquad z(t_0) = z_0 \tag{A.6}$$

with $\hat{A} \in C([t_0, T], L(\mathbb{R}^m, \mathbb{R}^m))$ and $b \in L^2(t_0, T; \mathbb{R}^m)$. Since $\hat{A}$ is linear and continuous, the map

$$x \mapsto \hat{A}(t)x$$

is Lipschitz continuous as map from $L^2(t_0, T; \mathbb{R}^m)$ into $L^2(t_0, T; \mathbb{R}^m)$ with a Lipschitz constant that is independent of t. Consequently (see e.g. [GGZ74], pp. 166-167), the IVP (A.6) has a unique solution $z \in C([t_0, T], \mathbb{R}^m)$ with $z' \in L^2(t_0, T; \mathbb{R}^m)$. The solution z satisfies (A.6) for almost all $t \in [t_0, T]$ and it is differentiable for almost all $t \in [t_0, T]$. Furthermore, there is a constant $C_1 > 0$ such that

$$\|z\|_{C([t_0,T],\mathbb{R}^m)} + \|z'\|_{L^2(t_0,T;\mathbb{R}^m)} \le C_1 \left(\|z_0\| + \|b\|_{L^2(t_0,T;\mathbb{R}^m)} \right). \tag{A.7}$$

In [GGZ74], this was proven not only for maps into the finite-dimensional space $\mathbb{R}^m$ but also for maps into any Banach space. In the finite-dimensional case, the unique solvability of (A.6) and the validity of the estimation (A.7) follow also from the theorem of Carathéodory (see e.g. [Zei90b], [Kam60]), an a priori estimate and the generalized Gronwall lemma (see e.g. [Zei86]).

For convenience reasons, we omit an extended explanation of the second way. Multiplying (A.4) by $I - R(t)$, we obtain that

$$((I - R)z)'(t) = -R'(t)((I - R)z)(t)$$

for the solution z and almost all $t \in [t_0, T]$. Since z_0 belongs to $\operatorname{im} D(t_0)$, we get

$$((I - R)z)(t_0) = 0.$$

Using again the unique solvability, we obtain that

$$((I - R)z)(t) = 0 \qquad \text{for almost all } t \in [t_0, T]. \tag{A.8}$$

Regarding (A.5), we see that all solutions of (A.2)-(A.3) are given by

$$x(t) = D^-(t)z(t) - (Q_0 G_1^{-1} B D^-)(t)z(t) + (Q_0 G_1^{-1} r)(t), \tag{A.9}$$

where z is the unique solution of (A.6). Obviously, $Dx = z$ belongs to $C([t_0, T], \mathbb{R}^m)$. Since D, R and P_0 are continuous on $[t_0, T]$, the generalized inverse D^- is continuous. It implies $x \in L^2(t_0, T; \mathbb{R}^n)$ since $r \in L^2(t_0, T; \mathbb{R}^n)$. Recall that G_1 is continuous due to the index-1 assumption. If, additionally, $Q_0 G^{-1} r$ is continuous on $[t_0, T]$, then the whole solution x belongs to $C([t_0, T], \mathbb{R}^n)$. The estimations of the Theorems A.9 and A.10 are a simple conclusion of the solution representation (A.9) and the estimation (A.7).

□

Remark A.11 Similar L_2-results for DAEs of the form

$$A(t)x'(t) + B(t)x(t) = q(t)$$

are given in [Han89].

A.3.2 Nonlinear DAEs

Regarding the form of the network equations (1.17)-(1.19), we consider quasi-linear DAEs

$$A\frac{\mathrm{d}}{\mathrm{d}t}d(x(t), t) + b(x(t), t) = 0 \tag{A.10}$$

with continuous matrix functions

$$b(x, t) : \ \Omega \times [t_0, T] \to \mathbb{R}^n, \quad d(x, t) : \ \Omega \times [t_0, T] \to \mathbb{R}^m$$

and $\Omega \in \mathbb{R}^m$ is an open, connected domain. The partial derivatives b_x, d_x, d_t, d_{xt} and d_{xx} are assumed to exist and to be continuous. Having positive

definite capacitance and inductance matrices for network equation systems in mind, we only need to consider DAEs with a constant nullspace $\ker AD(x,t)$ for $D(x,t) := d_x(x,t)$. Additionally, we may assume the leading term to be properly stated, i.e., the decomposition

$$\ker A \oplus \operatorname{im} D(x,t) = \mathbb{R}^m \quad \text{for all} \quad x \in \Omega,\ t \in [t_0, T]$$

is valid and $\operatorname{im} D(x,t)$ is constant. Obviously, we have again

$$\operatorname{im} AD(x,t) = \operatorname{im} AR = \operatorname{im} A$$

and

$$\ker AD(x,t) = \ker RD(x,t) = \ker D(x,t)$$

for the projector R onto $\operatorname{im} D(x,t)$ along $\ker A$.

Remark A.12 If one uses

$$d(x,t) := \begin{pmatrix} q_C(A_C^{\mathrm{T}} e, t) \\ \phi_L(j_L, t) \end{pmatrix}$$

for the network equations (1.17)-(1.19), then the image space $\operatorname{im} D(x,t)$ does not need to be constant. However, the formulation with

$$d(x,t) := \begin{pmatrix} \bar{P}_C q_C(A_C^{\mathrm{T}} e, t) \\ \phi_L(j_L, t) \end{pmatrix}, \qquad A := \begin{pmatrix} A_C & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix}$$

yields a DAE of the form (A.10) with a properly stated leading term for any projector $\bar{P}_C$ along the nullspace $\ker A_C$.

