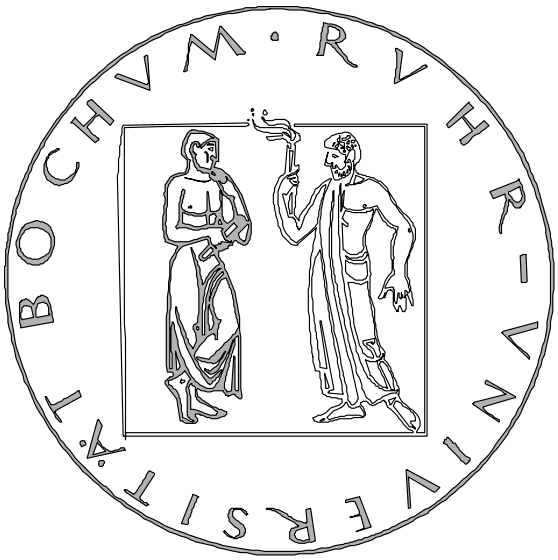
Michael Vollmer

Automatische Code-Erzeugung zur numerischen Integration partieller Differentialgleichungen für sicherheitskritische Anwendungen

Cuvillier Verlag Göttingen

Ruhr–Universität Bochum

Fakultät für Elektrotechnik und Informationstechnik



Automatische Code-Erzeugung zur numerischen Integration partieller Differentialgleichungen für sicherheitskritische Anwendungen

Dissertation

zur
Erlangung des Grades eines
Doktor-Ingenieurs

vorgelegt von
Michael Vollmer

Hamm

Bochum, den 20.04.2004

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2004
Zugl.: Bochum, Univ., Diss., 2004
ISBN 3-86537-221-X

Dissertation eingereicht:	20.04.2004
Referent:	Prof. Dr.-Ing. habil. H. D. Fischer
Korreferent:	Prof. Dr.-Ing. R. Martin
Tag der mündlichen Prüfung:	16.07.2004

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2004
Gedruckt auf säurefreiem Papier

ISBN 3-86537-221-X

Vorwort

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Nachrichtentechnik der Ruhr-Universität Bochum.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. habil. H. D. Fischer für die Unterstützung und Betreuung meiner Arbeit. Auch danke ich ihm für die interessante Aufgabenstellung, seine wertvollen Anregungen und die sorgfältige und kritische Durchsicht dieses Manuskripts. Vor allem danke ich ihm jedoch für den mir zur Verfügung gestellten Freiraum und die angenehme Arbeitsatmosphäre.

Für die Übernahme des Korreferats und der damit verbundenen Mühe bedanke ich mich bei Herrn Prof. Dr.-Ing. R. Martin.

Weiterhin bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Lehrstuhls für Nachrichtentechnik und der Arbeitsgruppe Digitale Signalverarbeitung sowie den Studien- und Diplomarbeitern und studentischen Hilfskräften, die alle zum Gelingen dieser Arbeit in freundlicher Atmosphäre beigetragen haben. Herrn Prof. Dr. sc. techn. Dr. h.c. mult. A. Fettweis danke ich für das Halten von Vorlesungen zur numerischen Integration partieller Differentialgleichungen mit dem Wellendigital-Konzept über den Zeitraum von vier Jahren. Besonders bedanken möchte ich mich bei meiner Kollegin Katrin Luhmann. Ihre wertvollen Ratschläge bei kleineren und größeren Problemen ersparten mir oft viel Zeit.

Zu guter Letzt möchte ich mich herzlich bei meiner Familie für die Unterstützung während der Erstellung meiner Arbeit bedanken.

Bochum, August 2004 Michael Vollmer

Inhaltsverzeichnis

1 Einführung **1**

2 Mehrdimensionale Wellendigitalfilter **3**
2.1 Abhängige und unabhängige Variablen . . . 3
2.2 Koordinatentransformation . . . 4
2.3 Abtastung . . . 10
2.4 Berechnungsgebiete . . . 13
2.5 Anordnung der Gitterpunkte und Abmessungen des Berechnungsgebietes . . . 15
2.6 Mehrdimensionale Kirchhoff'sche Netze . . . 15
2.7 Eigenschaften mehrdimensionaler Wellendigitalfilter . . . 21
2.8 Wellendigitalfilter Bauelemente . . . 21
2.8.1 Energieneutrale Bauelemente . . . 21
2.8.2 Verallgemeinerte Verbindungsnetze . . . 24
2.8.3 Quellen . . . 34
2.8.4 Dissipative Bauelemente . . . 37
2.8.5 Dynamische Bauelemente . . . 37
2.9 Zweckmäßige Beschreibung des Wellendigitalfilters . . . 40
2.10 Randbehandlung . . . 44
2.11 Von einem mehrdimensionalen Wellendigitalfilter zu einem eindimensionalen . . . 47

3 Das Engineeringsystem SPACE **53**
3.1 Der Aufbau des vom Engineeringsystem SPACE erzeugten C-Codes . . . 53
3.2 Die Einbindung der Wellendigitalfilter in das Engineeringsystem SPACE . . . 54
3.3 Zur Wahl der Abtastperioden . . . 57
3.4 Realisierungsaspekte . . . 57
3.5 Verknüpfung der Gleichungen des MDWDFs mit dem Funktionsbaustein . . . 58

4 Syntheseverfahren **65**
4.1 Der behandelte Typ von PDGLn . . . 65
4.1.1 Einschränkungen an die PDGLn . . . 65
4.1.2 Energiebetrachtungen . . . 66
4.1.3 Systemklassifizierung . . . 70
4.1.4 Eingeschwungener Zustand . . . 70
4.2 Synthese der Referenzschaltung . . . 71
4.2.1 Vorbetrachtungen . . . 71
4.2.2 Synthese des reaktiven Teils . . . 73
4.2.3 Synthese des konstanten Teils . . . 78
4.2.4 Gesamtreferenzschaltung . . . 79
4.3 Umsetzung der elektrischen Bauelemente in WD-Elemente . . . 80

4.4 Umsetzung der Gesamtschaltung in ein MDWDF . . . 82
4.5 Nachweis der Berechenbarkeit . . . 86
4.6 Alternative Methoden . . . 88
4.7 Randbehandlung . . . 88
4.8 Zusammenfassung . . . 92

5 Verifikation von Software mittels formaler Methoden 93
5.1 Definitionen, Notationen und Vereinbarungen . . . 94
5.2 Das Konzept der formalen Methoden . . . 95
5.3 Eine axiomatische Definition von Elementaranweisungen der Programmiersprache C . . . 97
5.3.1 Einzelanweisungen . . . 98
5.3.2 Blockanweisungen . . . 101
5.4 Behandlung von Ausdrücken . . . 101
5.4.1 Einzelanweisungen . . . 102
5.4.2 Blockanweisungen . . . 103
5.5 Einige Sätze . . . 103
5.5.1 Skalarprodukt mit Variablen . . . 103
5.5.2 Skalarprodukt mit Feldvariablen . . . 103

6 Der Algorithmus 105
6.1 Gesamtanweisung und Nachbedingung . . . 105
6.2 Berechnung der Verzögererwerte im Randfall . . . 118
6.3 Normalberechnung der Verzögererwerte . . . 125
6.4 Berechnung der Eingangswellen der Verzögerer . . . 134
6.5 Berechnung der Ausgangssignale . . . 140
6.6 Berechnung der Wellengrößen der nichtdynamischen Elemente . . . 144
6.7 Variablendeklarationen . . . 148
6.8 Zusammenfassung . . . 151

7 Zusammenfassung 153

A Die Operatoren min **und** max **155**

B Unitär beschränkte Matrizen 157

C Beschränktheit der Wellengrößen 159

D Auswirkungen endlicher Rechengenauigkeit 165
D.1 Zahlendarstellungen und Rundungsfehler . . . 165
D.2 Idealer und realer Zweitor-Parallel-Adaptor . . . 170
D.3 Direkte Umsetzung einer Zweitor-Adaptor-Streumatrix . . . 177

Systematik, Formelzeichen und Abkürzungen

Systematik

In der vorliegenden Arbeit werden Vektoren mit fetten Kleinbuchstaben gekennzeichnet, Großbuchstaben in Fettdruck stehen für Matrizen, und skalare Größen werden allgemein durch Buchstaben in Normalschrift dargestellt. Variablen werden i.d.R. schräg gestellt (Schriftart Italic) und Konstanten werden i.d.R. gerade gesetzt (Schriftart Roman).

Allgemeine Festlegungen

$\boldsymbol{M}_m^n$	Matrix mit m Zeilen und n Spalten (m und n meist unterdrückt)
m_{ij}	Element in der Zeile i und der Spalte j der Matrix $\boldsymbol{M}$
$\boldsymbol{m}_{i,\bullet}$	Zeile i der Matrix $\boldsymbol{M}$; Zeilenvektor
$\boldsymbol{m}_{\bullet,j}$	Spalte j der Matrix $\boldsymbol{M}$; Spaltenvektor
$\boldsymbol{M}^{\mathrm{T}}$	transponierte Matrix, das Element in der Zeile i und der Spalte j ist m_{ji}
$\boldsymbol{M}^{\mathrm{H}}$	transjugierte Matrix, das Element in der Zeile i und der Spalte j ist m_{ji}^*
$\boldsymbol{M}^{-\mathrm{T}}$	transponiert inverse Matrix, $\boldsymbol{M}^{-\mathrm{T}} = [\boldsymbol{M}^{\mathrm{T}}]^{-1} = [\boldsymbol{M}^{-1}]^{\mathrm{T}}$
$\boldsymbol{S}_*(\boldsymbol{p}) = [\boldsymbol{S}(-\boldsymbol{p}^*)]^{\mathrm{H}}$	Parakonjugierte der Matrix $\boldsymbol{S}$
$\boldsymbol{D} = \mathbf{diag}(d_1, \ldots, d_m)$	Diagonalmatrix
${}_{-i}^{-j}\boldsymbol{M}$	Matrix $\boldsymbol{M}$, in der die Zeile i und die Spalte j gestrichen wurde
$\mathbf{0}$	Nullvektor
$\mathbf{e}_\nu^{\mathrm{T}}$	Standardeinheits-Zeilenvektor $[\overset{1}{0}, \overset{2}{0}, \ldots, 0, \overset{\nu}{1}, 0, \ldots, \overset{n}{0}]$
$\mathbf{e}$	Vektor, deren sämtliche Koordinaten 1 sind
$\mathbf{1}_m$	Einheitsmatrix der Dimension m
$\mathbf{0}_m^n$	Nullmatrix mit m Zeilen und n Spalten (m und n meist unterdrückt)
$\lceil a \rceil$	größte ganze Zahl, die kleiner gleich a ist
$\|\boldsymbol{x}\|$	euklidische Vektornorm von $\boldsymbol{x}$
$\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker Produkt mit $c_{ij} = a_{ij}\boldsymbol{B}$
P	Bezeichner eines prädikatenlogischen Ausdrucks

S	Bezeichner eines Programmsegments
$\equiv$	Identität

$$\sum_{\mu=N}^{M} \boldsymbol{A}_\mu = \begin{cases} \boldsymbol{A}_N + \boldsymbol{A}_{N+1} + \cdots + \boldsymbol{A}_{M-1} + \boldsymbol{A}_M & \text{für } M \geq N \\ \boldsymbol{0} & \text{für } M < N \end{cases}$$

$$\prod_{\mu=N}^{M} \boldsymbol{A}_\mu = \begin{cases} \boldsymbol{A}_N \boldsymbol{A}_{N+1} \cdots \boldsymbol{A}_{M-1} \boldsymbol{A}_M & \text{für } M \geq N \\ \boldsymbol{1} & \text{für } M < N \end{cases}$$

$$\bigwedge_{\mu=N}^{M} \mathsf{P}_\mu \equiv \begin{cases} \mathsf{P}_N \wedge \mathsf{P}_{N+1} \wedge \cdots \wedge \mathsf{P}_{M-1} \wedge \mathsf{P}_M & \text{für } M \geq N \\ \text{wahre Aussage} & \text{für } M < N \end{cases}$$

$$\bigwedge_{\boldsymbol{l}=\boldsymbol{a}}^{\boldsymbol{b}} \mathsf{P} \equiv \bigwedge_{l_1=a_1}^{b_1} \bigwedge_{l_2=a_2}^{b_2} \cdots \bigwedge_{l_n=a_n}^{b_n} \mathsf{P}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \equiv \bigwedge_{\nu=1}^{n} x_\nu \geq 0$$

$\boldsymbol{A} \geq 0$	symbolische Darstellung für eine nicht negativ definite (n. n. d.) Matrix $\boldsymbol{A}$
$\boldsymbol{A} > 0$	symbolische Darstellung für eine positiv definite (p. d.) Matrix $\boldsymbol{A}$

Ausgewählte Formelzeichen

Skalare

Anzahlen des Wellendigitalfilters

n_e	Gesamtzahl der Tore der nichtdynamischen Bauelemente
$n_g = n_q + n_v + n_e$	Gesamttorzahl des Wellendigitalfilters
n_q	Anzahl der Quellen
n_v	Anzahl der dynamischen Bauelemente
n_v^κ	Anzahl der bzgl. t_κ dynamischen Bauelemente
N_e	Anzahl der nichtdynamischen Bauelemente

Anzahlen des Funktionsbausteins

n_{ea}	Anzahl der analogen Eingänge des Funktionsbausteins

n_{aa}	Anzahl der analogen Ausgänge des Funktionsbausteins
m	Anzahl der Zustandsgrößen

Vektoren
abhängige Variablen des Wellendigitalfilters

$\boldsymbol{a}_q$	Vektor der einfallenden Wellen der Quellen
$\boldsymbol{a}_v$	Vektor der einfallenden Wellen der dynamischen Elemente
$\boldsymbol{a}_e$	Vektor der einfallenden Wellen der nichtdynamischen Elemente
$\boldsymbol{a} = [\boldsymbol{a}_q^{\mathrm{T}}\ ,\ \boldsymbol{a}_v^{\mathrm{T}}\ ,\ \boldsymbol{a}_e^{\mathrm{T}}]^{\mathrm{T}}$	Vektor der einfallenden Wellen aller Elemente
$\boldsymbol{b}_q$	Vektor der ausfallenden Wellen der Quellen
$\boldsymbol{b}_v$	Vektor der ausfallenden Wellen der dynamischen Elemente
$\boldsymbol{b}_e$	Vektor der ausfallenden Wellen der nichtdynamischen Elemente
$\boldsymbol{b} = [\boldsymbol{b}_q^{\mathrm{T}}\ ,\ \boldsymbol{b}_v^{\mathrm{T}}\ ,\ \boldsymbol{b}_e^{\mathrm{T}}]^{\mathrm{T}}$	Vektor der ausfallenden Wellen aller Elemente
$b_{v\mu}^{\kappa}$	Koordinate μ des Vektors $\boldsymbol{b}_v^{\kappa}$

Vektoren des Funktionsbausteins

$\boldsymbol{a}_{\mathrm{FB}}$	Vektor der Ausgangssignale des Funktionsbausteins
$\boldsymbol{e}_{\mathrm{FB}}$	Vektor der Eingangssignale des Funktionsbausteins
$\boldsymbol{m}_{\mathrm{FB}}$	Vektor der Zustandsgrößen des Funktionsbausteins

Matrizen
Matrizen des Wellendigitalfilters

$\boldsymbol{P}$	Matrix, die die Verknüpfungen der einzelnen Tore des Wellendigitalfilters beschreibt

$$\boldsymbol{P} = \begin{array}{c} \\ q \\ v \\ e \end{array} \begin{array}{c} \begin{array}{ccc} q & v & e \end{array} \\ \begin{bmatrix} \boldsymbol{P}_{qq} & \boldsymbol{P}_{qv} & \boldsymbol{P}_{qe} \\ \boldsymbol{P}_{vq} & \boldsymbol{P}_{vv} & \boldsymbol{P}_{ve} \\ \boldsymbol{P}_{eq} & \boldsymbol{P}_{ev} & \boldsymbol{P}_{ee} \end{bmatrix} \end{array}$$

$\boldsymbol{P}_b$	Permutationsmatrix zur Festlegung der Berechnungsreihenfolge
$\boldsymbol{S}$	Streumatrix der nichtdynamischen Bauelemente

unabhängige Variablen Referenzschaltung/Wellendigitalfilter

$\mathrm{D}_{\kappa'} = \frac{\partial}{\partial t_{\kappa'}}$	Differentialoperatoren nach den neuen Koordinaten
$\mathbf{D}_{\boldsymbol{t}} = [\mathrm{D}_1, \ldots, \mathrm{D}_{k'}]^{\mathrm{T}}$	Gradient in den Koordinaten $\boldsymbol{t}$
$\mathbf{D}_{\boldsymbol{x}} = [\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{k-1}}, \frac{1}{v_k}\frac{\partial}{\partial t}]^{\mathrm{T}}$	Gradient in den Koordinaten $\boldsymbol{x}$
$\mathbf{D} = [\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{k-1}}]^{\mathrm{T}}$	Gradient in den Ortskoordinaten
$\boldsymbol{H}$, $\boldsymbol{x} = v_0 \boldsymbol{H} \boldsymbol{t}$	Transformationsmatrix
k	Dimension des ursprünglichen Koordinatensystems
k'	Dimension des neuen Koordinatensystems
$\boldsymbol{\mu} = [\mu_1, \ldots, \mu_k]^{\mathrm{T}}$	Diskrete Variablen des ursprünglichen Koordinatensystems
$\Delta\boldsymbol{\mu} = [\Delta\mu_1, \ldots, \Delta\mu_k]^{\mathrm{T}}$	Verschiebe-Vektor in den diskreten Variablen des ursprünglichen Koordinatensystems
$\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{k'}]^{\mathrm{T}}$	Diskrete Variablen des neuen Koordinatensystems
$\boldsymbol{t} = [t_1, \ldots, t_{k'}]^{\mathrm{T}}$	Vektor mit den unabhängigen Variablen des neuen Koordinatensystems
$\Delta\boldsymbol{t} = [\Delta t_1, \ldots, \Delta t_{k'}]^{\mathrm{T}}$	Verschiebe-Vektor des neuen Koordinatensystems
$\boldsymbol{T}_A$	Abtastmatrix des neuen Koordinatensystems
$\boldsymbol{x} = [x_1, \ldots, x_{k-1}, v_k t]^{\mathrm{T}}$	Vektor mit den unabhängigen Variablen des ursprünglichen Koordinatensystems
${}_{-k}\boldsymbol{x} = [x_1, \ldots, x_{k-1}]^{\mathrm{T}}$	Vektor mit den unabhängigen Ortskoordinaten des ursprünglichen Koordinatensystems (Echtraumkoordinaten)
$\boldsymbol{X}_A$	Abtastmatrix des ursprünglichen Koordinatensystems
$\Delta\boldsymbol{x} = [\Delta x_1, \ldots, v_k \Delta t]^{\mathrm{T}}$	Verschiebe-Vektor des ursprünglichen Koordinatensystems

Abmessungen

$\mathcal{G}$	Berechnungsgebiet (nur Ortskoordinaten)
$\mathcal{G}_{\boldsymbol{x}}$	Gebiet im ursprünglichen Koordinatensystem
$\mathcal{G}_{\boldsymbol{t}}$	Gebiet im neuen Koordinatensystem

$\mathcal{G}_0$	Gebiet außerhalb des Berechnungsgebietes
$\partial\mathcal{G}$	Rand des Berechnungsgebietes
$\mathcal{G}_n$	Teilberechnungsgebiet $n > 0$
l_κ	Länge des Berechnungsgebietes in Richtung x_κ
$l_{\kappa n}$	Länge des Teilberechnungsgebietes n in Richtung x_κ
N	Anzahl der Teilberechnungsgebiete
N_κ	Anzahl der Teilberechnungsgebiete in Richtung der unabhängigen Variablen x_κ
P	Anzahl der Abtastpunkte einer Abtastschicht
P_{x_κ}	Anzahl der Abtastpunkte in Richtung x_κ
$P_{x_{\kappa n}}$	Anzahl der Abtastpunkte in Richtung x_κ des Teilberechnungsgebietes n
ΔT_κ	Abstand zweier Abtastpunkte in Richtung t_κ
$x_{\kappa\,\min}$	untere Grenze des Berechnungsgebietes in Richtung x_κ
$x_{\kappa\,\max}$	obere Grenze des Berechnungsgebietes in Richtung x_κ
ΔX_κ	Abstand zweier Abtastpunkte in Richtung x_κ

Kapitel 1

Einführung

Ein Großteil naturwissenschaftlicher und technischer Prozesse wird mithilfe von partiellen Differentialgleichungen (PDGLn) beschrieben. Da für PDGLn i.Allg. keine Verfahren zur analytischen Lösung bekannt sind, setzt man numerische Verfahren zur Berechnung von Näherungslösungen mittels eines Digitalrechners ein. Sowohl die weiter ansteigende Rechenleistung der Digitalrechner, als auch neue, effizientere Algorithmen erschließen weitere Anwendungsgebiete dieses wachsenden Teilbereichs der Mathematik. Die numerischen Verfahren sollten dabei die wesentlichen Eigenschaften des ursprünglichen Systems auf das digitale System übertragen, insbesondere die Stabilität, die Lokalität und die Passivität. Die Übertragung dieser Eigenschaften leisten die von Fettweis entwickelten Wellendigitalfilter (WDF), die ursprünglich zur digitalen Nachbildung analoger Filter dienten [Fett70], [Fett86]. Es hat sich gezeigt, dass diese Filter hervorragende Stabilitätseigenschaften auch unter realen Bedingungen (endliche Wortlänge) besitzen. Diese günstigen Eigenschaften übertragen sich auch auf die mehrdimensionalen Wellendigitalfilter. Die mehrdimensionalen Wellendigitalfilter sind digitale Nachbildungen analoger(auch nichtlinearer) mehrdimensionaler Kirchhoffscher Schaltungen und wurden erstmals in [Fisc84] zur numerischen Integration von Differentialgleichungen vorgeschlagen. Bekanntermaßen kann eine mehrdimensionale Kirchhoffsche Schaltung äquivalent durch partielle Differentialgleichungen beschrieben werden. Somit lässt sich die Wellendigitalmethode, welche in dieser Arbeit ausschließlich verwendet wird, zur numerischen Integration von partiellen Differentialgleichungen nutzen, [FN90a], [FN90b], [FN91a], [FN91b], [Nits93], [Heme95], [Feld95], [Frie95], [Krau97], [Pott98]. Mittlerweile existiert eine Vielzahl weiterer Arbeiten, die über die erfolgreiche Anwendung des Verfahrens berichten. Eine zusammenfassende Darstellung in englischer Sprache findet sich in [Bilb01] und [Bilb04]. Die praktische Anwendbarkeit der Wellendigitalmethode erfordert aber die effiziente Erzeugung eines Codes zur Simulation des Wellendigitalfilters. Der Ausgangspunkt der Erzeugung eines Codes sollte das Ausgangsproblem selber sein, also die zu lösende PDGL. In dieser Arbeit wird ein Verfahren zur automatischen Codeerzeugung vorgestellt, welches auf PDGLn anwendbar ist, die lineare, zeitinvariante, symmetrisch hyperbolische Systeme beschreiben.

Eine automatische Codeerzeugung hat nicht nur den Vorteil geringeren Herstellungsaufwands, sondern liefert auch -geeignete Implementierung vorausgesetzt- eine höhere Qualität der erzeugten Software, da der Anteil der manuellen Software-Herstellung reduziert wird. Es eröffnet sich zudem die Möglichkeit des Einsatzes der Algorithmen in sicherheitskritischen Bereichen. Wie oben bereits angedeutet, sind für eine derartige Anwendung auch strengere Qualitätsansprüche an die Implementierung zu stellen. Um diesen nachzukommen, verwenden wir die so genannten formalen Methoden der Softwaretechnik zur Verifikation der entwickelten Codes gegenüber der Spezifikation. Die Theorie der Programmverifikation wurde durch McCarthy angeregt, [McCa62], [McCa63]. Seine Intention war es, die gewünschten Eigenschaften von Programmen mittels mathematischer Methoden nachzuweisen, anstatt die Programme durch Testläufe auf Fehlerlosigkeit zu prüfen. Floyd nahm die Idee auf und schlug ein Verfahren vor mit dem ein gegebenes Software-System analysiert werden konnte, [Floy67]. Dieses Konzept wurden in

[Dijk68] erweitert und zur Synthese beweisbarer korrekter Programme genutzt. In den weiteren Jahren folgten Arbeiten, die im Wesentlichen auf den zuvor genannten Artikeln aufbauen. Hervorzuheben hieraus ist [Hoar69], in der eine Programmanweisung bzw. eine zusammengesetzte Anweisung als Transformation von Prädikaten aufgefasst wird. Weiterhin sind dort die wichtigsten Semantikregeln zu finden. Auf dieser Basis wird auch der in dieser Arbeit angegebene Beweis geführt.

Die wesentlichen Ziele dieser Arbeit sind zum einen die Entwicklung eines Verfahrens zur Synthese einer Referenzschaltung und anschließender Umsetzung in ein mehrdimensionales Wellendigitalfilter, welches die formale Spezifikation der Software festgelegt. Zum anderen wird ein Code angegeben, der das mehrdimensionale Wellendigitalfilter simuliert und mittels eines formalen Korrektheitsbeweises gegenüber der Spezifikation verifiziert. Der Code soll dabei so beschaffen sein, dass er sich in das -für Steuerungs- und Regelungsfunktionen mit sicherheitsrelevanter Bedeutung in Kernkraftwerken entwickelte- digitale Leitsystem TELEPERM XS einbinden lässt.

Intention des Kapitels 2 ist es, die notwendigen Grundlagen der Theorie mehrdimensionaler Wellendigitalfilter zu rekapitulieren und zwar in einer auf unsere Aufgabenstellung angepassten Form.

Anschließend erfolgt im Kapitel 3 eine Einarbeitung in das Programmpaket SPACE und zwar einerseits aus Sicht des Benutzers (i.d.R. der Leittechniker) und andererseits aus Sicht des Entwicklers. Nach Analyse und Darstellung der erarbeiteten Erkenntnisse wird die Möglichkeit der Einbindung der Wellendigitalfilter in das Programmpaket SPACE untersucht. Dabei auftretende Probleme und deren Lösung werden aufgezeigt.

Kapitel 4 bildet den ersten Hauptteil dieser Arbeit. Dort werden wir zunächst die Klasse der in dieser Arbeit behandelten Systeme einschränken und daraus Eigenschaften der PDGLn ableiten. Im Anschluss daran werden wir das neu entwickelte Syntheseverfahren zur Gewinnung einer mehrdimensional passiven Referenzschaltung vorstellen. Ferner werden wir zu der systematisch gewonnenen Referenzschaltung ein mehrdimensionales Wellendigitalfilter angegeben und deren Berechenbarkeit aufzeigen.

In Kapitel 5 werden die aus der Literatur bekannten Grundlagen zur Anwendung formaler Methoden in der Softwaretechnik behandelt. Dabei werden wir uns auf die Inhalte beschränken, die für den weiteren Verlauf der Arbeit relevant sind. Zudem werden in diesem Kapitel die für den Algorithmus notwendigen Programmanweisungen axiomatisch definiert.

Kapitel 6 beinhaltet den zweiten Hauptteil dieser Arbeit. In diesem Kapitel werden wir den Übergang zur formalen Spezifikation des Algorithmus durchführen. Zudem wird die Implementierung des Algorithmus in der Programmiersprache C unter ausschließlicher Verwendung der axiomatisch definierten Programmanweisungen durchgeführt. Weiterhin führen wir den formalen Korrektheitsbeweis in diesem Kapitel. In formalen Korrektheitsbeweisen für Algorithmen der Signalverarbeitung kommt der Verhinderung eines Überlaufs des Darstellungsbereiches eine besondere Bedeutung zu, die ebenfalls in diesem Kapitel Berücksichtigung findet.

Die Ergebnisse der Kapitel 4 und 6 werden mit eigenen Zusammenfassungen gewürdigt. Die Zusammenfassung der gesamten Arbeit befindet sich im Kapitel 7.

Kapitel 2

Mehrdimensionale Wellendigitalfilter

In diesem Kapitel werden wir die für das Verständnis der Arbeit notwendigen Grundlagen der Theorie mehrdimensionaler Wellendigitalfilter zusammenfassen. Die Zusammenfassung der Grundlagen beginnt mit der Einführung von unabhängigen und abhängigen Variablen, wobei die unabhängigen Variablen danach einer Koordinatentransformation unterworfen und diskretisiert werden. Anschließend werden die Form des Berechnungsgebietes, also der Bereich der unabhängigen Variablen für die die Lösung der PDGL erfolgen soll, festgelegt. In diesem Zusammenhang wird zudem die Lage der Abtastpunkte diskutiert.

Nachdem in den ersten Unterkapiteln die unabhängigen Variablen im Vordergrund standen, widmet sich der weitere Teil den abhängigen Variablen. Diese abhängigen Variablen einer PDGL werden als Spannungen und Ströme von Toren eines Kirchhoff'schen Netzes interpretiert. Ähnlich den unabhängigen Variablen erfolgt bei den abhängigen Variablen eine Koordinatentransformation. Diese Transformation vollzieht sich in der Form, dass anstelle von Spannung und Strom Wellengrößen genutzt werden. Von zentraler Bedeutung für Stabilitätsfragen Kirchhoff'scher Netze sind die energetischen Eigenschaften ihrer Bauelemente. Aufbauend auf der torweisen Betrachtung der Bauelemente, werden aus Energiebetrachtungen heraus Eigenschaften wie MD-Passivität und MD-Energieneutralität erläutert. Gemäß diesen Eigenschaften werden dann typische Bauelemente der Theorie elektrischer Netze, die in dieser Arbeit Verwendung finden, eingeführt, qualifiziert und in MDWDF-Bauelemente überführt. Auf den Nachweis der Übertragung der Eigenschaften der elektrischen Bauelemente auf die WD-Elemente verzichten wir und verweisen auf die existierende Literatur, [Meer79], [MF92], [Fett92].

Weiterhin wird von den bekannten Verfahren zur Randwertbehandlung eines, welches sich besonders gut in die Zielsetzung dieser Arbeit einpasst und auch später Verwendung findet, ausführlich erläutert und kompakt beschrieben.

Ergänzend wird für die torweise Verschaltung der einzelnen WD-Elemente eine übersichtliche Darstellung, in Form eines Differenzen-Gleichungssystems, angegeben und mit dessen Hilfe die Berechnungsreihenfolge der einzelnen Wellengrößen systematisch bestimmt.

Zum Abschluss wird eine Transformation der unabhängigen Variablen dargelegt, die es ermöglicht, die endliche Anzahl örtlicher Abtastpunkte auf eine natürliche Zahl eineindeutig abzubilden. Dieses Konzept wird später bei der Einbindung der MDWDF in das Programmpaket SPACE angewendet.

Um eine einfachere Recherche zu ermöglichen, findet man die Quellenangaben in den einzelnen Unterkapiteln.

2.1 Abhängige und unabhängige Variablen

Die Variablen eines Differentialgleichungssystems unterscheiden wir nach abhängigen und unabhängigen Größen. Die unabhängigen Größen sind die Ortsvariablen und die Zeit. Die abhängigen Größen werden

wir als die Feldgrößen bezeichnen. Beispiele hierfür sind Druck oder Temperatur. Die Feldgrößen sind proportional zu einer Zweigspannung oder einem Zweigstrom der Referenzschaltung. Wir fassen die unabhängigen Variablen in dem Vektor

$$\boldsymbol{x} = [x_1\ ,\ \ldots\ ,\ x_{k-1}\ ,\ v_k t]^{\mathrm{T}} \quad ,\quad v_k > 0\ , \tag{2.1}$$

zusammen. Hierin ist t die physikalische Zeit und die restlichen Variablen sind beispielsweise physikalische Ortskoordinaten. Die mit der Einheit einer Geschwindigkeit behaftete Größe v_k kann eine Funktion der Zeit sein, wird aber in dieser Arbeit als konstant vorausgesetzt. Somit sind die Einheiten aller Koordinaten des Vektors $\boldsymbol{x}$ gleich. Mit $\mathbf{D}_{\boldsymbol{x}}$ bezeichnen wir den Vektor, der die partiellen Ableitungen bzgl. x_κ enthält, d. h.

$$\mathbf{D}_{\boldsymbol{x}} = \left[\frac{\partial}{\partial x_1}\ ,\ \ldots\ ,\ \frac{\partial}{\partial x_k}\right]^{\mathrm{T}} = \left[\frac{\partial}{\partial x_1}\ ,\ \ldots\ ,\ \frac{\partial}{\partial x_{k-1}}\ ,\ \frac{1}{v_k}\frac{\partial}{\partial t}\right]^{\mathrm{T}}\ . \tag{2.2}$$

Weiterhin definieren wir einen Vektor, der als Koordinaten die komplexen Frequenzen enthält

$$\boldsymbol{p}_{\boldsymbol{x}} = [p_{x_1}\ ,\ \ldots\ ,\ p_{x_k}]^{\mathrm{T}} \quad ,\quad p_{x_k} = p_t/v_k. \tag{2.3}$$

2.2 Koordinatentransformation

Die in der Theorie der mehrdimensionalen Wellendigitalfilter oft verwendete Koordinatentransformation werden wir zunächst durch Überlegungen zur Passivität[1] stützen. Die wesentlichen Quellen sind [FN91b] und [Nits93]. Im eindimensionalen Fall bezieht sich die Passivität nur auf die Zeit als unabhängige Variable. In der klassischen Literatur zu mehrdimensionalen elektrischen Schaltungen und in der mehrdimensionalen Signalverarbeitung werden alle unabhängigen Variablen als untereinander gleichberechtigt aufgefasst [Koga69], [Rao69], [Bose79], [Fett79], [Huan81], [Bose82], [FB87], [BF87], [Kumm88]. Die Passivität muss bezüglich aller unabhängigen Variablen gewährleistet sein. Diese Passivität wird auch als mehrdimensionale Passivität bezeichnet. Die durch Energiebetrachtungen abgeleitete Definition der Passivität physikalischer Systeme bezieht sich hingegen nur auf die Zeit. Die Ortskoordinaten finden hierbei keine Berücksichtigung. Insofern ist zu erwarten, dass eine mehrdimensionale elektrische Schaltung, die aus einer partiellen Differentialgleichung eines passiven Systems direkt hergeleitet wurde, i.Allg. nicht mehrdimensional passiv ist.

Nichtsdestotrotz ist es wünschenswert, eine mehrdimensional passive elektrische Schaltung aus der partiellen Differentialgleichung eines passiven Systems anzugeben. Dies hat den Vorteil, dass die Erkenntnisse der mehrdimensionalen elektrischen Schaltungen und der mehrdimensionalen Signalverarbeitung genutzt werden können. Um dieses Ziel zu erreichen, führen wir eine Transformation der unabhängigen Variablen durch.

Im engen Zusammenhang mit der Passivität steht die Kausalität. Im eindimensionalen Fall ist die Kausalität so definiert, dass der aktuelle Zustand nur von Zuständen zurückliegender Zeitpunkte und vom Eingangssignal des aktuellen Zeitpunktes und zurückliegender Zeitpunkte abhängt.

In der mehrdimensionalen Signalverarbeitung wird ebenfalls eine Ablaufrichtung festgelegt. Diese Ablaufrichtung teilt das Gebiet der unabhängigen Variablen in einen Abhängigkeitsbereich (Ursachenbereich), einen Wirkungsbereich und einen Bereich, der zu keinem der beiden gehört. Bei einer bestimmten Ablaufrichtung könnte beispielsweise der Zustand im Punkt $\boldsymbol{t}_0$ nur von Zuständen und Eingangssignalen des Abhängigkeitsbereiches $\boldsymbol{t} \leq \boldsymbol{t}_0$ abhängen. Der Zustand im Punkt $\boldsymbol{t}_0$ hingegen beeinflusst nur Zustände im Wirkungsbereich $\boldsymbol{t} \geq \boldsymbol{t}_0$.

[1] Die Passivitätsbegriffe werden im Kapitel 2.7 genau definiert.

Bei physikalischen Systemen bezieht sich die Kausalität hingegen nur auf die Zeit. Ein Zusammenhang zu dem gerade beschriebenen Sachverhalt in der mehrdimensionalen Signalverarbeitung ist daher nicht unmittelbar erkennbar. Insbesondere existiert kein Bereich, der nicht dem Wirkungs- oder dem Abhängigkeitsbereich zugeordnet werden kann. Allerdings zeigen sich Parallelen auf, wenn das System nur lokale Abhängigkeiten besitzt. In dem Fall bilden sich durch die endliche Ausbreitungsgeschwindigkeit des Vorgangs auch Bereiche aus, die weder dem Abhängigkeits- noch dem Wirkungsbereich zugeordnet werden können. Wir halten als Fazit dieser Überlegungen fest, dass die Darstellung eines physikalischen Systems mittels einer mehrdimensional passiven Schaltung eine endliche Ausbreitungsgeschwindigkeit voraussetzt. Im weiteren Verlauf dieses Kapitels werden wir sehen, welchen Einfluss die maximale Ausbreitungsgeschwindigkeit auf die Koordinatentransformation hat.

Da wir jeder der neuen Variablen die Bedeutung einer Zeit beimessen wollen, wählen wir als Bezeichner für die Variablen des neuen Koordinatensystems

$$\boldsymbol{t} = [t_1 \ , \ \ldots \ , \ t_{k'}]^{\mathrm{T}} \ . \tag{2.4}$$

Die Variablen $\boldsymbol{t}$ sind über

$$\boldsymbol{x} = v_0 \, \boldsymbol{H} \, \boldsymbol{t} \quad , \quad v_0 > 0 \tag{2.5}$$

mit den ursprünglichen unabhängigen Variablen verknüpft ($\dim\{\boldsymbol{H}\} = k \times k'$). Die Variable v_0 hat die Einheit einer Geschwindigkeit und wird in dieser Arbeit ebenfalls als konstant vorausgesetzt. Der Vektor $\boldsymbol{t}$ findet dann als eine mehrdimensionale Zeit eine willkommene Interpretation. Die Dimension des neuen Koordinatensystems soll nicht kleiner als die Dimension des alten Koordinatensystems sein, d.h. $k' \geq k$. Wir sprechen von einer verallgemeinerten Koordinatentransformation, wenn das neue Koordinatensystem eine höhere Dimension als das ursprüngliche Koordinatensystem hat. Weiterhin fordern wir, dass $\boldsymbol{H}$ zeilenregulär ist, was zur Folge hat, dass der Rang von $\boldsymbol{H}$ gleich k ist.

Nach den eingangs gemachten Bemerkungen bzgl. der Kausalität erscheint es sinnvoll, die Koordinatentransformation so durchzuführen, dass der Abhängigkeitsbereich und der Wirkungsbereich der mehrdimensionalen Zeit auf den Abhängigkeitsbereich und den Wirkungsbereich der physikalischen Zeit abgebildet werden. Zudem sollten Abhängigkeits- und Wirkungsbereich der Zeit auf den Abhängigkeits- und den Wirkungsbereich der mehrdimensionalen Zeit abgebildet werden. Aufgrund der i. Allg. rechteckförmigen Matrix $\boldsymbol{H}$ kann nicht von einer bijektiven Abbildung gesprochen werden. Die bislang an die Koordinatentransformation gestellten Forderungen können wie folgt formuliert werden :

- Nimmt keine der Variablen des neuen Koordinatensystems ab und mindestens eine der neuen Variablen zu, so muss dies eine Zunahme der Zeit t bewirken, d. h.

$$[\Delta t_\kappa \geq 0 \text{ für } \kappa = 1 \ldots k'] \wedge [\text{ mind. ein } \Delta t_\kappa > 0] \Longrightarrow \Delta t > 0. \tag{2.6}$$

- Nimmt die Zeit t zu und sind die restlichen Variablen des ursprünglichen Koordinatensystems konstant, so muss dies eine Zunahme jeder Variablen des neuen Koordinatensystems bewirken, d. h.

$$\Delta t > 0 \wedge \Delta x_\kappa = 0 \text{ für } \kappa = 1 \ldots k-1 \Longrightarrow \Delta t_\kappa > 0 \text{ für } \kappa = 1 \ldots k' \ . \tag{2.7}$$

Die Forderungen erscheinen notwendig für die gewünschte mehrdimensionale Passivität zu sein, falls das System passiv bzgl. t ist. Aus diesen Forderungen resultieren Bedingungen an die Elemente der Matrix $\boldsymbol{H}$. Wir werden später genauer darauf eingehen.

Normiert man den letzten Zeilenvektor von $\boldsymbol{H}$ auf eine euklidische Norm von 1 und wählt seine Koordinaten gleich, d. h. zu $\sqrt{1/k'}$, so lässt sich zeigen (siehe z.B. [Fett99]), dass die Ungleichung

$$v_k \geq \sqrt{k-1} v_{\max} \tag{2.8}$$

notwendig und hinreichend zur Erfüllung der Forderungen Gleichung (2.6) und Gleichung (2.7) ist. Die Geschwindigkeit $v_{\max}$ stellt die maximal mögliche Ausbreitungsgeschwindigkeit des physikalischen Vorgangs dar.

Die Differentialoperatoren im neuen Koordinatensystem sind in dem Vektor

$$\mathbf{D}_{\boldsymbol{t}} = \left[\frac{\partial}{\partial t_1}\,,\ \ldots\ ,\ \frac{\partial}{\partial t_{k'}}\right]^{\mathrm{T}} \tag{2.9}$$

zusammengefasst. Wir werden im Folgenden die Beziehung zwischen den Ableitungen zu bestimmen haben und gehen dazu von der Kettenregel

$$\boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{J}_{\boldsymbol{g}}^{\boldsymbol{f}} \cdot \boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{g}}, \tag{2.10}$$

aus, die der Differentiation der beliebigen differenzierbaren Funktion $\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{t}))$ nach $\boldsymbol{t}$ entspricht. Die drei Matrizen sind die Jacobi-Matrizen, z.B.

$$\boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{f}} = \left[\frac{\partial \boldsymbol{f}}{\partial t_1}\,,\ \ldots\ ,\ \frac{\partial \boldsymbol{f}}{\partial t_{k'}}\right] = [\mathbf{D}_{\boldsymbol{t}}\, \boldsymbol{f}^{\mathrm{T}}]^{\mathrm{T}}\,. \tag{2.11}$$

In unserem Fall gilt $\boldsymbol{g} = \boldsymbol{x} = v_0 \boldsymbol{H} \boldsymbol{t}$ und $\boldsymbol{f}$ ist eine skalare Funktion. Die Differentiation der skalaren Funktion f nach $\boldsymbol{t}$ lautet somit

$$\boldsymbol{J}_{\boldsymbol{t}}^{f} = \left[\frac{\partial f}{\partial t_1}\,,\ \ldots\ ,\ \frac{\partial f}{\partial t_{k'}}\right] = \underbrace{\left[\frac{\partial}{\partial t_1}\,,\ \ldots\ ,\ \frac{\partial}{\partial t_{k'}}\right]}_{=\mathbf{D}_{\boldsymbol{t}}^{\mathrm{T}}} f = \underbrace{\left[\frac{\partial}{\partial x_1}\,,\ \ldots\ ,\ \frac{\partial}{\partial x_k}\right]}_{=\mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}} f \cdot \underbrace{\left[\frac{\partial \boldsymbol{x}}{\partial t_1}\,,\ \ldots\ ,\ \frac{\partial \boldsymbol{x}}{\partial t_{k'}}\right]}_{=\boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{x}}}. \tag{2.12}$$

Die letzte Jacobi-Matrix berechnet man mit $\boldsymbol{x} = v_0\, \boldsymbol{H}\, \boldsymbol{t}$ und

$$\frac{\partial \boldsymbol{x}}{\partial t_\nu} = \frac{\partial}{\partial t_\nu} v_0\, \boldsymbol{H}\, \boldsymbol{t} = v_0\, \boldsymbol{H} \frac{\partial \boldsymbol{t}}{\partial t_\nu} = v_0\, \boldsymbol{H}\, \mathbf{e}_\nu \tag{2.13}$$

zu

$$\boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{x}} = v_0 \boldsymbol{H}. \tag{2.14}$$

Die Transposition von Gleichung (2.12) liefert letztendlich das gewünschte Ergebnis

$$\mathbf{D}_{\boldsymbol{t}}\, f = v_0 \boldsymbol{H}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{x}} f, \tag{2.15}$$

wobei wir im Folgenden nur die Operatoren ohne die Funktion f verwenden werden, d.h.

$$\mathbf{D}_{\boldsymbol{t}} = v_0 \boldsymbol{H}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{x}}\,. \tag{2.16}$$

Wir werden im Folgenden Bedingungen an die Matrix $\boldsymbol{H}$ aus der ersten Forderung an die Koordinatentransformation herleiten. Aus dem totalen Differential

$$dx_\kappa = \sum_{\mu=1}^{k'} \frac{\partial x_\kappa}{\partial t_\mu} dt_\mu \tag{2.17}$$

ergibt sich näherungsweise für die Zuwächse

$$\Delta x_\kappa = \sum_{\mu=1}^{k'} \frac{\partial x_\kappa}{\partial t_\mu} \Delta t_\mu. \tag{2.18}$$

Die Ableitung $\frac{\partial x_\kappa}{\partial t_\mu}$ ist das Element der Zeile κ und der Spalte μ der Jacobi-Matrix $\boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{x}}$ und lautet $\frac{\partial x_\kappa}{\partial t_\mu} = v_0 h_{\kappa\mu}$. Zusammenfassen von Gleichung (2.18) für alle κ in Vektorform liefert

$$\Delta \boldsymbol{x} = \boldsymbol{J}_{\boldsymbol{t}}^{\boldsymbol{x}}\, \Delta \boldsymbol{t} = v_0\, \boldsymbol{H}\, \Delta \boldsymbol{t}\ . \tag{2.19}$$

Aus der ersten Forderung Gleichung (2.6) folgt, dass die letzte Zeile von $\boldsymbol{H}$ positive Elemente besitzen muss.

Es empfiehlt sich, auch für die neuen Koordinaten einen Vektor der komplexen Frequenzen einzuführen

$$\boldsymbol{p_t} = [p_1\ ,\ \dots\ ,\ p_{k'}]^{\mathrm{T}}\ . \tag{2.20}$$

Dieser Vektor soll der Beziehung $\boldsymbol{p}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{p}_{\boldsymbol{t}}^{\mathrm{T}}\boldsymbol{t}$ genügen, was durch die Interpretation von Gleichung (2.16) als Operatorgleichung sichergestellt ist. Bemerkenswert ist an dieser Stelle, dass sich für $\mathrm{Re}\{p_{x_\kappa}\} = 0\ ,\ \kappa = 1 \dots k-1$ die linke offene p_t-Halbebene in die linke offene $\boldsymbol{p_t}$-Polyhalbebene abbildet, d. h.

$$\mathrm{Re}\{p_t\} < 0 \wedge \mathrm{Re}\{p_{x_\kappa}\} = 0 \text{ für } \kappa = 1 \dots k-1 \Longrightarrow \mathrm{Re}\{\boldsymbol{p_t}\} < \boldsymbol{0}\ . \tag{2.21}$$

Nachdem die alten Koordinaten durch die neuen Koordinaten und die neuen Ableitungen durch die alten Ableitungen ausgedrückt wurden, sollen nun die umgekehrten Beziehungen hergeleitet werden.

Wir bezeichnen eine Matrix $\boldsymbol{H}^{-\mathrm{R}}$ als die Rechtsinverse von $\boldsymbol{H}$, wenn sie eine Lösung der Gleichung

$$\boldsymbol{H}\boldsymbol{H}^{-\mathrm{R}} = \boldsymbol{1}_k. \tag{2.22}$$

ist.

Von den Rechtsinversen gibt es im Falle $k' > k$ unendlich viele. Genauer gesagt, der Freiheitsgrad jedes Spaltenvektors von $\boldsymbol{H}^{-\mathrm{R}}$ ist $k' - k$. Geht man davon aus, dass Gleichung (2.22) ein konsistentes Gleichungssystem ist, so lautet die allgemeine Lösung des Gleichungssystems

$$\boldsymbol{H}^{-\mathrm{R}} = \boldsymbol{H}_1^{\mathrm{S}} + [\boldsymbol{1}_{k'} - \boldsymbol{H}_1^{\mathrm{S}}\boldsymbol{H}]\boldsymbol{M}_1 \tag{2.23}$$

wobei $\boldsymbol{M}_1$ eine beliebige Matrix konsistenter Dimension ist, und die so genannte Semiinverse $\boldsymbol{H}_1^{\mathrm{S}}$ ist eine Matrix, die der Gleichung

$$\boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}_1^{\mathrm{S}}\boldsymbol{H} \tag{2.24}$$

genügt. $[\boldsymbol{1}_{k'} - \boldsymbol{H}_1^{\mathrm{S}}\boldsymbol{H}]\boldsymbol{M}_1$ beschreibt die homogene Lösung von Gleichung (2.22). $\boldsymbol{H}_1^{\mathrm{S}}$ ist hingegen die inhomogene Lösung. Sie ist allerdings i. Allg. durch $\boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}_1^{\mathrm{S}}\boldsymbol{H}$ allein nicht eindeutig bestimmt. Wir fordern zusätzlich $\boldsymbol{H}_{1,2}^{\mathrm{S}}\boldsymbol{H} = [\boldsymbol{H}_{1,2}^{\mathrm{S}}\boldsymbol{H}]^{\mathrm{T}}$. Zudem gelten $\boldsymbol{H}\boldsymbol{H}_1^{\mathrm{S}} = [\boldsymbol{H}\boldsymbol{H}_1^{\mathrm{S}}]^{\mathrm{T}}$ und $\boldsymbol{H}_1^{\mathrm{S}} = \boldsymbol{H}_1^{\mathrm{S}}\boldsymbol{H}\boldsymbol{H}_1^{\mathrm{S}}$, da $\boldsymbol{H}$ eine zeilenreguläre Matrix ist. Die Semiinverse $\boldsymbol{H}_{1,2}^{\mathrm{S}}$ ist somit zur Moore-Penrose-Inversen $\boldsymbol{H}^{+}$ festgelegt, die bekanntermaßen eindeutig ist und sich im Fall einer zeilenregulären Matrix $\boldsymbol{H}$ zu $\boldsymbol{H}^{+} = \boldsymbol{H}^{\mathrm{T}}[\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}]^{-1}$ berechnet. Der volle Lösungsraum von $\boldsymbol{H}^{-\mathrm{R}}$ bleibt aber trotz dieser Festlegung erhalten.

Wir fragen uns nun, welche Dimension der Raum hat, den die Spaltenvektoren der Matrix $[\boldsymbol{1}_{k'} - \boldsymbol{H}^{+}\boldsymbol{H}]$ aufspannen. Zunächst stellen wir fest, dass $\boldsymbol{H}^{+}\boldsymbol{H}$ eine idempotente Matrix vom Rang k ist. Sie hat k Eigenwerte bei 1 und $k' - k$ Eigenwerte bei 0. Da zur Einheitsmatrix jeder Vektor ein Eigenvektor ist, sind die Eigenvektoren von $\boldsymbol{H}^{+}\boldsymbol{H}$ auch Eigenvektoren der Matrix $[\boldsymbol{1}_{k'} - \boldsymbol{H}^{+}\boldsymbol{H}]$. Somit gilt für die Eigenwerte $\lambda\{[\boldsymbol{1}_{k'} - \boldsymbol{H}^{+}\boldsymbol{H}]\} = 1 - \lambda\{[\boldsymbol{H}^{+}\boldsymbol{H}]\}$. Die Matrix $[\boldsymbol{1}_{k'} - \boldsymbol{H}^{+}\boldsymbol{H}]$ hat also k Eigenwerte bei 0 und $k' - k$ Eigenwerte bei 1 und der aufgespannte Spaltenraum hat damit die Dimension $k' - k$, was dem Spaltendefekt von $\boldsymbol{H}$ entspricht.

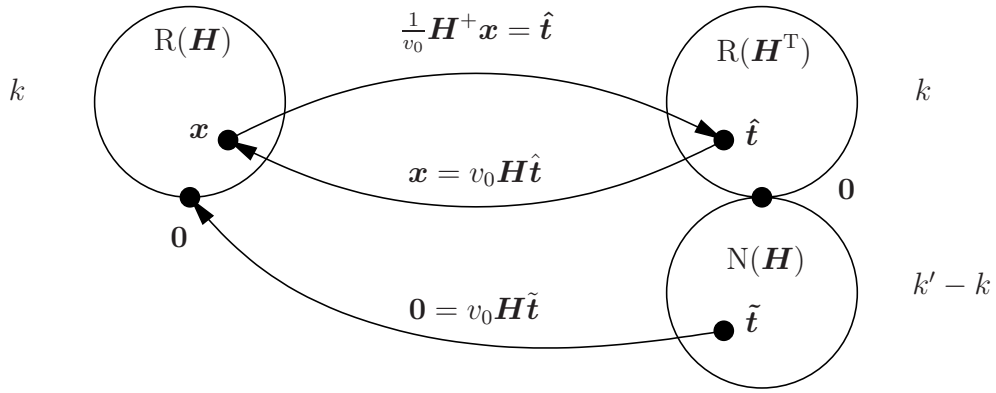


Bild 2.1: Koordinatentransformation

Nach diesen Vorbetrachtungen werden wir nun die Eigenschaften der Koordinatentransformation diskutieren. Setzt man in $\boldsymbol{x}=v_0\,\boldsymbol{H}\,\boldsymbol{t}$ den Vektor

$$\boldsymbol{t}=\underbrace{\frac{1}{v_0}\boldsymbol{H}^{-\mathrm{R}}\boldsymbol{x}}_{\hat{\boldsymbol{t}}}+\underbrace{[\mathbf{1}_{k'}-\boldsymbol{H}^{-\mathrm{R}}\boldsymbol{H}]\,\check{\boldsymbol{t}}}_{\tilde{\boldsymbol{t}}},\quad \check{\boldsymbol{t}}\text{ beliebig} \tag{2.25}$$

ein, so findet man

$$\boldsymbol{x}=v_0\boldsymbol{H}\boldsymbol{t}=\boldsymbol{H}\boldsymbol{H}^{-\mathrm{R}}\boldsymbol{x}+v_0\underbrace{[\boldsymbol{H}-\boldsymbol{H}\boldsymbol{H}^{-\mathrm{R}}\boldsymbol{H}]}_{=0}\tilde{\boldsymbol{t}}=\boldsymbol{x}, \tag{2.26}$$

d.h. jeder über Gleichung (2.25) aus $\boldsymbol{x}$ gebildete Vektor $\boldsymbol{t}$ wird über Gleichung (2.5) wieder eindeutig auf den ursprünglichen Vektor $\boldsymbol{x}$ abgebildet. Dies heißt aber nicht, dass der Vektor $\boldsymbol{t}$ zu einem Vektor $\boldsymbol{x}$ eindeutig ist, denn zum einen ist die Rechtsinverse $\boldsymbol{H}^{-\mathrm{R}}$ nicht eindeutig, und zum anderen ist $\check{\boldsymbol{t}}$ ein beliebiger Vektor konsistenter Dimension. Als Konsequenz daraus halten wir fest, dass sich zwar die alten Koordinaten eindeutig aus den neuen gewinnen lassen, aber mehrere verschiedene neue Koordinaten auf den gleichen Vektor im alten Koordinatensystem führen. Setzen wir in Gleichung (2.25) für $\boldsymbol{H}^{-\mathrm{R}}$ die allgemeine Lösung (Gleichung (2.23)) mit $\boldsymbol{H}_1^{\mathrm{S}}=\boldsymbol{H}^{+}$ ein, so erhalten wir

$$\boldsymbol{t}=\underbrace{\frac{1}{v_0}\boldsymbol{H}^{+}\boldsymbol{x}}_{\hat{\boldsymbol{t}}}+[\mathbf{1}_{k'}-\boldsymbol{H}^{+}\boldsymbol{H}]\underbrace{\left[\frac{1}{v_0}\boldsymbol{M}_1\boldsymbol{x}+\check{\boldsymbol{t}}-\boldsymbol{M}_1\boldsymbol{H}\check{\boldsymbol{t}}\right]}_{\boldsymbol{t}'}=\hat{\boldsymbol{t}}+\tilde{\boldsymbol{t}},\quad \check{\boldsymbol{t}}\,\text{beliebig}. \tag{2.27}$$

Der Vektor $\boldsymbol{t}'$ wird durch Matrix $[\mathbf{1}_k'-\boldsymbol{H}^{+}\boldsymbol{H}]$ in den Nullraum von $\boldsymbol{H}$ projiziert, somit liegt der Vektor $\tilde{\boldsymbol{t}}$ im Nullraum von $\boldsymbol{H}$, d.h. $\tilde{\boldsymbol{t}}\in\mathrm{N}(\boldsymbol{H})$. Der Vektor $\hat{\boldsymbol{t}}$ hingegen liegt im Spaltenraum von $\boldsymbol{H}^{\mathrm{T}}$, d.h. $\hat{\boldsymbol{t}}\in\mathrm{R}(\boldsymbol{H}^{\mathrm{T}})$. Der Vektor $\boldsymbol{x}$ liegt im Spaltenraum von $\boldsymbol{H}$. Der Nullraum $\mathrm{N}(\boldsymbol{H}^{\mathrm{T}})$ hat die Dimension 0, da wir die Zeilenregularität von $\boldsymbol{H}$ vorausgesetzt haben. Eineindeutigkeit zwischen den Vektoren $\boldsymbol{t}$ und $\boldsymbol{x}$ besteht nur für den im Spaltenraum von $\boldsymbol{H}^{\mathrm{T}}$ liegenden Teil des Vektors $\boldsymbol{t}$. Zur Verdeutlichung des Sachverhalts dient Bild 2.1 ([Fisc99], [Stra80]).

Wir werden nun aus der zweiten Forderung an die Koordinatentransformation Bedingungen an die Matrix $\boldsymbol{H}^{-\mathrm{R}}$ herleiten. Näherungsweise gilt

$$\Delta\boldsymbol{t}=\boldsymbol{J}_{\boldsymbol{x}}^{\boldsymbol{t}}\Delta\boldsymbol{x}=\frac{1}{v_0}\boldsymbol{H}^{-\mathrm{R}}\Delta\boldsymbol{x}\;. \tag{2.28}$$

Aus der zweiten Forderung an die Koordinatentransformation Gleichung (2.6) folgt unmittelbar, dass die letzte Spalte von $\boldsymbol{H}^{-\mathrm{R}}$ ausschließlich positive Elemente besitzen muss. Bei einer orthogonalen Matrix $\boldsymbol{H}$ fallen die beiden Bedingungen an $\boldsymbol{H}$, wegen $\boldsymbol{H}^{-\mathrm{R}}=\boldsymbol{H}^{\mathrm{T}}$, zusammen.

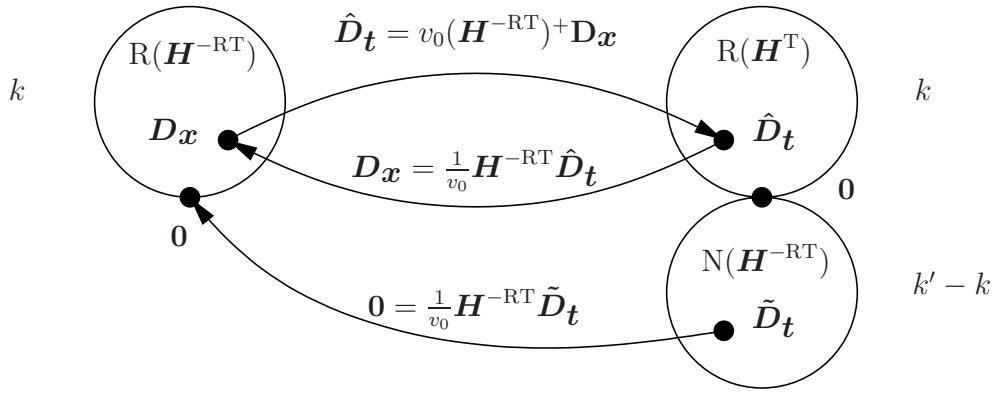


Bild 2.2: Transformation der Gradienten

Wir werden nun auf die Transformation der Gradienten im allgemeinen Fall eingehen. Dazu sei zunächst festgestellt, dass $\boldsymbol{D}_{\boldsymbol{t}} = v_0\boldsymbol{H}^{\mathrm{T}}\boldsymbol{D}_{\boldsymbol{x}}$ ein überbestimmtes Gleichungssystem mit Spaltendefekt 0 ist, von dem wir fordern, dass es konsistent ist. Aus $\boldsymbol{t} = \boldsymbol{H}^{-\mathrm{R}}\boldsymbol{x}/v_0$ gewinnt man wiederum die Beziehung zwischen den Differentialoperatoren

$$\mathbf{D}_{\boldsymbol{x}} = \frac{1}{v_0}[\boldsymbol{H}^{-\mathrm{R}}]^{\mathrm{T}}\mathbf{D}_{\boldsymbol{t}}\,. \tag{2.29}$$

Ein konsistentes, überbestimmtes Gleichungssystem mit Spaltendefekt 0 hat zwar eine eindeutige Lösung, aber es existieren im Fall $k' > k$ verschiedene Möglichkeiten, diese Lösung darzustellen. Dies äußert sich wiederum dadurch, dass die transponierte Rechtsinverse $\boldsymbol{H}^{-\mathrm{RT}} = [\boldsymbol{H}^{-\mathrm{R}}]^{\mathrm{T}}$ nicht eindeutig ist. Man verifiziert leicht, dass Gleichung (2.16) eine Lösung von Gleichung (2.29) ist und erhält durch Einsetzen von Gleichung (2.29) in Gleichung (2.16) die für eine speziell gewählte Rechtsinverse gültige Beziehung für die Differentialoperatoren im neuen Koordinatensystem

$$\mathbf{D}_{\boldsymbol{t}} = [\boldsymbol{H}^{-\mathrm{R}}\boldsymbol{H}]^{\mathrm{T}}\mathbf{D}_{\boldsymbol{t}}. \tag{2.30}$$

Die allgemeine Lösung lautet

$$\mathbf{D}_{\boldsymbol{t}} = v_0(\boldsymbol{H}^{-\mathrm{RT}})^{+}\mathbf{D}_{\boldsymbol{x}} + \underbrace{[\mathbf{1}_k - (\boldsymbol{H}^{-\mathrm{RT}})^{+}\,\boldsymbol{H}^{-\mathrm{RT}}]\boldsymbol{D'}_{\boldsymbol{t}}}_{=\tilde{\boldsymbol{D}}_{\boldsymbol{t}}} = \hat{\boldsymbol{D}}_{\boldsymbol{t}} + \tilde{\boldsymbol{D}}_{\boldsymbol{t}}\,. \tag{2.31}$$

Die Matrix $\boldsymbol{H}^{\mathrm{T}}$ ist selber Moore-Penrose-Inverse einer transponierten Rechtsinversen, d. h. $\boldsymbol{H}^{\mathrm{T}} = (\boldsymbol{H}^{-\mathrm{RT}})^{+}$. Somit sind die Spaltenräume gleich $\mathrm{R}(\boldsymbol{H}^{\mathrm{T}}) = \mathrm{R}((\boldsymbol{H}^{-\mathrm{RT}})^{+})$.

Die Ableitungsoperatoren des neuen Koordinatensystems liegen in einem durch die Spaltenvektoren von $\boldsymbol{H}^{-\mathrm{R}}$ aufgespannten k-dimensionalen Unterraum des $\mathbb{R}^{k'}$. Eineindeutigkeit besteht hier nur zwischen $\mathbf{D}_{\boldsymbol{x}}$ und den Ableitungsoperatoren des neuen Koordinatensystems $\mathbf{D}_{\boldsymbol{t}}$, die im Spaltenraum von $\boldsymbol{H}^{\mathrm{T}}$ liegen, siehe Bild 2.2.

In der Praxis werden beim Übergang von der Differentialgleichung im ursprünglichen Koordinatensystem in das neue Koordinatensystem i. d. R. mehrere verschiedene Rechtsinverse benutzt, um so möglichst einfache Referenznetze zu erhalten.

Als Beispiel sei

$$\boldsymbol{H} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{2.32}$$

gegeben und dazu die drei Rechtsinversen

$$\boldsymbol{H}_1^{-\mathrm{R}} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{H}_2^{-\mathrm{R}} = \begin{bmatrix} 0 & 1/3 \\ -1 & 1/3 \\ 1 & 1/3 \end{bmatrix}, \quad \boldsymbol{H}_3^{-\mathrm{R}} = \begin{bmatrix} -1 & 1/3 \\ -2 & 1/3 \\ 3 & 1/3 \end{bmatrix} \tag{2.33}$$

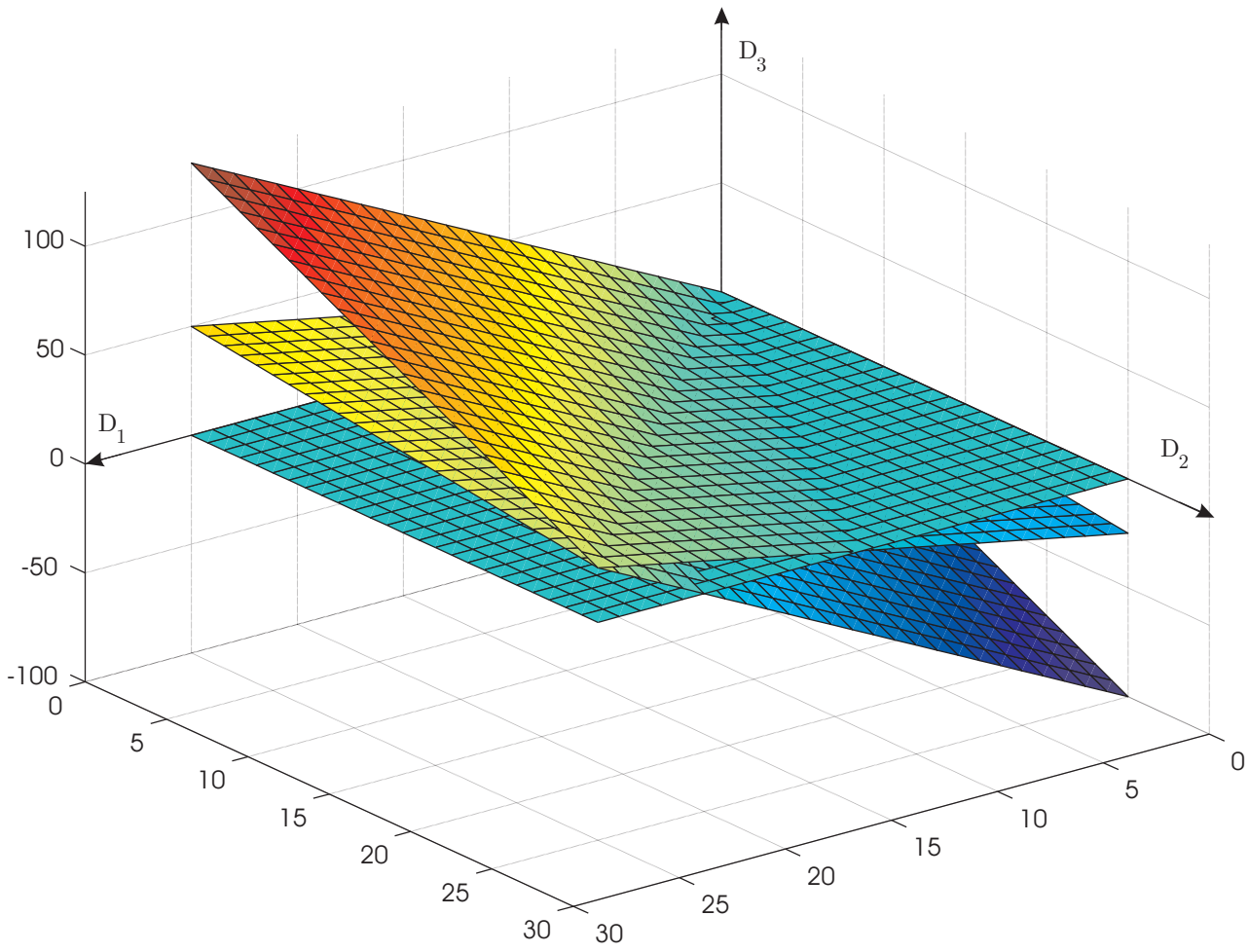


Bild 2.3: Durch die Spaltenvektoren der drei Rechtsinversen aufgespannten Ebenen (Spaltenräume von $\boldsymbol{H}^{-\mathrm{R}}$)

Bild 2.3 zeigt die Ebenen, auf denen die Lösungen $\mathbf{D}_{\boldsymbol{t}}$ von Gleichung (2.29) liegen.

2.3 Abtastung

Zur Erläuterung des Abtastvorgangs betrachten wir ein von der vektoriellen, kontinuierlichen Variablen $\boldsymbol{x}$ abhängiges Signal $f(\boldsymbol{x})$. In der Regel erfolgt die Diskretisierung der Variablen $\boldsymbol{x}$ durch

$$\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{X}_A\,\boldsymbol{\mu} \text{ mit } \boldsymbol{\mu} = [\mu_1\,,\,\cdots\,,\,\mu_{k-1}\,,\,\mu_k]^{\mathrm{T}} \in \mathbb{Z}^k\,. \tag{2.34}$$

Weiterhin benötigen wir später

$$\boldsymbol{\mu}' = [\mu_1\,,\,\cdots\,,\,\mu_{k-2}\,,\,\mu_{k-1}]^{\mathrm{T}}\,. \tag{2.35}$$

Die quadratische Matrix $\boldsymbol{X}_A$ bestimmt die Form des Abtastgitters. Wir bezeichnen sie als Abtastmatrix und setzen sie stets als regulär voraus. Sämtliche Ortspunkte zu einem konstanten Zeitpunkt μ_k wollen wir als eine Abtastschicht bezeichnen. Die durch die Abtastung entstandene Folge bezeichnen wir mit

$$\hat{f}(\boldsymbol{\mu}) = f(\boldsymbol{x}_0 + \boldsymbol{X}_A\,\boldsymbol{\mu}). \tag{2.36}$$

Der Abtastvorgang wird als linear bezeichnet, wenn $\boldsymbol{x}_0 = \boldsymbol{0}$ gilt, [Linn84]. Für die weitere Arbeit nehmen wir den Abtastvorgang als linear an. Die Fouriertransformierte der Funktion $f(\boldsymbol{x})$ lautet

$$F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}) = \mathcal{F}\{f(\boldsymbol{x})\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x})\,\mathrm{e}^{\,-\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{x}\ . \tag{2.37}$$

Die zugehörige Formel zur Rücktransformation ist

$$f(\boldsymbol{x}) = \mathcal{F}^{-1}\{F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}})\} = \frac{1}{(2\pi)^k}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}})\,\mathrm{e}^{\,\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{\omega}_{\boldsymbol{x}}\ . \tag{2.38}$$

Die orts-/zeitdiskrete Fouriertransformierte der Folge $f(\boldsymbol{X_A\mu})$ berechnet sich folgendermaßen

$$F_D(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}) = \mathcal{F}_D\{f(\boldsymbol{X}_A\boldsymbol{\mu})\} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^k} f(\boldsymbol{X}_A\boldsymbol{\mu})\,\mathrm{e}^{\,-\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{X}_A\boldsymbol{\mu}}\ . \tag{2.39}$$

Mit

$$f(\boldsymbol{X}_A\boldsymbol{\mu}) = \frac{|\det(\boldsymbol{X}_A)|}{(2\pi)^k}\int_{\boldsymbol{A}} F_D(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}})\,\mathrm{e}^{\,\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{X}_A\boldsymbol{\mu}}\,\mathrm{d}\boldsymbol{\omega}_{\boldsymbol{x}} \tag{2.40}$$

erhält man die Folge aus dem Frequenzspektrum, wobei $\boldsymbol{A}$ das Grundintervall ist. Das Frequenzspektrum der abgetasteten Funktion lässt sind durch $F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}})$ darstellen

$$\hat{f}(\boldsymbol{\mu}) \quad \overset{d}{\circ\!\!-\!\!\bullet} \quad \frac{1}{|\det \boldsymbol{X}_A|}\sum_{\boldsymbol{n}\in\mathbb{Z}^k} F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}} - \mathrm{j}2\pi\boldsymbol{X}_A^{-\mathrm{T}}\boldsymbol{n})\ . \tag{2.41}$$

Offenbar beeinflusst die Abtastmatrix $\boldsymbol{X}_A$ auch das Frequenzspektrum des abgetasteten Signals. Die orts-/zeitkontinuiertliche Funktion $f(\boldsymbol{x})$ lässt sich durch lineare Filterung zurückgewinnen, wenn sich die einzelnen Summanden nicht überlappen. Bei der Wahl der Abtastmatrix haben wir dies zu berücksichtigen. Die neue kontinuierliche unabhängige Variable $\boldsymbol{t}$ wird zu

$$\boldsymbol{t} = \boldsymbol{t}_0 + \boldsymbol{T}_A\,\boldsymbol{\nu} \text{ mit } \boldsymbol{\nu} = [\nu_1\ ,\ \cdots\ ,\ \nu_{k'-1}\ ,\ \nu_{k'}]^{\mathrm{T}}\ \in \mathbb{Z}^{k'} \tag{2.42}$$

diskretisiert. Weiterhin soll der Koordinatenursprung des Koordinatensystems $\boldsymbol{\nu}$ mit dem des Koordinatensystems $\boldsymbol{\mu}$ zusammenfallen, d. h. es gilt

$$\boldsymbol{x}_0 = v_0\,\boldsymbol{H}\,\boldsymbol{t}_0\ . \tag{2.43}$$

Setzt man die Abtastpunkte in die Transformationsvorschrift ein, so ergibt sich

$$\boldsymbol{X}_A\,\boldsymbol{\mu} = v_0\,\boldsymbol{H}\,\boldsymbol{T}_A\boldsymbol{\nu}\ . \tag{2.44}$$

Diese Beziehung offenbart die Tatsache, dass bei einer gegebenen Matrix $\boldsymbol{H}$ die Abtastmatrizen $\boldsymbol{X}_A$ und $\boldsymbol{T}_A$ nicht unabhängig voneinander wählbar sind. Die Abtastgitter in beiden Koordinatensystemen sind somit auch miteinander verbunden.

Der einfachste Fall einer Abtastung ist die Rechteckabtastung [Linn84]. Die Abtastmatrix lautet

$$\boldsymbol{X}_A = \mathbf{diag}(\Delta X_1\ ,\ \ldots\ ,\ \Delta X_k)\ , \tag{2.45}$$

wobei ΔX_κ den Abstand zweier Abtastpunkte in Richtung x_κ bezeichnet. Die Dreiecksabtastung im Zusammenhang mit Wellendigitalfiltern ist in [LF90] beschrieben. Im Fall $k = 3$ ist die Abtastmatrix durch

$$\boldsymbol{X}_A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \tag{2.46}$$

gegeben. Mit der Dreiecksabtastung erreicht man die höchste spektrale Packungsdichte, wenn das Frequenzspektrum $F(\mathrm{j}\boldsymbol{\omega}_{\boldsymbol{x}})$ (hyper-)kugelförmig bandbegrenzt ist. Allerdings wird dieser Vorteil dadurch erkauft, dass die Implementierung eines derartigen Abtastmusters einen erhöhten Realisierungsaufwand (insbesondere bei Hardware) erfordert [Frie00], [Lisc91], [Bose01]. Führen wir die Abtastung im Koordinatensystem $\boldsymbol{t}$ durch und wählen die Abtastmatrix $\boldsymbol{X}_A$ proportional zu $\boldsymbol{H}$, d.h.

$$\boldsymbol{X}_A = v_0 \boldsymbol{H} \boldsymbol{T}_A\ ,\ \boldsymbol{T}_A = \mathbf{1}_{k'} T\ , \tag{2.47}$$

so können wir bei geeigneter Wahl von $\boldsymbol{H}$ insbesondere Offset-und Rechteckraster erzeugen. Im Folgenden schränken wir uns auf $k = 4$ ein. Nachdem die üblicherweise verwendeten Abtastgitter kurz erläutert wurden, stellen wir nun sinnvolle Forderungen an die Abtastung.

- Jeder Verschiebe-Vektor, der von einem Abtastpunkt startet, muss wieder auf einem Abtastpunkt enden.
- Es darf nur eine Informationsübertragung von einer Abtastschicht zur unmittelbar darauf folgenden stattfinden. Diese Forderung findet ihre Begründung im Kapitel 3.
- Die Randbehandlung sollte möglichst einfach durchführbar sein.
- Eine systematische Synthese der Referenzschaltung sollte möglich sein.

Die erste Forderung lässt sich leicht erreichen, wenn die Abtastmatrix zu

$$\boldsymbol{X}_A = v_0\, T\, \boldsymbol{H} \tag{2.48}$$

gewählt wird. Der zweiten Forderung kann dadurch genügt werden, dass die letzte Zeile von $\boldsymbol{H}$ gleiche Elemente enthält. Die letzten beiden Forderungen lassen sich durch Rechteckabtastung in $\boldsymbol{x}$ erreichen. Wir wählen die Transformationsmatrix zu

$$\boldsymbol{H} = \mathbf{diag}\left(\frac{\Delta X_1}{v_0 T}\ ,\ \ldots\ ,\ \frac{\Delta X_k}{v_0 T}\right) \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}\ . \tag{2.49}$$

Zur prinzipiellen Erläuterung des Verfahrens wählt man üblicherweise $\Delta X_\kappa = v_0 T\,, \kappa = 1, 2, \ldots, k$, um die Darstellung zu entlasten. Die Abtastmatrix liefert bei dieser Wahl nicht das Frequenzspektrum mit der höchsten Packungsdichte. Insofern liegt es nahe, dass wir nicht den effizientesten Algorithmus erhalten. Es soll aber nochmals darauf hingewiesen werden, dass dies nicht Ziel dieser Arbeit ist. Vielmehr steht in dieser Arbeit automatische fehlervermeidende Codegenerierung im Vordergrund.

Zum Schluss sei noch erwähnt, dass wir im weiteren Verlauf der Arbeit den Begriff Raster anstatt Abtastgitter verwenden werden, da wir im Grunde kein Signal abtasten, sondern zu bestimmten Rasterpunkten Werte durch numerische Integration berechnen.

2.4 Berechnungsgebiete

In der Regel beschreiben die ersten k–1 Koordinaten des Vektors $\boldsymbol{x}$ Ortskoordinaten des physikalischen Raumes. Der Vektor der unabhängigen Ortsvariablen

$$ {}_{-k}\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \end{bmatrix} \tag{2.50} $$

liegt im $k-1$ dimensionalen Raum über dem Körper der reellen Zahlen, d. h. ${}_{-k}\boldsymbol{x} \in \mathbb{R}^{k-1}$ und beschreibt dann einen Ortspunkt im so genannten Echtraum. [2] Dabei setzen wir voraus, dass die ersten $k-1$ unabhängigen Variablen des Koordinatensystems $\boldsymbol{x}$ berandet sind.

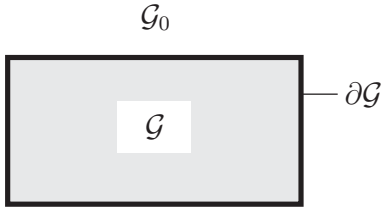


Bild 2.4: Das Berechnungsgebiet $\mathcal{G}$ berandet durch $\partial\mathcal{G}$ mit Außengebiet $\mathcal{G}_0$

Die Variable t setzen wir in positiver Richtung als unbegrenzt voraus, da die Laufzeit des Betriebsmittels nicht bekannt ist. Die obere Grenze der Koordinate x_κ lautet $x_{\kappa\,\max}$. Die untere Grenze bezeichnen wir mit $x_{\kappa\,\min}$. Im Kapitel 2.10 wird deutlich, dass bei dem in dieser Arbeit eingesetzten Verfahren zur Randbehandlung, die untere Grenze nicht grundsätzlich zu $x_{\kappa\,\min} = 0$ gewählt werden kann. Das durch die zuvor eingeführten Grenzen beschriebene Berechnungsgebiet bezeichnen wir mit $\mathcal{G}$. Wir nehmen an, dass es sich bei $\mathcal{G}$ um ein einfach zusammenhängendes Gebiet handelt. Ein Vektor $\boldsymbol{x}$ liegt genau dann in dem Berechnungsgebiet $\mathcal{G}$ falls

$$ x_{\kappa\,\min} < x_\kappa < x_{\kappa\,\max} \quad \forall\, \kappa = 1, 2, \ldots, k-1 \tag{2.51} $$

erfüllt ist. Durch die hier eingeführte untere und obere Grenze in Richtung x_κ wird in kartesischen Koordinaten für $k = 4$ eine rechteckige Box beschrieben. Im Falle allgemeiner Berechnungsgebiete wären die Grenzen selber noch Funktionen des Ortes. Der Einfachheit halber werden wir im Folgenden feste Grenzen annehmen.

Leider kann bei allgemeiner Abtastung nur mit expliziter Kenntnis der Abtastmatrix von den Grenzen in den kontinuierlichen Variablen auf die Anzahl Abtastpunkte in einer Richtung geschlossen werden. Die folgenden Ausführungen gelten daher nur für den Fall eines Rechteckgitters im Koordinatensystem $\boldsymbol{x}$. Die Anzahl der Abtastpunkte in Richtung x_κ bezeichnen wir mit P_{x_κ}. Wir definieren den Vektor

$$ \boldsymbol{P_x} = [P_{x_1}, P_{x_2}, \ldots, P_{x_{k-1}}]^{\mathrm{T}} . \tag{2.52} $$

Im Fall eines Rechteckgitters gibt es einen kleinsten Abtastpunkt μ_κ, der zu null gewählt wird und einen größten, der zu $P_{x_\kappa} - 1$ festgelegt ist. Wohlgemerkt korrespondiert der kleinste (größte) Abtastpunkt einer Richtung nicht mit $x_{\kappa\,\min}$ ($x_{\kappa\,\max}$), sondern stellt lediglich den grenznächsten Abtastpunkt innerhalb $\mathcal{G}$ dar. Die Gesamtanzahl der Abtastpunkte des Berechnungsgebietes $\mathcal{G}$ lautet

$$ P = \prod_{\kappa=1}^{k-1} P_{x_\kappa}. \tag{2.53} $$

[2] Wenn wir (unkorrekterweise) sagen, x ist eine reelle Zahl, so ist dies so zu verstehen, dass der Zahlenwert der physikalischen Größe x reell ist. Soweit die konkreten Zusammenhänge nichts anderes ergeben, gilt diese Vereinbarung für die gesamte Arbeit.