A computation of such a projector $\bar{P}_C$ is not necessary in practice but it simplifies the analysis significantly.

The natural solution space for DAEs of the form (A.10) is given by

$$C_d^1([t_0,T],\mathbb{R}^n) := \{x(\cdot) \in C([t_0,T],\mathbb{R}^n) :\ d(x(\cdot),\cdot) \in C^1([t_0,T],\mathbb{R}^m)\}.$$

Remark A.13 If P_0 is any constant projector along the constant nullspace $\ker AD(x,t)$, then the natural solution space can be formulated, equivalently, as (see [Mär02b])

$$C_d^1([t_0,T],\mathbb{R}^n) = \{x \in C([t_0,T],\mathbb{R}^n) :\ P_0 x \in C^1([t_0,T],\mathbb{R}^n)\}.$$

This follows, via the implicit function theorem, from the facts that $AD(x,t)$ maps $\operatorname{im} P_0$ bijectively onto $\operatorname{im} A$ and $\ker AD(x,t) = \ker D(x,t)$.

Definition A.14
The DAE (A.10) with a properly stated leading term has index μ [Mär02b], if there is a continuous matrix function sequence $G_i(x^{i-1}, ..., x^1, x, t)$ and a continuous projector function sequence $Q_i(x^{i-1}, ..., x^1, x, t)$ such that, for all $x^{i-1}, ..., x^1 \in \mathbb{R}^n$, $x \in \Omega$, $t \in [t_0, T]$:

(i) $Q_i(x^{i-1}, ..., x^1, x, t)$ is a projector onto $\ker G_i(x^{i-1}, ..., x^1, x, t)$,

(ii) $G_i(x^{i-1}, ..., x^1, x, t)$ has constant rank $r_i > 0$,

(iii) $r_{\mu-1} < r_\mu = n$,

(iv) $Q_i(x^{i-1}, ..., x^1, x, t)Q_j(x^{j-1}, ..., x^1, x, t) = 0$ for $j = 0, ..., i-1, \quad i > 0$,

(v) $DP_0...P_iD^-$ is continuously differentiable,

where

$$
\begin{aligned}
G_{0(x,t)} &= AD_{(x,t)}, \\
G_{i+1(x^i,...,x^1,x,t)} &= G_{i(x^{i-1},...,x^1,x,t)} + B_{i(x^i,...,x^1,x,t)}Q_{i(x^{i-1},...,x^1,x,t)}, \\
B_{0(x,t)} &= \tfrac{\partial b(x,t)}{\partial x}, \\
B_{i+1(x^{i+1},...,x^1,x,t)} &= B_{i(x^i,...,x^1,x,t)}P_{i(x^{i-1},...,x^1,x,t)} \\
&\quad - G_{i+1}D^-\mathrm{Diff}_{i+1(x^{i+1},...,x^1,x,t)}(DP_0...P_i)_{(x^{i-1},...,x^1,x,t)}, \\
\mathrm{Diff}_{1(x^1,x,t)} &= \tfrac{\partial(DP_0P_1D^-)(x,t)}{\partial x}x^1 + \tfrac{\partial(DP_0P_1D^-)(x,t)}{\partial t}, \\
\mathrm{Diff}_{i+1(x^{i+1},...,x^1,x,t)} &= \sum_{j=1}^{i} \tfrac{\partial(DP_0...P_{i+1}D^-)(x^i,...,x^1,x,t)}{\partial x_j}x^{j+1} \\
&\quad + \tfrac{\partial(DP_0...P_{i+1}D^-)(x^i,...,x^1,x,t)}{\partial x}x^1 + \tfrac{\partial(DP_0...P_{i+1}D^-)(x^i,...,x^1,x,t)}{\partial t}, \\
P_{i(x^{i-1},...,x^1,x,t)} &= I - Q_{i(x^{i-1},...,x^1,x,t)},
\end{aligned}
$$

and $D^-(x,t)$ denotes the reflexive generalized inverse of $D(x,t)$ such that

$$D^-D^- = D^-, \quad DD^-D = D, \quad DD^- = R, \quad D^-D = P_0.$$

This index definition is closely related to the index of linearizations of DAEs of the form (A.10). Let x_* belong to the natural solution space $C_d^1([t_0, T], \mathbb{R}^n)$ with $x(t) \in \Omega$ for all $t \in [t_0, T]$. Then, the linearization of (A.10) along x_* is given by

$$A\frac{\mathrm{d}}{\mathrm{d}t}(D_*(t)x(t)) + B_*(t)x(t) = q(t) \tag{A.11}$$

with

$$D_*(t) := D(x_*(t), t) \quad \text{and} \quad B_*(t) := \tfrac{\partial}{\partial x} b(x_*(t), t).$$

Theorem A.15 *[Mär02b] Assume the DAE (A.10) to have index μ and x_* to belong to $C^{\mu-1}([t_0, T], \mathbb{R}^n)$ with $x(t) \in \Omega$ for all $t \in [t_0, T]$. Then, the linearized DAE* (A.11) *has also index μ. Furthermore, the characteristic values r_{*i} of the linearization equal the characteristic values r_i of (A.10).*

The next theorem shows that index-1 DAEs of the form (A.10) completed by consistent initial conditions are uniquely solvable provided that $d(x, t)$ is linear in x. Furthermore, the solution depends continuously on perturbations of the right hand side and the initial data.

Theorem A.16 *[HM] Let the DAE* (A.10) *have the index 1 and $d(x, t) = D(t)x(t)$ be given.*

(i) Through each x_0 in $\{x \in \mathcal{D}(t_0) : \ b(x, t) \in \operatorname{im} A\}$ passes exactly one solution of (A.10).