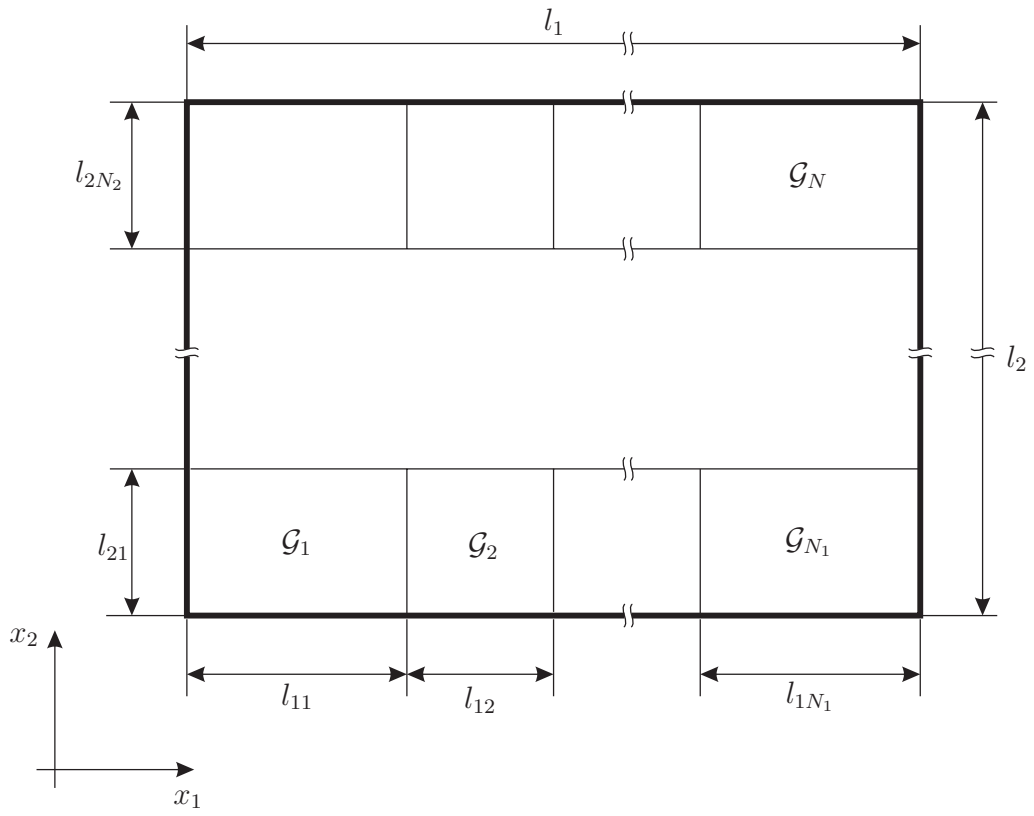


Bild 2.5: Abmessungen der Teilgebiete

Das Berechnungsgebiet $\mathcal{G}$ ist eine echte Teilmenge des $\mathbb{R}^{k-1}$, d. h.

$$\mathcal{G} \subset \mathbb{R}^{k-1}. \tag{2.54}$$

Den nicht zum Berechnungsgebiet gehörenden Teil des $\mathbb{R}^{k-1}$ bezeichnen wir mit $\mathcal{G}_0$, d.h.

$$\mathcal{G}_0 = \mathbb{R}^{k-1} \setminus \mathcal{G} \Longleftrightarrow \mathbb{R}^{k-1} = \mathcal{G}_0 \cup \mathcal{G}, \tag{2.55}$$

vgl. Bild 2.4. Teilgebiete homogener Materialverteilung bezeichnen wir mit $\mathcal{G}_n$. Das Berechnungsgebiet $\mathcal{G}$ setzt sich aus N disjunkten Teilgebieten $\mathcal{G}_n$ zusammen, d.h.

$$\mathcal{G} = \bigcup_{n=1}^{N} \mathcal{G}_n, \tag{2.56}$$

vgl. Bild 2.5.

Die Länge des Berechnungsgebietes in Richtung κ bezeichnen wir mit l_κ , $\kappa = 1, 2, \ldots, k$–1. Die Länge der Teilgebiete $\mathcal{G}_n$ in Richtung κ bezeichnen wir mit $l_{\kappa n}$, $n = 1, 2, \ldots, N_\kappa$. Die Anzahl der Teilgebiete $\mathcal{G}_n$ in Richtung κ lautet N_κ. Die Gesamtanzahl der Teilgebiete $\mathcal{G}_n$ ergibt sich zu

$$\prod_{\kappa=1}^{k-1} N_\kappa = N. \tag{2.57}$$

Zur Verdeutlichung der Situation im Fall $k = 3$ dient Bild 2.5. Wir unterscheiden zwischen $\mathcal{G}$ und $\mathcal{G}_0$, da uns die Lösung der Differentialgleichung nur für das Gebiet $\mathcal{G}$ interessiert. Die Parameter der Differentialgleichung innerhalb des Gebietes $\mathcal{G}$ unterscheiden sich oft sehr stark von denen des Gebietes $\mathcal{G}_0$. Der einfachste Fall liegt dann vor, wenn nur der Rand $\partial\mathcal{G}$ selber und nicht das Gebiet $\mathcal{G}_0$ Einfluss auf das Verhalten der Lösung innerhalb des Gebietes $\mathcal{G}$ nimmt.

2.5 Anordnung der Gitterpunkte und Abmessungen des Berechnungsgebietes

Die Lage der Abtastpunkte resultiert i. W. aus der Lage der Randpunkte. Diese haben den Abstand einer halben Abtastlänge zum Rand. Die Begründung hierfür liefern wir in 2.10.

Wir betrachten nun den Fall gleich großer Teilberechnungsgebiete und rechteckförmiger Abtastung innerhalb der Berechnungsgebiete. Zwischen den Abtastpunkten $\mu_\kappa = 0$ und $\mu_\kappa = P_{x_\kappa} - 1$ beträgt der Abstand $(P_{x_\kappa} - 1)\Delta X_\kappa$. Wir berücksichtigen noch, dass aufgrund obiger Argumentation zwischen dem Rand $\partial\mathcal{G}$ und den ihm nächsten Gitterpunkten ein Abstand von einer halben Abtastlänge $\Delta X_\kappa/2$ liegt (siehe Bild 2.6). Die Längen l_κ betragen

$$l_\kappa = x_{\kappa\max} - x_{\kappa\min} = \Delta X_\kappa P_{x_\kappa} \iff \Delta X_\kappa = \frac{l_\kappa}{P_{x_\kappa}}. \tag{2.58}$$

Sind die Teilgebiete unterschiedlich groß, so ergibt sich die Länge des Teilgebietes n in Richtung μ_κ zu

$$l_{\kappa n} = \begin{cases} \Delta X_\kappa \left[P_{x_\kappa n} - \dfrac{1}{2}\right] & \text{für} \quad n = 1, N_\kappa \\ \Delta X_\kappa \left[P_{x_\kappa n} - 1\right] & \text{für} \quad n = 2, \ldots, N_\kappa - 1, \end{cases} \tag{2.59}$$

wobei $P_{x_\kappa n}$ die Anzahl Abtastpunkte des Teilgebietes n in Richtung x_κ ist. Die Summe der Längen der Teilgebiete in Richtung κ muss die Gesamtlänge des Berechnungsgebietes in Richtung κ ergeben, d.h.

$$l_\kappa = \sum_{n=1}^{N_\kappa} l_{\kappa n}. \tag{2.60}$$

Setzen wir Gleichung (2.59) ein, so erhalten wir zunächst

$$l_\kappa = \Delta X_\kappa \left[-1 - (N_\kappa - 2) + \sum_{n=1}^{N_\kappa} P_{x_\kappa n}\right] = \Delta X_\kappa \left[\sum_{n=1}^{N_\kappa} P_{x_\kappa n} \; - (N_\kappa - 1)\right] \tag{2.61}$$

und durch Vergleich mit Gleichung (2.58)

$$P_{x_\kappa} = \sum_{n=1}^{N_\kappa} P_{x_\kappa n} \; - (N_\kappa - 1), \tag{2.62}$$

was durch die Tatsache bestätigt wird, dass genau die $N_\kappa - 1$ inneren Abtastpunkte in Richtung κ zu zwei Gebieten gemeinsam gehören.

2.6 Mehrdimensionale Kirchhoff'sche Netze

In diesem Abschnitt werden die für diese Arbeit benötigten Definitionen und Eigenschaften mehrdimensionaler Kirchhoff'scher Netze angegeben. Hierzu führen wir zunächst den Torbegriff ein. Wir verstehen unter einem Tor ein Klemmenpaar $1 - 1'$ gemäß Bild 2.7, welches durch die Spannung $u(\boldsymbol{t})$ und den Strom $i(\boldsymbol{t})$ (diese sind proportional zu den Feldgrößen in dem im Kapitel 2.1 definierten Sinne) charakterisiert ist. Zu beachten ist, dass hineinfließender und herausfließender Strom gleich sind. Neben der Beschreibung mit Spannung und Strom verwenden wir die Beschreibung mit Wellengrößen. Die Verwendung der Wellengrößen stellt eine bijektive Koordinatentransformation der unabhängigen Größen dar.

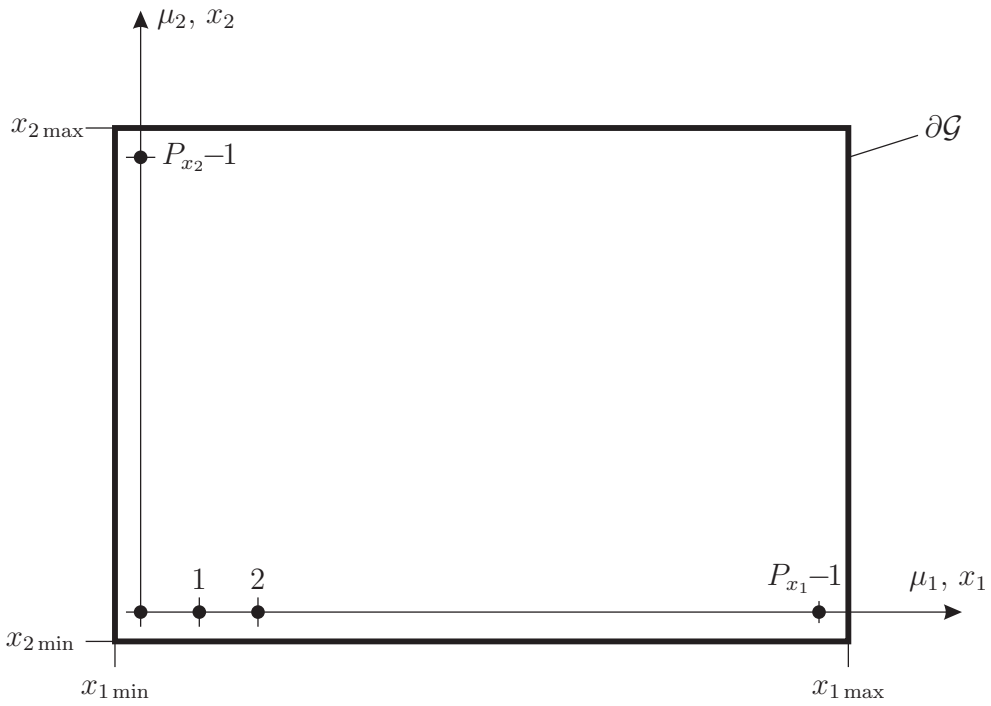


Bild 2.6: Zur Lage der Abtastpunkte

Diese Koordinatentransformation ist durch den positiv angenommenen so genannten Torwiderstand R bestimmt. Die gebräuchlichsten Wellengrößen sind die Spannungswellen

$$a' = u + Ri\,,\ b' = u - Ri \tag{2.63}$$

und die Leistungswellen

$$a = \frac{u + Ri}{2\sqrt{R}}\,,\ b = \frac{u - Ri}{2\sqrt{R}}\,. \tag{2.64}$$

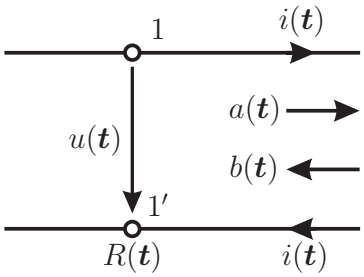


Bild 2.7: Zum Torbegriff

Nun betrachten wir ein n-Tor. Zweckmäßigerweise fassen wir zur Beschreibung des n-Tors alle Spannungen, Ströme und Wellengrößen in Vektoren zusammen. Weiterhin definieren wir eine reelle, p. d., symmetrische Torwiderstandsmatrix $\boldsymbol{R}$. Sofern nichts anderes gesagt ist, gehen wir von einer Diagonalmatrix aus, d. h. $\boldsymbol{R} = \boldsymbol{G}^{-1} = \mathbf{diag}(R_1, \ldots, R_n)$. Die inverse Matrix $\boldsymbol{G}$ bezeichnen wir als Torleitwert-

matrix. Wir erhalten die kompakte Darstellung

$$\begin{bmatrix} \boldsymbol{b}' \\ \boldsymbol{a}' \end{bmatrix} = \begin{bmatrix} \mathbf{1} & -\boldsymbol{R} \\ \mathbf{1} & \boldsymbol{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{i} \end{bmatrix} \Longleftrightarrow \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{i} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ -\boldsymbol{G} & \boldsymbol{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}' \\ \boldsymbol{a}' \end{bmatrix} \tag{2.65}$$

$$\begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \boldsymbol{G}^{1/2} & -\boldsymbol{R}^{1/2} \\ \boldsymbol{G}^{1/2} & \boldsymbol{R}^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{i} \end{bmatrix} \Longleftrightarrow \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}^{1/2} & \boldsymbol{R}^{1/2} \\ -\boldsymbol{G}^{1/2} & \boldsymbol{G}^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} . \tag{2.66}$$

Falls die einfallenden Wellen des n-Tores unabhängig voneinander gewählt werden können, gelten die Beziehungen

$$\boldsymbol{b} = \boldsymbol{S}\boldsymbol{a} \quad \text{und} \quad \boldsymbol{b}' = \boldsymbol{S}'\boldsymbol{a}' , \tag{2.67}$$

wobei $\boldsymbol{S}$ die Leistungswellenstreumatrix und $\boldsymbol{S}'$ die Spannungswellenstreumatrix darstellt. Die beiden Arten von Streumatrizen sowie die Wellengrößen lassen sich ineinander umrechnen

$$\boldsymbol{S}' = \boldsymbol{R}^{1/2}\boldsymbol{S}\boldsymbol{G}^{1/2} \quad , \quad \boldsymbol{b}' = 2\,\boldsymbol{R}^{1/2}\boldsymbol{b} \quad , \quad \boldsymbol{a}' = 2\,\boldsymbol{R}^{1/2}\boldsymbol{a} . \tag{2.68}$$

Leistungswellen garantieren die volle Robustheit des Wellendigitalalgorithmus im nicht konstanten und nicht linearen Fall. Hingegen ist bei linearen, konstanten Bauelementen die volle Robustheit auch schon durch Verwendung von Spannungswellen sichergestellt, [Fett98].

Die vom n-Tor aufgenommene Leistungsdichte ist die Summe der über die n Tore übertragene Leistungsdichte

$$p(\boldsymbol{t}) = \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{t})\,\boldsymbol{i}(\boldsymbol{t}) = \|\boldsymbol{a}\|^2 - \|\boldsymbol{b}\|^2 = \boldsymbol{a}'^{\mathrm{T}}\boldsymbol{G}\boldsymbol{a}' - \boldsymbol{b}'^{\mathrm{T}}\boldsymbol{G}\boldsymbol{b}' = \sum_{\mu=1}^{n}[a_\mu^2 - b_\mu^2] = \sum_{\mu=1}^{n}\sum_{\nu=1}^{n}[a'_\mu G_{\mu\nu} a'_\nu - b'_\mu G_{\mu\nu} b'_\nu] . \tag{2.69}$$

Korrekterweise können wir im mehrdimensionalen Fall nicht mehr von Spannungen und Strömen sprechen, da es sich hier auch um Dichten handelt. Wir wollen aber trotzdem an den eingeführten Begriffen Spannung und Strom festhalten. In [Fett99] und [Heme95] wird die Leistungsdichte auch als mehrdimensionale Leistung bezeichnet.

In diesem Zusammenhang sprechen wir von externen Eigenschaften, wenn wir das Verhalten des Systems nur von außen betrachten. Die inneren Zustände werden dabei nicht berücksichtigt, insbesondere kann der Fall auftreten, dass das externe System eine Eigenschaft besitzt, die ein Teilsystem nicht besitzt. Ein System besteht in der Regel aus mehreren (Elementar-)elementen. Als Element bezeichnen wir ein System, welches nicht weiter zerlegt werden kann. Falls ein System keine Quellen besitzt und alle Elemente passiv (verlustfrei, energieneutral) sind, dann heißt das System intern passiv (verlustfrei, energieneutral). Zur geeigneten Definition der Passivität greifen wir auf [Nits93], [Fett99] und [Heme95] zurück. Dabei werden wir zunächst Energiebetrachtungen im Koordinatensystem $\boldsymbol{x}$ durchführen. Wir wollen uns dabei auf den Fall $k = 4$ einschränken und betrachten ein 4-dimensionales Gebiet $\mathcal{G}_{\boldsymbol{x}}$ mit dem Rand $\partial\mathcal{G}_{\boldsymbol{x}}$, einem Flächenelement $\mathrm{d}A_{\boldsymbol{x}}$ und dem nach außen gerichteten Normalenvektor $\boldsymbol{n}_{\boldsymbol{x}}$. Im Folgenden wollen wir die auf ein System wirkenden energetischen Einflüsse in einer Energiebilanz quantitativ beschreiben. Dazu führen wir die im System gespeicherte Energiedichte

$$\boldsymbol{W}_{\mathrm{s}} = [W_x\ ,\ W_y\ ,\ W_z\ ,\ W_t]^{\mathrm{T}} \quad \text{mit} \quad W_t \geq 0 . \tag{2.70}$$

ein. Die physikalische Einheit ist für alle Koordinaten Joule pro Kubikmeter, d. h. $[\boldsymbol{W}_{\mathrm{s}}] = [\mathrm{J/m^3}]$. Auf die Energiebilanz haben die folgenden Größen einen Einfluss. Zunächst lassen sich die Einflüsse unterteilen nach Ursachen innerhalb des Gebietes $\mathcal{G}_{\boldsymbol{x}}$ und nach Ursachen, die ihren Ursprung außerhalb des Gebiets

haben. Von Außen geht die durch die Gebietsoberfläche (diese räumliche und zeitliche Begrenzung wird auch als Segmentgrenze bezeichnet) transportierte Energiedichte $\boldsymbol{W}_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{n}_{\boldsymbol{x}}$ in integraler Form als die Energie

$$\int_{\partial\mathcal{G}_{\boldsymbol{x}}} \boldsymbol{W}_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{n}_{\boldsymbol{x}}\,\mathrm{d}A_{\boldsymbol{x}} \tag{2.71}$$

in die Energiebilanz ein. Falls das Integral positiv ist, sprechen wir von einer nach Außen abgeführten Energie. Innerhalb des Gebietes kann ein Energieaustausch mit anderen Systemen über die so genannte Prozessgrenze (das sind die Tore des Systems) erfolgen. Die über die Tore zugeführte Leistungsdichte ist p, welche die Einheit Watt pro Kubikmeter hat, d. h. $[p] = [\mathrm{W/m^3}]$.

Die über die Tore zugeführte Energie lautet mit $\mathrm{d}V_{\boldsymbol{x}} = \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}x_4 = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,v_4\,\mathrm{d}t$

$$\frac{1}{v_4}\int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}V_{\boldsymbol{x}} = \int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}t\ . \tag{2.72}$$

Die durch die Quellen über die Quellengrenze zugeführte Leistungsdichte ist p^Q. Die dem Gebiet $\mathcal{G}_{\boldsymbol{x}}$ durch Quellen zugeführte Energie ist

$$\frac{1}{v_4}\int_{\mathcal{G}_{\boldsymbol{x}}} p^Q\mathrm{d}V_{\boldsymbol{x}}\ . \tag{2.73}$$

Letztendlich entweicht über die Dissipationsgrenze die nicht negative Leistungsdichte $p^V \geq 0$, sodass die im Gebiet $\mathcal{G}_{\boldsymbol{x}}$ dissipierte Energie

$$\frac{1}{v_4}\int_{\mathcal{G}_{\boldsymbol{x}}} p^V\mathrm{d}V_{\boldsymbol{x}} \tag{2.74}$$

ist. Alle energetischen Einflüsse auf das System sind im Bild 2.8 dargestellt. Für die Energiebilanz

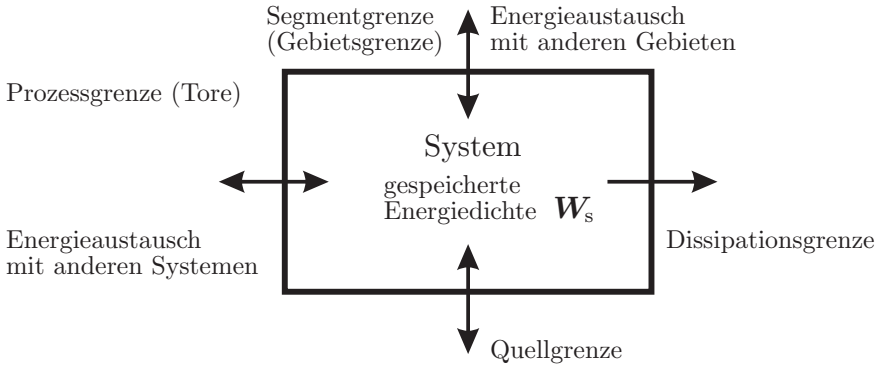


Bild 2.8: Energiebilanz eines Systems (z.B. n-Tor)

vereinbaren wir die folgende Vorzeichenkonvention

$$\int_{\partial\mathcal{G}_{\boldsymbol{x}}} \boldsymbol{W}_{\mathrm{s}}^{\mathrm{T}}\,\boldsymbol{n}\,\mathrm{d}A_{\boldsymbol{x}} + \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p^V}{v_4}\mathrm{d}V_{\boldsymbol{x}} = \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p^Q}{v_4}\mathrm{d}V_{\boldsymbol{x}} + \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p}{v_4}\mathrm{d}V_{\boldsymbol{x}}\ . \tag{2.75}$$

Mit dem Gauß'schen Satz gilt unter sehr allgemeinen Bedingungen (Stetigkeit der Integranden und endliche Grenzen sind bereits hinreichend)

$$\int_{\mathcal{G}_{\boldsymbol{x}}} \mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}}\mathrm{d}V_{\boldsymbol{x}} + \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p^V}{v_4}\mathrm{d}V_{\boldsymbol{x}} = \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p^Q}{v_4}\mathrm{d}V_{\boldsymbol{x}} + \int_{\mathcal{G}_{\boldsymbol{x}}} \frac{p}{v_4}\mathrm{d}V_{\boldsymbol{x}}\ . \tag{2.76}$$

Da wir keine Einschränkungen bzgl. des Gebietes $\mathcal{G}_{\boldsymbol{x}}$ gemacht haben, muss die Energiebilanz für beliebige $\mathcal{G}_{\boldsymbol{x}}$ gültig sein. Hieraus folgt die Gleichheit der Integranden, die als Energieerhaltung in differentieller Form oder Leistungsdichtebilanz bezeichnet wird

$$v_4\,\mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}} + p^V = p^Q + p\;. \tag{2.77}$$

Passive Systeme definieren wir nun durch Quellenfreiheit $p^Q = 0$. Dann gilt mit $p^V \geq 0$ die Ungleichung

$$\int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}V_{\boldsymbol{x}} - v_4 \int_{\mathcal{G}_{\boldsymbol{x}}} \mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}}\mathrm{d}V_{\boldsymbol{x}} = \int_{\mathcal{G}_{\boldsymbol{x}}} p^V\,\mathrm{d}V_{\boldsymbol{x}} \geq 0\;, \tag{2.78}$$

bzw. in differentieller Form

$$p - v_4\mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}} = p^V \geq 0 \Longrightarrow p \geq v_4\mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}}\;. \tag{2.79}$$

Verlustfreie Systeme sind passive Systeme, die keine Leistung dissipieren, d. h.

$$p^V = 0 \Longleftrightarrow p = v_4\mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_{\mathrm{s}}\;. \tag{2.80}$$

Energieneutrale Systeme sind dadurch gekennzeichnet, dass sie verlustfrei sind und keine Energie speichern, d. h. $\boldsymbol{W}_{\mathrm{s}} = \mathbf{0}$. Sie nehmen daher auch keine Energie auf $p = 0$.

Um eine andere Interpretation der Passivitätsbedingung zu erhalten, teilen wir die Energiedichte in einen räumlichen und einen zeitlichen Anteil auf

$$\boldsymbol{W}_{\mathrm{s}} = \left[\boldsymbol{W}_{\mathrm{sp}}^{\mathrm{T}}\,,\, W_t\right]^{\mathrm{T}}\;. \tag{2.81}$$

Zudem definieren wir die in einem örtlichen Gebiet $\mathcal{G}$ mit Rand $\partial\mathcal{G}$, einem Flächenelement $\mathrm{d}A$ und dem nach außen gerichteten Normalenvektor $\boldsymbol{n}$ gespeicherte Energie zu

$$E = \int_{\mathcal{G}} W_t\,\mathrm{d}V \quad,\quad \mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\;. \tag{2.82}$$

Die Energie hat die Einheit Joule $[E] = [\mathrm{J}]$. Nehmen wir stetige Funktionen und endliche Integralgrenzen an, so ist die Vertauschbarkeit der Integrationsreihenfolge gegeben. Dann können wir die transportierte Energie durch

$$\begin{aligned}
\int_{\mathcal{G}_{\boldsymbol{x}}} \mathbf{D}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{W}_s\,\mathrm{d}V_{\boldsymbol{x}} &= \int_{\mathcal{G}_{\boldsymbol{x}}} \operatorname{div}\boldsymbol{W}_{\mathrm{sp}}\,\mathrm{d}V_{\boldsymbol{x}} \qquad + \frac{1}{v_4}\int_{\mathcal{G}_{\boldsymbol{x}}} \mathrm{D}_t\,W_t\,\mathrm{d}V_{\boldsymbol{x}} \\
&= v_4\int_{t_0}^{t_1}\int_{\mathcal{G}} \operatorname{div}\boldsymbol{W}_{\mathrm{sp}}\mathrm{d}V\,\mathrm{d}t + \int_{\mathcal{G}}\int_{t_0}^{t_1} \mathrm{D}_t\,W_t\,\mathrm{d}t\,\mathrm{d}V \\
&= v_4\int_{t_0}^{t_1}\int_{\mathcal{G}} \operatorname{div}\boldsymbol{W}_{\mathrm{sp}}\mathrm{d}V\,\mathrm{d}t + \int_{\mathcal{G}} [W_t(t_1) - W_t(t_0)]\,\mathrm{d}V \\
&= v_4\int_{t_0}^{t_1}\int_{\partial\mathcal{G}} \boldsymbol{W}_{\mathrm{sp}}^{\mathrm{T}}\boldsymbol{n}\mathrm{d}A\mathrm{d}t \quad + E(t_1) - E(t_0)
\end{aligned} \tag{2.83}$$

ausdrücken. Hierin ist

$$v_4\int_{t_0}^{t_1}\int_{\partial\mathcal{G}} \boldsymbol{W}_{\mathrm{sp}}^{\mathrm{T}}\boldsymbol{n}\mathrm{d}A\mathrm{d}t \tag{2.84}$$

die im Zeitintervall $t_0 < t < t_1$ über die örtliche Segmentgrenze zugeführte Energie. Die im 4-dimensionalem Raum transportierte Energie entspricht somit der örtlich transportierten Energie und der Differenz der gespeicherten Energie.

Wir wollen nun eine notwendige Bedingung für passive Systeme herleiten. Dazu nehmen wir an, dass zwischen den Zeitpunkten t_0 und t_1 dem System über die örtliche Segmentgrenze keine Energie zugeführt wird. Dies ist dann der Fall, wenn die Normalkomponente der Energiedichte $\boldsymbol{W}_{\mathrm{sp}}$ auf $\partial\mathcal{G}$ verschwindet, d. h. $\boldsymbol{W}_{\mathrm{sp}}^{\mathrm{T}}\boldsymbol{n} = 0\,\mathrm{auf}\,\partial\mathcal{G}$. Nehmen wir noch $E(t_0) = 0$ an, berücksichtigen $W_t \geq 0$ und Gleichung (2.82), so folgt aus der Passivitätsbedingung Gleichung (2.78) dass die zugeführte Energie nicht negativ ist.

$$0 \leq \frac{1}{v_4}\int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}V_{\boldsymbol{x}}\ . \tag{2.85}$$

Um nun eine geeignete Definition für mehrdimensional passive Systeme in den Koordinaten $\boldsymbol{t}$ anzugeben, betrachten wir ein k'-dimensionales Gebiet $\mathcal{G}_{\boldsymbol{t}}$ mit dem Rand $\partial\mathcal{G}_{\boldsymbol{t}}$, einem Flächenelement $\mathrm{d}A_{\boldsymbol{t}}$ und dem nach außen gerichteten Normalenvektor $\boldsymbol{n}_{\boldsymbol{t}}$. Die vom n-Tor aufgenommene Energie ergibt sich durch Integration der zugeführten Leistungsdichte über $\mathcal{G}_{\boldsymbol{t}}$, d. h.

$$W = \int_{\mathcal{G}_{\boldsymbol{t}}} p'\ \mathrm{d}t_1 \ldots \mathrm{d}t_{k'}\ . \tag{2.86}$$

Diese Energie soll nach Definition die Einheit einer physikalischen Energie besitzen. Zusätzlich zur aufgenommen Energie definieren wir eine (vektorielle) gespeicherte Energiedichte

$$\hat{\boldsymbol{W}}_s(\boldsymbol{t}) = [W_1'\ ,\ W_2'\ ,\ \ldots\ ,\ W_{k'}']^{\mathrm{T}}\ . \tag{2.87}$$

Mit diesen Vorbetrachtungen können wir nun die MD-Passivität eines Systems näher beschreiben. Ein System heißt MD-extern passiv, falls eine (vektorielle) Funktion $\hat{\boldsymbol{W}}_s$ existiert, für die gilt

$$\begin{aligned} &1.\quad \hat{\boldsymbol{W}}_s(\boldsymbol{t}) \geq \boldsymbol{0} \quad (\text{koordinatenweise } \geq 0\,) \text{ und} \\ &2.\quad W \geq \int_{\partial\mathcal{G}_{\boldsymbol{t}}} \hat{\boldsymbol{W}}_s^{\mathrm{T}}\boldsymbol{n}_{\boldsymbol{t}}\,\mathrm{d}A_{\boldsymbol{t}} \quad (\text{für beliebige Gebiete})\ . \end{aligned} \tag{2.88}$$

Ein System heißt MD-extern verlustfrei, falls das System passiv ist und für beliebige Gebiete $\mathcal{G}_{\boldsymbol{t}}$ gilt

$$W = \int_{\partial\mathcal{G}_{\boldsymbol{t}}} \hat{\boldsymbol{W}}_s^{\mathrm{T}}\boldsymbol{n}_{\boldsymbol{t}}\,\mathrm{d}A_{\boldsymbol{t}} \tag{2.89}$$

Ein System heißt MD-extern energieneutral, falls das System verlustfrei ist und für beliebige Zeitpunkte $\boldsymbol{t}$ gilt

$$\hat{\boldsymbol{W}}_s(\boldsymbol{t}) = \boldsymbol{0}\ . \tag{2.90}$$

Ein System heißt MD-extern dissipativ, falls das System passiv ist und ein Vorgang existiert, sodass

$$W > \int_{\partial\mathcal{G}} \hat{\boldsymbol{W}}_s^{\mathrm{T}}\boldsymbol{n}_{\boldsymbol{t}}\,\mathrm{d}A_{\boldsymbol{t}} \tag{2.91}$$

gilt.

Wir betrachten nun den Sonderfall eines linearen konstanten Systems. Im stationären Zustand folgen die Spannungen und Ströme den Beziehungen

$$\boldsymbol{u}(\boldsymbol{t}) = \mathrm{Re}\{\boldsymbol{U}\,\mathrm{e}^{\boldsymbol{p}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{t}}\}\ \ ,\ \ \boldsymbol{i}(\boldsymbol{t}) = \mathrm{Re}\{\boldsymbol{I}\,\mathrm{e}^{\boldsymbol{p}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{t}}\}\,. \tag{2.92}$$

Wir definieren die auf die gesamte $\boldsymbol{p}_{\boldsymbol{t}}$-Polyebene verallgemeinerte Wirkleistung zu $P = \frac{1}{2}\mathrm{Re}\{\boldsymbol{I}^{\mathrm{H}}\boldsymbol{U}\}$. Wohlgemerkt ist P aber nur für $\mathrm{Re}\,\boldsymbol{p}_{\boldsymbol{t}} = \boldsymbol{0}$ die im Mittel über die n Tore übertragene Leistung. Es ergeben sich die folgenden notwendigen und hinreichenden Bedingungen für P

- Passives System : $P \geq 0$ für $\mathrm{Re}\,\boldsymbol{p_t} > 0$
- Verlustfreies System : passiv und $P = 0$ für $\mathrm{Re}\,\boldsymbol{p_t} = 0$
- Energieneutrales System : verlustfrei und $P = 0 \;\forall\; \boldsymbol{p_t}$.

2.7 Eigenschaften mehrdimensionaler Wellendigitalfilter

Ein System heißt diskret MD-extern passiv, falls eine (vektorielle) Funktion der gespeicherten Energie

$$\hat{\boldsymbol{W}}_s = [\hat{W}_{s1}\ ,\ \ldots\ ,\ \hat{W}_{sk'}]^{\mathrm{T}} \tag{2.93}$$

existiert, die koordinatenweise nicht negativ ist. Zudem darf die am Rasterpunkt $\boldsymbol{T\nu}$ zugeführte Energie $Tp(\boldsymbol{T\nu})$ nicht kleiner als die Summe der Energiedifferenzen des aktuellen und des unmittelbar zurückliegenden Rasterpunktes über alle Raumkoordinaten sein, d. h. es muss gelten

$$\begin{aligned} &1.\quad \hat{\boldsymbol{W}}_s(\boldsymbol{T\nu}) \geq \boldsymbol{0} \quad (\text{koordinatenweise } \geq 0\,) \text{ und} \\ &2.\quad Tp(\boldsymbol{T\nu}) \geq \sum_{\kappa=1}^{k'} \hat{W}_{s\kappa}(\boldsymbol{T\nu}) - \sum_{\kappa=1}^{k'} \hat{W}_{s\kappa}(\boldsymbol{T\nu} - \Delta\boldsymbol{t})\ . \end{aligned} \tag{2.94}$$

Ein System heißt diskret MD-extern verlustfrei, falls das System passiv ist und die zum Zeitpunkt $\boldsymbol{T\nu}$ zugeführte Energie $Tp(\boldsymbol{T\nu})$ gleich der Summe der Energiedifferenzen des aktuellen und des unmittelbar zurückliegenden Rasterpunktes über alle Raumkoordinaten ist

$$Tp(\boldsymbol{T\nu}) = \sum_{\kappa=1}^{k'} \hat{W}_{s\kappa}(\boldsymbol{T\nu}) - \sum_{\kappa=1}^{k'} \hat{W}_{s\kappa}(\boldsymbol{T\nu} - \Delta\boldsymbol{t})\ . \tag{2.95}$$

Ein System heißt diskret MD-extern energieneutral, falls das System verlustfrei ist und darüberhinaus die gespeicherte Energie an allen Abtastpunkten null ist, d. h.

$$\hat{\boldsymbol{W}}_s(\boldsymbol{T\nu}) = \boldsymbol{0}\ . \tag{2.96}$$

Die in 2.6 gemachten Aussagen bzgl. interner Eigenschaften gelten auch hier.

2.8 Wellendigitalfilter Bauelemente

2.8.1 Energieneutrale Bauelemente

Energieneutrale n-Tore sind gemäß Gleichung (2.90) dadurch gekennzeichnet, dass die aufgenommene MD-Leistung zu jedem Zeitpunkt null ist. Die zum n-Tor zugehörige Streumatrix genügt der Eigenschaft

$$\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} = \boldsymbol{1}_n\ . \tag{2.97}$$

Man beachte, dass die Streumatrix eine Funktion der Zeit und der Feldgrößen sein darf.

Die direkte Implementierung der Streumatrix auf einem Digitalrechner ist nicht sinnvoll, da bei endlicher Wortlänge die Sicherstellung der Passivität einen erheblichen Aufwand erfordert. Zur Lösung des Problems greifen wir auf die bekannte Tatsache zurück, dass die Leistungswellen-Streumatrix eines energieneutralen n-Tors mittels QR-Zerlegung nach Givens durch Streumatrizen von $(n^2 - n)/2 = \binom{n}{2}$

Zweitoren und n skalaren unimodularen Multiplizierern ρ_ν dargestellt werden kann [MF96], [Frän97], d. h.

$$\boldsymbol{S} = \left[\prod_{\mu=1}^{\binom{n}{2}} \boldsymbol{H}_\mu\right] \boldsymbol{R} \quad , \ \tbinom{n}{2}=\tfrac{n(n-1)}{2} \ , \text{ wobei} \tag{2.98}$$

$$\begin{aligned} \boldsymbol{H}_\mu = \ & \mathbf{1}_n + \mathbf{e}_\alpha \mathbf{e}_\beta^{\mathrm{T}} s_\mu + \mathbf{e}_\beta \mathbf{e}_\alpha^{\mathrm{T}} s_\mu^* + [\mathbf{e}_\alpha \mathbf{e}_\alpha^{\mathrm{T}}][c_\mu - 1] - [\mathbf{e}_\beta \mathbf{e}_\beta^{\mathrm{T}}][c_\mu + 1] \\ = \ & \begin{bmatrix} \mathbf{1}_{\alpha-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c_\mu & \mathbf{0} & s_\mu & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{\beta-\alpha-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & s_\mu^* & \mathbf{0} & -c_\mu & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}_{n-\beta} \end{bmatrix} \end{aligned} \tag{2.99}$$

die so genannten Givens-Matrizen sind mit geeignet gewählten α, β für ein bestimmtes μ und reelles c_μ. Die Zuordnung von α, β zu einem speziellem μ kann auf verschiedene Weisen erfolgen. Im reellen Fall können die Matrix-Elemente als $c_\mu = \cos(\varphi_\mu)$ und $s_\mu = \sin(\varphi_\mu)$ interpretiert werden. Die Givens-Matrizen können in reduzierter Form als

$$\begin{aligned} \hat{\boldsymbol{H}}_\mu = \ & \mathbf{e}_1 \mathbf{e}_2^{\mathrm{T}} s_\mu + \mathbf{e}_2 \mathbf{e}_1^{\mathrm{T}} s_\mu^* + [\mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} - \mathbf{e}_2 \mathbf{e}_2^{\mathrm{T}}] c_\mu \\ = \ & \begin{bmatrix} c_\mu & s_\mu \\ s_\mu^* & -c_\mu \end{bmatrix} = \mathbf{1}_2 - \boldsymbol{\gamma}_\mu \boldsymbol{\gamma}_\mu^{\mathrm{H}} \end{aligned} \tag{2.100}$$

geschrieben werden mit $\boldsymbol{\gamma}_\mu^{\mathrm{H}} \boldsymbol{\gamma}_\mu = 2$. Wegen der Unitarität der Matrix $\boldsymbol{S}$ ist die untere oder obere Dreiecksmatrix $\boldsymbol{R}$ eine Diagonalmatrix,

$$\boldsymbol{R} = \mathbf{diag}(\rho_1, \ldots, \rho_n) \ , \tag{2.101}$$

mit unimodularen Skalaren $\rho_1, \ldots, \rho_n$, [Frän97]. Wählen wir die gleiche Faktorisierung wie in [Frän97] auf Seite 55, erhalten wir die Wellen-Digital-Struktur für $n = 4$ im Bild 2.9. Ein mögliche Referenzschal-

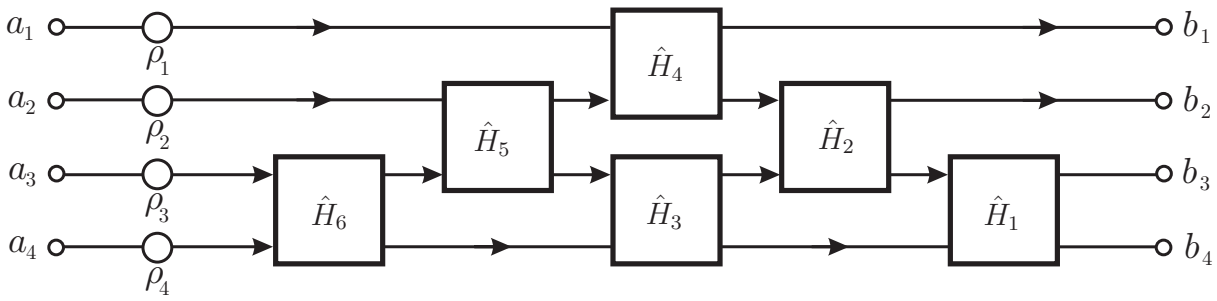


Bild 2.9: Realisierung eines energieneutralen M-Tors mit QR-Zerlegung nach Givens für $M = n = 4$

tung findet sich in [Voll04c].

Es folgen einige spezielle energieneutrale Bauelemente.

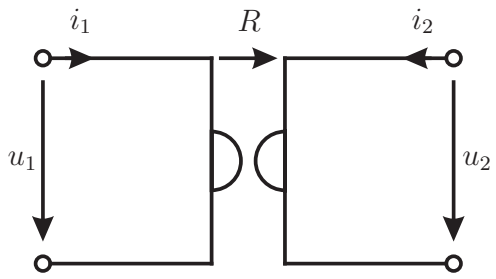


Bild 2.10: Idealer Gyrator

Idealer Gyrator

Der ideale Gyrator ist ein nichtreziprokes Bauelement, das durch die Gleichungen

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 & -R \\ R & 0 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{2.102}$$

bestimmt ist. Die Streumatrizen lauten

$$\boldsymbol{S} = \begin{bmatrix} \frac{R^2G_1G_2-1}{1+R^2G_1G_2} & \frac{-2R\sqrt{G_1G_2}}{1+R^2G_1G_2} \\ \frac{2R\sqrt{G_1G_2}}{1+R^2G_1G_2} & \frac{R^2G_1G_2-1}{1+R^2G_1G_2} \end{bmatrix} \quad \text{und} \quad \boldsymbol{S'} = \begin{bmatrix} \frac{R^2G_1G_2-1}{1+R^2G_1G_2} & \frac{-2RG_2}{1+R^2G_1G_2} \\ \frac{2RG_1}{1+R^2G_1G_2} & \frac{R^2G_1G_2-1}{1+R^2G_1G_2} \end{bmatrix}. \tag{2.103}$$

Wählt man die Torwiderstände beider Tore gleich der Gyrationskonstanten R, so berechnen sich die Streumatrizen zu

$$\boldsymbol{S} = \boldsymbol{S'} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{2.104}$$

Bei dieser Wahl der Torwiderstände hängt die Streumatrix von keinem Parameter ab. Das Bild 2.11 zeigt das Wellenflussdiagramm.

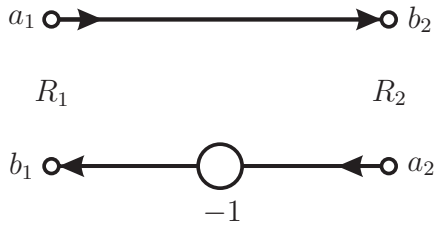


Bild 2.11: Wellenflussdiagramm eines idealen Gyrators für $R = R_1 = R_2$

n-Tor-Zirkulator

Ein weiteres nichtreziprokes Bauelement ist der Zirkulator, siehe Bild 2.12 a). Den Zirkulator werden wir über seine Streumatrix definieren, wobei wir annehmen, dass alle Torwiderstände gleich sind. Im Gegensatz zu [Bele68] definieren wir das Element S_{1n} zu $S_{1n} = 1$. Die gesamte Streumatrix lautet

$$\boldsymbol{S} = \boldsymbol{S'} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & & & & & & \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix}. \tag{2.105}$$

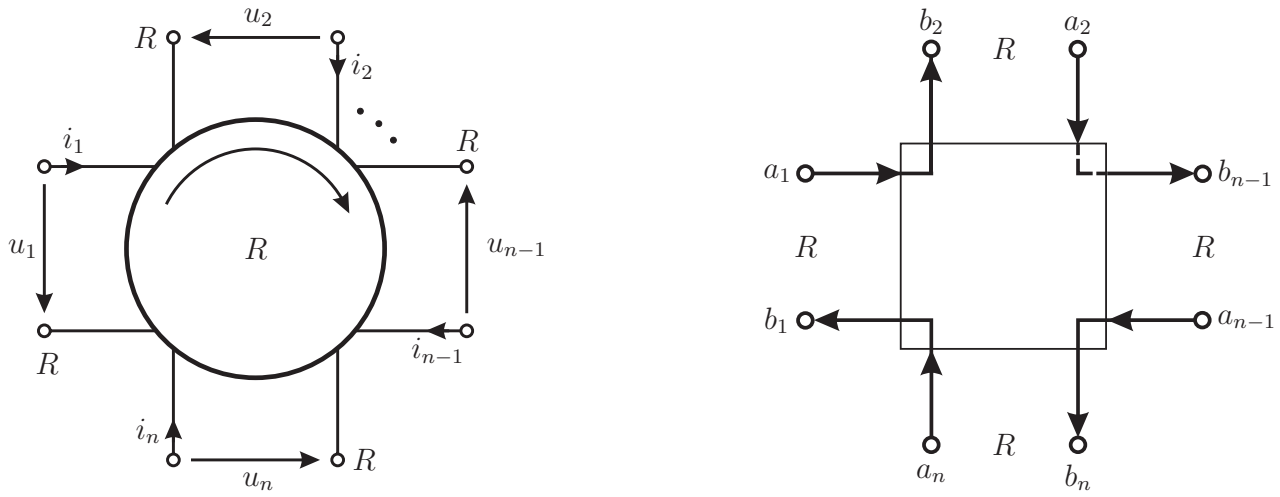


Bild 2.12: n-Tor Zirkulator: a) Kirchhoff'sche Schaltung b) Wellenflussdiagramm

Dass die Streumatrizen gleich sind, erkennt man an der Beziehung Gleichung (2.68) unter Berücksichtigung der Gleichheit der Torwiderstände. Bild 2.12 b) zeigt das Wellenflussdiagramm.

2.8.2 Verallgemeinerte Verbindungsnetze

Die Spannungen und Ströme verallgemeinerter Verbindungsnetze sind unabhängig voneinander Lösungen der homogenen Gleichungssysteme

$$\boldsymbol{A}\boldsymbol{i} = \boldsymbol{0} \quad \text{und} \quad \boldsymbol{B}\boldsymbol{u} = \boldsymbol{0}\,. \tag{2.106}$$

Die zeilenreguläre Matrix $\boldsymbol{A}$ hat die Dimension $r \times e$. $\boldsymbol{B}$ ist ebenfalls zeilenregulär und hat die Dimension $m \times e$. Weiterhin gilt $r + m = e$ und

$$\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{B}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{0}\,. \tag{2.107}$$

Die r (bzw. m) Spalten von $\boldsymbol{A}^{\mathrm{T}}$ ($\boldsymbol{B}^{\mathrm{T}}$) bilden eine vollständige Basis des Nullraumes von $\boldsymbol{B}$ ($\boldsymbol{A}$). Somit können durch $\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_r$ ($\boldsymbol{i} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{i}_m$) mit geeigneten Vektoren $\boldsymbol{u}_r$ und $\boldsymbol{i}_m$ sämtliche Lösungen des homogenen Gleichungssystems dargestellt werden [Fisc00a]. Aus der Orthogonalität von $\boldsymbol{u}$ und $\boldsymbol{i}$ folgt unmittelbar die Energieneutralität des verallgemeinerten Verbindungsnetzes. Ein verallgemeinertes Verbindungsnetz ist folglich ein spezielles energieneutrales n-Tor. Betrachten wir

$$\boldsymbol{b} = \boldsymbol{S}\boldsymbol{a} \iff \boldsymbol{G}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_r - \boldsymbol{R}^{1/2}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{i}_m = \boldsymbol{S}[\boldsymbol{G}^{1/2}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_r + \boldsymbol{R}^{1/2}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{i}_m] \tag{2.108}$$

so wird durch spezielle Wahl von $\boldsymbol{u}_r$ und $\boldsymbol{i}_m$ die Eigenschaft der Unitarität der Streumatrix durch

$$\boldsymbol{S}\boldsymbol{S} = \boldsymbol{1}_n \quad \text{und} \quad \boldsymbol{S}^{\mathrm{H}} = \boldsymbol{S} \tag{2.109}$$

weiter eingeschränkt, wobei die dritte der drei Eigenschaften aus den jeweils anderen beiden hervorgeht [Fränz97].

Die Streumatrizen eines verallgemeinerten Verbindungsnetzes berechnen sich zu

$$\boldsymbol{S} = 2\boldsymbol{G}^{1/2}\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^{\mathrm{T}}]^{-1}\boldsymbol{A}\boldsymbol{G}^{1/2} - \boldsymbol{1}_n = \boldsymbol{1}_n - 2\boldsymbol{R}^{1/2}\boldsymbol{B}^{\mathrm{T}}[\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}]^{-1}\boldsymbol{B}\boldsymbol{R}^{1/2} \tag{2.110}$$

$$\boldsymbol{S}' = 2\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^{\mathrm{T}}]^{-1}\boldsymbol{A}\boldsymbol{G} - \boldsymbol{1}_n = \boldsymbol{1}_n - 2\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}[\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}]^{-1}\boldsymbol{B} \tag{2.111}$$

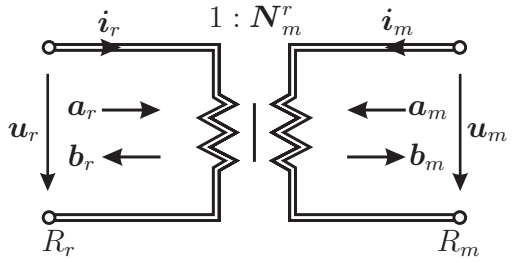


Bild 2.13: Idealer e-Tor Übertrager

vgl. [Shek74], [Meer79], [Frie95]. Aufgrund der Zeilenregularität von $\boldsymbol{A}$ ist $\boldsymbol{G}^{1/2}\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^{\mathrm{T}}]^{-1}$ offenbar die Moore-Penrose-Inverse von $\boldsymbol{A}\boldsymbol{G}^{1/2}$. Für den Fall gleicher Torwiderstände ergibt sich

$$\boldsymbol{S} = \boldsymbol{S}' = 2\boldsymbol{A}^{+}\boldsymbol{A} - \mathbf{1} = \mathbf{1} - 2\boldsymbol{B}^{+}\boldsymbol{B}\,. \tag{2.112}$$

Die Bestimmung der Matrix $\boldsymbol{A}$ aus der Streumatrix kann wie folgt durchgeführt werden, [Ochs02]. Bekanntermaßen können die zulässigen Zweigspannungen und Zweigströme unabhängig von einander gewählt werden. Mit der Wahl der offenbar zulässigen Zweigspannungen $\boldsymbol{u} = \mathbf{0}$ geht $\boldsymbol{b}' = \boldsymbol{S}'\boldsymbol{a}'$ in $-\boldsymbol{R}\,\boldsymbol{i} = \boldsymbol{S}'\,\boldsymbol{R}\,\boldsymbol{i}$ über. Die (ggf. noch zu normierende) Matrix $[\boldsymbol{S}' + \mathbf{1}]\boldsymbol{R}$, bzw. jede andere durch reguläre Transformation daraus erzeugbare Matrix, stellt eine gültige Matrix $\boldsymbol{A}$ dar. Bei reinen Verbindungsnetzen ist der maximale Rang der Matrix $\boldsymbol{A}$ durch die Anzahl Eigenwerte von $\boldsymbol{S}'$ mit dem Wert 1 festgelegt. Die Matrix $\boldsymbol{A}$ besitzt allerdings erst nach geeigneter regulärer Transformation die Eigenschaften einer Inzidenzmatrix, wie z.B. nur Elemente und Minoren der Form $\{-1, 0, 1\}$.

Ein Bauelement, welches den Eigenschaften von verallgemeinerten Verbindungsnetzen genügt, ist der ideale reelle e-Tor Übertrager, vgl. Bild 2.13. Der ideale Übertrager soll nach Definition durch die Matrizen

$$\begin{array}{lll} & \overset{r}{} \quad \overset{m}{} & \\ \boldsymbol{A} = & [\mathbf{1}_r \quad \boldsymbol{N}^{\mathrm{T}}] \;, & \boldsymbol{B} = \quad \overset{r}{} \quad \overset{m}{} \\ & & [-\boldsymbol{N} \quad \mathbf{1}_m] \end{array} \tag{2.113}$$

beschrieben werden. Wir nehmen die folgenden Partitionierungen vor

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_r \\ \boldsymbol{u}_m \end{bmatrix}\,,\; \boldsymbol{i} = \begin{bmatrix} \boldsymbol{i}_r \\ \boldsymbol{i}_m \end{bmatrix}\,,\; \boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_r \\ \boldsymbol{b}_m \end{bmatrix}\,,\; \boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_r \\ \boldsymbol{a}_m \end{bmatrix}\,,\; \boldsymbol{G} = \mathbf{diag}(\boldsymbol{G}_r, \boldsymbol{G}_m)\,. \tag{2.114}$$

Auswerten von Gleichung (2.110) liefert die Streumatrix

$$\boldsymbol{S} = \left[\begin{array}{l|l} 2\boldsymbol{G}_r^{1/2}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{G}_r^{1/2} - \mathbf{1}_r & 2\boldsymbol{G}_r^{1/2}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m^{1/2} \\ \hline 2\boldsymbol{G}_m^{1/2}\boldsymbol{N}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{G}_r^{1/2} & 2\boldsymbol{G}_m^{1/2}\boldsymbol{N}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m^{1/2} - \mathbf{1}_m \end{array}\right]\,. \tag{2.115}$$

Im weiteren Verlauf der Arbeit werden wir den idealen Übertrager benötigen, dessen zugehörige Streumatrix die Form

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{0} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \mathbf{0} \end{bmatrix} \tag{2.116}$$

besitzt. Hierbei setzen wir $\boldsymbol{N}$ als quadratisch und regulär voraus. Aus Gleichung (2.115) erhalten wir die simultan zu erfüllenden Bedingungen

$$\begin{array}{lll} 2\boldsymbol{G}_r^{1/2}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{G}_r^{1/2} = \mathbf{1}_r & \Longleftrightarrow & 2\boldsymbol{G}_r = \boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N} \\ 2\boldsymbol{G}_m^{1/2}\boldsymbol{N}[\boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m^{1/2} = \mathbf{1}_m & \Longleftrightarrow & 2\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N} = \boldsymbol{G}_r + \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}\,. \end{array} \tag{2.117}$$

Diese Bedingungen zerfallen -offenbar aufgrund der Eigenschaften von $\boldsymbol{S}$- zu der einen Bedingung

$$\boldsymbol{G}_r = \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N} \Longleftrightarrow \boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}} = \boldsymbol{N}^{-1}\boldsymbol{R}_m \Longleftrightarrow \boldsymbol{N}\boldsymbol{R}_r = \boldsymbol{R}_m\boldsymbol{N}^{-\mathrm{T}} \; . \tag{2.118}$$

Die zugehörige Spannungswellenstreumatrix wollen wir auf einem alternativen Weg herleiten. Offenbar ist $\boldsymbol{u}$ ($\boldsymbol{i}^{\mathrm{H}}$) Rechtseigenvektor (Linkseigenvektor) zum Eigenwert 1 (-1) der Matrix

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} , \tag{2.119}$$

d. h. es gilt

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} \boldsymbol{u} \quad , \quad \boldsymbol{i} = - \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{\mathrm{T}} \\ \boldsymbol{N}^{-\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \boldsymbol{i} \; . \tag{2.120}$$

Somit ergibt sich

$$\boldsymbol{b}' = \boldsymbol{u} - \boldsymbol{R}\boldsymbol{i} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} \boldsymbol{u} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}} \\ \boldsymbol{R}_m\boldsymbol{N}^{-\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \boldsymbol{i} \; . \tag{2.121}$$

Mit Gleichung (2.118) erhalten wir

$$\boldsymbol{b}' = \boldsymbol{u} - \boldsymbol{R}\boldsymbol{i} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} [\boldsymbol{u} + \boldsymbol{R}\boldsymbol{i}] = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} \boldsymbol{a}' \tag{2.122}$$

und somit die gewünschte Spannungswellenstreumatrix

$$\boldsymbol{S}' = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \boldsymbol{0} \end{bmatrix} \; . \tag{2.123}$$

n-Tor-Paralleladaptor

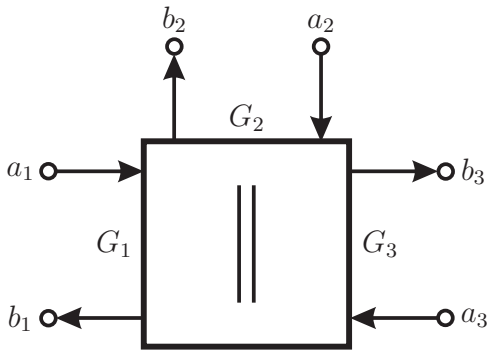


Bild 2.14: 3-Tor-Paralleladaptor

Der n-Tor-Paralleladaptor ist durch die Gleichheit seiner n Torspannungen und durch die Tatsache, dass die n Torströme eine Nullsumme bilden, gekennzeichnet, d. h.

$$u = u_\nu \, , \, \nu = 1, 2, \, \ldots \, , n \quad \wedge \quad \sum_{\nu=1}^{n} i_\nu = 0. \tag{2.124}$$

Die Torwiderstände $R_\nu = 1/G_\nu$ des Adaptors bestimmen das Verhalten des Wellendigitalfilter-Bausteins. Drückt man nun die $2n$ Variablen in den oben angegebenen n linear unabhängigen Kirchhoff'schen Gleichungen durch die $2n$ Wellengrößen aus und führt die Hilfsvariablen

$$G_0 = \sum_{\nu=1}^{n} G_\nu\,, \quad \gamma_\nu = \sqrt{\frac{2G_\nu}{G_0}} \quad \text{und} \quad \gamma'_\nu = \gamma_\nu^2 \tag{2.125}$$

ein, so erhalten wir durch Anwendung von Gleichung (2.110) mit $\boldsymbol{A} = [1\ \ 1 \cdots 1\ \ 1]$ die Streumatrizen des Paralleladaptors

$$\boldsymbol{S} = \begin{bmatrix} \gamma_1^2-1 & \gamma_1\gamma_2 & \dots & \gamma_1\gamma_n \\ \gamma_1\gamma_2 & \gamma_2^2-1 & & \gamma_2\gamma_n \\ \vdots & & \ddots & \vdots \\ \gamma_1\gamma_n & \gamma_2\gamma_n & \dots & \gamma_n^2-1 \end{bmatrix} \text{ und } \boldsymbol{S'} = \begin{bmatrix} \gamma'_1-1 & \gamma'_2 & \dots & \gamma'_n \\ \gamma'_1 & \gamma'_2-1 & & \gamma'_n \\ \vdots & & \ddots & \vdots \\ \gamma'_1 & \gamma'_2 & \dots & \gamma'_n-1 \end{bmatrix}. \tag{2.126}$$

Werden die Koeffizienten γ_ν und γ'_ν in den Vektoren $\boldsymbol{\gamma} = [\gamma_1\ \gamma_2\ \dots\ \gamma_n]^{\mathrm{T}}$ und $\boldsymbol{\gamma'} = [\gamma'_1\ \gamma'_2\ \dots\ \gamma'_n]^{\mathrm{T}}$ zusammengefasst, so kann man die Streumatrizen kompakt durch

$$\boldsymbol{S} = \boldsymbol{\gamma}\,\boldsymbol{\gamma}^{\mathrm{T}} - \mathbf{1}_n \ \text{ und } \ \boldsymbol{S'} = \mathbf{e}\,\boldsymbol{\gamma'}^{\mathrm{T}} - \mathbf{1}_n \tag{2.127}$$

beschreiben.

Berücksichtigt man, dass $\|\boldsymbol{\gamma}\|^2 = 2$ gilt, so erkennen wir wieder, dass $\boldsymbol{S}$ eine Householdermatrix ist. Der nichtgebundene Paralleladaptor besitzt zudem die Eigenschaft, dass alle Torwiderstände unabhängig voneinander vorgebbar sind, wobei wir uns auf positive, endliche Torwiderstände beschränken wollen. Das zugehörige Wellendigitalfiltersymbol ist für $n = 3$ im Bild 2.14 dargestellt.

Gebundener n-Tor-Paralleladaptor

Der gebundene n-Tor-Adaptor ist nur ein Spezialfall des allgemeinen Leistungswellen-Paralleladaptors, der aber eine so herausragende Bedeutung besitzt, dass er als ein eigenes Wellendigitalfilter-Bauelement aufgefasst wird. Das Besondere ist, dass ein Tor des Adaptors reflexionsfrei ist und sich somit zum Aufbrechen von verzögerungsfreien gerichteten Schleifen eignet. Wir nehmen nun an, dass das reflexionsfreie Tor die Nummer ν hat. Folglich gilt für das Diagonalelement $s_{\nu\nu}$ und für den Adaptorkoeffizienten γ_ν

$$s_{\nu\nu} = 0 \Leftrightarrow \gamma_\nu^2 = 1. \tag{2.128}$$

Mit der Berechnungsvorschrift der Adaptorkoeffizienten

$$\gamma_\nu^2 = 2\frac{G_\nu}{G_0} \quad \text{folgt} \quad G_0 = 2G_\nu. \tag{2.129}$$

Aus

$$G_0 = G_\nu + \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} G_\mu \quad \text{erhalten wir} \quad G_\nu = \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} G_\mu. \tag{2.130}$$

Die Hilfsgröße G_0, die Summe aller Torleitwerte, kann somit durch die Torleitwerte der nicht reflexionsfreien Tore berechnet werden, d.h.

$$G_0 = 2\sum_{\substack{\mu=1\\ \mu\neq\nu}}^{n} G_\mu. \tag{2.131}$$

Die Adaptorkoeffizienten ergeben sich zu

$$\gamma_\mu = \begin{cases} \sqrt{2G_\mu/G_0} & \text{für} \quad \mu \neq \nu \\ 1 & \text{für} \quad \mu = \nu. \end{cases} \quad \text{und} \quad \gamma'_\mu = \gamma_\mu^2 \,. \tag{2.132}$$

Von $\gamma_\nu^2 = 1$ wird nur die Lösung mit dem positive Vorzeichen akzeptiert, da γ_μ über Gleichung (2.125) als positiv definiert ist. Die Streumatrizen lauten

$$\boldsymbol{S} = \begin{bmatrix} \gamma_1^2-1 & \dots & \gamma_1 & \dots & \gamma_1\gamma_n \\ \vdots & \ddots & & & \vdots \\ \gamma_1 & \dots & 0 & \dots & \gamma_n \\ \vdots & & & \ddots & \vdots \\ \gamma_1\gamma_n & \dots & \gamma_n & \dots & \gamma_n^2-1 \end{bmatrix} \quad \text{und} \quad \boldsymbol{S}' = \begin{bmatrix} \gamma'_1-1 & \dots & 1 & \dots & \gamma'_n \\ \vdots & \ddots & & & \vdots \\ \gamma'_1 & \dots & 0 & \dots & \gamma'_n \\ \vdots & & & \ddots & \vdots \\ \gamma'_1 & \dots & 1 & \dots & \gamma'_n-1 \end{bmatrix}. \tag{2.133}$$

Das zugehörige Wellendigitalfiltersymbol ist für $n = 3$ im Bild 2.15 dargestellt, wobei der Übersicht halber das reflexionsfreie Tor (hier Tor 3) besonders gekennzeichnet ist, damit auf einem Plan mit Wellendigitalfilterelementen die reflexionsfreien Tore auch ohne Kenntnis der Werte der Torwiderstände sofort erkennbar sind.

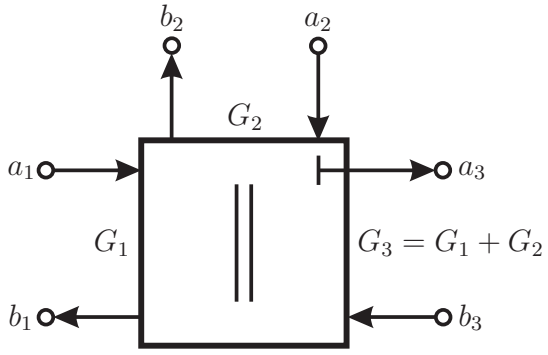


Bild 2.15: Gebundener 3-Tor-Paralleladaptor (hier: Tor 3 reflexionsfrei)

Zum Abschluss wollen wir noch den Fall einer positiv definiten, symmetrischen Torwiderstandsmatrix untersuchen. Dazu definieren wir $\bar{\boldsymbol{G}} = \boldsymbol{G}^{1/2}$ und $\boldsymbol{\beta} = [\beta_1\,,\,\cdots\,,\,\beta_n]^{\mathrm{T}} = \bar{\boldsymbol{G}}\boldsymbol{A}^{\mathrm{T}} = \bar{\boldsymbol{G}}[1\,,\,\cdots\,,\,1]^{\mathrm{T}}$ mit $\beta_l = \sum_{k=1}^{n} \bar{g}_{lk}$. Nach Gleichung (2.110) lautet die Streumatrix dann

$$\boldsymbol{S} = 2\boldsymbol{\beta}[\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}]^{-1}\boldsymbol{\beta}^{\mathrm{T}} - \mathbf{1}_n \,. \tag{2.134}$$

Die Forderung nach Reflexionsfreiheit am Tor ν, d. h. $s_{\nu\nu} = 0$ liefert

$$2\beta_\nu^2 = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta} = \sum_{\mu=1}^{n} \beta_\mu^2 \Longleftrightarrow \beta_\nu^2 = \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} \beta_\mu^2 \; . \tag{2.135}$$

Die Frage, die sich nun aufdrängt, ist, ob sich mehr als ein Tor reflexionsfrei wählen lässt. Wir beantworten die Frage für einen später noch relevanten Spezialfall. Und zwar nehmen wir $n = 4$ an und fordern, dass die ersten beiden Tore eine diagonale Torwiderstandsmatrix besitzen. Der Vektor $\boldsymbol{\beta}$ lautet dann

$$\boldsymbol{\beta} = [\bar{g}_{11} \, , \, \bar{g}_{22} \, , \, \bar{g}_{33}+\bar{g}_{34} \, , \, \bar{g}_{43}+\bar{g}_{44}]^{\mathrm{T}} \; . \tag{2.136}$$

Die Forderungen $s_{11} = 0$ und $s_{22} = 0$ gehen in die simultan zu erfüllenden Beziehungen

$$\bar{g}_{11}^2 = \bar{g}_{22}^2 + \beta_3^2 + \beta_4^2 \quad \text{und} \quad \bar{g}_{22}^2 = \bar{g}_{11}^2 + \beta_3^2 + \beta_4^2 \tag{2.137}$$

über, die letztendlich in

$$\beta_3^2 + \beta_4^2 = 0 \iff (\bar{g}_{33}+\bar{g}_{43})^2 + (\bar{g}_{34}+\bar{g}_{44})^2 = 0 \tag{2.138}$$

münden. Eine Forderung, die aufgrund der Reellwertigkeit von $\boldsymbol{G}^{1/2}$ nur durch $\bar{g}_{33} = -\bar{g}_{43}$ und $\bar{g}_{34} = -\bar{g}_{44}$ erfüllt werden kann. Diese einzige Lösung bedeutet aber eine lineare Abhängigkeit der Zeilenvektoren von $\boldsymbol{G}^{1/2}$, was im Widerspruch zur vorausgesetzten Regularität von $\boldsymbol{G}$ steht. In dem hier betrachteten Fall existieren somit keine 2 reflexionsfreien Tore.

n-Tor-Serienadaptor

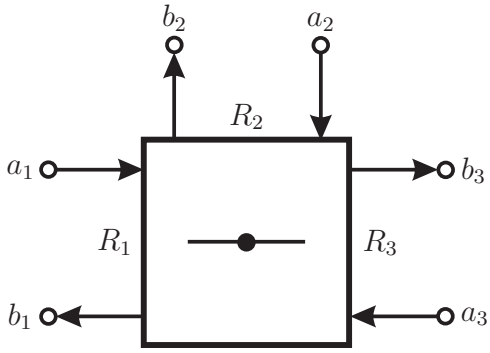


Bild 2.16: 3-Tor-Serienadaptor

Der n-Tor-Serienadaptor ist durch die Gleichheit seiner n Torströme und durch die Tatsache, dass die n Torspannungen eine Nullsumme bilden, gekennzeichnet, d. h.

$$i = i_\nu \, , \, \nu = 1, 2, \, \ldots \, , n \quad \wedge \quad \sum_{\nu=1}^{n} u_\nu = 0. \tag{2.139}$$

Drückt man nun wiederum die $2n$ Variablen in den oben angegebenen n linear unabhängigen Kirchhoff'schen Gleichungen durch die $2n$ Wellengrößen aus und führt die Hilfsvariablen

$$R_0 = \sum_{\nu=1}^{n} R_\nu \, , \quad \gamma_\nu = \sqrt{\frac{2R_\nu}{R_0}} \quad \text{und} \quad \gamma_\nu' = \gamma_\nu^2 \tag{2.140}$$

ein, so erhalten wir durch Anwendung von Gleichung (2.110) die Streumatrizen des Serienadaptors zu

$$\boldsymbol{S} = \begin{bmatrix} 1-\gamma_1^2 & -\gamma_1\gamma_2 & \dots & -\gamma_1\gamma_n \\ -\gamma_1\gamma_2 & 1-\gamma_2^2 & & -\gamma_2\gamma_n \\ \vdots & & \ddots & \vdots \\ -\gamma_1\gamma_n & -\gamma_2\gamma_n & \dots & 1-\gamma_n^2 \end{bmatrix} \quad \text{und} \quad \boldsymbol{S}' = \begin{bmatrix} 1-\gamma_1' & -\gamma_1' & \dots & -\gamma_1' \\ -\gamma_2' & 1-\gamma_2' & & -\gamma_2' \\ \vdots & & \ddots & \vdots \\ -\gamma_n' & -\gamma_n' & \dots & 1-\gamma_n' \end{bmatrix}, \tag{2.141}$$

bzw.

$$\boldsymbol{S} = \mathbf{1}_n - \boldsymbol{\gamma}\,\boldsymbol{\gamma}^{\mathrm{T}} \text{ und } \boldsymbol{S}' = \mathbf{1}_n - \boldsymbol{\gamma}'\,\mathbf{e}^{\mathrm{T}}\,. \tag{2.142}$$

Das zugehörige Wellendigitalfiltersymbol findet sich für $n = 3$ im Bild 2.16.

Gebundener n-Tor-Serienadaptor

Wir nehmen für den gebundenen n-Tor-Serienadaptor auch an, dass sein reflexionsfreies Tor die Nummer ν hat. Folglich gilt für das Diagonalelement $s_{\nu\nu}$ und für den Adaptorkoeffizienten γ_ν

$$s_{\nu\nu} = 0 \Leftrightarrow \gamma_\nu^2 = 1. \tag{2.143}$$

Mit der Berechnungsvorschrift der Adaptorkoeffizienten

$$\gamma_\mu^2 = 2\frac{R_\mu}{R_0} \quad \text{folgt} \quad R_0 = 2R_\nu. \tag{2.144}$$

Aus

$$R_0 = R_\nu + \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} R_\mu \quad \text{erhalten wir} \quad R_\nu = \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} R_\mu. \tag{2.145}$$

Die Hilfsgröße R_0, die Summe aller Torwiderstände, kann somit durch die Torwiderstände der nicht reflexionsfreien Tore berechnet werden, d.h.

$$R_0 = 2\sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} R_\mu. \tag{2.146}$$

Die Adaptorkoeffizienten ergeben sich zu

$$\gamma_\mu = \begin{cases} \sqrt{2R_\mu/R_0} & \text{für} \quad \mu \neq \nu \\ 1 & \text{für} \quad \mu = \nu. \end{cases}, \; \gamma_\mu' = \gamma_\mu^2\,. \tag{2.147}$$

Die negative Lösung für γ_μ entfällt aus den selben Gründen wie im Falle des gebundenen Paralleladaptors. Die Streumatrizen lauten

$$\boldsymbol{S} = \begin{bmatrix} 1-\gamma_1^2 & \dots & -\gamma_1 & \dots & -\gamma_1\gamma_n \\ \vdots & \ddots & & & \vdots \\ -\gamma_1 & \dots & 0 & \dots & -\gamma_n \\ \vdots & & & \ddots & \vdots \\ -\gamma_1\gamma_n & \dots & -\gamma_n & \dots & 1-\gamma_n^2 \end{bmatrix} \quad \text{und} \quad \boldsymbol{S}' = \begin{bmatrix} 1-\gamma_1' & \dots & -\gamma_1' & \dots & -\gamma_1' \\ \vdots & \ddots & & & \vdots \\ -1 & \dots & 0 & \dots & -1 \\ \vdots & & & \ddots & \vdots \\ -\gamma_n' & \dots & -\gamma_n' & \dots & 1-\gamma_n' \end{bmatrix}. \tag{2.148}$$

Das zugehörige Wellendigitalfiltersymbol findet sich für $n = 3$ im Bild 2.17. Zu bemerken ist auch hier, dass der Torwiderstand R_ν nicht in die Streumatrix eingeht.

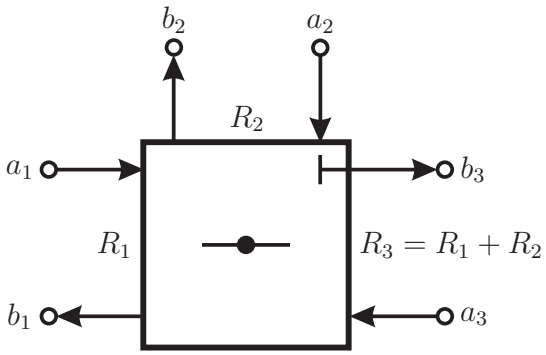


Bild 2.17: Gebundener-3-Tor-Serienadaptor (hier: Tor 3 reflexionsfrei)

2-Tor-Übertrager

Der 2-Tor-Übertrager wird durch

$$u_2 = nu_1 \quad \text{und} \quad i_1 = -ni_2 \tag{2.149}$$

definiert (vgl. Bild 2.18). Die Streumatrizen lauten

$$\boldsymbol{S} = \begin{bmatrix} \frac{R_2-n^2R_1}{R_2+n^2R_1} & \frac{2n\sqrt{R_1R_2}}{R_2+n^2R_1} \\ \frac{2n\sqrt{R_1R_2}}{R_2+n^2R_1} & \frac{n^2R_1-R_2}{R_2+n^2R_1} \end{bmatrix}, \quad \boldsymbol{S'} = \begin{bmatrix} \frac{R_2-n^2R_1}{R_2+n^2R_1} & \frac{2nR_1}{R_2+n^2R_1} \\ \frac{2nR_2}{R_2+n^2R_1} & \frac{n^2R_1-R_2}{R_2+n^2R_1} \end{bmatrix}. \tag{2.150}$$

Wählen wir das Übersetzungsverhältnis zu $n = \sqrt{R_2/R_1}$, so vereinfachen sich die Streumatrizen zu

$$\boldsymbol{S} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{und} \quad \boldsymbol{S'} = \begin{bmatrix} 0 & \frac{1}{n} \\ n & 0 \end{bmatrix}. \tag{2.151}$$

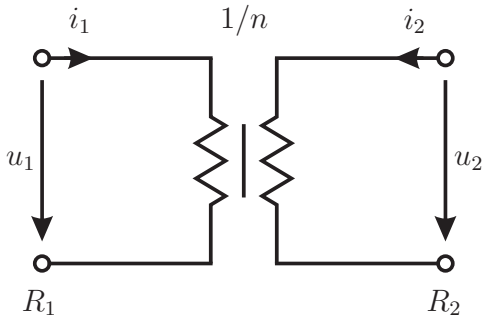


Bild 2.18: 2-Tor-Übertrager

Jaumann-Adaptor

Der im Bild 2.19 dargestellte Jaumann-Adaptor stellt ein verallgemeinertes Verbindungsnetz gemäß Gleichung (2.113) mit

$$\boldsymbol{N} = \frac{1}{1+n}\begin{bmatrix} 1 & -1 \\ n & 1 \end{bmatrix} \quad , \quad \boldsymbol{N}^{-1} = \begin{bmatrix} 1 & 1 \\ -n & 1 \end{bmatrix} \tag{2.152}$$

dar. Der freie Parameter ist das Übersetzungsverhältnis des Übertragers. Wir nehmen an, es gilt $n \neq -1$, d.h. wir schließen den Fall aus, dass der Übertrager eine reine Parallelverbindung ist. $\boldsymbol{N}$ ist somit regulär. Die Streumatrix des Jaumann-Adaptors sollte eine möglichst einfache Struktur besitzen. Aus diesem

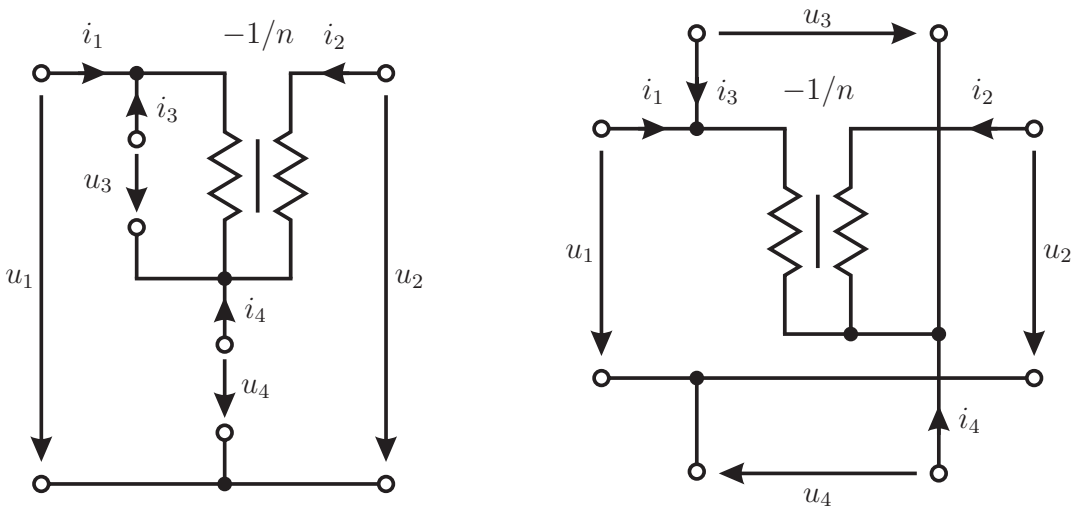


Bild 2.19: Jaumann-Adaptor

Grund wählen wir die Torwiderstandsmatrix gemäß Gleichung (2.118) zu

$$\boldsymbol{R} = R_1\,\mathbf{diag}\left(1, n, \frac{1}{n+1}, \frac{n}{n+1}\right). \tag{2.153}$$

Durch Anwendung von Gleichung (2.123) lässt sich die Spannungswellen-Streumatrix zu

$$\boldsymbol{S}' = \begin{bmatrix} \mathbf{0}_2 & \boldsymbol{N}^{-1} \\ \boldsymbol{N} & \mathbf{0}_2 \end{bmatrix} \tag{2.154}$$

ablesen. Mit Gleichung (2.68) ermitteln wir die Leistungswellen-Streumatrix zu

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{0} & \boldsymbol{S_1} \\ \boldsymbol{S_1}^{\mathrm{T}} & \mathbf{0} \end{bmatrix}, \tag{2.155}$$

wobei sich die Untermatrix $\boldsymbol{S_1}$ zu

$$\boldsymbol{S_1} = \begin{bmatrix} \sqrt{\frac{1}{n+1}} & \sqrt{\frac{n}{n+1}} \\ -\sqrt{\frac{n}{n+1}} & \sqrt{\frac{1}{n+1}} \end{bmatrix} \tag{2.156}$$

ergibt. Im Bild 2.20 ist das in dieser Arbeit verwendete WDF-Symbol des Jaumann-Adaptors dargestellt. Die dickere Kante zwischen den Toren 2 und 3 der Raute trägt der Tatsache Rechnung, dass nur die Matrixelemente s_{23} und s_{32} negativ sind.

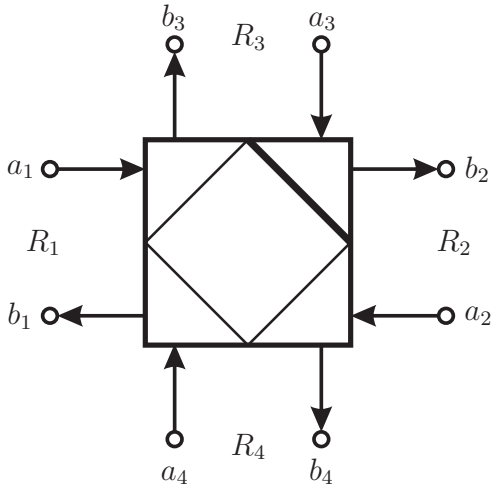


Bild 2.20: Wellendigital-Realisierung des Jaumann-Adaptors

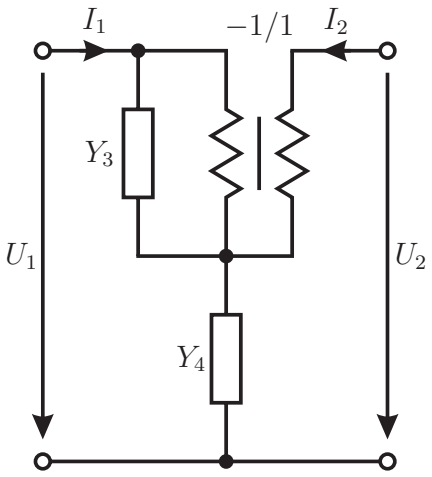


Bild 2.21: Beschalteter Jaumann-Adaptor

In der Regel nutzt man einen Jaumann-Adaptor in Verbindung mit zwei Impedanzen Z_3 und Z_4, mit denen die Tore 3 und 4 abgeschlossen werden (Bild 2.21), d. h. es gilt

$$\begin{bmatrix} U_3 \\ \\ U_4 \end{bmatrix} = \begin{bmatrix} -Z_3 & 0 \\ \\ 0 & -Z_4 \end{bmatrix} \begin{bmatrix} I_3 \\ \\ I_4 \end{bmatrix} . \tag{2.157}$$

Durch Verwendung von Gleichung (2.152) i.V.m. Gleichung (2.120) erhält man zunächst

$$\begin{bmatrix} U_1 \\ \\ U_2 \end{bmatrix} = \boldsymbol{N}^{-1} \begin{bmatrix} -Z_3 & 0 \\ \\ 0 & -Z_4 \end{bmatrix} (-\boldsymbol{N}^{-\mathrm{T}}) \begin{bmatrix} I_1 \\ \\ I_2 \end{bmatrix} , \tag{2.158}$$

woraus man die Impedanzmatrix bezogen auf die Tore 1 und 2 zu

$$\boldsymbol{Z} = \begin{bmatrix} Z_4+Z_3 & Z_4-nZ_3 \\ \\ Z_4-nZ_3 & Z_4+n^2Z_3 \end{bmatrix} \tag{2.159}$$

ermittelt. Zur Berechnung der Admittanzmatrix schreibt man zunächst

$$\begin{bmatrix} I_1 \\ \\ I_2 \end{bmatrix} = -\boldsymbol{N}^{\mathrm{T}} \begin{bmatrix} -\frac{1}{Z_3} & 0 \\ \\ 0 & -\frac{1}{Z_4} \end{bmatrix} \boldsymbol{N} \begin{bmatrix} U_1 \\ \\ U_2 \end{bmatrix} \tag{2.160}$$

und erhält die Admittanzmatrix bezogen auf die Tore 1 und 2 zu

$$\boldsymbol{Y} = \frac{1}{(1+n)^2} \begin{bmatrix} \frac{n^2}{Z_4}+\frac{1}{Z_3} & \frac{n}{Z_4}-\frac{1}{Z_3} \\ \\ \frac{n}{Z_4}-\frac{1}{Z_3} & \frac{1}{Z_4}+\frac{1}{Z_3} \end{bmatrix} . \tag{2.161}$$

2.8.3 Quellen

Wir betrachten die resistive Spannungsquelle im Bild 2.22 a) mit Innenwiderstand R_i, Quellenspannung e und Ausgangsspannung u_q. Wir ersetzen in der Maschengleichung $u_q = e + R_i i_q$ die auftretenden Spannungen und Ströme durch die Wellengrößen und erhalten

$$a'_q + b'_q = 2e + R_i G_q [a'_q - b'_q] \Longleftrightarrow b'_q = \frac{2e}{1 + R_i G_q} + \rho\, a'_q \quad \text{mit} \quad \rho = \frac{R_i - R_q}{R_i + R_q} . \tag{2.162}$$

Das zugehörige Wellenflussdiagramm zeigt Bild 2.22 b). Das Wellenflussdiagramm vereinfacht sich, wenn wir den Torwiderstand gleich dem Innenwiderstand wählen $R_q = R_i$. Die ausfallende Welle ist nun unabhängig von der einfallenden Welle $b'_q = 2\sqrt{R_q} b_q = e$, vgl. Bild 2.22 c). Die ideale Spannungsquelle ($R_i = 0$) wird in den Wellengrößen durch $b'_q = 2e - a'_q$ beschrieben.

Analoge Ergebnisse erhalten wir bei der Untersuchung von Stromquellen. Für die im Bild 2.23 a) dargestellte resistive Stromquelle lautet die Knotengleichung $j + i_q = G_i u_q$. Wir erhalten

$$2j + G_q [a'_q - b'_q] = G_i [a'_q + b'_q] \Longleftrightarrow b'_q = \frac{2j}{G_i + G_q} + \rho\, a'_q \quad \text{mit} \quad \rho = \frac{R_i - R_q}{R_i + R_q} . \tag{2.163}$$

Als Wellendigitalfiltersymbol verwenden wir das gleiche Symbol wie bei den Spannungswellen, nur, dass in diesem Fall sich die Quellenwelle anders berechnet, vgl. Bild 2.23 b) und c).

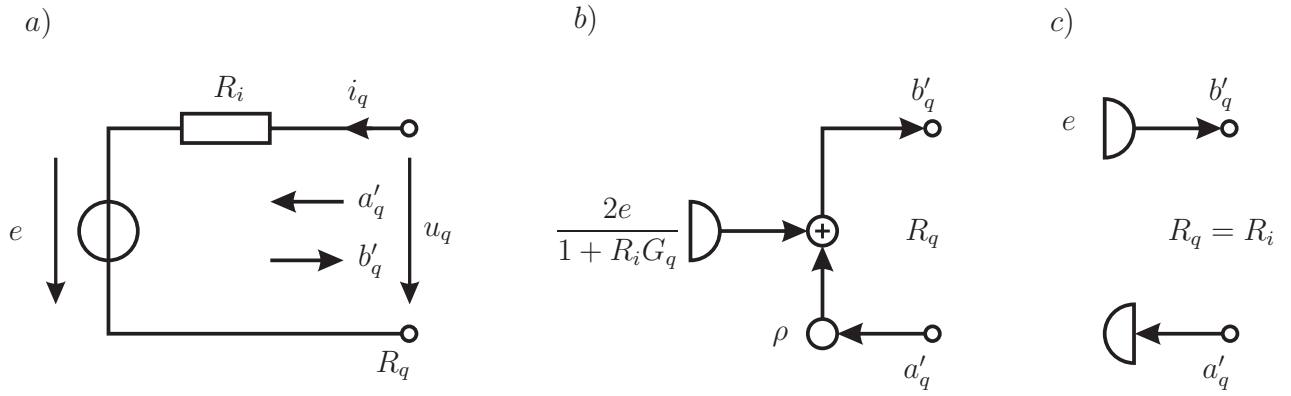


Bild 2.22: Resistive Spannungsquelle a) Kirchhoff'sche Schaltung b) und c) Wellenflussdiagramm

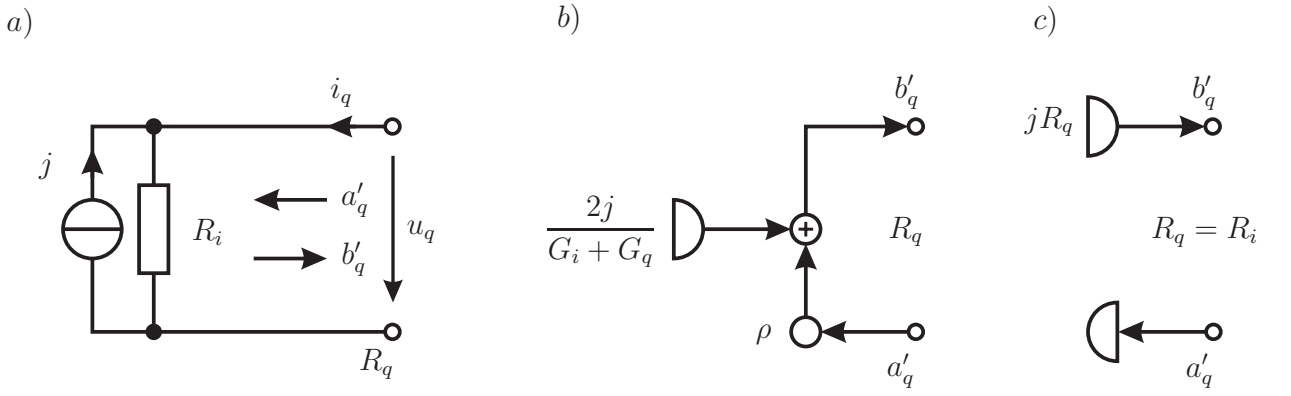


Bild 2.23: Resistive Stromquelle a) Kirchhoff'sche Schaltung b) und c) Wellenflussdiagramm

Neben den idealen Quellen kann eine Vielzahl weiterer Quellen betrachtet werden. Von besonders praktischer Relevanz sind jedoch ideale Quellen, die in einem verallgemeinerten Verbindungsnetz eingebettet sind. Um dies zu untersuchen, betrachten wir den mit einer vektoriellen idealen Stromquelle abgeschlossenen idealen e-Tor-Übertrager im Bild 2.24. Im Folgenden ist die Beschreibung durch die Wellengrößen gesucht. Der Übertrager soll Gleichung (2.113) mit den Bezeichnungen aus Gleichung (2.114) folgen. Wir nehmen $m \geq r$ an, und $\boldsymbol{N}^{\mathrm{T}}$ sei zeilenregulär. Obwohl die Ströme $\boldsymbol{i}_m$ unabhängig sind, wollen wir die Stromquellen in die r-Zweige legen. Der Vektor $\boldsymbol{j} = \boldsymbol{i}_r$ liegt dann im Spaltenraum von $\boldsymbol{N}^{\mathrm{T}}$. Gesucht ist nun eine Beziehung der Form

$$\boldsymbol{b'}_m = \boldsymbol{S'a'}_m + \boldsymbol{Q'j}\,. \tag{2.164}$$

Wir multiplizieren dazu die Gleichung $\boldsymbol{a'}_m = \boldsymbol{u}_m + \boldsymbol{R}_m\boldsymbol{i}_m$ von links mit $\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m$, drücken $\boldsymbol{u}_m$, $\boldsymbol{i}_m$ durch

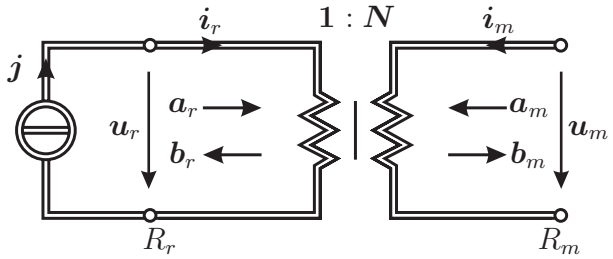


Bild 2.24: e-Tor Übertrager mit idealen Stromquellen

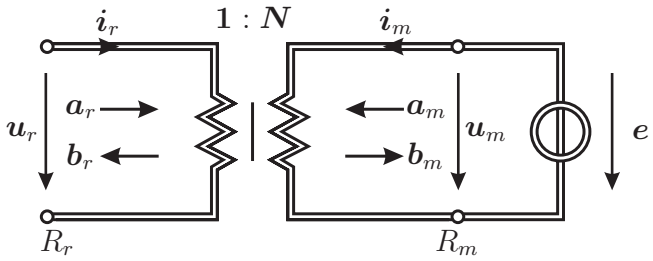


Bild 2.25: e-Tor Übertrager mit idealen Spannungsquellen

$\boldsymbol{u}_r$, $\boldsymbol{i}_r$ aus und lösen anschließend nach $\boldsymbol{u}_r$ auf

$$\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{a'}_m = \boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}\boldsymbol{u}_r - \boldsymbol{i}_r \iff \boldsymbol{u}_r = [\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{a'}_m + \boldsymbol{i}_r]\,. \tag{2.165}$$

Dieses setzen wir in $\boldsymbol{b'}_m = 2\boldsymbol{u}_m - \boldsymbol{a'}_m = 2\boldsymbol{N}\boldsymbol{u}_r - \boldsymbol{a'}_m$ ein und erhalten

$$\boldsymbol{b'}_m = \underbrace{[2\boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m - \boldsymbol{1}_m]}_{\boldsymbol{S'}}\boldsymbol{a'}_m + \underbrace{2\boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}}_{\boldsymbol{Q'}}\boldsymbol{j}\,. \tag{2.166}$$

Wir wollen nun die Freiheitsgrade diskutieren. Die Spannungen $\boldsymbol{u}_m$ müssen im Spaltenraum von $\boldsymbol{N}$ liegen, d.h. r Spannungen sind frei vorgebbar. Die Ströme $\boldsymbol{i}_m$ müssen $\boldsymbol{N}^{\mathrm{T}}\,\boldsymbol{i}_m = -\boldsymbol{j}$ bei gegebener Quellenverteilung genügen, d.h.

$$\begin{aligned} \boldsymbol{i}_m &= -\boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}]^{-1}\boldsymbol{j} + [\boldsymbol{1}_m - \boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}]\boldsymbol{z} = -(\boldsymbol{N}^{\mathrm{T}})^{+}\boldsymbol{j} + [\boldsymbol{1}_m - (\boldsymbol{N}^{\mathrm{T}})^{+}\boldsymbol{N}^{\mathrm{T}}]\boldsymbol{z} \\ &= -(\boldsymbol{N}^{\mathrm{T}})^{+}\boldsymbol{j} + [\boldsymbol{1}_m - \boldsymbol{N}\boldsymbol{N}^{+}]^{\mathrm{T}}\boldsymbol{z} = -(\boldsymbol{N}^{\mathrm{T}})^{+}\boldsymbol{j} + [\boldsymbol{1}_m - \boldsymbol{N}\boldsymbol{N}^{+}]\boldsymbol{z}\,,\ \boldsymbol{z} \quad \text{beliebig.} \end{aligned} \tag{2.167}$$

Die Ströme $\boldsymbol{i}_m$ liegen daher im r-dimensionalen Unterraum des $\mathbb{R}^m$. Dieser Teil ist durch die Quellenverteilung festgelegt. Somit verbleiben noch die $m - r$ Freiheitsgrade des orthogonalen Raumes. Insgesamt haben wir m Freiheitsgrade, die durch die äußere Beschaltung am rechten Tor festgelegt werden. Wird z.B. $\boldsymbol{a'}_m$ durch die äußere Beschaltung vorgegeben, so liegt $\boldsymbol{b'}_m$ eindeutig fest. (Vgl.: in einem Netz lassen sich die r Baumzweigspannungen und die $m = e - r$ Cobaumströme unabhängig vorgeben, also e Größen insgesamt. Durch die Quellenvorgabe sind hier r Größen festgelegt. Es verbleiben $e - r = m$ Freiheitsgrade.)

Der Vollständigkeit halber wollen wir noch die Beziehung

$$\boldsymbol{b}_m = \boldsymbol{S}\boldsymbol{a}_m + \boldsymbol{Q}\boldsymbol{j}\,. \tag{2.168}$$

für die Leistungswellen herleiten. Mit Gleichung (2.68) folgt aus Gleichung (2.166)

$$\boldsymbol{b}_m = \underbrace{[2\boldsymbol{G}_m^{1/2}\boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m^{1/2} - \boldsymbol{1}_m]}_{\boldsymbol{S}}\boldsymbol{a}_m + \underbrace{\boldsymbol{G}_m^{1/2}\boldsymbol{N}[\boldsymbol{N}^{\mathrm{T}}\boldsymbol{G}_m\boldsymbol{N}]^{-1}}_{\boldsymbol{Q}}\boldsymbol{j}\,. \tag{2.169}$$

Bei Leerlauf der r-Tore reduziert sich Gleichung (2.169) auf Gleichung (2.110) mit $\boldsymbol{A} = \pm\boldsymbol{N}^{\mathrm{T}}$.

Nun betrachten wir den mit einer vektoriellen idealen Spannungsquelle abgeschlossenen idealen e-Tor-Übertrager aus Bild 2.25. Es gilt $\boldsymbol{u}_m = \boldsymbol{e}$. Wir nehmen $m \leq r$ an, und $\boldsymbol{N}$ sei zeilenregulär. Wir erhalten in den Wellengrößen

$$\boldsymbol{b'}_r = \underbrace{[\boldsymbol{1}_r - 2\boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}}[\boldsymbol{N}\boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}}]^{-1}\boldsymbol{N}]}_{\boldsymbol{S'}}\boldsymbol{a'}_r + \underbrace{2\boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}}[\boldsymbol{N}\boldsymbol{R}_r\boldsymbol{N}^{\mathrm{T}}]^{-1}}_{\boldsymbol{Q'}}\,\boldsymbol{e}\,. \tag{2.170}$$

Bei Kurzschluss der m-Tore, d.h. $\boldsymbol{e} = \boldsymbol{0}$ reduziert sich Gleichung (2.170) zu Gleichung (2.111) mit $\boldsymbol{B} = \pm\boldsymbol{N}$.

2.8.4 Dissipative Bauelemente

Widerstand

Der Widerstand R kann als reale Spannungsquelle mit der Quellenspannung $E = 0$ aufgefasst werden, womit der Widerstand schon erfasst wäre. Wir werden aber dennoch ein eigenes Bauelement einführen, da die Quellen bei uns die Eingänge eines Funktionsbausteins (der Begriff wird später noch erläutert) sind. Die Streumatrix ist im Falle des Widerstandes ein Skalar, da der Widerstand nur ein Tor besitzt

$$\boldsymbol{S} = \boldsymbol{S}' = \frac{R - R_T}{R + R_T}. \tag{2.171}$$

Üblicherweise bezeichnet man diese Streumatrix als Reflektanz ρ. Das Wellendigitalfiltersymbol ist erneut aus dem Signalflussgraph abgeleitet und ist ein Multiplizierer (siehe Bild 2.26).

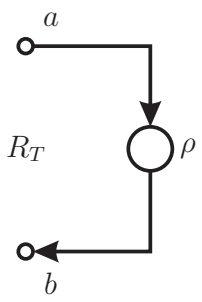


Bild 2.26: Wellenflussdiagramm eines Widerstands

2.8.5 Dynamische Bauelemente

Das einzige hier verwendete dynamische Bauelement ist der ideale Kondensator. Ideale Spulen lassen sich aus Gyratoren und idealen Kondensatoren realisieren. Die idealen Spulen haben gegenüber den idealen Kondensatoren den Nachteil einer zusätzlichen Negation in Wellenflussdiagramm. Der ideale Kondensator wird durch

$$\hat{i} = \sqrt{C}\ \mathrm{D}\{\sqrt{C}\hat{u}\} \tag{2.172}$$

beschrieben, wobei C eine Funktion der Zustandsgrößen und der unabhängigen Größen ist. Der Ableitungsoperator ist $\mathrm{D} = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_{\boldsymbol{t}}$ mit $\boldsymbol{\alpha} = \text{const.}$ und $\|\boldsymbol{\alpha}\| = 1$. Diese aus [Fett92] stammende Darstellung ist völlig gleichwertig zur ursprünglichen in [MF92] vorgeschlagenen Definition des nichtlinearen, zeitinvarianten idealen Kondensators, bietet aber den Vorteil übersichtlicherer Stromlaufpläne. [3]

Wir wollen zunächst eine diskrete Approximation für einen konstanten idealen Kondensator herleiten, der nur bzgl. der Richtung t_κ reaktiv ist, d. h. $\boldsymbol{\alpha} = \mathbf{e}_\kappa$, siehe Bild 2.27. Integrieren wir den Strom entlang einer Geraden vom Punkt $\boldsymbol{t}$ nach $\boldsymbol{t} + T_0\mathbf{e}_\kappa = \boldsymbol{t} + \Delta\boldsymbol{t}_\kappa$ und werten das Integral mittels der Trapezregel [4] aus, so erhalten wir einen Näherungswert für die Spannung am Punkt $\boldsymbol{t} + T_0\mathbf{e}_\kappa$

$$\hat{u}(\boldsymbol{t} + T_0\mathbf{e}_\kappa) \approx \hat{u}(\boldsymbol{t}) + R\left[\hat{i}(\boldsymbol{t} + T_0\mathbf{e}_\kappa) + \hat{i}(\boldsymbol{t})\right] \quad \text{mit} \quad R = \frac{\|\Delta\boldsymbol{t}_\kappa\|}{2\,C} = \frac{T_0}{2\,C}\,. \tag{2.173}$$

[3] Genau genommen haben die angegebenen Literaturstellen anstelle des idealen Kondensators eine ideale Spule benutzt.

[4] Wir verwenden hier die Trapezregel, weil sie einfach anzuwenden ist. Zudem hat die Trapezregel die größte Konsistenzordnung und den kleinsten lokalen Fehler aller A-stabilen (passiven) linearen Mehrschritt-Verfahren, [Dahl63], [Ochs01a]. Die Verwendung von Runge-Kutta-Verfahren unter Beibehaltung der Passivität wurde im eindimensionalen Fall in [Ochs01b] gezeigt. Dem Autor der hier vorliegenden Arbeit sind jedoch keine mehrstufigen passiven Runge-Kutta-Verfahren bekannt, die auf allgemeine PDGLn anwendbar sind.

Von nun an werden nicht mehr die kontinuierlichen unabhängigen Variablen verwendet, sondern die diskreten. Außerdem werden die mit einem Dach versehenen Größen durch Näherungsgrößen ersetzt

$$u(\boldsymbol{\nu}+\mathbf{e}_\kappa) = u(\boldsymbol{\nu}) + R\,[i(\boldsymbol{\nu}+\mathbf{e}_\kappa) + i(\boldsymbol{\nu})]\,. \tag{2.174}$$

Nach Umsortieren der letzten Gleichung und Verwendung der Wellengrößen mit dem Torwiderstand R erhalten wir

$$b(\boldsymbol{\nu}+\mathbf{e}_\kappa) = a(\boldsymbol{\nu})\ , \tag{2.175}$$

vgl. Bild 2.28. Wir fragen uns nun, wie sich die gefundenen Zusammenhänge im ursprünglichen Koordi-

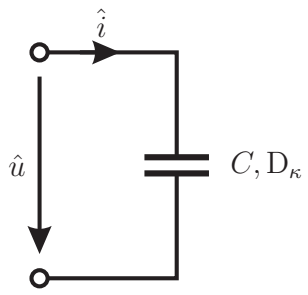


Bild 2.27: Idealer Kondensator, der bzgl. t_κ reaktiv ist

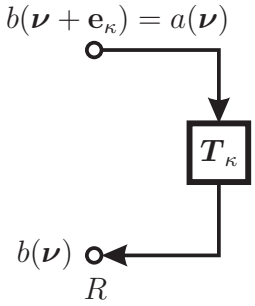


Bild 2.28: Näherungsweise Nachbildung durch ein diskretes System

natensystem $\boldsymbol{x}$ darstellen. Die Beschreibung des idealen Kondensators im ursprünglichen Koordinatensystem gewinnen wir durch Anwendung von Gleichung (2.16) auf Gleichung (2.172) zu

$$\hat{i} = C\,\mathrm{D}\{\hat{u}\} = \tilde{C}\,\frac{\boldsymbol{h}_\kappa^{\mathrm{T}}}{\|\boldsymbol{h}_\kappa\|}\,\mathbf{D}_{\boldsymbol{x}}\{\hat{u}\} \quad \text{mit} \quad \tilde{C} = C v_0 \|\boldsymbol{h}_\kappa\|\,. \tag{2.176}$$

Der zuvor durchgeführten Integration im Koordinatensystem $\boldsymbol{t}$ entspricht die Integration vom Punkt $\boldsymbol{x}$ nach $\boldsymbol{x}+\Delta\boldsymbol{x}_\kappa$ mit $\Delta\boldsymbol{x}_\kappa = v_0 T_0 \boldsymbol{h}_\kappa$ über das Kurvenstück

$$K = \{\boldsymbol{\xi}(\lambda) = \boldsymbol{x} + \lambda\,\Delta\boldsymbol{x}_\kappa \mid 0 \le \lambda \le 1\}\,. \tag{2.177}$$

Nach Umformung des Integrals in Parameterform und Berechnung mittels der Trapezregel erhalten wir einen Näherungswert für die Spannung am Punkt $\boldsymbol{x}+\Delta\boldsymbol{x}_\kappa$

$$\tilde{u}(\boldsymbol{x}+\Delta\boldsymbol{x}_\kappa) = \tilde{u}(\boldsymbol{x}) + \|\Delta\boldsymbol{x}_\kappa\|\frac{1}{2\tilde{C}}[\tilde{i}(\boldsymbol{x}+\Delta\boldsymbol{x}_\kappa) + \tilde{i}(\boldsymbol{x})]\ . \tag{2.178}$$

Der Torwiderstand lautet hier

$$R = \frac{\|\Delta\boldsymbol{x}_\kappa\|}{2\tilde{C}} = \frac{T_0}{2C}\ . \tag{2.179}$$

In den in dieser Arbeit verwendeten Zeichnungen bezieht sich der Kapazitätswert immer auf die Darstellung im neuen Koordinatensystem, vgl. Bild 2.27. Im Sinne der obigen Diskussion ist darunter also immer C und nicht $\tilde{C}$ zu verstehen. Wie gezeigt werden kann, sind sowohl der ideale Kondensator als auch die diskrete Nachbildung mehrdimensional intern verlustfreie Bauelemente, siehe z. B. [Fett98].

Der Verschiebevektor $\Delta\boldsymbol{\nu}_\kappa = \mathbf{e}_\kappa$ geht im Koordinatensystem $\boldsymbol{\mu}$ zu

$$\Delta\boldsymbol{\mu}_\kappa = \boldsymbol{X}_A^{-1}\,v_0\,\boldsymbol{H}\,T_0\,\mathbf{e}_\kappa = \boldsymbol{X}_A^{-1}\,v_0\,T_0\,\boldsymbol{h}_\kappa \tag{2.180}$$

über. Da dieser Vektor im Zusammenhang mit der Randbehandlung eine wichtige Rolle spielt, führen wir für den aus den ersten 3 Koordinaten gebildeten Untervektor die besondere Bezeichnung

$$\Delta\boldsymbol{\mu}_\kappa = \begin{bmatrix} \Delta\boldsymbol{\mu}'_\kappa \\ \\ 1 \end{bmatrix} \tag{2.181}$$

ein. Ebenso teilen wir die Spalten von $\boldsymbol{H}$ wie folgt auf

$$\boldsymbol{h}_{\bullet,\kappa} = \begin{bmatrix} \boldsymbol{h}'_\kappa \\ h_{k,\kappa} \end{bmatrix} . \tag{2.182}$$

Im Folgenden soll eine Beziehung zwischen der Spannung des idealen Kondensators und den Wellengrößen hergeleitet werden, die im Zusammenhang mit der Randbehandlung erhebliche Bedeutung besitzt, vgl. [Frie95], [Fett98] und [Sera00]. Dazu entwickeln wir zunächst die Funktion $f(\boldsymbol{t}) = f[(\boldsymbol{t}\pm\boldsymbol{T}/2)\mp \boldsymbol{T}/2]$ bezüglich der Variablen $\mp\boldsymbol{T}/2$ um den Punkt $\boldsymbol{t}\pm\boldsymbol{T}/2$ in eine Taylor-Reihe

$$f(\boldsymbol{t}) = f(\boldsymbol{t}\pm\boldsymbol{T}/2) \mp \frac{\boldsymbol{T}^{\mathrm{T}}}{2}\boldsymbol{D}_{\boldsymbol{\tau}} f(\boldsymbol{\tau}) \mid_{\boldsymbol{\tau}=\boldsymbol{t}\pm\boldsymbol{T}/2} + \mathcal{O}(\|\boldsymbol{T}\|^2), \tag{2.183}$$

wobei

$$\boldsymbol{D}_{\boldsymbol{\tau}} f(\boldsymbol{\tau}) \mid_{\boldsymbol{\tau}=\boldsymbol{t}\pm\boldsymbol{T}/2} = \mathrm{D}_{\boldsymbol{t}} f(\boldsymbol{t}\pm\boldsymbol{T}/2) \tag{2.184}$$

gilt. Aus dieser Beziehung gewinnen wir zum einen durch Wahl der oberen Vorzeichen und der Substitution $\boldsymbol{t} \Rightarrow \boldsymbol{t} - \boldsymbol{T}/2$

$$f(\boldsymbol{t}-\boldsymbol{T}/2) = f(\boldsymbol{t}) - \frac{\boldsymbol{T}^{\mathrm{T}}}{2}\mathrm{D}_{\boldsymbol{t}} f(\boldsymbol{t}) + \mathcal{O}(\|\boldsymbol{T}\|^2) \tag{2.185}$$

und zum anderen durch Wahl der unteren Vorzeichen und der Substitution $\boldsymbol{t} \Rightarrow \boldsymbol{t} + \boldsymbol{T}/2$

$$f(\boldsymbol{t}+\boldsymbol{T}/2) = f(\boldsymbol{t}) + \frac{\boldsymbol{T}^{\mathrm{T}}}{2}\mathrm{D}_{\boldsymbol{t}} f(\boldsymbol{t}) + \mathcal{O}(\|\boldsymbol{T}\|^2) . \tag{2.186}$$

Wir betrachten nun die Definition der Leistungswellen an einem idealen Kondensator gemäß Gleichung (2.172) mit dem Torwiderstand $R = \frac{T_0}{2C}$

$$\hat{b}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}[\hat{u}(\boldsymbol{t}) - R\hat{i}(\boldsymbol{t})] \quad , \quad \hat{a}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}[\hat{u}(\boldsymbol{t}) + R\hat{i}(\boldsymbol{t})] . \tag{2.187}$$

Durch Einsetzen des Stromes $\hat{i} = C\mathrm{D}\hat{u}$ erhalten wir

$$\hat{b}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}\left[\hat{u}(\boldsymbol{t}) - \frac{T_0}{2}\,\mathrm{D}\,\hat{u}(\boldsymbol{t})\right] \quad , \quad \hat{a}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}\left[\hat{u}(\boldsymbol{t}) + \frac{T_0}{2}\,\mathrm{D}\,\hat{u}(\boldsymbol{t})\right] . \tag{2.188}$$

Mit $\mathrm{D} = \boldsymbol{\alpha}^{\mathrm{T}}\mathrm{D}_{\boldsymbol{t}}$ und $\boldsymbol{T} = \boldsymbol{\alpha}T_0$ ergibt sich dann

$$\hat{b}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}\left[\hat{u}(\boldsymbol{t}) - \frac{\boldsymbol{T}^{\mathrm{T}}}{2}\mathrm{D}_{\boldsymbol{t}}\,\hat{u}(\boldsymbol{t})\right] \quad , \quad \hat{a}(\boldsymbol{t}) = \frac{1}{2\sqrt{R}}\left[\hat{u}(\boldsymbol{t}) + \frac{\boldsymbol{T}^{\mathrm{T}}}{2}\mathrm{D}_{\boldsymbol{t}}\,\hat{u}(\boldsymbol{t})\right] . \tag{2.189}$$

Durch Nutzung der Taylorreihenentwicklungen Gleichung (2.185) und Gleichung (2.186) erhalten wir die gewünschten approximativen Beziehungen

$$\hat{b}(\boldsymbol{t}) = \sqrt{\frac{C}{2T_0}}u(\boldsymbol{t}-\boldsymbol{T}/2) \approx \sqrt{\frac{C}{2T_0}}\hat{u}(\boldsymbol{t}-\boldsymbol{T}/2) \quad \text{und} \quad \hat{a}(\boldsymbol{t}) = \sqrt{\frac{C}{2T_0}}u(\boldsymbol{t}+\boldsymbol{T}/2) \approx \sqrt{\frac{C}{2T_0}}\hat{u}(\boldsymbol{t}+\boldsymbol{T}/2) ,$$

(2.190)

bzw. im diskreten ursprünglichen Koordinatensystem

$$\tilde{b}(\boldsymbol{\mu}) \approx \sqrt{\frac{C}{2T_0}}\,\tilde{u}(\boldsymbol{\mu} - \Delta\boldsymbol{\mu}/2) \quad \text{und} \quad \tilde{a}(\boldsymbol{\mu}) \approx \sqrt{\frac{C}{2T_0}}\,\tilde{u}(\boldsymbol{\mu} + \Delta\boldsymbol{\mu}/2)\,. \tag{2.191}$$

Die Ergebnisse können im linearen, konstanten Fall auch durch Betrachtung des Frequenzbereichs gewonnen werden. Dazu wenden wir die Approximationen

$$\begin{aligned} z^{1/2} &= \mathrm{e}^{\,pT_0/2} = \sum_{\nu=0}^{\infty} \frac{1}{\nu!}\left(\frac{pT_0}{2}\right)^{\nu} \approx 1 + pT_0/2 \Longleftrightarrow pT_0/2 \approx z^{1/2} - 1 \\ z^{-1/2} &= \mathrm{e}^{\,-pT_0/2} = \sum_{\nu=0}^{\infty} \frac{1}{\nu!}\left(-\frac{pT_0}{2}\right)^{\nu} \approx 1 - pT_0/2 \Longleftrightarrow pT_0/2 \approx 1 - z^{-1/2} \end{aligned} \tag{2.192}$$

auf $C\,p\,U = I$ an. Wir erhalten näherungsweise mit dem Torwiderstand $R = \frac{T_0}{2C}$

$$[z^{1/2} - 1]U = RI \quad \text{und} \quad [1 - z^{-1/2}]U = RI\,. \tag{2.193}$$

Hieraus gewinnen wir durch Verwendung von Leistungswellen

$$\frac{1}{2\sqrt{R}}\,z^{-1/2}U = A \quad \text{und} \quad \frac{1}{2\sqrt{R}}\,z^{1/2}U = B\,. \tag{2.194}$$

Durch Übergang zu den zeitabhängigen Größen ergibt sich Gleichung (2.191).

2.9 Zweckmäßige Beschreibung des Wellendigitalfilters

Das Ziel dieses Unterkapitels ist es, ein Wellendigitalmodell, welches aus den einzelnen WDF-Bauelementen des Unterkapitels 2.8 und deren Verbindungen besteht, in Form eines Gleichungssystems darzustellen. Das Gleichungssystem soll dabei so geartet sein, dass nach einer bloßen Festlegung der Berechnungsreihenfolge eine explizite Berechnung erfolgen kann. Durch die Festlegung der Berechnungsreihenfolge erhalten wir einen Algorithmus, der uns später als Grundlage für die formale Spezifikation und die Bestimmung eines Codes zur Simulation dienen wird.

Die zweckmäßige Beschreibung eines mehrdimensionalen Wellendigitalfilters geschieht in Anlehnung an die Darstellung im Eindimensionalen [Rumm98]. Diese Beschreibung ist besonders geeignet für die Führung eines formalen Korrektheitsbeweises. Wir nehmen zunächst an, alle verwendeten Quellen sind reflexionsfrei, d. h. die Torwiderstände sind gleich den Innenwiderständen der Quellen. Der Vektor $\boldsymbol{b}_q$ beinhaltet die skalaren Quellenwellen der Strom- und Spannungsquellen und hat die Länge n_q

$$\boldsymbol{b}_q = [b_{q1}, b_{q2}, \ldots, b_{qn_q}]^{\mathrm{T}}\,. \tag{2.195}$$

Die einfallenden Wellen der Quellen lauten

$$\boldsymbol{a}_q = [a_{q1}, a_{q2}, \ldots, a_{qn_q}]^{\mathrm{T}}\,. \tag{2.196}$$

Die Wellengrößen der dynamikfreien Elemente sind die Koordinaten der Vektoren

$$\boldsymbol{b}_e = [b_{e1}, b_{e2}, \ldots, b_{en_e}]^{\mathrm{T}} \quad \text{und} \quad \boldsymbol{a}_e = [a_{e1}, a_{e2}, \ldots, a_{en_e}]^{\mathrm{T}}\,. \tag{2.197}$$

Die Wellengrößen $\boldsymbol{a}_e$ und $\boldsymbol{b}_e$ sind über die Streumatrix $\boldsymbol{S}$ der dynamikfreien Elemente miteinander verknüpft, d. h.

$$\boldsymbol{b}_e = \boldsymbol{S}\,\boldsymbol{a}_e\,. \tag{2.198}$$

Hierin ist die Blockdiagonalmatrix $\boldsymbol{S}$ die direkte Summe der N_e Streumatrizen $\boldsymbol{S}_{\boldsymbol{\nu}}$ eines jeden dynamikfreien Elementes ν

$$\boldsymbol{S} = \mathbf{diag}(\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_{\boldsymbol{\nu}}, \ldots, \boldsymbol{S}_{N_e})\,. \tag{2.199}$$

Die Anzahl der Tore des dynamikfreien Elementes ν ist $n_{e\nu}$. Die Anzahl aller Tore aller dynamikfreien Elemente ist $n_e = \sum_{\nu=1}^{N_e} n_{e\nu}$. Die Verzögerer bzgl. t_κ werden durch

$$\boldsymbol{b}_v^\kappa([\nu_1, \ldots, \nu_{\kappa-1}, \nu_\kappa + 1, \nu_{\kappa+1}, \ldots, \nu_{k'}]^{\mathrm{T}}) = \boldsymbol{a}_v^\kappa(\boldsymbol{\nu}) \tag{2.200}$$

beschrieben, wobei alle Ausgangswellen der Verzögerer bzgl. t_κ in dem Vektor

$$\boldsymbol{b}_v^\kappa = [b_{v1}^\kappa, b_{v2}^\kappa, \ldots, b_{vn_v^\kappa}^\kappa]^{\mathrm{T}} \tag{2.201}$$

und alle Eingangswellen der Verzögerer bzgl. t_κ in dem Vektor

$$\boldsymbol{a}_v^\kappa = [a_{v1}^\kappa, a_{v2}^\kappa, \ldots, a_{vn_v^\kappa}^\kappa]^{\mathrm{T}} \tag{2.202}$$

zusammengefasst sind. Die Anzahl Verzögerer bzgl. t_κ bezeichnen wir mit n_v^κ. Die Anzahl aller Verzögerer ist $n_v = \sum_{\kappa=1}^{k'} n_v^\kappa$. Mit Hilfe des Standardeinheitsvektors erhalten wir die kompakte Darstellung

$$\boldsymbol{b}_v^\kappa(\boldsymbol{\nu} + \mathbf{e}_\kappa) = \boldsymbol{a}_v^\kappa(\boldsymbol{\nu}) \quad , \quad \kappa = 1, 2, \ldots, k'\,. \tag{2.203}$$

Weiterhin fassen wir die Vektoren der Verzögerer zu den Vektoren

$$\boldsymbol{a}_v = \begin{bmatrix} \boldsymbol{a}_v^1 \\ \boldsymbol{a}_v^2 \\ \vdots \\ \boldsymbol{a}_v^{k'} \end{bmatrix} \quad \text{und} \quad \boldsymbol{b}_v = \begin{bmatrix} \boldsymbol{b}_v^1 \\ \boldsymbol{b}_v^2 \\ \vdots \\ \boldsymbol{b}_v^{k'} \end{bmatrix} \tag{2.204}$$

zusammen.

Um das gesamte Wellendigitalfilter zu beschreiben, erweist es sich als günstig die Übervektoren

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_q \\ \boldsymbol{a}_v \\ \boldsymbol{a}_e \end{bmatrix} \quad \text{und} \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ \boldsymbol{b}_e \end{bmatrix} \tag{2.205}$$

zu bilden. Die Übervektoren $\boldsymbol{a}$ und $\boldsymbol{b}$ besitzen die Länge $n_g = n_q + n_v + n_e$.

Die Verschaltung der einzelnen Bauelemente wird durch die Permutationsmatrix $\boldsymbol{P}$ beschrieben. Die einzelnen Wellengrößen sind miteinander über

$$\boldsymbol{a} = \boldsymbol{P}\,\boldsymbol{b} \tag{2.206}$$

verknüpft. Ausführlich sind die Verbindungen der Bauelemente durch

$$\begin{bmatrix} \boldsymbol{a}_q \\ \boldsymbol{a}_v^1 \\ \vdots \\ \boldsymbol{a}_v^{k'} \\ \boldsymbol{a}_e \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_{qq} & \boldsymbol{P}_{qv^1} & \cdots & \boldsymbol{P}_{qv^{k'}} & \boldsymbol{P}_{qe} \\ \boldsymbol{P}_{v^1q} & \boldsymbol{P}_{v^1v^1} & \cdots & \boldsymbol{P}_{v^1v^{k'}} & \boldsymbol{P}_{v^1e} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{P}_{v^{k'}q} & \boldsymbol{P}_{v^{k'}v^1} & \cdots & \boldsymbol{P}_{v^{k'}v^{k'}} & \boldsymbol{P}_{v^{k'}e} \\ \boldsymbol{P}_{eq} & \boldsymbol{P}_{ev^1} & \cdots & \boldsymbol{P}_{ev^{k'}} & \boldsymbol{P}_{ee} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_{v^1} \\ \vdots \\ \boldsymbol{b}_{v^{k'}} \\ \boldsymbol{b}_e \end{bmatrix} \tag{2.207}$$

definiert. Durch Zusammenfassen von Matrizen zu Übermatrizen erhalten wir

$$\begin{bmatrix} \boldsymbol{a}_q \\ \boldsymbol{a}_v \\ \boldsymbol{a}_e \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_{qq} & \boldsymbol{P}_{qv} & \boldsymbol{P}_{qe} \\ \boldsymbol{P}_{vq} & \boldsymbol{P}_{vv} & \boldsymbol{P}_{ve} \\ \boldsymbol{P}_{eq} & \boldsymbol{P}_{ev} & \boldsymbol{P}_{ee} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ \boldsymbol{b}_e \end{bmatrix} \tag{2.208}$$

mit den entsprechend zu bildenden Übermatrizen $\boldsymbol{P}_{qv}, \boldsymbol{P}_{vq}, \boldsymbol{P}_{ev}, \boldsymbol{P}_{ve}$ und $\boldsymbol{P}_{vv}$. Die Matrix $\boldsymbol{P}$ ist eine quadratische Matrix der Dimension $n_g = n_q + n_v^1 + \cdots + n_v^{k'} + n_e$.

Permutationsmatrizen haben die Eigenschaft orthogonal zu sein. Hier ist $\boldsymbol{P}$ sogar symmetrisch, da die Verbindung der Bauelemente torweise erfolgt. Es gilt somit $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{1}_{n_g}$ und $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}$.

Das Wellendigitalkonzept geht von einer Referenzschaltung aus. Dies wollen wir beibehalten, d. h. jedes Bauelement der Referenzschaltung ist durch das Verbindungsnetz mit einem anderen verbunden. Für das digitale Verbindungsnetz bedeutet dies, dass mindestens ein WDF-Bauelement einer Verbindung ein dynamikfreies WDF-Bauelement ist, d.h. $\boldsymbol{P}_{vv} = \boldsymbol{0}$, $\boldsymbol{P}_{vq} = \boldsymbol{0}$, $\boldsymbol{P}_{qv} = \boldsymbol{0}$, $\boldsymbol{P}_{qq} = \boldsymbol{0}$. Dies gilt nach Definition für die weitere Arbeit. Die Permutationsmatrix lautet dann

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{P}_{qe} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{P}_{ve} \\ \boldsymbol{P}_{eq} & \boldsymbol{P}_{ev} & \boldsymbol{P}_{ee} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}_q \\ \boldsymbol{P}_v \\ \boldsymbol{P}_e \end{bmatrix}. \tag{2.209}$$

Im Bild 2.29 ist die Beschreibung des MDWDFs graphisch dargestellt. Eine Detailansicht in Form eines Signalflussdiagramms zeigt Bild 2.30.

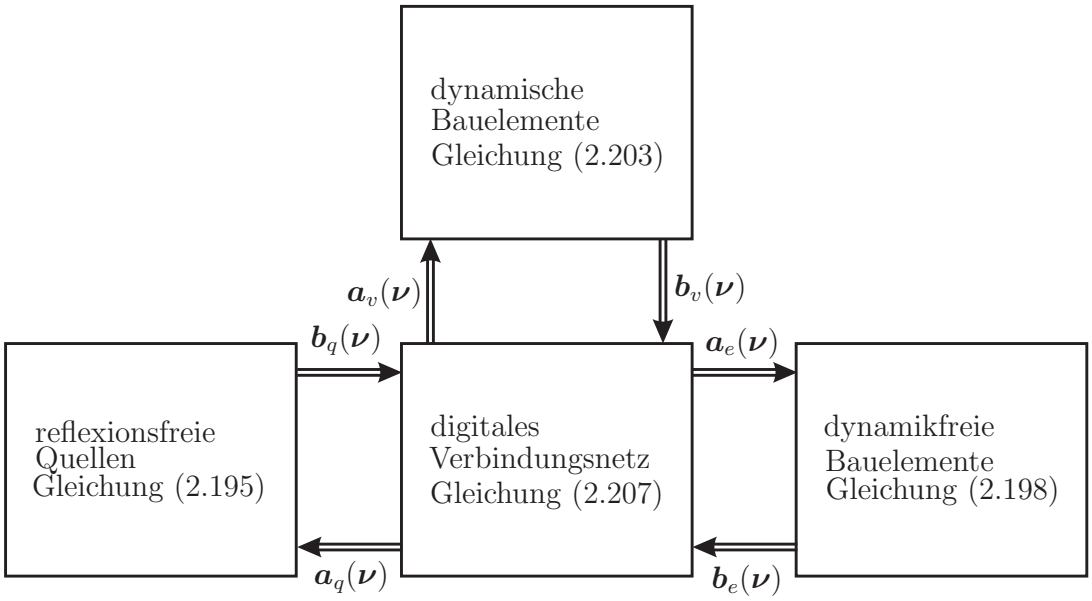


Bild 2.29: Algorithmus gültig in $\mathcal{G}$

Die Ausgangswellen der Quellen $\boldsymbol{b}_q(\boldsymbol{\nu})$ und der Verzögerer $\boldsymbol{b}_v^{\kappa}(\boldsymbol{\nu})$ sind zu einem Zeitpunkt für alle Gitterpunkte innerhalb $\mathcal{G}$ bekannt. Die nötige Berechnung von $\boldsymbol{b}_e(\boldsymbol{\nu})$ erfolgt durch Einsetzen der letzten Zeile von Gleichung (2.207) in Gleichung (2.198)

$$\boldsymbol{b}_e(\boldsymbol{\nu}) = \boldsymbol{S}\,[\boldsymbol{P}_{eq}\,\boldsymbol{b}_q(\boldsymbol{\nu}) + \boldsymbol{P}_{ev^1}\,\boldsymbol{b}_{v^1}(\boldsymbol{\nu}) + \cdots + \boldsymbol{P}_{ev^{k'}}\,\boldsymbol{b}_{v^{k'}}(\boldsymbol{\nu}) + \boldsymbol{P}_{ee}\,\boldsymbol{b}_e(\boldsymbol{\nu})]. \tag{2.210}$$

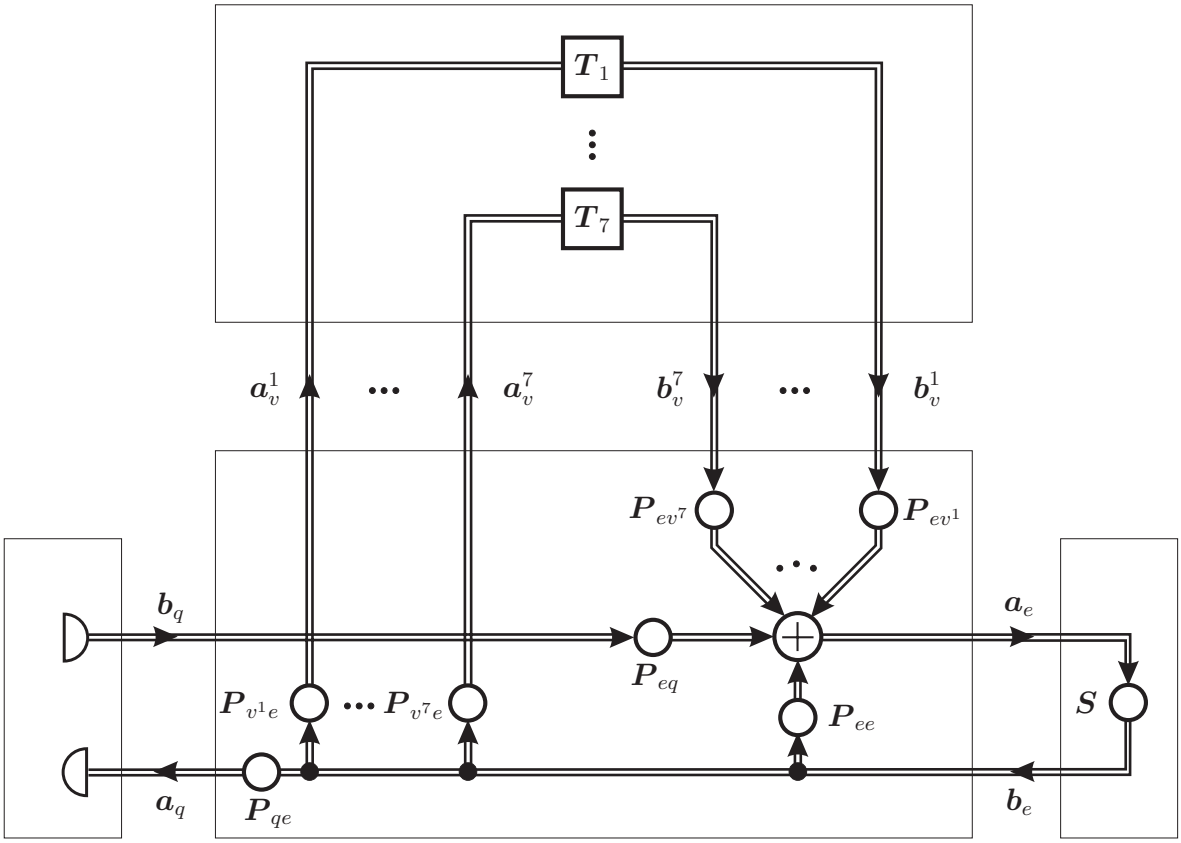


Bild 2.30: Detailansicht vom Bild 2.29

Die Matrix $\boldsymbol{S}\,\boldsymbol{P}_{ee}$ muss nun durch symmetrische Zeilen- und Spaltenpermutationen auf untere Dreiecksgestalt gebracht werden, bei der zudem noch die Hauptdiagonalelemente null sind. Eine untere Dreiecksmatrix, deren Hauptdiagonalelemente null sind, werden wir im weiteren Verlauf der Arbeit als strikte untere Dreiecksmatrix bezeichnen. Existiert eine Permutationsmatrix, die die Matrix $\boldsymbol{S}\,\boldsymbol{P}_{ee}$ auf strikte untere Dreiecksgestalt transformiert, so liegt keine verzögerungsfreie gerichtete Schleife in dem Wellendigitalfilter vor und die Berechenbarkeit ist gegeben.

Es stellt sich also die Frage, ob wenigstens eine Permutationsmatrix $\boldsymbol{P}_b$ existiert, sodass

$$\boldsymbol{P}_b\,\boldsymbol{S}\,\boldsymbol{P}_{ee}\,\boldsymbol{P}_b^{\mathrm{T}} = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \bullet & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \bullet & \cdots & \bullet & 0 \end{bmatrix} \tag{2.211}$$

gilt. Wir nehmen im Folgenden an, dass ein korrekt spezifiziertes MDWDF vorliegt. Wir führen die permutierten Vektoren

$$\boldsymbol{b'}_e = \boldsymbol{P}_b\,\boldsymbol{b}_e \quad \text{und} \quad \boldsymbol{b'} = [\boldsymbol{b}_q^{\mathrm{T}} \quad \boldsymbol{b}_v^{\mathrm{T}} \quad \boldsymbol{b'}_e^{\mathrm{T}}]^{\mathrm{T}} \tag{2.212}$$

ein und erhalten die in ihrer Reihenfolge getauschten Wellengrößen durch

$$\boldsymbol{b}'_e = \boldsymbol{P}_b\boldsymbol{b}_e = \boldsymbol{P}_b\,\boldsymbol{S}\,\boldsymbol{P}_b^{\mathrm{T}}\,[\,\boldsymbol{P}_b\boldsymbol{P}_{eq} \quad \boldsymbol{P}_b\boldsymbol{P}_{ev^1} \quad \cdots \quad \boldsymbol{P}_b\boldsymbol{P}_{ev^{k'}} \quad \boldsymbol{P}_b\boldsymbol{P}_{ee}\boldsymbol{P}_b^{\mathrm{T}}\,]\,\boldsymbol{b'} = \boldsymbol{L'}\boldsymbol{b'}. \tag{2.213}$$

Das weitere Einführen von permutierten Matrizen

$$\boldsymbol{S'} = \boldsymbol{P}_b\,\boldsymbol{S}\,\boldsymbol{P}_b^{\mathrm{T}} \quad , \quad \boldsymbol{P'}_{eq} = \boldsymbol{P}_b\,\boldsymbol{P}_{eq} \quad , \quad \boldsymbol{P'}_{ev} = \boldsymbol{P}_b\,\boldsymbol{P}_{ev} \quad , \quad \boldsymbol{P'}_{ee} = \boldsymbol{P}_b\,\boldsymbol{P}_{ee}\boldsymbol{P}_b^{\mathrm{T}} \tag{2.214}$$

führt uns zunächst zu

$$\boldsymbol{S'P'}_{eq} = \boldsymbol{P}_b\boldsymbol{S}\ \boldsymbol{P}_{eq} \quad , \quad \boldsymbol{S'P'}_{ev} = \boldsymbol{P}_b\boldsymbol{S}\ \boldsymbol{P}_{ev} \quad , \quad \boldsymbol{S'P'}_{ee} = \boldsymbol{P}_b\boldsymbol{S}\ \boldsymbol{P}_{ee}\,\boldsymbol{P}_b^{\mathrm{T}}, \tag{2.215}$$

so dass wir den permutierten Vektor $\boldsymbol{b'}_e$ durch

$$\begin{aligned} \boldsymbol{b'}_e &= \boldsymbol{S'P'}_{eq}\,\boldsymbol{b}_q + \boldsymbol{S'P'}_{ev}\,\boldsymbol{b}_v + \boldsymbol{S'}\,\boldsymbol{P'}_{ee}\,\boldsymbol{b'}_e \\ &= \underbrace{[\boldsymbol{S'P'}_{eq} \quad \boldsymbol{S'P'}_{ev} \quad \boldsymbol{S'P'}_{ee}]}_{\boldsymbol{L'}_=}\,\boldsymbol{b'} \end{aligned} \tag{2.216}$$

darstellen können. Dieses Gleichungssystem kann aufgrund der Struktur der Matrix $\boldsymbol{L'}$ durch Vorwärtssubstitution berechnet werden. Der neue Zustandsvektor ergibt sich zu

$$\boldsymbol{b}_v(\boldsymbol{\nu} + \mathbf{e}_\nu) = \boldsymbol{P}_{ve}\,\boldsymbol{b}_e(\boldsymbol{\nu}) = \boldsymbol{P}_{ve}\,\boldsymbol{P}_b^{\mathrm{T}} = \boldsymbol{P'}_{ve}\,\boldsymbol{b'}_e(\boldsymbol{\nu})\,. \tag{2.217}$$

Definieren wir abschliessend

$$\boldsymbol{L'}_q = \boldsymbol{S'}\,\boldsymbol{P'}_{eq}\,, \quad \boldsymbol{L'}_v = \boldsymbol{S'}\,\boldsymbol{P'}_{ev}\,, \quad \boldsymbol{L'}_e = \boldsymbol{S'}\,\boldsymbol{P'}_{ee}, \tag{2.218}$$

so können letztendlich die zu implementierenden Gleichungen zu

$$\begin{aligned} \boldsymbol{b'}_e(\boldsymbol{\nu}) &= \boldsymbol{L'}_q\,\boldsymbol{b}_q(\boldsymbol{\nu}) + \boldsymbol{L'}_v\,\boldsymbol{b}_v(\boldsymbol{\nu}) + \boldsymbol{L'}_e\,\boldsymbol{b'}_e(\nu) \\ \boldsymbol{b}_v^{\kappa}(\boldsymbol{\nu} + \boldsymbol{e}_\kappa) &= \boldsymbol{P}_{v^\kappa e}\boldsymbol{b}_e(\boldsymbol{\nu}) \qquad \forall\,\kappa = 1, 2, \ldots, k' \end{aligned} \tag{2.219}$$

notiert werden. Dabei hat die Berechnung der Vektoren $\boldsymbol{b'}_e(\boldsymbol{\nu})$ und $\boldsymbol{b}_v(\boldsymbol{\nu})$ zu allen Abtastpunkten einer Abtastschicht zu erfolgen, wobei allerdings die zweite Gleichung nur für eingeschränkte $\boldsymbol{\nu}$ gültig ist. Auf diese Problematik wird im Zusammenhang mit den Randwertfragen im Kapitel 2.10 eingegangen.

2.10 Randbehandlung

Die Verwendung von Digitalfiltern zur numerische Integration erfordert eine Berücksichtigung der Ränder. Ein Signalflussdiagramm (nach Bild 2.29) ist eine äquivalente Darstellung eines Differenzengleichungssystems (Gleichung (2.198), Gleichung (2.203) und Gleichung (2.207). Die Differenzengleichung wird dabei aus der zu lösenden Differentialgleichung gewonnen. Der Algorithmus eines mehrdimensionalen Wellendigitalfilters zur numerischen Integration ist nun nicht allein durch die Differenzengleichung bestimmt, sondern erfordert zusätzlich eine Beschreibung der Abmessungen des Berechnungsgebietes und die Behandlung des Randes.

Bevor wir die Diskussion fortsetzen, wollen wir den Begriff Randbehandlung präzisieren. Dazu betrachten wir eine beliebige Fläche im Raum und stellen uns auf eine Seite der Fläche. Das Verhalten der Feldgrößen auf dieser Seite der Fläche wird vom gesamten Raum hinter der Fläche beeinflusst. Oft machen es bestimmte Materialübergänge möglich, das Verhalten der Feldgrößen hinter der Fläche durch ein Modell zu approximieren, welches keine Abhängigkeit in Richtung des Normalenvektors der Grenzfläche besitzt. Wir definieren eine Fläche, auf der eine Approximation des Verhaltens der Feldgrößen innerhalb Raumes hinter dieser Fläche durchgeführt wird, als eine Randfläche. Wir unterscheiden bei denen durch die Approximation entstandenen Modelle i.W. zwei Fälle. Zum einen betrachten wir Modelle, die ausschließlich algebraische Beziehungen zwischen den Feldgrößen auf dem Rand liefern und zum anderen modellieren wir das Randverhalten durch differentielle Beziehungen zwischen den Feldgrößen.

Im Folgenden soll der Grund für die besondere Behandlung des Randes eines Berechnungsgebietes bei der numerischen Integration mit mehrdimensionalen Wellendigitalfiltern erläutert werden. Dazu betrachten wir noch einmal das Vorgehen bei der Simulation eines mehrdimensionalen Wellendigitalfilters. Zunächst werden gemäß Gleichung (2.219) alle fehlenden Wellengrößen der dynamikfreien Elemente des Berechnungsgebietes ermittelt. Danach werden die ausfallenden Wellen der dynamischen Elemente der nächsten Abtastschicht, d. h.

$$\boldsymbol{b}_v^\kappa(\boldsymbol{\nu} + \mathbf{e}_\kappa) = \boldsymbol{a}_v^\kappa(\boldsymbol{\nu}) \quad \text{für} \quad \kappa = 1, \ldots, k' \tag{2.220}$$

ermittelt. Wir gehen dazu von Gleichung (2.203) aus, wobei wir zuvor $\boldsymbol{a}_v^\kappa(\boldsymbol{\nu})$ aus allen bereits vorliegenden Vektoren $\boldsymbol{b}_q(\boldsymbol{\nu})$, $\boldsymbol{b}_v(\boldsymbol{\nu})$ und $\boldsymbol{b}_e(\boldsymbol{\nu})$ gemäß

$$\boldsymbol{a}_v^\kappa(\boldsymbol{\nu}) = \boldsymbol{P}_{v^\kappa e}\, \boldsymbol{b}_e(\boldsymbol{\nu}) \tag{2.221}$$

berechnet haben. Aus dieser Gleichung lässt sich allerdings nicht ohne Weiteres eine Aussage über die Randproblematik ableiten, da das Berechnungsgebiet im neuen Koordinatensystem nicht mehr die Form eines Quaders hat. Wir ziehen daher für die weitere Diskussion eine Beschreibung der dynamischen Elemente im ursprünglichen Koordinatensystem vor, d. h.

$$\tilde{\boldsymbol{b}}_v^\kappa(\boldsymbol{\mu}) = \boldsymbol{b}_v^\kappa(\boldsymbol{\nu}) = \boldsymbol{a}_v^\kappa(\boldsymbol{\nu} - \mathbf{e}_\kappa) = \tilde{\boldsymbol{a}}_v^\kappa(\boldsymbol{\mu} - \boldsymbol{h}_\kappa)\ . \tag{2.222}$$

Über diese Beziehung ist nun der Vektor $\tilde{\boldsymbol{b}}_v^\kappa(\boldsymbol{\mu})$ für alle örtlichen Gitterpunkte $\boldsymbol{\mu}' \in \mathcal{G}$ zu bestimmen, siehe Bild 2.31. Dafür sind die Wellengrößen $\tilde{\boldsymbol{a}}_v^\kappa$ zu den Gitterpunkten $(\boldsymbol{\mu}' - \boldsymbol{h}'_\kappa)$ notwendig. Die Wellengrößen liegen aber nur für die Gitterpunkte innerhalb des Berechnungsgebietes vor. Wir haben daher bei der Bestimmung von $\tilde{\boldsymbol{b}}_v^\kappa$ zwischen den Punkten $(\boldsymbol{\mu}' - \boldsymbol{h}_\kappa) \in \mathcal{G}$ und den Punkten $(\boldsymbol{\mu}' - \boldsymbol{h}_\kappa) \in \mathcal{G}_0$ zu unterscheiden. (Hier wurde selbstverständlich wieder angenommen, dass $\boldsymbol{\mu}'$ selber im Berechnungsgebiet $\mathcal{G}$ liegt.) Ein Algorithmus zur Bestimmung der Wellengrößen $\tilde{\boldsymbol{b}}_v^\kappa(\boldsymbol{\mu})$ ist somit in zwei Teilalgorithmen zu zerlegen, die Normalbehandlung und die Randbehandlung.

Es stellt sich daher die Frage, wie wir die notwendigen Wellengrößen am Rand bestimmen. Wir verwenden für die Randbehandlung einen Vorschlag von Fettweis und Seraji, [FS99], [Sera00]. In [Sera00] ist die Behandlung von Rändern im $(1+1)$D-Fall am Beispiel der Telegraphengleichung dargestellt. Es wird darauf hingewiesen, dass die vorgestellte Methode auf Problemstellungen mit beliebiger Dimension erweiterbar ist. Wir werden im Folgenden den für uns wesentlichen Teil aus dieser Arbeit rekapitulieren und das beschriebene Verfahren verallgemeinern. Hierbei beschränken wir uns auf die Transformationsmatrix und Abtastmatrix gemäß Gleichung (2.49). Untersuchungen haben ergeben, dass dies die einfachste Möglichkeit der Randbehandlung darstellt [Voll02]. Aus diese Festlegung resultiert zudem der Vorteil einer sich ergebenden festen Codestruktur, wie sich im Kapitel 6 zeigen wird.

Die generelle Schwierigkeit der Randbehandlung resultiert daraus, dass der Rand in den Feldgrößen beschrieben wird, aber das MDWDF als Zustandsgrößen die Wellengrößen verwendet. Zwar lassen sich die Feldgrößen aus den Wellengrößen bestimmen, die umgekehrte Richtung ist allerdings nur in Sonderfällen möglich. Das Grundprinzip zur Vermeidung dieser Schwierigkeit beruht auf der zuerst in [Frie95] dargestellten Tatsache, dass sich die Wellengrößen eines idealen Kondensators bzw. einer idealen Spule durch die Spannung bzw. den Strom zu einem um den halben Verschiebe-Vektor versetzten Punkt mit einer lokalen Konsistenzordnung von eins approximieren lassen, siehe Gleichung (2.190). Da die globale Konsistenzordnung der Trapezregel auch eins beträgt [HNW93] fügt sich diese Randbehandlung in den Rest des Verfahrens ein. Diese Methode der Randbehandlung erfordert einen Abstand von einem halben Verschiebe-Vektor zwischen dem randnächstem Punkt und der Grenzebene, so wie es im Kapitel 2.5 beschrieben wurde.

In [Sera00] wird ein weiteres, ähnliches Verfahren zur Randbehandlung beschrieben. Dabei liegen die Randpunkte auf der Grenzebene. Um nun die in Gleichung (2.190) beschriebene Approximation nutzen

zu können, wird die Abtastrate am Rand geändert. Dies hat allerdings zur Folge, dass die Torwiderstände der ein- und ausfallenden Wellen unterschiedlich sind. Für ein Werkzeug, welches eine automatische Codegenerierung ermöglichen soll, erscheint dieses Verfahren zu aufwendig und wird deshalb in dieser Arbeit nicht verwendet.

Um nun das Verfahren zu beschreiben, gehen wir der Einfachheit halber, ohne Einschränkung der Allgemeinheit, davon aus, dass die behandelte Referenzschaltung keine idealen Spulen, sondern nur ideale Kondensatoren als reaktive Bauelemente besitzt. Weiterhin ist festzustellen, dass sich bei der Randbehandlung nur diejenigen idealen Kondensatoren gegenseitig beeinflussen, deren zugehörige Verschiebe-Vektoren einen gemeinsamen Punkt auf der Grenzebene besitzen. Die Verschiebe-Vektoren mit einem gemeinsamen Punkt auf der Grenzebene müssen wir zusätzlich unterscheiden und zwar zwischen Verschiebe-Vektoren, die aus dem örtlichen[5] Berechnungsgebiet hinaus gerichtet sind, denen, die in das Berechnungsgebiet herein zeigen und denen, die orthogonal zum Normalenvektor sind. Dazu betrachten wir nun den Normalenvektor $\boldsymbol{n}$, der senkrecht auf der Grenzfläche steht und aus dem Berechnungsgebiet hinaus zeigt. Im (3+1)D-Fall haben wir, falls $\mathcal{G}$ ein Quader ist, die 6 verschiedenen Normalenvektoren $\boldsymbol{n}_1 = -\boldsymbol{n}_4 = \mathbf{e}_1$, $\boldsymbol{n}_2 = -\boldsymbol{n}_5 = \mathbf{e}_2$, $\boldsymbol{n}_3 = -\boldsymbol{n}_6 = \mathbf{e}_3$. Ein Orts- Verschiebe-Vektor $\boldsymbol{h}'_\kappa$ zeigt dann aus dem Berechnungsgebiet hinaus (herein), wenn das Skalarprodukt $\boldsymbol{n}^{\mathrm{T}}\,\boldsymbol{h}'_\kappa$ positiv (negativ) ist. Ist das Skalarprodukt $\boldsymbol{n}^{\mathrm{T}}\,\boldsymbol{h}'_\kappa$ gleich null, so ist zum einen für Zustandsgrößen mit dem Argument $\boldsymbol{\mu} + \boldsymbol{h}_\kappa$ keine Randbehandlung erforderlich und zum anderen liefern diese Zustandsgrößen keinen Beitrag zur Berechnung der anderen Randwerte. Die Spannungen derjenigen idealen Kondensatoren, deren zugehöriger Verschiebe-Vektor aus dem Berechnungsgebiet hinaus zeigt, sind die bekannten Feldgrößen eines Modells zur Randbehandlung. Die Spannungen derjenigen idealen Kondensatoren, deren zugehöriger Verschiebe-Vektor in das Berechnungsgebiet herein zeigt, müssen über das den Rand beschreibende Modell bestimmt werden.

Als Beispiel betrachten wir erneut ein Berechnungsgebiet in Form eines Quaders und nehmen eine Randfläche $x_1 = x_{1\,\max}$ an, siehe Bild 2.31. Wir betrachten zwei vektorielle ideale Kondensatoren deren diskrete Approximationen die Verschiebe-Vektoren $\boldsymbol{h}_1$ und $\boldsymbol{h}_4$ besitzen. Sie sollen den gemeinsamen Gitterpunkt $\boldsymbol{\mu} + \boldsymbol{h}_1/2$ haben. Der Verschiebe-Vektor $\boldsymbol{h}_1$ zeigt wegen $\boldsymbol{n}_1^{\mathrm{T}}\boldsymbol{h}'_1 = 1 > 0$ aus dem Berechnungsgebiet hinaus. Der Verschiebe-Vektor $\boldsymbol{h}_4$ zeigt in das Berechnungsgebiet herein.

Als Basis für das Aufstellen der benötigten Gleichungen dient Gleichung (2.191) in vektorieller Form mit $\boldsymbol{C}^\kappa = \mathbf{diag}(C_1^\kappa, \ldots, C_n^\kappa)$

$$\tilde{\boldsymbol{b}}^\kappa(\boldsymbol{\mu}) = [\boldsymbol{C}^\kappa]^{1/2}\sqrt{1/2T_0}\;\tilde{\boldsymbol{u}}^\kappa(\boldsymbol{\mu} - \boldsymbol{h}_\kappa/2) \quad \text{und} \quad \tilde{\boldsymbol{a}}^\kappa(\boldsymbol{\mu}) = [\boldsymbol{C}^\kappa]^{1/2}\sqrt{1/2T_0}\;\tilde{\boldsymbol{u}}^\kappa(\boldsymbol{\mu} + \boldsymbol{h}_\kappa/2)\;, \tag{2.223}$$

deren Argumente zu unserem Beispiel so verschoben werden müssen, dass die Spannungen als Argument den betrachteten Punkt $\boldsymbol{\mu} + \boldsymbol{h}_1/2$ haben. Der Vektor $\tilde{\boldsymbol{a}}^1(\boldsymbol{\mu})$ ist bekannt. Mit der Gleichung (2.191)

$$\tilde{\boldsymbol{a}}^1(\boldsymbol{\mu}) \;=\; [\boldsymbol{C}^1]^{1/2}\sqrt{1/2T_0}\;\tilde{\boldsymbol{u}}^1(\boldsymbol{\mu} + \boldsymbol{h}_1/2) \tag{2.224}$$

können wir den Vektor der Feldgrößen $\tilde{\boldsymbol{u}}^1(\boldsymbol{\mu} + \boldsymbol{h}_1/2)$ bestimmen. Die gesuchten Wellengrößen sind

$$\tilde{\boldsymbol{b}}^4(\boldsymbol{\mu} + \boldsymbol{h}_1/2 + \boldsymbol{h}_4/2) \;=\; [\boldsymbol{C}^4]^{1/2}\sqrt{1/2T_0}\;\tilde{\boldsymbol{u}}^4(\boldsymbol{\mu} + \boldsymbol{h}_1/2)\;. \tag{2.225}$$

Die anderen vorhandenen idealen Kondensatoren benötigen keine Randbehandlung und tragen auch nichts zur Randbehandlung bei, da ihre Verschiebe-Vektoren $\boldsymbol{n}_2$, $\boldsymbol{n}_3$, $\boldsymbol{n}_5$ und $\boldsymbol{n}_6$ orthogonal zum Normalenvektor $\boldsymbol{n}_1$ sind. Wir beachten $[\boldsymbol{h}_1 + \boldsymbol{h}_4]/2 = \mathbf{e}_4$.

Für den Randpunkt $\boldsymbol{\mu}+\boldsymbol{h}_1/2$ ist nun ein Abbild der physikalischen Realität anzugeben. Dieses Modell muss so geartet sein, dass nur die approximierten Größen $\tilde{\boldsymbol{u}}^1$ und $\tilde{\boldsymbol{u}}^4$ miteinander verknüpft sind, z. B. in der Form

$$\tilde{\boldsymbol{u}}^4 = \boldsymbol{M}\tilde{\boldsymbol{u}}^1\;. \tag{2.226}$$

[5] Im weiteren Verlauf dieses Unterkapitels ist immer das örtliche Berechnungsgebiet gemeint.

In diesem Fall haben wir

$$\tilde{\boldsymbol{b}}^4(\boldsymbol{\mu}+\mathbf{e}_4) \;=\; [\boldsymbol{C}^4]^{1/2}\boldsymbol{M}[\boldsymbol{C}^1]^{-1/2}\,\tilde{\boldsymbol{a}}^1(\boldsymbol{\mu})\;. \tag{2.227}$$

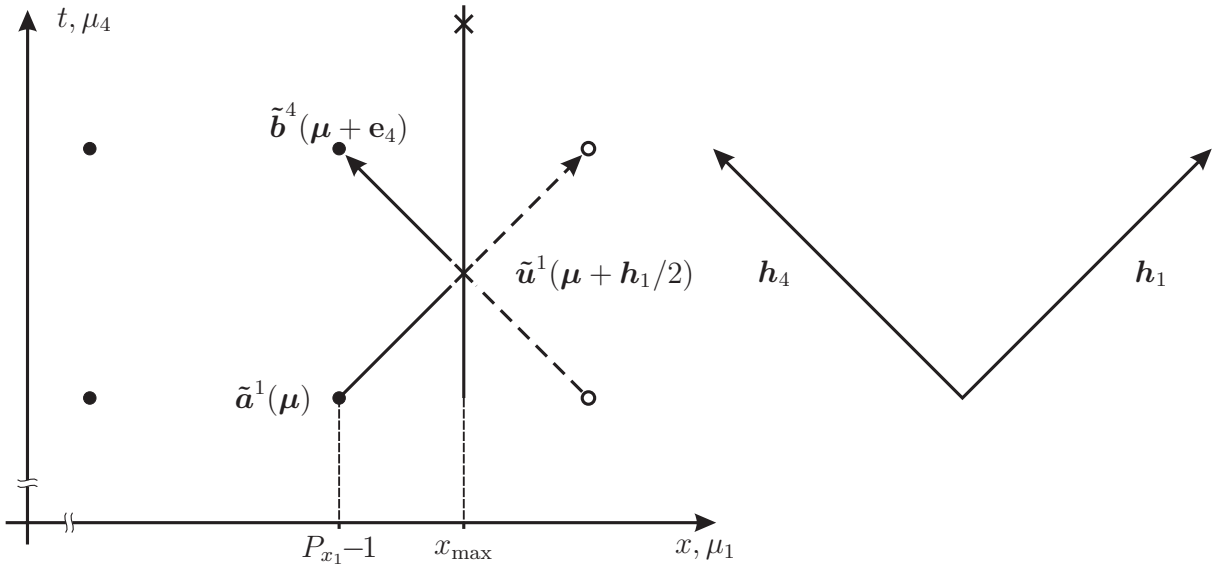


Bild 2.31: Randbehandlung

Im Falle eines nicht konstanten Modells am Rand ist das implizite Randlösungsmodell durch ein (ggf. nichtlineares) Wellendigitalfilter zu approximieren. Die Frage, die sich in jedem konkreten Einzelfall erneut stellt, ist, ob ein Randmodell gefunden werden kann, mit dem am Punkt $\boldsymbol{\mu}+\boldsymbol{h}_1/2$ aus den Spannungen $\tilde{\boldsymbol{u}}^1$ die gesuchten Spannungen $\tilde{\boldsymbol{u}}^4$ ermittelbar sind. Wir werden nicht weiter darauf eingehen, wie man ein derartiges Wellendigitalfilter des Randes findet, sondern es als gegeben annehmen.

Im weiteren Verlauf der Arbeit werden wir uns auf algebraische Randmodelle beschränken, da zu diesem Zeitpunkt noch keine ausreichenden Erkenntnisse über die Behandlung differentieller Ränder im (3+1)D Fall vorliegen. Im Falle algebraischer Randbedingungen können die Wellengrößen mittels Matrizen miteinander verknüpft werden

$$\tilde{\boldsymbol{b}}_v^{\kappa}(\boldsymbol{\mu}+\mathbf{e}_k) = \boldsymbol{R}^{\kappa}\,\tilde{\boldsymbol{a}}_v^{(\kappa+3)}(\boldsymbol{\mu}) \text{ für } \kappa=1,2,3 \quad \text{und} \quad \tilde{\boldsymbol{b}}_v^{\kappa}(\boldsymbol{\mu}+\mathbf{e}_k) = \boldsymbol{R}^{\kappa}\,\tilde{\boldsymbol{a}}_v^{(\kappa-3)}(\boldsymbol{\mu}) \text{ für } \kappa=4,5,6. \tag{2.228}$$

Die Matrizen $\boldsymbol{R}^1$ bis $\boldsymbol{R}^6$ stellen einen Teil der MDWDF-Spezifikation dar. Anhand der Spezifikation des Wellendigitalfilters für das Berechnungsgebiet kann eine Prüfung auf Vollständigkeit der Matrizen $\boldsymbol{R}^1$ bis $\boldsymbol{R}^6$ für die Randbehandlung erfolgen.

2.11 Von einem mehrdimensionalen Wellendigitalfilter zu einem eindimensionalen

Wir wollen nun auf den eigentlichen Ablauf der Berechnung eines Wellendigitalfilters eingehen. Wir betrachten dazu die notwendigen Rechenschritte zur Berechnung aller Wellengrößen zu einem festen Zeitpunkt μ_k. Zunächst haben wir die Quellenwerte $\boldsymbol{b}_q(\boldsymbol{\mu})$ für $\boldsymbol{\mu}' \in \mathcal{G}$ als bekannt vorausgesetzt. Die Verzögererwerte $\boldsymbol{b}_v(\boldsymbol{\nu})$ aller Abtastpunkte des Berechnungsgebietes der Zeitschicht μ_k können durch Auslesen der Speicherwerte aus den vorhandenen Werten der Zeitschicht μ_k-1 ermittelt werden. Liegt die Matrix $\boldsymbol{L}'$ aus Gleichung (2.219) in der benötigten Form vor, so kann der Vektor $\boldsymbol{b'}_e(\boldsymbol{\nu})$ mittels Vorwärtssubstitution berechnet werden. Die Berechnung des Vektors $\boldsymbol{b'}_e(\boldsymbol{\nu})$ kann dabei für jeden Abtastpunkt der

Zeitschicht μ_k parallel erfolgen, da wie oben bereits erwähnt die benötigten Vektoren $\boldsymbol{b}_q(\boldsymbol{\nu})$ und $\boldsymbol{b}_v(\boldsymbol{\nu})$ der Abtastschicht μ_k vorliegen. Wir haben also eine Kombination aus paralleler und serieller Berechnung. Die Ausnutzung der möglichen Parallelität erfordert allerdings eine spezielle Hardware, die für jeden Abtastpunkt des Berechnungsgebietes eine Recheneinheit vorsieht, die Gleichung (2.219) implementiert. Bei der Realisierung eines mehrdimensionalen Wellendigitalfilters auf einem Universalrechner werden aber alle Rechenoperationen zumindest im Quellcode sequentiell programmiert. (Inwieweit letztendlich eine Berechnung dann parallel ausgeführt wird, hängt natürlich von dem Universalrechner selber und dem verwendeten Compiler ab.) Durch das Verzichten auf die massiv parallele Abarbeitung gehen aber keine weiteren Vorteile des Verfahrens der numerischen Integration nach dem Wellendigitalfilterprinzip verloren [Fett99].

Im Folgenden soll eine analytische Beschreibung für die sequentielle Berechnung eines mehrdimensionalen Wellendigitalfilters hergeleitet werden, deren Grundlagen auf [Kumm88] zurückgehen. Wir bilden dazu das mehrdimensionale System auf ein eindimensionales System ab. Dabei kommt uns die Tatsache zu Hilfe, dass wir ein endliches Berechnungsgebiet vorliegen haben. Wir bilden jeden Ortsvektor der P Abtastpunkte einer Abtastschicht auf eine ganze Zahl δ ab. Sollen dann alle Abtastpunkte eines Berechnungsgebietes berechnet werden, so kann dies dadurch geschehen, dass die neue Variable δ die Werte 0 bis $P-1$ durchläuft und die Wellengrößen an diesen Punkten berechnet werden.

Die Verknüpfung von δ mit dem Vektor ${}_{-k}\boldsymbol{\mu} = \boldsymbol{\mu}'$ erfolgt durch die eineindeutige nichtlineare Abbildung

$$\begin{aligned} \boldsymbol{\mu}' = \boldsymbol{g}(\delta) \quad &: \quad \delta \in [0, P-1] \subset \mathbb{Z} \quad \longmapsto \quad \boldsymbol{\mu}' \in \mathcal{G} \subset \mathbb{Z}^{k-1} \\ \delta = f(\boldsymbol{\mu}') \quad &: \quad \boldsymbol{\mu}' \in \mathcal{G} \subset \mathbb{Z}^{k-1} \quad \longmapsto \quad \delta \in [0, P-1] \subset \mathbb{Z} \end{aligned} \tag{2.229}$$

die durch

$$\begin{aligned} \mu_\kappa &= \left\lceil \frac{\delta - \sum_{n=\kappa+1}^{k-1} \mu_n c_n}{c_\kappa} \right\rceil \quad , \quad \kappa = 1, 2 \ldots, k-1 \\ \delta &= \sum_{n=1}^{k-1} \mu_n c_n = {\boldsymbol{\mu}'}^{\mathrm{T}} \boldsymbol{c} \\ c_n &= \prod_{d=1}^{n-1} P_{x_d} \end{aligned} \tag{2.230}$$

definiert ist. Bei der Ermittlung des Vektors μ_κ ist man auf die μ_η , $\eta > \kappa$ angewiesen. Bei einer konkreten Berechnung beginnt man dann bei der letzten Koordinate und arbeitet sich zur ersten Koordinate durch, was einer Art Rückwärtssubstitution bei linearen Gleichungssystemen entspricht. Zur Verdeutlichung dient Bild 2.32.

Zum Beweis der Eineindeutigkeit der Abbildung benutzen wir zunächst

Satz 1

$$\sum_{n=1}^{N} (P_{x_n} - 1) \prod_{d=1}^{n-1} P_{x_d} = -1 + \prod_{d=1}^{N} P_{x_d}. \tag{2.231}$$

Beweis:

$$\sum_{n=1}^{N} (P_{x_n} - 1) \prod_{d=1}^{n-1} P_{x_d} = \sum_{n=1}^{N} P_{x_n} \prod_{d=1}^{n-1} P_{x_d} - \sum_{n=1}^{N} \prod_{d=1}^{n-1} P_{x_d} = \sum_{n=1}^{N} \prod_{d=1}^{n} P_{x_d} - \sum_{n=0}^{N-1} \prod_{d=1}^{n} P_{x_d} = \prod_{d=1}^{N} P_{x_d} - \prod_{d=1}^{0} P_{x_d}$$

(2.232)

Wir benötigen außerdem den

Satz 2

$$\sum_{n=1}^{N} \mu_n c_n < c_{N+1}. \tag{2.233}$$

Beweis:
Offenbar nimmt die Summe ihren maximalen Wert an, wenn in allen Richtungen der letzte Abtastpunkt innerhalb $\mathcal{G}$ eingesetzt wird, d. h. $\mu_n = P_n - 1, n = 1, 2, \ldots, N$ gilt. Wir können somit die Beweisaufgabe auf das Zeigen der Richtigkeit von

$$\sum_{n=1}^{N} (P_{x_n} - 1) c_n < c_{N+1} \tag{2.234}$$

zurückführen. Wir setzen die Definition der c_n aus Gleichung (2.230) ein und erhalten

$$\sum_{n=1}^{N} (P_{x_n} - 1) \prod_{d=1}^{n-1} P_{x_d} < \prod_{d=1}^{N} P_{x_d}. \tag{2.235}$$

Durch Verwendung von Satz 1 erkennt man unmittelbar die Korrektheit von Satz 2.

Wir kommen nun zur eigentlichen Aufgabe zurück, die Eineindeutigkeit der Abbildung zu zeigen. Um $\boldsymbol{\mu}' = \boldsymbol{g}(f(\boldsymbol{\mu}'))$ zu belegen, setzen wir

$$\delta = \sum_{n=1}^{k-1} \mu_n c_n \tag{2.236}$$

in

$$\mu_\kappa = \left\lceil \frac{\delta - \sum_{n=\kappa+1}^{k-1} \mu_n c_n}{c_\kappa} \right\rceil \tag{2.237}$$

ein und bekommen

$$\mu_\kappa = \left\lceil \frac{\sum_{n=1}^{k-1} \mu_n c_n - \sum_{n=\kappa+1}^{k-1} \mu_n c_n}{c_\kappa} \right\rceil = \left\lceil \frac{\sum_{n=1}^{\kappa} \mu_n c_n}{c_\kappa} \right\rceil. \tag{2.238}$$

Durch Aufteilen der Summe und Verwendung von Satz 2 finden wir

$$\mu_\kappa = \left\lceil \frac{\sum_{n=1}^{\kappa-1} \mu_n c_n + \mu_\kappa c_\kappa}{c_\kappa} \right\rceil = \left\lceil \underbrace{\frac{\sum_{n=1}^{\kappa-1} \mu_n c_n}{c_\kappa}}_{<1} + \mu_\kappa \right\rceil = \mu_\kappa. \tag{2.239}$$

Nachdem die Eineindeutigkeit bewiesen ist, kann die Berechnung der Punkte einer Abtastschicht auf das Berechnen der Wellengrößen zu den einzelnen Punkten der skalaren, unabhängigen Variablen δ zurückgeführt werden.