(ii) For a solution in $C_D^1([t_0, T], \mathbb{R}^n)$, all perturbed IVPs

$$A\frac{d}{dt}D(t)x(t) + b(x(t), t) = q(t), \qquad D(t_0)(x(t_0) - x^0) = 0, \quad \text{(A.12)}$$

$x^0 \in \mathbb{R}^n$, $q \in C([t_0, T], \mathbb{R}^n)$, are uniquely solvable on $C_D^1([t_0, T], \mathbb{R}^n)$ supposed the perturbations $|D(t_0)(x_*(t_0) - x^0)|$ and $\|q\|_\infty$ are sufficiently small.

(iii) For the solution x of (A.12) *it holds that*

$$\|x - x_*\|_\infty \le const(|D(t_0)(x_*(t_0) - x(t_0)| + \|q\|_\infty)$$

Remark A.17 Theorem A.16 was also shown for more general leading functions A depending on x and t in [HM].

Remark A.18 The matrix $G_2(x^1, x, t)$ is nonsingular if and only if the matrix

$$\bar{G}_2(x, t) := G_1(x, t) + B_0(x, t)P_0(x, t)Q_1(x, t)$$

is nonsingular since

$$G_2(x^1, x, t) = \bar{G}_2(x, t)[I - (P_1 D^-)(x, t)\mathrm{Diff}_1(x^1, x, t)(DP_0Q_1)(x, t)]$$

and $I - P_1(x, t)MQ_1(x, t)$ is nonsingular with the inverse $I + P_1(x, t)MQ_1(x, t)$ for any matrix M with appropriate matrix dimensions. Obviously, the computation of $\bar{G}_2$ is easier than that one of G_2. Therefore, we have used $\bar{G}_2$ for the index determination of the network equation systems in Chapter 1.

A.4 Basics for Evolution Equations

This section summarizes basic spaces and their properties for the treatment of evolution equations (see e.g. [Zei90a]).

1. Dual space. Let V be a real Banach space. Then, V^* denotes the set of all linear continuous functionals on V, i.e., the set of all linear continuous maps $f: V \to \mathbb{R}$. Furthermore,

$$\langle f, v\rangle := f(v) \qquad \text{for all} \quad v \in V$$

and

$$\|f\|_{V^*} := \sup_{\|v\|_V \le 1} |\langle f, v\rangle|.$$

This way, V^* becomes a real Banach space. It is called the dual space to V.

2. Reflexive Banach space. Let V be a real Banach space. Then, V is called *reflexive* if $V = V^{**}$.

3. Evolution triple. The spaces $V \subseteq H \subseteq V^*$ are called an evolution triple if

   (i) V is a real, separable, and reflexive Banach space,

   (ii) H is a real, separable Hilbert space,

   (iii) The embedding $V \subseteq H$ is continuous, i.e.,

$$\|v\|_H \le \text{const}\|v\|_V \qquad \text{for all} \quad v \in V,$$

   and V is dense in H.

   Below, Proposition A.20 explains how the inclusion $H \subseteq V^*$ is to be understood.

4. The Lebesgue space $L_p(t_0, T; V)$ of vector-valued functions. Let V be a Banach space, $1 < p < \infty$, and $t_0 < T < \infty$. The space $L_p(t_0, T; V)$ consists of all measurable functions $v: (t_0, T) \to V$ for which

$$\|v\|_p := \left(\int_{t_0}^{T} \|v(t)\|_V^p \, \mathrm{d}t\right)^{\frac{1}{p}} < \infty.$$

   The dual space of $L_p(t_0, T; V)$ is given by $L_q(t_0, T; V^*)$ where $p^{-1} + q^{-1} = 1$.

5. Generalized Derivatives. Let X and Y be Banach spaces. Furthermore, let $u \in L_1(t_0, T; X)$ and $w \in L_1(t_0, T; Y)$. Then, the function w is called the generalized derivative of the function u on (t_0, T) if

$$\int_{t_0}^{T} \varphi'(t)u(t)\ \mathrm{d}t = -\int_{t_0}^{T} \varphi(t)w(t)\ \mathrm{d}t \qquad \text{for all} \quad \varphi \in C_0^\infty(t_0, T).$$

The last equation includes the requirement that the integrals on both sides belong to $X \cap Y$.

6. The Sobolev space $W_2^1(t_0, T; V, H)$. Let $V \subseteq H \subseteq V^*$ be an evolution triple and $t_0 < T < \infty$. Then, the Sobolev space

$$W_2^1(t_0, T; V, H) := \{u \in L_2(t_0, T; V) :\ u' \in L_2(t_0, T; V^*)\}$$

forms a Banach space with the norm

$$\|u\|_{W_2^1} = \|u\|_{L_2(t_0,T;V)} + \|u'\|_{L_2(t_0,T;V^*)}.$$

The following proposition is a consequence of Riesz theorem.

Proposition A.19 *Let H be a Hilbert space. Then for each $u \in H$, there is a unique linear continuous functional Ju on V with*

$$\langle Ju, v\rangle = (u|v) \qquad \textit{for all} \quad u, v \in V,$$

where $(\cdot|\cdot)$ denotes the scalar product of H. The operator $J : V \to V^*$ is linear, bijective, and norm isomorphic, i.e.,

$$\|Ju\|_{V^*} = \|u\|_V \qquad \textit{for all} \quad u \in V.$$

Therefore, one can identify Ju with u for all $u \in V$. This way we get $H = H^*$ and

$$\langle u, v\rangle = (u|v) \qquad \text{for all} \quad u, v \in V.$$

The next proposition explains how the relation $H \subseteq V^*$ has to be understood.