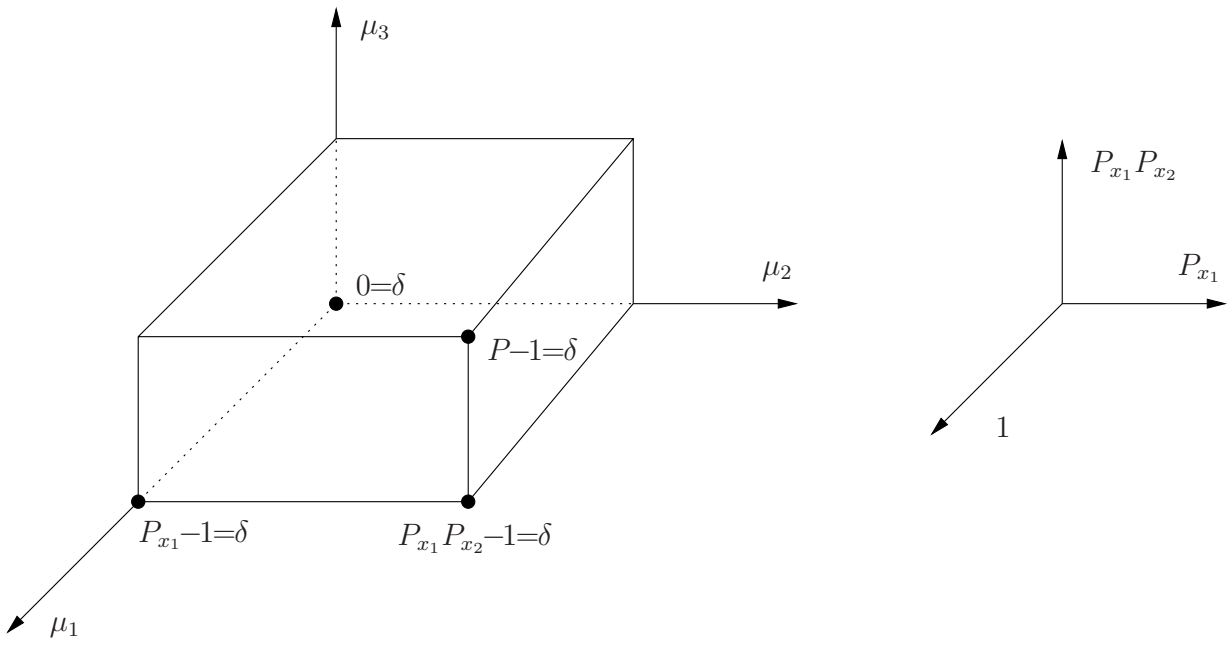


Bild 2.32: Koordinatensystem (links) und symbolische Darstellung der Gewichtungsfaktoren (rechts)

Abschließend soll noch einmal der nichtlineare Charakter der Abbildung veranschaulicht werden. Dazu nehmen wir drei konkrete Punkte δ aus dem Berechnungsgebiet an, die jeweils mit einem Punkt $\boldsymbol{\mu}'$ korrespondieren sollen, d.h.

$$\delta^{(1)} = f(\boldsymbol{\mu}'^{(1)}) \ , \ \delta^{(2)} = f(\boldsymbol{\mu}'^{(2)}) \ , \ \delta^{(3)} = f(\boldsymbol{\mu}'^{(3)}). \tag{2.240}$$

Aus der Summe bzw. der Differenz zweier δ folgt i. Allg. nicht, dass diese der Summe bzw. Differenz der zugehörigen $\boldsymbol{\mu}'$ entspricht, d.h.

$$\delta^{(1)} \pm \delta^{(2)} = \delta^{(3)} \Longrightarrow \boldsymbol{\mu}'^{(1)} \pm \boldsymbol{\mu}'^{(2)} \overset{\text{i.Allg.}}{\neq} \boldsymbol{\mu}'^{(3)}. \tag{2.241}$$

Für die umgekehrte Richtung trifft dies ebenso zu

$$\boldsymbol{\mu}'^{(1)} \pm \boldsymbol{\mu}'^{(2)} = \boldsymbol{\mu}'^{(3)} \Longrightarrow \delta^{(1)} \pm \delta^{(2)} \overset{\text{i.Allg.}}{\neq} \delta^{(3)}. \tag{2.242}$$

Schränkt man sich allerdings auf Vektoren $\boldsymbol{\mu}'^{(1)}$ und $\boldsymbol{\mu}'^{(2)}$ ein, deren Summe

$$0 \leq \mu_\kappa^{(1)} + \mu_\kappa^{(2)} \leq P_{x_\kappa} - 1 \quad \forall\, \kappa = 1, 2, \ldots, k-1 \tag{2.243}$$

genügt, so gilt

$$f(\boldsymbol{\mu}'^{(1)} + \boldsymbol{\mu}'^{(2)}) = \delta^{(1)} + \delta^{(2)} \tag{2.244}$$

$$\Longleftrightarrow \quad g(\delta^{(1)} + \delta^{(2)}) = \boldsymbol{\mu}'^{(1)} + \boldsymbol{\mu}'^{(2)}. \tag{2.245}$$

Wir führen den Beweis der Übersicht halber nur für den relevanten Fall $k = 4$. Mit

$$\delta^{(1)} = \mu_1^{(1)} + \mu_2^{(1)} P_{x_1} + \mu_3^{(1)} P_{x_1} P_{x_2} \quad \text{und} \quad \delta^{(2)} = \mu_1^{(2)} + \mu_2^{(2)} P_{x_1} + \mu_3^{(2)} P_{x_1} P_{x_2} \tag{2.246}$$

erhält man die rechte Gleichungsseite zu

$$\delta^{(1)} + \delta^{(2)} = \mu_1^{(1)} + \mu_2^{(1)} P_{x_1} + \mu_3^{(1)} P_{x_1} P_{x_2} + \mu_1^{(2)} + \mu_2^{(2)} P_{x_1} + \mu_3^{(2)} P_{x_1} P_{x_2}. \tag{2.247}$$

Da $\boldsymbol{\mu}'^{(1)} + \boldsymbol{\mu}'^{(2)}$) ebenfalls im Berechnungsgebiet liegt, lässt sich die linke Gleichungsseite zu

$$f(\boldsymbol{\mu}'^{(1)} + \boldsymbol{\mu}'^{(2)}) = (\mu_1^{(1)} + \mu_1^{(2)}) + (\mu_2^{(1)} + \mu_2^{(2)})P_{x_1} + (\mu_3^{(1)} + \mu_3^{(2)})P_{x_1}P_{x_2}. \tag{2.248}$$

berechnen.

Die Richtigkeit der umgekehrten Beziehung werden wir im Folgenden zeigen. Nach Gleichung (2.230) berechnen wir zunächst die dritte Koordinate zu

$$\mu_3^{(3)} = \left\lceil \frac{\delta^{(1)} + \delta^{(2)}}{P_{x_1}P_{x_2}} \right\rceil = \left\lceil \frac{\mu_1^{(1)} + \mu_1^{(2)}}{P_{x_1}P_{x_2}} + \frac{\mu_2^{(1)} + \mu_2^{(2)}}{P_{x_2}} + \mu_3^{(1)} + \mu_3^{(2)} \right\rceil = \mu_3^{(1)} + \mu_3^{(2)}\,, \tag{2.249}$$

wobei die letzte Identität durch mehrfache Anwendung von Gleichung (C.35) gewonnen wurde, d. h.

$$\frac{\mu_1^{(1)} + \mu_1^{(2)} + P_{x_1}[\mu_2^{(1)} + \mu_2^{(2)}]}{P_{x_1}P_{x_2}} \leq \frac{\mu_1^{(1)} + \mu_1^{(2)} + P_{x_1}[P_{x_2} - 1]}{P_{x_1}P_{x_2}} = 1 + \frac{\mu_1^{(1)} + \mu_1^{(2)} - P_{x_1}}{P_{x_1}P_{x_2}} \leq 1 - \frac{1}{P_{x_1}P_{x_2}} < 1\,. \tag{2.250}$$

Die zweite Koordinate berechnet man zu

$$\mu_2^{(3)} = \left\lceil \frac{\delta^{(1)} + \delta^{(2)} - (\mu_3^{(1)} + \mu_3^{(2)})P_{x_1}P_{x_2}}{P_{x_1}} \right\rceil = \left\lceil \underbrace{\frac{\mu_1^{(1)} + \mu_1^{(2)}}{P_{x_1}}}_{<1} + \mu_2^{(1)} + \mu_2^{(2)} \right\rceil = \mu_2^{(1)} + \mu_2^{(2)}. \tag{2.251}$$

Die erste Koordinate kann man dann zu

$$\mu_1^{(3)} = \left\lceil \delta^{(1)} + \delta^{(2)} - (\mu_2^{(1)} + \mu_2^{(2)})P_{x_1} - (\mu_3^{(1)} + \mu_3^{(2)})P_{x_1}P_{x_2} \right\rceil = \left\lceil \mu_1^{(1)} + \mu_1^{(2)} \right\rceil = \mu_1^{(1)} + \mu_1^{(2)}. \tag{2.252}$$

ermitteln, womit Gleichung (2.245) auch bewiesen ist.

Kapitel 3

Das Engineeringsystem SPACE

Für Steuerungs- und Regelungsfunktionen mit sicherheitstechnischer Bedeutung wird das digitale Leitsystem Teleperm XS verwendet. Die typischen Anwendungen von Teleperm XS liegen im Bereich des Reaktorschutzes und der ESFAS-Funktionen (Engineered Safety Features Actuation System). Des Weiteren können sicherheitsrelevante Funktionen wie etwa die Reaktorregelung oder die Aufgaben des Steuerstabfahrrechners übernommen werden.

Die Projektierung und Wartung von Teleperm XS Systemen erfolgt mit dem Engineeringsystem SPACE (SPecification And Coding Environment). Das Engineeringsystem SPACE stellt eine graphische Oberfläche zur Verfügung, unter der sowohl die Leittechnikfunktionen, als auch die Hardwarearchitektur spezifiziert werden können. Aus der Spezifikation kann automatisch C-Code generiert werden, der anschließend übersetzt, gebunden und in das Zielsystem (mehrere Verarbeitungseinheiten der Teleperm XS Hardware) geladen werden kann. [KWU 98]

Zur Zeit wird am Lehrstuhl für Nachrichtentechnik der Ruhr-Universität-Bochum an einem Zusatzpaket zum Engineeringsystem SPACE gearbeitet, welches die i. W. von Fettweis entwickelten Wellendigitalfilter, unter Beibehaltung aller Qualitätsmerkmale in Bezug auf die Korrektheit der erstellten Software, in das Programmpaket SPACE einbindet. Die Wellendigitalfilter sollen zur numerischen Integration von partiellen Differentialgleichungen, insbesondere der Neutronendiffusionsgleichung, genutzt werden, um das Potential digitaler Leittechnik zur weiteren Optimierung von Steuerungs- und Regelungsstrategien nutzbar zu machen.

3.1 Der Aufbau des vom Engineeringsystem SPACE erzeugten C-Codes

Das Engineeringsystem SPACE ermöglicht die graphische Projektierung einerseits der Hardwarekonfiguration und andererseits der Funktionen des zu realisierenden Leittechniksystems. Ausschlaggebend für die Funktion des Leittechniksystems ist die graphische Spezifikation der Software, was i. W. durch Verschalten von vorgegebenen Funktionsbausteinen (FB) erfolgt, wodurch Funktionspläne (FP) entstehen. Die Funktionsbausteine sind manuell codiert, qualifiziert und anlagenunabhängig. Jeder Funktionsplan eines Projektes ist einer in der Hardwarekonfiguration angegebenen Verarbeitungseinheit zuzuweisen. Auf jeder Verarbeitungseinheit läuft im Betrieb der Anlage eine Ablaufumgebung, in die eine Gruppe von Funktionsplänen eingebettet ist. Eine Funktionsplangruppe (FPG) besteht aus einem oder mehreren Funktionsplänen, die wiederum die Funktionsbausteine beinhalten. Der C-Code-Generator erzeugt zu jedem Funktionsplan und zu jeder Funktionsplangruppe automatisch eine C-Code-Datei und ein zugehöriges Header-File. Der C-Code der Funktionspläne ruft in seinem Code die fertig vorliegenden Funktionsbausteine auf, wobei die Funktionspläne selbst von der Funktionsplangruppe aufgerufen werden. Somit liegt ein kompletter C-Code für eine Verarbeitungseinheit vor, der nun compiliert und auf das

Zielsystem TELEPERM XS geladen werden kann oder von einer Simulationsumgebung zur Verifizierung compiliert wird. Die Aufrufhierarchie verdeutlicht Bild 3.1.

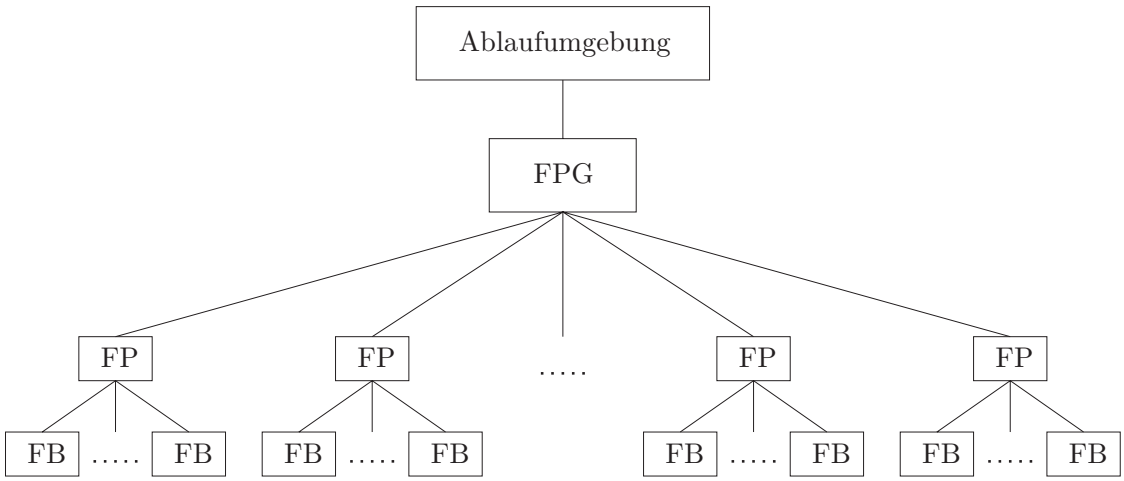


Bild 3.1: Aufrufhierarchie Funktionsplangruppe-Funktionsplan-Funktionsbaustein

3.2 Die Einbindung der Wellendigitalfilter in das Engineeringsystem SPACE

Vielen Bauelementen der klassischen Theorie elektrischer Netze, können entsprechende Wellendigitalfilter-Elemente zugeordnet werden. Um nun das numerische Integrationsverfahren nach dem Wellendigitalfilterprinzip in das Engineeringsystem SPACE zu integrieren, könnte man jedes Wellendigitalfilter-Element als einen Funktionsbaustein realisieren. Dieses Vorgehen führt allerdings zu den folgenden drei schwerwiegenden Problemen:

- Das Programmpaket SPACE detektiert verzögerungsfreie gerichtete Schleifen, obwohl keine vorhanden sind:

  Die Prüfung der Funktionsplangruppen und Funktionspläne auf verzögerungsfreie gerichtete Schleifen erfolgt ohne Berücksichtigung der inneren Struktur der miteinander verschalteten Funktionsbausteine [Waed94]. Dieses Vorgehen kommt in nahezu allen Problemstellungen einer Detektion von echten verzögerungsfreien gerichteten Schleifen gleich, da davon ausgegangen wird, dass jeder Eingang eines Funktionsbausteins alle Ausgänge beeinflusst. In der Theorie der Wellendigitalfilter existieren mehrere Bauelemente, bei denen das Ausgangssignal eines Tores unabhängig von dem Eingangsignal des gleichen Tores ist. An dieser Stelle wird eine mögliche verzögerungsfreie gerichtete Schleife durch eine spezielle Wahl der inneren Struktur der Bauelemente aufgebrochen. In nahezu jeder praktisch relevanten WDF-Nachbildung einer analogen Referenzschaltung tritt eines dieser Bauelemente auf, da diese Bauelemente gerade für das Aufbrechen der verzögerungsfreien gerichteten Schleifen Verwendung finden. Bei dem Wellendigitalfilter nach Bild 3.2 würde das Programmpaket SPACE eine verzögerungsfreie gerichtete Schleifen detektieren, obwohl das linke Bauelement diese Schleife aufbricht.

- Im Programmpaket SPACE ist nur eine unabhängige Variable vorgesehen:

  Das Programmpaket SPACE stellt nur Totzeitglieder in der Zeit zur Verfügung. Möchte man aber eine partielle Differentialgleichung mit der Zeit und den drei Ortskoordinaten als unabhängige Variablen lösen, so hat das zugehörige zeitdiskrete System mindestens vier unabhängige Variablen.

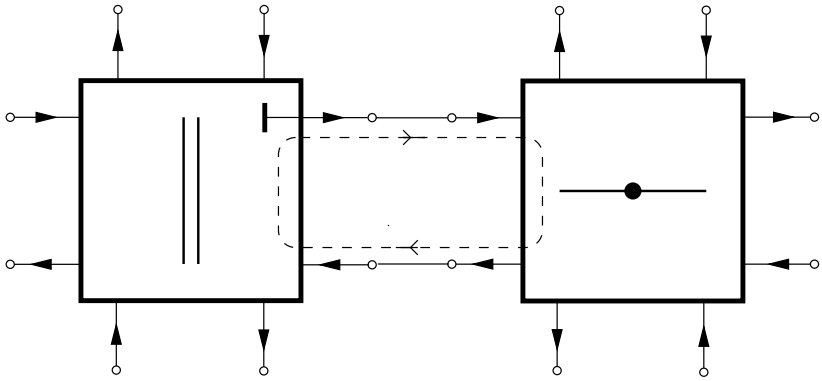


Bild 3.2: Aufbrechen von verzögerungsfreien gerichteten Schleifen

- Einem WDF-Baustein kann nicht ohne weiteres einem C-Code-Modul zugeordnet werden:

  Der Grund liegt darin, dass die Berechnung eines Wellendigitalfilters nicht bausteinweise erfolgt. Stattdessen wird die Berechnungsreihenfolge nach Gleichung (2.216) genutzt.

Um diesen beiden Problemen entgegentreten zu können, werden wir zunächst den Aufbau eines Funktionsbausteins näher untersuchen. Im Bild 3.3 ist die Aufrufschnittstelle eines Funktionsbausteins

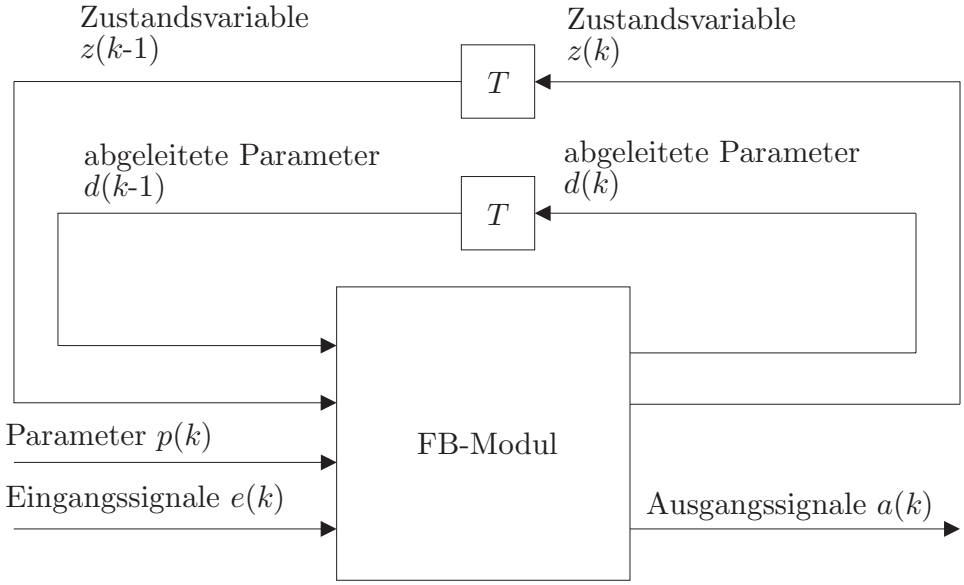


Bild 3.3: Funktionsbaustein eingebettet in einen Funktionsplan

dargestellt. Fasst man die dort auftretenden Vektoren wie folgt zusammen

$$\boldsymbol{w}(k+1) = \begin{bmatrix} \boldsymbol{z}(k) \\ \\ \boldsymbol{d}(k) \end{bmatrix} \quad , \quad \boldsymbol{x}(k) = \begin{bmatrix} \boldsymbol{p}(k) \\ \\ \boldsymbol{e}(k) \end{bmatrix} \quad , \quad \boldsymbol{y}(k) = \boldsymbol{a}(k) \tag{3.1}$$

und ändert den Laufindex ab, d.h. $k = \nu$, so erhalten wir die in der digitalen Signalverarbeitung übliche Zustandsraumdarstellung eines eindimensionalen Systems

$$\begin{aligned} \boldsymbol{w}(\nu+1) &= \boldsymbol{A}\,\boldsymbol{w}(\nu) + \boldsymbol{B}\,\boldsymbol{x}(\nu) \\ \boldsymbol{y}(\nu) &= \boldsymbol{C}\,\boldsymbol{w}(\nu) + \boldsymbol{D}\,\boldsymbol{x}(\nu). \end{aligned} \tag{3.2}$$

Folglich kann jedes zeitdiskrete eindimensionale, lineare mathematische Modell mittels eines Funktionsbaustein-Moduls in Verbindung mit den zugehörigen Zustandsspeichern gebildet werden. Ein 1D-Wellendigitalfilter ist eine i. d. R. nichtkanonische Zustandsraumdarstellung eines Digitalfilters.

Wir halten als Ergebnis dieses Abschnitts fest, dass eine Realisierung eines Wellendigitalfilters mit einzelnen Wellendigitalfilter-Elementen, ausgeführt als Funktionsbaustein, i. Allg. nicht möglich ist, ein 1D-Wellendigitalfilter aber selber immer durch ein Funktionsbaustein-Modul aufgebaut werden kann.

Wir werden von dieser Erkenntnis Gebrauch machen und ein komplettes Wellendigitalfilter als einen Funktionsbaustein realisieren. Der Leittechniker in der Praxis kann dann auf fertige Funktionsbausteine zurückgreifen, welche z. B. Löser einer Differentialgleichung repräsentieren. Dieses Vorgehen erscheint auch aus dem Grund sinnvoll, da das Auffinden eines Wellendigitalfilters zu einer gegebenen Differentialgleichung (z. B. Neutronendiffusionsgleichung) einen gewissen Sachverstand voraussetzt [Luhm04].

Im Folgenden soll beispielhaft der Original C-Code des Funktionsbausteins `fb_154.c` erläutert werden. Es handelt sich hierbei i. W. um einen Integrierer, der durch verschiedene Steuersignale noch Sonderfunktionen, wie z. B. Schnellangleich, externen Halt usw. ausführen kann.

Der Original C-Code ist für einen Integrierer recht umfangreich. Dies liegt zum einen an den realisierten Sonderfunktionen und zum anderen an eingebauten Konsistenzprüfungen. Wir interessieren uns in diesem Abschnitt nur für den eigentlichen Verarbeitungsteil des Integrierers. Aus diesem Grunde haben wir den C-Code zu diesem Programmpfad kompakt dargestellt:

```
dlOut    = ddlC * (DL_t) eas1.v + mdlState ;
aas1.v   = (FL_t) dlOut ;
mdlState = ddlC * (DL_t) eas1.v + (DL_t) aas1.v;
```

Hinter `DL_t`, `FL_t` verbirgt sich Double und Float. Mit (`DL_t` wird nicht multipliziert, sondern `DL_t` bezeichnet den so genannten Cast-(Umwandlungs-)Operator d. h. die Variablen `eas1.v`, `dlOut` und `aas1.v` werden explizit in den entsprechenden Datentyp umgewandelt.

Bevor wir mit der signaltheoretischen Untersuchung beginnen, sei bemerkt, dass jedes Aufrufen des C-Codes einem Fortschreiten um einen Abtastpunkt entspricht, d. h. die Betriebsperiode des Funktionsbausteins ist durch die Betriebsperiode des Funktionsplanes vorgegeben.

Nun zum eigentlichen Programmcode. In der ersten Programmzeile wird das Eingangssignal `eas1.v` mit einer Konstanten `ddlC` multipliziert und der aktuelle Wert der Zustandsgröße `mdlState` hinzuaddiert. Das Ergebnis wird der temporären Variablen `dlOut` zugewiesen. Die temporäre Variable `dlOut` stellt gleichzeitig das Ausgangssignal dar. Deshalb wird der Wert von `dlOut` der Ausgangsvariablen `aas1.v` in der zweiten Zeile zugewiesen. In der dritten Zeile wird dann der neue Wert der Zustandsgröße aus dem Ein- und dem Ausgangssignal berechnet.

Die temporäre Konstante `ddlC` besteht aus der Betriebsperiode T_a und einer Integrierkonstanten T_i: `ddlC`$= T_a/2T_i$.

Fassen wir den Programmcode als ein lineares konstantes zeitdiskretes System auf und eliminieren wir `dlOut`, so können die Systemgleichungen in die übliche Standardnotation Gleichung (3.2) gebracht werden

$$\begin{aligned} \texttt{aas1.v} &= \texttt{mdlState} + \tfrac{T_a}{2T_i}\,\texttt{eas1.v} \\ \texttt{mdlState} &= \texttt{mdlState} + \tfrac{T_a}{T_i}\,\texttt{eas1.v}\,. \end{aligned} \tag{3.3}$$

An dieser Stelle ist es von größter Wichtigkeit auf Folgendes hinzuweisen. Die Variable `mdlState` ist eine skalare Variable. Möchte man die Werte dieser Variablen zu mehreren Zeitpunkten speichern, so muss die Variable ein Feld sein. In dem Fall liegt kein Problem bei der Zuordnung vor. Eine mögliche Zuordnung ist `mdlState[k]`$= x(k)$ und `mdlState[k+1]`$= x(k+1)$. Erfordert die Anwendung nur den aktuellen Wert der Zustandsvariablen, so lässt sich die Zustandsgröße in einem Programm durch eine skalare Variable

realisieren. Fraglich ist nun, ob man der Programmvariablen den Wert $x(k)$ oder $x(k+1)$ zuordnet. Die Frage kann folgendermaßen beantwortet werden. Solange $x(k)$ benötigt wird, muss `mdlState`$= x(k)$ gelten. Wird $x(k)$ nicht mehr benötigt, so kann die Zuweisung `mdlState`$= x(k+1)$ erfolgen. Dies erfordert eine Reihenfolge der Berechnung und zwar werden erst alle Leseoperationen, dann der Wechsel von k nach $k+1$ durchgeführt und zum Schluss erfolgt die Schreiboperation. Diese Reihenfolge ist der Grund dafür, dass die Zeilen in Gleichung (3.3) gegenüber denen in Gleichung (3.2) vertauscht sind.

Die Übertragungsfunktion berechnet sich mit $\boldsymbol{H}(z) = \boldsymbol{C}[z\mathbf{1} - \boldsymbol{A}]^{-1}\boldsymbol{B} + \boldsymbol{D}$ zu

$$H(z) = \frac{T_a}{2T_i}\left[\frac{2}{z-1} + 1\right] = \frac{T_a}{2T_i}\left[\frac{z+1}{z-1}\right] = \frac{T_a}{2T_i\psi}. \tag{3.4}$$

Der Übertragungsfunktion ist zu entnehmen, dass es sich um einen Integrierer nach der Trapezregel handelt.

3.3 Zur Wahl der Abtastperioden

Die Wahl der Abtastmatrix $\boldsymbol{X}_A$ ist zunächst durch das Abtasttheorem festgelegt. Zusätzliche Randbedingungen sind der vorhandene Speicherplatz, die zur Verfügung stehende Rechenzeit und Forderungen des Lastenhefts. Eine Forderung des Pflichtenhefts ist die Festlegung eines Zeitintervalls, in dem die Berechnung der Lösung zu erfolgen hat. Die bisher eingeführten Zeiten sind die Zeit t als unabhängige Variable der Differentialgleichung und die mehrdimensionale Zeit $\boldsymbol{t}$. Keine der beiden Zeiten hat allerdings unmittelbar mit dem Fortschreiten der Rechenzeit des implementierten Algorithmus zu tun. Das Fortschreiten der Rechenzeit hängt unter anderem von der verwendeten Hardware ab. Die Simulationszeit und die unabhängige Variable t stimmen i. Allg. nicht überein. Bei unserer vorliegenden Echtzeitanwendung (siehe Gesamtkonzept) heißt dies nun, dass alle notwendigen Berechnungen zur Abarbeitung einer Abtastschicht innerhalb der Abtastperiode der Zeit durchzuführen sind. Natürlich kann man, wenn nicht genügend Rechenzeit zur Verfügung steht, die Abtastperiode T erhöhen oder die Anzahl der Abtastpunkte P reduzieren, so dass alle nötigen Berechnungen innerhalb der Abtastperiode T ausgeführt werden können. Dieses Vorgehen, der Erhöhung der Abtastperioden, steht allerdings diametral dem Abtasttheorem gegenüber. Eine andere Möglichkeit ist eine Abtastratenumsetzung, d. h. der Löser der Differentialgleichung hat nicht die gleiche Abtastrate wie der Versorgungsblock.

3.4 Realisierungsaspekte

Die vorliegende Beschreibung der Verzögerer bedeutet im ursprünglichen Koordinatensystem, dass der Vektor $\tilde{\boldsymbol{b}}_v(\boldsymbol{\mu})$ zu jedem festen Zeitpunkt μ_k und beliebigem Abtastpunkt nur von Werten eines Zeitpunktes zuvor, d.h. $\mu_k - 1$, abhängt und bietet somit den Vorteil, dass nur Werte einer Abtastschicht gespeichert werden müssen. Die dabei zum Zeitpunkt μ_k benötigten Randwerte werden aus den Werten des Berechnungsgebietes zum Zeitpunkt $\mu_k - 1$ durch Gleichung (2.228) ermittelt. Die Tatsache, dass alle zur Zeitschicht μ_k zu berechnenden Werte durch Kenntnis der Werte des physikalischen Zeitpunktes zuvor berechnet werden können, ermöglicht die in 2.9 vorgestellte sehr einfache, systematische Beschreibungsweise eines mehrdimensionalen Wellendigitalfilters . Dies hat noch den weiteren Vorteil, dass nach jedem Zeitschritt globale Größen des Berechnungsgebietes $\mathcal{G}$ wie z.B. die Gesamtleistung, der Eigenwert, berechnet werden können, bevor der nächste Zeitschritt beginnt.

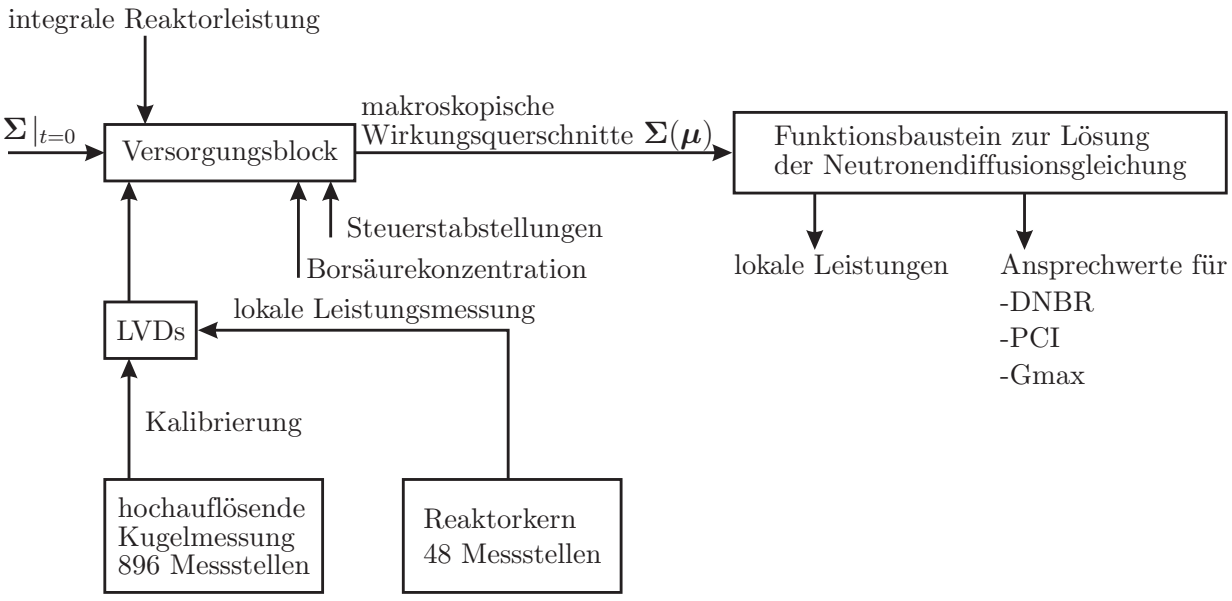


Bild 3.4: Blockschaltbild

3.5 Verknüpfung der Gleichungen des MDWDFs mit dem Funktionsbaustein

Bevor wir auf die Verknüpfung eingehen wollen, muss noch diskutiert werden, ob und an welcher Stelle der Anwender (i. d. R. der Leittechniker) Einfluss auf die Parameter des Wellendigitalfilters nehmen kann. Zunächst soll anhand eines typischen Anwendungsbeispiels die Frage geklärt werden, ob der Anwender überhaupt Parameter an das Wellendigitalfilter übergeben muss. Dazu betrachten wir Bild 3.4. Das dort angegebene Blockschaltbild erläutert u. a. die Berechnung der Leistungsverteilung im Reaktorkern, eine typische Aufgabe eines Zusatzpaketes für TELEPERM XS. Wie dem Bild zu entnehmen ist, setzt die Berechnung der lokalen Leistung die Lösung der Neutronendiffusionsgleichung voraus, was aber wiederum ein typisches Einsatzgebiet der Wellendigitalfilter ist. Legt man zwei Energiegruppen zugrunde, so berechnet sich die lokale thermische Reaktorleistung P_n innerhalb des Volumens V zu

$$P_n = \kappa \int_V (\nu\, \Sigma_{f1}\, \varphi_1 + \nu\, \Sigma_{f2}\, \varphi_2)\mathrm{d}V. \tag{3.5}$$

Die Neutronenflussdichten φ_1 und φ_2 werden durch Lösung der Neutronendiffusionsgleichung

$$\begin{bmatrix} \operatorname{div} \boldsymbol{j_1} \\ \operatorname{div} \boldsymbol{j_2} \end{bmatrix} + \left[\mathbf{diag} \left\{ \frac{1}{v_1}, \frac{1}{v_2} \right\} \frac{\partial}{\partial t} + \boldsymbol{A} \right] \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} = \mathbf{0}\,, \quad \boldsymbol{A} = \begin{bmatrix} \Sigma_{a1} + \Sigma_{\overrightarrow{12}} - \frac{\nu_1}{\lambda}\Sigma_{f1} & -\frac{\nu_2}{\lambda}\Sigma_{f2} \\ -\Sigma_{\overrightarrow{12}} & \Sigma_{a2} \end{bmatrix}$$

$$\begin{aligned} (\tau_1 \tfrac{\partial}{\partial t} + 1)\, \boldsymbol{j}_1 \; + \mathcal{D}_1 \operatorname{grad} \varphi_1 = 0 \\ (\tau_2 \tfrac{\partial}{\partial t} + 1)\, \boldsymbol{j}_2 \; + \mathcal{D}_2 \operatorname{grad} \varphi_2 = 0 \end{aligned} \qquad \text{(Ficksches Gesetz)} \tag{3.6}$$

ermittelt [Pott98], [Smid75]. Die Vektoren $\boldsymbol{j}_1$ und $\boldsymbol{j}_2$ sind die Neutronenstromdichten der Energiegruppen 1 und 2. Die Parameter Σ beschreiben die makroskopischen Wirkungsquerschnitte: Σ_a ist der Absorptionsquerschnitt, Σ_f ist der Wirkungsquerschnitt für die Kernspaltung und $\Sigma_{\overrightarrow{12}}$ beschreibt die Streuung von der Energiegruppe 1 in die Energiegruppe 2. $\mathcal{D}_1$ und $\mathcal{D}_2$ stellen die Diffusionskonstanten dar. Die in der Matrix $\boldsymbol{A}$ auftretenden Größen ν_1 und ν_2 sind in diesem Zusammenhang die mittleren Anzahlen der Spaltneutronen einer Energiegruppe und nicht die diskreten Variablen des neuen Koordinatensystems.

Die Größen τ_1, τ_2, v_1 und v_2 dienen der Hyperbolisierung der üblicherweise verwendeten Form der Differentialgleichung. Alle auftretenden Parameter sind i. Allg. Funktionen des Ortes, werden aber lokal als konstant angenommen. Räume mit konstanten Parametern werden wir den Teilberechnungsgebieten $\mathcal{G}_n$ zuordnen.

Zur weiteren Erläuterung des Blockschaltbildes betrachten wir eine Betriebsperiode des Reaktors in der keine Änderung des Beladungszustands vorgenommen wird (i.d.R. 1 Jahr). Vor dem Reaktoranlauf (t=0) wird dieser mit Brennstäben beladen. Aus den bekannten Daten dieser Brennstäbe lassen sich die makroskopischen Wirkungsquerschnitte berechnen $\boldsymbol{\Sigma}|_{t=0}$. Mit diesen Parametern lässt sich nun die Leistungsverteilung berechnen. Während des Betriebs werden sich die Materialeigenschaften des Brennstoffes ändern, was sich in der Differentialgleichung durch Veränderung der Parameter äußert. Die makroskopischen Wirkungsquerschnitte sind somit nicht nur Funktionen des Ortes, sondern auch der Zeit. Die Berechnung der makroskopischen Wirkungsquerschnitte übernimmt der Versorgungsblock. Der Versorgungsblock benötigt dazu die integrale Reaktorleistung, die makroskopischen Wirkungsquerschnitte zur Zeit der Neubeladung und die lokalen Leistungen an (z.B. 48) Messstellen. Die lokalen Leistungen werden durch LVDs (Leistungsverteilungsdetektoren) gemessen. Die Leistungsverteilungsdetektoren wiederum werden ca. alle 14 Tage durch hochauflösende Kugelmessungen an den $28 \times 32 = 896$ Messstellen kalibriert. Durch Vergleich der gemessenen und der für $t = 0$ berechneten Leistungsverteilung kann der Versorgungsblock eine Adaption der makroskopischen Wirkungsquerschnitte $\boldsymbol{\Sigma}(\boldsymbol{x})$ vornehmen [Fisc00b].

Als Ergebnis der Diskussion halten wir fest, dass zur Lösung der Differentialgleichung die makroskopischen Wirkungsquerschnitte, die Anzahlen der Spaltneutronen und die Diffusionskonstanten erforderlich sind. Ein entsprechender Funktionsbaustein, der den Löser der Neutronendiffusionsgleichung repräsentiert, muss demnach von außen die Parameter des zugehörigen Wellendigitalfilters übergeben bekommen.

Nun stellt sich die Frage, auf welche Art man generell Parameter an die Funktionsbausteine übergibt. Betrachten wir erneut Bild 3.3, so stellen wir fest, dass die Übergabe prinzipiell über den dort angegebenen Eingangsvektor $\boldsymbol{e}(k)$ und den Parametervektor $\boldsymbol{p}(k)$ erfolgen kann. Die Werte des Parametervektors $\boldsymbol{p}(k)$ werden über eine so genannten Parametriermaske bei der Erstellung eines Funktionsplans eingegeben und können somit während des Betriebs nicht mehr verändert werden. Dies liegt darin begründet, dass die Parameter des Vektors $\boldsymbol{p}(k)$ zwar Übergabeparameter eines FB-Softwaremoduls sind, aber keine Eingangssignale eines Funktionsbausteins, wie er im Schaltplaneditor auftritt. Da unsere Aufgabenstellung, wie oben bereits diskutiert, eine Änderung der Parameter des Wellendigitalfilters im Betrieb erfordert, müssen diese Parameter als Eingänge des Funktionsbausteins aufgefasst werden.

Eine Änderung der Parameter des Wellendigitalfilters hat aber eine Änderung der Streumatrix $\boldsymbol{S}$ zur Folge, was wiederum eine Änderung der Matrizen $\boldsymbol{L'}_q$, $\boldsymbol{L'}_v$ und $\boldsymbol{L'}_e$ in Gleichung (2.219) bewirkt. Aus Gleichung (2.219) wird aber unmittelbar der C-Code der Funktionsbausteine generiert, d. h. der gemäß Gleichung (2.219) generierte C-Code hängt von den Übergabeparametern des Funktionsbausteins ab. Die sich daraus ergebenden Änderungen zur Generierung des C-Codes ist Gegenstand der Arbeit [Juss00].

Die während des Betriebs sich nicht ändernden Parameter des Wellendigitalfilters können über eine Parametriermaske in SPACE eingegeben werden. Zudem sind bei der eigentlichen Erstellung eines Funktionsbausteins natürlich beliebig viele Parameter und die Struktur selber änderbar.

Syntaktische Grundlage für die Einbindung eines Wellendigitalfilters in einen Funktionsbaustein ist die Übergabeschnittstelle. Beim Aufruf eines Funktionsbausteins durch einen Funktionsplan wird grundsätzlich die Adresse einer Variablen vom Typ `fb_<log_id>_t` übergeben. Der Datentyp dieser Variablen wird mithilfe des C-Schlüsselwortes `struct` zur Definition von Datenstrukturen festgelegt. Die Datenstruktur lautet [KWU 99] :

```
typedef  struct
{ const  TX_t     *txExtIdent_p;       /* externe Identifikation
```

```
  const <E/A-Typ> *e<e/a-Typ>l_p;      /* Zeiger auf 1. Eingangssignal      */
  const <E/A-Typ> *e<e/a-Typ>2_p;      /* Zeiger auf 2. Eingangssignal      */
                                       /* ... */
  const <E/A-Typ> *e<e/a-Typ><i>_p;    /* Zeiger auf i. Eingangssignal      */
        <E/A-Typ> *a<e/a-Typ><i+l>_p;  /* Zeiger auf 1. Ausgangssignal      */
        <E/A-Typ> *a<e/a-Typ><i+2>_p;  /* Zeiger auf 2. Ausgangssignal      */
                                       /* ... */
        <E/A-Typ> *a<e/a-Typ><i+o>_p;  /* Zeiger auf o. Ausgangssignal      */
  const <Par-Typ> *p<par-Typ>l_p;      /* Zeiger auf 1. Parameter           */
  const <Par-Typ> *p<par-Typ>2_p;      /* Zeiger auf 2. Parameter           */
                                       /* ... */
  const <Par-Typ> *p<par-Typ><p>_p;    /* Zeiger auf p. Parameter           */
        <Mem-Typ> *d<mem-Typ>l_p;      /* Zeiger auf 1. abgeleiteten Parameter */
        <Mem-Typ> *d<mem-Typ>2_p;      /* Zeiger auf 2. abgeleiteten Parameter */
                                       /* ... */
        <Mem-Typ> *d<mem-Typ><d>_p;    /* Zeiger auf d. abgeleiteten Parameter */
        <Mem-Typ> *m<mem-Typ>l_p;      /* Zeiger auf 1. Zustandsspeicher    */
        <Mem-Typ> *m<mem-Typ>2_p;      /* Zeiger auf 2. Zustandsspeicher    */
                                       /* ... */
        <Mem-Typ> *m<mem-Typ><m>_p;    /* Zeiger auf m. Zustandsspeicher    */
}fb_<log_id>_t ;
```

Da nach Verlassen des Funktionsbausteins alle internen Variablen wieder gelöscht werden, müssen die Zustände in den Zustandsspeichern `m<mem-Typ><nr>_p` abgelegt werden. Die Anzahl der Zustandsspeicher ergibt sich als Produkt aus den Zustandsgrößen des mehrdimensionalen Wellendigitalfilters und der Anzahl der Abtastpunkte einer Abtastschicht: $n_v P$. Legen wir im Fall $k = 4$ eine Zahl von 100 Berechnungspunkten in jede Richtung zugrunde, so benötigen wir $n_v 10^6$ skalare Übergabeparameter, was nicht sinnvoll zu realisieren ist. Um eine derartige Menge an Variablen kompakt beschreiben zu können, gibt es in der Programmiersprache C das Konzept der vektoriellen Datentypen. Die Deklaration von Feldern ist aber im Programmpaket SPACE nicht vorgesehen. *Die Einbindung von vektoriellen Datentypen sollte zukünftig in das Programmpaket SPACE integriert sein.* Dazu sind entsprechende Erweiterungen an dem Codegenerator des Programmpaketes SPACE vorzunehmen [Waed00].

Wir werden im weiteren Verlauf dieser Arbeit davon ausgehen, dass innerhalb des Codegenerators auch vektorielle Signale definiert werden können. Für die Quellen, Verzögerer und Ausgänge sind die an einen Funktionsbaustein übergebenen Variablen Vektoren, deren Länge gleich der Anzahl Abtastpunkte einer Abtastschicht ist. Der Wert eines Feldelementes entspricht `(*(args->mdl1_p))[elementnr]`. Um eine bessere Übersicht in den Quellcodes zu ermöglichen, werden wir in den Quellcodes anstatt `(*(args->mdl1_p))[elementnr]` Bezeichner nutzen, die den verwendeten Variablen des Wellendigitalfilters entsprechen. Diese Bezeichner werden dann mittels den in den Header-Dateien definierten Präprozessordirektiven `#define` aufgelöst, in diesem Fall durch `#define b_v_1_1 (*(args->mdl1_p))`. Näheres dazu findet sich in [Voll02].

Auf die einzelnen Koordinaten kann dann mittels der Adressierung `[delta]` zugegriffen werden. Die Länge der Vektoren der variablen Parameter ist gleich der Anzahl der Teilberechnungsgebiete, wobei wir der Einfachheit halber annehmen, dass die Länge der Vektoren der variablen Parameter auch P ist. Eine Koordinate des übergebenen Vektors von Zeigern ist dabei ein Zeiger, der auf einen Vektor mit den ortsabhängigen Parametern des Wellendigitalfilters zeigt. Die neue Datenstruktur mit vektoriellen Signalen lautet

```
typedef struct{
  const TX_t *txExtIdent_p;              /* externe Identifikation            */
  const <E/A-Typ> *e<e/a-Typ>l_p[P];     /* Zeiger auf 1. Quellensignal       */
```

```
  const <E/A-Typ> *e<e/a-Typ>2_p[P];       /* Zeiger auf 2. Quellensignal          */
                                           /* ...                                  */
  const <E/A-Typ> *e<e/a-Typ><n_q>_p[P];   /* Zeiger auf n_q. Quellensignal        */
  const <E/A-Typ> *e<e/a-Typ><n_q+1>_p[P]; /* Zeiger auf 1. variablen Parameter    */
                                           /* ...                                  */
  const <E/A-Typ> *e<e/a-Typ><i>_p[P];     /* Zeiger auf n_vP. variablen Parameter */
        <E/A-Typ> *a<e/a-Typ><i+l>_p[P];   /* Zeiger auf 1. Ausgangssignal         */
        <E/A-Typ> *a<e/a-Typ><i+2>_p[P];   /* Zeiger auf 2. Ausgangssignal         */
                                           /* ...                                  */
        <E/A-Typ> *a<e/a-Typ><i+o>_p[P];   /* Zeiger auf o. Ausgangssignal         */
  const <Par-Typ> *p<par-Typ>l_p;          /* Zeiger auf 1. Parameter              */
  const <Par-Typ> *p<par-Typ>2_p;          /* Zeiger auf 2. Parameter              */
                                           /* ...                                  */
  const <Par-Typ> *p<par-Typ><p>_p;        /* Zeiger auf p. Parameter              */
        <Mem-Typ> *d<mem-Typ>l_p;          /* Zeiger auf 1. abgeleiteten Parameter */
        <Mem-Typ> *d<mem-Typ>2_p;          /* Zeiger auf 2. abgeleiteten Parameter */
                                           /* ...                                  */
        <Mem-Typ> *d<mem-Typ><d>_p;        /* Zeiger auf d. abgeleiteten Parameter */
        <Mem-Typ> *m<mem-Typ>l_p[P];       /* Zeiger auf 1. Zustandsspeicher       */
        <Mem-Typ> *m<mem-Typ>2_p[P];       /* Zeiger auf 2. Zustandsspeicher       */
                                           /* ...                                  */
        <Mem-Typ> *m<mem-Typ><m>_p[P];     /* Zeiger auf m. Zustandsspeicher       */
}fb_<log_id>_t ;
```

Die Eingangs-, Ausgangs- und Zustandssignale bekommen in Übereinstimmung mit der Header-Datei der Funktionsbausteine fb_t.h des Programmpaketes SPACE fortlaufende Nummern, und zwar in der Reihenfolge, dass die ersten i Nummern von den Eingangssignalen belegt sind und die Ausgangssignale von den darauf folgenden Nummern. Die Zustandsgrößen erhalten neue Nummern. Wir werden die Variablennamen der Übersicht halber für die theoretische Beschreibung in Vektoren zusammenfassen:

$$\begin{aligned} \boldsymbol{e}_{\mathrm{FB}} &= \left[\begin{array}{l} \texttt{e<e/a-Typ>1\_p[P]} \\ \vdots \\ \texttt{e<e/a-Typ><i>\_p[P]} \end{array}\right]\Bigg\}\,\texttt{i}, \quad \boldsymbol{a}_{\mathrm{FB}} = \left[\begin{array}{l} \texttt{a<e/a-Typ><i+1>\_p[P]} \\ \vdots \\ \texttt{a<e/a-Typ><i+o>\_p[P]} \end{array}\right]\Bigg\}\,\texttt{o} \\ \boldsymbol{m}_{\mathrm{FB}} &= \left[\begin{array}{l} \texttt{m<mem-Typ>1\_p[P]} \\ \vdots \\ \texttt{m<mem-Typ><m>\_p[P]} \end{array}\right]\Bigg\}\,\texttt{m}\,, \end{aligned} \tag{3.7}$$

hierbei steht der erste Buchstabe für den Eingang (e), Ausgang (a) oder Zustand (m).

Wir werden im Folgenden die Dimensionen der oben definierten Vektoren diskutieren. Die Eingangssignale des Funktionsbausteins seien Spannungen und Ströme von idealen Quellen und die Parameter für das Wellendigitalfilter. Der Vektor $\boldsymbol{e}_{\mathrm{FB}}$ mit den Eingangssignalen bildet sich folglich aus einem Vektor mit den Quellenwellen und einem Vektor der Länge n_{vP} mit variablen Parametern. Die Anzahl der Eingänge des Funktionsbausteins entspricht der Summe der Quellen und der änderbaren Parameter des Wellendigitalfilters, d. h.

$$\texttt{i} = n_q + n_{vP}\ . \tag{3.8}$$

Die Anzahl der Zustandsvariablen des Funktionsbausteins entspricht natürlich der Anzahl der dynamischen Bauelemente des Wellendigitalfilters, d.h.

$$\texttt{m} = n_v\ . \tag{3.9}$$

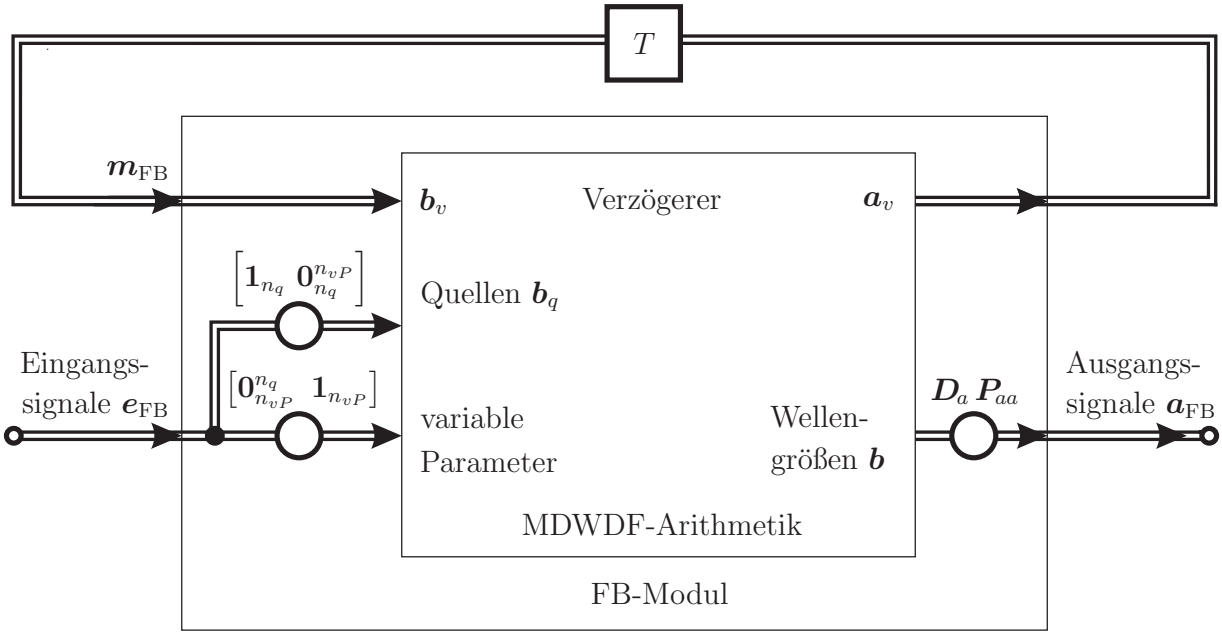


Bild 3.5: Einbettung der Wellendigitalfilter in das Programmpaket SPACE

Da die Anzahl Tore $n_g = n_q + n_v + n_e$ beträgt (wobei sich jeweils zwei Tore nur von der Richtung des Stromes und der Wellengrößen unterscheiden), können maximal n_g verschiedene Signale die Ausgangssignale des Funktionsbausteins darstellen, d.h.

$$n_{aa} \leq n_g. \tag{3.10}$$

Da das Wellendigitalfilter integraler Bestandteil des Funktionsbausteins ist, besteht nun die Notwendigkeit, diese Einbettung mathematisch zu beschreiben. Zweckmäßigerweise erfolgt die Verknüpfung der Wellengrößen $\boldsymbol{a}$ und $\boldsymbol{b}$ mit den Ein- und Ausgangssignalen $\boldsymbol{e}_{\mathrm{FB}}$ und $\boldsymbol{a}_{\mathrm{FB}}$ des Funktionsbausteins durch Matrizen. Wir bezeichnen die Verknüpfungsmatrix mit $\boldsymbol{P}_{aa}$. Das Wellendigitalfilter wird vollständig durch

$$\boldsymbol{b}_q = \left[\mathbf{1}_{n_q}\ \mathbf{0}_{n_q}^{n_{vP}}\right] \boldsymbol{e}_{\mathrm{FB}} \quad , \quad \boldsymbol{a}_{\mathrm{FB}} = \boldsymbol{D}_a\,\boldsymbol{P}_{aa} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \quad , \quad \boldsymbol{m}_{\mathrm{FB}} = \boldsymbol{b}_v \tag{3.11}$$

in einen Funktionsbaustein eingebunden. Die Einbindung der Wellendigitalfilter-Arithmetik in einen Funktionsbaustein ist im Bild 3.5 dargestellt. In dieser Arbeit wollen wir nur lineare zeit-/ortsinvariante Systeme betrachten. Ergebnisse zur Behandlung von linearen zeit-/ortsvarianten Systemen finden sich in [Juss00].

Wir werden nun die Diagonalmatrix $\boldsymbol{D}_a$ näher erläutern. Ist der Ausgang eine Spannung, so ergibt sich das Ausgangssignal als gewichtete Summe der zugehörigen Wellengrößen

$$u = (a+b)\sqrt{R} = (a'+b')/2\,. \tag{3.12}$$

Ist hingegen das Ausgangssignal ein Strom, so ergibt sich das Ausgangssignal als gewichtete Differenz der zugehörigen Wellengrößen

$$i = (a-b)/\sqrt{R} = (a'-b')/(2R)\,. \tag{3.13}$$

Der Vektor der Ausgangssignale $\boldsymbol{a}_{\mathrm{FB}}$ berechnet sich somit als Produkt einer Diagonalmatrix mit einem Vektor, der die Summen oder Differenzen der Wellengrößen beinhaltet. Bei Verwendung von Leistungswellen ist ein Element der Diagonalmatrix $\boldsymbol{D}_a$ durch die Wurzel des Torwiderstandes gegeben, wenn das

entsprechende Ausgangssignal eine Spannung ist oder im Falle des Stromes durch das Reziproke der Wurzel des Torwiderstandes. Bei Verwendung von Spannungswellen sind die Elemente von $\boldsymbol{D}_a$ entsprechend durch $1/2$ oder $1/(2R)$ gegeben.

Die Matrix $\boldsymbol{P}_{aa}$ kann durch $\boldsymbol{P}_{aa} = [\boldsymbol{P}_{aaa}\ \boldsymbol{P}_{aab}]$ dargestellt werden, wobei die Zeilenanzahl beider Untermatrizen gleich n_{aa} ist und die Spaltenanzahl gleich n_g. Die Elemente der Matrix $\boldsymbol{P}_{aaa}$ besitzen nur die Werte 0 und 1

$$p_{aaa\,\mu\nu} = \begin{cases} 1 & \text{falls der Ausgang } \mu \text{ des Funktionsbausteins eine Spannung} \\ & \text{oder einen Strom des WDF-Tores } \nu \text{ darstellt} \\ 0 & \text{sonst.} \end{cases} \tag{3.14}$$

Dagegen nehmen die Elemente von $\boldsymbol{P}_{aab}$ nicht nur 0 oder 1 an, sondern auch -1, und zwar in Abhängigkeit davon, ob das Ausgangssignal eine Spannung oder ein Strom ist

$$p_{aab\,\mu\nu} = \begin{cases} 1 & \text{falls der Ausgang } \mu \text{ des Funktionsbausteins eine Spannung} \\ & \text{des WDF-Tores } \nu \text{ ist} \\ -1 & \text{falls der Ausgang } \mu \text{ des Funktionsbausteins ein Strom des} \\ & \text{WDF-Tores } \nu \text{ ist} \\ 0 & \text{sonst.} \end{cases} \tag{3.15}$$

Jede Zeile der Matrizen $\boldsymbol{P}_{aaa}$ und $\boldsymbol{P}_{aab}$ hat nur einen Eintrag, aber eine Spalte kann mehrere Einträge besitzen, nämlich dann, wenn sowohl die Spannung als auch der Strom eines Tores ein Ausgangssignal ist.

Sämtliche einfallenden Wellen $\boldsymbol{a}_q, \boldsymbol{a}_v$ und $\boldsymbol{a}_e$ lassen sich durch die Permutationsmatrix $\boldsymbol{P}$ aus den Wellen $\boldsymbol{b}_q, \boldsymbol{b}_v$ und $\boldsymbol{b}_e$ berechnen. Folglich können die Ausgangsgrößen allein durch Kenntnis der ausfallenden Wellen $\boldsymbol{b}_q, \boldsymbol{b}_v$ und $\boldsymbol{b}_e$ ermittelt werden. Mit Gleichung (2.206) erhalten wir somit

$$\boldsymbol{a}_{\mathrm{FB}} = \boldsymbol{D}_a[\boldsymbol{P}_{aaa}\boldsymbol{a} + \boldsymbol{P}_{aab}\,\boldsymbol{b}] = \boldsymbol{D}_a[\boldsymbol{P}_{aaa}\,\boldsymbol{P} + \boldsymbol{P}_{aab}]\,\boldsymbol{b} = \underbrace{[\boldsymbol{D}_a\boldsymbol{P}_{aaa}\,\boldsymbol{P} + \boldsymbol{D}_a\boldsymbol{P}_{aab}]}_{\boldsymbol{A}}\,\boldsymbol{b}\,. \tag{3.16}$$

Kapitel 4

Syntheseverfahren

Dieses Kapitel widmet sich der systematischen Generierung mehrdimensional intern passiver Wellendigitalfilter zur numerischen Integration einer Klasse von PDGLn. In [Kumm88] wurde der Beweis erbracht, dass i. Allg. keine intern passive Schaltung zu einem drei- und höherdimensionalen Reaktanzmehrtor existiert. Da die Reaktanzmehrtore eine Untermenge passiver Mehrtore sind, existiert somit auch für die passiven Mehrtore i. Allg. keine intern passive Schaltung für drei-und mehr Dimensionen. Es ist allerdings gelungen zu vielen PDGLn eine Referenzschaltung zu finden. Da ein systematisches Vorgehen zur Durchführung der Synthese aber nicht bekannt ist, ist viel Erfahrung zur Durchführung erforderlich. Die Synthese erfordert das Aufsuchen von mehreren geeigneten Rechtsinversen zur Koordinatentransformationsmatrix unter gleichzeitiger Berücksichtigung der mehrdimensionalen konkreten Passivität einer möglichen Referenzschaltung. Anschließend wird aus der Referenzschaltung ein mehrdimensionales Wellendigitalfilter gewonnen. Die soeben erläuterte Vorgehensweise stellt die erste Phase des Softwareerstellungsprozesses dar. Diese Phase ist manuell durchzuführen. In der zweiten Phase wird das mehrdimensionale Wellendigitalfilter als graphische Spezifikation der zur Simulation verwendeten Software genutzt. Die Umsetzung der graphischen Spezifikation in eine Anweisungssequenz kann mittels Rechnerunterstützung automatisiert durchgeführt werden. Diese zweistufige Vorgehensweise bringt i.W. zwei Nachteile mit sich. Zum einen wird der Kreis der Anwender auf Sachkundige stark eingeschränkt. Zum zweiten ist die manuelle Phase mit hohen Kosten und erhöhter Gefahr eines Fehlers verbunden. Wünschenswert wäre daher ein komplett automatisierter Softwareerstellungsprozess ausgehend von der PDGL. Dieses ist aber im allgemeinen Fall nicht möglich, wie bereits erläutert wurde. Möglicherweise kann man aber die Klasse der PDGLn so einschränken, dass einerseits eine sinnvolle Klasse physikalischer Probleme behandelbar ist und andererseits ein komplett automatisierter Softwareerstellungsprozess möglich ist. Es soll daher hier ein Verfahren zur automatisierten Synthese vorgestellt werden, welches auf eine eingeschränkte Klasse PDGLn anwendbar ist, [Voll04a].

4.1 Der behandelte Typ von PDGLn

4.1.1 Einschränkungen an die PDGLn

Ausgangspunkt der Diskussion ist das folgende System nichtlinearer, nicht konstanter partieller Differentialgleichungen

$$\sum_{\xi=x,y,z,t} \sum_{\nu=1}^{N} \operatorname{sgn}(c^{\xi}_{\mu\nu}) |c^{\xi}_{\mu\nu}|^{1/2} \frac{\partial}{\partial \xi} \left[|c^{\xi}_{\mu\nu}|^{1/2} u_\nu \right] + \sum_{\nu=1}^{N} y^{k}_{\mu\nu} u_\nu = j_\mu \ , \ \mu = 1, 2, \ldots, N \ , \tag{4.1}$$

wobei die $c^{\xi}_{\mu\nu}$ und $y^{k}_{\mu\nu}$ reellwertige Funktionen von $u_1, u_2, \ldots, u_N$ und $\boldsymbol{x}$ sind. Ein derartiges System wird auch als quasilinear bezeichnet, da die Ableitungen der abhängigen Variablen nur linear auftreten.

Man beachte die Beziehung

$$|c^{\xi}_{\mu\nu}|^{1/2}\frac{\partial}{\partial\xi}\left[|c^{\xi}_{\mu\nu}|^{1/2}u_\nu\right] = \frac{1}{2}\left[|c^{\xi}_{\mu\nu}|\frac{\partial u_\nu}{\partial\xi} + \frac{\partial}{\partial\xi}\left(|c^{\xi}_{\mu\nu}|u_\nu\right)\right] . \tag{4.2}$$

Im Falle eines Kondensators stellt diese Definition den Mittelwert aus der globalen und der differentiellen Kapazität dar, [Fett98].

Die unabhängigen Variablen des partiellen Differentialgleichungs-Systems sind $\boldsymbol{x}$. Weiter definieren wir $\boldsymbol{u}$ als den Vektor der abhängigen Feldgrößen u_ν. Die eingeprägten Größen j_μ sind in dem Vektor $\boldsymbol{j}$ zusammengefasst. Im Falle konstanter $c^{\xi}_{\mu\nu}$ lautet das PDGL-System

$$\left[\boldsymbol{C}^x\mathrm{D}_x + \boldsymbol{C}^y\mathrm{D}_y + \boldsymbol{C}^z\mathrm{D}_z + \boldsymbol{C}^t\mathrm{D}_t + \boldsymbol{Y}^k\right]\boldsymbol{u} = \boldsymbol{j} , \tag{4.3}$$

wobei $c^{\xi}_{\mu\nu}$, $y^{k}_{\mu\nu}$ die reellen Elemente der Matrizen $\boldsymbol{C}^\xi$ und $\boldsymbol{Y}^k$ sind. Im weiteren Verlauf der Arbeit werden wir nur diesen Fall behandeln. Die Teilströme der einzelnen Elemente definieren wir durch $\boldsymbol{i}^x_c = \boldsymbol{C}^x\mathrm{D}_x\boldsymbol{u}$, ... , $\boldsymbol{i}^t_c = \boldsymbol{C}^t\mathrm{D}_t\boldsymbol{u}$, $\boldsymbol{i}_k = \boldsymbol{Y}^k\boldsymbol{u}$.

Die Eigenschaften der Matrizen $\boldsymbol{C}^\xi$ und $\boldsymbol{Y}^k$ werden wir im Folgenden durch Energiebetrachtungen weiter einschränken.

4.1.2 Energiebetrachtungen

Der quellenfreie Teil der PDGL, das ist ihr homogener Teil,

$$\boldsymbol{i} = \left[\boldsymbol{C}^x\mathrm{D}_x + \boldsymbol{C}^y\mathrm{D}_y + \boldsymbol{C}^z\mathrm{D}_z + \boldsymbol{C}^t\mathrm{D}_t + \boldsymbol{Y}^k\right]\boldsymbol{u} . \tag{4.4}$$

soll ein passives System beschreiben. Voraussetzung für die folgenden Energiebetrachtungen ist, dass die Koordinaten mit gleichem Index der Vektoren $\boldsymbol{u}$ und $\boldsymbol{i}$ ein Paar komplementärer Feldgrößen ist. Mit anderen Worten, jeder Summand des Skalarprodukts $p = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{i}$ muss eine Energiedichte sein. Die Tatsache, dass jeder Summand die Dimension einer Energiedichte besitzt, reicht nicht aus. Zur Erläuterung betrachten wir 2 konzentrierte Elementarkörper mit den Dichten ρ_1 und ρ_2, die sich mit den Geschwindigkeiten v_1 bzw. v_2 bewegen. Die kinetische Energiedichte ist $\rho_1 v_1^2/2 + \rho_2 v_2^2/2$. Hingegen ist $\rho_2 v_1^2/2 + \rho_1 v_2^2/2$ keine Energiedichte, obwohl die Dimensionen gleich sind.

Fassen wir die Größen $\boldsymbol{u}$ und $\boldsymbol{i}$ als Spannung und Strom in einer elektrischen Schaltung auf, so wird implizit angenommen, dass es sich hier um ein Paar komplementärer Feldgrößen handelt. Die Energiebetrachtungen beziehen sich dann aber auf die elektrische Schaltung und nicht auf das zugrunde liegende System.

Satz 1 *Der quellenfreie Teil des durch Gleichung (4.3) beschriebenen Systems ist genau dann passiv, wenn $\boldsymbol{C}^x$, $\boldsymbol{C}^y$, $\boldsymbol{C}^z$, $\boldsymbol{C}^t$ symmetrisch sind und sowohl $\boldsymbol{C}^t$ als auch der symmetrische Teil der Matrix $\boldsymbol{Y}^k$ n. n. d.* [1] *sind.*

Beweis 1

Der Beweis ist dem aus dem eindimensionalen bekannten ähnlich, vgl. [Tell54], [FH92].

Notwendigkeit

Für jeden beliebigen Vorgang ist nach Gleichung (2.85) zu zeigen, dass die über die Systemgrenze aufgenommene Energie nicht negativ ist, wenn die über die örtliche Segmentgrenze des Systems transportierte Energiedichte null ist, d. h.

$$\boldsymbol{W}^{\mathrm{T}}_{\mathrm{sp}}\boldsymbol{n} = 0 \,\text{auf}\, \partial\mathcal{G} . \tag{4.5}$$

[1] nicht negativ definit

Weiterhin soll zum Zeitpunkt t_0 im System keine Energie gespeichert sein, d. h. $E(t_0) = 0$. Mit $p = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{i}$ lautet die Bedingung

$$0 \le \int_{t_0}^{t_1} \int_{\mathcal{G}} p \, \mathrm{d}V \mathrm{d}t \quad \text{für} \quad t_1 \ge t_0 \,. \tag{4.6}$$

Da die Passivitätsforderungen für beliebige Vorgänge gelten, müssen sie auch für einen von uns speziell gewählten Vorgang gelten. Dazu nehmen für $\mathcal{G}$ einen Quader an. Setzen wir zudem $\boldsymbol{i}^{ct} = \boldsymbol{i}^{cy} = \boldsymbol{i}^{cz} = \boldsymbol{i}_k = \boldsymbol{0}$, so muss $\boldsymbol{i}^{cx} = \boldsymbol{C}^x \mathbf{D}_{\boldsymbol{x}} \boldsymbol{i}$ schon allein passiv sein.

Für die Spannungen und die Ströme $\boldsymbol{i}^{cx}$ nehmen wir 2 spezielle Vorgänge an. Für den 1.Vorgang soll $u_\sigma = 0$ und $i_\sigma = 0$ für $\mu \neq \sigma \neq \nu$ gelten. Die anderen beiden Spannungen und Ströme folgen dem Kreisprozess (mit als stetig angenommenen Leistungsdichten)

Intervall	Spannung u_μ	Spannung u_ν	$\mathrm{D}_x u_\mu$	$\mathrm{D}_x u_\nu$
$x_0 \le x \le x_1$	$u_\mu = \begin{cases} 0 & \text{für } x = x_0 \\ \text{beliebig} & \text{für } x_0 < x < x_1 \\ u_{\mu 0} > 0 & \text{für } x = x_1 \end{cases}$	$u_\nu = 0$		0
$x_1 \le x \le x_2$	$u_\mu = u_{\mu 0}$	$u_\nu = \begin{cases} 0 & \text{für } x = x_1 \\ \text{beliebig} & \text{für } x_1 < x < x_2 \\ u_{\nu 0} > 0 & \text{für } x = x_2 \end{cases}$	0	
$x_2 \le x \le x_3$	$u_\mu = \begin{cases} u_{\mu 0} & \text{für } x = x_2 \\ \text{beliebig} & \text{für } x_2 < x < x_3 \\ 0 & \text{für } x = x_3 \end{cases}$	$u_\nu = u_{\nu 0}$		0
$x_3 \le x \le x_4$	$u_\mu = 0$	$u_\nu = \begin{cases} u_{\nu 0} & \text{für } x = x_3 \\ \text{beliebig} & \text{für } x_3 < x < x_4 \\ 0 & \text{für } x = x_4 \end{cases}$	0	

der im Bild 4.1 a) graphisch dargestellt ist.

Die eingangs gestellte Bedingung Gleichung (4.5) ist wegen $\boldsymbol{u}(x_0) = \boldsymbol{u}(x_4) = \boldsymbol{0}$ erfüllt. Aufgrund des quaderförmigen örtlichen Berechnungsgebietes hat das Volumenintegral mit $t_1 = t_0 + T$ feste Grenzen

$$\int_{t_0}^{t_1} \int_V p \, \mathrm{d}V \, \mathrm{d}t = \int_{t_0}^{t_1} \int_0^Z \int_0^Y \int_{x_0}^{x_4} p \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \mathrm{d}t = [t_1 - t_0] ZY \int_{x_0}^{x_4} p \; = TZY \int_{x_0}^{x_4} p \, \mathrm{d}x \,. \tag{4.7}$$

Im Folgenden wird das innere Integral

$$\int_{x_0}^{x_4} \boldsymbol{u}^{\mathrm{T}} \, \boldsymbol{i} \, \mathrm{d}x = \int_{x_0}^{x_4} c^x_{\mu\mu} \, u_\mu \, \mathrm{D}_x u_\mu \, \mathrm{d}x + \int_{x_0}^{x_4} c^x_{\mu\nu} \, u_\mu \, \mathrm{D}_x u_\nu \, \mathrm{d}x + \int_{x_0}^{x_4} c^x_{\nu\mu} \, u_\nu \, \mathrm{D}_x u_\mu \, \mathrm{d}x + \int_{x_0}^{x_4} c^x_{\nu\nu} \, u_\nu \, \mathrm{D}_x u_\nu \, \mathrm{d}x \tag{4.8}$$

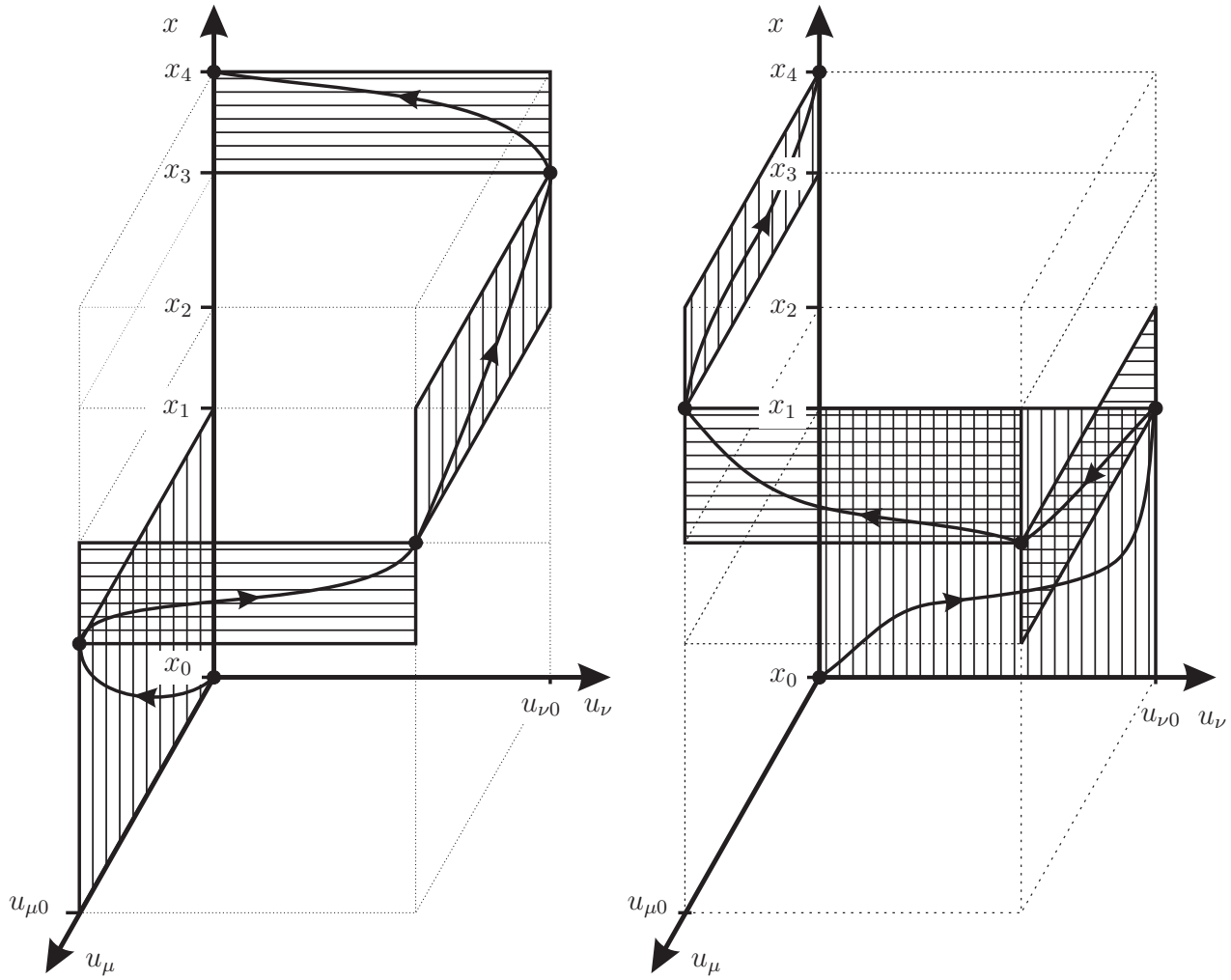


Bild 4.1: a) Erster Kreisprozess b) Zweiter Kreisprozess

untersucht. Die Auswertung der Integrale über die einzelnen Teilintervalle ergibt

$$\begin{aligned}
\int_{x_0}^{x_1} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x &= c^x_{\mu\mu}\,\frac{u^2_{\mu 0}}{2} \\
\int_{x_1}^{x_2} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x &= c^x_{\mu\nu}\,u_{\mu 0}\,u_{\nu 0} + c^x_{\nu\nu}\,\frac{u^2_{\nu 0}}{2} \\
\int_{x_2}^{x_3} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x &= -c^x_{\mu\mu}\,\frac{u^2_{\mu 0}}{2} - c^x_{\nu\mu}\,u_{\mu 0}\,u_{\nu 0} \\
\int_{x_3}^{x_4} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x &= -c^x_{\nu\nu}\,\frac{u^2_{\nu 0}}{2}\ .
\end{aligned} \tag{4.9}$$

Für den 1. speziellen Vorgang erhalten wir schließlich

$$\int_{x_0}^{x_4} p\,\mathrm{d}x = \int_{x_0}^{x_4} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x = [c^x_{\mu\nu} - c^x_{\nu\mu}]u_{\mu 0}u_{\nu 0} \geq 0\ . \tag{4.10}$$

Für den 2.Vorgang gilt $u_\sigma = 0$ und $i_\sigma = 0$ für $\mu \neq \sigma \neq \nu$. Die anderen beiden Spannungen und Ströme folgen dem ersten Kreisprozess in umgekehrter Richtung, siehe Bild 4.1 b).

Für den 2. speziellen Vorgang erhalten wir

$$\int_{x_0}^{x_4} p\,\mathrm{d}x = \int_{x_0}^{x_4} \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{i}\,\mathrm{d}x = [c^x_{\nu\mu} - c^x_{\mu\nu}]u_{\mu 0}u_{\nu 0} \geq 0\ . \tag{4.11}$$

Dieses Ergebnis hätten wir auch durch Vertauschung von μ und ν ermitteln können, da Gleichung (4.6) für beliebige Vorgänge gelten muss und die ursprüngliche Bezeichnung ja willkürlich gewählt wurde.

Mit $t_1 > t_0$, $Z > 0$, $Y > 0$, $u_{\mu 0} >$ und $u_{\nu 0} > 0$ erhalten wir aus den beiden Vorgängen die simultan zu erfüllenden Ungleichungen

$$[c^x_{\nu\mu} - c^x_{\mu\nu}] \geq 0 \quad \text{und} \quad [c^x_{\mu\nu} - c^x_{\nu\mu}] \geq 0\ . \tag{4.12}$$

Die einzige Lösung ist $c^x_{\mu\nu} = c^x_{\nu\mu}$. Da μ und ν beliebig gewählt wurden, muss $c^x_{\mu\nu} = c^x_{\nu\mu}$ für alle μ und ν gelten. Somit muss $\boldsymbol{C}^x$ symmetrisch sein. Durch ähnliches Vorgehen ergibt sich als notwendige Bedingung an $\boldsymbol{C}^y$, $\boldsymbol{C}^z$ und $\boldsymbol{C}^t$ ebenfalls die Symmetrie. Aufgrund der Symmetrie lassen sich die aufgenommenen Teilleistungsdichten durch die partielle Ableitung einer quadratischen Form ausdrücken

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^{\xi}\mathrm{D}_{\xi}\boldsymbol{u} = \frac{1}{2}\mathrm{D}_{\xi}\{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^{\xi}\,\boldsymbol{u}\} \quad \text{für} \quad \xi = x, y, z \text{ und } t\ . \tag{4.13}$$

Die gesamte aufgenommene Leistungsdichte kann daher durch

$$p = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{i}\ = \frac{1}{2}[\mathrm{D}_x\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^x\boldsymbol{u} + \mathrm{D}_y\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^y\boldsymbol{u} + \mathrm{D}_z\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^z\boldsymbol{u} + \mathrm{D}_t\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^t\boldsymbol{u}] + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u} \tag{4.14}$$

dargestellt werden. Die Energiedichte $\boldsymbol{W}_s$ ist folglich durch

$$\boldsymbol{W}_s = \frac{1}{2}\left[\boldsymbol{u}^{\mathrm{T}}\frac{\boldsymbol{C}^x}{v_4}\boldsymbol{u}\ ,\ \boldsymbol{u}^{\mathrm{T}}\frac{\boldsymbol{C}^y}{v_4}\boldsymbol{u}\ ,\ \boldsymbol{u}^{\mathrm{T}}\frac{\boldsymbol{C}^z}{v_4}\boldsymbol{u}\ ,\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}^t\boldsymbol{u}\right]^{\mathrm{T}} \tag{4.15}$$

gegeben. Das Integral

$$\begin{aligned}
\int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}V\mathrm{d}t &= v_4\int_{\mathcal{G}}\int_{t_0}^{t_1}\mathbf{D}^{\mathrm{T}}_{\boldsymbol{x}}\boldsymbol{W}_s\,\mathrm{d}V\mathrm{d}t + \int_{\mathcal{G}}\int_{t_0}^{t_1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u}\,\mathrm{d}V\mathrm{d}t \\
&= v_4\int_{t_0}^{t_1}\int_{\mathcal{G}}\operatorname{div}\boldsymbol{W}_{\mathrm{sp}}\mathrm{d}V\mathrm{d}t + \int_{\mathcal{G}}\int_{t_0}^{t_1}\mathrm{D}_t\,W_t\,\mathrm{d}t\,\mathrm{d}V + \int_{\mathcal{G}}\int_{t_0}^{t_1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u}\,\mathrm{d}V\mathrm{d}t \\
&= v_4\int_{t_0}^{t_1}\int_{\partial\mathcal{G}}\boldsymbol{W}^{\mathrm{T}}_{\mathrm{sp}}\boldsymbol{n}\,\mathrm{d}A\,\mathrm{d}t\ + \int_{\mathcal{G}}\int_{t_0}^{t_1}\mathrm{D}_t\,W_t\,\mathrm{d}t\,\mathrm{d}V + \int_{\mathcal{G}}\int_{t_0}^{t_1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u}\,\mathrm{d}V\mathrm{d}t \\
&= \int_{\mathcal{G}}\left[\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}(t_1)\boldsymbol{C}^t\boldsymbol{u}(t_1) \quad + \int_{t_0}^{t_1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u}\,\mathrm{d}t\right]\mathrm{d}V
\end{aligned} \tag{4.16}$$

darf nach Gleichung (2.85) für beliebige Vorgänge $\boldsymbol{u}$ und $t_1 \geq t_0$ nicht negativ werden. Dies ist nur dann der Fall, wenn die Matrizen $\boldsymbol{C}^t$ und $\boldsymbol{Y}^k$ n. n. d. sind.

Hinlänglichkeit

Nach Gleichung (2.78) müssen wir zeigen, dass die Differenz aus der über die Tore aufgenommenen Energie und der über die Gebietsgrenze aufgenommenen Energie nicht kleiner als null ist. Bilden dieser Differenz liefert

$$\int_{\mathcal{G}_{\boldsymbol{x}}} p\,\mathrm{d}V\,\mathrm{d}t - v_4\int_{\mathcal{G}_{\boldsymbol{x}}}\mathbf{D}^{\mathrm{T}}_{\boldsymbol{x}}\boldsymbol{W}_s\,\mathrm{d}V\,\mathrm{d}t = \int_{\mathcal{G}}\int_{t_0}^{t_1}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Y}^k\boldsymbol{u}\,\mathrm{d}t\,\mathrm{d}V\ . \tag{4.17}$$

Der Integrand ist aufgrund der nicht negativen Definitheit von $\boldsymbol{Y}^k$ nicht negativ. Wegen $t_1 \geq t_0$ und der Tatsache, dass über ein Volumen integriert wird, ist die Differenz für alle Vorgänge nicht negativ. Damit ist der Satz bewiesen.

Abschließend wollen wir noch einmal die Ergebnisse zusammentragen. Notwendig und hinreichend für die Passivität sind symmetrische Matrizen $\boldsymbol{C}^t$, $\boldsymbol{C}^x$, $\boldsymbol{C}^y$ und $\boldsymbol{C}^z$ sowie nicht negativ definite Matrizen $\boldsymbol{C}^t$ und $\boldsymbol{Y}^k$.