Proposition A.20 *Let $V \subseteq H \subseteq V^*$ be an evolution triple. Then, the following is satisfied*

(i) To each $u \in H$, there corresponds a linear continuous functional $\bar{u} \in V^*$ with

$$\langle \bar{u}, v\rangle_V = (u|v)_H \quad \textit{for all} \quad v \in V.$$

(ii) The mapping $u \mapsto \bar{u}$ from H into V^* is linear, injective, and continuous.

PROOF: Ad(i) Let $u \in H$. Then:

$$|(u|v)_H| \leq \|u\|_H \|v\|_H \leq \text{const} \|u\|_H \|v\|_V$$

is fulfilled for all $v \in V$. Therefore, there exists a $\bar{u} \in V^*$ with

$$\langle \bar{u}, v \rangle_V = (u|v)_H \qquad \text{and} \qquad \|\bar{u}\|_{V^*} \leq \text{const} \|u\|_H.$$

Ad(ii) The mapping $u \mapsto \bar{u}$ is obviously linear and continuous. In order to show injectivity, we assume that $\bar{u} = 0$. This implies

$$(u|v)_H = 0 \qquad \text{for all} \quad v \in V.$$

Since V is dense in H, we get $u = 0$.

□

This allows us to identify $\bar{u}$ with u such that

$$\begin{aligned} \langle u, v \rangle_V &= (u|v)_H \qquad \text{for all} \quad u \in H,\ v \in V, \\ \|u\|_{V^*} &\leq \text{const} \|u\|_H \qquad \text{for all} \quad u \in H. \end{aligned}$$

In this sense, the relation $H \subseteq V^*$ is to be understood. Obviously, this embedding is continuous.

Appendix B

Notations

B.1 Symbols

$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
C	capacitance
D	diffusivity
D	electric displacement
E	electric field
J	current density
L	inductance
N	doping concentration
N_D^+	donor concentration
N_A^-	acceptor concentration
R	1. resistance (network element)
	2. generation/recombination rate (drift-diffusion model)
T	1. temperature
	2. end of a time interval
V	electrostatic potential
V_{bi}	built-in potential
e	nodal potential
j	current
n	electron density
n_i	intrinsic density
p	hole density
q	charge

t	time
x	position variable
Ω	domain
ε	permittivity
μ	permeability/mobility
ν	outer unit normal vector
ϱ	charge density
σ	surface charge
ϕ	flux
ϕ_B	barrier height
$\forall$	for all
$\exists$	it exists

B.2 General Notations

$x \in S$	x is an element of the set S
$x \notin S$	x is not an element of S
$S \subseteq X$	S is contained in X
$X \times Y$	product set, $= \{(x, y) : \ x \in X, \, y \in Y\}$
$X \oplus Y$	direct sum of X and Y
$L(X, Y)$	space of linear continuous operators from X into Y
$C(X, Y)$	space of continuous operators from X into Y
$C^k(X, Y)$	space of k-times Frèchet-differentiable functions from X into Y
$\dim X$	dimension of a linear space X
$\operatorname{rank} A$	rank of a linear operator A, $\operatorname{rank} A = \dim \operatorname{im} A$
A^{T}	transpose of a matrix A
$f : \ X \to Y$	mapping from X into Y
$\operatorname{im} f$	image of f, $\operatorname{im} f = \{f(x) : \ x \in X\}$
$\ker f$	kernel of f, $\ker f = \{x : \ f(x) = 0\}$
f injective	$\ker f = \{0\}$
f surjective	$\operatorname{im} f = Y$
f bijective	f is injective and surjective
$\partial_x f$	partial derivative of f in direction of x
$\nabla f = \operatorname{grad} f$	gradient of f, $\operatorname{grad} f = (\partial_{x_1} f, ..., \partial_{x_n} f)$
$\operatorname{div} f$	divergence of f, $\operatorname{div} f = \partial_{x_1} f_1 + ... + \partial_{x_n} f_n$
Δf	Laplace operator, $\Delta f = \operatorname{div} \operatorname{grad} f$
$\langle f, x \rangle$	value of the linear functional f at the point x
$(x \mid y)$	scalar product of x and y in a Hilbert space
$x \cdot y$	scalar product in $\mathbb{R}^m$

Ω	nonempty open bounded set in $\mathbb{R}^N$
$\partial\Omega$	boundary of Ω
$\bar{\Omega}$, cl(Ω)	closure of Ω, $\bar{\Omega} = \Omega \cup \partial\Omega$
$C^\infty(\Omega)$	space of infinitely continuously differentiable functions $u : \Omega \to \mathbb{R}$
$C_0^\infty(\Omega)$	space of all functions $u \in C^\infty(\Omega)$ with compact support in Ω
$L_p(\Omega)$	Lebesgue space of all measurable functions $u : \Omega \to \mathbb{R}$ with $\|u\|_p = \left(\int_\Omega \|u\|^p \,\mathrm{d}x\right)^{1/p} < \infty$
$H^1(\Omega)$	Sobolev space of all measurable functions $u : \Omega \to \mathbb{R}$, for which the generalized derivatives $\partial_i u$ $(i = 1, ..., N)$ exist and $\|u\|_{1,2} = \left(\int_\Omega (\|u\|^2 + \sum_{i=1}^N \|\partial_i u\|^2) \,\mathrm{d}x\right)^{1/2} < \infty$
$H_0^1(\Omega)$	closure of $C_0^\infty(\Omega)$ in $H^1(\Omega)$
$H(\mathrm{div}\,;\Omega)$	space of all measurable functions $u : \Omega \to \mathbb{R}^k$, for which $u \in L_2(\Omega)^k$ and $\mathrm{div}\, u \in L_2(\Omega)$
$H^2(\Omega)$	Sobolev space of all measurable functions $u : \Omega \to \mathbb{R}$, for which the generalized derivatives $D^\alpha u$ $(\|\alpha\| \le 2)$ exist and $\|u\|_{2,2} = \left(\int\limits_\Omega \Big(\sum\limits_{\|\alpha\|\le 2} \|D^\alpha u\|^2\Big) \,\mathrm{d}x\right)^{1/2} < \infty$
X^*	dual space of X
f^*	dual mapping of f
$x_n \to x$	convergence of x_n to x
$x_n \rightharpoonup x$	weak convergence of x_n to x

B.3 Physical constants

Boltzmann constant	k	$1.38066 \cdot 10^{-23}$ J/K
Elementary charge	q	$1.60218 \cdot 10^{-19}$ C
Thermal voltage at 300 K	$\frac{kT}{q}$	0.0259 V
Permeability in vacuum	μ_0	$1.25663 \cdot 10^{-8}$ H/cm
Permittivity in vacuum	ε_0	$8.85418 \cdot 10^{-14}$ F/cm

Bibliography

[ABGT02] G. Alì, A. Bartel, M. Günther, and C. Tischendorf. Elliptic partial differential-algebraic multiphysics models in electrical network design. Technical Report 02/05, Institute of Scientific Computing and Mathematical Modeling, University of Karlsruhe, 2002. To appear in *Math. Models Meth. Appl. Sci. 2003.*

[AGH02] G. Albinus, H. Gajewski, and R. Hünlich. Thermodynamic design of energy models of semiconductor devices. *Nonlinearity*, 15(2):367–383, 2002.

[AP91] U. Ascher and L. Petzold. Stability of computational methods for constrained dynamics systems. *SIAM J. Sci. Stat. Comput.*, 1:95–120, 1991.

[BC95] J. Butcher and P. Chartier. Parallel general linear methods for stiff ordinary differential and differential algebraic equations. *Applied Numerical Mathematics*, 17:213–222, 1995.

[BC98] J. C. Butcher and R.P.K. Chan. Efficient Runge-Kutta integrators for index-2 differential algebraic equations. *Math. Comp.*, 67:1001–1021, 1998.

[BCP89] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *The Numerical Solution of Initial Value Problems in Ordinary Differential-Algebraic Equations.* North Holland Publishing Co., 1989.

[BF98] V. Barbu and A. Favini. Existence for an implicit nonlinear differential equations. *Nonlinear Anal.*, 32(1):33–40, 1998.

[BM64] R. K.. Brayton and J. K. Moser. A theory of nonlinear networks I, II. *Quart. Appl. Math.*, 22:1–33, 81–104, 1964.

[BM02] K. Balla and R. März. A unified approach to linear differential algebraic equations and their adjoints. *Z. Anal. Anwend.*, 21(3):783–802, 2002.

[Cam80] S. L. Campbell. *Singular Systems of Differential Equations I.* Research Notes in Math.; 40. Pitman, Marshfield, 1980.

[Cam81] S. L. Campbell. A procedure for analyzing a class of nonlinear semistate equations that arise in circuit and control problems. *IEEE Transactions on Circuits and Systems*, CAS-28:256–261, 1981.

[Cam82] S. L. Campbell. *Singular Systems of Differential Equations II.* Research Notes in Math.; 61. Pitman, Marshfield, 1982.

[Cam85] S. L. Campbell. The numerical solution of higher index linear time varying singular systems of differential equations. *SIAM J. Sci. Stat. Comput.*, 6:334–348, 1985.

[Cam87] S. L. Campbell. A general form for solvable linear time varying singular systems of differential equations. *SIAM J. Math. Anal.*, 18(4):1101–1115, 1987.

[CDK87] L. O. Chua, Ch. A. Desoer, and E. S. Kuh. *Linear and nonlinear Circuits.* McGraw-Hill Book Co., Singapore, 1987.

[CG95a] S. L. Campbell and C. W. Gear. The index of general nonlinear DAEs. *Numerische Mathematik*, 72:173–196, 1995.

[CG95b] S. L. Campbell and E. Griepentrog. Solvability of general differential algebraic equations. *SIAM J. Sci. Comp.*, 16:257–270, 1995.

[CL75] L. O. Chua and Pen-Min Lin. *Computer-Aided Analysis of Electronic Circuits.* Prentice Hall, Englewood Cliffs, 1975.

[CM99a] S. L. Campbell and W. Marszalek. The index of an infinite dimensional implicit system. *Math. Comput. Model. Dyn. Syst.*, 5(1):18–42, 1999.

[CM99b] S. L. Campbell and W. Marszalek. The interaction between simulation and modeling in infinte dimensional systems, 1999. Department of Mathematics, North Carolina State University, Raleigh, USA.

[CR65] L. O. Chua and R. A. Rohrer. On the dynamic equations of a class of nonlinear RLC-networks. *IEEE Trans. Circuit Theory*, 12:475–489, 1965.

[DHZ87] P. Deuflhard, E. Hairer, and J. Zugck. One-step and extrapolation methods for differential-algebraic systems. *Numer. Math.*, 51:501–516, 1987.