4.1.3 Systemklassifizierung

Ein partielles Differentialgleichungssystem Gleichung (4.3) wird nach [Frie54] symmetrisch hyperbolisches System bezeichnet, wenn die Matrizen $\boldsymbol{C}^x$, $\boldsymbol{C}^y$, $\boldsymbol{C}^z$ und $\boldsymbol{C}^t$ symmetrisch sind und eine von ihnen positiv definit ist.

Wie wir später sehen werden, benötigt das in diesem Kapitel vorgestellte Syntheseverfahren eine positiv definite Matrix $\boldsymbol{C}^t$. Es liegt daher nahe, die aus der Passivität resultierende Forderung an $\boldsymbol{C}^t$ von der nicht negativen Definitheit auf die positive Definitheit zu verstärken. Das vorliegende partielle Differentialgleichungssystem gehört dann bei beliebigen symmetrischen Matrizen $\boldsymbol{C}^x$, $\boldsymbol{C}^y$ und $\boldsymbol{C}^z$ zur Klasse der symmetrisch hyperbolischen.

Wir machen noch eine weitere Einschränkung an die Matrizen $\boldsymbol{C}^x$, $\boldsymbol{C}^y$ und $\boldsymbol{C}^z$. Wir fordern, dass deren Hauptdiagonalelemente verschwinden. Eine Synthese und anschließende Umsetzung in eine WD-Struktur ließe sich zwar auch bei nichtverschwindenden Hauptdiagonalen erreichen, aber bislang konnte keine Methode zur Randbehandlung dieser Anteile gefunden werden. Eine erwähnenswerte Eigenschaft einer reellen, symmetrischen Matrix $\boldsymbol{M}$ mit verschwindender Hauptdiagonale ist es, indefinit zu sein (abgesehen von der Nullmatrix). Die Summe der Eigenwerte ist nämlich gleich der Spur der Matrix und diese verschwindet, d. h. spur $\boldsymbol{M} = \sum \lambda_\nu = 0$. Es existieren daher mindestens zwei (reelle) Eigenwerte mit unterschiedlichem Vorzeichen, d.h. die Matrix ist indefinit.

4.1.4 Eingeschwungener Zustand

Aufgrund der Linearität existiert bei Anregung der Form

$$\boldsymbol{j} = \mathrm{Re}\{\boldsymbol{J}\,\mathrm{e}^{\boldsymbol{p}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}}\}\,, \tag{4.18}$$

bis auf endlich viele Frequenzen $\boldsymbol{p}_{\boldsymbol{x}}$, eine Lösung der Form

$$\boldsymbol{u} = \mathrm{Re}\{\boldsymbol{U}\,\mathrm{e}^{\boldsymbol{p}_{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}}\}\,, \tag{4.19}$$

wobei $\boldsymbol{J}$ und $\boldsymbol{U}$ Vektoren mit komplexen konstanten Koordinaten sind. Bei Betrachtung des stationären Zustands und Verwendung der komplexen Größen anstatt der zeit- und ortsabhängigen Größen tritt die Frequenzvariable $\boldsymbol{p}_{\boldsymbol{x}}$ an die Stelle des Differentialoperators $\mathbf{D}_{\boldsymbol{x}}$. Im stationärem Zustand geht Gleichung (4.3) in

$$\left[\boldsymbol{C}^x p_x + \boldsymbol{C}^y p_y + \boldsymbol{C}^z p_z + \boldsymbol{C}^t p_t + \boldsymbol{Y}^k\right]\boldsymbol{U} = \boldsymbol{J} \tag{4.20}$$

über. Mit der Interpretation der Koordinaten von $\boldsymbol{u}$ als Spannungen und der Koordinaten von $\boldsymbol{j}$ als eingeprägte Ströme stellen

$$\boldsymbol{Y}^{cx}(p_x) = p_x\boldsymbol{C}^x\,, \boldsymbol{Y}^{cy}(p_y) = p_y\boldsymbol{C}^y\,, \boldsymbol{Y}^{cz}(p_z) = p_z\boldsymbol{C}^z \text{ und } \boldsymbol{Y}^{ct}(p_t) = p_t\boldsymbol{C}^t \tag{4.21}$$

Admittanzmatrizen dar. Der reaktive Teil besitzt die Admittanzmatrix

$$\boldsymbol{Y}^c(\boldsymbol{p}_x) = \boldsymbol{Y}^{cx}(p_x) + \boldsymbol{Y}^{cy}(p_y) + \boldsymbol{Y}^{cz}(p_z) + \boldsymbol{Y}^{ct}(p_t)\,. \tag{4.22}$$

Die durch $\boldsymbol{Y}(\boldsymbol{p}_x)\boldsymbol{U} = \boldsymbol{J}$ festgelegte gesamte Admittanzmatrix $\boldsymbol{Y}(\boldsymbol{p}_x)$ ist die Summe aus dem reaktiven und dem konstanten Teil $\boldsymbol{Y}(\boldsymbol{p}_x) = \boldsymbol{Y}^c(\boldsymbol{p}_x) + \boldsymbol{Y}^k$.

Wir wollen nun untersuchen, wie die Einschränkungen an die partielle Differentialgleichung sich im Frequenzbereich auswirken und welche Eigenschaften daraus für die zuvor definierten Admittanzmatrizen resultieren. Da die partielle Differentialgleichung ein reelles System beschreiben soll, gilt $\boldsymbol{C}^x$, $\boldsymbol{C}^y$, $\boldsymbol{C}^z$, $\boldsymbol{C}^t$, $\boldsymbol{Y}^k \in \mathbb{R}^{N\times N}$. Folglich ist jedes Element der Admittanzmatrix gleich dem bikonjugierten Element, d. h. $\boldsymbol{Y}(\boldsymbol{p}_x) = \boldsymbol{Y}^*(\boldsymbol{p}_x^*)$. Die Admittanzmatrix $\boldsymbol{Y}^c$ soll ein verlustfreies System beschreiben. Somit ist die Summe aus der Matrix und ihrer parakonjugierten Matrix gleich der Nullmatrix, d.h. $\boldsymbol{Y}^c(\boldsymbol{p}_x) + \boldsymbol{Y}^c_*(\boldsymbol{p}_x) = \boldsymbol{0}$, [Bele68], [Bose82].

Offenbar ist die Admittanzmatrix des reaktiven Teils eine ungerade Funktion in $\boldsymbol{p_x}$. Aus dieser Beziehung $\boldsymbol{Y}^c(\boldsymbol{p}_x) = -\boldsymbol{Y}^c(-\boldsymbol{p}_x)$, den Eigenschaften eines reellen Systems und den Eigenschaften eines verlustfreien Systems, folgt das schon aus dem Orts-/Zeitbereich bekannte Ergebnis, dass der reaktive Teil durch eine symmetrische Matrix beschrieben werden kann

$$\boldsymbol{Y}^c(\boldsymbol{p}_x) = \boldsymbol{Y}^{c\,\mathrm{T}}(\boldsymbol{p}_x)\,. \tag{4.23}$$

Die partielle Differentialgleichung soll ein extern passives, symmetrisch hyperbolisches System beschreiben. Im Koordinatensystem $\boldsymbol{x}$ ist $\boldsymbol{Y}^k$ eine positive Matrix in der Frequenzvariablen p_t, d. h. $\boldsymbol{Y}^k$ ist analytisch und der symmetrische Teil ist n.n.d. für $\mathrm{Re}\{p_t\} > 0$

$$\begin{array}{ll} \boldsymbol{Y}^k & \text{analytisch in } \mathrm{Re}\{p_t\} > 0 \\ \boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}} \geq 0 & \text{in } \mathrm{Re}\{p_t\} > 0\,. \end{array} \tag{4.24}$$

Für $\boldsymbol{Y}^{ct}$ gilt hingegen die schärfere Bedingung

$$\begin{array}{ll} \boldsymbol{Y}^{ct} & \text{analytisch in } \mathrm{Re}\{p_t\} > 0 \\ \boldsymbol{Y}^{ct} + \boldsymbol{Y}^{ct\mathrm{T}} > 0 & \text{in } \mathrm{Re}\{p_t\} > 0\,. \end{array} \tag{4.25}$$

Die Matrizen $\boldsymbol{Y}^{cx}$, $\boldsymbol{Y}^{cy}$ und $\boldsymbol{Y}^{cz}$ unterliegen bis auf die schon festgelegte Symmetrie keinen weiteren Restriktionen.

Im Koordinatensystem $\boldsymbol{t}$ ergeben sich für $\boldsymbol{Y}$ die folgenden notwendigen und hinreichenden Eigenschaften für die MD-Passivität. $\boldsymbol{Y}$ ist eine positive Matrix in der Variablen $\boldsymbol{p}_t$, d. h. $\boldsymbol{Y}$ ist analytisch und $\boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}}$ ist n.n.d. in der rechten offenen Polyhalbebene $\mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0}$, d. h.

$$\begin{array}{ll} \boldsymbol{Y} & \text{analytisch in } \mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0} \\ \boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}} \geq 0 & \text{in } \mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0}\,. \end{array} \tag{4.26}$$

Die letzte Eigenschaft setzt allerdings die Wahl einer geeigneten Konstanten v_4 voraus. Wir werden später auf die Wahl der Konstanten und den Beweis der Äquivalenz der Aussagen in den beiden Koordinatensystemen zu sprechen kommen.

4.2 Synthese der Referenzschaltung

4.2.1 Vorbetrachtungen

Jede quadratische Matrix $\boldsymbol{M}$ mit den Elementen $m_{\mu\nu}$ und N Zeilen kann durch

$$\boldsymbol{M} = \sum_{\mu=1}^{N}\sum_{\nu=1}^{N} \mathbf{e}_\mu\,\mathbf{e}_\nu^{\mathrm{T}}\,m_{\mu\nu} \tag{4.27}$$

dargestellt werden. Aus der Summe spalten wir die Summe mit den Hauptdiagonalelementen ab

$$\boldsymbol{M} = \sum_{\mu=1}^{N} \mathbf{e}_\mu \, \mathbf{e}_\mu^{\mathrm{T}} \, m_{\mu\mu} \; + \sum_{\mu=1}^{N} \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{N} \mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \, m_{\mu\nu} \, . \tag{4.28}$$

Den zweiten Teil zerlegen wir in die Summe einer strikten linken unteren Dreiecksmatrix und einer strikten rechten oberen Dreiecksmatrix

$$\sum_{\mu=1}^{N} \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{N} \mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \, m_{\mu\nu} = \underbrace{\sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \, m_{\mu\nu}}_{\text{strikte linke untere Dreiecksmatrix}} + \underbrace{\sum_{\nu=2}^{N} \sum_{\mu=1}^{\nu-1} \mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \, m_{\mu\nu}}_{\text{strikte rechte obere Dreiecksmatrix}} \, . \tag{4.29}$$

Die innere Summe der linken unteren Dreiecksmatrix hat $\mu - 1$ Summanden. Somit lautet die Zahl der Summanden der Doppelsumme

$$\sum_{\mu=2}^{N} \mu - 1 = \sum_{\mu=1}^{N-1} \mu = \frac{N-1}{2}[(N-1)+1] = \frac{N^2-N}{2} = \frac{N(N-1)}{2} = \frac{N!}{2!(N-2)!} = \binom{N}{2} . \tag{4.30}$$

Dies entspricht der Hälfte aus der Differenz der Zahl der Matrixelemente und der Zahl der Hauptdiagonalelemente.

Wir nehmen im Folgenden eine symmetrische Matrix $\boldsymbol{M}$ an. Die strikte rechte obere Dreiecksmatrix formen wir unter Ausnutzung von $m_{\mu\nu} = m_{\nu\mu}$ und Tauschen der (bzgl. der Summen lokalen) Indizes um, d.h.

$$\sum_{\nu=2}^{N} \sum_{\mu=1}^{\nu-1} \mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \, m_{\mu\nu} \; = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \mathbf{e}_\nu \, \mathbf{e}_\mu^{\mathrm{T}} \, m_{\mu\nu} \, . \tag{4.31}$$

Die Matrix $\boldsymbol{M}$ lautet dann

$$\boldsymbol{M} = \sum_{\mu=1}^{N} \mathbf{e}_\mu \, \mathbf{e}_\mu^{\mathrm{T}} \, m_{\mu\mu} \; + \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \; + \mathbf{e}_\nu \, \mathbf{e}_\mu^{\mathrm{T}}] \, m_{\mu\nu} \, . \tag{4.32}$$

Wir werden im Folgenden die Summe der Dyaden $\mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}}$ und $\mathbf{e}_\nu \, \mathbf{e}_\mu^{\mathrm{T}}$ aufbereiten

$$\mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \; + \; \mathbf{e}_\nu \, \mathbf{e}_\mu^{\mathrm{T}} = \begin{bmatrix} \mathbf{e}_\mu & \mathbf{e}_\nu \end{bmatrix} \begin{bmatrix} \mathbf{e}_\nu^{\mathrm{T}} \\ \\ \mathbf{e}_\mu^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_\mu & \mathbf{e}_\nu \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_\mu^{\mathrm{T}} \\ \\ \mathbf{e}_\nu^{\mathrm{T}} \end{bmatrix} . \tag{4.33}$$

Wir definierten nun eine Matrix $\boldsymbol{T}_{\mu\nu} = [\mathbf{e}_\mu \; \mathbf{e}_\nu]$, sodass

$$\mathbf{e}_\mu \, \mathbf{e}_\nu^{\mathrm{T}} \; + \; \mathbf{e}_\nu \, \mathbf{e}_\mu^{\mathrm{T}} = \boldsymbol{T}_{\mu\nu} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \tag{4.34}$$

gilt. In diesem Zusammenhang nutzen wir die folgende Zerlegung, die einer Eigenwertzerlegung ähnlich ist

$$2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\boldsymbol{S}_{0=}} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} , \tag{4.35}$$

sodass

$$\mathbf{e}_\mu \mathbf{e}_\nu^{\mathrm{T}} + \mathbf{e}_\nu \mathbf{e}_\mu^{\mathrm{T}} = \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} -1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \tag{4.36}$$

resultiert. Die mit $\boldsymbol{S}_0$ definierte Hadamard-Matrix wird im weiteren Verlauf noch eine große Rolle spielen. Weiterhin werden wir später die leicht verifizierbare Beziehung der speziellen Diagonalmatrix

$$\mathbf{e}_\mu \mathbf{e}_\mu^{\mathrm{T}} + \mathbf{e}_\nu \mathbf{e}_\nu^{\mathrm{T}} = \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \tag{4.37}$$

benötigen.

4.2.2 Synthese des reaktiven Teils

Ziel der weiteren Überlegungen ist es, die Matrix $\boldsymbol{Y}^c$ als Knotenleitwertmatrix einer mehrdimensional konkret passiven Schaltung darzustellen, d.h.

$$\boldsymbol{Y}^c = \boldsymbol{A}\, \boldsymbol{Y}_{\mathrm{D}}^c\, \boldsymbol{A}^{\mathrm{T}} . \tag{4.38}$$

Hierin ist $\boldsymbol{Y}_{\mathrm{D}}^c$ eine Diagonalmatrix mit den Elementen $p_\kappa C_\sigma\,, \kappa = 1, 2, \ldots 7\,, C_\sigma \geq 0$. $\boldsymbol{A}$ ist die Knoten-Zweig-Inzidenzmatrix eines verallgemeinerten Verbindungsnetzes. Es wird sich später herausstellen, dass das verallgemeinerte Verbindungsnetz mittels eines Verbindungsnetzes und Zweitorübertragern mit Übersetzungverhältnis $-1/1$ realisiert werden kann. Da die Elemente von $\boldsymbol{Y}_{\mathrm{D}}^c$ positive Funktionen sind, stellt $\boldsymbol{Y}_{\mathrm{D}}^c$ eine positive Matrix dar. Weiterhin ist sichergestellt, dass $\boldsymbol{Y}^c$ eine positive Matrix ist, weil die nicht negative Definitheit invariant bzgl. der Kongruenztransformation ist.

Unter den gemachten Voraussetzungen lautet die Matrix $\boldsymbol{Y}^{cx}$

$$\boldsymbol{Y}^{cx} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\, \mathbf{e}_\mu \mathbf{e}_\nu^{\mathrm{T}} + \mathbf{e}_\nu \mathbf{e}_\mu^{\mathrm{T}} \,]\, p_x\, c_{\mu\nu}^x. \tag{4.39}$$

Mit den Vorbemerkungen kann nun $\boldsymbol{Y}^{cx}$ durch Nutzen von Gleichung (4.36) und $c_{\mu\nu}^x = |c_{\mu\nu}^x| \operatorname{sgn}(c_{\mu\nu}^x)$ wie folgt dargestellt werden

$$\boldsymbol{Y}^{cx} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} -p_x \frac{|c_{\mu\nu}^x| \operatorname{sgn}(c_{\mu\nu}^x)}{2} & 0 \\ 0 & p_x \frac{|c_{\mu\nu}^x| \operatorname{sgn}(c_{\mu\nu}^x)}{2} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} . \tag{4.40}$$

Nun definieren wir die Diagonalmatrizen $\boldsymbol{Y}^{ct1,4}$ und $\boldsymbol{C}^{t1,4}$

$$\boldsymbol{Y}^{ct1,4} = p_t \boldsymbol{C}^{t1,4} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\mathbf{e}_\mu \mathbf{e}_\mu^{\mathrm{T}} + \mathbf{e}_\nu \mathbf{e}_\nu^{\mathrm{T}}]\, \frac{p_t}{v_4}\, |c_{\mu\nu}^x| = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}}\, \frac{p_t}{v_4}\, |c_{\mu\nu}^x| \,, \tag{4.41}$$

wobei die letzte Identität über Gleichung (4.37) zu gewinnen ist[2]. Die Matrix $\boldsymbol{Y}^{ct1,4}$ wird nun zu $\boldsymbol{Y}^{cx}$ aus Gleichung (4.40) addiert

$$\boldsymbol{Y}^{c1,4} = \boldsymbol{Y}^{cx} + \boldsymbol{Y}^{ct1,4} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} \frac{p_t}{v_4} - p_x \operatorname{sgn}(c_{\mu\nu}^x) & 0 \\ 0 & \frac{p_t}{v_4} + p_x \operatorname{sgn}(c_{\mu\nu}^x) \end{bmatrix} \frac{|c_{\mu\nu}^x|}{2}\, \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} . \tag{4.42}$$

[2]Die Indizes 1 und 4 stellen einen Bezug zu den Koordinaten 1 und 4 der mehrdimensionalen Zeit her.

Mögliche Darstellungen von $p_t/v_4 - p_x$ und $p_t/v_4 + p_x$ sind $p_t/v_4 - p_x = p_4/v_0$ bzw. $p_t/v_4 + p_x = p_1/v_0$. Diese Darstellungen entsprechen der Wahl zweier Spaltenvektoren einer Rechtsinversen von $\boldsymbol{H}$. Damit haben wir in den neuen Koordinaten

$$\boldsymbol{Y}^{c1,4} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} \left\{ {p_4 \atop p_1} \right\} \frac{|c^x_{\mu\nu}|}{2v_0} & 0 \\ 0 & \left\{ {p_1 \atop p_4} \right\} \frac{|c^x_{\mu\nu}|}{2v_0} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} . \tag{4.43}$$

Hier und im Folgenden sind die geschweiften Klammern so zu interpretieren, dass die obere (untere) Alternative zu nehmen ist, wenn die Signum-Funktion $\mathrm{sgn}(c^x_{\mu\nu})$ positiv (negativ) ist. Eine formale Darstellung kann mithilfe der Einheitssprungfunktion erreicht werden

$$\left\{ {p_4 \atop p_1} \right\} = p_4 \, \mathrm{u}(c^x_{\mu\nu}) + p_1 \, \mathrm{u}(-c^x_{\mu\nu}) \, . \tag{4.44}$$

Wir definieren

$$C^{\xi}_{\mu\nu} = \frac{2|c^{\xi}_{\mu\nu}|}{v_0} \quad \text{für} \quad \xi = x, y, z \tag{4.45}$$

die später Kapazitäten von Kondensatoren sind. Dann haben wir

$$\boldsymbol{Y}^{c1,4} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} \left\{ {p_4 \atop p_1} \right\} \frac{C^x_{\mu\nu}}{4} & 0 \\ 0 & \left\{ {p_1 \atop p_4} \right\} \frac{C^x_{\mu\nu}}{4} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} , \tag{4.46}$$

vgl. dazu Gleichung (2.160). Ebenso verfährt man mit den anderen Richtungen, die die Matrizen $\boldsymbol{Y}^{c2,5}$ und $\boldsymbol{Y}^{c3,6}$ definieren. Da wir die ursprüngliche Matrix $\boldsymbol{Y}^{c}$ erhalten wollen, erfolgt eine Kompensation der Additionen der Matrix $\boldsymbol{Y}^{ct}$, und zwar wie folgt

$$\boldsymbol{Y}^{c} = \boldsymbol{Y}^{c1,4} + \boldsymbol{Y}^{c2,5} + \boldsymbol{Y}^{c3,6} + \underbrace{\boldsymbol{Y}^{ct} \; - \; \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\mathbf{e}_\mu \, \mathbf{e}_\mu^{\mathrm{T}} + \mathbf{e}_\nu \, \mathbf{e}_\nu^{\mathrm{T}}] \, \frac{p_t}{v_4} \, [|c^x_{\mu\nu}| + |c^y_{\mu\nu}| + |c^z_{\mu\nu}|]}_{\boldsymbol{Y}^{c7} = p_7 \boldsymbol{C}^7 = p_t \frac{v_0}{v_4} \boldsymbol{C}^7} \, . \tag{4.47}$$

Durch Wahl einer hinreichend großen Konstanten v_4 kann die nicht negative Definitheit von

$$\boldsymbol{C}^7 = \frac{v_4}{v_0} \boldsymbol{C}^t \; - \; \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\mathbf{e}_\mu \, \mathbf{e}_\mu^{\mathrm{T}} + \mathbf{e}_\nu \, \mathbf{e}_\nu^{\mathrm{T}}] \, \frac{1}{v_0} \, [|c^x_{\mu\nu}| + |c^y_{\mu\nu}| + |c^z_{\mu\nu}|] \, , \tag{4.48}$$

die für die MD-Passivität notwendig und hinreichend ist, erreicht werden. Zum Beweis berücksichtigen wir, dass die Eigenwerte einer symmetrischen reellen p.d. Matrix $\boldsymbol{C}^t$ reell und positiv sind. Die Eigenwerte einer symmetrischen reellen n.n.d. Matrix

$$\boldsymbol{C} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} [\mathbf{e}_\mu \, \mathbf{e}_\mu^{\mathrm{T}} + \mathbf{e}_\nu \, \mathbf{e}_\nu^{\mathrm{T}}] \, \frac{1}{v_0} \, [|c^x_{\mu\nu}| + |c^y_{\mu\nu}| + |c^z_{\mu\nu}|] \tag{4.49}$$

sind reell und nicht negativ. Weiterhin besitzt der Rayleigh-Quotient einer reellen symmetrischen Matrix als untere (obere) Schranke ihren kleinsten (größten) Eigenwert. Der kleinste (größte) Eigenwert der Matrix $\boldsymbol{C}^t$ ($\boldsymbol{C}$) soll $\lambda^{Ct}_{\min}$ ($\lambda^{C}_{\max}$) sein. Dann gilt für $\boldsymbol{x} \neq \mathbf{0}$

$$\frac{1}{\lambda^{Ct}_{\min}} \frac{\boldsymbol{x}^{\mathrm{H}} \boldsymbol{C}^t \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \; \geq \; 1 \; \geq \; \frac{1}{\lambda^{C}_{\max}} \frac{\boldsymbol{x}^{\mathrm{H}} \boldsymbol{C} \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \quad \Longleftrightarrow \quad \boldsymbol{x}^{\mathrm{H}} \left[\frac{\lambda^{C}_{\max}}{\lambda^{Ct}_{\min}} \boldsymbol{C}^t \; - \; \boldsymbol{C} \right] \boldsymbol{x} \; \geq \; 0 \, . \tag{4.50}$$

Wir müssen somit

$$v_4 \;\geq\; v_0 \,\frac{\lambda^{C}_{\max}}{\lambda^{C^t}_{\min}} \tag{4.51}$$

wählen um die mehrdimensionale Passivität zu garantieren.

Die gesamte Admittanzmatrix des reaktiven Teils kann nun in Abhängigkeit der Frequenzvariablen des neuen Koordinatensystems durch

$$\boldsymbol{Y}^{c} = \sum_{\kappa=1}^{7} \boldsymbol{C}^{\kappa} p_{\kappa} \tag{4.52}$$

ausgedrückt werden.

Durch Darstellen der Doppelsumme in Gleichung (4.46) mittels eines Skalarprodukts

$$\boldsymbol{Y}^{c1,4} = \underbrace{[\boldsymbol{T}_{21}\boldsymbol{S}_0, \ldots, \boldsymbol{T}_{N,N-1}\boldsymbol{S}_0]}_{\boldsymbol{A}} \underbrace{\mathbf{diag}\left(\left\{ \begin{matrix} p_4 \\ p_1 \end{matrix} \right\} \frac{C^x_{21}}{4}, \ldots, \left\{ \begin{matrix} p_1 \\ p_4 \end{matrix} \right\} \frac{C^x_{N,N-1}}{4} \right)}_{\boldsymbol{Y}^{c1,4}_{\mathrm{D}}} [\boldsymbol{T}_{21}\boldsymbol{S}_0, \ldots, \boldsymbol{T}_{N,N-1}\boldsymbol{S}_0]^{\mathrm{T}} \tag{4.53}$$

haben wir für die Matrix $\boldsymbol{Y}^{c1,4}$ die Struktur als Knotenleitwertmatrix gewonnen, die für die weiteren Überlegungen günstig ist. In der Regel sind $\boldsymbol{C}^x$, $\boldsymbol{C}^y$, $\boldsymbol{C}^z$, $\boldsymbol{C}^t$ und $\boldsymbol{Y}^k$ lichte Matrizen. Aus diesem Grunde kann eine Reduktion der Matrizen durch Streichen der verschwindenden Einträge in $\boldsymbol{Y}^{c1,4}_{\mathrm{D}}$ durchgeführt werden, d. h.

$$\boldsymbol{Y}^{c1,4} = \boldsymbol{A}^{x}\, \boldsymbol{Y}^{c1,4}_{\mathrm{D,red}}\, \boldsymbol{A}^{x\mathrm{T}} \,, \tag{4.54}$$

mit regulärer Diagonalmatrix $\boldsymbol{Y}^{c1,4}_{\mathrm{D,red}}$. Wir werden nun die Matrix des verallgemeinerten Verbindungsnetzes wie folgt aufteilen

$$\boldsymbol{A} = [\boldsymbol{T}_{21}\,, \ldots, \boldsymbol{T}_{N,N-1}]\,\mathbf{diag}(\boldsymbol{S}_0\,, \ldots, \boldsymbol{S}_0) = \boldsymbol{A}_{\mathrm{rvn}}\,\mathbf{diag}(\boldsymbol{S}_0\,, \ldots, \boldsymbol{S}_0)\,. \tag{4.55}$$

Die Matrix $\boldsymbol{A}_{\mathrm{rvn}}$ hat N Zeilen, $N^2{-}N$ Spalten und ist Knoten-Zweig-Inzidenzmatrix eines topologischen Verbindungsnetzes. (Da $\boldsymbol{Y}^{c1,4}$ als Summe von Matrizen ausdrückbar ist, handelt es sich um eine Parallelschaltung.) Die zweite Matrix $\mathbf{diag}\{\boldsymbol{S}_0\,, \ldots, \boldsymbol{S}_0\}$ wird als das Übersetzungsverhältnis von Jaumann-Adaptoren interpretiert, die wiederum selber aufgrund der speziellen Struktur des Übersetzungsverhältnises durch $\binom{N}{2}$ 2-Tor-Übertrager mit Übersetzungsverhältnis –1/1 realisiert werden können. Um dies zu erläutern stellen wir $\boldsymbol{Y}^{c1,4}$ aus Gleichung (4.46) durch

$$\boldsymbol{Y}^{c1,4} = \boldsymbol{A}_{\mathrm{rvn}}\,\mathbf{diag}(\boldsymbol{Y}^{c1,4}_{21}, \ldots, \boldsymbol{Y}^{c1,4}_{N,N-1})\boldsymbol{A}^{\mathrm{T}}_{\mathrm{rvn}} \quad \text{mit} \quad \boldsymbol{Y}^{c1,4}_{\mu\nu} = \boldsymbol{S}_0 \begin{bmatrix} \left\{ \begin{matrix} p_4 \\ p_1 \end{matrix} \right\} \frac{C^x_{\mu\nu}}{4} & 0 \\ 0 & \left\{ \begin{matrix} p_1 \\ p_4 \end{matrix} \right\} \frac{C^x_{\mu\nu}}{4} \end{bmatrix} \boldsymbol{S}^{\mathrm{T}}_0 \tag{4.56}$$

dar. Die Admittanzmatrix $\boldsymbol{Y}^{c1,4}_{\mu\nu}$ entspricht der Admittanzmatrix aus Gleichung (2.160), bei Wahl von $\frac{1}{Z_3} = \left\{ \begin{matrix} p_4 \\ p_1 \end{matrix} \right\} C^x_{\mu\nu}$, $\frac{1}{Z_4} = \left\{ \begin{matrix} p_1 \\ p_4 \end{matrix} \right\} C^x_{\mu\nu}$, $n = 1$ und $\boldsymbol{S}_0 = 2\boldsymbol{N}^{\mathrm{T}}$. Die Realisierung der Jaumann-Schaltung ist im Bild 2.21 angegeben.

Es verbleibt, eine Realisierung der Matrix $\boldsymbol{Y}^{c7}$ zu finden. Wir suchen ebenfalls eine Zerlegung der Form

$$\boldsymbol{Y}^{c7} = \boldsymbol{A}^{7}\, \boldsymbol{Y}^{c7}_{\mathrm{D}}\, \boldsymbol{A}^{7\mathrm{T}} \,. \tag{4.57}$$

Um eine Aufwandsreduzierung zu erreichen ist es sinnvoll, einen möglicherweise vorhandenen Diagonalteil der Matrix $\boldsymbol{Y}^{c7}$ abzuspalten. Zu diesem Zweck sortieren wir die Zeilen und Spalten $\boldsymbol{C}^7$ mithilfe der Permutationsmatrix $\boldsymbol{P}^7$ zu

$$\boldsymbol{C}^7 = \boldsymbol{P}^7 \begin{bmatrix} \bar{\boldsymbol{C}}^7 & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{C}}^7 \end{bmatrix} \boldsymbol{P}^{7\mathrm{T}} \tag{4.58}$$

um, wobei $\tilde{\boldsymbol{C}}^7$ eine Diagonalmatrix mit Kapazitäten ist. Die verbleibende Matrix $\bar{\boldsymbol{C}}^7$ hat die Dimension $\bar{n} \times \bar{n}$. Da $\bar{\boldsymbol{C}}^7$ reell und symmetrisch ist, existiert eine zugehörige reelle, orthogonale Matrix $\boldsymbol{N} = \boldsymbol{N}^{-\mathrm{T}}$ der Eigenwertzerlegung

$$\bar{\boldsymbol{C}}^7 = \boldsymbol{N}^{\mathrm{T}}\, \mathbf{diag}(\bar{C}_1, \ldots, \bar{C}_{\bar{n}})\boldsymbol{N} = \boldsymbol{N}^{\mathrm{T}} \bar{\boldsymbol{C}}^7_{\mathrm{D}} \boldsymbol{N}\ . \tag{4.59}$$

Hierbei sind aufgrund der nicht negativen Definitheit die Eigenwerte nicht negativ, $\bar{C}_\nu \geq 0$. Die Eigenwerte sind Kapazitäten von Kondensatoren. Wir fassen nun $\boldsymbol{N}$ als das Übersetzungsverhältnis eines idealen $2\bar{n}$-Tor Übertragers gemäß Gleichung (2.113) auf, der mit $\bar{n}$ idealen Kondensatoren abgeschlossen ist, d. h. es gilt $\boldsymbol{I}_m = -p_7 \bar{\boldsymbol{C}}^7 \boldsymbol{U}_m$. Mit $\boldsymbol{I}_r = -\boldsymbol{N}^{\mathrm{T}} \boldsymbol{I}_m$ und $\boldsymbol{U}_m = \boldsymbol{N}\boldsymbol{U}_r$ haben wir

$$\boldsymbol{I}_r = \boldsymbol{N}^{\mathrm{T}} p_7 \bar{\boldsymbol{C}}^7 \boldsymbol{N} \boldsymbol{U}_r \tag{4.60}$$

und somit eine geeignete Synthese von $\bar{\boldsymbol{Y}}^{c7} = p_7 \bar{\boldsymbol{C}}^7$ gefunden. Kompakt lässt sich $\boldsymbol{Y}^{c7}$ durch

$$\boldsymbol{C}^7 = \boldsymbol{P}^7 \begin{bmatrix} \boldsymbol{N}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \mathbf{diag}(\bar{\boldsymbol{C}}^7_{\mathrm{D}}, \tilde{\boldsymbol{C}}^7) \begin{bmatrix} \boldsymbol{N} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \boldsymbol{P}^{7\mathrm{T}} \tag{4.61}$$

beschreiben. Hierin ist

$$\tilde{\boldsymbol{N}} = \begin{bmatrix} \boldsymbol{N} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \boldsymbol{P}^{7\mathrm{T}} \tag{4.62}$$

das Übersetzungsverhältnis eines idealen Übertragers.

Ziel unserer Überlegungen ist es, eine konkret mehrdimensional passive Schaltung für $\boldsymbol{Y}^c$ zu finden. Aus den erarbeiteten Darstellungen können wir zwei wesentliche Resultate festhalten. Zum einen haben wir eine mehrdimensional konkret passive Realisierung für $\boldsymbol{Y}^{c1,4}, \boldsymbol{Y}^{c2,5}$ und $\boldsymbol{Y}^{c3,6}$ gefunden. Zum anderen haben wir für eine n. n. d. Matrix $\boldsymbol{C}^7$ eine mehrdimensional konkret passive Realisierung von $\boldsymbol{Y}^{c7}$ angegeben. Wir haben gezeigt, dass $\boldsymbol{C}^7$ durch Wahl eines genügend großen v_4 n. n. d. ist.

Es folgt der nachzutragende Beweis für die im Kapitel 4.1.4 gemachte Aussage in Gestalt von Gleichung (4.26). Zur Beweisvorbereitung gehen wir von

$$\boldsymbol{C}^x = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} -1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} c^x_{\mu\nu} \tag{4.63}$$

und

$$\boldsymbol{C}^{t1,4} = \sum_{\mu=2}^{N} \sum_{\nu=1}^{\mu-1} \boldsymbol{T}_{\mu\nu} \boldsymbol{S}_0 \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \frac{|c^x_{\mu\nu}|}{v_4} \tag{4.64}$$

aus. Damit können wir durch

$$\boldsymbol{Y}^{c1,4} = p_x \boldsymbol{C}^x + p_t \boldsymbol{C}^{t1,4} = \frac{1}{2v_0}[p_1 - p_4]\boldsymbol{C}^x + \frac{v_4}{2v_0}[p_1 + p_4]\boldsymbol{C}^{t1,4} = p_1 \underbrace{\left[\frac{v_4 \boldsymbol{C}^{t1,4}}{2v_0} + \frac{\boldsymbol{C}^x}{2v_0}\right]}_{\boldsymbol{C}^1} + p_4 \underbrace{\left[\frac{v_4 \boldsymbol{C}^{t1,4}}{2v_0} - \frac{\boldsymbol{C}^x}{2v_0}\right]}_{\boldsymbol{C}^4}$$

(4.65)

die Trennung nach $\boldsymbol{C}^1$ und $\boldsymbol{C}^4$ durchführen. $\boldsymbol{C}^1$ berechnet sich zu

$$\boldsymbol{C}^1 = \frac{1}{4v_0}\sum_{\mu=2}^{N}\sum_{\nu=1}^{\mu-1}\boldsymbol{T}_{\mu\nu}\boldsymbol{S}_0\begin{bmatrix} |c^x_{\mu\nu}| - c^x_{\mu\nu} & 0 \\ 0 & |c^x_{\mu\nu}| + c^x_{\mu\nu}\end{bmatrix}\boldsymbol{S}_0^{\mathrm{T}}\boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \tag{4.66}$$

und ist n. n. d., da die Diagonalmatrix keine negativen Elemente enthält und die Kongruenztransformation die nicht negative Definitheit nicht beeinflusst. Offenbar ist auch

$$\boldsymbol{C}^4 = \frac{1}{4v_0}\sum_{\mu=2}^{N}\sum_{\nu=1}^{\mu-1}\boldsymbol{T}_{\mu\nu}\boldsymbol{S}_0\begin{bmatrix} |c^x_{\mu\nu}| + c^x_{\mu\nu} & 0 \\ 0 & |c^x_{\mu\nu}| - c^x_{\mu\nu}\end{bmatrix}\boldsymbol{S}_0^{\mathrm{T}}\boldsymbol{T}_{\mu\nu}^{\mathrm{T}} \tag{4.67}$$

nicht negativ definit. Ebenso können wir zeigen, dass $\boldsymbol{C}^2$, $\boldsymbol{C}^3$, $\boldsymbol{C}^5$ und $\boldsymbol{C}^6$ n. n. d. sind.

Nach diesen Vorbetrachtungen kommen wir nun zum eigentlichen Beweis. Wir zeigen zunächst

$$\left[\ \boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}} \geq 0 \text{ für } \mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0}\ \right] \Longrightarrow \left[\ \boldsymbol{Y}^{ct} + \boldsymbol{Y}^{ct\mathrm{H}} > 0 \text{ für } \mathrm{Re}\{p_t\} > 0\ \right]\ . \tag{4.68}$$

Wir gehen dazu von der verallgemeinerten Wirkleistung aus

$$2P = \mathrm{Re}\{\boldsymbol{I}^{\mathrm{H}}\boldsymbol{U}\} = \boldsymbol{U}^{\mathrm{H}}\frac{1}{2}[\boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}}]\boldsymbol{U} = \sum_{\kappa=1}^{7}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^{\kappa}\,\mathrm{Re}\{p_\kappa\}\boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}}\frac{1}{2}[\boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}}]\boldsymbol{U} \geq 0\,. \tag{4.69}$$

Da die Ungleichung für alle $\mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0}$ gilt, folgt aus der Prämisse insbesondere die nicht negative Definitheit von $\boldsymbol{C}^7$. Die Matrix $\boldsymbol{C}^7 = \frac{v_4}{v_0}\boldsymbol{C}^t - \sum_\kappa^6 \boldsymbol{C}^\kappa$ ist also für beliebige symmetrische n. n. d. $\boldsymbol{C}^\kappa$ nicht negativ definit. Dies kann nur dann der Fall sein, wenn $\boldsymbol{C}^t$ p.d. ist, die nicht negative Definitheit von $\boldsymbol{C}^t$ reicht nicht aus. Denn bei nur n. n. d. Matrix $\boldsymbol{C}^t$ ist nicht auszuschließen, dass ein Vektor $\boldsymbol{v} \neq \boldsymbol{0}$ existiert für den

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}^7\boldsymbol{v} = \underbrace{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}^t\boldsymbol{v}}_{=0} - \frac{v_0}{v_4}\sum_{\kappa=1}^{6}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}^\kappa\boldsymbol{v} < 0 \tag{4.70}$$

gilt und somit $\boldsymbol{C}^7$ der unmittelbar aus der Prämisse abgeleiteten Aussage widerspräche.

Eine weitere Darstellung der verallgemeinerten Wirkleistung erhalten wir mit

$$\hat{\boldsymbol{W}}_s = [\,\boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^1\boldsymbol{U}\ ,\ \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^2\boldsymbol{U}\ ,\ \ldots\ ,\ \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^7\boldsymbol{U}\,]^{\mathrm{T}} \tag{4.71}$$

zu

$$\begin{aligned} 2P &= \mathrm{Re}\{\boldsymbol{p_t}^{\mathrm{T}}\}\hat{\boldsymbol{W}}_s + \boldsymbol{U}^{\mathrm{H}}\tfrac{1}{2}[\boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}}]\boldsymbol{U} \\ &= \mathrm{Re}\{\boldsymbol{p_x}^{\mathrm{T}}\}\boldsymbol{H}v_0\hat{\boldsymbol{W}}_s + \boldsymbol{U}^{\mathrm{H}}\tfrac{1}{2}[\boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}}]\boldsymbol{U} = \mathrm{Re}\{\boldsymbol{p_x}^{\mathrm{T}}\}v_0\begin{bmatrix} \boldsymbol{U}^{\mathrm{H}}[\boldsymbol{C}^1 - \boldsymbol{C}^4]\boldsymbol{U} \\ \boldsymbol{U}^{\mathrm{H}}[\boldsymbol{C}^2 - \boldsymbol{C}^5]\boldsymbol{U} \\ \boldsymbol{U}^{\mathrm{H}}[\boldsymbol{C}^3 - \boldsymbol{C}^6]\boldsymbol{U} \\ \boldsymbol{U}^{\mathrm{H}}\sum_{\nu=1}^{7}\boldsymbol{C}^\nu\boldsymbol{U}\end{bmatrix} + \boldsymbol{U}^{\mathrm{H}}\tfrac{1}{2}[\boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}}]\boldsymbol{U}\,. \end{aligned} \tag{4.72}$$

Drücken wir nun wieder $\boldsymbol{C}^1$ bis $\boldsymbol{C}^7$ durch $\boldsymbol{C}^x$, $\boldsymbol{C}^y$, $\boldsymbol{C}^z$ und $\boldsymbol{C}^t$ aus, so erhalten wir

$$2P = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^x\,\mathrm{Re}\{p_x\}\boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^y\,\mathrm{Re}\{p_y\}\boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^z\,\mathrm{Re}\{p_z\}\boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}^t\,\mathrm{Re}\{p_t\}\boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}}\frac{1}{2}[\boldsymbol{Y}^k + \boldsymbol{Y}^{k\mathrm{T}}]\boldsymbol{U}\,.$$

(4.73)

Nun zeigen wir

$$\exists v_4 \quad \left[\, \boldsymbol{Y}^{ct} + \boldsymbol{Y}^{ct\mathrm{H}} > 0 \text{ für } \mathrm{Re}\{p_t\} > 0 \,\right] \Longrightarrow \left[\, \boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}} \geq 0 \text{ für } \mathrm{Re}\{\boldsymbol{p_t}\} > \boldsymbol{0} \,\right] . \tag{4.74}$$

Wie bereits nachgewiesen wurde, folgt aus $\boldsymbol{Y}^{ct}+\boldsymbol{Y}^{ct\mathrm{H}} > 0$ für $\mathrm{Re}\{p_t\} > 0$ eine positiv definite Matrix $\boldsymbol{C}^t$. Die nicht negative Definitheit von $\boldsymbol{C}^1$ bis $\boldsymbol{C}^6$ ist aus Gleichung (4.66) und Gleichung (4.67) ersichtlich. Aus Gleichung (4.52) i. V. m. Gleichung (4.64) folgt zudem, dass bei hinreichend großer Wahl von v_4 die Matrix $\boldsymbol{C}^7$ n. n. d. ist. Mit

$$\boldsymbol{Y} = \sum_{\nu=1}^{7} \boldsymbol{C}^{\nu} p_{\nu} + \boldsymbol{Y}^{k} \tag{4.75}$$

folgt

$$\left[\boldsymbol{Y} + \boldsymbol{Y}^{\mathrm{H}}\right] = 2 \sum_{\nu=1}^{7} \boldsymbol{C}^{\nu} \,\mathrm{Re}\{p_{\nu}\} + \left[\boldsymbol{Y}^{k} + \boldsymbol{Y}^{k\mathrm{T}}\right] . \tag{4.76}$$

Da nach Voraussetzung $\mathrm{Re}\,\boldsymbol{p_t} > \boldsymbol{0}$ gilt und die Matrizen in der Summe n. n. d. sind, ist die gesamte Summe n. n. d., was zu zeigen war.

4.2.3 Synthese des konstanten Teils

Zunächst werden wir wie auch im Fall der Matrix $\boldsymbol{Y}^{c7}$ einen schon vorhandenen Diagonalteil der Matrix $\boldsymbol{Y}^{k}$ abspalten. Zu diesem Zweck sortieren wir die Zeilen und Spalten von $\boldsymbol{Y}^{k}$ mithilfe der Permutationsmatrix $\boldsymbol{P}^{k}$ um

$$\boldsymbol{Y}^{k} = \boldsymbol{P}^{k} \begin{bmatrix} \bar{\boldsymbol{Y}}^{k} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{Y}}^{k} \end{bmatrix} \boldsymbol{P}^{k\mathrm{T}} , \tag{4.77}$$

wobei $\bar{\boldsymbol{Y}}^{k}$ eine reelle Matrix der Dimension $\tilde{n} \times \tilde{n}$ ist und einen n. n. d. symmetrischen Teil besitzt.

Die verbleibende reelle Diagonalmatrix $\tilde{\boldsymbol{Y}}^{k}$ hat dann die Dimension $N-\tilde{n} \times N-\tilde{n}$ und ist ebenfalls n. n. d.. Die Matrix $\bar{\boldsymbol{Y}}^{k}$ wird in einen symmetrischen und einen schiefsymmetrischen Teil zerlegt

$$\bar{\boldsymbol{Y}}^{k} = \underbrace{[\bar{\boldsymbol{Y}}^{k} + \bar{\boldsymbol{Y}}^{k\mathrm{T}}]\frac{1}{2}}_{\boldsymbol{Y}_R} + \underbrace{[\bar{\boldsymbol{Y}}^{k} - \bar{\boldsymbol{Y}}^{k\mathrm{T}}]\frac{1}{2}}_{\boldsymbol{Y}_G} . \tag{4.78}$$

Der Index R steht für einen Widerstand und der Index G für einen Gyrator. Der symmetrische Teil $\boldsymbol{Y}_R$ kann mittels eines mit positiven Widerständen abgeschlossenen Mehrtorübertragers synthetisiert werden. Dieses Vorgehen entspricht dem, welches zur Synthese von $\bar{\boldsymbol{C}}^{7}$ angewendet wurde. Der schiefsymmetrische Teil $\boldsymbol{Y}_G$ kann durch Gyratoren synthetisiert werden. Aufgrund der Addition von $\boldsymbol{Y}_R$ und $\boldsymbol{Y}_G$ müssen beide Teile mittels Parallelschaltung miteinander verbunden werden. Ohne dies im Detail erläutern zu wollen, sei gesagt, dass dieses Vorgehen zu verzögerungsfreien gerichteten Schleifen in den Wellengrößen führt. Wir suchen daher nach einer Synthese des konstanten Teils, welche in eine Wellen-Digital-Struktur mündet, die keine verzögerungsfreien Schleifen enthält. Zudem fordern wir, dass die Wellen-Digital-Struktur nur aus Elementarbausteinen mit 2 Toren im Wellenbereich besteht, da es so einfacher ist die Passivität sicherzustellen [Meer79].

Zur Lösung des Problems greifen wir auf die Erkenntnisse aus Abschnitt 2.8.1 zurück. Wir wählen nun $M = 2\tilde{n}$. Schließen wir das M-Tor an den letzten $\tilde{n}$ Toren mit positiven Widerständen ab, so stellt die Gesamtschaltung bzgl. der ersten $\tilde{n}$ Tore eine passive dynamikfreie, nichtreziproke Schaltung dar und lässt sich unter Einhaltung der obigen Forderungen realisieren. Zur Einhaltung der Forderung nach Nichtauftreten von verzögerungsfreien Schleifen, wählen wir die Torwiderstände gleich den Abschlusswiderständen, sodass die Schaltung im Bild 4.2 resultiert. Die Frage ist nun, wie man sich die Streu- bzw.

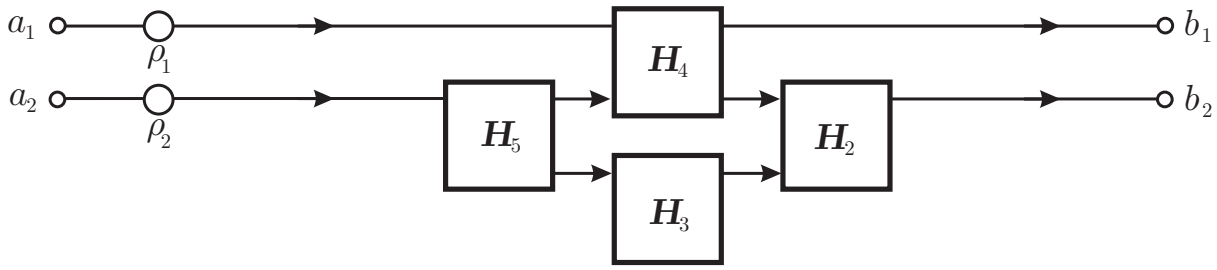


Bild 4.2: Realisierung eines passiven nichtreziproken $\tilde{n}$-Tors für $\tilde{n} = 2$

die Impedanzmatrix der energieneutralen Schaltung beschafft, sodass die gewünschte resultierende dissipative Schaltung bzgl. der ersten $\tilde{n}$ Tore entsteht. Geht man von der partionierten Streumatrix aus, so entstehen kompliziert zu lösende Beziehungen. Wir gehen deshalb von der Admittanzmatrix aus

$$\begin{bmatrix} \boldsymbol{I}_1 \\ \boldsymbol{I}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{bmatrix} . \tag{4.79}$$

Der Abschluss der letzten $\tilde{n}$ Tore wird durch $\boldsymbol{U}_2 = -\boldsymbol{Z}_\mathrm{t}\boldsymbol{I}_2$ beschrieben, sodass

$$\boldsymbol{I}_1 = [\boldsymbol{Y}_{11} - \boldsymbol{Y}_{12}[\boldsymbol{Y}_\mathrm{t} + \boldsymbol{Y}_{22}]^{-1}\boldsymbol{Y}_{21}]\boldsymbol{U}_1 \tag{4.80}$$

resultiert. Nun bringen wir die Bedingungen der Energieneutralität für die Admittanzmatrix mit ein $\boldsymbol{Y}_{11} = -\boldsymbol{Y}_{11}^\mathrm{H}$, $\boldsymbol{Y}_{12} = -\boldsymbol{Y}_{21}^\mathrm{H}$, $\boldsymbol{Y}_{22} = -\boldsymbol{Y}_{22}^\mathrm{H}$. Setzen wir dies ein, so haben wir

$$\bar{\boldsymbol{Y}}_k = \boldsymbol{Y}_{11} + \boldsymbol{Y}_{21}^\mathrm{H}[\boldsymbol{Y}_\mathrm{t} + \boldsymbol{Y}_{22}]^{-1}\boldsymbol{Y}_{21} . \tag{4.81}$$

Wir setzen $\boldsymbol{Y}_{22} = \boldsymbol{0}$ und $\boldsymbol{Y}_\mathrm{t} = G\boldsymbol{1}_{\tilde{n}}$. Dadurch ist der symmetrische Teil von $\bar{\boldsymbol{Y}}_k$ durch $\boldsymbol{Y}_{21}^\mathrm{H}\boldsymbol{Y}_{21}/G$ bestimmt und der schiefsymmetrische Teil durch $\boldsymbol{Y}_{11}$, d.h.

$$\boldsymbol{Y}_{11} = \frac{1}{2}[\bar{\boldsymbol{Y}}_k - \bar{\boldsymbol{Y}}_k^\mathrm{T}] \quad \text{und} \quad \boldsymbol{Y}_{21}^\mathrm{H}\boldsymbol{Y}_{21} = \frac{G}{2}[\bar{\boldsymbol{Y}}_k + \bar{\boldsymbol{Y}}_k^\mathrm{T}] . \tag{4.82}$$

Wobei wir nun $\boldsymbol{Y}_{21}$ z.B. durch die Cholesky-Zerlegung ermitteln können. Da $\bar{\boldsymbol{Y}}_k$ reell ist, folgt aus der nicht negativen Definitheit des symmetrischen Teils auch die Reellwertigkeit von $\boldsymbol{Y}_{21}$. Es ist offensichtlich, dass die Cholesky-Zerlegung nicht die optimale Lösung bzgl. des Rechenaufwands liefert. Insbesondere könnte eine Optimierung dahingehend durchgeführt werden, dass die Anzahl der Multiplizierer der Wellen-Digital-Struktur minimiert wird.

4.2.4 Gesamtreferenzschaltung

Nachdem die Teilschaltungen synthetisiert wurden, suchen wir nun nach einer geeigneten Realisierung der Gesamtschaltung. Insbesondere ist darauf zu achten, dass das MDWDF keine verzögerungsfreien gerichteten Schleifen enthält. Von den untersuchten Gesamtschaltungen wird im Folgenden eine vorgestellt. Zur Herleitung der Schaltung gehen wir von Gleichung (4.20) in der Form

$$\boldsymbol{I}^{c1,4} + \boldsymbol{I}^{c2,5} + \boldsymbol{I}^{c3,6} + \boldsymbol{I}^{c7} + \boldsymbol{I}_k = \boldsymbol{J} \tag{4.83}$$

mit

$$\boldsymbol{I}^{c1,4} = \boldsymbol{Y}^{c1,4}\boldsymbol{U}\ ,\ \boldsymbol{I}^{c2,5} = \boldsymbol{Y}^{c2,5}\boldsymbol{U}\ ,\ \boldsymbol{I}^{c3,6} = \boldsymbol{Y}^{c3,6}\boldsymbol{U}\ ,\ \boldsymbol{I}^{c7} = \boldsymbol{Y}^{c7}\boldsymbol{U} \text{ und } \boldsymbol{I}_k = \boldsymbol{Y}^{k}\boldsymbol{U} \tag{4.84}$$

aus. Zudem hatten wir die folgenden Zerlegungen

$$\boldsymbol{Y}^{k} = \boldsymbol{P}^{k}\,\mathbf{diag}(\bar{\boldsymbol{Y}}^{k}, \boldsymbol{0})\boldsymbol{P}^{k\mathrm{T}} + \boldsymbol{P}^{k}\,\mathbf{diag}(\boldsymbol{0}, \tilde{\boldsymbol{Y}}^{k})\boldsymbol{P}^{k\mathrm{T}}, \tag{4.85}$$

$$\boldsymbol{Y}^{c7} = p_7\boldsymbol{P}^{7}\,\mathbf{diag}(\bar{\boldsymbol{C}}^{7}, \boldsymbol{0})\boldsymbol{P}^{7\mathrm{T}} + p_7\boldsymbol{P}^{7}\,\mathbf{diag}(\boldsymbol{0}, \tilde{\boldsymbol{C}}^{7})\boldsymbol{P}^{7\mathrm{T}} \tag{4.86}$$

vorgenommen. Die Addition der Admittanzmatrizen spiegelt die Parallelschaltung der Jaumann-Adaptoren, des Übertragers, der idealen Stromquelle und des energieneutralen $2\tilde{n}$-Tors wieder. Bild 4.3 zeigt die Referenzschaltung mit später noch zu bestimmenden Torleitwerten. [3] Leider enthält, wie wir sehen

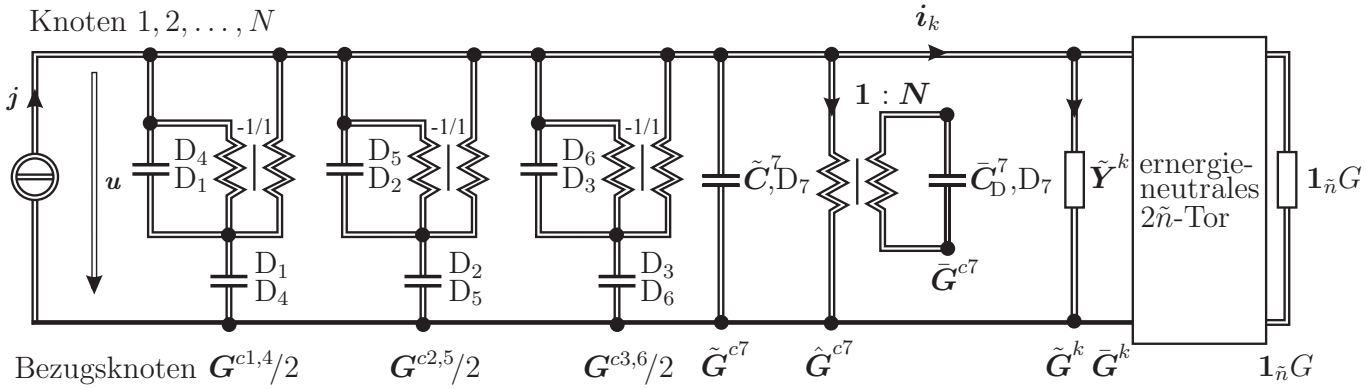


Bild 4.3: Symbolische Darstellung der Gesamtreferenzschaltung

werden, das zugehörige MDWDF, bei direkter Umsetzung, verzögerungsfreie gerichtete Schleifen.

4.3 Umsetzung der elektrischen Bauelemente in WD-Elemente

Bei der Umsetzung der Referenzschaltung in ein MDWDF gehen wir so vor, dass wir zunächst die einzelnen Elemente in WD-Elemente umsetzen und anschließend die Verknüpfung der Teilstrukturen vornehmen.

Umsetzung des reaktiven Anteils in WD-Elemente

Für die Umsetzung der Matrizen $\boldsymbol{Y}^{c1,4}_{\mathrm{D}}$, $\boldsymbol{Y}^{c2,5}_{\mathrm{D}}$ und $\boldsymbol{Y}^{c3,6}_{\mathrm{D}}$ betrachten wir exemplarisch

$$\boldsymbol{Y}^{c1,4} = \boldsymbol{A}_{\mathrm{rvn}}\,\mathbf{diag}(\boldsymbol{Y}^{c1,4}_{21}, \ldots, \boldsymbol{Y}^{c1,4}_{N,N-1})\boldsymbol{A}^{\mathrm{T}}_{\mathrm{rvn}}\ . \tag{4.87}$$

Die Knoten-Zweig-Inzidenzmatrix $\boldsymbol{A}_{\mathrm{rvn}}$ beschreibt die topologische Verschaltung der Zweitore, die selbst durch die Admittanzmatrizen

$$\boldsymbol{Y}^{c1,4}_{\mu\nu} = \boldsymbol{S}_0 \begin{bmatrix} \{^{p_4}_{p_1}\}\frac{C^x_{\mu\nu}}{4} & 0 \\ 0 & \{^{p_1}_{p_4}\}\frac{C^x_{\mu\nu}}{4} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \tag{4.88}$$

[3] Es handelt sich hier und im Folgenden immer um symbolische Darstellungen. Falls Unklarheiten auftreten sollten, sind immer die zu den Zeichnungen gehörigen Gleichungen maßgebend.

beschrieben werden. Die Synthese erfolgte durch eine mit $\frac{1}{Z_3} = \{^{p_4}_{p_1}\} C^x_{\mu\nu}$ und $\frac{1}{Z_4} = \{^{p_1}_{p_4}\} C^x_{\mu\nu}$ abgeschlossene Jaumann-Schaltung mit $n = 1$. Die Nachbildung der idealen Kondensatoren durch Verzögererelemente im Wellenbereich erfolgt gemäß Kapitel 2.8.5 durch Wahl der Torwiderstände nach Gleichung (2.173) zu

$$R_3 = R_4 = \frac{T}{2C^x_{\mu\nu}} \,. \tag{4.89}$$

Fassen wir die Torleitwerte aller $N^2 - N$ idealen Kondensatoren der Matrix $\boldsymbol{Y}^{c1,4}$ in der Diagonalmatrix

$$\boldsymbol{G}^{c1,4} = \frac{2}{T}\,\mathbf{diag}\left(C^x_{21}\,,\,\ldots\,,\,C^x_{N,N-1}\right) \tag{4.90}$$

zusammen, so erhalten wir durch Anwendung der Trapezregel auf die reaktiven Bauelemente mit

$$z_\nu = \mathrm{e}^{\,p_\nu T} \quad , \quad z_\nu = \frac{1+\psi_\nu}{1-\psi_\nu} \quad , \quad p_\nu = \frac{2}{T}\,\mathrm{artanh}(\psi_\nu) \approx \frac{2\psi_\nu}{T} \tag{4.91}$$

dann näherungsweise

$$\boldsymbol{R}^{c1,4}\boldsymbol{Y}^{c1,4}_{\mathrm{D}} \approx \mathbf{diag}\left(\left\{\begin{matrix}\psi_4\\ \psi_1\end{matrix}\right\},\,\ldots\,,\left\{\begin{matrix}\psi_1\\ \psi_4\end{matrix}\right\}\right) . \tag{4.92}$$

Wir werden im Folgenden das Näherungszeichen durch das Gleichheitszeichen ersetzen. Die zu $\boldsymbol{Y}^{c1,4}_{\mathrm{D}}$ gehörige Streumatrix lautet dann

$$\boldsymbol{S}^{c1,4}_{\mathrm{D}} = [\mathbf{1} - \boldsymbol{R}^{c1,4}\boldsymbol{Y}^{c1,4}_{\mathrm{D}}][\mathbf{1} + \boldsymbol{R}^{c1,4}\boldsymbol{Y}^{c1,4}_{\mathrm{D}}]^{-1} = \mathbf{diag}\left(\left\{\begin{matrix}z_4^{-1}\\ z_1^{-1}\end{matrix}\right\},\,\ldots\,,\left\{\begin{matrix}z_1^{-1}\\ z_4^{-1}\end{matrix}\right\}\right) . \tag{4.93}$$

Die Matrizen $\boldsymbol{Y}^{c2,5}_{\mathrm{D}}$ und $\boldsymbol{Y}^{c3,6}_{\mathrm{D}}$ werden in gleicher Weise in eine WD-Struktur umgesetzt.

Nun suchen wir eine WDF-Realisierung für $\boldsymbol{Y}^{c7}$. Diese Matrix wurde in Gleichung (4.61) aufbereitet. Um für die darin enthaltenen idealen Kondensatoren $\mathbf{diag}(\bar{\boldsymbol{C}}_{\mathrm{D}}, \tilde{\boldsymbol{C}}^7)$ einfache Signalflussdiagramme zu erhalten, legen wir die Torleitwertmatrizen zu

$$\bar{\boldsymbol{G}}^{c7} = \frac{2}{T}\bar{\boldsymbol{C}}_{\mathrm{D}} \,. \tag{4.94}$$

und

$$\tilde{\boldsymbol{G}}^{c7} = \frac{2}{T}\tilde{\boldsymbol{C}}^7 \,. \tag{4.95}$$

fest. Die idealen Kondensatoren haben somit i. S. der Trapezregel die Streumatrix $\mathbf{1}_N z_7^{-1}$.

Umsetzung der Jaumann-Adaptoren in WD-Elemente

Die einzelnen Jaumann-Strukturen, die durch

$$\boldsymbol{S}_0 \begin{bmatrix} \{^{p_4}_{p_1}\}\frac{C^x_{\mu\nu}}{4} & 0 \\ 0 & \{^{p_1}_{p_4}\}\frac{C^x_{\mu\nu}}{4} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \tag{4.96}$$

beschrieben werden, sind im Kapitel 2.8.2 erläutert worden. Durch die Festlegungen der Torwiderstände der idealen Kondensatoren an den jeweiligen Toren 3 und 4 gemäß Gleichung (4.90) ist der erste Teil

der Torwiderstände der Jaumann-Adaptoren bestimmt. Durch die Forderung nach Reflexionsfreiheit resultieren die Torleitwerte der jeweils anderen Tore 1 und 2 aus Gleichung (2.153) mit $n = 1$ zu

$$\frac{1}{2}\boldsymbol{G}^{c1,4}\,. \tag{4.97}$$

Nachdem alle Torwiderstände festliegen, ergibt sich die Streumatrix der Jaumann-Adaptoren als direkte Summe der Streumatrizen eines Jaumann-Adaptors gemäß Gleichung (2.155) zu

$$\boldsymbol{S}^{jau} = \mathbf{diag}\left(\left[\begin{array}{cc}\mathbf{0} & \boldsymbol{S}_1\\ \boldsymbol{S}_1^{\mathrm{T}} & \mathbf{0}\end{array}\right], \cdots, \left[\begin{array}{cc}\mathbf{0} & \boldsymbol{S}_1\\ \boldsymbol{S}_1^{\mathrm{T}} & \mathbf{0}\end{array}\right]\right). \tag{4.98}$$

Umsetzung des idealen Übertragers in WD-Elemente

Für die WD-Struktur des $\mathbf{1}:\boldsymbol{N}$-Übertragers fordern wir eine reflexionsfreie Realisierung. Die Torleitwerte der rechten Seite liegen gemäß Gleichung (4.94) fest. Die Torleitwerte $\hat{\boldsymbol{G}}^{c7}$ der linken Seite des Übertragers ergeben sich dann gemäß Gleichung (2.118) zu $\hat{\boldsymbol{G}}^{c7} = \boldsymbol{N}^{\mathrm{T}}\,\mathbf{diag}\,\bar{\boldsymbol{G}}^{c7}\boldsymbol{N}$. Wir bemerken, dass hier die Torwiderstandsmatrix $\hat{\boldsymbol{G}}^{c7}$ keine Diagonalmatrix mehr ist. Welche Auswirkungen dies hat, werden wir später sehen.

Umsetzung des konstanten Teils in WD-Elemente

Die Torleitwertmatrix von $\boldsymbol{P}^k\boldsymbol{Y}^k\boldsymbol{P}^{k\mathrm{T}}$ ist

$$\boldsymbol{G}^k = \mathbf{diag}(\bar{\boldsymbol{G}}^k, \tilde{\boldsymbol{G}}^k) \quad . \tag{4.99}$$

Die Torleitwerte der Widerstände werden zu $\tilde{\boldsymbol{G}}^k = \tilde{\boldsymbol{Y}}^k$ gewählt. Dann sind die Streumatrizen der Widerstände $\tilde{\boldsymbol{S}}^k = \mathbf{0}$. Die Torwiderstände $\bar{\boldsymbol{G}}^k$ sind frei wählbar.

4.4 Umsetzung der Gesamtschaltung in ein MDWDF

Bei der Gesamtschaltung handelt es sich i. W. um N Parallelschaltungen von $2N$-Toren, die mit idealen Stromquellen abgeschlossen sind. Die zur Umsetzung in ein MDWDF notwendigen N Paralleladaptoren mit angeschlossenen idealen Stromquellen wurden im Abschnitt 2.8.3 hergeleitet.

Im Bild 4.4 ist das MDWDF der direkten Umsetzung der Referenzschaltung aus Bild 4.3 dargestellt. Hierbei wurde die Tatsache ausgenutzt, dass sich ein n-Tor-Paralleladaptor durch einen (n-1)-Tor-Paralleladaptor und einen 3-Tor-Paralleladaptor darstellen lässt. Dies ist aus dem Grund nötig, da die gemeinsamen Tore von Paralleladaptoren und idealen Übertragern keine diagonale Torwiderstandsmatrix mehr besitzen. Die rechten N skalaren Dreitor-Paralleladaptoren müssen als ein (vektorieller) Paralleladaptor aufgefasst werden. Im Kapitel 2.8.2 wurde gezeigt, dass nur ein Tor eines Paralleladaptors reflexionsfrei sein kann. Demnach sind zwischen dem rechten vektoriellen Paralleladaptor und dem energieneutralen $2\tilde{n}$-Tor verzögerungsfreie gerichtete Schleifen vorhanden.

Um die verzögerungsfreien Schleifen zu beseitigen, fassen wir den rechten vektoriellen Paralleladaptor, die MDWDF-Realisierung des Übertragers und das energieneutrale $2\tilde{n}$-Tor zu einem energieneutralen $(\check{n}+\tilde{n}+\bar{n})$-Tor zusammen. Die Toranzahl $\check{n}$ wird später erläutert. Das Bild 4.5 zeigt das MDWDF. Zu diesem energieneutralen $(\check{n}+\tilde{n}+\bar{n})$-Tor berechnen wir nun die Streumatrix. Dazu betrachten wir die Matrix

$$\begin{aligned}\boldsymbol{Y}_r &= \boldsymbol{P}^{c7}\,\mathbf{diag}(p_7\bar{\boldsymbol{C}}^7, \mathbf{0})\boldsymbol{P}^{c7\mathrm{T}} + \boldsymbol{P}^k\,\mathbf{diag}(\bar{\boldsymbol{Y}}^k, \mathbf{0})\boldsymbol{P}^{k\mathrm{T}}\\ &= \boldsymbol{P}^{c7}\left[\begin{array}{cc}\bar{\boldsymbol{C}}^7 p_7 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\boldsymbol{P}^{c7\mathrm{T}} + \boldsymbol{P}^k\left[\begin{array}{cc}\bar{\boldsymbol{Y}}^k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\boldsymbol{P}^{k\mathrm{T}} \quad , \quad \bar{\boldsymbol{C}}^7 = \boldsymbol{N}^{\mathrm{T}}\bar{\boldsymbol{C}}^7_{\mathrm{D}}\boldsymbol{N}\,.\end{aligned} \tag{4.100}$$

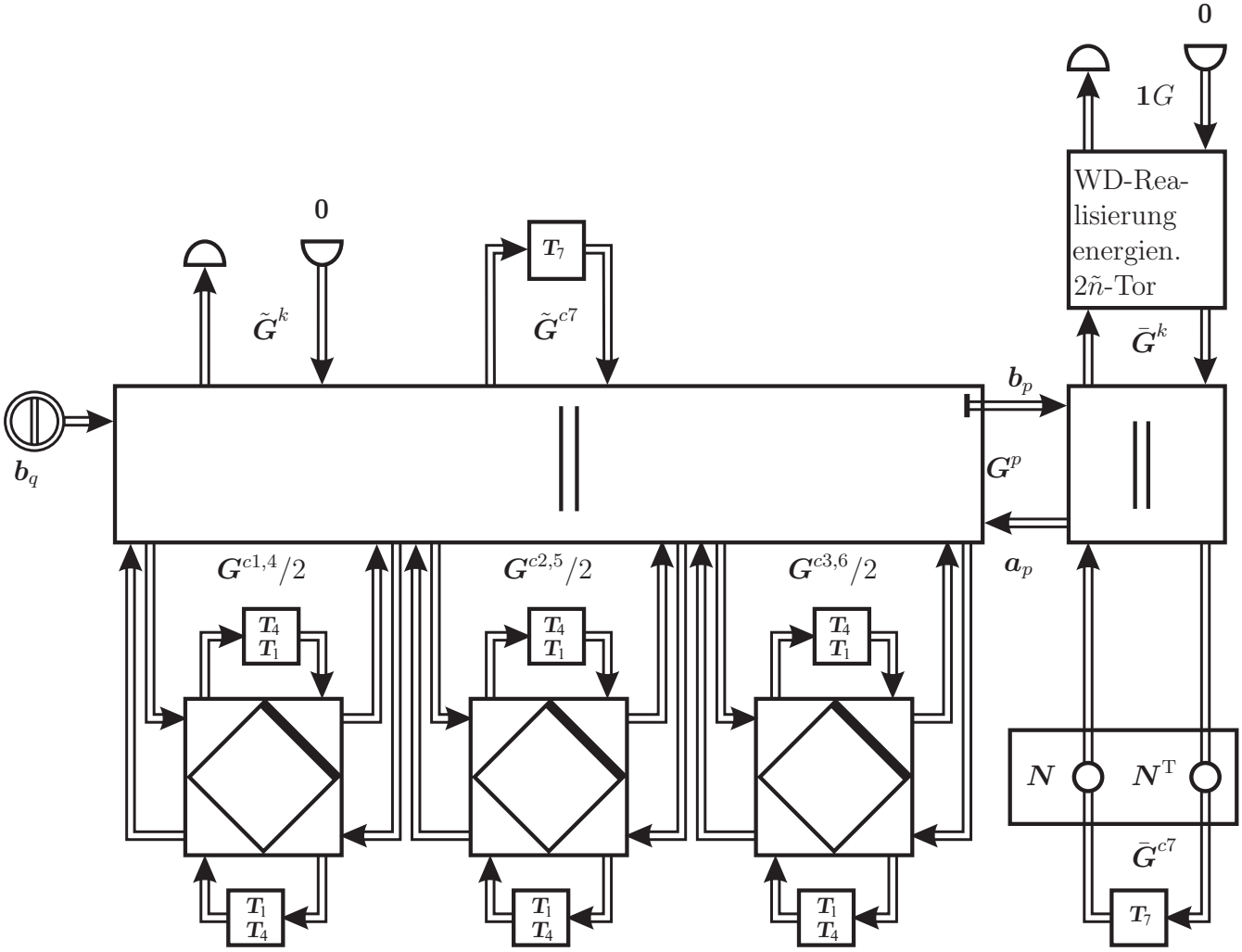


Bild 4.4: Wellendigitalrealisierung

Die Matrix besitzt möglicherweise Nullzeilen und Nullspalten mit gleichem Index. Ihre Anzahl soll $N-\check{n}$ sein. Sie werden im Folgenden eliminiert. Dazu multiplizieren wir die Matrix von links und von rechts mit einer Permutationsmatrix, sodass die Matrix

$$\begin{aligned}\hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{Y}_r\hat{\boldsymbol{P}} &= \hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{P}^{c7}\,\mathbf{diag}(p_7\bar{\boldsymbol{C}}^7,\mathbf{0})\boldsymbol{P}^{c7\mathrm{T}}\hat{\boldsymbol{P}} + \hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{P}^{k}\,\mathbf{diag}(\bar{\boldsymbol{Y}}^k,\mathbf{0})\boldsymbol{P}^{k\mathrm{T}}\hat{\boldsymbol{P}} \\ &= \hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{P}^{c7}\begin{bmatrix}\bar{\boldsymbol{C}}^7 p_7 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{P}^{c7\mathrm{T}}\hat{\boldsymbol{P}} + \hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{P}^{k}\begin{bmatrix}\bar{\boldsymbol{Y}}^k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{P}^{k\mathrm{T}}\hat{\boldsymbol{P}}\end{aligned} \tag{4.101}$$

resultiert. Die letzten $N-\check{n}$ Zeilen und Spalten der zu addierenden Matrizen haben ebenfalls nur Elemente, die null sind. Wir betrachten nun nur noch die ersten $\check{n}$ Zeilen und Spalten von $\hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{Y}_r\hat{\boldsymbol{P}}$, d. h. die Matrix

$$\boldsymbol{Y}_1 = \begin{bmatrix}\mathbf{1}_{\check{n}} & \mathbf{0}\end{bmatrix}\hat{\boldsymbol{P}}^{\mathrm{T}}\boldsymbol{Y}_r\hat{\boldsymbol{P}}\begin{bmatrix}\mathbf{1}_{\check{n}}\\ \mathbf{0}\end{bmatrix} \quad , \quad \boldsymbol{I}_1 = \boldsymbol{Y}_1\boldsymbol{U}_1\,. \tag{4.102}$$

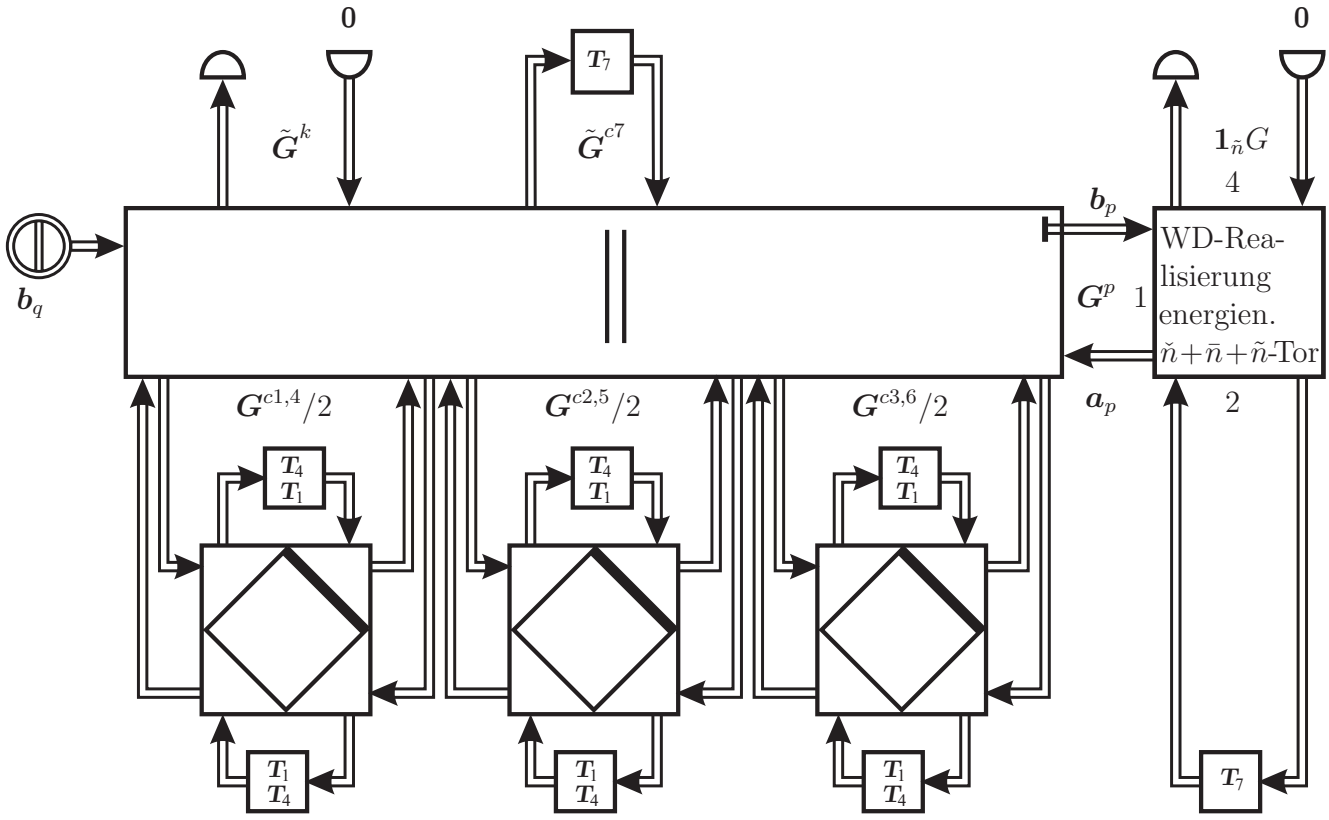


Bild 4.5: Wellendigitalrealisierung ohne verzögerungsfreie gerichtete Schleifen

Nun schreiben wir die Beziehungen zwischen Strom und Spannung neu auf

$$\begin{aligned}\boldsymbol{I}_1 = \boldsymbol{Y}_1\boldsymbol{U}_1 = \; & \left[\begin{array}{cc} \mathbf{1}_{\check{n}} & \mathbf{0} \end{array}\right] \hat{\boldsymbol{P}}^{\mathrm{T}} \boldsymbol{P}^{c7} \left[\begin{array}{c} \mathbf{1}_{\bar{n}} \\ \mathbf{0} \end{array}\right] \boldsymbol{N}^{\mathrm{T}} p_7 \bar{\boldsymbol{C}}_{\mathrm{D}}^7 \boldsymbol{N} \left[\begin{array}{cc} \mathbf{1}_{\bar{n}} & \mathbf{0} \end{array}\right] \boldsymbol{P}^{c7\mathrm{T}} \hat{\boldsymbol{P}} \left[\begin{array}{c} \mathbf{1}_{\check{n}} \\ \mathbf{0} \end{array}\right] \boldsymbol{U}_1 + \\ & \left[\begin{array}{cc} \mathbf{1}_{\check{n}} & \mathbf{0} \end{array}\right] \hat{\boldsymbol{P}}^{\mathrm{T}} \boldsymbol{P}^{k} \left[\begin{array}{c} \mathbf{1}_{\tilde{n}} \\ \mathbf{0} \end{array}\right] \bar{\boldsymbol{Y}}^{k} \left[\begin{array}{cc} \mathbf{1}_{\tilde{n}} & \mathbf{0} \end{array}\right] \boldsymbol{P}^{k\mathrm{T}} \hat{\boldsymbol{P}} \left[\begin{array}{c} \mathbf{1}_{\check{n}} \\ \mathbf{0} \end{array}\right] \boldsymbol{U}_1 \, . \end{aligned} \tag{4.103}$$

Führen wir die Matrizen

$$\boldsymbol{C}_1 = \left[\begin{array}{cc} \mathbf{1}_{\check{n}} & \mathbf{0} \end{array}\right] \hat{\boldsymbol{P}}^{\mathrm{T}} \boldsymbol{P}^{c7} \left[\begin{array}{c} \mathbf{1}_{\bar{n}} \\ \mathbf{0} \end{array}\right] \boldsymbol{N}^{\mathrm{T}} \quad \text{und} \quad \boldsymbol{C}_2 = \left[\begin{array}{cc} \mathbf{1}_{\check{n}} & \mathbf{0} \end{array}\right] \hat{\boldsymbol{P}}^{\mathrm{T}} \boldsymbol{P}^{k} \left[\begin{array}{c} \mathbf{1}_{\tilde{n}} \\ \mathbf{0} \end{array}\right] \tag{4.104}$$

als wesentlichen Teil von Inzidenzmatrizen ein, so lautet dies

$$\boldsymbol{I}_1 = \boldsymbol{Y}_1\boldsymbol{U}_1 = \boldsymbol{C}_1 p_7 \bar{\boldsymbol{C}}_{\mathrm{D}}^7 \boldsymbol{C}_1^{\mathrm{T}} \boldsymbol{U}_1 + \boldsymbol{C}_2 \bar{\boldsymbol{Y}}^k \boldsymbol{C}_2^{\mathrm{T}} \boldsymbol{U}_1 = \boldsymbol{C}_1 p_7 \bar{\boldsymbol{C}}_{\mathrm{D}}^7 \boldsymbol{U}_2 + \boldsymbol{C}_2 \bar{\boldsymbol{Y}}^k \boldsymbol{U}_3 = \boldsymbol{C}_1(-\boldsymbol{I}_2) + \boldsymbol{C}_2 \boldsymbol{I}_3 \, . \tag{4.105}$$

Das Verbindungsnetz kann somit durch

$$\left[\begin{array}{ccc} \mathbf{1}_{\check{n}} & \boldsymbol{C}_1 & \boldsymbol{C}_2 \end{array}\right] \left[\begin{array}{c} \boldsymbol{I}_1 \\ \boldsymbol{I}_2 \\ -\boldsymbol{I}_3 \end{array}\right] = \mathbf{0} \quad \text{und} \quad \left[\begin{array}{ccc} -\boldsymbol{C}_1^{\mathrm{T}} & \mathbf{1}_{\bar{n}} & \mathbf{0} \\ -\boldsymbol{C}_2^{\mathrm{T}} & \mathbf{0} & \mathbf{1}_{\tilde{n}} \end{array}\right] \left[\begin{array}{c} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \boldsymbol{U}_3 \end{array}\right] = \mathbf{0} \tag{4.106}$$

beschrieben werden, wobei wir uns von der Orthogonalitätsbeziehung zwischen den Inzidenzmatrizen überzeugen

$$\left[\begin{array}{ccc} \mathbf{1}_{\check{n}} & \boldsymbol{C}_1 & \boldsymbol{C}_2 \end{array}\right] \left[\begin{array}{cc} -\boldsymbol{C}_1 & -\boldsymbol{C}_2 \\ \mathbf{1}_{\bar{n}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{\tilde{n}} \end{array}\right] = \mathbf{0} \, . \tag{4.107}$$

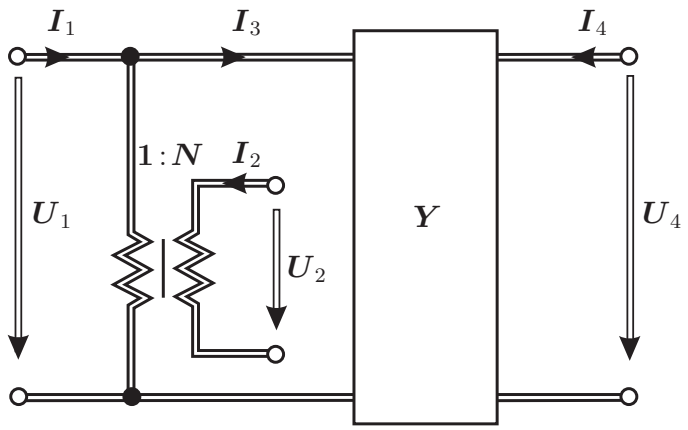


Bild 4.6: Symbolische Darstellung zur Berechnung der Streumatrix des energieneutralen $\check{n}+\tilde{n}+\bar{n}$-Tors

Die Bezugsrichtung des Stromes $\boldsymbol{I}_3$ wurde entgegen der Bezugsrichtung der Spannung $\boldsymbol{U}_3$ gewählt. Nun wird Tor 3 mit dem energieneutralen $2\tilde{n}$-Tor

$$\begin{bmatrix} \boldsymbol{I}_3 \\ \boldsymbol{I}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_3 \\ \boldsymbol{U}_4 \end{bmatrix} \tag{4.108}$$

entsprechend Bild 4.5 verbunden. Es ist die Streumatrix bezogen auf die Tore 1,2 und 4 gesucht, vgl. Bild 4.6. Diese Zeichnung ist wieder symbolisch zu verstehen, d.h. die auftretenden Knoten sind keine im üblichen Sinne, sondern werden durch Gleichung (4.106) beschrieben. Es werden zunächst $\boldsymbol{I}_3$ und $\boldsymbol{U}_3$ durch Nutzen von Gleichung (4.106) eliminiert

$$\begin{bmatrix} \boldsymbol{C}_2\boldsymbol{I}_3 \\ \boldsymbol{I}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_2\boldsymbol{Y}_{11} & \boldsymbol{C}_2\boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_3 \\ \boldsymbol{U}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_1 + \boldsymbol{C}_1\boldsymbol{I}_2 \\ \boldsymbol{I}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_2\boldsymbol{Y}_{11} & \boldsymbol{C}_2\boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{U}_1 \\ \boldsymbol{U}_4 \end{bmatrix} . \tag{4.109}$$

Zu diesen 2 Gleichungen nehmen wir noch das Spannungsübertragungsverhältnis $\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{U}_1 = \boldsymbol{U}_2$ in der Form $\boldsymbol{0} = -\boldsymbol{G}_2\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{U}_1 + \boldsymbol{G}_2\boldsymbol{U}_2$ mit $\boldsymbol{G}_2$ als Torleitwertmatrix der Tore 2 hinzu und erhalten

$$\underbrace{\begin{bmatrix} \boldsymbol{1}_{\check{n}} & \boldsymbol{C}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1}_{\tilde{n}} \end{bmatrix}}_{\boldsymbol{z}} \begin{bmatrix} \boldsymbol{I}_1 \\ \boldsymbol{I}_2 \\ \boldsymbol{I}_4 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{C}_2\boldsymbol{Y}_{11}\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{C}_2\boldsymbol{Y}_{12} \\ -\boldsymbol{G}_2\boldsymbol{C}_1^{\mathrm{T}} & \boldsymbol{G}_2 & \boldsymbol{0} \\ \boldsymbol{Y}_{21}\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}}_{\check{\boldsymbol{Y}}} \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \boldsymbol{U}_4 \end{bmatrix} . \tag{4.110}$$

Übrigens äußert sich die Energieneutralität in $\boldsymbol{z}\check{\boldsymbol{Y}}^{\mathrm{H}} + \check{\boldsymbol{Y}}\boldsymbol{z}^{\mathrm{H}} = \boldsymbol{0}$. Die eigentlich gesuchte Streumatrix berechnet sich durch

$$\boldsymbol{S} = \boldsymbol{G}^{1/2}[\boldsymbol{z}\boldsymbol{G} + \check{\boldsymbol{Y}}]^{-1}[\boldsymbol{z}\boldsymbol{G} - \check{\boldsymbol{Y}}]\boldsymbol{R}^{1/2} . \tag{4.111}$$

Im Folgenden werden wir zeigen, dass die Inverse von

$$[\boldsymbol{z}\boldsymbol{G} + \check{\boldsymbol{Y}}] = \begin{bmatrix} \boldsymbol{G}_1 + \boldsymbol{C}_2\boldsymbol{Y}_{11}\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{C}_1\boldsymbol{G}_2 & \boldsymbol{C}_2\boldsymbol{Y}_{12} \\ -\boldsymbol{G}_2\boldsymbol{C}_1^{\mathrm{T}} & \boldsymbol{G}_2 & \boldsymbol{0} \\ \boldsymbol{Y}_{21}\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{G}_4 \end{bmatrix} \tag{4.112}$$

existiert. Dazu rechnen wir ihren symmetrischen Teil aus

$$[\boldsymbol{z}\boldsymbol{G} + \check{\boldsymbol{Y}}] + [\boldsymbol{z}\boldsymbol{G} + \check{\boldsymbol{Y}}]^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{G}_1 + \boldsymbol{G}_1 + \boldsymbol{C}_2(\boldsymbol{Y}_{11} + \boldsymbol{Y}_{11}^{\mathrm{T}})\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{C}_1(\boldsymbol{G}_2 - \boldsymbol{G}_2) & \boldsymbol{C}_2(\boldsymbol{Y}_{12} + \boldsymbol{Y}_{21}^{\mathrm{T}}) \\ -(\boldsymbol{G}_2 - \boldsymbol{G}_2)\boldsymbol{C}_1^{\mathrm{T}} & \boldsymbol{G}_2 + \boldsymbol{G}_2 & \boldsymbol{0} \\ (\boldsymbol{Y}_{12}^{\mathrm{T}} + \boldsymbol{Y}_{21})\boldsymbol{C}_2^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{G}_4 + \boldsymbol{G}_4 \end{bmatrix} = 2\boldsymbol{G}$$

(4.113)

also die p.d. Matrix der doppelten Torleitwerte. Wenn der symmetrische Teil einer reellen Matrix $\boldsymbol{M}$ p.d. ist, ist die Matrix selbst regulär. Zu dieser Erkenntnis gelangt man durch Einsetzen von $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{x}$ in die Definition der positiven Definitheit des symmetrischen Teils $\boldsymbol{x}^{\mathrm{T}}[\boldsymbol{M}/2 + \boldsymbol{M}^{\mathrm{T}}/2]\boldsymbol{x} > 0\ \forall \boldsymbol{x} \neq \boldsymbol{0}$ d.h. $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} > 0\ \forall \boldsymbol{x} \neq \boldsymbol{0}$. Die Matrix $\boldsymbol{M}$ muss regulär sein, andernfalls würde ein Vektor $\boldsymbol{x} \neq \boldsymbol{0}$ existieren, der $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{0}$ erfüllt und die Ungleichung verletzt.