[EFM+03] D. Estévez Schwarz, U. Feldmann, R. März, S. Sturtzel, and C. Tischendorf. Finding beneficial DAE structures in circuit simulation. In W. Jäger and H.-J. Krebs, editors, *Mathematics-Key Technology for the Future: Joint Projects Between Universities and Industry*, pages 413–428. Springer, Berlin, 2003.

[ELD82] W. L. Engl, R. Laur, and H. K. Dirks. MEDUSA – A simulator for modular circuits. *IEEE Trans. CAD*, 1(2):85–93, 1982.

[ESF98] E. Eich-Soellner and C. Führer. *Numerical Methods in Multibody Dynamics*. B.G.Teubner, Stuttgart, 1998.

[Est00] D. Estévez Schwarz. *Consistent initialization for index-2 differential algebraic equations and its application to circuit simulation*. PhD thesis, Humboldt-Univ. Berlin, 2000.

[Est02] D. Estévez Schwarz. A step-by-step approach to compute a consistent initialization for the MNA. *Int. J. Circ. Theor. Appl.*, 30(1):1–6, 2002.

[ESTZ96] K. Einwich, P. Schwarz, P. Trappe, and H. Zojer. Simulatorkopplung für den Entwurf komplexer Schaltkreise der Nachrichtentechnik. In *7. ITG-Fachtagung "Mikroelektronik für die Informationstechnik"*, pages 139–144, Chemnitz, 1996.

[ET00] D. Estévez Schwarz and C. Tischendorf. Structural analysis of electric circuits and consequences for MNA. *Int. J. Circ. Theor. Appl.*, 28:131–162, 2000.

[Fos92] M. Fosséprez. *Non-linear Circuits: Qualitative Analysis of Non-linear, Non-reciprocal Circuits*. John Wiley & Sons, Chichester, 1992.

[FP86] A. Favini and P. Plazzi. Some results concerning the abstract nonlinear equation $D_t M u(t) + Lu(t) = f(t, Ku(t))$. *Circuits Systems Signal Process.*, 5:261–274, 1986.

[FP89] A. Favini and P. Plazzi. On some abstract degenerate problems of parabolic type 2. the nonlinear case. *Nonlinear Anal.*, 13:23–31, 1989.

[FR99] A. Favini and A. Rutkas. Existence and uniqueness of solutions of some abstract degenerate nonlinear equations. *Diff. Int. Eqs*, 12(3):373–394, 1999.

[FY99] A. Favini and A. Yagi. *Degenerate differential equations in Banach spaces*. Pure and Applied Mathematics, Marcel Dekker, New York, 1999.

[Gaj93] H. Gajewski. Analysis und Numerik des Ladungstransports in Halbleitern. *GAMM Mitteilungen*, 16:35–57, 1993.

[Gea71] C. W. Gear. Simultaneous numerical solution of differential-algebraic equations. *IEEE Trans. Circuit Theory*, CT-18(1):89–95, 1971.

[GF95] M. Günther and U. Feldmann. The DAE-index in electric circuit simulation. *Mathematics and Computers in Simulation*, 39:573–582, 1995.

[GF99a] M. Günther and U. Feldmann. CAD based electric modeling in industry. Part I: Mathematical structure and index of network equations. *Surv. Math. Ind.*, 8:97–129, 1999.

[GF99b] M. Günther and U. Feldmann. CAD based electric modeling in industry. Part II: Impact of circuit configurations and parameters. *Surv. Math. Ind.*, 8:131–157, 1999.

[GG86] H. Gajewski and K. Gröger. On the basic equations for carrier transport in semiconductors. *J. Math. Anal. Appl.*, 113:12–35, 1986.

[GG89] H. Gajewski and K. Gröger. Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics. *Math. Nachr.*, 140:7–36, 1989.

[GGZ74] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*. Akademie-Verlag, Berlin, 1974.

[GM86] E. Griepentrog and R. März. *Differential-Algebraic Equations and Their Numerical Treatment*. Teubner-Texte zur Mathematik No. 88. BSB B.G. Teubner Verlagsgesellschaft, Leipzig, 1986.

[GP84] C. W. Gear and L. R. Petzold. ODE methods for the solution of differential/algebraic systems. *SIAM J. Numer. Anal.*, 21:716–728, 1984.

[GR96] M. Günther and P. Rentrop. The differential-algebraic index concept in electric circuit simulation. *Z. f. angew. Math. u. Mech.*, 76, Suppl.1:91–94, 1996.

[Gri83] R.D. Grigorieff. Stability of multistep-methods on variable grids. *Numer. Math.*, 42:359–377, 1983.

[Grö87] K. Gröger. Initial boundary value problems from semiconductor device theory. *ZAMM*, 67(8):345–355, 1987.

[Gün01a] M. Günther. *Partielle differential-algebraische Systeme in der numerischen Zeitbereichsanalyse elektrischer Schaltungen*. Number 343 in Fortschritt-Berichte VDI, Reihe 20, Rechnerunterstützte Verfahren. VDI Verlag, Düsseldorf, 2001. Habilitation.

[Gün01b] M. Günther. A PDAE model for interconnected linear RLC networks. *Mathematical and Computer Modelling of Dynamical Systems*, 7(2):189–203, 2001.

[Han89] M. Hanke. Linear differential-algebraic equations in spaces of integrable functions. *J. Diff. Eqs.*, 79:14–30, 1989.

[Han01] M. Hanke. A new implementation of a BDF method within the method of lines. Technical Report 2001:01, Royal Instiute of Technology and Uppsala University, Sweden, 2001.