Wir berechnen nun die Streumatrix der N Paralleladaptoren aus Bild 4.5. Jeder Paralleladaptor hat maximal $M = \{3[N^2 - N] + 3\}$ Tore. Die Torleitwerte der Paralleladaptoren liegen bis auf $\boldsymbol{G}^p$ fest. Dies Torleitwerte werden nun so gewählt, dass die zugehörigen Tore reflexionsfrei sind. Den Torleitwert des Tores ν bezeichnen wir mit $g^p_{\nu\nu}$. Nach Gleichung (2.130) berechnet sich dieser Torleitwert durch die Summe der Torleitwerte der restlichen Tore, die an diesen Knoten inzidieren. Diese Summe kann sehr einfach mittels Skalarprodukt aus den Inzidenzmatrizen dargestellt werden. Die Torleitwertmatrix lautet

$$\boldsymbol{G}^p = \frac{1}{2}\boldsymbol{A}_{\mathrm{rvn}}\left[\boldsymbol{G}^{c1,4} + \boldsymbol{G}^{c2,5} + \boldsymbol{G}^{c3,6}\right]\boldsymbol{A}^{\mathrm{T}}_{\mathrm{rvn}} + \boldsymbol{P}^k\,\mathbf{diag}(\boldsymbol{0}_{\bar{n}}, \tilde{\boldsymbol{G}}^k)\boldsymbol{P}^{k\mathrm{T}} + \boldsymbol{P}^7\,\mathbf{diag}(\boldsymbol{0}_{\hat{n}}, \tilde{\boldsymbol{G}}^{c7})\boldsymbol{P}^{7\mathrm{T}}\ . \tag{4.114}$$

Diese Torleitwerte bestimmen dann auch die des energieneutralen $\check{n} + \tilde{n} + \bar{n}$-Tors.

4.5 Nachweis der Berechenbarkeit

Zum Nachvollziehen der folgenden Überlegungen empfiehlt sich die Betrachtung von Bild 4.5. Zunächst stellen wir fest, dass die ausfallenden Wellengrößen der Quellen, der Verzögererelemente und der Widerstände bekannt sind. Aufgrund der Reflexionsfreiheit der Jaumann-Adaptoren lassen sich die in Richtung der Paralleladaptoren (mit idealer Stromquelle) aus den Jaumann-Adaptoren ausfallenden Wellen (wir wollen diese hier $\boldsymbol{b}_{v\nu}$ nennen) unmittelbar bestimmen. Die Paralleladaptoren nach Gleichung (2.166) können durch

$$\begin{bmatrix} \boldsymbol{a}_{v\nu} \\ b_{p\nu} \end{bmatrix} = \boldsymbol{S}_\nu \begin{bmatrix} \boldsymbol{b}_{v\nu} \\ a_{p\nu} \end{bmatrix} + \boldsymbol{b}_{q\nu} \tag{4.115}$$

beschrieben werden. Die Vektoren $\boldsymbol{b}_{v\nu}$ und $\boldsymbol{b}_{q\nu}$ sind bekannt. Aufgrund der Reflexionsfreiheit erfährt $a_{p\nu}$ keine Berücksichtigung bei der Berechnung von $b_{p\nu}$. Somit kann $b_{p\nu}$ für $\nu = 1, 2, \ldots, N$ parallel berechnet werden. Der Vektor der ausfallenden Wellen der Paralleladaptoren $\boldsymbol{b}_p$ ist nun bekannt. Da die von den Widerständen ausfallenden Wellen null sind und die aus den Verzögerern bzgl. t_7 ausfallenden Wellen bekannt sind, können sämtliche aus dem energieneutralen $(\check{n} + \tilde{n} + \bar{n})$-Tor ausfallenden Wellen $\boldsymbol{a}_p$ bestimmt werden. Diese Berechnung stellt die einzige Kopplung der jeweiligen WD-Realisierungen, der an die N Netzknoten inzidierenden Kirchhoff'schen Elemente dar. Diese Berechnung kann nicht knotenweise parallel ausgeführt werden. Da $\boldsymbol{a}_p$ nun bekannt ist, können die restlichen einfallenden Wellen der Jaumann-Adaptoren bestimmt werden und damit $\boldsymbol{a}_{v\nu}$ für $\nu = 1, 2, \ldots, N$. Zusammengefasst ergibt sich somit die Reihenfolge

1. berechne alle in Richtung der Paralleladaptoren aus den Jaumann-Adaptoren ausfallenden Wellen,
2. berechne $b_{p\nu}$ für jeden Knoten $\nu = 1, 2, \ldots, N$ parallel,
3. verknüpfe die Knoten durch Berechnung von $\boldsymbol{a}_p$ und gleichzeitig Berechnung der restlichen aus dem $(\check{n} + \tilde{n} + \bar{n})$-Tor ausfallenden Wellen,
4. berechne die ausfallenden Wellen der Paralleladaptoren,
5. berechne $\boldsymbol{a}_{v\nu}$ für jeden Knoten $\nu = 1, 2, \ldots, N$ parallel.

Somit ist gezeigt, dass die Berechenbarkeit sichergestellt ist. Die Tatsache, dass die einzelnen Ortspunkte parallel berechnet werden können (bis auf den Rand) bleibt hiervon unberührt. Durch die vorliegende Darstellung wird deutlich, dass innerhalb der Parallelberechnung der Ortspunkte auch noch die Berechnung der Knoten (des elektrischen Netzes) zum Großteil parallel erfolgen kann.

Aus diesen Überlegungen heraus kann sehr einfach eine Einbindung des Konzepts in die zweckmäßige Beschreibung eines MDWDF aus dem Abschnitt 2.9 erfolgen. Als Grundlage dieser Beschreibung dient das MDWDF in der im Bild 4.7 dargestellten Form. Hierbei wurde eine Indizierung verwendet, welche eine Berechnungsreihenfolge unmittelbar erkennen lässt. Um nun die Matrix $\boldsymbol{L}$ aufstellen zu können,

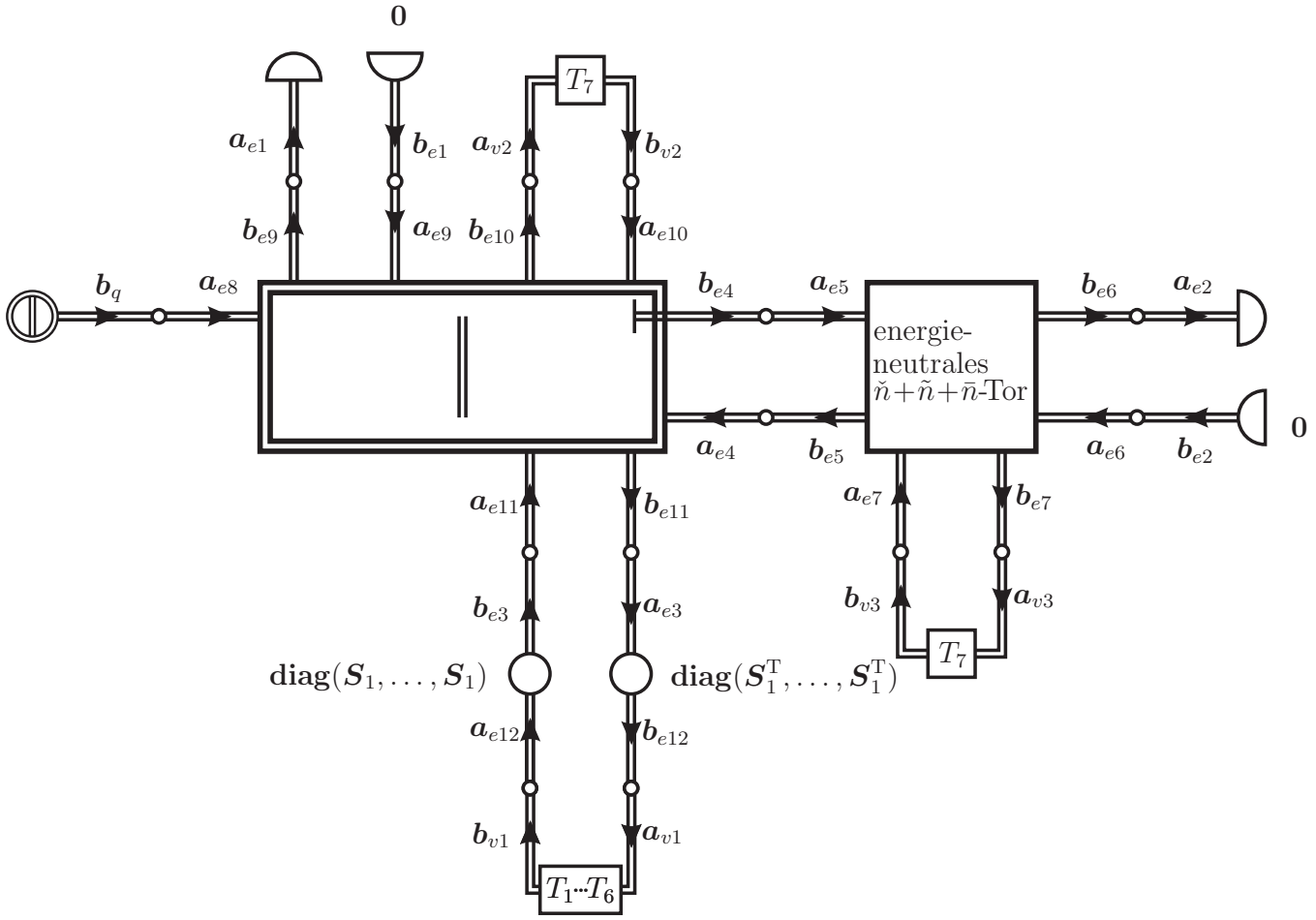


Bild 4.7: Zweckmäßige Beschreibung eines MD-Wellendigitalfilters

empfiehlt es sich, zuerst die Streumatrizen der dynamikfreien Bauelemente anders darzustellen. Zunächst schreiben wir die Streumatrix der N Paralleladaptoren

$$\begin{bmatrix} \boldsymbol{b}_{e4} \\ \boldsymbol{b}_{e8} \\ \boldsymbol{b}_{e9} \\ \boldsymbol{b}_{e10} \\ \boldsymbol{b}_{e11} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{12}^p & \boldsymbol{S}_{13}^p & \boldsymbol{S}_{14}^p & \boldsymbol{S}_{15}^p \\ \boldsymbol{S}_{21}^p & \boldsymbol{S}_{22}^p & \boldsymbol{S}_{23}^p & \boldsymbol{S}_{24}^p & \boldsymbol{S}_{25}^p \\ \boldsymbol{S}_{31}^p & \boldsymbol{S}_{32}^p & \boldsymbol{S}_{33}^p & \boldsymbol{S}_{34}^p & \boldsymbol{S}_{35}^p \\ \boldsymbol{S}_{41}^p & \boldsymbol{S}_{42}^p & \boldsymbol{S}_{43}^p & \boldsymbol{S}_{44}^p & \boldsymbol{S}_{45}^p \\ \boldsymbol{S}_{51}^p & \boldsymbol{S}_{52}^p & \boldsymbol{S}_{53}^p & \boldsymbol{S}_{54}^p & \boldsymbol{S}_{55}^p \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{e4} \\ \boldsymbol{a}_{e8} \\ \boldsymbol{a}_{e9} \\ \boldsymbol{a}_{e10} \\ \boldsymbol{a}_{e11} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{12}^p & \boldsymbol{S}_{13}^p & \boldsymbol{S}_{14}^p & \boldsymbol{S}_{15}^p \\ \boldsymbol{S}_{21}^p & \boldsymbol{S}_{22}^p & \boldsymbol{S}_{23}^p & \boldsymbol{S}_{24}^p & \boldsymbol{S}_{25}^p \\ \boldsymbol{S}_{31}^p & \boldsymbol{S}_{32}^p & \boldsymbol{S}_{33}^p & \boldsymbol{S}_{34}^p & \boldsymbol{S}_{35}^p \\ \boldsymbol{S}_{41}^p & \boldsymbol{S}_{42}^p & \boldsymbol{S}_{43}^p & \boldsymbol{S}_{44}^p & \boldsymbol{S}_{45}^p \\ \boldsymbol{S}_{51}^p & \boldsymbol{S}_{52}^p & \boldsymbol{S}_{53}^p & \boldsymbol{S}_{54}^p & \boldsymbol{S}_{55}^p \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{e5} \\ \boldsymbol{b}_q \\ \boldsymbol{b}_{e1} \\ \boldsymbol{b}_{v2} \\ \boldsymbol{b}_{e3} \end{bmatrix} . \tag{4.116}$$

Die Streumatrix des energieneutralen $\check{n}+\tilde{n}+\bar{n}$-Tors ist durch

$$\begin{bmatrix} \boldsymbol{b}_{e5} \\ \boldsymbol{b}_{e6} \\ \boldsymbol{b}_{e7} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_{11}^{en} & \boldsymbol{S}_{12}^{en} & \boldsymbol{S}_{13}^{en} \\ \boldsymbol{S}_{21}^{en} & \boldsymbol{S}_{22}^{en} & \boldsymbol{S}_{23}^{en} \\ \boldsymbol{S}_{31}^{en} & \boldsymbol{S}_{32}^{en} & \boldsymbol{S}_{33}^{en} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{e5} \\ \boldsymbol{a}_{e6} \\ \boldsymbol{a}_{e7} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_{11}^{en} & \boldsymbol{S}_{12}^{en} & \boldsymbol{S}_{13}^{en} \\ \boldsymbol{S}_{21}^{en} & \boldsymbol{S}_{22}^{en} & \boldsymbol{S}_{23}^{en} \\ \boldsymbol{S}_{31}^{en} & \boldsymbol{S}_{32}^{en} & \boldsymbol{S}_{33}^{en} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{e4} \\ \boldsymbol{b}_{e2} \\ \boldsymbol{b}_{v3} \end{bmatrix} \tag{4.117}$$

festgelegt. Die Streumatrix der Jaumann-Adaptoren lautet

$$\begin{bmatrix} \boldsymbol{b}_{e3} \\ \boldsymbol{b}_{e12} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{12}^{j} \\ \boldsymbol{S}_{21}^{j} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{e3} \\ \boldsymbol{a}_{e12} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{12}^{j} \\ \boldsymbol{S}_{21}^{j} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{e11} \\ \boldsymbol{b}_{v1} \end{bmatrix} . \tag{4.118}$$

Die Streumatrizen der resistiven Abschlüsse sind gleich der Nullmatrix, d. h.

$$\boldsymbol{b}_{e1} = \boldsymbol{0}\,\boldsymbol{a}_{e1} \quad , \quad \boldsymbol{b}_{e2} = \boldsymbol{0}\,\boldsymbol{a}_{e2} . \tag{4.119}$$

Auf eine Beschreibung der torweisen Verbindungen mittels der Permutationsmatrix $\boldsymbol{P}$ verzichten wir, da sich die Matrix $\boldsymbol{L}$ unmittelbar aus Bild 4.7 und den zuvor definierten Matrizen zu

$$\boldsymbol{L} = \left[\begin{array}{cccc|cccccccccccc} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{12}^{j} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}_{12}^{p} & \boldsymbol{0} & \boldsymbol{S}_{14}^{p} & \boldsymbol{0} & \boldsymbol{S}_{13}^{p} & \boldsymbol{0} & \boldsymbol{S}_{15}^{p} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{13}^{en} & \boldsymbol{0} & \boldsymbol{S}_{12}^{en} & \boldsymbol{0} & \boldsymbol{S}_{11}^{en} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{23}^{en} & \boldsymbol{0} & \boldsymbol{S}_{22}^{en} & \boldsymbol{0} & \boldsymbol{S}_{21}^{en} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{33}^{en} & \boldsymbol{0} & \boldsymbol{S}_{32}^{en} & \boldsymbol{0} & \boldsymbol{S}_{31}^{en} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}_{22}^{p} & \boldsymbol{0} & \boldsymbol{S}_{24}^{p} & \boldsymbol{0} & \boldsymbol{S}_{23}^{p} & \boldsymbol{0} & \boldsymbol{S}_{25}^{p} & \boldsymbol{0} & \boldsymbol{S}_{21}^{p} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}_{32}^{p} & \boldsymbol{0} & \boldsymbol{S}_{34}^{p} & \boldsymbol{0} & \boldsymbol{S}_{33}^{p} & \boldsymbol{0} & \boldsymbol{S}_{35}^{p} & \boldsymbol{0} & \boldsymbol{S}_{31}^{p} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}_{42}^{p} & \boldsymbol{0} & \boldsymbol{S}_{44}^{p} & \boldsymbol{0} & \boldsymbol{S}_{43}^{p} & \boldsymbol{0} & \boldsymbol{S}_{45}^{p} & \boldsymbol{0} & \boldsymbol{S}_{41}^{p} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}_{52}^{p} & \boldsymbol{0} & \boldsymbol{S}_{54}^{p} & \boldsymbol{0} & \boldsymbol{S}_{53}^{p} & \boldsymbol{0} & \boldsymbol{S}_{55}^{p} & \boldsymbol{0} & \boldsymbol{S}_{51}^{p} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{21}^{j} & \boldsymbol{0} \end{bmatrix} , \; \boldsymbol{b}_e = \boldsymbol{L} \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ \boldsymbol{b}_e \end{bmatrix} \tag{4.120}$$

ablesen lässt. Die strikte untere Dreiecksgestalt der Matrix $\boldsymbol{SP}_{ee}$ liegt offenbar vor, sodass die Berechenbarkeit gegeben und unmittelbar erkennbar ist.

4.6 Alternative Methoden

Als Alternative können wir Referenzschaltungen herleiten, die auf einer Zustandstransformation basieren. Dazu transformieren wir $\boldsymbol{Y}^{c7}$ auf Diagonalgestalt. Die meisten Eigenschaften von $\boldsymbol{Y}$ bleiben durch die reguläre Transformation erhalten. Der wesentliche Nachteil dieser Synthese ist der rapide ansteigende Aufwand, da die i.d.R. lichten Matrizen i. Allg. zu voll besetzten entarten. Aus diesem Grunde sollen diese Referenzschaltungen hier nicht weiter betrachtet werden, stattdessen wird auf die Untersuchungsergebnisse in [Voll02] verwiesen.

4.7 Randbehandlung

Wie im Kapitel 2.10 dargestellt, benötigen die Verzögerer in Richtung $\boldsymbol{T}_1$ bis $\boldsymbol{T}_6$ am Rand eine spezielle Behandlung. Die Randbehandlung erfolgt jedoch nicht für jeden Verzögerer einzeln, sondern immer für eine Jaumann-Struktur. Wir erläutern die Randbehandlung an Hand

$$\boldsymbol{T}_{\mu\nu}\boldsymbol{S}_0 \begin{bmatrix} \left\{ {p_4 \atop p_1} \right\} \frac{C_{\mu\nu}^{x}}{4} & 0 \\ 0 & \left\{ {p_1 \atop p_4} \right\} \frac{C_{\mu\nu}^{x}}{4} \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}}\, \boldsymbol{T}_{\mu\nu}^{\mathrm{T}} . \tag{4.121}$$

Neben den schon festliegenden Knotenspannungen definieren wir noch die zugehörigen Torströme $I_\mu^{C_{\mu\nu}^x}$ und $I_\nu^{C_{\mu\nu}^x}$ wie folgt

$$\boldsymbol{S}_0 \begin{bmatrix} \{^{p_4}_{p_1}\}C_{\mu\nu}^x/4 & 0 \\ 0 & \{^{p_1}_{p_4}\}C_{\mu\nu}^x/4 \end{bmatrix} \boldsymbol{S}_0^{\mathrm{T}} \begin{bmatrix} U_\mu^{C_{\mu\nu}^x} \\ U_\nu^{C_{\mu\nu}^x} \end{bmatrix} = \begin{bmatrix} I_\mu^{C_{\mu\nu}^x} \\ I_\nu^{C_{\mu\nu}^x} \end{bmatrix} . \tag{4.122}$$

Im Folgenden betrachten wir nur die obere Alternative die untere Alternative erfährt eine ähnliche Behandlung. Zudem soll auf das Abdrucken des hochgestellten $C_{\mu\nu}^x$ verzichtet werden. Ferner ersetzen wir für die restlichen Überlegungen dieses Kapitels die Indizes der unabhängigen Größen durch 1 und 2. Die Tornummerierungen und die Bezugspfeilrichtungen in diesem Kapitel entsprechen denen des Jaumann-Adaptors im Kapitel 2.8.2.

Zunächst wollen wir das Problem konkretisieren, indem wir feststellen, welcher Verzögerer an welchem Rand eine Randbehandlung benötigt. Dazu formen wir die letzte Gleichung zu

$$\begin{bmatrix} p_4 C_{\mu\nu}^x & 0 \\ 0 & p_1 C_{\mu\nu}^x \end{bmatrix} \frac{1}{2} \boldsymbol{S}_0^{\mathrm{T}} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \boldsymbol{S}_0^{\mathrm{T}} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{4.123}$$

um. Mit den aus Kapitel 2.8.2 bekannten Beziehungen

$$\begin{bmatrix} U_3 \\ U_4 \end{bmatrix} = \frac{1}{2} \boldsymbol{S}_0^{\mathrm{T}} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \quad , \quad \begin{bmatrix} I_3 \\ I_4 \end{bmatrix} = -\boldsymbol{S}_0^{\mathrm{T}} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{4.124}$$

haben wir

$$\begin{bmatrix} p_4 C_{\mu\nu}^x & 0 \\ 0 & p_1 C_{\mu\nu}^x \end{bmatrix} \begin{bmatrix} U_3 \\ U_4 \end{bmatrix} = - \begin{bmatrix} I_3 \\ I_4 \end{bmatrix} . \tag{4.125}$$

Die Anwendung der Trapezregel entspricht der Näherung $p_\nu \approx 2\psi_\nu/T$. Mit dem Torwiderstand $R = T/(2C_{\mu\nu}^x)$ an den idealen Kondensatoren haben wir

$$\begin{bmatrix} \psi_4/R & 0 \\ 0 & \psi_1/R \end{bmatrix} \begin{bmatrix} U_3 \\ U_4 \end{bmatrix} \approx - \begin{bmatrix} I_3 \\ I_4 \end{bmatrix} . \tag{4.126}$$

Im Folgenden werden wir anstatt des Näherungszeichens ein Gleichheitszeichen verwenden. Mit $\psi_\nu = \frac{z_\nu - 1}{z_\nu + 1}$ erhalten wir in den Wellengrößen

$$\begin{bmatrix} A_3 \\ A_4 \end{bmatrix} = \mathbf{diag}(z_4^{-1}, z_1^{-1}) \begin{bmatrix} B_3 \\ B_4 \end{bmatrix} . \tag{4.127}$$

Mit den aus Kapitel 2.8.2 bekannten Beziehungen

$$\begin{bmatrix} A_3 \\ A_4 \end{bmatrix} = \boldsymbol{S}_1^{\mathrm{T}} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad , \quad \begin{bmatrix} B_3 \\ B_4 \end{bmatrix} = \boldsymbol{S}_1^{\mathrm{T}} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{4.128}$$

schreiben wir diese Gleichungen als

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \boldsymbol{S}_1 \, \mathbf{diag}(z_4^{-1}, z_1^{-1}) \boldsymbol{S}_1^{\mathrm{T}} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} . \tag{4.129}$$

Nun lässt sich das Signalflussdiagramm leicht angeben, vgl. Bild 4.8. Durch Betrachtung der Gleichungen $A_3 = z_4^{-1} B_3$ und $A_4 = z_1^{-1} B_4$ im Zeitbereich, d. h. $a_3(\boldsymbol{\nu} + \mathbf{e}_4) = b_3(\boldsymbol{\nu})$ und $a_4(\boldsymbol{\nu} + \mathbf{e}_1) = b_4(\boldsymbol{\nu})$ erkennen

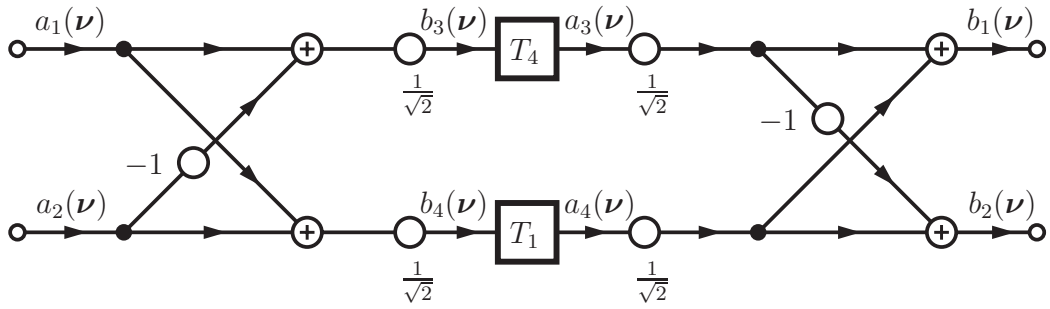


Bild 4.8: Signalflussdiagramm bei Normalberechnung

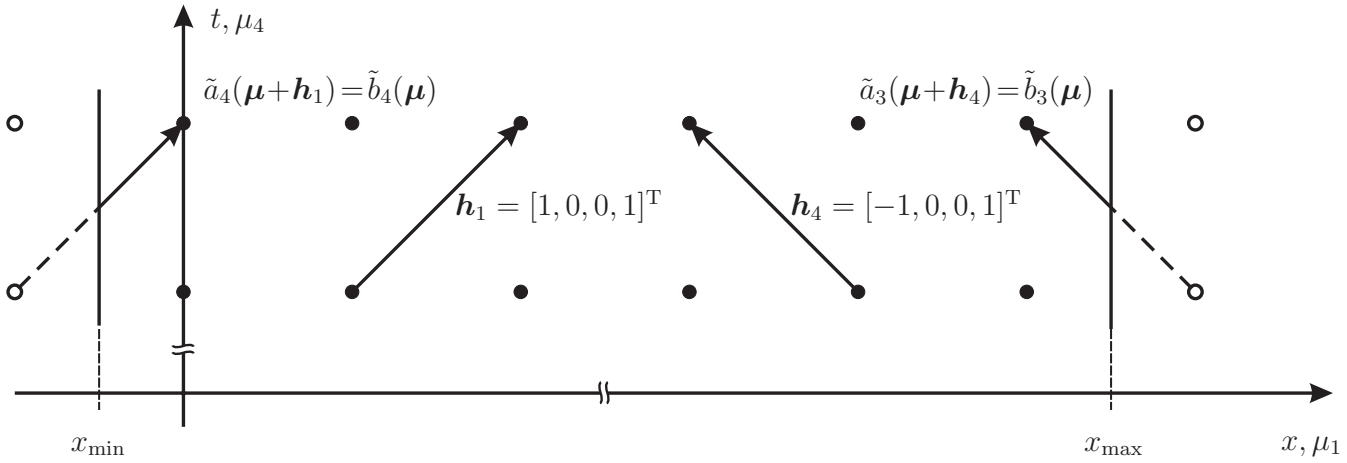


Bild 4.9: Verschiebevektoren bei Normalberechnung

wir, dass der erste Kondensator eine Randbehandlung an $x = x_{\max}$ und der zweite Kondensator eine Randbehandlung an $x = x_{\min}$ erfordert, vgl. Bild 4.9.

Zur eigentlichen Randbehandlung wenden wir Gleichung (2.191) auf Gleichung (4.125) im Zeitbereich an. Hierbei haben wir die mit negativen Vorzeichen versehenen Ströme durch Vertauschen der ein- und ausfallenden Wellen zu berücksichtigen. Mit Wahl der Torwiderstände zu $R = T/(2C^x_{\mu\nu})$ erhalten wir für den ersten Kondensator

$$a_3(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_3(\boldsymbol{\nu} - \mathbf{e}_4/2) \quad , \quad b_3(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_3(\boldsymbol{\nu} + \mathbf{e}_4/2) \tag{4.130}$$

und für den zweiten Kondensator

$$a_4(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_4(\boldsymbol{\nu} - \mathbf{e}_1/2) \quad , \quad b_4(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_4(\boldsymbol{\nu} + \mathbf{e}_1/2) \; . \tag{4.131}$$

Von diesen 4 Gleichungen werden jeweils 2 zur Randbehandlung an $x = x_{\min}$ und $x = x_{\max}$ herangezogen. Um die Randfläche $x = x_{\max}$ zu behandeln, nutzen wir die 2 Gleichungen

$$a_3(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_3(\boldsymbol{\nu} - \mathbf{e}_4/2) \quad \text{und} \quad b_4(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_4(\boldsymbol{\nu} + \mathbf{e}_1/2) \; . \tag{4.132}$$

Um die Randfläche $x = x_{\min}$ zu behandeln, nutzen wir die verbleibenden 2 Gleichungen

$$a_4(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_4(\boldsymbol{\nu} - \mathbf{e}_1/2) \quad \text{und} \quad b_3(\boldsymbol{\nu}) = \frac{1}{2\sqrt{R}} u_3(\boldsymbol{\nu} + \mathbf{e}_4/2) \; . \tag{4.133}$$

Die am Rand $x = x_{\max}$ durch die normale Berechnung nicht ermittelbare Welle $a_3(\boldsymbol{\nu})$ können wir uns nun aus der bekannten Welle $b_4(\boldsymbol{\nu})$ beschaffen, wenn wir einen Zusammenhang zwischen den Feldgrößen u_1 und u_2 auf dem Rand haben. Dabei greifen wir zwei Spezialfälle heraus. Zum einen soll eine konstante Beziehung zwischen u_1 und u_2 auf dem Rand und zum anderen eine lineare differentielle Beziehung bzgl. t zwischen u_1 und u_2 auf dem Rand bestehen. Beide Fälle sind mit dem Ansatz $U_1 = F(x, p_t)U_2$ behandelbar. In Gleichung (4.132) können nun die Feldgrößen eliminiert werden und wir erhalten die gewünschte Beziehung zwischen den Zustandsgrößen auf der Randfläche $x = x_{\max}$

$$A_3 = \underbrace{\frac{F(x_{\max}, p_t) - 1}{F(x_{\max}, p_t) + 1}}_{H_1(p_t)} z_1^{-1/2} z_4^{-1/2} B_4 \quad , \tag{4.134}$$

bzw. im Zeitbereich

$$a_3(\boldsymbol{\nu} + \mathbf{e}_7) = h_1(\nu_7) * b_4(\boldsymbol{\nu}) \tag{4.135}$$

mit dem Faltungsoperator bzgl. ν_7. Im Bild 4.10 ist das WDF für die Randfläche $x = x_{\max}$ dargestellt, wobei das System S_1 die Impulsantwort $h_1(\nu_7)$ besitzt.

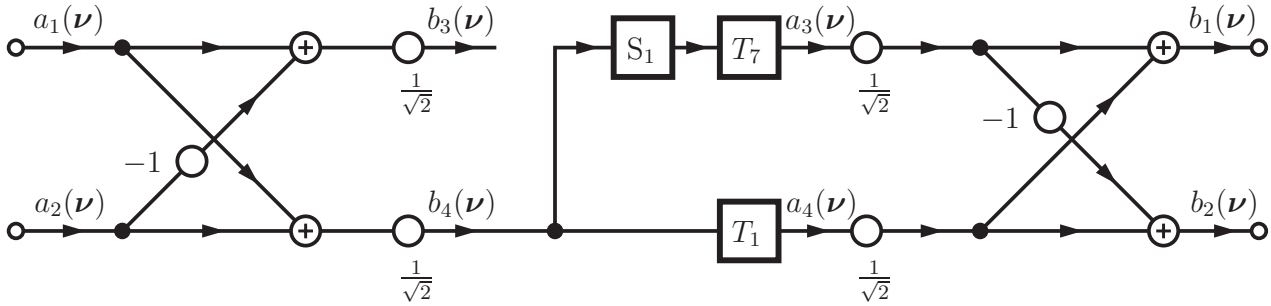


Bild 4.10: Zur Erläuterung der Randbehandlung an $x = x_{\max}$

Für $x = x_{\min}$ wird aus Gleichung (4.133)

$$A_4 = \underbrace{\frac{F(x_{\min}, p_t) + 1}{F(x_{\min}, p_t) - 1}}_{H_2(p_t)} z_1^{-1/2} z_4^{-1/2} B_3 \ , \tag{4.136}$$

bzw. im Zeitbereich

$$a_4(\boldsymbol{\nu} + \mathbf{e}_7) = h_2(\nu_7) * b_3(\boldsymbol{\nu}) \ . \tag{4.137}$$

Im Bild 4.11 ist das WDF für die Randfläche $x = x_{\min}$ dargestellt, wobei das System S_2 die Impulsantwort $h_2(\nu_7)$ besitzt.

Aufgrund des Verzögerers $z_1^{-1/2} z_4^{-1/2} = z_t^{-1} = z_7^{-1}$, ist das MDWDF auch am Rand explizit berechenbar. Bild 4.12 zeigt die Verschiebevektoren bei Randbehandlung.

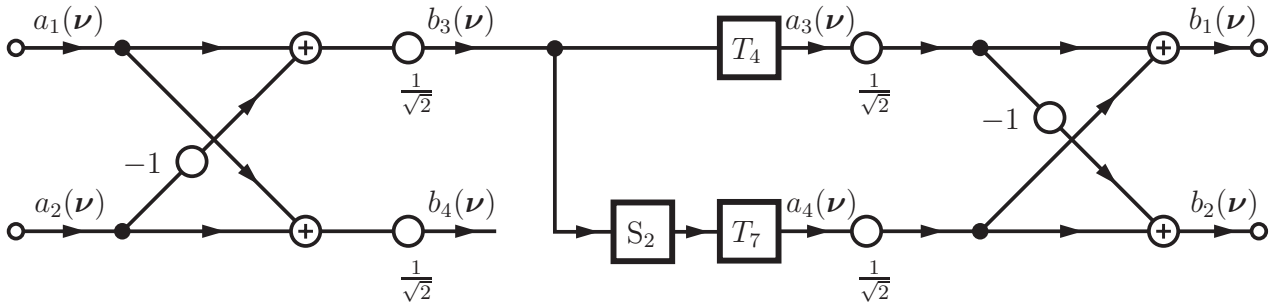


Bild 4.11: Zur Erläuterung der Randbehandlung an $x = x_{\min}$

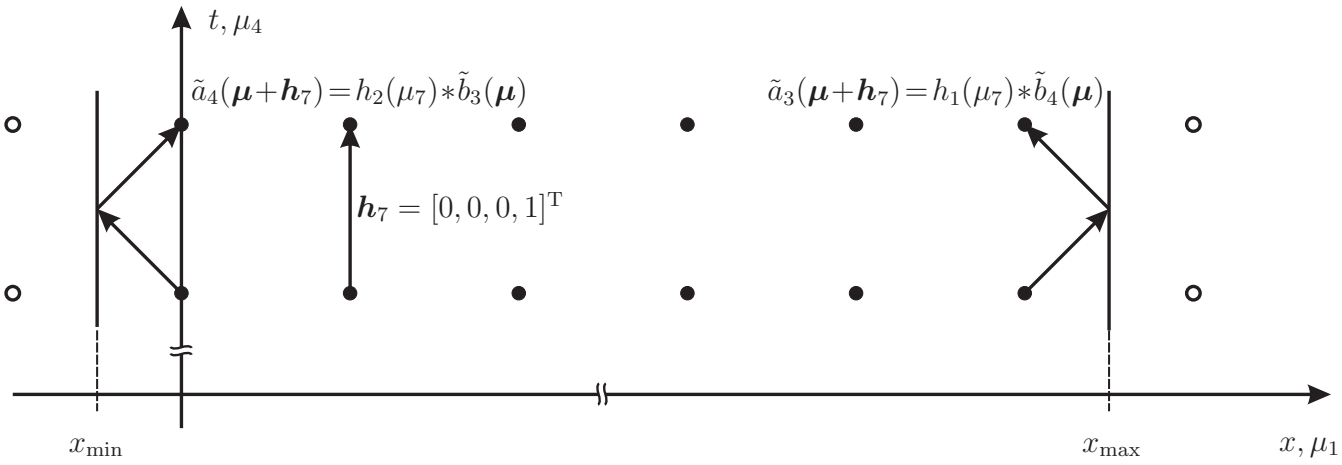


Bild 4.12: Verschiebevektoren bei Randbehandlung

4.8 Zusammenfassung

Zunächst wurde in diesem Kapitel ein System aus linearen partiellen Differentialgleichungen betrachtet und dieses auf die Eigenschaft der Passivität untersucht. Aus der Forderung nach Passivität des quellenfreien Teils haben wir gezeigt, dass die Symmetrie der Matrizen, die den reaktiven Teil beschreiben, notwendig ist. Zudem forderte die Passivität die nicht negative Definitheit des konstanten Teils und des reaktiven Teils bzgl. der Zeit, wobei wir letztere Eigenschaft zur positiven Definitheit stärkten. Systeme mit diesen Eigenschaften werden symmetrisch hyperbolische Systeme genannt. Hyperbolische Systeme folgen insbesondere dem Nahwirkungsprinzip und ihre Vorgänge pflanzen sich mit endlicher Ausbreitungsgeschwindigkeit fort. Die vorhandene Beschreibung des bzgl. der Zeit passiven PDGLn-Systems wurde geeignet umgeformt. Anschließend wurde dieses System durch eine geeignete Koordinatentransformation in ein MD-passives System überführt. Dabei überträgt sich die Eigenschaft der Lokalität auf das System in den neuen Koordinaten. Diese vorliegende Beschreibung wurde zur Synthese einer MD-passiven Kirchhoff'schen Schaltung (der so genannten Referenzschaltung) genutzt. Das kontinuierliche System in Form einer MD-passiven Kirchhoff'schen Schaltung setzten wir durch Anwendung der verallgemeinerten Trapezregel und Verwendung von Wellengrößen in ein MDWDF um, welches die Eigenschaft der MD-Passivität im Diskreten besitzt. Es folgte das Aufbrechen von verzögerungsfreien gerichteten Schleifen und der anschließende Nachweis der Rekursierbarkeit. Zum Abschluss wurde, ausgehend von bekannten Ergebnissen im 1+1D-Fall, eine kompakte und übersichtliche Behandlung von Rändern, für die durch die Synthese entstandenen Strukturen, aufgezeigt. Hierbei haben wir uns auf konstante Randmodelle eingeschränkt.

Kapitel 5

Verifikation von Software mittels formaler Methoden

Ziel dieses Kapitels ist es, die Anwendung formaler Methoden zur Erstellung hochzuverlässiger Software darzulegen. Weiterhin werden die notwendigen Grundlagen und verwendeten Notationen für den im Kapitel 6 geführten Korrektheitsbeweis erläutert. Die herkömmliche Softwareerstellung erfolgt i.d.R. mittels des Wasserfallmodells, Spiralmodells oder des Phasenmodells [Royc87], [Boeh76], [Boeh81], [Boeh86], [Bun96]. Eine Verifikation wird zwischen den einzelnen Stufen dieser Modelle durchgeführt. Obwohl die Schwachstelle der Software-Zuverlässigkeit im Übergang von der formalen Anforderungsspezifikation (das ist Teil des Lastenheftes) zur Software-Spezifikation (das ist Teil des Pflichtenheftes) liegt,[Fisc03], steht in dieser Arbeit die Verifikation der Implementierung gegenüber der formalen Software-Spezifikation im Vordergrund.

Als Maß für die Zuverlässigkeit technischer Systeme wird oft die Ausfallrate, d.h. die Anzahl der Ausfälle innerhalb einer Zeitspanne gewählt. Je nach Anwendung ergeben sich unterschiedliche Anforderungen an die Ausfallrate. Als Methoden zur Bestimmung der Ausfallrate kommen insbesondere Tests und die Schätzung von Parametern eines Zuverlässigkeitsmodells in Betracht. Die Schätzung der Parameter erfolgt durch mehrfaches Durchlaufen eines Zyklusses, der auch zur Qualitätsverbesserung eingesetzt wird. Innerhalb eines Zyklusses wird zunächst ein Test durchgeführt. Die auftretenden Fehler werden behoben und die korrigierte Software erneut dem Test unterzogen. Dieses Vorgehen ist aber mit prinzipiellen Nachteilen behaftet. Aus dem Durchlauf eines Zyklusses muss nicht notwendigerweise eine niedrigere Ausfallrate folgen. Möglicherweise entstehen sogar durch das Beheben des Fehlers neue Fehler. Zudem werden durch den Test nicht alle Fehler aufgedeckt. Ein fehlerfreier Test hat immer nur für die gewählten Testzustände und Testeingänge Gültigkeit. Schon bei Systemen geringer Komplexität kann die Software nicht für ihr gesamtes Definitionsgebiet getestet werden. Aus den eingeschränkten Testeingaben resultiert eine nicht unterschreitbare Grenze der Ausfallrate. Weiterer Nachteil ist, dass eine nicht formale Spezifikation unvollständig, widersprüchlich oder gar falsch sein kann. Abhilfe schafft hier nur die Verwendung anderer Konzepte, z.B. die Verwendung von formalen Methoden zur Verifikation. Schon die Spezifikation der Software erfolgt mittels formaler Methoden, [Floy67], [Dijk68], [Hoar69]. Der Grundgedanke dieses Verfahrens besteht darin, eine mathematische Beschreibung für jede Teilanweisung zu finden, wie es auch in anderen ingenieurwissenschaftlichen Disziplinen üblich ist. Vergleichbare Konzepte sind z.B. Signalflussdiagramme oder Schaltpläne. In diesen Fällen ist jeweils die symbolische Darstellung in Form von Blöcken oder elektrischen Schaltelementen völlig äquivalent zu einer mathematischen Beschreibung, z.B. mittels Differentialgleichungen. Bei den formalen Methoden ist nun jeder symbolischen Darstellung in Form des Quellcodes eine mathematische Beschreibung (die Semantik des Programmcodes) zugeordnet. Ebenso wie für Signalflussdiagramme und Schaltpläne, Regeln für die Beschreibung zusammengeschalteter Elemente existieren, sind auch Regeln für die Verknüpfung von Elementaranweisungen vorhanden. Dem gut dokumentierten, systematischen Erstellen von möglichst

allgemeinen wiederverwendbaren Modulen kommt bei Anwendung formaler Methoden im Hinblick auf Qualität und schnelle Entwicklungszeiten noch größere Bedeutung zu. Der formale Korrektheitsbeweis muss für jedes Modul geführt werden. Aus diesen Modulen kann ein System erzeugt werden. Für den Korrektheitsbeweis des Systems kann auf die Beweise der Module zurückgegriffen werden.

Obwohl die Theorie der Programm-Korrektheits-Beweisführung hinsichtlich sequentieller Programme weitgehend entwickelt worden ist, findet sie in der Praxis nur wenig Anwendung. Die wesentlichen Kritikpunkte, die in [Babe95] aufgelistet sind, wollen wir kurz zusammenfassen. Formale Verifikation ist zu kompliziert, erfordert zu viel hochqualifiziertes Personal, ist zu teuer in der Anwendung, wenig verständlich und nur auf kleine, meist künstliche Problemen anwendbar. Zudem bezieht sich die Beweisführung nur auf den Quellcode. Übersetzer, Betriebssystem, andere systemnahe Software, Beweiswerkzeug und die Hardware können weiterhin fehlerhaft sein. Die angesprochenen Kritikpunkte sind die Gründe dafür, dass formale Verifikationsmethoden bisher fast nur in sicherheitsrelevanten Anwendungen eingesetzt werden. Als Beitrag zur Umsetzung der bekannten theoretischen Konzepte in die Praxis ist [Babe95] zu verstehen. Eine weitere wichtige Quelle, die auch für die vorliegende Arbeit verwendet wurde, ist [Rumm98]. Die letztgenannte Arbeit beinhaltet die Programm-Korrektheits-Beweisführung für eindimensionale Wellendigitalfilter.

Das vorliegende Kapitel stellt die Grundlage für den im Kapitel 6 geführten Korrektheitsbeweis dar. Wir erläutern die Grundbegriffe und Grundkonzepte der Programm-Korrektheits-Beweisführung. Es folgt die Festlegung einer axiomatischen Basis für die in dieser Arbeit notwendigen Anweisungen der Programmiersprache C. Auf eine Berücksichtigung der Laufzeit wie sie in [Rumm98] durchgeführt wurde, wird in dieser Arbeit verzichtet, da die Übersichtlichkeit des Beweises stark darunter leiden würde. Die Berücksichtigung der Zeit kann leicht ergänzt werden. Dies läuft darauf hinaus, dass sich die Laufzeit als eine Summe der Ausführungszeiten für Additionen, Multiplikationen, For-Schleifen und Vergleichsoperationen ergibt.

5.1 Definitionen, Notationen und Vereinbarungen

In diesem Abschnitt werden wir die für den Korrektheitsbeweis benötigten Begriffe erläutern. Hierbei geben wir allerdings keine allgemeine Einführung, sondern beschränken uns auf die Dinge, die im weiteren Verlauf der Arbeit verwendet werden. Die wichtigsten Quellen sind [Rumm98], [Babe95], [HA28], [Novi73] und [Meye88].

Die Menge der Variablennamen bezeichnen wir mit $\mathcal{N}$. Die Menge der durch die Programmiersprache C und von uns benötigten Datentypen ist $\mathcal{D} = \{$ `int, float`$\}$. Die Menge $\mathcal{B}$ beinhaltet die vom Programm schon genutzten Variablenbezeichner. Als Abkürzungen für die kleinsten und größten positiven Zahlen nutzen wir

$$\begin{array}{lcl} \mathrm{Max_i} & : & \text{Darstellbare ganze Zahl mit dem größten Betrag} \\ \mathrm{Max_f} & : & \text{Darstellbare float-Zahl mit dem größten Betrag}. \end{array} \tag{5.1}$$

Eine skalare Programmvariable `var` ist semantisch ein Quadrupel, bestehend aus den vier PL-Variablen [1] Variablenbezeichner *var.name*, Variablentyp *var.typ*, und Variablenwert (Value) *var.v* und $var.l = -1$. Letzte kennzeichnet die Eigenschaft einer skalaren Variablen. Ist die Programmvariable ein Feld, so liegt auf semantischer Ebene ein Quadrupel vor, wobei für die Feldlänge $var.l > 0$ gilt. Zudem ist *var.v* nun ein Vektor. Sofern sich durch konkrete Umstände nichts anderes ergibt, soll zur Vereinfachung im

[1] Das Akronym PL steht für Prädikatenlogik. Die Prädikatenlogik untersucht den Zusammenhang zwischen den Axiomen und Sätzen mathematischer Theorien. Die Prädikatenlogik versteht sich also als eine formale Sprache, die zur Darstellung mathematischer Aussagen benutzt wird. Die logischen Sachverhalte, die zwischen Urteilen, Begriffen usw. bestehen, finden ihre Darstellung durch Formeln, deren Interpretation frei ist von Unklarheiten, die beim sprachlichen Ausdruck leicht auftreten können [HA28].

Folgenden die Programmvariable *var* abkürzend nur noch durch die PL-Variablen *name.typ*, *name.v* und *name .l* vertreten werden. Zwischen den Bezeichnern der PL-Variablen und den Bezeichnern der Programmvariablen ist streng zu unterscheiden. Nichtsdestotrotz besteht per Konvention zwischen den beiden Bezeichnern ein eineindeutiger Zusammenhang. Für die Bezeichner soll gelten, dass die Programmvariablen in Schreibmaschinenschrift gesetzt werden. Die PL-Variablen werden normal kursiv gesetzt. Die PL-Variable *name.v* beschreibt immer den Ist-Wert. I.d.R. sind in Bedingungen Variablen enthalten, z.B. soll die Bedingung garantieren, dass die Variable f den so genannten Sollwert x annimmt, d.h. $f.v = x$. Der Übersicht halber wollen wir vereinbaren, dass sich die Bezeichner der PL-Variablen aus den Bezeichnern der Sollwerte und Anfügen von *.v* ergeben. In diesem Fall hätten wir $f.v = f$. Somit sind aus Kenntnis eines der Bezeichner von Sollwert, PL-Variable oder Programm-Variablen-Bezeichner die anderen zwei jeweils eindeutig bestimmbar. Wir greifen also bei der Wahl der Programm-Variablen-Bezeichner de facto auf die in der Spezifikation enthaltenen Bezeichner der Sollwerte zurück.

Bei der Betrachtung von Variablen mit Indizes stellt sich heraus, dass das angegebene Vorgehen nicht möglich ist, da die Syntax der Programm-Variablen-Bezeichner der Programmiersprache C keine Indizes erlaubt. Ist die PL-Variable ein Matrixelement, so soll der zugehörige Programm-Variablen-Bezeichner

```
<Matrixkleinbuchstabe><Matrixindexoben><Matrixindexunten>_<Zeilenindex>_<Spaltenindex>
```

lauten. Als Beispiel betrachten wir die PL-Variable $\boldsymbol{R}_a.v$. Für das Element der Zeile 5 und Spalte 2 lautet die PL-Variable $r_{a\,52}.v$ und die Programmvariable `ra_5_2`.

Im Falle von Feldern muss sowohl aus den PL-Variablen als auch aus den Programmvariablen hervorgehen, welches Feldelement angesprochen wird.

Für die Programmiersprache C liegt die Programmvariable zu `a[k]` fest. Die zugehörige PL-Variable trägt die Bezeichnung $a(k)$. Der Feldindex $k.v$ liegt im Intervall $0 \leq k.v \leq a\,.l - 1$.

Wir besprechen nun noch die Behandlung griechischer Buchstaben, die als PL-Variablen Verwendung finden. Der Programm-Variablen-Bezeichner ist der in kleinen arabischen Lettern ausgeschriebene griechische Buchstabe. Der Ist-Wert der PL-Variablen ist der griechische Buchstabe mit der zusätzlichen Endung *.v* .

Als abschließendes Beispiel betrachten wir das Element der Zeile m und der Spalte n der Matrix $\boldsymbol{A}^e_R$, deren Elemente Felder sind und vom Feldindex `delta` abhängen. Der Sollwert lautet $a^e_{R\,mn}(\delta)$, die PL-Variable $a^e_{R\,mn}(\delta.v).v$ und die Programmvariable `aeR_m_n[delta]`.

Im Zusammenhang mit prädikatenlogischen Ausdrücken werden wir die Identität

$$\mathsf{P}_1 \;\equiv\; \mathsf{P}_2\,[{}^{a.v}_{b.v}] \tag{5.2}$$

verwenden, die so zu verstehen ist, dass der Ausdruck P_1 sich aus dem Ausdruck P_2 ergibt, indem dort jede auftretende prädikatenlogische Variable $a.v$ durch $b.v$ ersetzt wird.

5.2 Das Konzept der formalen Methoden

Um das Konzept der formalen Methoden zu erläutern, betrachten wir eine Folge von Programmvariablen, die wir als Datenumgebung bezeichnen wollen. Als Vorbedingung wird ein Zustand einer Datenumgebung bezeichnet, in dem sich die Datenumgebung vor der Ausführung der Anweisung befinden muss. Die Nachbedingung gibt den Zustand an, in dem sich die Datenumgebung nach der Ausführung der Anweisung befinden muss. Die Nachbedingung ist i.d.R. abhängig vom Datenzustand vor Ausführung der Anweisung. Das Prinzip der formalen Methoden basiert nun darauf, das Verhalten einer Anweisung auf den Datenzustand vor und nach der Anweisung durch die PL-Formel

$$\{\mathsf{V}\}S\{\mathsf{P}\} \tag{5.3}$$

zu beschreiben. Hierin sind die Vorbedingung V und die Nachbedingung P selber PL-Formeln. An Stelle des Begriffs Bedingung wird auch der Begriff Zusicherung verwandt.

Der fundamentale Satz der formalen Methoden ist der folgende: eine Bedingung V ist eine (gewöhnliche) Vorbedingung der Nachbedingung P bezüglich der Programmanweisung S, falls die Wahrheit von V vor der Ausführung von S die Wahrheit von P nach Ausführung sicherstellt, d.h. die PL-Formel

$$\mathsf{V} \Longrightarrow \{\mathsf{V}\}S\{\mathsf{P}\} \tag{5.4}$$

wahr ist. In dem Fall erfüllt die Anweisung die Spezifikation und wird als korrekt bezeichnet. In der Literatur wird noch zwischen partieller Korrektheit und totaler Korrektheit differenziert. Eine Korrektheitsformel $\{\mathsf{V}\}S\{\mathsf{P}\}$ heißt partiell korrekt, wenn die Anweisung aus jeder zulässigen Vorbedingung in eine beliebige Nachbedingung $\{\mathsf{P}\}$ terminiert. Eine Korrektheitsformel $\{\mathsf{V}\}S\{\mathsf{P}\}$ heißt total korrekt, wenn die Anweisung aus jeder zulässigen Vorbedingung in eine vorgegebene Nachbedingung $\{\mathsf{P}\}$ terminiert. In dieser Arbeit interessieren wir uns nur für die totale Korrektheit. Aus diesem Grund werden wir auf das Adjektiv total in Zukunft verzichten .

Für die Beweisführung sind prinzipiell 2 Varianten anwendbar. Ausgehend von einer gegebenen Vorbedingung kann man den Nachweis führen, dass die Anweisung der gegebenen Nachbedingung genügt. Andererseits kann man von der Nachbedingung im Rückschluss den Nachweis führen, dass die Anweisung der Vorbedingung genügt. Oft ist es auch so, dass nur die Nachbedingung vorgeschrieben ist. Durch Anwendung der formalen Methoden wird dann eine Vorbedingung gefunden, die einen Teil der Spezifikation bildet.

In der Praxis sind die Bedingungen durch logische Verknüpfungen einzelner Formeln definiert. Die Formeln können abhängig von reellen Zahlen sein, wie das folgende Beispiel zeigt. Wir betrachten eine Anweisung, die die Wurzel einer Zahl vom Datentyp `float` ziehen soll.

Anweisung S	:	`y = sqrt(x)`
Vorbedingung V	:	$\mathrm{Max_f} \geq x.v \geq 0 \wedge \alpha = \sqrt{x.v}$
Nachbedingung P	:	$y.v = \alpha$

In der Kompaktnotation

$$\{\mathrm{Max_f} \geq x.v \geq 0 \wedge \alpha = \sqrt{x.v}\}\texttt{y = sqrt(x)}\{y.v = \alpha\}\,. \tag{5.5}$$

In diesem einführenden Beispiel wurden einige Bedingungen unterdrückt.

Wie jede mathematische Theorie, benötigt auch die Verifikation von Software als Basis Axiome. Diese Basis bildet sich aus Elementaranweisungen, Axiomen zur Verknüpfung von Elementaranweisungen und Axiomen zur Vereinfachung der Beweisführung.

Elementaranweisungen

Elementaranweisungen sind Anweisungen, die axiomatisch definiert sind. Man unterscheidet nach Einzelanweisungen und Blockanweisungen. Blockanweisungen sind dadurch gekennzeichnet, dass sich innerhalb der Blockanweisung noch Anweisungen befinden.

Die von uns benötigten Elementaranweisungen sind Variablendeklarationsanweisungen, Zuweisungen, algebraische Operatoren (Addition, Subtraktion, Multiplikation), Boolsche Operatoren (Und, Oder, Negation), Klammeroperatoren und for-Schleifenanweisungen. Eine axiomatische Definition dieser Elementaranweisungen in der Programmiersprache C erfolgt im Kapitel Gleichung (5.3).

Anweisungssequenzen

Mittels einer Elementaranweisung lassen sich keine komplexen Aufgaben lösen. Ein Programm wird daher aus einer Sequenz von Elementaranweisungen bestehen. Ein Axiom welches die PL-Ausdrücke zweier Anweisungen zu einem PL-Ausdruck der Anweisungssequenz verknüpft, wird nun eingeführt.

Dazu betrachten wir die beiden Anweisungen S_1 und S_2, die nacheinander abgearbeitet werden sollen. Die Frage ist, wie die Korrektheit der zusammengesetzten Anweisung $S_1;\ S_2$ bewiesen werden kann. Die Vorbedingung dieser Anweisungssequenz sei V und die Nachbedingung P, d.h.

$$\{\mathsf{V}\}\, S_1; S_2\, \{\mathsf{P}\}\ . \tag{5.6}$$

Die Nachbedingung von S_1 bezeichnen wir mit R. Sie stellt gleichzeitig die Vorbedingung von S_2 dar. Für den Nachweis der Korrektheit kann man auf das folgende Axiom zurückgreifen

$$\{\mathsf{V}\}\, S_1\, \{\mathsf{R}\} \wedge \{\mathsf{R}\}\, S_2\, \{\mathsf{P}\} \Longrightarrow \{\mathsf{V}\}\, S_1; S_2\, \{\mathsf{P}\}\ , \tag{5.7}$$

d. h. der Nachweis kann auf das Zeigen der Richtigkeit von $\{\mathsf{V}\}\, S_1\, \{\mathsf{R}\}$ und $\{\mathsf{R}\}\, S_2\, \{\mathsf{P}\}$ zurückgeführt werden.

Durch Sequenzen von Elementaranweisungen kann man neue Anweisungen definieren. Für den Nachweis der Korrektheit der Anweisung darf man nur auf Axiome oder als korrekt bewiesene Anweisungen zurückgreifen. Für die effiziente Durchführung von Korrektheitsbeweisen bietet es sich daher an, eine Bibliothek neuer Anweisungen zu erstellen oder zu erwerben. Diese Module sollten einen möglichst allgemein gültigen Charakter besitzen.

Stärkung, Schwächung

Die Axiome „Stärkung der Vorbedingung“und „Schwächung der Nachbedingung“, welche auch gelegentlich als Konsequenzaxiome bezeichnet werden, können zur Vereinfachung von Beweisen genutzt werden.

Zur Erläuterung der Axiome gehen wir von einer Vorbedingung V_1 und einer Nachbedingung P_1 bzgl. der Anweisung S aus. Das Axiom über die „Stärkung der Vorbedingung“ besagt, dass V auch eine gültige Vorbedingung bzgl. S und P_1 ist, wenn V schärfer als V_1 ist. Anders ausgedrückt, dass die Vorbedingung V die Vorbedingung V_1 impliziert , d.h.

$$\mathsf{V} \Longrightarrow \mathsf{V}_1\,. \tag{5.8}$$

Die Menge der Zustände, die V erfüllen, ist dann nicht größer als die Menge der Zustände, die V_1 erfüllen. Ist z z.B. eine komplexe Zahl und die Vorbedingung lautet $\mathsf{V}_1 \equiv |z| < 3$, dann ist mit der schärferen Vorbedingung $\mathsf{V} \equiv |z| < 2$ auch V_1 erfüllt.

Das Axiom über die „Schwächung der Nachbedingung“ besagt, dass P auch eine gültige Nachbedingung bzgl. S und $\{\mathsf{V}_1\}$ ist, wenn P schwächer als P_1 ist. Anders ausgedrückt, dass die Bedingung P_1 die Bedingung P impliziert , d.h.

$$\mathsf{P}_1 \Longrightarrow \mathsf{P}\,. \tag{5.9}$$

Ist z z. B. eine komplexe Zahl und die Nachbedingung lautet $\mathsf{P}_1 \equiv |x| < 1$, dann ist die schwächere Nachbedingung $\mathsf{P} \equiv |x| < 2$ auch eine gültige Nachbedingung.

Beide Axiome lassen sich wie folgt kompakt darstellen

$$[(\{\mathsf{V}_1\}\, S\, \{\mathsf{P}_1\}) \wedge (\mathsf{V} \Longrightarrow \mathsf{V}_1) \wedge (\mathsf{P}_1 \Longrightarrow \mathsf{P})] \Longrightarrow \{\mathsf{V}\}\, S\, \{\mathsf{P}\}\,. \tag{5.10}$$

5.3 Eine axiomatische Definition von Elementaranweisungen der Programmiersprache C

Selbst bei so populären Sprachen wie C oder C++ ist die Syntax im Detail beschrieben. Eine formale Spezifikation der Semantik existiert aber nicht. Die Bedeutung der Anweisungen wird meist intuitiv den

Anweisungsbezeichnern entnommen und nicht einer formalen Definition. Wir werden daher die Semantik axiomatisch, formal und praxisgerecht definieren. Als Programmiersprache verwenden wir dazu C im ANSI-C-Standard, da die entwickelte Software in das Programmpaket SPACE eingebunden werden soll, das wir im Kapitel 3 vorgestellt haben. Die Definition erfolgt auf Grundlage von [Inte99]. Dabei werden wir uns auf die Definition derjenigen Elementaranweisungen beschränken, die wir benötigen, um unsere Aufgabenstellung zu lösen. Da in unserem Fall der Sicherheitsaspekt im Vordergrund steht, werden wir nicht jede Anweisung, die zur Verkürzung der Programmlaufzeit führt, verwenden.

5.3.1 Einzelanweisungen

Deklaration

Die Deklarationsanweisung einer skalaren Variablen lautet

$$vartyp\ \mathtt{a;}\ . \tag{5.11}$$

Die Vorbedingung ergibt sich aus der abgeänderten Nachbedingung. Konjunktiv kommt die Existenz des angeforderten Datentyps und die Tatsache, dass der Bezeichner gültig und noch nicht verwendet worden ist, hinzu. Die PL-Variable $a.typ$ ist durch $vartyp$ zu ersetzen. Zusätzlich wird $a\,.l$ durch -1 ersetzt, um zu kennzeichnen, dass es sich um eine skalare Variable handelt. Formal erhalten wir das Deklarationsaxiom zu

$$\{\mathsf{P}\begin{bmatrix} a\,.typ \\ vartyp \end{bmatrix}\begin{bmatrix} a\,.l \\ -1 \end{bmatrix} \wedge vartyp \in \mathcal{D} \wedge \mathtt{a} \notin \mathcal{N} \wedge \mathtt{a} \in \mathcal{B}\}\ vartyp\ \mathtt{a;}\ \{\mathsf{P}\}\ . \tag{5.12}$$

Deklaration von Feldern

Die Deklarationsanweisung einer Feldvariablen lautet

$$vartyp\ \mathtt{x}[laenge];\ . \tag{5.13}$$

Ein eindimensionales Feld stellt einen Vektor dar. Wir verwenden für die PL-Variablen die Bezeichnungen

$$\boldsymbol{x}.v = [x_0.v\ ,\ x_1.v\ ,\ \ldots\ , x_{x\,.l-2}.v, x_{x\,.l-1}.v]^{\mathrm{T}} \quad \text{und} \quad x_\nu.v = \boldsymbol{x}(\nu).v\ \ . \tag{5.14}$$

Jede vektorielle PL-Variable trägt neben den Attributen der Variablen zusätzlich die Länge $x\,.l$. Die vektorielle PL-Variable stellt dann ein Quadrupel dar, wobei $\boldsymbol{x}.v$ selber ein Vektor mit den Koordinaten $x.v_\nu\ ,\ \nu = 1, 2, \ldots, x\,.l - 1$ ist. Die Vorbedingungen sind denen der Deklaration von skalaren Variablen ähnlich. Hinzu kommt noch die Bedingung, dass die Länge des Vektors positiv sein muss. Das Axiom lautet daher

$$\mathsf{P}\begin{bmatrix} x\,.typ \\ vartyp \end{bmatrix}\begin{bmatrix} x\,.l \\ laenge \end{bmatrix} \wedge vartyp \in \mathcal{D} \wedge x \notin \mathcal{N} \wedge x \in \mathcal{B} \wedge laenge > 0\, vartyp\ \mathtt{x[}laenge\mathtt{]\ ;}\ \{\mathsf{P}\}\ . \tag{5.15}$$

Zuweisung

Die Zuweisungsanweisung der Form

$$\mathtt{a\ =\ b;}\ . \tag{5.16}$$

Hierbei muss unterschieden werden zwischen einer Programmvariablen `b` und einer Konstanten b. Gehen wir zunächst davon aus, dass `b` eine Programmvariable ist, lautet das Axiom

$$\{\mathsf{P}[^{a.v}_{b.v}] \wedge \mathrm{Max}_{a.typ} \geq |b.v| \wedge a\,.l = b\,.l = -1 \wedge a.typ \in \mathcal{D} \wedge b.typ = a.typ\}\,\mathtt{a\ =\ b;}\ \{\mathsf{P}\}\ . \tag{5.17}$$

Im Falle einer Konstanten b lautet das Axiom für den Datentyp `int`

$$\{\mathsf{P}[{}^{a.v}_{b}] \wedge \mathrm{Max_i} \geq |b| \wedge a.l = -1 \wedge a.typ = \texttt{int}\}\,\texttt{a = b;}\,\{\mathsf{P}\}\,. \tag{5.18}$$

Die Zuweisung gilt nicht für float-Variablen. Im Falle von float-Variablen ist der Konstanten das Postfix f hinzuzufügen

$$\{\mathsf{P}[{}^{a.v}_{b.v}] \wedge \mathrm{Max_f} \geq |b| \wedge a.l = -1 \wedge a.typ = \texttt{float}\}\,\texttt{a = bf;}\,\{\mathsf{P}\}\,. \tag{5.19}$$

Zuweisungen bei Feldern

Die Zuweisung der Form `x[k]=y[m];` folgt dem Axiom

$$\begin{aligned}&\{\mathsf{P}\left[{}^{x_{k.v}.v}_{y_{m.v}.v}\right] \wedge \mathrm{Max}_{a.typ} \geq |y_{m.v}.v| \wedge 0 \leq k.v \leq x.l - 1 \wedge 0 \leq m.v \leq y.l - 1 \wedge x.typ = y.typ \wedge x.typ \in \mathcal{D} \wedge \\ &1 \leq k.v \leq \mathrm{Max_i} \wedge 1 \leq m.v \leq \mathrm{Max_i} \wedge k.l = -1 \wedge m.l = -1 \wedge k.typ = \texttt{int} \wedge m.typ = \texttt{int}\} \\ &\texttt{x[k]=y[m];}\,\{\mathsf{P}\}\,.\end{aligned} \tag{5.20}$$

Algebraische Grundoperationen

Die konsequente Erweiterung der Zuweisung entsteht, wenn die rechte Seite nun nicht mehr durch eine Konstante oder Variable gegeben ist, sondern durch mehrere mit algebraischen Grundoperationen verknüpfte Variablen bestimmt ist. Weiterhin ist es so, dass die Variablen der rechten Seite auch durch Konstanten ersetzt werden dürfen.

Unter den Variablen verstehen wir hier sowohl skalare Variablen, als auch Komponenten von Feldern, wobei dann in der Vorbedingung noch konjunktive Verknüpfungen bezüglich der PL-Variablen der Längen und der Typen hinzukommen. Dies soll nicht wiederholt werden, da es bei der Zuweisung schon erläutert wurde.

In der Programmiersprache C sind Operationen und Zuweisungen zwischen 'gemischten' Datentypen erlaubt. Insofern könnte man daraus weitere Axiome für die algebraischen Grundoperationen ableiten. Da wir diese im Verlauf der Arbeit nicht benötigen, verzichten wir darauf. Ebenso verzichten wir auf die Division, da der WDF-Algorithmus keine enthält.

Addition / Subtraktion von Ganzzahlen

Rundungsfehler treten nur dann auf, wenn eine Bereichsüberschreitung vorliegt

$$\{\mathsf{P}[{}^{a.v}_{(b.v \pm c.v)}] \wedge \mathrm{Max_i} \geq |b.v \pm c.v| \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\,\texttt{a=b} \pm \texttt{c}\,\{\mathsf{P}\}\,. \tag{5.21}$$

Addition / Subtraktion von Gleitkommazahlen des Typs float

Rundungsfehler treten i.Allg. auf. Zudem ist eine Bereichsüberschreitung möglich

$$\begin{aligned}&\{\mathsf{P}[{}^{a.v}_{(b.v \pm c.v + \varepsilon(b.v \pm c.v))}] \wedge \mathrm{Max_f} \geq |b.v \pm c.v| \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{float}\} \\ &\texttt{a=b} \pm \texttt{c}\,\{\mathsf{P}\}\,.\end{aligned} \tag{5.22}$$

Multiplikation von Ganzzahlen

Rundungsfehler treten nur dann auf, wenn eine Bereichsüberschreitung vorliegt

$$\{\mathsf{P}[^{a.v}_{(b.v \cdot c.v)}] \wedge \mathrm{Max_i} \geq |b.v \cdot c.v| \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}$$
$$\texttt{a=b*c}\,\{\mathsf{P}\}\,. \tag{5.23}$$

Multiplikation von Gleitkommazahlen des Typs float

Sowohl Rundungsfehler als auch eine Bereichsüberschreitung treten i.Allg. auf.

$$\{\mathsf{P}[^{a.v}_{(b.v \cdot c.v + \varepsilon(b.v \cdot c.v))}] \wedge \mathrm{Max_f} \geq |b.v \cdot c.v| \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{float}\}$$
$$\texttt{a=b*c}\,\{\mathsf{P}\}\,. \tag{5.24}$$

Die im Axiom (5.22) und im Axiom (5.24) auftretende Funktion ε stellt den Rundungsfehler dar. Dieser ist zwar kaum explizit auszudrücken, jedoch kann eine betragsmäßige Obergrenze angegeben werden [Rumm98], [Babe95], [Inte99] und Anhang D.

Logische Operatoren

Im Zusammenhang mit den hier definierten logischen Operatoren wird der Variablen $a.v$ der Wert 1 zugewiesen, falls die Aussage wahr ist, andernfalls 0

$$\{\mathsf{P}[^{a.v}_{(b.v \wedge c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b \&\& c;}\,\{\mathsf{P}\} \tag{5.25}$$

$$\{\mathsf{P}[^{a.v}_{(b.v \vee c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b || c;}\,\{\mathsf{P}\} \tag{5.26}$$

$$\{\mathsf{P}[^{a.v}_{(b.v < c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b < c;}\,\{\mathsf{P}\} \tag{5.27}$$

$$\{\mathsf{P}[^{a.v}_{(b.v \leq c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b <= c;}\,\{\mathsf{P}\} \tag{5.28}$$

$$\{\mathsf{P}[^{a.v}_{(b.v > c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b > c;}\,\{\mathsf{P}\} \tag{5.29}$$

$$\{\mathsf{P}[^{a.v}_{(b.v \geq c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b >= c;}\,\{\mathsf{P}\} \tag{5.30}$$

$$\{\mathsf{P}[^{a.v}_{(b.v = c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b == c;}\,\{\mathsf{P}\} \tag{5.31}$$

$$\{\mathsf{P}[^{a.v}_{(b.v \neq c.v)}] \wedge a.l = b.l = c.l = -1 \wedge a.typ = b.typ = c.typ = \texttt{int}\}\texttt{a = b != c;}\,\{\mathsf{P}\} \tag{5.32}$$

$$\{\mathsf{P}[^{a.v}_{[\neg b.v]}] \wedge a.l = b.l = -1 \wedge a.typ = b.typ = \texttt{int}\}\texttt{a = !b;}\,\{\mathsf{P}\}\,. \tag{5.33}$$

5.3.2 Blockanweisungen

For-Schleifenanweisung

Die For-Schleifenanweisung lautet

$$\texttt{for (}\, S_1 \,\texttt{;}\, b \,\texttt{;}\, S_3 \,\texttt{)\{}\; S_4 \;\texttt{\};} \;. \tag{5.34}$$

Hierin stellt S_1 die Anweisung zur Initialisierung einer Schleifenvariablen, $b.v \neq 0$ die Schleifenbedingung und S_3 die Anweisung zur Reinitialisierung der Schleifenvariablen dar. S_4 ist die Anweisung, die innerhalb der Schleife ausgeführt wird.

Der Programmablaufplan der For-Schleifenanweisung ist mit Bedingungen im Bild 5.1 dargestellt. Mithilfe der so genannten Schleifeninvarianten I gelingt die Beweisführung bei Schleifenanweisungen. Wie der Name schon sagt, ändert sich die Schleifeninvariante während eines Schleifendurchlaufs nicht. Diese Eigenschaft macht man sich beim Beweis der Korrektheit zu nutze. Die konjunktive Verknüpfung der Invarianten und der negierten Schleifenbedingung stellt die Nachbedingung dar. Zur Erfüllung der gestellten Aufgabe (Erreichung der Nachbedingung P aus der Vorbedingung V) sind mehrere Schleifeninvarianten I geeignet. Die Schleifeninvariante muss vor der Programmausführung bekannt sein. Zur Förderung des Verständnisses dient die Darstellung der Anweisung mit den Bedingungen

$\{\mathsf{V}\}$	
	`for (`S_1 `;` b `;` S_3 `){`
$\{\mathsf{I}\}$	
$\{\mathsf{I} \wedge (b.v \neq 0)\}$	
	S_4
$\{\mathsf{Z}\}$	
	`};`
$\{\mathsf{P}\}$	

.

Das Axiom lautet in der Kompaktnotation

$$\begin{aligned} &\{\mathsf{V}\}\, S_1\, \{\mathsf{I}\} \wedge \{\mathsf{I} \wedge (b.v \neq 0)\}\, S_4;\; S_3\, \{\mathsf{I}\} \wedge b.typ = \texttt{int} \wedge b\,.l = -1 \\ &\Rightarrow\; \{\mathsf{V}\}\, \texttt{for(}S_1\texttt{;}b\texttt{;}S_3\texttt{)\{}S_4\texttt{\}}\; \{\mathsf{I} \wedge (b.v = 0)\} \end{aligned} \tag{5.35}$$

und es gilt $\mathsf{P} \equiv \mathsf{I} \wedge b.v = 0$. Zu beachten ist, dass die Klammern um S_4 syntaktisch zum C-Code gehören und nicht zu einer Bedingung.

5.4 Behandlung von Ausdrücken

In die bislang eingeführten Axiome lassen sich nur Elementaranweisungen einsetzen, welche keine arithmetischen Ausdrücke, sondern nur einfache Programmvariablen besitzen. Um nun auch arithmetische und logische Ausdrücke, wie sie im Abschnitt 5.3.1 beschrieben sind verwenden zu können, werden noch weitere Axiome nötig. Mithilfe dieser Axiome kann eine Programmanweisung, welche einen arithmetische Ausdruck besitzt, in mehrere einfachere Anweisungen zerlegt werden. Die in diesen Axiomen auftretenden Variablen b und c stellen PL-Variablen dar, welche nicht in den PL-Formeln V, P, oder I vorkommen dürfen.

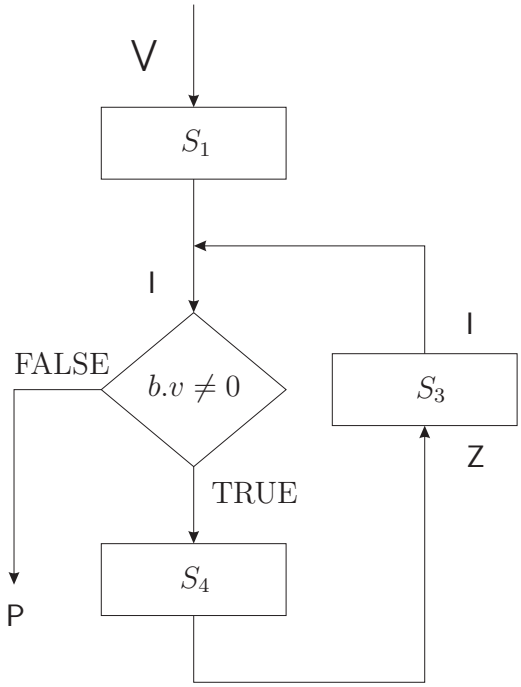


Bild 5.1: for-schleife

5.4.1 Einzelanweisungen

Wendet man den unären Negations-Operator auf einen Ausdruck *expr* an, so kann der Ausdruck vor der Anwendung des Operators ausgewertet werden

$$\{\mathsf{V} \wedge b.l = -1 \wedge b.typ = \texttt{int}\}\ \texttt{b = } expr\ \texttt{; a = !b;}\{\mathsf{P}\} \Rightarrow \{\mathsf{V}\}\texttt{a = !}expr\,\{\mathsf{P}\}\ . \tag{5.36}$$

Ebenso können die Ausdrücke $expr_1$ und $expr_2$ ausgewertet werden, bevor man einen beliebigen der definierten binären logischen Operatoren *op* auf die Ausdrücke anwendet. Wir erhalten dabei drei Fälle. Zunächst zwei Ausdrücke als Operanden

$$\begin{aligned}&\{\mathsf{V} \wedge b.typ = vartype1 \wedge c.typ = vartype2 \wedge b.l = -1 \wedge c.l = -1\}\\ &\texttt{b = }expr_1\texttt{; c = }expr_2\texttt{; a = b }\ op\ \texttt{ c}\,\{\mathsf{P}\}\\ &\Rightarrow \{\mathsf{V}\}\,\texttt{a=(}expr_1\texttt{)}\ op\ \texttt{(}expr_2\texttt{)}\,\{\mathsf{P}\}\ .\end{aligned} \tag{5.37}$$

Für $vartype1$ bzw.$vartype2$ darf hier ein beliebiger Variablentyp aus $\mathcal{D}$ eingesetzt werden. Die Typüberprüfung erfolgt dann später bei der Auflösung von $expr1$ und $expr2$ sowie bei der Zuweisung `a=b` op `c`. Die anderen beiden Fälle sind durch jeweils einen Operanden als Ausdruck und einen als Variable gekennzeichnet, d. h.

$$\{\mathsf{V} \wedge b.typ = vartype \wedge b.l = -1\}\,\texttt{b = }expr\texttt{; a = b }\ op\ \texttt{ c}\,\{\mathsf{P}\} \Rightarrow \{\mathsf{V}\}\,\texttt{a=(}expr\texttt{)}\ op\,\texttt{c}\,\{\mathsf{P}\} \tag{5.38}$$

und

$$\{\mathsf{V} \wedge c.typ = vartype \wedge c.l = -1\}\,\texttt{c = }expr\texttt{; a = b }\ op\ \texttt{ c}\,\{\mathsf{P}\} \Rightarrow \{\mathsf{V}\}\,\texttt{a=b}\ op\ \texttt{(}expr\texttt{)}\,\{\mathsf{P}\}\ . \tag{5.39}$$

5.4.2 Blockanweisungen

5.5 Einige Sätze

5.5.1 Skalarprodukt mit Variablen

Der folgende Satz stammt aus [Rumm98]. Eine Anweisungssequenz zur Berechnung des Skalarprodukts der Vektoren $\boldsymbol{b}.v$ und $\boldsymbol{c}.v$ mit der Länge N und den reellen Koordinaten $b_1.v, b_2.v, \ldots, b_N.v$, $c_1.v, c_2.v, \ldots, c_N.v$ lautet

$$\begin{array}{l}\{\mathsf{V}\}\\ \qquad \texttt{a = ((}\cdots\texttt{(b\_1 * c\_1)}\cdots\texttt{)+(b\_}N\texttt{-1 * c\_}N\texttt{-1))+(b\_}N\texttt{ * c\_}N\texttt{)}\\ \{\mathsf{P}\}\,.\end{array} \tag{5.40}$$

Satz: Eine gültige Vorbedingung ist

$$\begin{aligned}\mathsf{V} \equiv\ & \mathsf{P}\left[\begin{matrix}a.v\\ (\boldsymbol{b}^{\mathrm{T}}.v\,\boldsymbol{c}.v)\end{matrix}\right] \wedge \bigwedge_{n=1}^{N} |b_n.v\ c_n.v| \le \mathrm{Max_f} \bigwedge_{n=2}^{N} \Big|\sum_{\nu=1}^{n} b_\nu.v\ c_\nu.v\Big| \le \mathrm{Max_f} \wedge a.l = -1\\ & \bigwedge_{n=1}^{N} b_n.l = -1 \bigwedge_{n=1}^{N} c_n.l = -1 \wedge a.typ = \texttt{float} \wedge \bigwedge_{n=1}^{N} a.typ = b_n.typ \bigwedge_{n=1}^{N} a.typ = c_n.typ\end{aligned} \tag{5.41}$$

wobei

$$\boldsymbol{b}^{\mathrm{T}}.v\,\boldsymbol{c}.v = \sum_{n=1}^{N} b_n.v\ c_n.v$$

gilt. Der angegebene Satz lässt sich mit Hilfe von Axiom (5.22) und Axiom (5.24) beweisen [Rumm98], S.121-124. Wichtig ist in diesem Zusammenhang, dass für den Programmausdruck weder das Kommutativgesetz noch das Assoziativgesetz gelten. Die Klammern dürfen nicht weggelassen werden, da die Reihenfolge der Auswertung von Unterausdrücken gleicher Priorität in C nicht definiert ist [Brey96], [Hero99].

5.5.2 Skalarprodukt mit Feldvariablen

Eine Anweisungssequenz zur Berechnung des Skalarprodukts der Vektoren $\boldsymbol{b}(x).v$ und $\boldsymbol{c}(y).v$ mit der Länge N und den reellen Koordinaten $b_1(x).v, b_2(x).v, \ldots, b_N(x).v, c_1(y).v, c_2(y).v, \ldots, c_N(y).v$ lautet

$$\begin{array}{l}\{\mathsf{V}\}\\ \qquad \texttt{a[z] = ((}\cdots\texttt{(b\_1[x] * c\_1[y])}\cdots\texttt{)+(b\_N[x] * c\_N[y])}\\ \{\mathsf{P}\}\,.\end{array} \tag{5.42}$$

Satz: Eine gültige Vorbedingung ist

$$\begin{aligned}
\mathsf{V} \equiv\ & \mathsf{P}\left[\begin{matrix} a(z.v) \\ \boldsymbol{b}^{\mathrm{T}}_{(x.v).v}\,\boldsymbol{c}_{(y.v).v} \end{matrix}\right] \wedge \bigwedge_{n=1}^{N} |b_n(x.v).v\ c_n(y.v).v| \leq \mathrm{Max_f} \bigwedge_{n=2}^{N} \Big|\sum_{\nu=1}^{n} b_\nu(x.v).v\ c_\nu(y.v).v\Big| \leq \mathrm{Max_f} \\
& \wedge \quad 0 \leq z.v \leq a.l - 1 \wedge 0 \leq x.v \leq b.l - 1 \wedge 0 \leq y.v \leq c.l - 1 \\
& \wedge \quad a.typ = \texttt{float} \wedge \bigwedge_{n=1}^{N} a.typ = b_n.typ \bigwedge_{n=1}^{N} a.typ = c_n.typ \\
& \wedge \quad 0 \leq z.v \leq \mathrm{Max_i} \wedge 0 \leq x.v \leq \mathrm{Max_i} \wedge 0 \leq y.v \leq \mathrm{Max_i} \\
& \wedge \quad z.l = y.l = x.l = -1 \wedge z.typ = x.typ = y.typ = \texttt{int}
\end{aligned} \tag{5.43}$$

wobei

$$\boldsymbol{b}^{\mathrm{T}}(x.v).v\boldsymbol{c}(y.v).v = \sum_{n=1}^{N} b_n(x.v).v\ c_n(y.v).v$$

gilt. Einer der beiden Vektoren kann auch durch einen Vektor mit Konstanten ersetzt werden. In dem Fall erübrigen sich die Forderungen an den ersetzten Vektor.

Kapitel 6

Der Algorithmus

Ziel dieses Kapitels ist es, einen durch formale Methoden verifizierten Algorithmus durch Verknüpfen von Elementaranweisungen der Programmiersprache C anzugeben, der ein lineares, konstantes MDWDF für ein quaderförmiges Berechnungsgebiet simuliert. Die Anweisungssequenz soll eine Abtastschicht berechnen und einen, gemäß Kapitel 3.5 spezifizierten, Funktionsbaustein des Programmpaketes SPACE darstellen, [Voll04b].

Inhalt des Kapitels 6.1 ist es, Vor- und Nachbedingungen des Algorithmus anzugeben. Dies entspricht im Softwareentwicklungsprozess dem Übergang von der Anforderungsspezifikation zur formalen Spezifikation. Für den formalen Korrektheitsbeweis verzichten wir allerdings aus Gründen der Übersichtlichkeit auf eine Berücksichtigung der Laufzeit. In den Kapiteln 6.2 bis 6.7 erfolgt zu den Teilanweisungen die manuelle Softwaresynthese. Zudem wird die Erfüllung der in Kapitel 6.1 festgelegten Vor- und Nachbedingungen jeweils nachgewiesen. Im Kapitel 6.8 erfolgt schließlich eine Zusammenfassung und Bewertung der Ergebnisse.

6.1 Gesamtanweisung und Nachbedingung

In diesem Unterkapitel werden wir zunächst die relevanten Ergebnisse der vorherigen Kapitel rekapitulieren, insbesondere die Voraussetzungen für die Gültigkeit der folgenden Ergebnisse wiederholen. Um einen leichteren Übergang zur formalen Spezifikation zu ermöglichen, werden anschließend die Gleichungen aufbereitet. Zum einen werden wir die Wellengrößen zu dimensionslosen Größen normieren und zum anderen wird die Darstellung der Gültigkeitsbereiche mittels Allquantoren auf konjunktiv verknüpfte Prädikate umgeformt. Es folgen Definitionen von Prädikaten, die eine einfacherer Schreibweise ermöglichen. Danach geben wir einen Algorithmus an und illustrieren diesen anhand von Programmablaufplänen. Diesen Grobentwurf zerlegen wir wiederum in sequentiell abzuarbeitende, syntaktisch zusammengehörige Teilanweisungen. Abschließend legen wir die im Kapitel 6.2 bis 6.7 benötigten Nach-, Zwischen- und Vorbedingungen fest.

Das Vorgehen und die Voraussetzungen zur Einbindung mehrdimensionaler linearer konstanter Wellendigitalfilter bei quaderförmigen Berechnungsgebiet in das Programmpaket SPACE nach Kapitel 3 wird wie folgt zusammengefasst :

- Ein Funktionsbaustein stellt den algebraischen Teil eines mehrdimensionalen Wellendigitalfilters dar und jeder Aufruf eines Funktionsbausteins entspricht der Berechnung einer Abtastschicht des Wellendigitalfilters und somit dem Fortschreiten der physikalischen Zeit.
- Die Eingänge des Funktionsbausteins sind die Quellen-Wellen des Wellendigitalfilters.
- Die Ausgänge des Funktionsbausteins sind die Spannungen und Ströme an den Toren des Referenznetzes.

- Wir gehen davon aus, dass das Verhalten der Feldgrößen am Rand algebraisch modelliert wird.
- Die abgeänderte Schnittstelle für Funktionsbausteine des Programmpaketes SPACE nach Kapitel 3.5 wird genutzt.

Die Gleichungen (2.219), (2.221), (2.228) und (3.16) aus den Kapiteln 2.9, 2.10 und 3.5, die ein mehrdimensionales Wellendigitalfilter beschreiben, lauten in den Variablen $\boldsymbol{\mu}$

$$\boldsymbol{b}_e(\boldsymbol{\mu}) = \boldsymbol{L}_q\boldsymbol{b}_q(\boldsymbol{\mu}) + \boldsymbol{L}_v\boldsymbol{b}_v(\boldsymbol{\mu}) + \boldsymbol{L}_e\boldsymbol{b}_e(\boldsymbol{\mu}) \quad \forall \quad \boldsymbol{\mu} \in \mathcal{G} \tag{6.1}$$

$$\boldsymbol{a}_{\mathrm{FB}}(\boldsymbol{\mu}) = \boldsymbol{A}\boldsymbol{b}(\boldsymbol{\mu}) \quad \forall \quad \boldsymbol{\mu} \in \mathcal{G}, \tag{6.2}$$

$$\boldsymbol{a}_v^{\kappa}(\boldsymbol{\mu}) = \boldsymbol{P}_{v^{\kappa}e}\boldsymbol{b}_e(\boldsymbol{\mu}) \quad \forall \quad (\kappa = 1, 2, \ldots k'-1) \wedge (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_{\kappa} \in \mathcal{G}_0) \tag{6.3}$$

$$\boldsymbol{b}_v^{\kappa}(\boldsymbol{\mu} + \boldsymbol{h}_{\kappa}) = \boldsymbol{P}_{v^{\kappa}e}\boldsymbol{b}_e(\boldsymbol{\mu}) \quad \forall \quad (\kappa = 1, 2, \ldots k') \wedge (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_{\kappa} \in \mathcal{G}) \tag{6.4}$$

$$\boldsymbol{b}_v^{\kappa}(\boldsymbol{\mu} + \mathbf{e}_7) = \boldsymbol{R}^{\kappa}\boldsymbol{a}_v^{(\kappa+3)}(\boldsymbol{\mu}) \quad \forall \quad (\kappa = 1, 2, 3) \wedge (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_{\kappa} \in \mathcal{G}_0) \quad \text{und} \tag{6.5}$$

$$\boldsymbol{b}_v^{\kappa}(\boldsymbol{\mu} + \mathbf{e}_7) = \boldsymbol{R}^{\kappa}\boldsymbol{a}_v^{(\kappa-3)}(\boldsymbol{\mu}) \quad \forall \quad (\kappa = 4, 5, 6) \wedge (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_{\kappa} \in \mathcal{G}_0),$$

wobei wir der Übersicht halber die Tilde weglassen und die gestrichenen Größen in Gleichung (6.1) durch ungestrichene Größen ersetzen. Zu beachten ist, dass die Struktur für ein beliebiges mehrdimensionales Wellendigitalfilter mit der in Gleichung (2.49) gewählten Transformationsmatrix $\boldsymbol{H}$ gleich ist. Diese Regelmäßigkeit eröffnet gerade die Möglichkeit einer strukturierten Programmsynthese und darüber hinaus eine systematische und allgemeingültige Verifizierbarkeit der Code-Struktur gegenüber der Spezifikation.

Für den Übergang von dimensionsbehafteten zu dimensionslosen Größen definieren wir

$$\begin{aligned}
&\boldsymbol{u}_N = \frac{\boldsymbol{u}}{u_B}, && u_B \quad \ldots \text{Bezugsspannung}\\
&\boldsymbol{i}_N = \frac{\boldsymbol{i}}{i_B}, && i_B \quad \ldots \text{Bezugsstrom}\\
&\boldsymbol{R}_N = \frac{\boldsymbol{R}}{R_B}, && R_B \quad \ldots \text{Bezugswiderstand}\\
&\boldsymbol{G}_N = \boldsymbol{G}R_B\\
&\boldsymbol{b}_N = \frac{\boldsymbol{b}}{b_B}, && b_B \quad \ldots \text{Bezugswelle}\\
&\boldsymbol{a}_N = \frac{\boldsymbol{a}}{b_B}\,.
\end{aligned} \tag{6.6}$$

Die Wellengrößen sind mit Kirchhoff'schen Größen über Gleichung (2.66) miteinander verknüpft. Wir fordern, dass diese Beziehung für normierte Größen erhalten bleiben soll, d. h.

$$\begin{bmatrix} \boldsymbol{a}_N \\ \boldsymbol{b}_N \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \boldsymbol{G}_N^{1/2} & \boldsymbol{R}_N^{1/2} \\ \boldsymbol{G}_N^{1/2} & -\boldsymbol{R}_N^{1/2} \end{bmatrix}\begin{bmatrix} \boldsymbol{u}_N \\ \boldsymbol{i}_N \end{bmatrix}, \tag{6.7}$$

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} = b_B \begin{bmatrix} \boldsymbol{a}_N \\ \boldsymbol{b}_N \end{bmatrix} = b_B \frac{1}{2} \begin{bmatrix} \boldsymbol{G}_N^{1/2} & \boldsymbol{R}_N^{1/2} \\ \boldsymbol{G}_N^{1/2} & -\boldsymbol{R}_N^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_N \\ \boldsymbol{i}_N \end{bmatrix} = b_B \frac{1}{2} \begin{bmatrix} \boldsymbol{G}_N^{1/2}/u_B & \boldsymbol{R}_N^{1/2}/i_B \\ \boldsymbol{G}_N^{1/2}/u_B & -\boldsymbol{R}_N^{1/2}/i_B \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{i} \end{bmatrix}, \quad (6.8)$$

$$\boldsymbol{G}^{1/2} = \boldsymbol{G}_N^{1/2} \frac{b_B}{u_B} \text{ und } \boldsymbol{R}^{1/2} = \boldsymbol{R}_N^{1/2} \frac{b_B}{i_B} . \quad (6.9)$$

Ihre Multiplikation liefert

$$\mathbf{1} = \boldsymbol{G}^{1/2}\boldsymbol{R}^{1/2} = \boldsymbol{G}_N^{1/2} \frac{b_B}{u_B} \boldsymbol{R}_N^{1/2} \frac{b_B}{i_B} \Rightarrow b_B^2 = u_B i_B \quad (6.10)$$

und mit invertierter 2. Gleichung (6.9) liefert die Multiplikation

$$\boldsymbol{G} = \boldsymbol{G}_N \frac{i_B}{u_B} \Rightarrow R_B = \frac{u_B}{i_B} . \quad (6.11)$$

Somit sind nur 2 der 4 Bezugsgrößen frei wählbar. Im Folgenden werden wir den Algorithmus in den normierten Größen entwickeln. Um die Übersicht zu wahren, wollen wir allerdings den Zusatz N für die normierten Größen in dem Algorithmus weglassen.

Nun werden die Gleichungen (6.3) - (6.5) so umgeformt, dass die Allquantoren verschwinden. Insbesondere ist es das Ziel der folgenden Überlegungen, festzustellen, in welchen Intervallen die Variablen μ_1, μ_2 und μ_3 liegen, wenn die Vektoren $\boldsymbol{\mu}$ den beiden folgenden Restriktionen unterliegen:
a) $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}_0)$
b) $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G})$.
Diese Bereiche geben uns an, für welche Werte der Variablen μ_1, μ_2 und μ_3 wir welche Gleichung zu erfüllen haben.
Gemäß Kapitel 2.4 sind $\mathcal{G}$ und $\mathcal{G}_0$ durch
$\boldsymbol{\mu} \in \mathcal{G}$: $0 \le \mu_1 \le P_{x_1} - 1 \wedge 0 \le \mu_2 \le P_{x_2} - 1 \wedge 0 \le \mu_3 \le P_{x_3} - 1$ und
$\boldsymbol{\mu} \in \mathcal{G}_0 : (0 > \mu_1 \vee \mu_1 > P_{x_1} - 1) \vee (0 > \mu_2 \vee \mu_2 > P_{x_2} - 1) \vee (0 > \mu_3 \vee \mu_3 > P_{x_3} - 1)$
definiert. Zur weiteren Beantwortung der Frage betrachten wir als Vorüberlegung die Koordinate σ des Vektors $\boldsymbol{\mu}$, wobei später die Koordinaten für $\sigma = 1, 2, 3, 4$ miteinander verknüpft werden müssen um das Gebiet zu beschreiben. Hierbei unterscheiden wir danach, ob das Element der Matrix $\boldsymbol{H}$ den Wert 1, −1 oder 0 besitzt.

1. $h_{\sigma\kappa} = 1$

$$\begin{aligned} &\text{a)} \quad (0 \le \mu_\sigma \le P_{x_\sigma} - 1) \wedge (0 > \mu_\sigma + 1 \vee \mu_\sigma + 1 > P_{x_\sigma} - 1) \equiv \mu_\sigma = P_{x_\sigma} - 1 \\ &\text{b)} \quad (0 \le \mu_\sigma \le P_{x_\sigma} - 1) \wedge (0 \le \mu_\sigma + 1 \le P_{x_\sigma} - 1) \equiv 0 \le \mu_\sigma \le P_{x_\sigma} - 2 \end{aligned} \quad (6.12)$$

2. $h_{\sigma\kappa} = -1$

$$\begin{aligned} &\text{a)} \quad (0 \le \mu_\sigma \le P_{x_\sigma} - 1) \wedge (0 > \mu_\sigma - 1 \vee \mu_\sigma - 1 > P_{x_\sigma} - 1) \equiv \mu_\sigma = 0 \\ &\text{b)} \quad (0 \le \mu_\sigma \le P_{x_\sigma} - 1) \wedge (0 \le \mu_\sigma - 1 \le P_{x_\sigma} - 1) \equiv 1 \le \mu_\sigma \le P_{x_\sigma} - 1 \end{aligned} \quad (6.13)$$

3. $h_{\sigma\kappa} = 0$
b) $(0 \le \mu_\sigma \le P_{x_\sigma} - 1) \wedge (0 \le \mu_\sigma \le P_{x_\sigma} - 1) \equiv (0 \le \mu_\sigma \le P_{x_\sigma} - 1)$
a) Dieser Fall ist wesentlich komplizierter. Wir müssen hier die Fälle unterscheiden, in denen die anderen beiden der ersten 3 Spaltenelemente gleich null sind und die, in denen sie es nicht sind.

κ	$h_{\sigma\kappa}$	$a) \quad (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}_0)$	$b) \quad (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G})$
1	$h_{11} = 1$	$\mu_1 = P_{x_1} - 1$	$0 \le \mu_1 \le P_{x_1} - 2$
	$h_{21} = 0$	$0 \le \mu_2 \le P_{x_2} - 1$	$0 \le \mu_2 \le P_{x_2} - 1$
	$h_{31} = 0$	$0 \le \mu_3 \le P_{x_3} - 1$	$0 \le \mu_3 \le P_{x_3} - 1$
2	$h_{12} = 0$	$0 \le \mu_1 \le P_{x_1} - 1$	$0 \le \mu_1 \le P_{x_1} - 1$
	$h_{22} = 1$	$\mu_2 = P_{x_2} - 1$	$0 \le \mu_2 \le P_{x_2} - 2$
	$h_{32} = 0$	$0 \le \mu_3 \le P_{x_3} - 1$	$0 \le \mu_3 \le P_{x_3} - 1$
3	$h_{13} = 0$	$0 \le \mu_1 \le P_{x_1} - 1$	$0 \le \mu_1 \le P_{x_1} - 1$
	$h_{23} = 0$	$0 \le \mu_2 \le P_{x_2} - 1$	$0 \le \mu_2 \le P_{x_2} - 1$
	$h_{33} = 1$	$\mu_3 = P_{x_3} - 1$	$0 \le \mu_3 \le P_{x_3} - 2$
4	$h_{14} = -1$	$\mu_1 = 0$	$1 \le \mu_1 \le P_{x_1} - 1$
	$h_{24} = 0$	$0 \le \mu_2 \le P_{x_2} - 1$	$0 \le \mu_2 \le P_{x_2} - 1$
	$h_{34} = 0$	$0 \le \mu_3 \le P_{x_3} - 1$	$0 \le \mu_3 \le P_{x_3} - 1$
5	$h_{15} = 0$	$0 \le \mu_1 \le P_{x_1} - 1$	$0 \le \mu_1 \le P_{x_1} - 1$
	$h_{25} = -1$	$\mu_2 = 0$	$1 \le \mu_2 \le P_{x_2} - 1$
	$h_{35} = 0$	$0 \le \mu_3 \le P_{x_3} - 1$	$0 \le \mu_3 \le P_{x_3} - 1$
6	$h_{16} = 0$	$0 \le \mu_1 \le P_{x_1} - 1$	$0 \le \mu_1 \le P_{x_1} - 1$
	$h_{26} = 0$	$0 \le \mu_2 \le P_{x_2} - 1$	$0 \le \mu_2 \le P_{x_2} - 1$
	$h_{36} = -1$	$\mu_3 = 0$	$1 \le \mu_3 \le P_{x_3} - 1$
7	$h_{17} = 0$	$FALSE$	$0 \le \mu_1 \le P_{x_1} - 1$
	$h_{27} = 0$		$0 \le \mu_2 \le P_{x_2} - 1$
	$h_{37} = 0$		$0 \le \mu_3 \le P_{x_3} - 1$
			$0 \le \delta \le P_{x_1} P_{x_2} P_{x_3} - 1$

Tabelle 6.1: Grenzen

1) Wir wollen dies am Beispiel $\kappa = 1$ diskutieren. Hier gilt $h_{11} = 1$ und für die anderen beiden Spaltenelemente $h_{12} = h_{13} = 0$. Für das Prädikat $\boldsymbol{\mu} + \boldsymbol{h}_1 \in \mathcal{G}_0$ ergibt sich als konjunktive Bedingung

$$(0 \le \mu_2 \le P_{x_2} - 1) \wedge (0 \le \mu_3 \le P_{x_3} - 1) \; . \tag{6.14}$$

Das Prädikat $\boldsymbol{\mu} + \boldsymbol{h}_1 \in \mathcal{G}_0$ ist mit $\boldsymbol{\mu} \in \mathcal{G}$ konjunktiv zu verknüpfen und liefert für die Spalten 2 und 3

$$(0 \le \mu_2 \le P_{x_2} - 1) \wedge (0 \le \mu_3 \le P_{x_3} - 1) \; . \tag{6.15}$$

Hinzu kommt die konjunktive Verknüpfung mit den in Gleichung (6.12) ermittelten Ergebnissen für die erste Spalte, d.h. $\mu_1 = P_{x_1} - 1$.
2) Nun betrachten wir das Beispiel $\kappa = 7$. Hier gilt neben $h_{17} = 0$ auch für die anderen beiden Spaltenelemente $h_{72} = h_{73} = 0$.

Da $\mathcal{G}$ und $\mathcal{G}_0$ für ein konstantes μ_4 disjunkt sind, gilt

$$(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_7 \in \mathcal{G}_0) \equiv FALSE \tag{6.16}$$

Die Bedeutung dieses Resultates ist, dass für den in 1) genannten Bereich die Aussage wahr ist und für den in 2) genannten Bereich die Aussage falsch ist.

Für jedes κ und der Kenntnis von $\boldsymbol{H}$ muss jeweils für ein Spaltenelement $0, -1, 1$ unterschieden werden. Das ergibt die in der Tabelle 6.1 dargestellten Grenzen.
Bemerkung : Sowohl an $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}_0) \vee (\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}) \equiv \boldsymbol{\mu} \in \mathcal{G}$ (die disjunktive Verknüpfung wirkt wie eine Vereinigung der Mengen), als auch an den Werten in der Tabelle ist zu

erkennen, dass sämtliche Punkte, für die $\boldsymbol{\mu} \in \mathcal{G}$ gilt, abgearbeitet werden. Mit der Tabelle sind wir später in der Lage, den Schleifenanfang und die Schleifenabbruchbedingung in den Variablen μ_1, μ_1, μ_3 festzulegen.

Wir wollen nun die Auswirkungen einer Verschiebung des Vektors $\boldsymbol{\mu}$ nach $\boldsymbol{\mu} + \boldsymbol{h}_\kappa$ auf δ untersuchen. Konkret möchten wir die wie folgt definierten δ_κ bestimmen

$$\delta(\boldsymbol{\mu}) + \delta_\kappa = \delta(\boldsymbol{\mu} + \boldsymbol{h}_\kappa) \; . \tag{6.17}$$

Nach Gleichung (2.230) gilt

$$\delta(\boldsymbol{\mu}) = \mu_1 + \mu_2 P_{x_1} + \mu_3 P_{x_1} P_{x_2} = \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{c} \quad , \quad \boldsymbol{c}^{\mathrm{T}} = [1 \, , \, P_{x_1} \, , \, P_{x_1} P_{x_2}] \tag{6.18}$$

und es folgt

$$\delta_\kappa = \delta(\boldsymbol{\mu} + \boldsymbol{h}_\kappa) - \delta(\boldsymbol{\mu}) = \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{c} + \boldsymbol{h}_\kappa^{\mathrm{T}} \boldsymbol{c} - \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{c} = \boldsymbol{h}_\kappa^{\mathrm{T}} \boldsymbol{c} \, . \tag{6.19}$$

Im Einzelnen ergeben sich die folgenden Werte

$$\delta_1 = -\delta_4 = 1 \; , \; \delta_2 = -\delta_5 = P_{x_1} \; , \; \delta_3 = -\delta_6 = P_{x_1} P_{x_2} \; , \; \delta_7 = 0 \, . \tag{6.20}$$

Wir führen nun eine Reihe von Prädikaten ein, deren Gegenstände der Ortspunkt des Sollwertes, ausgedrückt durch μ_1, μ_2, μ_3 und der Ortspunkt des Ist-Wertes δ ist. Die Wahrheit eines Prädikates entspricht der Tatsache, dass der Wert der Programmvariablen (der Ist-Wert) identisch mit dem geforderten Wert (dem Sollwert) ist. Wir benötigen jeweils ein Prädikat für die ausfallenden Wellen der nichtdynamischen Bauelemente

$$\mathsf{P}_{e\,k}\,(\mu_1, \mu_2, \mu_3, \delta) \equiv \; b_{e\,k}(\boldsymbol{\mu}) = b_{e\,k}(\delta).v \; , \tag{6.21}$$

die Ausgangswellen der Quellen

$$\mathsf{P}_{q\,k}\,(\mu_1, \mu_2, \mu_3, \delta) \equiv \; b_{q\,k}(\boldsymbol{\mu}) = b_{q\,k}(\delta).v \; , \tag{6.22}$$

die Ausgangssignale des Funktionsbausteins

$$\mathsf{P}_{a\mathrm{FB}\,k}\,(\boldsymbol{\mu}, \delta) \equiv \; a_{\mathrm{FB}\,k}(\boldsymbol{\mu}) = a_{\mathrm{FB}\,k}(\delta).v \; , \tag{6.23}$$

die einfallenden Wellen der Verzögerer am Rand

$$\mathsf{P}^{\kappa}_{a\,k}\,(\mu_1, \mu_2, \mu_3, \delta) \equiv \; a^{\kappa}_{v\,k}(\boldsymbol{\mu}) = a^{\kappa}_{v\,k}(\delta).v \tag{6.24}$$

und die ausfallenden Wellen der Verzögerer

$$\mathsf{P}^{\kappa}_{b\,k}\,(\mu_1, \mu_2, \mu_3, \delta) \equiv \; b^{\kappa}_{v\,k}(\boldsymbol{\mu}) = b^{\kappa}_{v\,k}(\delta).v \; . \tag{6.25}$$

Diese Prädikate nutzen wir, um weitere Prädikate zu definieren. Wir verknüpfen konjunktiv die soeben definierten Prädikate, ausgewertet an Ortspunkten eines Gebietes, um die Gleichheit von Soll-und Istwert in einem Gebiet auszudrücken. Zunächst definieren wir Prädikate für die ausfallenden Wellen der

Verzögerer innerhalb des Gebietes $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G})$

$$
\begin{aligned}
\mathsf{P}^1_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-2} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1+1, \mu_2, \mu_3, \delta+\delta_1) \\
\mathsf{P}^2_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-2} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^2} \mathsf{P}^2_{bk}(\mu_1, \mu_2+1, \mu_3, \delta+\delta_2) \\
\mathsf{P}^3_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-2} \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{bk}(\mu_1, \mu_2, \mu_3+1, \delta+\delta_3) \\
\mathsf{P}^4_{\text{norm}} &\equiv \bigwedge_{\mu_1=1}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^4} \mathsf{P}^4_{bk}(\mu_1-1, \mu_2, \mu_3, \delta+\delta_4) \\
\mathsf{P}^5_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=1}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^5} \mathsf{P}^5_{bk}(\mu_1, \mu_2-1, \mu_3, \delta+\delta_5) \\
\mathsf{P}^6_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=1}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^6} \mathsf{P}^6_{bk}(\mu_1, \mu_2, \mu_3-1, \delta+\delta_6) \\
\mathsf{P}^7_{\text{norm}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^7} \mathsf{P}^7_{bk}(\mu_1, \mu_2, \mu_3, \delta+\delta_7)\ ,
\end{aligned}
\tag{6.26}
$$

oder kompakt

$$
\begin{aligned}
\mathsf{P}_{\text{norm}} \equiv & \bigwedge_{\kappa=1}^{3} \bigwedge_{-4\boldsymbol{\mu}=\boldsymbol{0}}^{\boldsymbol{P}\boldsymbol{x}-\mathbf{e}-{}_{-4}\boldsymbol{h}_\kappa} \bigwedge_{k=1}^{n_v^\kappa} \mathsf{P}^\kappa_{bk}(\mu_1+h_{1\kappa}, \mu_2+h_{2\kappa}, \mu_3+h_{3\kappa}, \delta+\delta_\kappa) \wedge \\
& \bigwedge_{\kappa=4}^{7} \bigwedge_{-4\boldsymbol{\mu}={}_{-4}\boldsymbol{h}_\kappa}^{\boldsymbol{P}\boldsymbol{x}-\mathbf{e}} \bigwedge_{k=1}^{n_v^\kappa} \mathsf{P}^\kappa_{bk}(\mu_1+h_{1\kappa}, \mu_2+h_{2\kappa}, \mu_3+h_{3\kappa}, \delta+\delta_\kappa)\ .
\end{aligned}
\tag{6.27}
$$

Weiterhin definieren wir Prädikate für die ausfallenden Wellen der Verzögerer des Gebietes $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}_0)$

$$
\begin{aligned}
\mathsf{P}_{\text{x}}(\mu_2, \mu_3) &\equiv \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{b\,k}\left(0, \mu_2, \mu_3, \delta([0, \mu_2, \mu_3]^{\text{T}})\right) \wedge \bigwedge_{k=1}^{n_v^4} \mathsf{P}^4_{b\,k}\left(P_{x_1}-1, \mu_2, \mu_3, \delta([P_{x_1}-1, \mu_2, \mu_3]^{\text{T}})\right) \\
\mathsf{P}_{\text{y}}(\mu_1, \mu_3) &\equiv \bigwedge_{k=1}^{n_v^2} \mathsf{P}^2_{b\,k}\left(\mu_1, 0, \mu_3, \delta([\mu_1, 0, \mu_3]^{\text{T}})\right) \wedge \bigwedge_{k=1}^{n_v^5} \mathsf{P}^5_{b\,k}\left(\mu_1, P_{x_2}-1, \mu_3, \delta([\mu_1, P_{x_2}-1, \mu_3]^{\text{T}})\right) \\
\mathsf{P}_{\text{z}}(\mu_1, \mu_2) &\equiv \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{b\,k}\left(\mu_1, \mu_2, 0, \delta([\mu_1, \mu_2, 0]^{\text{T}})\right) \wedge \bigwedge_{k=1}^{n_v^6} \mathsf{P}^6_{b\,k}\left(\mu_1, \mu_2, P_{x_3}-1, \delta([\mu_1, \mu_2, P_{x_3}-1]^{\text{T}})\right)\ .
\end{aligned}
\tag{6.28}
$$

Der nächste große Block von Prädikatsdefinitionen gilt den einfallenden Wellen der Verzögerer des Ge-

bietes $(\boldsymbol{\mu} \in \mathcal{G}) \wedge (\boldsymbol{\mu} + \boldsymbol{h}_\kappa \in \mathcal{G}_0)$

$$\begin{aligned}
\mathsf{P}_{\mathrm{ax}}(\mu_2,\mu_3) &\equiv \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{a\,k}(P_{x_1}-1,\mu_2,\mu_3,\delta([P_{x_1}-1,\mu_2,\mu_3]^{\mathrm{T}})) \wedge \bigwedge_{k=1}^{n_v^4} \mathsf{P}^4_{a\,k}(0,\mu_2,\mu_3,\delta([0,\mu_2,\mu_3]^{\mathrm{T}})) \\
\mathsf{P}_{\mathrm{ay}}(\mu_1,\mu_3) &\equiv \bigwedge_{k=1}^{n_v^2} \mathsf{P}^2_{a\,k}(\mu_1,P_{x_2}-1,\mu_3,\delta([\mu_1,P_{x_2}-1,\mu_3]^{\mathrm{T}})) \wedge \bigwedge_{k=1}^{n_v^5} \mathsf{P}^5_{a\,k}(\mu_1,0,\mu_3,\delta([\mu_1,0,\mu_3]^{\mathrm{T}})) \\
\mathsf{P}_{\mathrm{az}}(\mu_1,\mu_2) &\equiv \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{a\,k}(\mu_1,\mu_2,P_{x_3}-1,\delta([\mu_1,\mu_2,P_{x_3}-1]^{\mathrm{T}})) \wedge \bigwedge_{k=1}^{n_v^6} \mathsf{P}^6_{a\,k}(\mu_1,\mu_2,0,\delta([\mu_1,\mu_2,0]^{\mathrm{T}})) \,.
\end{aligned} \tag{6.29}$$

Über die ganzen Randflächen ergibt sich

$$\begin{aligned}
\mathsf{P}'_{\mathrm{avx}} &\equiv \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \mathsf{P}_{\mathrm{ax}}(\mu_2,\mu_3) \\
\mathsf{P}'_{\mathrm{avy}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \mathsf{P}_{\mathrm{ay}}(\mu_1,\mu_3) \\
\mathsf{P}'_{\mathrm{avz}} &\equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1,\mu_2) \,.
\end{aligned} \tag{6.30}$$

Außerdem benötigen wir Prädikate, die genau dann wahr sind, wenn die Werte der Variablen der Quellenwellen

$$\mathsf{P}_{\mathrm{q}} \equiv \bigwedge_{\delta=0}^{P-1} \bigwedge_{k=1}^{n_q} \mathsf{P}_{qk}\left(\mu_1(\delta),\mu_2(\delta),\mu_3(\delta),\delta\right) \tag{6.31}$$

und die Werte der Verzögerer

$$\mathsf{P}_{\mathrm{v}} \equiv \bigwedge_{\delta=0}^{P-1} \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} \mathsf{P}^\kappa_{bk}\left(\mu_1(\delta),\mu_2(\delta),\mu_3(\delta),\delta\right) \tag{6.32}$$

den tatsächlichen Werten im gesamten Berechnungsgebiet entsprechen. Im Folgenden werden wir eine Reihe von häufig verwendeten logischen Prädikaten definieren, die uns später bei dem eigentlichen Korrektheitsbeweis eine einfachere Schreibweise ermöglichen. Hierbei dient

$$\mathsf{P}_{\mathrm{dec}} \equiv \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge \mathsf{P}_{\mathrm{decFB}} \wedge \mathsf{P}_{\mathrm{decuv}} \wedge \mathsf{P}_{\mathrm{decbe}} \wedge \mathsf{P}_{\mathrm{decav}} \tag{6.33}$$

mit

$$
\begin{aligned}
\mathsf{P}_{\text{decbq}} &\equiv \bigwedge_{n=1}^{n_q} \left[\, b_{q\,n}.typ = \texttt{float} \wedge P \le b_{q\,n}.l \le \text{Max}_\text{i} \,\right] \\
\mathsf{P}_{\text{decbv}} &\equiv \bigwedge_{\kappa=1}^{7} \bigwedge_{n=1}^{n_v^\kappa} \left[\, b_{v\,n}^\kappa.typ = \texttt{float} \wedge P \le b_{v\,n}^\kappa.l \le \text{Max}_\text{i} \,\right] \\
\mathsf{P}_{\text{decbe}} &\equiv \bigwedge_{n=1}^{n_e} \left[\, b_{e\,n}.typ = \texttt{float} \wedge P \le b_{e\,n}.l \le \text{Max}_\text{i} \,\right] \\
\mathsf{P}_{\text{decaFB}} &\equiv \bigwedge_{n=1}^{n_{aa}} \left[\, a_{\text{FB}\,n}.typ = \texttt{float} \wedge P \le a_{\text{FB}\,n}.l \le \text{Max}_\text{i} \,\right] \\
\mathsf{P}_{\text{decav}} &\equiv \bigwedge_{\kappa=1}^{6} \bigwedge_{n=1}^{n_v^\kappa} \left[\, a_{v\,n}^\kappa.typ = \texttt{float} \wedge P \le a_{v\,n}^\kappa.l \le \text{Max}_\text{i} \,\right] \\
\mathsf{P}_{\text{decuv}} &\equiv \delta.typ = \texttt{int} \wedge \delta.l = -1 \wedge \bigwedge_{\kappa=1}^{3} \left[\mu_\kappa.typ = \texttt{int} \wedge \mu_\kappa.l = -1\right]
\end{aligned}
\tag{6.34}
$$

für die Deklariertheit von vektoriellen Variablen der richtigen Länge. Dabei unterscheiden wir zwischen denen, die vor der Ausführung von S_{alg} schon deklariert sind und denen, die noch deklariert werden müssen. Das wichtigste Prädikat ist das korrekte WDF, welches aus Gleichung (6.1) bis Gleichung (6.5) entsteht

$$
\mathsf{P}_{\text{WDF}} \equiv \bigwedge_{\boldsymbol{\mu}=\mathbf{0}}^{\boldsymbol{P}\boldsymbol{x}-\mathbf{e}} \boldsymbol{b}_e(\boldsymbol{\mu}) = \boldsymbol{L}_q\boldsymbol{b}_q(\boldsymbol{\mu}) + \boldsymbol{L}_v\boldsymbol{b}_v(\boldsymbol{\mu}) + \boldsymbol{L}_e\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{a}_{\text{FB}}(\boldsymbol{\mu}) = \boldsymbol{A}\boldsymbol{b}(\boldsymbol{\mu})
$$

$$
\bigwedge_{\kappa=1}^{3} \bigwedge_{{}_{-4}\boldsymbol{\mu}=\mathbf{0}}^{\boldsymbol{P}\boldsymbol{x}-\mathbf{e}-{}_{-4}\boldsymbol{h}_\kappa} \bigwedge_{k=1}^{n_v^\kappa} \boldsymbol{b}_v^\kappa(\boldsymbol{\mu}+\boldsymbol{h}_\kappa) = \boldsymbol{P}_{v^\kappa e}\boldsymbol{b}_e(\boldsymbol{\mu}) \wedge \bigwedge_{\kappa=4}^{7} \bigwedge_{{}_{-4}\boldsymbol{\mu}\,=\,-{}_{-4}\boldsymbol{h}_\kappa}^{\boldsymbol{P}\boldsymbol{x}\,-\,\mathbf{e}} \bigwedge_{k=1}^{n_v^\kappa} \boldsymbol{b}_v^\kappa(\boldsymbol{\mu}+\boldsymbol{h}_\kappa) = \boldsymbol{P}_{v^\kappa e}\boldsymbol{b}_e(\boldsymbol{\mu})
$$

$$
\bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \boldsymbol{a}_v^1(\boldsymbol{\mu}) = \boldsymbol{P}_{v^1 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^1(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^1\boldsymbol{a}_v^4(\boldsymbol{\mu}) \wedge \mu_1 = P_{x_1} - 1
$$

$$
\bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \boldsymbol{a}_v^2(\boldsymbol{\mu}) = \boldsymbol{P}_{v^2 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^2(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^2\boldsymbol{a}_v^5(\boldsymbol{\mu}) \wedge \mu_2 = P_{x_2} - 1
$$

$$
\bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \boldsymbol{a}_v^3(\boldsymbol{\mu}) = \boldsymbol{P}_{v^3 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^3(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^3\boldsymbol{a}_v^6(\boldsymbol{\mu}) \wedge \mu_3 = P_{x_3} - 1
$$

$$
\bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \boldsymbol{a}_v^6(\boldsymbol{\mu}) = \boldsymbol{P}_{v^6 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^6(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^6\boldsymbol{a}_v^3(\boldsymbol{\mu}) \wedge \mu_3 = 0
$$

$$\bigwedge_{\mu=1}^{n_e}\bigwedge_{\nu=\mu}^{n_e} l_{e\mu\nu}=0 \wedge \|\boldsymbol{x}\| \geq \|\boldsymbol{L}\boldsymbol{x}\| \;\forall\, \boldsymbol{x} \;\wedge \bigwedge_{\kappa=1}^{6} \|\boldsymbol{x}\| \geq \|\boldsymbol{R}^{\kappa}\boldsymbol{x}\| \;\forall\, \boldsymbol{x} \wedge$$

$$\bigwedge_{\mu_2=0}^{P_{x_2}-1}\bigwedge_{\mu_3=0}^{P_{x_3}-1} \boldsymbol{a}_v^4(\boldsymbol{\mu}) = \boldsymbol{P}_{v^4 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^4(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^4\boldsymbol{a}_v^1(\boldsymbol{\mu}) \wedge \mu_1 = 0$$

$$\bigwedge_{\mu_1=0}^{P_{x_1}-1}\bigwedge_{\mu_3=0}^{P_{x_3}-1} \boldsymbol{a}_v^5(\boldsymbol{\mu}) = \boldsymbol{P}_{v^5 e}\boldsymbol{b}_e(\boldsymbol{\mu}) \;\wedge\; \boldsymbol{b}_v^5(\boldsymbol{\mu}+\mathbf{e}_7) = \boldsymbol{R}^5\boldsymbol{a}_v^2(\boldsymbol{\mu}) \wedge \mu_2 = 0$$

$$\bigwedge_{\kappa=1}^{3} [\mathrm{Max_i} \geq P_{x_\kappa} > 0 \;\wedge\; P_{x_\kappa} \in \mathbb{N}] \wedge P = P_{x_1}P_{x_2}P_{x_3} \wedge \mathrm{Max_i} \geq P\;. \tag{6.35}$$

Weiterhin definieren wir ein Prädikat, welches im Zusammenhang mit der Überschreitung des Wertebereichs der abhängigen Variablen verwendet wird

$$\mathrm{P_{max}} \equiv \bigwedge_{\delta=0}^{P-1} \sum_{\nu=1}^{n_q} |b_{q\,\nu}(\delta).v|^2 + \sum_{\nu=1}^{n_v} |b_{v\,\nu}(\delta).v|^2 < {\mathrm{Max_f}}^2/(\beta^2 4)\;, \tag{6.36}$$

wobei die Konstante $\beta > 0$ sich aus

$$\beta^2 = \max_{\mu}\{\max\{R_{N\mu}, G_{N\mu}\}\} \tag{6.37}$$

ergibt und $R_{N\mu}, G_{N\mu}$ die (normierten) Torwiderstände bzw. Torleitwerte sind. Unter Berücksichtigung von Gleichung (A.4) gilt $\beta \geq 1$. Somit kann das Radizieren in den Operator hineingezogen werden, so dass

$$\beta = \max_{\mu}\left\{\max\left\{\sqrt{R_{N\mu}}, \sqrt{G_{N\mu}}\right\}\right\} \tag{6.38}$$

entsteht. Zudem benötigen wir

$$\mathsf{P}_{\mathrm{avx\,max}} \equiv \bigwedge_{\mu_2=0}^{P_{x_2}-1}\bigwedge_{\mu_3=0}^{P_{x_3}-1} \left[\bigwedge_{\nu=1}^{n_v^1} |a_{v\,\nu}^1(\delta([0,\mu_2,\mu_3]^{\mathrm{T}})| < \mathrm{Max_f} \;\wedge\; \bigwedge_{\nu=1}^{n_v^4} |a_{v\,\nu}^4(\delta([P_{x_1}-1,\mu_2,\mu_3]^{\mathrm{T}})| < \mathrm{Max_f}\right],$$

$$\mathsf{P}_{\mathrm{avy\,max}} \equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1}\bigwedge_{\mu_3=0}^{P_{x_3}-1} \left[\bigwedge_{\nu=1}^{n_v^2} |a_{v\,\nu}^2(\delta([\mu_1,0,\mu_3]^{\mathrm{T}})| < \mathrm{Max_f} \;\wedge\; \bigwedge_{\nu=1}^{n_v^5} |a_{v\,\nu}^5(\delta([\mu_1,P_{x_2}-1,\mu_3]^{\mathrm{T}})| < \mathrm{Max_f}\right],$$

$$\mathsf{P}_{\mathrm{avz\,max}} \equiv \bigwedge_{\mu_1=0}^{P_{x_1}-1}\bigwedge_{\mu_2=0}^{P_{x_2}-1} \left[\bigwedge_{\nu=1}^{n_v^3} |a_{v\,\nu}^3(\delta([\mu_1,\mu_2,0]^{\mathrm{T}})| < \mathrm{Max_f} \;\wedge\; \bigwedge_{\nu=1}^{n_v^6} |a_{v\,\nu}^6(\delta([\mu_1,\mu_2,P_{x_3}-1]^{\mathrm{T}})| < \mathrm{Max_f}\right],$$

$$\mathsf{P}_{\mathrm{av\,max}} \equiv \mathsf{P}_{\mathrm{avx\,max}} \wedge \mathsf{P}_{\mathrm{avy\,max}} \wedge \mathsf{P}_{\mathrm{avz\,max}} \;\text{ und}$$

$$\mathsf{P}_{\mathrm{be\,max}} \equiv \bigwedge_{\delta=0}^{P-1}\bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta).v|2\beta < \mathrm{Max_f}\;.$$

(6.39)

Zum Abschluss benötigen wir für die korrekte Abbildung des Vektors $\boldsymbol{\mu}$ ins Eindimensionale das Prädikat

$$\mathsf{P}_{\mathrm{MDED}} \equiv \mu_1 + \mu_2 P_{x_1} + \mu_3 P_{x_1} P_{x_2} = \delta(\boldsymbol{\mu}) \; . \tag{6.40}$$

Im engen Zusammenhang damit steht der folgende

Satz 2 *Sei* $\{\mathsf{V}\}$ `delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;` $\{\mathsf{P}\}$.
Wenn dazu die Nachbedingung $\mathsf{P} \equiv \mathsf{P}' \wedge 0 \leq \delta.v \leq P - 1 \wedge \mathsf{P}_{\mathrm{dec}}$ *ist, dann ist*

$$0 \leq \mu_1.v \leq P_{x_1} - 1 \wedge 0 \leq \mu_2.v \leq P_{x_2} - 1 \wedge 0 \leq \mu_3.v \leq P_{x_3} - 1 \wedge \mathsf{P}' \begin{bmatrix} \delta.v \\ \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \end{bmatrix} \wedge \mathsf{P}_{\mathrm{dec}} \tag{6.41}$$

eine gültige Vorbedingung $\{\mathsf{V}\}$.

Beweis 2

Durch Verwendung von Gleichung (2.233) i.V.m. Gleichung (2.230) erhalten wir

$$\begin{aligned} &0 \leq \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \leq P - 1 \wedge \mathsf{P}' \begin{bmatrix} \delta.v \\ \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \end{bmatrix} \wedge \\ &\mathsf{P}_{\mathrm{dec}} \wedge \delta.l = -1 \wedge \delta.typ = \texttt{int} \wedge |\mu_1.v| \leq \mathrm{Max_i} \wedge |\mu_1.v + P_{x_1} \mu_2.v| \leq \mathrm{Max_i} \wedge \\ &|\mu_1.v + P_{x_1} \mu_2.v + P_{x_1} P_{x_2} \mu_3.v| \leq \mathrm{Max_i} \wedge \bigwedge_{l=1}^{3} [\mu_l.l = -1 \wedge \mu_l.typ = \texttt{int}] \; . \end{aligned} \tag{6.42}$$

Zunächst stellen wir fest, dass die auftretenden Typ- und Längenbedingungen abgetrennt werden können, da sie von $\mathsf{P}_{\mathrm{dec}}$ impliziert werden.

Hinreichend für $0 \leq \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \leq P - 1$ mit $P = P_{x_1} P_{x_2} P_{x_3}$ ist, wie im Kapitel 2.11 gezeigt

$$0 \leq \mu_1.v \leq P_{x_1} - 1 \wedge 0 \leq \mu_2.v \leq P_{x_2} - 1 \wedge 0 \leq \mu_3.v \leq P_{x_3} - 1 \tag{6.43}$$

Zudem erübrigt sich die weitere Betrachtung der Teilsummen, da die Summanden dieses Skalarproduktes nicht negativ sind. $\mathsf{P}_{\mathrm{dec}}$ impliziert die Aussage $P - 1 \leq \mathrm{Max_i}$, so dass mit $0 \leq \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \leq P - 1$ die Aussage $\mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} \leq \mathrm{Max_i}$ auch aus Gleichung (6.42) abgetrennt werden kann.

Wir kommen nun zur eigentlichen Problemstellung zurück, der Formulierung eines Algorithmus zur Berechnung eines mehrdimensionalen Wellendigitalfilters, welcher sich in das Programmpaket SPACE einfügt. Hierbei berücksichtigen wir, dass der Aufruf der Funktionsbausteine durch die Ablaufumgebung erfolgt. Bild 6.1 zeigt den Teil der Ablaufumgebung, der die Funktionsbausteine aufruft. Die äußere nichtterminierte Schleife stellt die Ablaufumgebung dar, die über die Funktionsplangruppen und den Funktionsplänen die einzelnen Funktionsbausteine aufruft. Die diskrete Variable μ_k des Wellendigitalfilters entspricht der diskreten Variablen der Ablaufumgebung, d.h. das Fortschreiten von einer Abtastschicht zur nächsten entspricht dem Aufruf des MDWDF-Funktionsbausteins, wie im Kapitel 3 gefordert. Der Funktionsplan umfasst den MDWDF-Funktionsbaustein und weitere Funktionsbausteine. Der Teil des Programmablaufplans innerhalb der gestrichelten Linie ist eine detailliertere Darstellung des mehrdimensionalen Wellendigitalfilters. Hinter dem Anweisungsblock zur Ermittlung der Verzögererwerte der neuen Abtastschicht verbergen sich die im Bild 6.2 dargestellten Anweisungsblöcke, die wiederum später im Detail exemplarisch im Bild 6.5 dargestellt werden. Beim Aufruf des MDWDF-Funktionsbausteins

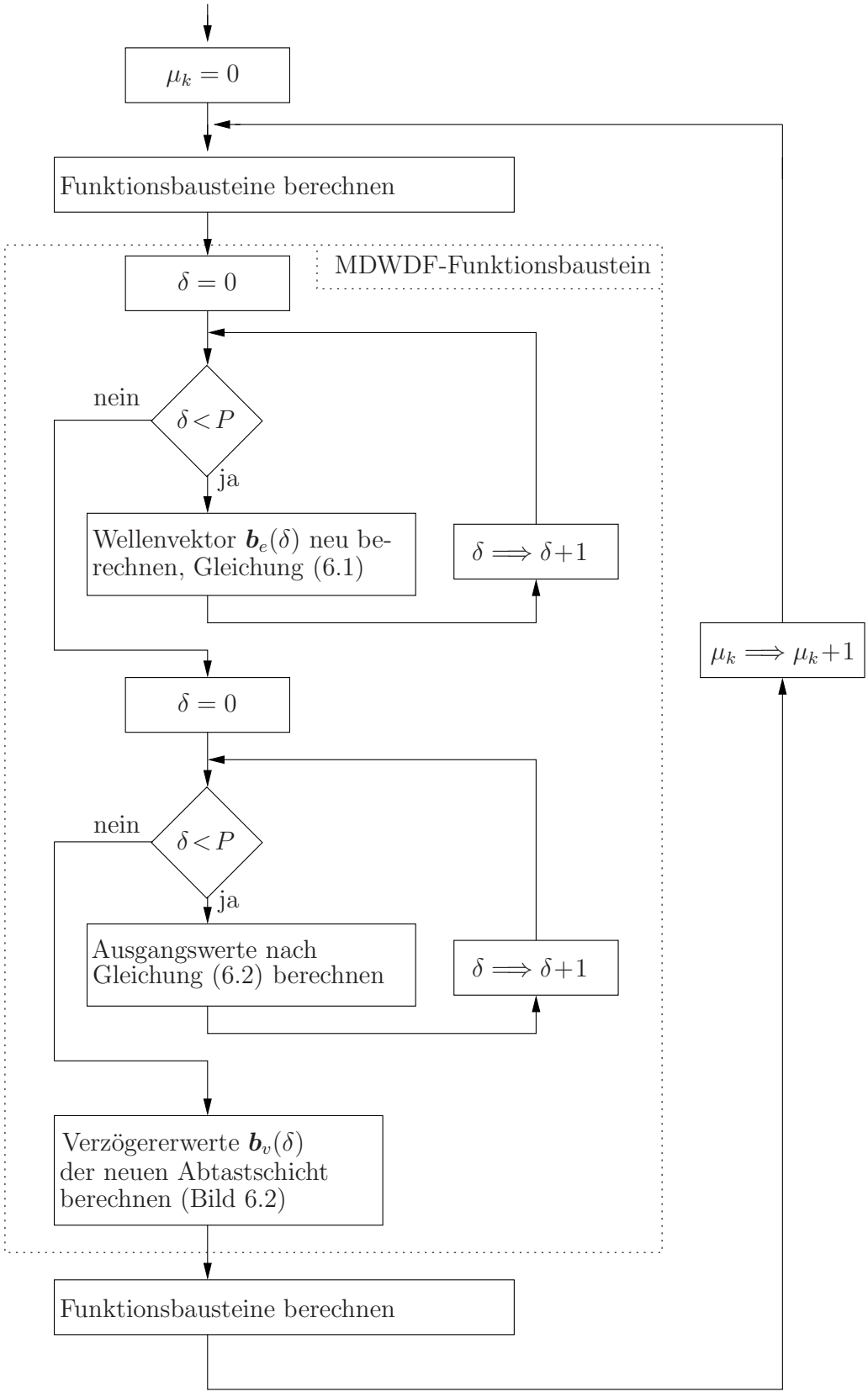


Bild 6.1: Einbindung eines MDWDF-Funktionsbausteins in eine Ablaufumgebung

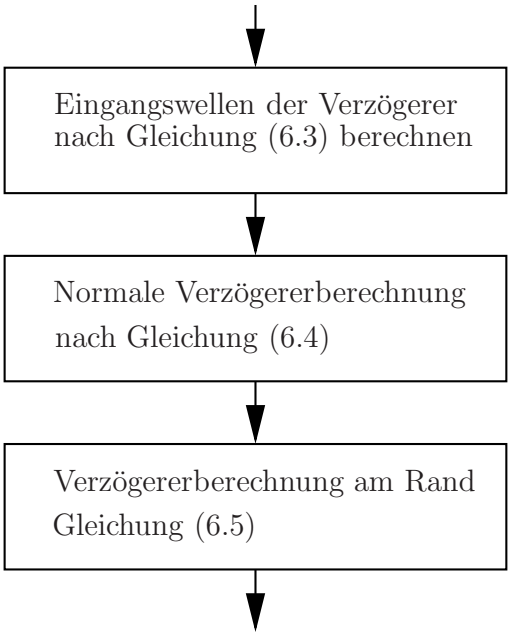


Bild 6.2: Verzögererwerte $\boldsymbol{b}_v(\delta)$ der neuen Abtastschicht ermitteln

werden zunächst die Wellengrößen der nichtdynamischen Elemente nach Gleichung (6.1) berechnet. Im Anschluss können die Ausgangswerte des Funktionsbausteins nach Gleichung (6.2) ermittelt werden. Darauf folgend berechnen wir die Eingangswellen der Verzögerer am Rand gemäß Gleichung (6.3).

Wir wollen an dieser Stelle noch einen wichtigen Aspekt erläutern. Bis einschließlich Gleichung (6.3) wurde nur lesend auf die Ausgangswellen der Verzögerer $\boldsymbol{m}_{\mathrm{FB}} = \boldsymbol{b}_v([\bullet, \bullet, \bullet, \mu_k.v]^{\mathrm{T}})$ zugegriffen. Nach Gleichung (6.3) wird nicht mehr lesend darauf zugegriffen, d. h., ab hier werden die Verzögererwerte der aktuellen Abtastschicht nicht mehr benötigt. Bei der Berechnung der Verzögererwerte der nächsten Abtastschicht starten wir mit den Verzögererwellen, die keine Randbehandlung benötigen, d. h. also mit Gleichung (6.4). Hierbei betrachten wir nun $\boldsymbol{m}_{\mathrm{FB}}$ als die Verzögererwerte der nächsten Abtastschicht, d. h. es gilt $\boldsymbol{m}_{\mathrm{FB}} = \boldsymbol{b}_v([\bullet, \bullet, \bullet, \mu_k.v + 1]^{\mathrm{T}})$ Der Zugriff auf $\boldsymbol{m}_{\mathrm{FB}}$ erfolgt nun schreibend. Dass wir die Verzögererwerte der nächsten Abtastschicht berechnen, erkennen wir daran, dass $\mu_k.v$ immer noch den gleichen Wert hat und im Argument von $\boldsymbol{b}_v$ der Wert $\mu_k.v + 1$ steht. Insbesondere ist zu beachten, dass erst nach Inkrementierung von $\mu_k.v$ die Beziehung $\boldsymbol{m}_{\mathrm{FB}} = \boldsymbol{b}_v([\bullet, \bullet, \bullet, \mu_k.v]^{\mathrm{T}})$ gilt und erst dann $\boldsymbol{m}_{\mathrm{FB}}$ wieder benutzt werden darf.

Zum Abschluss berechnet man sequentiell die Verzögererwellen des Randes nach Gleichung (6.5). Außerhalb des Funktionsbausteins wird $\mu_k.v$ inkrementiert. Für den lesenden Zugriff auf $\boldsymbol{b}_v$ gilt nun $\boldsymbol{m}_{\mathrm{FB}} = \boldsymbol{b}_v([\bullet, \bullet, \bullet, \mu_k.v]^{\mathrm{T}})$.

Das erläuterte Vorgehen entspricht einem Grobentwurf, also einer Zerlegung der Gesamtanweisung S_{alg} in aufeinanderfolgende Teilanweisungen, d. h.

$$S_{\mathrm{alg}} \equiv S_{\mathrm{dekl}};\ S_{\mathrm{be}};\ S_{\mathrm{aus}};\ S_{\mathrm{av}};\ S_{\mathrm{norm}};\ S_{\mathrm{rand}}\,. \tag{6.44}$$

Aus bereits im Kapitel 5 diskutierten Gründen werden wir den Korrektheitsbeweis führen, indem wir eine Vorbedingung aus einer vorgegebenen Nachbedingung ermitteln.[1] Insofern ist es sinnvoll, an dieser Stelle die Nachbedingung des Algorithmus formal zu spezifizieren, um auf dieser Basis dann die

[1]Die Bedingungen stellen Prädikate dar, die von ungebundenen Variablen abhängen. Streng genommen müssten die Bezeichner der Prädikate mit diesen Gegenständen erfolgen. Der Übersicht halber wollen wir dennoch darauf verzichten. Beispiel : Sei $\mathsf{P}(x) \equiv x > 0$. Dann schreiben wir anstatt $\mathsf{P}(x)$ nur P. Dieses unpräzise Vorgehen wirft aber insbesondere dann Probleme auf, wenn die Ersetzungsregel genutzt wird. Falls im weiteren Verlauf der Arbeit auf ein Prädikat die Ersetzungsregel angewendet wird, werden wir abweichend von der obigen Vereinbarung die präzise Darstellung wählen.

Teilanweisungen zu entwerfen. Die Nachbedingung an den Algorithmus stellt den Datenzustand des Programms dar, der nach Ausführung des Algorithmus erreicht werden soll. Wir führen hierfür weitere Zwischenbedingungen ein

Bedingung	Anweisung	Kommentar
$\{\mathsf{V}\}$		
	S_{dekl}	Variablendeklarationen
$\{\mathsf{P}_{\text{dekl}}\}$		
	S_{be}	Berechnung des Vektors $\boldsymbol{b}_e$
$\{\mathsf{P}_{\text{be}}\}$		
	S_{aus}	Berechnung der Ausgangssignale $\boldsymbol{a}_{\text{FB}}$
$\{\mathsf{P}_{\text{aus}}\}$		
	S_{av}	Berechnung der Verzögererwerte $\boldsymbol{a}_v$ am Rand
$\{\mathsf{P}_{\text{av}}\}$		
	S_{norm}	Berechnung der Verzögerer $\boldsymbol{b}_v$ im Normalfall
$\{\mathsf{P}_{\text{norm}}\}$		
	S_{rand}	Berechnung der Verzögererwerte $\boldsymbol{b}_v$ am Rand
$\{\mathsf{P}\}$		

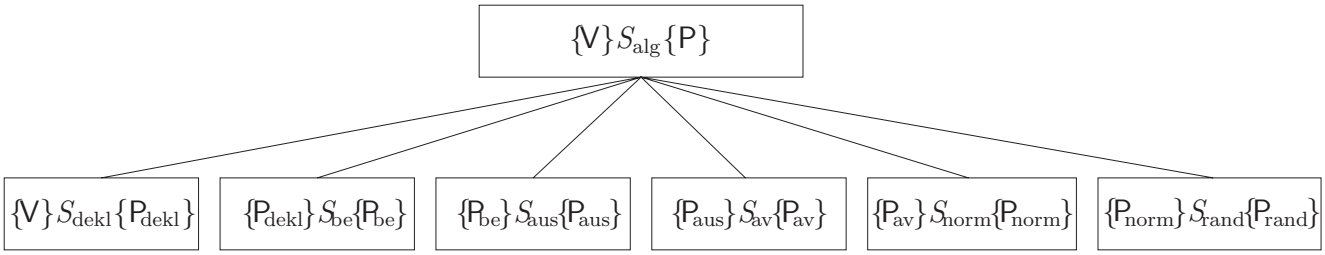


Bild 6.3: Übersicht über den Korrektheitsbeweis

Im Bild 6.3 ist der Algorithmus und die Zerlegung in sequentielle Teilanweisungen dargestellt. Dieses Diagramm und die folgenden Diagramme sind so zu verstehen, dass aus der Korrektheit aller untergeordneten Prädikate die Korrektheit des übergeordneten Prädikats folgt. Die einzelnen Teilanweisungen werden in den Kapiteln 6.2-6.7 festgelegt. Jede Nachbedingung dieser Teilanweisungen soll aus der konjunktiven Verknüpfung der Vorbedingung und einem bestimmten zusätzlichem Teil bestehen. Um diese zusätzlichen Nachbedingungen von den Nachbedingungen zu unterscheiden, kennzeichnen wir die Bezeichner der Prädikate dieser zusätzlichen Nachbedingungen mit einem $'$.

Mithilfe der definierten Prädikate können wir nun die zusätzlichen Nachbedingungen der 6 Teilanweisungen aus Gleichung (6.44) in der folgenden Tabelle darstellen

Anweisung	zusätzliche Nachbedingung
S_{dekl}	$\mathsf{P}'_{\mathrm{be}} \equiv \mathsf{P}_{\mathrm{decbe}} \wedge \mathsf{P}_{\mathrm{decav}} \wedge \mathsf{P}_{\mathrm{decuv}}$
S_{be}	$\mathsf{P}'_{\mathrm{be}} \equiv \bigwedge_{\delta=0}^{P-1} \bigwedge_{k=1}^{n_e} \mathsf{P}_{e\,k}\,(\mu_1(\delta), \mu_2(\delta), \mu_3(\delta), \delta) \wedge \mathsf{P}_{\mathrm{be\,max}}$
S_{aus}	$\mathsf{P}'_{\mathrm{aus}} \equiv \bigwedge_{\delta=0}^{P-1} \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\,\mathrm{FB}k}\,(\boldsymbol{\mu}(\delta), \delta)$
S_{av}	$\mathsf{P}'_{\mathrm{av}} \equiv \mathsf{P}'_{\mathrm{avx}} \wedge \mathsf{P}'_{\mathrm{avx}} \wedge \mathsf{P}'_{\mathrm{avx}} \wedge \mathsf{P}_{\mathrm{av\,max}}$
S_{norm}	$\mathsf{P}'_{\mathrm{norm}} \equiv \bigwedge_{\kappa=1}^{7} \mathsf{P}^{\kappa}_{\mathrm{norm}}$
S_{rand}	$\mathsf{P}'_{\mathrm{rand}} \equiv \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \mathsf{P}_{\mathrm{x}}(\mu_2, \mu_3) \wedge \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \mathsf{P}_{\mathrm{y}}(\mu_1, \mu_3) \wedge \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{z}}(\mu_1, \mu_2)$

Die Nachbedingung des Gesamtalgorithmus ist nun die konjunktive Verknüpfung mehrerer dieser Bedingungen, d.h.

$$\mathsf{P} \equiv \mathsf{P}'_{\mathrm{aus}} \wedge \mathsf{P}'_{\mathrm{norm}} \wedge \mathsf{P}'_{\mathrm{randx}} \wedge \mathsf{P}'_{\mathrm{randy}} \wedge \mathsf{P}'_{\mathrm{randz}} \,. \tag{6.45}$$

Die in der Tabelle auftretenden und nicht in P berücksichtigten Bedingungen können beispielsweise die Speicherung von Zwischenwerten fordern.

Bevor wir zur eigentlichen Beweisführung übergehen, geben wir die Grobübersicht noch einmal an, diesmal aber mit den Zwischenbedingungen

Bedingung		Anweisung
$\{\mathsf{V}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge \mathsf{P}_{\mathrm{decaFB}} \wedge \mathsf{P}_{\mathrm{name}}$	S_{dekl}
$\{\mathsf{P}_{\mathrm{dekl}}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}}$	S_{be}
$\{\mathsf{P}_{\mathrm{be}}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}'_{\mathrm{be}}$	S_{aus}
$\{\mathsf{P}_{\mathrm{aus}}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}'_{\mathrm{be}} \wedge \mathsf{P}'_{\mathrm{aus}}$	S_{av}
$\{\mathsf{P}_{\mathrm{av}}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}'_{\mathrm{be}} \wedge \mathsf{P}'_{\mathrm{aus}} \wedge \mathsf{P}'_{\mathrm{av}}$	S_{norm}
$\{\mathsf{P}_{\mathrm{norm}}\}$	$\equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}'_{\mathrm{aus}} \wedge \mathsf{P}'_{\mathrm{av}} \wedge \mathsf{P}'_{\mathrm{norm}}$	S_{rand}
$\{\mathsf{P}\}$	$\equiv \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}'_{\mathrm{aus}} \wedge \mathsf{P}'_{\mathrm{norm}} \wedge \mathsf{P}'_{\mathrm{randx}} \wedge \mathsf{P}'_{\mathrm{randy}} \wedge \mathsf{P}'_{\mathrm{randz}}$	

6.2 Berechnung der Verzögererwerte im Randfall

Die Anweisung zur Berechnung der Verzögererwerte im Randfall S_{rand} zerlegen wir in

$$S_{\mathrm{rand}} \equiv S_{\mathrm{randx}}; S_{\mathrm{randy}}; S_{\mathrm{randz}} \,. \tag{6.46}$$

Wir werden nur die Anweisung S_{randz} ausführlich behandeln. Die anderen zwei Anweisungen sind zu S_{randz} analog, vgl. Bild 6.4. Die Nachbedingung der Anweisung S_{randz} entspricht der des gesamten Algorithmus S_{alg}, also P gemäß Gleichung (6.45). Der Beitrag, den die Anweisung S_{randz} zu P leistet, ist die Berechnung der Verzögererwerte am Rand $z = \mathrm{const.}$, formal repräsentiert durch die zusätzliche Nachbedingung $\mathsf{P}'_{\mathrm{randz}}$. Um das geforderte Ziel zu erreichen, nutzen wir den im Bild 6.5 dargestellten

Algorithmus. Die Implementierung in C-Code lautet :

Anweisungssequenz $\{\mathsf{P}_{\text{randy}}\}S_{\text{randz}}\{\mathsf{P}_{\text{randz}}\}$

```
for(mu1 = 0 ; mu1 < P_{x_1} ; mu1=mu1+1){
   for(mu2 = 0 ; mu2 < P_{x_2} ; mu2=mu2+1){
      mu3 = 0;
      delta = mu1 + mu2 *  P_{x_1}  +  mu3 *  P_{x_1}  *  P_{x_2} ;
      S_{b^3_{v1}}
      ⋮
      S_{b^3_{vn^3_v}}
      mu3 =  P_{x_1}  - 1;
      delta = mu1 + mu2 *  P_{x_1}  +  mu3 *  P_{x_1}  *  P_{x_2} ;
      S_{b^6_{v1}}
      ⋮
      S_{b^6_{vn^6_v}}
   }
}
```

wobei die Zuweisungsoperationen durch

$$S_{b^3_{v1}} \equiv \texttt{b\_v\_3\_1 [delta]=}\, r^3_{11}\, \texttt{*a\_v\_6\_1[delta]+...+}r^3_{1n^6_v}\, \texttt{*a\_v\_6\_}n^6_v\texttt{[delta];}$$
$$\vdots$$
$$S_{b^3_{vn^3_v}} \equiv \texttt{b\_v\_3\_}n^3_v\texttt{[delta]=}\, r^3_{n^3_v 1}\texttt{*a\_v\_6\_1[delta]+...+}r^3_{n^3_v n^6_v}\texttt{*a\_v\_6\_}n^6_v\texttt{[delta];}$$
$$S_{b^6_{v1}} \equiv \texttt{b\_v\_6\_1 [delta]=}\, r^6_{11}\, \texttt{*a\_v\_3\_1[delta]+...+}r^3_{1n^3_v}\, \texttt{*a\_v\_1\_}n^3_v\texttt{[delta];}$$
$$\vdots$$
$$S_{b^6_{vn^6_v}} \equiv \texttt{b\_v\_6\_}n^6_v\texttt{[delta]=}\, r^6_{n^6_v 1}\texttt{*a\_v\_3\_1[delta]+...+}r^3_{n^6_v n^3_v}\texttt{*a\_v\_3\_}n^3_v\texttt{[delta];}$$

gegeben sind.

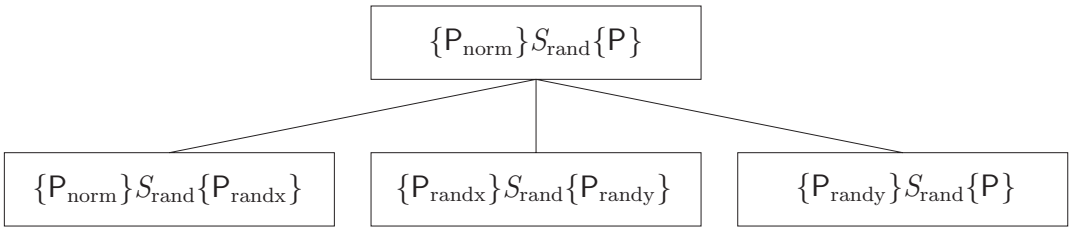


Bild 6.4: Die Berechnung der Verzögererwerte im Randfall

Wir ermitteln nun die Vorbedingung $\mathsf{P}_{\text{randy}}$. Im Bild 6.6 sind die notwendigen Schritte zum Nachweis der Korrektheit der Anweisung S_{randz} dargestellt. Bei S_{randz} handelt es sich um zwei verschachtelte for-Schleifen mit den Schleifenbedingungen $b_a \equiv \mu_1.v < P_{x_1}$ und $b_i \equiv \mu_2.v < P_{x_2}$. Dem Bild 6.6 ist weiter zu entnehmen, dass jeweils 2 Mal Axiom (5.35) und Axiom (5.7) zur Anwendung kommen. Die verbleibende Aufgabe besteht nun darin, die Korrektheit der im Bild 6.6 dargestellten Anweisungen nachzuweisen. Vorab wollen wir uns Gedanken über die Reihenfolge der Beweisführung machen. Wir werden zuerst die Invariante der äußeren Schleife I_{a} in einem Satz angeben. Im Anschluss daran weisen wir durch die Initialisierungsanweisung die Vorbedingung

$$\mathsf{P}_{\text{randy}} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{dec}} \wedge \mathsf{P}'_{\text{aus}} \wedge \mathsf{P}'_{\text{av}} \wedge \mathsf{P}'_{\text{norm}} \wedge \mathsf{P}'_{\text{randx}} \wedge \mathsf{P}'_{\text{randy}} \tag{6.47}$$

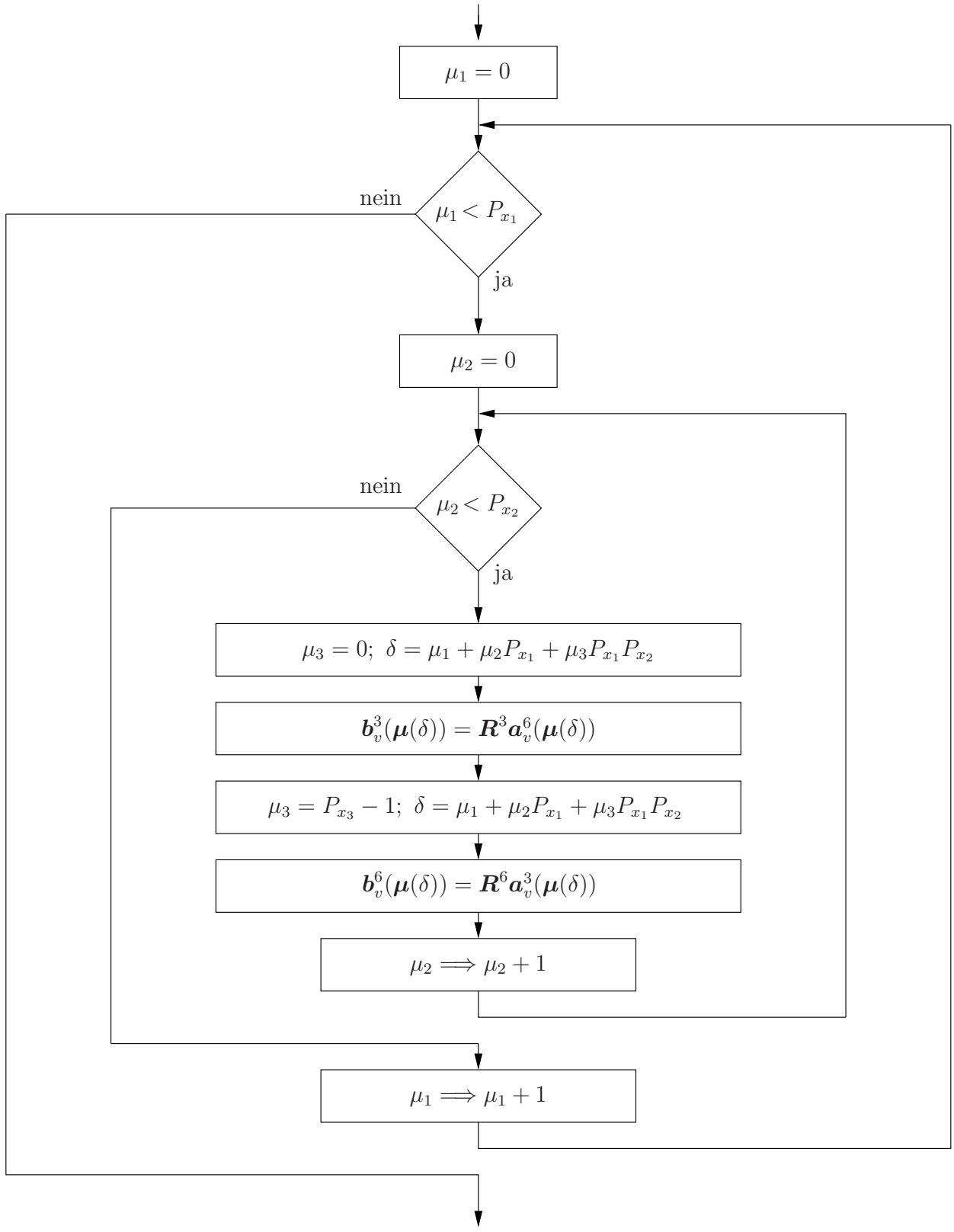


Bild 6.5: Berechnung der neuen Verzögererwerte $\boldsymbol{b}_v^3$ und $\boldsymbol{b}_v^6$ an den Grenzflächen $\mu_3 = 0$ und $\mu_3 = P_{x_3} - 1$

und über die Abbruchbedingung die Nachbedingung P nach. Das Zeigen der Invarianz von $\mathsf{I_a}$ bzgl. $S_{4a}; S_{3a}$ führt uns schließlich zur inneren Schleife.

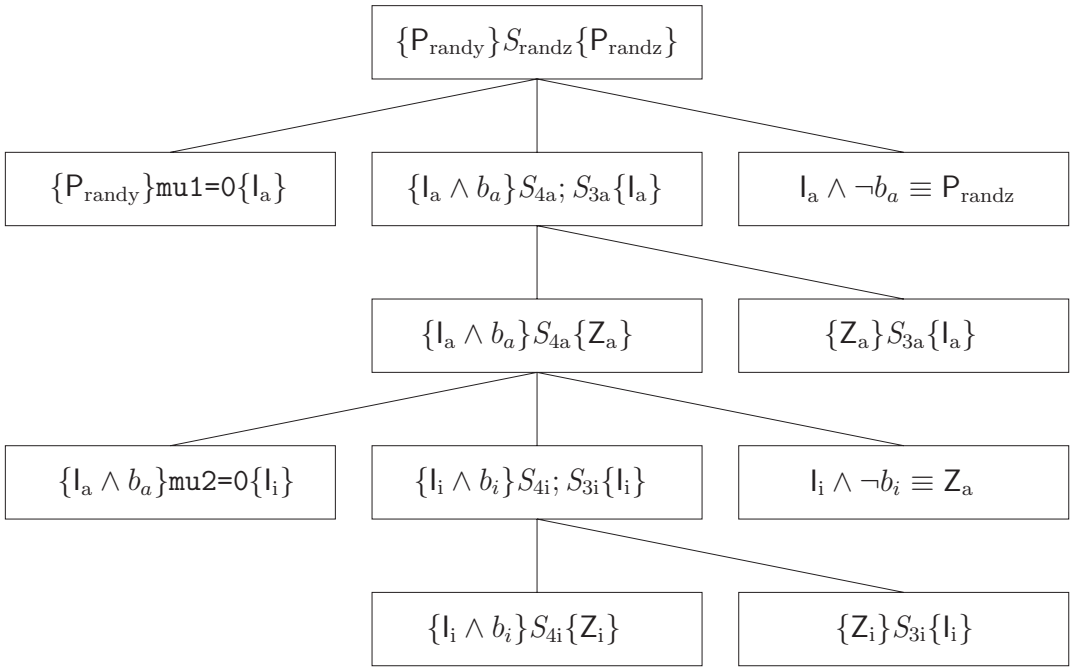


Bild 6.6: Übersicht zum Korrektheitsbeweis der Anweisung S_{randz}.

Satz 3 *Der Ausdruck*

$$\mathsf{I_a} \equiv 0 \leq \mu_1.v \leq P_{x_1} \wedge \mathsf{P}_{\text{randy}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P_z}(\mu_1, \mu_2) \tag{6.48}$$

ist eine gültige Schleifeninvariante der äußeren Schleife.

Beweis 3

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu1=0` liefert die Vorbedingung

$$0 \leq P_{x_1} \wedge \mathsf{P}_{\text{randy}} \wedge \bigwedge_{\mu_1=0}^{-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P_z}(\mu_1, \mu_2) \wedge 0 \leq \text{Max}_\text{i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int} \equiv \mathsf{P}_{\text{randy}}\,, \tag{6.49}$$

da $\mathsf{P}_{\text{randy}}$ den Ausdruck $\mu_1.typ = \texttt{int} \wedge \mu_1.l = -1$ impliziert. Die Nachbedingung bestimmt sich zu

$$\mathsf{I_a} \wedge \mu_1.v \geq P_{x_1} \equiv \mathsf{P}_{\text{randy}} \wedge \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P_z}(\mu_1, \mu_2) \wedge \mu_1.v = P_{x_1} \equiv \mathsf{P}_{\text{randz}} \wedge \mu_1.v = P_{x_1} \tag{6.50}$$

und kann gemäß Gleichung (5.10) zu $\mathsf{P}_{\text{randz}}$ geschwächt werden, weil der Wert der Variablen `mu1` nicht weiter benötigt wird. Die Zwischenbedingung der äußeren Schleife bestimmt sich zu

$$\mathsf{Z_a} \equiv 0 \leq \mu_1.v{+}1 \leq P_{x_1} \wedge \mathsf{P}_{\text{randy}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P_z}(\mu_1, \mu_2) \wedge \mu_1.v{+}1 \leq \text{Max}_\text{i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int}\,.$$

(6.51)

Wir stärken die Zwischenbedingung um $0 \leq \mu_1.v$ zu

$$\mathsf{Z}'_{\mathrm{a}} \equiv \mathsf{Z}_{\mathrm{a}} \wedge 0 \leq \mu_1.v \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2)\,. \tag{6.52}$$

Der Beweis von $\{\mathsf{I}_{\mathrm{a}} \wedge b_a\} S_{4\mathrm{a}} \{\mathsf{Z}'_{\mathrm{a}}\}$ wird erneut gemäß Axiom (5.35) durchgeführt, da die Anweisung $S_{4\mathrm{a}}$ die innere for-Schleife ist.

Satz 4 *Der Ausdruck*

$$\mathsf{I}_{\mathrm{i}} \equiv 0 \leq \mu_2.v \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v-1} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2) \tag{6.53}$$

ist eine gültige Schleifeninvariante der inneren Schleife.