[Has86] M. Hasler. Non-linear non-reciprocal resistive circuits with a structurally unique solution. *Int. J. Circ. Theor. Appl.*, 14:237–262, 1986.

[HB84] B. C. Haggman and P. R. Bryant. Solutions of singular constrained differential equations: A generalization of circuits containing capacitor-only loops and inductor-only cutsets. *IEEE Trans. CAS*, 31:1015–1029, 1984.

[HL03] M. Hanke and R. Lamour. Consistent initialization for nonlinear index-2 differential-algebraic equation: large sparse systems in matlab. *Numerical Algorithms*, 32:67–85, 2003.

[HLR89] E. Hairer, Ch. Lubich, and M. Roche. *The Numerical Solution of Differential-Algebraic Equations by Runge-Kutta Methods*. Lecture Notes in Mathematics Vol. 1409. Springer, Heidelberg, 1989.

[HM] I. Higueras and R. März. Differential-algebraic equations with properly stated leading terms. To appear in Computers Math. Appl.

[HMTa] I. Higueras, R. März, and C. Tischendorf. Stability preserving intgeration of index-1 DAEs. To appear in APNUM.

[HMTb] I. Higueras, R. März, and C. Tischendorf. Stability preserving intgeration of index-2 DAEs. To appear in APNUM.

[HS02] I. Higueras and G. Söderlind. Logarithmic norms and nonlinear DAE stability. *BIT*, 42(4):823–841, 2002.

[HW91] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and differential-algebraic problems*. Springer Series in Computational Mathematics 14. Springer-Verlag, Berlin, Heidelberg, 1991.

[Jün01] A. Jüngel. *Quasi-hydrodynamic Semiconductor Equations*. Birkhäuser, Basel, 2001.

[Kam60] E. Kamke. *Das Lebesgue-Stieltjes Integral*. Teubner, Leipzig, 1960.

[KM94] P. Kunkel and V. Mehrmann. Canonical forms for linear differential-algebraic equations with variable coefficients. *J. Comput. Appl. Math.*, 56:225–251, 1994.

[KM98] P. Kunkel and V. Mehrmann. Regular solutions of nonlinear differential-algebraic equations and their numerical determination. *Numer. Math.*, 79:581–600, 1998.

[KR65] E. S. Kuh and R. A. Rohrer. The state-variable approach to network analysis. *Proc. IEEE*, 53:672–686, 1965.

[KR99] A. Kvaernø and P. Rentrop. Low order multirate Runge-Kutta methods in electric circuit simulation. Technical Report 2/1999, Norwegian Univ. of Science and Technology, Trondheim, Norway, 1999.

[Kur93] G.A. Kurina. Singular perturbations of control problems with equation of state not solved for the derivative (a survey). *Journal of Computer and System Science International*, 31(6):17–45, 93.

[Kva90] A. Kvaernø. Runge-Kutta methods applied to fully implicit differential-algebraic equations. *Math. Comp.*, 54(190):583–625, 1990.

[Lam97] R. Lamour. A shooting method for fully implicit index-2 differential algebraic equations. *SIAM J. Sci. Comput.*, 18(1):94–114, 1997.

[Lan01] J. Lang. *Adaptive multilevel solution of nonlinear parabolic PDE systems. Theory, algorithm, and applications*. Number 16 in Lecture Notes in Computational Science and Engineering. Springer Berlin, 2001.

[LMM93] R. Lamour, R. März, and R.M.M. Mattheij. On the stability behaviour of systems obtained by index-reduction. *J. Comp. Appl. Math.*, 56:305–319, 1993.

[LMT01] R. Lamour, R. März, and C. Tischendorf. PDAEs and further mixed systems as abstract differential algebraic systems. Technical Report 01-11, Institute of Mathematics, Humboldt-Univ. of Berlin, 2001.

[LSEL99] W. Lucht, K. Strehmel, and C. Eichler-Liebenow. Indexes and special discretization methods for linear partial differential algebraic equations. *BIT*, 39(3):484–512, 1999.

[LSK+96] J. Litsios, B. Schmithüsen, U. Krumbein, A. Schenk, E. Lyumkis, B. Polsky, and W. Fichtner. *DESSIS 3.0 Manual*. ISE Integrated Systems Engineering, Zürich, 1996.

[Mär92] R. März. Numerical methods for differential-algebraic equations. *Acta Numerica*, pages 141–198, 1992.

[Mär02a] R. März. Differential algebraic systems anew. *Appl. Num. Math.*, 42:315–335, 2002.

[Mär02b] R. März. Differential algebraic systems with properly stated leading term and MNA equations. Technical Report 2002-13, Institut für Mathematik, Humboldt-Univ. zu Berlin, 2002.

[Mär02c] R. März. The index of linear differential-algebraic equations with properly stated leading term. *Results in Mathematics*, 42:308–338, 2002.

[Mat76] T. Matsumoto. On the dynamics of electrical networks. *J. Differ. Equations*, 21:179–196, 1976.

[Mat87] W. Mathis. *Theorie nichtlinearer Netzwerke*. Springer Verlag Berlin, Heidelberg, New York, 1987.

[MP92] K. Mayaram and D. O. Pederson. Coupling algorithms for mixed-level circuit and device simulation. *IEEE Trans. CAD*, 11(8):1003–1012, 1992.

[MRS90] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser. *Semiconductor equations*. Springer Verlag Wien, 1990.

[MS93] S.E. Mattsson and G. Söderlind. Index reduction in differential-algebraic equations using dummy derivatives. *SIAM J. Sci. Comput.*, 14(3):677–692, 1993.

[MT97] R. März and C. Tischendorf. Recent results in solving index-2 differential-algebraic equations in circuit simulation. *SIAM J. Sci. Comput.*, 18(1):139–159, 1997.

[NC84] T. Nishi and L. O. Chua. Topological criteria for nonlinear resistive circuits containing controlled sources to have a unique solution. *IEEE Trans. Circuits Systems*, 31(8):722–741, 1984.

[OW69] T. Ohtsuki and H. Watanabe. State-variable analysis of RLC networks containing nonlinear coupling elements. *IEEE Trans. Circuit Theory*, 16:26–38, 1969.

[PG78] G. Da Prato and P. Grisvard. On an abstract singular Cauchy problem. *Comm. Partial. Diff. Eqs.*, 3:1077–1082, 1978.

[QRV01] A. Quarteroni, S. Ragni, and A. Veneziani. Coupling between lumped and distributed models for blood flow problems. *Comput. Vis. Sci.*, 4(2):111–124, 2001.

[Rei90] S. Reich. *Beitrag zur Theorie der Algebrodifferentialgleichungen*. PhD thesis, Techn. Univ. Dresden, 1990.

[Rei98] G. Reißig. The index of the standard circuit equations of passive RLCTG-networks does not exceed 2. *Proc. ISCAS'98*, 3:419–422, 1998.

[Rot00] F.M. Rotella. *Mixed circuit and device simulation for analysis, design, and optimization of opto-electronic, radio frequency, and high speed semiconductor devices*. PhD thesis, Stanford University, 2000.

[RR91] P.C. Rabier and W.C. Rheinboldt. A general existence and uniqueness theory for implicit differential-algebraic equations. *Differential and Integral Equations*, 4(3):563–582, 1991.

[RR96] P.J. Rabier and W.C. Rheinboldt. Classical and generalized solutions of time-dependent linear differential algebraic equations. *Linear Algebra and its Applications*, 245:259–293, 1996.

[RR99] P.J. Rabier and W.C. Rheinboldt. *Nonholonomic motion of rigid mechanical systems from a DAE viewpoint*. SIAM, Society for Industrial and Applied Mathematics, 1999.

[RR02] P.J. Rabier and W.C. Rheinboldt. Techniques of scientific computing (part 4). In Ciarlet, P. G. et al., editor, *Handbook of numerical analysis*, volume 8, pages 183–540. Amsterdam: North Holland/ Elsevier, 2002.

[Sch90] D. Schroeder. *Modelling of Interface Carrier Transport for Device Simulation*. Computational Microelectronics. Springer-Verlag, Wien New York, 1990.

[Sch96] S. Scharfenberg. *Mixed-Level Circuit-Device Simulation auf einem heterogenen Workstation-Cluster*. Shaker Verlag, 1996. PhD thesis.

[Sch02] S. Schulz. Ein PDAE-Netzwerkmodell als abstraktes differential-algebraisches System. Master's thesis, Humboldt University of Berlin, Inst. of Math., 2002.

[Sel84] S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, Wien New York, 1984.

[Sim98] B. Simeon. DAE's and PDE's in elastic multibody systems. *Numer. Algorithms*, 1998.

[Sim00] B. Simeon. *Numerische Simulation gekoppelter Systeme von partiellen und differential-algebraischen Gleichungen in der Mehrkörperdynamik*. Number 325 in Fortschritt-Berichte VDI, Reihe 20. VDI-Verlag, Düsseldorf, 2000. Habilitation.

[Sza82] A. Szatkowski. On the dynamic spaces and on the equations of motion of non-linear RLC networks. *Int. J. Circ. Theor. Appl.*, 10:99–122, 1982.

[Sze81] S. M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, New York, 1981.

[Tis96] C. Tischendorf. *Solution of index-2 differential algebraic equations and its application in circuit simulation*. PhD thesis, Humboldt-Univ. Berlin, 1996.

[Tis97] C. Tischendorf. Topological index-calculation of DAEs in circuit simulation. In *ZAMM, GAMM97*, volume 78, pages S1103–S1104, Regensburg, Germany, 1997.

[Tis99] C. Tischendorf. Topological index calculation of DAEs in circuit simulation. *Surv. Math. Ind.*, 8:187–199, 1999.

[Tis01] C. Tischendorf. Model design criteria for integrated circuits to have a unique solution and good numerical properties. In U. van Rienen, M. Günther, and D. Hecht, editors, *Scientific Computing in Electrical Engineering*, number 18 in Lecture Notes in Computational Science and Engineering. Springer Verlag Berlin Heidelberg, 2001.

[VS94] J. Vlach and K. Singhal. *Computer methods for circuit analysis and design*. Van Nostrand Reinhold, New York, 1994.

[Wac91] G. Wachutka. Unified framework for thermal, electrical, magnetic, and optical semiconductor device modeling. *COMPEL*, 5:311–321, 1991.

[Wac95] G. Wachutka. Consistent treatment of carrier emission and capture kinetics in electrothermal and energy transport models. *Microelectronics Journal*, 26:307–315, 1995.

[Zei86] E. Zeidler. *Nonlinear Functional Analysis and its Applications I. Fixed-Point Theorems*. Springer-Verlag, New York Inc., 1986.

[Zei90a] E. Zeidler. *Nonlinear Functional Analysis and its Applications II/A. Linear Monotone Operators*. Springer-Verlag, New York Inc., 1990.

[Zei90b] E. Zeidler. *Nonlinear Functional Analysis and its Applications II/B. Nonlinear Monotone Operators*. Springer-Verlag, New York Inc., 1990.

Index