Beweis 4

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu2=0` liefert die behauptete Vorbedingung von $S_{4\mathrm{a}}$

$$0 \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{-1} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2) \wedge 0 \leq \mathrm{Max}_{\mathrm{i}} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int} \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a\,. \tag{6.54}$$

Die Nachbedingung der inneren Schleife lautet

$$\mathsf{Z}_{\mathrm{a}} \equiv \mathsf{I}_{\mathrm{i}} \wedge \mu_2.v \geq P_{x_2} \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2)\,. \tag{6.55}$$

Die Zwischenbedingung der inneren Schleife lautet

$$\mathsf{Z}_{\mathrm{i}} \equiv 0 \leq \mu_2.v+1 \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2) \wedge \mu_2.v+1 \leq \mathrm{Max}_{\mathrm{i}} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int}\,. \tag{6.56}$$

Wir stärken Sie durch $0 \leq \mu_2.v$ zu

$$\mathsf{Z}'_{\mathrm{i}} \equiv \mathsf{Z}_{\mathrm{i}} \wedge 0 \leq \mu_2.v \equiv 0 \leq \mu_2.v \leq P_{x_2}-1 \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v} \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2) \equiv \mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2.v)\,. \tag{6.57}$$

Es verbleibt, die Richtigkeit von $\{\mathsf{I}_{\mathrm{i}} \wedge b_i\} S_{4\mathrm{i}} \{\mathsf{Z}'_{\mathrm{i}}\}$ zu zeigen. Wir ermitteln dazu zunächst die Vorbedingung der Anweisung $S_{b^6_{v\,n^6_v}}$. Mit Satz (5.43) erhalten wir die Vorbedingung zu

$$\begin{aligned} &\mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2.v) \begin{bmatrix} b^6_{v\,n^6_v}(\delta.v).v \\ \boldsymbol{r}^6_{n^6_v,\bullet}\, \boldsymbol{a}^3_v(\delta.v).v \end{bmatrix} \wedge |\boldsymbol{r}^6_{n^6_v,\bullet} \boldsymbol{a}^3_v(\delta.v).v| \leq \mathrm{Max}_{\mathrm{f}} \wedge 0 \leq \delta.v \leq b^6_{v\,n^6_v}.l - 1 \wedge \\ &\bigwedge_{\nu=1}^{n^3_v} 0 \leq \delta.v \leq a^3_{v\,\nu}.l - 1 \wedge \bigwedge_{\nu=1}^{n^3_v} b^6_{v\,n^6_v}.typ = a^3_{v\,\nu}.typ \wedge \delta\,.l = -1 \wedge \delta\,.typ = \texttt{int}\,. \end{aligned} \tag{6.58}$$

Die Bedingungen an die Beschränktheit der Teilsummen des Skalarprodukts treten nicht auf, da nur eine Koordinate von $\boldsymbol{r}^6_{n_v^6,\bullet}$ von null verschieden ist. Das Bestreben ist es im Folgenden, die Teile dieser Aussage abzuspalten, die $\mathsf{P}_{\text{randy}}$ impliziert. Wir werden dabei mit den hinteren Teilaussagen anfangen, d. h. mit den Bedingungen an den Datentyp. Die Verwendung von Gleichung (6.48) i.V.m. Gleichung (6.33) und Gleichung (6.34) liefert

$$\mathsf{P}_{\text{randy}} \Rightarrow b^6_{v\,n_v^6}.typ = \texttt{float} \wedge \bigwedge_{l=1}^{n_v^3} a^3_{v\,l}.typ = \texttt{float} \Rightarrow \bigwedge_{l=1}^{n_v^3} a^3_{v\,l}.typ = b^6_{v\,n_v^6}.typ\,. \tag{6.59}$$

Die Aussage

$$\bigwedge_{l=1}^{n_v^3} a^3_{v\,l}.typ = b^6_{v\,n_v^6}.typ$$

in Gleichung (6.58) ist somit wahr und kann gemäß $[(\mathsf{A} \rightarrow \mathsf{B}) \wedge \mathsf{A}] \rightarrow \mathsf{B}$ abgetrennt werden.

Um in der Vorbedingung Gleichung (6.58), die Bedingungen an die Länge der Felder abtrennen zu können, müssen wir uns eine Aussage über $\delta.v$ verschaffen. Dazu nehmen wir zunächst eine Stärkung von Gleichung (6.58) um $0 \leq \delta.v \leq P-1$ vor. Wir fordern also zusätzlich, dass $\delta.v$ innerhalb eines Intervalls liegt. Zudem nutzen wir Gleichung (6.48) i.V.m. Gleichung (6.33) und Gleichung (6.34) zur Gewinnung von

$$\mathsf{P}_{\text{randy}} \Rightarrow 0 \leq P-1 \leq b^6_{v\,n_v^6}.l - 1 \wedge \bigwedge_{l=1}^{n_v^3} 0 \leq P-1 \leq a^3_{v\,l}.l - 1\,, \tag{6.60}$$

d. h. die Feldlängen betragen mindestens P. Konjunktive Verknüpfung von $\mathsf{P}_{\text{randy}}$ und $0 \leq \delta.v \leq P-1$ liefert

$$\mathsf{P}_{\text{randy}} \wedge 0 \leq \delta.v \leq P-1 \Rightarrow 0 \leq \delta.v \leq b^6_{v\,n_v^6}.l - 1 \wedge \bigwedge_{l=1}^{n_v^3} 0 \leq \delta.v \leq a^3_{v\,l}.l - 1\,. \tag{6.61}$$

Die Aussage

$$0 \leq \delta.v \leq b^6_{v\,n_v^6}.l - 1 \wedge \bigwedge_{l=1}^{n_v^3} 0 \leq \delta.v \leq a^3_{v\,l}.l - 1 \tag{6.62}$$

in Gleichung (6.58) ist somit schon in $\mathsf{P}_{\text{randy}} \wedge 0 \leq \delta.v \leq P-1$ enthalten und kann entfernt werden. Zudem gilt

$$\mathsf{P}_{\text{randy}} \Rightarrow \delta\,.l = -1 \wedge \delta\,.typ = \texttt{int}\,. \tag{6.63}$$

Es verbleibt die Richtigkeit von $0 \leq \delta.v \leq P-1$ zu zeigen. Die Vorbedingung von S_{4i} impliziert die Ungleichungen $0 \leq \mu_1.v \leq P_{x_1} - 1 \wedge 0 \leq \mu_2.v \leq P_{x_2} - 1$ und P_{MDED}. Daher ist nach Satz 6.4 die Richtigkeit von $0 \leq \delta.v \leq P-1$ schon gezeigt, wenn $0 \leq \mu_3.v \leq P_{x_3} - 1$. gilt. An dieser Stelle kann allerdings keine weitere Vereinfachung des Ausdrucks vorgenommen werden. $0 \leq \mu_3.v \leq P_{x_3} - 1$ wird in die Vorbedingung übernommen.

Die Zuweisungsanweisungen in Gleichung (6.58) fordern, dass $|\boldsymbol{r}^6_{n_v^6,\bullet}\boldsymbol{a}^3_v(\delta.v).v| < \text{Max}_\text{f}$ gilt. Da die Matrizen $\boldsymbol{R}^\kappa$ unitär beschränkt sind, ist dies dann der Fall, wenn $\bigwedge_{\nu=1}^{n_v^3} |a^3_{v\,\nu}(\delta.v).v| < \text{Max}_\text{f}$ gilt.

Um den in Gleichung (6.58) auftretenden Ausdruck

$$\mathsf{P}_\text{z}(\mu_1.v, \mu_2.v) \begin{bmatrix} b^6_{v\,n_v^6}(\delta.v).v \\ \boldsymbol{r}^6_{n_v^6,\bullet}\boldsymbol{a}^3_v(\delta.v).v \end{bmatrix} \tag{6.64}$$

zu vereinfachen, schreiben wir $\mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2.v)$ ausführlich auf

$$\mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2.v) \equiv \bigwedge_{k=1}^{n_v^3} \mathsf{P}_{b\,k}^3\,(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})) \wedge \bigwedge_{k=1}^{n_v^6-1}$$

$$\mathsf{P}_{b\,k}^6\,(\mu_1.v, \mu_2.v, P_{x_3}-1, \delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})) \wedge \mathsf{P}_{b\,n_v^6}^6\,(\mu_1.v, \mu_2.v, P_{x_3}-1, \delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}}))\,, \tag{6.65}$$

wobei

$$\mathsf{P}_{b\,n_v^6}^6(\mu_1.v, \mu_2.v, P_{x_3}\text{–}1, \delta([\mu_1.v, \mu_2.v, P_{x_3}\text{–}1]^{\mathrm{T}})) \equiv b_{v\,n_v^6}^6(\delta([\mu_1.v, \mu_2.v, P_{x_3}\text{–}1]^{\mathrm{T}})).v = b_{v\,n_v^6}^6([\mu_1.v, \mu_2.v, P_{x_3}\text{–}1]^{\mathrm{T}}). \tag{6.66}$$

Offenbar erfährt nur der letzte Konjunktionsterm von $\mathsf{P}_{\mathrm{z}}(\mu_1.v, \mu_2.v)$ eine Änderung in Gleichung (6.64). Wir werden im Folgenden diesen untersuchen, d.h.

$$b_{v\,n_v^6}^6(\delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})).v = b_{v\,n_v^6}^6([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}}) \begin{bmatrix} b_{v\,n_v^6}^{6\,(\delta.v).v} \\ \boldsymbol{r}_{n_v^6,\bullet}^{6}\,\boldsymbol{a}_v^{3\,(\delta.v).v} \end{bmatrix} \tag{6.67}$$

bestimmen. Dazu stärken wir allerdings noch die Vorbedingung von $S_{b_{v\,n_v^6}^6}$ durch $\delta.v = \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2}$. Aus dem Teilprädikat Gleichung (6.67) wird durch Nutzen der Stärkung und Ersetzen des δ durch $\delta.v$ auf der linken Seite

$$\boldsymbol{r}_{n_v^6,\bullet}^{6}\,\boldsymbol{a}_v^3\,(\delta.v).v = b_{v\,k}^6([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})\,. \tag{6.68}$$

Nutzen wir nun die Stärkung in anderer Richtung, um in $\boldsymbol{a}_v^3\,(\delta.v).v$ das $\delta.v$ durch $\delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})$ zu ersetzen, so haben wir

$$\boldsymbol{r}_{n_v^6,\bullet}^{6}\,\boldsymbol{a}_v^3\,(\delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})).v = b_{v\,k}^6([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})\,. \tag{6.69}$$

Wir werden nun die Istwerte der abhängigen Variablen durch die Sollwerte ersetzen. Dazu nutzen wir Gleichung (6.48) i.V.m. Gleichung (6.33) und Gleichung (6.35) und ermitteln

$$\mathsf{P}_{\mathrm{randy}} \Rightarrow \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{avz}}(\mu_1, \mu_2) \Rightarrow \boldsymbol{a}_v^3\,(\delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})).v = \boldsymbol{a}_v^3\,([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})\,. \tag{6.70}$$

Aus Gleichung (6.69) wird dann

$$\boldsymbol{r}_{n_v^6,\bullet}^{6}\,\boldsymbol{a}_v^3\,([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}}) = b_{v\,k}^6([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}})\,. \tag{6.71}$$

Diese Beziehung ist offenbar in $\mathsf{P}_{\mathrm{WDF}}$ enthalten und kann somit auch aus der Vorbedingung Gleichung (6.58) abgetrennt werden. Die gestärkte Vorbedingung aus Gleichung (6.58) ergibt sich also zu

$$\mathsf{I}_{\mathrm{i}} \wedge\, b_i \,\wedge \delta.v = \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2} \wedge\, 0 \le \mu_3.v \le P_{x_3}-1 \,\wedge \bigwedge_{\nu=1}^{n_v^3} |a_{v\,\nu}^3(\delta.v).v| \le \mathrm{Max_f}$$

$$\bigwedge_{k=1}^{n_v^3} \mathsf{P}_{b\,k}^3\,(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})) \,\wedge \bigwedge_{k=1}^{n_v^6-1} \mathsf{P}_{b\,k}^6\,(\mu_1.v, \mu_2.v, P_{x_3}-1, \delta([\mu_1.v, \mu_2.v, P_{x_3}-1]^{\mathrm{T}}))\,. \tag{6.72}$$

Die aus den Anweisungen $S_{b^6_{vn^6_v-1}}$ bis einschließlich $S_{b^6_{v1}}$ resultierenden Vorbedingungen können ebenso umgeformt und abgetrennt werden. Als Vorbedingung von $S_{b^6_{v1}}$ erhalten wir

$$\begin{aligned}&\mathsf{I}_\mathrm{i} \wedge b_i \wedge \delta.v = \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2} \wedge 0 \le \mu_3.v \le P_{x_3}-1 \wedge \\ &\bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{b\,k}\left(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0]^\mathrm{T})\right) \wedge \bigwedge_{\nu=1}^{n_v^3} |a^3_{v\,\nu}(\delta.v).v| < \mathrm{Max_f}\,.\end{aligned} \tag{6.73}$$

Die Vorbedingung von `delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;` lautet mit Gleichung (6.41)

$$\begin{aligned}&\mathsf{I}_\mathrm{i} \wedge b_i \wedge \mu_1.v + \mu_2.vP_{x_1} + \mu_3.vP_{x_1}P_{x_2} = \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2} \wedge \\ &\bigwedge_{\nu=1}^{n_v^3} |\boldsymbol{a}^3_{v\,\nu}\left(\mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2}\right).v| \le \mathrm{Max_f} \wedge \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{b\,k}\left(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0]^\mathrm{T})\right) \wedge \\ &0 \le \mu_1.v \le P_{x_1}-1 \wedge 0 \le \mu_2.v \le P_{x_2}-1 \wedge 0 \le \mu_3.v \le P_{x_3}-1\,.\end{aligned} \tag{6.74}$$

Die Vorbedingung von `mu3=` P_{x_3} `-1;` ergibt sich leicht aus Gleichung (6.74) zu

$$\begin{aligned}&\mathsf{I}_\mathrm{i} \wedge b_i \wedge \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2} = \mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2} \wedge 0 \le P_{x_3}-1 \wedge \\ &\bigwedge_{\nu=1}^{n_v^3} |\boldsymbol{a}^3_{v\,\nu}\left(\mu_1.v + \mu_2.vP_{x_1} + (P_{x_3}-1)P_{x_1}P_{x_2}\right).v| \le \mathrm{Max_f} \wedge \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{b\,k}\left(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0]^\mathrm{T})\right) \wedge \\ &0 \le \mu_1.v \le P_{x_1}-1 \wedge 0 \le \mu_2.v \le P_{x_2}-1\,.\end{aligned} \tag{6.75}$$

Sowohl die wahren Aussagen, als auch die von $\mathsf{I}_\mathrm{i} \wedge b_i$ implizierten Aussagen können abgetrennt werden und es verbleibt

$$\mathsf{I}_\mathrm{i} \wedge b_i \wedge \bigwedge_{k=1}^{n_v^3} \mathsf{P}^3_{b\,k}\left(\mu_1.v, \mu_2.v, 0, \delta([\mu_1.v, \mu_2.v, 0])\right). \tag{6.76}$$

Es folgen die Anweisungen $S_{b^3_{vn^3_v}}$ bis einschließlich `mu3=0;` deren Beweis ähnlich dem vorherigen ist. Für die Vorbedingung der Anweisung $S_{4\mathrm{i}}$ verbleibt, wie behauptet, $\mathsf{I}_\mathrm{i} \wedge b_i$.

Die Programmablaufpläne und Anweisungssequenzen für S_randx und S_randy entsprechen bis auf kleine Änderungen denen der Anweisungssequenz S_randz und werden hier nicht angegeben. Die Beweise werden ebenfalls analog zu denen der Anweisung S_randz geführt.

6.3 Normalberechnung der Verzögererwerte

Die Anweisung S_norm dient der Berechnung der neuen Verzögererwerte innerhalb des Berechnungsgebietes, d. h., es erfolgt die Berechnung von Gleichung (6.4). Die zu erfüllende Nachbedingung lautet

$$\mathsf{P}_\mathrm{norm} \equiv \mathsf{P}_\mathrm{WDF} \wedge \mathsf{P}_\mathrm{dec} \wedge \mathsf{P}'_\mathrm{aus} \wedge \mathsf{P}'_\mathrm{av} \wedge \mathsf{P}'_\mathrm{norm}\,. \tag{6.77}$$

Die Anweisung S_norm erfolgt durch Zerlegung in 7 sequentielle Anweisungen von denen jeweils die Anweisung $S_{\kappa\mathrm{norm}}$ den Vektor $\boldsymbol{b}^\kappa_v$ berechnet. Dazu die folgende Übersicht

Bedingung	Anweisung
$\{\mathsf{P}_{\text{av}}\}$	
	$S_{1\text{norm}}$
$\{\mathsf{P}^1_{\text{norm}}\}$	
	$S_{2\text{norm}}$
	$\vdots$
	$S_{6\text{norm}}$
$\{\mathsf{P}^6_{\text{norm}}\} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{dec}} \wedge \mathsf{P}'_{\text{aus}} \wedge \mathsf{P}'_{\text{av}} \wedge \bigwedge_{\kappa=1}^{6} \mathsf{P}^{\kappa}_{\text{norm}}$	
	$S_{7\text{norm}}$
$\{\mathsf{P}_{\text{norm}}\} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{dec}} \wedge \mathsf{P}'_{\text{aus}} \wedge \mathsf{P}'_{\text{av}} \wedge \mathsf{P}'_{\text{norm}}$	

Im Bild 6.7 ist der Programmablaufplan für die Berechnung des Vektors $\boldsymbol{b}^1_v$ dargestellt. Die Grenzen der Schleifen der Anweisungen $S_{2\text{norm}}$ bis $S_{7\text{norm}}$ ergeben sich aus Tabelle 6.1. Im Bild 6.8 sind die für die Beweisführung nötigen Schritte für die Anweisung $S_{1\text{norm}}$ dargestellt. [2] Die Schleifenbedingungen sind $b_a \equiv \mu_1.v < P_{x_1} - 1 \equiv \mu_1.v + 1 \leq P_{x_1} - 1$, $b_m \equiv \mu_2.v < P_{x_2} \equiv \mu_2.v + 1 \leq P_{x_2}$ und $b_i \equiv \mu_3.v < P_{x_3} \equiv \mu_3.v + 1 \leq P_{x_3}$. Aus dem Programmablaufplan ermitteln wir den C-Code von $S_{1\text{norm}}$ zu

BedingungAnweisung	Kommentar
$\{\mathsf{P}_{\text{av}}\}$	
`for(mu1 = 0 ; mu1 <` P_{x_1} `-1 ; mu1=mu1+1){`	
`for(mu2 = 0 ; mu2 <` P_{x_2} `; mu2=mu2+1){`	
`for(mu3 = 0 ; mu3 <` P_{x_3} `; mu3=mu3+1){`	
`delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;`	
S_{bv11}	Wellen b^1_{v1} berechnen
$\vdots$	
$S_{\text{bv1}n^1_v}$	Wellen $b^1_{vn^1_v}$ berechnen
`}`	
`}`	
`}`	
$\{\mathsf{P}^1_{\text{norm}}\}$	

mit den Anweisungen

$S_{\text{bv1}} \equiv$ `b_v_1_1 [delta+1]=` $p^1_{v^1 e\,11}$ `*b_e_1[delta]+...+`$p^1_{v^1 e\,1 n_e}$ `*b_e_`n_e`[delta];`

$\vdots$

$S_{\text{bv}n^1_v} \equiv$ `b_v_1_`n^1_v`[delta+1]=` $p^1_{v^1 e\,n^1_v 1}$`*b_e_1[delta]+...+`$p^1_{v^1 e\,n^1_v n_e}$`*b_e_`n_e`[delta];`

Der Beweis wird nur für die erste der 7 Teilanweisungen geführt, da die Teilanweisungen einander ähnlich sind.

Satz 5 *Der Ausdruck*

$$\mathsf{I}_{\text{a}} \equiv 0 \leq \mu_1.v \leq P_{x_1} - 1 \wedge \mathsf{P}_{\text{av}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n^1_v} \mathsf{P}^1_{bk}(\mu_1+1, \mu_2, \mu_3, \delta(\boldsymbol{\mu})+\delta_1) \tag{6.78}$$

[2] Eigentlich müssten für die hier auftretenden Invarianten I_{a}, I_{i}, und Bedingungen b_a, b_i neue Bezeichnungen eingeführt werden, da sie schon im Kapitel 6.2 verwendet wurden. Um die Übersicht zu wahren, verzichten wir jedoch darauf und wollen die Gültigkeit dieser Bezeichner auf die Unterkapitel beschränken.

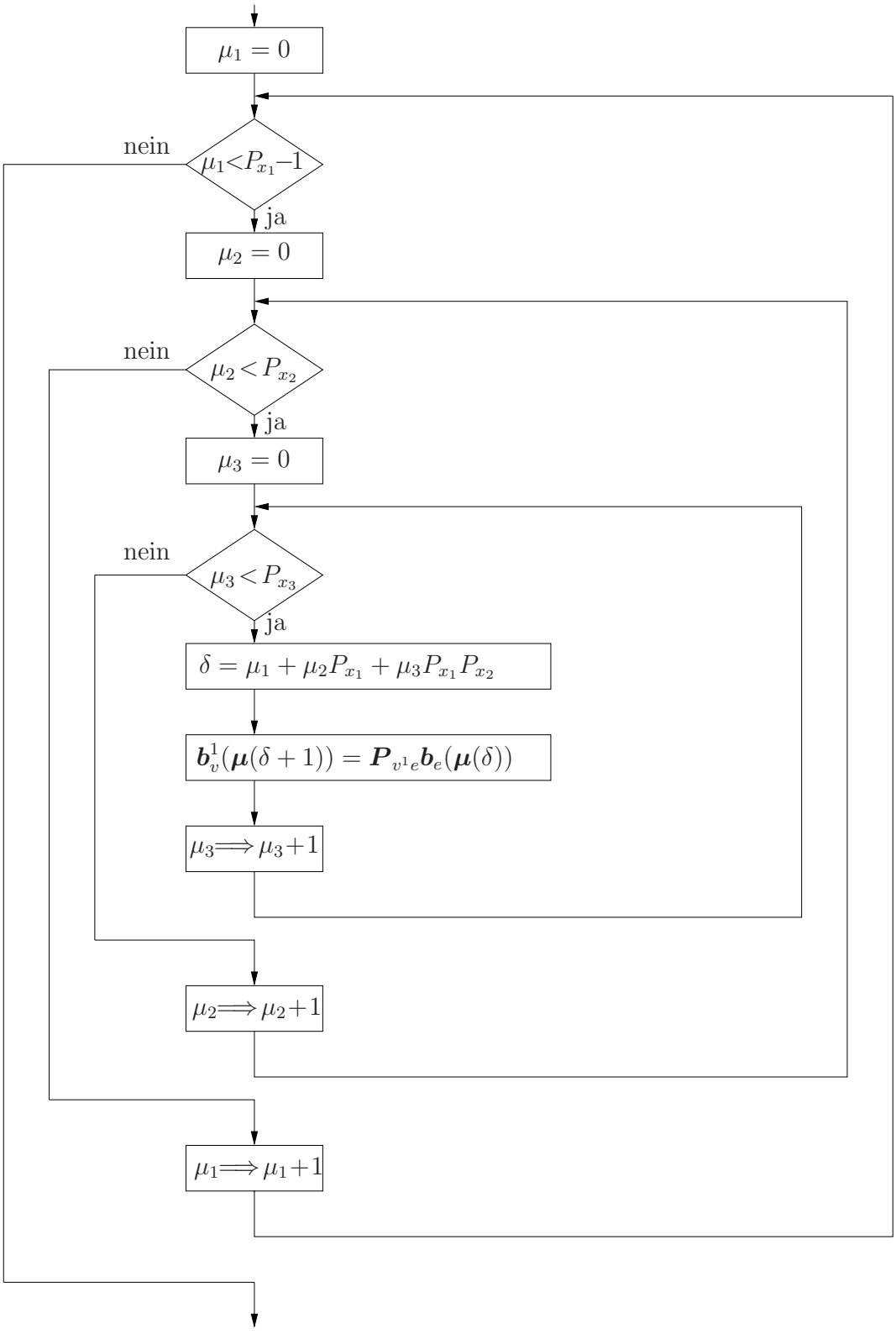


Bild 6.7: Berechnung der neuen Verzögererwerte innerhalb des Berechnungsgebietes (Gleichung (6.4))

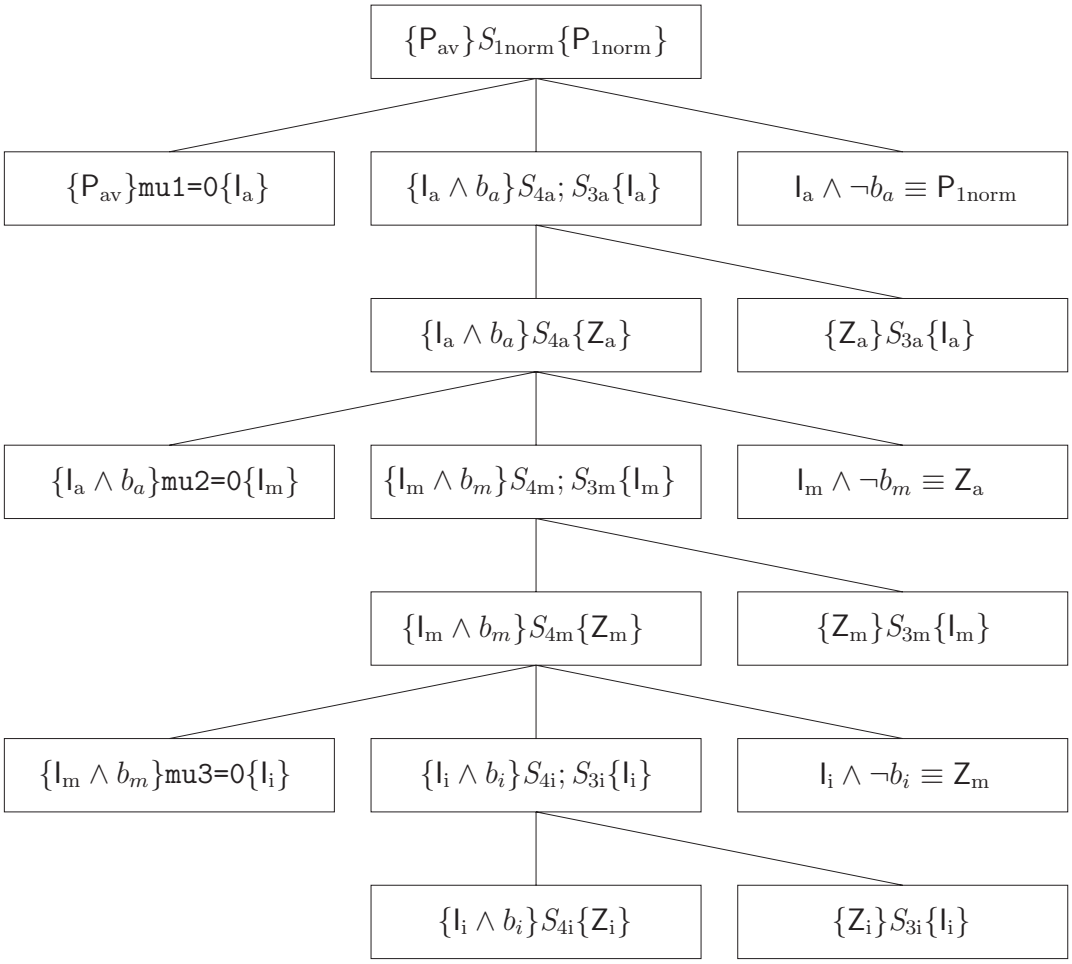


Bild 6.8: Berechnung der Verzögererwerte $\boldsymbol{b}_v^1$ im Normalfall

mit

$$\mathsf{P}_{\mathrm{av}} \equiv \mathsf{P}'_{\mathrm{aus}} \wedge \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}'_{\mathrm{av}} \wedge \mathsf{P}'_{\mathrm{be}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{dec}}\,. \tag{6.79}$$

ist eine gültige Schleifeninvariante der äußeren Schleife.

Beweis 5

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung liefert die Vorbedingung

$$\begin{aligned}\mathsf{P}_{\mathrm{av}} \equiv\ & 0 \le P_{x_1} - 1 \wedge \mathsf{P}_{\mathrm{av}} \wedge \bigwedge_{\mu_1=0}^{-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1{+}1, \mu_2, \mu_3, \delta(\boldsymbol{\mu}){+}\delta_1) \wedge \\ & 0 \le \mathrm{Max_i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int} \equiv \mathsf{P}_{\mathrm{av}}\,,\end{aligned} \tag{6.80}$$

da P_{av} die Bedingung $0 \le P_{x_1} - 1$ impliziert. Die Nachbedingung bestimmt sich (unter Beachtung von $P_{x_1} - 1 \le \mu_1.v \wedge 0 \le \mu_1.v \le P_{x_1} - 1 \equiv 0 \le \mu_1.v = P_{x_1} - 1$) zu

$$\mathsf{P}^1_{\mathrm{norm}} \equiv \mathsf{I}_a \wedge P_{x_1} - 1 \le \mu_1.v \equiv 0 \le \mu_1.v = P_{x_1} - 1 \wedge \mathsf{P}_{\mathrm{av}} \wedge \bigwedge_{\mu_1=0}^{P_{x_1}-2} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1{+}1, \mu_2, \mu_3, \delta{+}\delta_1) \tag{6.81}$$

und kann zu $\mathsf{P}_{\mathrm{av}} \wedge \mathsf{P}^1_{\mathrm{norm}}$ geschwächt werden. Die Zwischenbedingung bestimmt sich zu

$$\begin{aligned}\mathsf{Z}_{\mathrm{a}} \equiv\ & 0 \le \mu_1.v + 1 \le P_{x_1} - 1 \wedge \mathsf{P}_{\mathrm{av}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1{+}1, \mu_2, \mu_3, \delta(\boldsymbol{\mu}){+}\delta_1) \wedge \\ & \mu_1.v + 1 \le \mathrm{Max_i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int}\,.\end{aligned} \tag{6.82}$$

Stärkung um $0 \le \mu_1.v$ liefert

$$\mathsf{Z}'_{\mathrm{a}} \equiv \mathsf{Z}_{\mathrm{a}} \wedge 0 \le \mu_1.v \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2, \mu_3]^{\mathrm{T}}){+}\delta_1)\,. \tag{6.83}$$

Der Beweis von $\{\mathsf{I}_{\mathrm{a}} \wedge b_a\} S_{4\mathrm{a}} \{\mathsf{Z}'_{\mathrm{a}}\}$ wird gemäß Axiom (5.35) durchgeführt, da $S_{4\mathrm{a}}$ die mittlere Schleife darstellt.

Satz 6 *Der Ausdruck*

$$\mathsf{I}_{\mathrm{m}} \equiv 0 \le \mu_2.v \le P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2, \mu_3]^{\mathrm{T}}){+}\delta_1) \tag{6.84}$$

ist eine gültige Schleifeninvariante der mittleren Schleife.

Beweis 6

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu2=0` liefert

$$\begin{aligned} &0 \leq P_{x_2} \wedge \mathsf{I}_\mathrm{a} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v{+}1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2, \mu_3]^\mathrm{T}){+}\delta_1) \wedge \\ &0 \leq \mathrm{Max_i} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int} \equiv \mathsf{I}_\mathrm{a} \wedge b_a\,, \end{aligned} \tag{6.85}$$

da I_a die Aussage $0 \leq P_{x_2}$ impliziert, sodass wir, wie zu zeigen ist, die Vorbedingung von $S_{4\mathrm{a}}$ erhalten. Die Nachbedingung bestimmt sich zu

$$\mathsf{I}_\mathrm{m} \wedge P_{x_2} \leq \mu_2.v \equiv 0 \leq \mu_2.v = P_{x_2} \wedge \mathsf{I}_\mathrm{a} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v{+}1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2, \mu_3]^\mathrm{T}){+}\delta_1) \tag{6.86}$$

und stellt nach Abtrennung von $0 \leq \mu_2.v = P_{x_2}$, wie behauptet, Z'_a aus Gleichung (6.83) dar. Die Zwischenbedingung bestimmt sich zu

$$\begin{aligned} \mathsf{Z}_\mathrm{m} \equiv \quad & 0 \leq \mu_2.v + 1 \leq P_{x_2} \wedge \mathsf{I}_\mathrm{a} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v+1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2, \mu_3]^\mathrm{T}){+}\delta_1) \\ & \wedge \mu_2.v + 1 \leq \mathrm{Max_i} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int}\,. \end{aligned} \tag{6.87}$$

Stärkung um $0 \leq \mu_2.v$ und Vereinfachung liefert

$$\mathsf{Z}'_\mathrm{m} \equiv \mathsf{Z}_\mathrm{m} \wedge 0 \leq \mu_2.v \equiv \mathsf{I}_\mathrm{m} \wedge b_m \wedge \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v+1, \mu_2.v, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1)\,. \tag{6.88}$$

Der Beweis von $\{\mathsf{I}_\mathrm{m} \wedge b_m\} S_{4\mathrm{m}} \{\mathsf{Z}'_\mathrm{m}\}$ wird gemäß Axiom (5.35) durchgeführt, da $S_{4\mathrm{m}}$ die innere Schleife darstellt.

Satz 7 *Der Ausdruck*

$$\mathsf{I}_\mathrm{i} \equiv 0 \leq \mu_3.v \leq P_{x_3} \wedge \mathsf{I}_\mathrm{m} \wedge b_m \wedge \bigwedge_{\mu_3=0}^{\mu_3.v-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v{+}1, \mu_2.v, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1)\,. \tag{6.89}$$

ist eine gültige Schleifeninvariante der inneren Schleife.

Beweis 7

Zunächst bemerken wir $\mathsf{I}_\mathrm{m} \wedge b_m \Rightarrow 0 \leq \mu_2.v \leq P_{x_2} - 1$. Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu3=0` liefert

$$\begin{aligned} &0 \leq P_{x_3} \wedge \mathsf{I}_\mathrm{m} \wedge b_m \wedge \bigwedge_{\mu_3=0}^{-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}_{bk}^1(\mu_1.v{+}1, \mu_2.v, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1) \wedge \\ &0 \leq \mathrm{Max_i} \wedge \mu_3.l = -1 \wedge \mu_3.typ = \texttt{int} \equiv \mathsf{I}_\mathrm{m} \wedge b_m\,. \end{aligned} \tag{6.90}$$

Die Nachbedingung der inneren Schleife lautet

$$\mathsf{I}_\mathrm{i} \wedge P_{x_3} \leq \mu_3.v \equiv 0 \leq \mu_3.v = P_{x_3} \wedge \mathsf{I}_\mathrm{m} \wedge b_m \wedge \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2.v, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1)\,. \tag{6.91}$$

Einsetzen von I_m aus Gleichung (6.84) und b_m liefert

$$0 \leq \mu_3.v{=}P_{x_3} \wedge 0 \leq \mu_2.v{+}1 \leq P_{x_2} \wedge \mathsf{I}_\mathrm{a} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v} \bigwedge_{\mu_3=0}^{P_{x_3}-1} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1) \tag{6.92}$$

und ist wegen $\mathsf{P}_{x_3} > 0$ die Bedingung Z_m aus Gleichung (6.87). Die Zwischenbedingung der inneren Schleife lautet

$$\begin{aligned}\mathsf{Z}_\mathrm{i} \equiv\ & 0 \leq \mu_3.v + 1 \leq P_{x_3} \wedge \mathsf{I}_\mathrm{m} \wedge b_m \wedge \bigwedge_{\mu_3=0}^{\mu_3.v} \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2.v, \mu_3, \delta([\mu_1.v, \mu_2.v, \mu_3]^\mathrm{T}){+}\delta_1) \wedge \\ & \mu_3.v + 1 \leq \mathrm{Max}_\mathrm{i} \wedge \mu_3.l = -1 \wedge \mu_3.typ = \texttt{int}\,.\end{aligned} \tag{6.93}$$

Stärkung um $0 \leq \mu_3.v$ und Vereinfachung liefert

$$\mathsf{Z}'_\mathrm{i} \equiv \mathsf{Z}_\mathrm{i} \wedge 0 \leq \mu_3.v \equiv \mathsf{I}_\mathrm{i} \wedge b_i \wedge \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta([\mu_1.v, \mu_2.v, \mu_3.v]^\mathrm{T}){+}\delta_1)\,. \tag{6.94}$$

Es verbleibt die Richtigkeit von $\{\mathsf{I}_\mathrm{i} \wedge b_i\} S_{4\mathrm{i}} \{\mathsf{Z}'_\mathrm{i}\}$ zu zeigen. Zu diesem Zweck bereiten wir Z'_i weiter auf

$$\mathsf{Z}'_\mathrm{i} \equiv \mathsf{I}_\mathrm{i} \wedge b_i \wedge \bigwedge_{k=1}^{n_v^1-1} \mathsf{P}^1_{bk}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta(\boldsymbol{\mu}.v){+}\delta_1) \wedge \mathsf{P}^1_{bn_v^1}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta(\boldsymbol{\mu}.v){+}\delta_1)\,. \tag{6.95}$$

Die Vorbedingung von $S_{b_{vn_v^1}}$ erhalten wir durch Anwenden von Gleichung (5.43) und Stärkung zu

$$\begin{aligned}&\mathsf{I}_\mathrm{i} \wedge b_i \wedge \mathsf{P}^1_{bn_v^1}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta(\boldsymbol{\mu}.v){+}\delta_1) \begin{bmatrix} b^1_{v\,n_v^1}(\delta.v{+}\delta_1).v \\ \boldsymbol{p}^1_{e\,n_v^1,\bullet}\, \boldsymbol{b}_{e\,(\delta.v).v} \end{bmatrix} \wedge \\ &\bigwedge_{k=1}^{n_v^1-1} \mathsf{P}^1_{b\,k}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta(\boldsymbol{\mu}.v){+}\delta_1) \wedge \bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| \leq \mathrm{Max}_\mathrm{f} \wedge \\ &0 \leq \delta.v + \delta_1 \leq b^1_{v\,n_v^1}.l - 1 \wedge \bigwedge_{l=1}^{n_e} 0 \leq \delta.v \leq b_{e\,l}.l - 1 \wedge \bigwedge_{k=1}^{n_e} b^1_{v\,n_v^1}.typ = b_{e\,k}.typ = \texttt{float} \wedge \\ &0 \leq \delta.v + \delta_1 \leq \mathrm{Max}_\mathrm{i} \wedge 0 \leq \delta.v \leq \mathrm{Max}_\mathrm{i} \wedge \delta.l = -1 \wedge \delta.typ = \texttt{int}\,.\end{aligned} \tag{6.96}$$

Hierbei wurde berücksichtigt, dass $n_e - 1$ Elemente des Zeilenvektors $\boldsymbol{p}^\kappa_{ve\,\mu,\bullet}$ den Wert null haben und das verbleibende Element den Wert eins hat. Weitere Vereinfachungen können wegen $\mathsf{I}_\mathrm{i} \Rightarrow \delta.typ =$

$\texttt{int} \wedge \delta.l = -1$ vorgenommen werden, indem die in I_i enthaltenen Prädikate abgetrennt werden. Gleiches Vorgehen für die Anweisungen $S_{b_{v n_v^1 - 1}}, \ldots, S_{b_{v2}}, S_{b_{v1}}$ führt zur Vorbedingung von $S_{b_{v1}}$

$$\mathsf{I}_\mathrm{i} \wedge b_i \wedge \bigwedge_{k=1}^{n_v^1} \mathsf{P}^1_{b\,k}(\mu_1.v{+}1, \mu_2.v, \mu_3.v, \delta(\boldsymbol{\mu}.v){+}\delta_1) \begin{bmatrix} b^1_{v\,k}(\delta.v+\delta_1).v \\ \boldsymbol{p}^1_{e\,k,\bullet}\, \boldsymbol{b}_e\,(\delta.v).v \end{bmatrix} \wedge \bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| \leq \mathrm{Max_f} \wedge$$
$$\bigwedge_{l=1}^{n_v^1} \bigwedge_{k=1}^{n_e} b^1_{v\,l}.typ = b_{e\,k}.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_v^1} 0 \leq \delta.v + \delta_1 \leq b^1_{v\,k}.l - 1 \wedge \bigwedge_{k=1}^{n_e} 0 \leq \delta.v \leq b_{e\,k}.l - 1 \wedge$$
$$0 \leq \delta.v + \delta_1 \leq \mathrm{Max_i} \wedge 0 \leq \delta.v \leq \mathrm{Max_i}\,. \tag{6.97}$$

Diese Vorbedingung stärken wir durch $\delta.v = \delta(\boldsymbol{\mu}.v)$ und erhalten durch Auswertung und Nutzung von Gleichung (6.25)

$$\mathsf{I}_\mathrm{i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \bigwedge_{k=1}^{n_v^1} \boldsymbol{p}^1_{e\,k,\bullet}\, \boldsymbol{b}_e\,(\delta.v).v = b^1_{v\,k}(\boldsymbol{\mu}.v + \boldsymbol{h}_1) \wedge \bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| \leq \mathrm{Max_f} \wedge$$
$$\bigwedge_{l=1}^{n_v^1} \bigwedge_{k=1}^{n_e} b^1_{v\,l}.typ = b_{e\,k}.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_v^1} 0 \leq \delta.v + \delta_1 \leq b^1_{v\,k}.l - 1 \wedge \bigwedge_{k=1}^{n_e} 0 \leq \delta.v \leq b_{e\,k}.l - 1 \wedge$$
$$0 \leq \delta.v + \delta_1 \leq \mathrm{Max_i} \wedge 0 \leq \delta.v \leq \mathrm{Max_i}\,. \tag{6.98}$$

Wir werden zunächst zeigen, dass I_i die in Gleichung (6.98) auftretenden Aussagen bzgl. der Datentypen impliziert. Die Verwendung von Gleichung (6.89) i.V.m. Gleichung (6.33), Gleichung (6.34), Gleichung (6.78), Gleichung (6.79) und Gleichung (6.84), liefert

$$\mathsf{I}_\mathrm{i} \Rightarrow \bigwedge_{k=1}^{n_v^1} b^1_{v\,k}.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_e} b_{e\,k}.typ = \texttt{float}\,. \tag{6.99}$$

Alle Datentypen sind somit gleich. Insbesondere folgt daraus -wie behauptet- die Richtigkeit von

$$\mathsf{I}_\mathrm{i} \Rightarrow \bigwedge_{l=1}^{n_v^1} \bigwedge_{k=1}^{n_e} b^1_{v\,l}.typ = b_{e\,k}.typ\,, \tag{6.100}$$

sodass die Forderungen an die Datentypen aus Gleichung (6.97) abgetrennt werden können. Wie im Folgenden gezeigt wird, impliziert I_i auch die in Gleichung (6.98) auftretenden Ausdrücke bzgl. der Feldlängen.Um dies zu zeigen verwenden wir erneut Gleichung (6.89) i.V.m. Gleichung (6.33), Gleichung (6.34), Gleichung (6.78), Gleichung (6.79) und Gleichung (6.84). Wir erhalten

$$\mathsf{I}_\mathrm{i} \wedge b_i \;\Rightarrow\; 0 \leq \mu_1.v \leq P_{x_1} - 2 \;\wedge\; 0 \leq \mu_2.v \leq P_{x_2} - 1 \;\wedge\; 0 \leq \mu_3.v \leq P_{x_3} - 1\,. \tag{6.101}$$

Wegen Gleichung (6.40) folgt für $\delta.v$ die Beziehung

$$\mathsf{I}_\mathrm{i} \wedge b_i \Rightarrow 0 \leq \delta.v \leq P - 2\,. \tag{6.102}$$

Durch mehrfache Anwendung des Kettenschlusses und Gleichung (6.34) folgen die Implikationen

$$\mathsf{I}_\mathrm{i} \Rightarrow \bigwedge_{k=1}^{n_v^1} P \leq b^1_{v\,k}.l \leq \mathrm{Max_i}\,,\; \mathsf{I}_\mathrm{i} \Rightarrow \bigwedge_{k=1}^{n_e} P \leq b_{e\,k}.l \leq \mathrm{Max_i}\,. \tag{6.103}$$

Konjunktive Verknüpfung mit 6.102 liefert

$$\mathsf{I_i} \wedge b_i \Rightarrow \bigwedge_{k=1}^{n_v^1} 0 \le \delta.v \le b_{v\,k}^1.l - 2\ ,\ \mathsf{I_i} \wedge b_i \Rightarrow \bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 2\,. \tag{6.104}$$

Anschaulich bedeutet diese Aussage, dass der Wert der Variablen δ innerhalb eines Intervalls liegt. Nun betrachten wir die folgenden -offenbar richtigen- Aussagen

$$\begin{gathered}\left[\bigwedge_{k=1}^{n_v^1} 1 \le \delta.v + 1 \le b_{v\,k}^1.l - 1\right] \Rightarrow \left[\bigwedge_{k=1}^{n_v^1} 0 \le \delta.v + 1 \le b_{v\,k}^1.l - 1\right], \\ \left[\bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 2\right] \Rightarrow \left[\bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 1\right]\ .\end{gathered} \tag{6.105}$$

Offenbar sind auch die Aussagen bzgl. der Feldlänge in Gleichung (6.98) richtig und werden abgetrennt. Aufgrund von $\mathsf{I_i} \Rightarrow b_{e\,k}.l \le \mathrm{Max_i}$ folgt aus $\mathsf{I_i}$ auch $\mathsf{I_i} \Rightarrow 0 \le \delta.v + \delta_1 \le \mathrm{Max_i} \wedge 0 \le \delta.v \le \mathrm{Max_i}$, was ebenfalls in Gleichung (6.98) abgetrennt werden kann.

Da $\mathsf{I_i}$ die Beziehung $\bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| \le \mathrm{Max_f}$ impliziert, verbleibt von Gleichung (6.98) noch

$$\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \bigwedge_{k=1}^{n_v^1} \boldsymbol{p}_{e\,k,\bullet}^1\, \boldsymbol{b}_e(\delta.v).v = b_{v\,k}^1(\boldsymbol{\mu}.v + \boldsymbol{h}_1)\ . \tag{6.106}$$

Hierin ersetzen wir $\delta.v$ durch $\delta(\boldsymbol{\mu}.v)$. Zudem werden wir im Folgenden die Istwerte durch die Sollwerte ersetzen. Dazu ermitteln wir zunächst aus Gleichung (6.89), Gleichung (6.84), Gleichung (6.78), $\mathsf{P_{av}}$, $\mathsf{P_e}$ und $\mathsf{P'_{be}}$

$$\mathsf{I_i} \Rightarrow \bigwedge_{\delta=0}^{P-1} \boldsymbol{b}_e(\delta).v = \boldsymbol{b}_e(\boldsymbol{\mu}(\delta))\ . \tag{6.107}$$

Da diese Gleichung für alle δ des Berechnungsgebietes gilt, gilt sie auch für $\delta.v = \delta(\boldsymbol{\mu}.v)$, d. h.

$$\mathsf{I_i} \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \Rightarrow \boldsymbol{b}_e(\delta.v).v = \boldsymbol{b}_e(\boldsymbol{\mu}(\delta.v))\ . \tag{6.108}$$

Somit lautet die um $\delta.v = \delta(\boldsymbol{\mu}.v)$ gestärkte Vorbedingung von S_{bv1}

$$\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \bigwedge_{k=1}^{n_v^1} \boldsymbol{p}_{e\,k,\bullet}^1\, \boldsymbol{b}_e(\boldsymbol{\mu}(\delta.v)) = b_{v\,k}^1(\boldsymbol{\mu}.v + \boldsymbol{h}_1)\,. \tag{6.109}$$

Mit $\delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mathsf{P_{MDED}} \equiv \boldsymbol{\mu}(\delta.v) = \boldsymbol{\mu}.v \wedge \mathsf{P_{MDED}}$ haben wir

$$\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \bigwedge_{k=1}^{n_v^1} \boldsymbol{p}_{e\,k,\bullet}^1\, \boldsymbol{b}_e(\boldsymbol{\mu}.v) = b_{v\,k}^1(\boldsymbol{\mu}.v + \boldsymbol{h}_1)\,. \tag{6.110}$$

Mit $\mathsf{I_i} \Rightarrow \mathsf{P_{WDF}}$ folgt die Wahrheit der letzten Aussage. Durch Abtrennung der wahren Aussage erhalten wir die Vorbedingung von S_{bv1} letztendlich zu $\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v)$. Die Vorbedingung von `delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;` lautet mit Gleichung (6.41)

$$\begin{gathered}\mathsf{I_i} \wedge b_i \wedge \mu_1.v + \mu_2.v P_{x_1} + \mu_3.v P_{x_1} P_{x_2} = \delta(\boldsymbol{\mu}.v) \wedge \\ 0 \le \mu_1.v \le P_{x_1} - 1 \wedge 0 \le \mu_2.v \le P_{x_2} - 1 \wedge 0 \le \mu_3.v \le P_{x_3} - 1\,.\end{gathered} \tag{6.111}$$

Wegen $\mathsf{I}_\mathrm{i} \Rightarrow \mathsf{P}_\mathrm{MDED} \wedge \mathsf{P}_\mathrm{decuv}$ und $\mathsf{I}_\mathrm{i} \wedge b_i \Longrightarrow 0 \le \mu_1.v \le P_{x_1} - 1 \wedge 0 \le \mu_2.v \le P_{x_2} - 1 \wedge 0 \le \mu_3.v \le P_{x_3} - 1$ folgt schließlich, wie behauptet,

$$\mathsf{I}_\mathrm{i} \wedge b_i \ . \tag{6.112}$$

Die anderen 6 Anweisungen wollen wir nicht beweisen, da sie ähnlich sind. Die Unterschiede in den Nachbedingungen liegen darin, dass jeweils $\mathsf{P}^{\kappa}_{\mathrm{norm}}$ hinzukommt.

6.4 Berechnung der Eingangswellen der Verzögerer

Die Nachbedingung von S_av lautet

$$\mathsf{P}_\mathrm{av} \equiv \mathsf{P}_\mathrm{WDF} \ \wedge \mathsf{P}_\mathrm{dec} \ \wedge \mathsf{P}_\mathrm{q} \ \wedge \mathsf{P}'_\mathrm{be} \ \wedge \mathsf{P}'_\mathrm{aus} \ \wedge \mathsf{P}'_\mathrm{av} \tag{6.113}$$

und beinhaltet die korrekt berechneten Werte der Wellengrößen $\boldsymbol{a}_v$ am Rand. Diese Berechnung leistet die Anweisung S_av. Wie zuvor erfolgt zunächst eine Zerlegung S_av in sequentielle Teilanweisungen nach $S_\mathrm{av} \equiv S_\mathrm{avx}; S_\mathrm{avy}; S_\mathrm{avz};$. Dazu ist die Übersicht

Bedingung	Anweisung
$\{\mathsf{P}_\mathrm{aus}\} \equiv \mathsf{P}_\mathrm{WDF} \wedge \mathsf{P}_\mathrm{dec} \wedge \mathsf{P}_\mathrm{q} \wedge \mathsf{P}'_\mathrm{be} \wedge \mathsf{P}'_\mathrm{aus}$	
	S_avx
$\{\mathsf{P}_\mathrm{avx}\} \equiv \mathsf{P}_\mathrm{WDF} \wedge \mathsf{P}_\mathrm{dec} \wedge \mathsf{P}_\mathrm{q} \wedge \mathsf{P}'_\mathrm{be} \wedge \mathsf{P}'_\mathrm{aus} \wedge \mathsf{P}'_\mathrm{avx}$	
	S_avy
$\{\mathsf{P}_\mathrm{avy}\} \equiv \mathsf{P}_\mathrm{WDF} \wedge \mathsf{P}_\mathrm{dec} \wedge \mathsf{P}_\mathrm{q} \wedge \mathsf{P}'_\mathrm{be} \wedge \mathsf{P}'_\mathrm{aus} \wedge \mathsf{P}'_\mathrm{avx} \wedge \mathsf{P}'_\mathrm{avy}$	
	S_avz
$\{\mathsf{P}_\mathrm{av}\} \equiv \mathsf{P}_\mathrm{WDF} \wedge \mathsf{P}_\mathrm{dec} \wedge \mathsf{P}_\mathrm{q} \wedge \mathsf{P}'_\mathrm{be} \wedge \mathsf{P}'_\mathrm{aus} \wedge \mathsf{P}'_\mathrm{avx} \wedge \mathsf{P}'_\mathrm{avy} \wedge \mathsf{P}'_\mathrm{avz}$	

hilfreich. Die Anweisungen S_avx, S_avy und S_avz sind jeweils zweifach verschachtelte for-Schleifen in den Variablen μ_1, μ_2 und μ_3, von denen jeweils eine konstant ist. Die verbliebenen zwei unabhängigen Variablen bilden dann eine Grenzfläche. Innerhalb der Schleifen wird Gleichung (6.3) ausgewertet. Der zugehörige Programmablaufplan für S_avz ist im Bild 6.9 dargestellt. Der C-Code für S_avz lautet

Bedingung	Anweisung	Kommentar
$\{\mathsf{P}_\mathrm{avy}\}$		
	`for(mu1 = 0 ; mu1 <` P_{x_1} `; mu1=mu1+1){`	
	`  for(mu2 = 0 ; mu2 <` P_{x_2} `; mu2=mu2+1){`	
	`  mu3 = 0;`	
	`  delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;`	
S_av6		Wellen $\boldsymbol{a}_{v6}$ berechnen
	`  mu3 =` P_{x_3} `- 1;`	
	`  delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;`	
S_av3		Wellen $\boldsymbol{a}_{v3}$ berechnen
	`  }`	
	`}`	
$\{\mathsf{P}_\mathrm{avz}\}$		

wobei die Berechnungsanweisungen durch

$S_\mathrm{av1} \equiv$ `a_v_1_1 [delta]=` $p_{v^1 e\, 11}$ `*b_e_1[delta]+...+` $p_{v^1 e\, 1 n_e}$ `*b_e_`n_e`[delta];`

$\vdots$

`a_v_1_`n_v^1`[delta]=` $p_{v^1 e\, n_v^1 1}$`*b_e_1[delta]+...+` $p_{v^1 e\, n_v^1 n_e}$`*b_e_`n_e`[delta];`

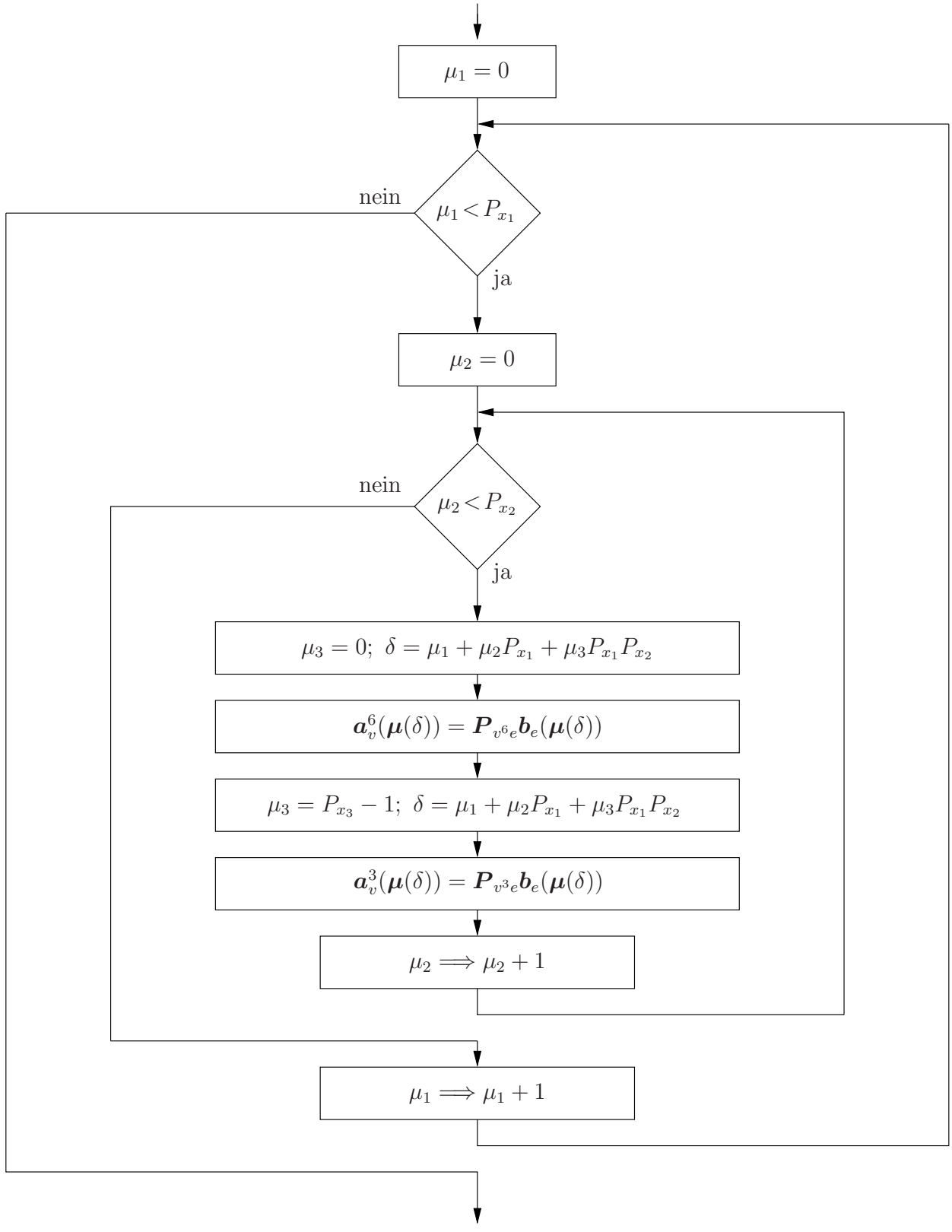


Bild 6.9: Berechnung der Wellen, die in der nächsten Abtastschicht durch die Grenzflächen $\mu_3 = 0$ und $\mu_3 = P_{x_3} - 1$ das Gebiet $\mathcal{G}$ verlassen (Gleichung (6.3) für $\kappa = 3, 6$)

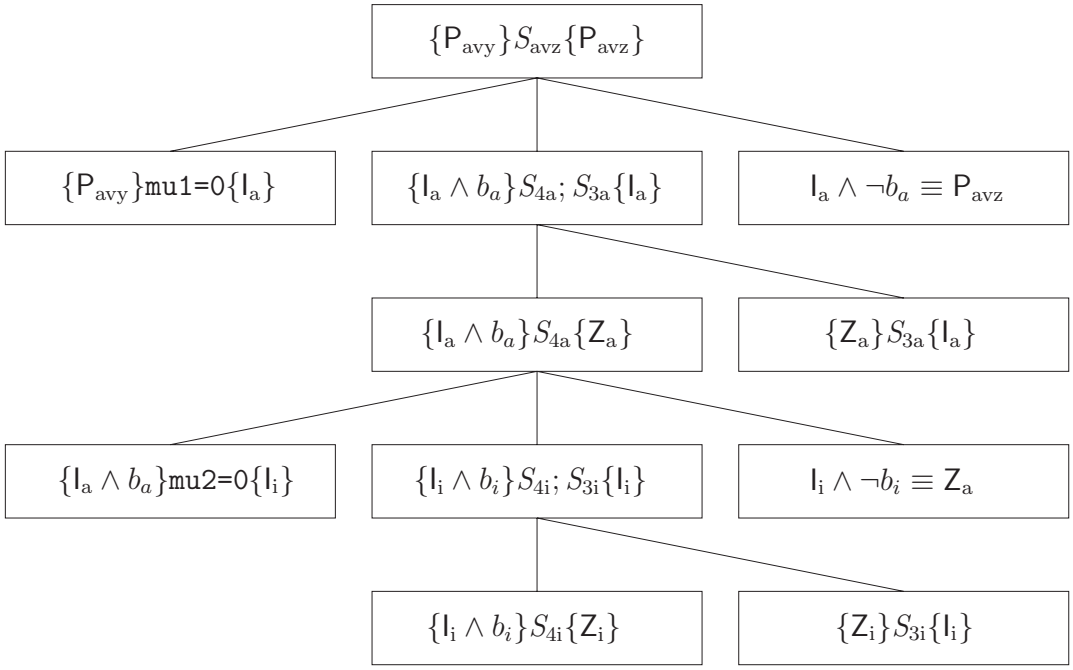


Bild 6.10: Berechnung der Eingangswellen der Verzögerer

bis

$$\begin{aligned} S_{\mathrm{av6}} \equiv\ & \texttt{a\_v\_6\_1 [delta]=}\ p_{v^6 e\,11}\ \texttt{*b\_e\_1[delta]+...+}p_{v^6 e\,1 n_e}\ \texttt{*b\_e\_}n_e\texttt{[delta];} \\ & \vdots \\ & \texttt{a\_v\_6\_}n_v^6\texttt{[delta]=}\ p_{v^6 e\, n_v^6 1}\texttt{*b\_e\_1[delta]+...+}p_{v^6 e\, n_v^6 n_e}\texttt{*b\_e\_}n_e\texttt{[delta];} \end{aligned}$$

gegeben sind. Zur Ermittlung der Vorbedingung von S_{avz} ist der Überblick im Bild 6.10 nützlich. Hierin ist $b_a \equiv \mu_1.v < P_{x_1} \equiv \mu_1.v + 1 \leq P_{x_1}$ und $b_i \equiv \mu_2.v < P_{x_2} \equiv \mu_2.v + 1 \leq P_{x_2}$. Wir werden nun die Beweisführung für S_{avz} vornehmen und beginnen mit der äußeren Schleife.

Satz 8 *Der Ausdruck*

$$\mathsf{I}_{\mathrm{a}} \equiv 0 \leq \mu_1.v \leq P_{x_1} \wedge \mathsf{P}_{\mathrm{avy}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1, \mu_2) \tag{6.114}$$

ist eine gültige Schleifeninvariante der äußeren Schleife.

Beweis 8

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu1=0` liefert

$$0 \leq P_{x1} \wedge \mathsf{P}_{\mathrm{avy}} \wedge \bigwedge_{\mu_1=0}^{-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1, \mu_2) \wedge 0 \leq \mathrm{Max_i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int} \equiv \mathsf{P}_{\mathrm{avy}}\,. \tag{6.115}$$

Die Nachbedingung bestimmt sich zu

$$\mathsf{I}_{\mathrm{a}} \wedge \mu_1.v \geq P_{x_1} \equiv \mu_1.v = P_{x_1} \wedge \mathsf{P}_{\mathrm{avy}} \wedge \bigwedge_{\mu_1=0}^{P_{x_1}-1} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}\,(\mu_1, \mu_2) \tag{6.116}$$

und kann um $\mu_1.v = P_{x_1}$ zu $\mathsf{P}_{\mathrm{avz}}$ geschwächt werden. Die Zwischenbedingung bestimmt sich zu

$$\begin{aligned}\mathsf{Z}_{\mathrm{a}} \equiv\ & 0 \leq \mu_1.v + 1 \leq P_{x_1} \wedge \mathsf{P}_{\mathrm{avy}} \wedge \bigwedge_{\mu_1=0}^{\mu_1.v} \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1,\mu_2) \wedge \\ & \mu_1.v + 1 \leq \mathrm{Max_i} \wedge \mu_1.l = -1 \wedge \mu_1.typ = \texttt{int}.\end{aligned} \tag{6.117}$$

Stärkung um $0 \leq \mu_1.v$ liefert

$$\mathsf{Z}'_{\mathrm{a}} \equiv \mathsf{Z}_{\mathrm{a}} \wedge 0 \leq \mu_1.v \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2)\,. \tag{6.118}$$

Der Beweis von $\{\mathsf{I}_{\mathrm{a}} \wedge b_a\} S_{4\mathrm{a}} \{\mathsf{Z}_{\mathrm{a}}\}$ wird wiederum gemäß Axiom (5.35) durchgeführt, da $S_{4\mathrm{a}}$ die innere Schleife darstellt.

Satz 9 *Der Ausdruck*

$$\mathsf{I}_{\mathrm{i}} \equiv 0 \leq \mu_2.v \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v-1} \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2) \tag{6.119}$$

ist eine gültige Schleifeninvariante der inneren Schleife.

Beweis 9

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `mu2=0` liefert

$$0 \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{-1} \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2) \wedge \mu_2.v + 1 \leq \mathrm{Max_i} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int} \equiv \mathsf{I}_{\mathrm{a}} \wedge b_a\,. \tag{6.120}$$

Die Nachbedingung der inneren Schleife lautet

$$\mathsf{I}_{\mathrm{i}} \wedge \mu_2.v \geq P_{x_2} \equiv \mu_2.v = P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{P_{x_2}-1} \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2) \tag{6.121}$$

und ist nach Schwächung um $\mu_2.v = P_{x_2}$, wie behauptet, Z'_{a} aus Gleichung (6.118). Die Zwischenbedingung der inneren Schleife lautet

$$\mathsf{Z}_{\mathrm{i}} \equiv 0 \leq \mu_2.v+1 \leq P_{x_2} \wedge \mathsf{I}_{\mathrm{a}} \wedge b_a \wedge \bigwedge_{\mu_2=0}^{\mu_2.v} \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2) \wedge \mu_2.v+1 \leq \mathrm{Max_i} \wedge \mu_2.l = -1 \wedge \mu_2.typ = \texttt{int}\,. \tag{6.122}$$

Stärkung um $0 \leq \mu_2.v$ liefert

$$\mathsf{Z}'_{\mathrm{i}} \equiv \mathsf{Z}_{\mathrm{i}} \wedge 0 \leq \mu_2.v \equiv \mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \mathsf{P}_{\mathrm{az}}(\mu_1.v,\mu_2.v)\,. \tag{6.123}$$

Es verbleibt die Richtigkeit von $\{\mathsf{I}_{\mathrm{i}} \wedge b_i\} S_{4\mathrm{i}} \{\mathsf{Z}'_{\mathrm{i}}\}$ zu zeigen. Dazu zerlegen wir Z'_{i} wie folgt

$$\begin{aligned}\mathsf{Z}'_{\mathrm{i}} \equiv\ & \mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \boldsymbol{a}^6_v(\delta([\mu_1.v,\mu_2.v,0]^{\mathrm{T}})).v = \boldsymbol{a}^6_v([\mu_1.v,\mu_2.v,0]^{\mathrm{T}}) \wedge \\ & \boldsymbol{a}^3_v(\delta([\mu_1.v,\mu_2.v,P_{x_3}-1]^{\mathrm{T}})).v = \boldsymbol{a}^3_v([\mu_1.v,\mu_2.v,P_{x_3}-1]^{\mathrm{T}})\,.\end{aligned} \tag{6.124}$$

Zudem wird die Nachbedingung von S_{av3} durch

$$\delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \equiv \delta.v = \delta([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}} \wedge \mu_3.v = P_{x_3} - 1 \tag{6.125}$$

gestärkt. Für die gestärkte Bedingung Z'_{i} notieren wir

$$\begin{aligned} &\mathsf{Z}'_{\mathrm{i}} \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \equiv \mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \wedge \\ &\boldsymbol{a}_v^6\left(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})\right).v = \boldsymbol{a}_v^6\left([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}\right) \wedge \boldsymbol{a}_v^3(\delta.v).v = \boldsymbol{a}_v^3\left([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}}\right). \end{aligned} \tag{6.126}$$

Nach diesen Vorbereitungen gewinnen wir die Vorbedingung von S_{av3} zu

$$\begin{aligned} &\mathsf{Z}'_{\mathrm{i}} \begin{bmatrix} \boldsymbol{a}_v^{3(\delta.v).v} \\ \boldsymbol{P}_{v^3 e}\, \boldsymbol{b}_{e\,(\delta.v).v} \end{bmatrix} \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3 = P_{x_3} - 1 \wedge \bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| < \mathrm{Max_f} \wedge \\ &\bigwedge_{k=1}^{n_v^3} 0 \le \delta.v \le a_{v\,k}^3.l - 1 \wedge \bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 1 \wedge \bigwedge_{l=1}^{n_v^3} \bigwedge_{k=1}^{n_e} a_{v\,l}^3.typ = b_{e\,k}.typ \wedge \\ &\delta.v \le \mathrm{Max_i} \wedge \delta.l = -1 \wedge \delta.typ = \texttt{int}\ . \end{aligned} \tag{6.127}$$

Zur Verdichtung des obigen Prädikats werden wir zeigen, dass $\mathsf{I}_{\mathrm{i}} \wedge b_i$ die Prädikate bzgl. der Datentypen und der Feldlänge impliziert.

Durch Gleichung (6.119) , Gleichung (6.114) , Gleichung (6.113), Gleichung (6.34) und Gleichung (6.33) erhalten wir für die Datentypen

$$\mathsf{I}_{\mathrm{i}} \wedge b_i \Rightarrow \bigwedge_{k=1}^{n_v^3} a_{v\,k}^3.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_e} b_{e\,k}.typ = \texttt{float}\ . \tag{6.128}$$

Mit dem Prädikatenkalkül folgt wie behauptet

$$\mathsf{I}_{\mathrm{i}} \Rightarrow \bigwedge_{l=1}^{n_v^3} \bigwedge_{k=1}^{n_e} a_{v\,l}^3.typ = b_{e\,k}.typ\ . \tag{6.129}$$

Zudem gilt $\mathsf{I}_{\mathrm{i}} \Rightarrow \delta.typ = \texttt{int} \wedge \delta.l = -1$, so dass die Forderungen bzgl. der Datentypen aus Gleichung (6.127) abgetrennt werden können. Nun werden wir zeigen, dass die Bedingungen an die Feldlängen abgetrennt werden können. Einerseits folgt aus Gleichung (6.119) , Gleichung (6.114) , Gleichung (6.113), Gleichung (6.34) und Gleichung (6.33)

$$\mathsf{I}_{\mathrm{i}} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \Rightarrow \bigwedge_{k=1}^{n_v^3} P - 1 \le a_{v\,k}^3.l - 1 \wedge \bigwedge_{k=1}^{n_e} P - 1 \le b_{e\,k}.l - 1 \tag{6.130}$$

und andererseits gilt wegen $\mathsf{Z}_{\mathrm{i}} \Rightarrow \mathsf{P}_{\mathrm{MDED}}$ die Beziehung

$$\mathsf{Z}'_{\mathrm{i}} \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \Rightarrow 0 \le \delta.v \le P - 1\ . \tag{6.131}$$

Somit erhalten wir

$$\mathsf{I}_{\mathrm{i}} \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \Rightarrow \bigwedge_{k=1}^{n_v^3} 0 \le \delta.v \le a_{v\,k}^3.l - 1 \wedge \bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 1\ . \tag{6.132}$$

Dies ist der in Gleichung (6.127) vorhandene Teil der Vorbedingung, der somit abgetrennt werden kann. Die Bedingung $\mathsf{I_i}$ impliziert

$$\bigwedge_{\nu=1}^{n_e} |b_{e\,\nu}(\delta.v).v| \leq \mathrm{Max_f} \,. \tag{6.133}$$

Von Gleichung (6.127) verbleibt somit als Vorbedingung

$$\begin{aligned} &\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \wedge \boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \wedge \\ &\boldsymbol{P}_{v^3e}\, \boldsymbol{b}_e(\delta.v).v = \boldsymbol{a}_v^3([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}}) \,. \end{aligned} \tag{6.134}$$

Den letzten Konjunktionsterm ersetzen wir mit

$$\boldsymbol{b}_e(\delta.v).v = \boldsymbol{b}_e(\boldsymbol{\mu}(\delta.v)) = \boldsymbol{b}_e([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}}) \tag{6.135}$$

zu

$$\begin{aligned} &\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \wedge \boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \wedge \\ &\boldsymbol{P}_{v^3e}\, \boldsymbol{b}_e([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}}) = \boldsymbol{a}_v^3([\mu_1.v, \mu_2.v, P_{x_3} - 1]^{\mathrm{T}}) \,. \end{aligned} \tag{6.136}$$

Berücksichtigen wir $\mathsf{I_i} \Rightarrow \mathsf{P_{WDF}}$, so haben wir schließlich

$$\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \wedge \boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \,. \tag{6.137}$$

Die Vorbedingung der Anweisung `delta = mu1 + mu2 *` P_{x_1} `+ mu3 *` P_{x_1} `*` P_{x_2} `;` lautet mit Gleichung (6.41)

$$\begin{aligned} &\mathsf{I_i} \wedge b_i \wedge \mu_1.v + P_{x_1}\mu_2.v + P_{x_1}P_{x_2}\mu_3.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = P_{x_3} - 1 \wedge \\ &\mu_1.v \leq P_{x_1} - 1 \wedge \mu_2.v \leq P_{x_2} - 1 \wedge \mu_3.v \leq P_{x_3} - 1 \wedge \\ &\boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \,. \end{aligned} \tag{6.138}$$

Trennen wir zudem die von $\mathsf{I_i} \wedge b_i$ implizierten Konjunktionsterme ab, so ist die Vorbedingung von `mu3=` P_{x_3} `-1`

$$\mathsf{I_i} \wedge b_i \wedge P_{x_3} - 1 = P_{x_3} - 1 \wedge \boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \wedge P_{x_3} - 1 \leq \mathrm{Max_i} \,. \tag{6.139}$$

Berücksichtigung von $\mathsf{I_i} \Longrightarrow P_{x_3} \leq \mathrm{Max_i}$ (nach Gleichung (6.34)) liefert schließlich

$$\mathsf{I_i} \wedge b_i \wedge \boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}}) \,. \tag{6.140}$$

Durch analoge Überlegungen erreichen wir durch S_{av6} eine Abtrennung von $\boldsymbol{a}_v^6(\delta([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})).v = \boldsymbol{a}_v^6([\mu_1.v, \mu_2.v, 0]^{\mathrm{T}})$. Wir bekommen als Vorbedingung von S_{av6} im Wesentlichen

$$\mathsf{I_i} \wedge b_i \wedge \delta.v = \delta(\boldsymbol{\mu}.v) \wedge \mu_3.v = 0 \,. \tag{6.141}$$

Durch die beiden zuvor ausgeführten Anweisungen können $\delta.v = \delta(\boldsymbol{\mu}.v)$ und $\mu_3.v = 0$ abgetrennt werden. Die Vorbedingung von `mu3=0;` und somit auch die von $S_{4\mathrm{i}}$ ist, wie behauptet, durch $\mathsf{I_i} \wedge b_i$ gegeben.

Die Vorbedingungen der Anweisungen S_{avx} und S_{avy} werden in gleicher Weise ermittelt. Auf die Beweisführung wird an dieser Stelle verzichtet.

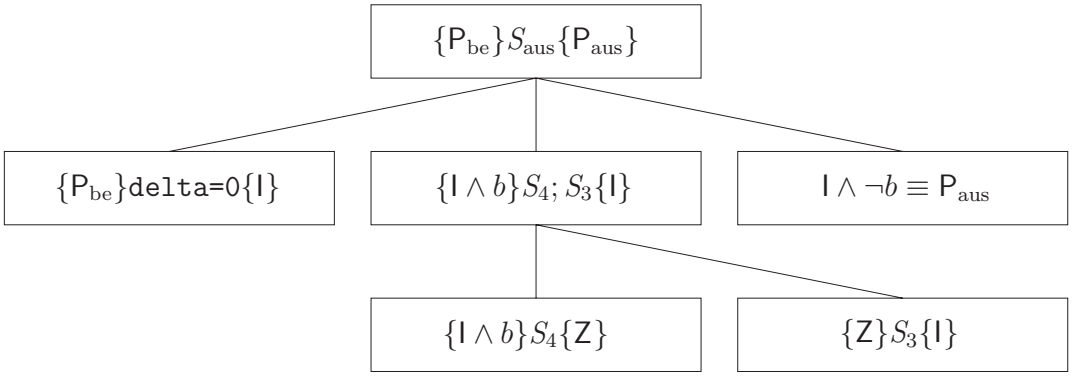


Bild 6.11: Berechnung der Eingangswellen der Verzögerer

6.5 Berechnung der Ausgangssignale

Die Berechnung der Ausgangssignale $\boldsymbol{a}_{\mathrm{FB}}$ des Funktionsbausteins für jeden räumlichen Gitterpunkt gemäß Gleichung (6.2) erfolgt mittels der Anweisung S_{aus}. Die Nachbedingung lautet

$$\mathsf{P}_{\mathrm{aus}} \equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}'_{\mathrm{be}} \wedge \mathsf{P}'_{\mathrm{aus}} \,. \tag{6.142}$$

Die Berechnung über alle Gitterpunkte erfolgt ebenfalls durch Verwendung einer for-Schleife der Länge P und der Laufvariablen δ. Wir verwenden dazu den folgenden C-Code

Bedingung	Anweisung	Kommentar
$\{\mathsf{P}_{\mathrm{be}}\}$		
	`for(delta=0 ; delta <` P `; delta=delta+1){`	
$\{\mathsf{I}\}$		Invariante
	S_{aFB1}	Berechnung von a_{FB1}
	$\vdots$	
	$S_{\mathrm{aFB}n_{aa}}$	Berechnung von $a_{FBn_{aa}}$
	`}`	
$\{\mathsf{P}_{\mathrm{aus}}\}$		

wobei die Berechnungsanweisungen durch

$S_{\mathrm{aFB1}} \equiv$ `a_FB_1[delta]=` a_{11} `*b_q_1[delta]` `+...+`a_{1n_q} `*b_q_`n_q `[delta]+`
a_{1n_q+1} `*b_v_1_1[delta]+...+`$a_{1n_q+n_v}$ `*b_v_7_`n_v^7`[delta];`
$a_{1n_q+n_v+1}$ `*b_e_1[delta]` `+...+`$a_{1n_q+n_v+n_e}$ `*b_e_`n_e `[delta];`

bis

$S_{\mathrm{aFB}n_{aa}} \equiv$ `a_FB_`n_{aa}`[delta]=` $a_{n_{aa}1}$ `*b_q_1[delta]` `+...+`$a_{n_{aa}n_q}$ `*b_q_`n_q `[delta]+`
$a_{n_{aa}n_q+1}$ `*b_v_1_1[delta]+...+`$a_{n_{aa}n_q+n_v}$ `*b_v_7_`n_v^7`[delta];`
$a_{n_{aa}n_q+n_v+1}$ `*b_e_1[delta]` `+...+`$a_{n_{aa}n_q+n_v+n_e}$ `*b_e_`n_e `[delta];`

gegeben sind. Im Bild 6.11 ist eine Übersicht des Beweises (der Ermittlung der Vorbedingung) dargestellt. Hierin ist die Schleifenbedingung $b \equiv \delta.v < P$.

Satz 10 *Der Ausdruck*

$$\mathsf{I} \equiv 0 \leq \delta.v \leq P \wedge \mathsf{P}_{\mathrm{be}} \wedge \bigwedge_{\delta=0}^{\delta.v-1} \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\mathrm{FB}k}(\boldsymbol{\mu}(\delta), \delta) \tag{6.143}$$

mit

$$\mathsf{P}_{\mathrm{be}} \equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}'_{\mathrm{be}} \tag{6.144}$$

ist eine gültige Schleifeninvariante der Schleife.

Beweis 10

Anwenden von Axiom (5.18) auf die Initialisierungsanweisung `delta=0` liefert

$$0 \leq P \wedge \mathsf{P}_{\mathrm{be}} \wedge \delta.l = -1 \wedge \delta.typ = \mathtt{int} \wedge \bigwedge_{\delta=0}^{-1} \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\mathrm{FB}k}(\boldsymbol{\mu}(\delta), \delta) \equiv \mathsf{P}_{\mathrm{be}}\,. \tag{6.145}$$

Die Nachbedingung der Schleife lautet

$$\mathsf{I} \wedge \delta.v \geq P \;\equiv\; \delta.v = P \wedge \mathsf{P}_{\mathrm{be}} \wedge \bigwedge_{\delta=0}^{P-1} \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\mathrm{FB}k}(\boldsymbol{\mu}(\delta), \delta) \tag{6.146}$$

und ist nach Schwächung um $\delta.v = P$ und P_{v} die Nachbedingung $\mathsf{P}_{\mathrm{aus}}$ nach Gleichung (6.142). Die Zwischenbedingung der Schleife lautet

$$\mathsf{Z} \;\equiv\; 0 \leq \delta.v{+}1 \leq P \wedge \mathsf{P}_{\mathrm{be}} \wedge \bigwedge_{\delta=0}^{\delta.v} \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\mathrm{FB}k}(\boldsymbol{\mu}(\delta), \delta) \wedge \delta.v{+}1 \leq \mathrm{Max_i} \wedge \delta.l = -1 \wedge \delta.typ = \mathtt{int}\,. \tag{6.147}$$

Wegen $\mathsf{P}_{\mathrm{be}} \Longrightarrow P \leq \mathrm{Max_i}$ kann $\delta.v + 1 \leq \mathrm{Max_i}$ abgetrennt werden. Mit der Stärkung um $0 \leq \delta.v$ und einigen Vereinfachungen erhalten wir

$$\mathsf{Z}' \;\equiv\; \mathsf{Z} \wedge 0 \leq \delta.v \;\equiv\; \mathsf{I} \wedge b \wedge \bigwedge_{k=1}^{n_{aa}} \mathsf{P}_{a\mathrm{FB}k}(\boldsymbol{\mu}(\delta.v), \delta.v)\,. \tag{6.148}$$

Es muss nun die Richtigkeit von $\{\mathsf{I} \wedge b\} S_4 \{\mathsf{Z}'\}$ mit $S_4 \equiv S_{\mathrm{aFB1}}; S_{\mathrm{aFB2}}; \ldots S_{\mathrm{aFB}n_{aa}}$ gezeigt werden. Wir bestimmen die Vorbedingung von S_{aFB1} mit Axiom (5.43) zu

$$\begin{aligned}
&\mathsf{I} \wedge b \wedge \bigwedge_{k=1}^{n_{aa}} a_{\mathrm{FB}\,k}(\delta.v).v = a_{\mathrm{FB}\,k}(\boldsymbol{\mu}(\delta.v)) \begin{bmatrix} a_{\mathrm{FB}\,k}(\delta.v).v \\ \boldsymbol{a}_{k,\bullet}[\boldsymbol{b}_q^{\mathrm{T}}(\delta.v).v, \boldsymbol{b}_v^{1\mathrm{T}}(\delta.v).v, \ldots, \boldsymbol{b}_v^{7\mathrm{T}}(\delta.v).v, \boldsymbol{b}_e^{\mathrm{T}}(\delta.v).v]^{\mathrm{T}} \end{bmatrix} \\
&\bigwedge_{k=1}^{n_{aa}} \left[\bigwedge_{m=1}^{n_q} \left| \sum_{\mu=1}^{m} a_{k\,\mu} b_{q\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \bigwedge_{l=1}^{7} \bigwedge_{m=1}^{n_v^l} \left| \sum_{\mu=1}^{n_q} a_{k\,\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{l} \sum_{\mu=1}^{m} a_{k\,*} b_{v\,\mu}^{\kappa}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \right. \\
&\left. \bigwedge_{m=1}^{n_e} \left| \sum_{\mu=1}^{n_q} a_{k\,\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{7} \sum_{\mu=1}^{n_v^\kappa} a_{k\,*} b_{v\,\mu}^{\kappa}(\delta.v).v + \sum_{\mu=1}^{m} a_{k\,*} b_{e\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \right] \wedge \\
&\bigwedge_{k=1}^{n_{aa}} \left[\bigwedge_{\mu=1}^{n_q} |a_{k\,\mu} b_{q\,\mu}(\delta.v).v| \leq \mathrm{Max_f} \wedge \bigwedge_{l=1}^{7} \bigwedge_{\mu=1}^{n_v^l} |a_{k\,*} b_{v\,\mu}^{l}(\delta.v).v| \leq \mathrm{Max_f} \wedge \bigwedge_{\mu=1}^{n_e} |a_{k\,*} b_{e\,\mu}(\delta.v).v| \leq \mathrm{Max_f} \right] \wedge \\
&\bigwedge_{k=1}^{n_{aa}} 0 \leq \delta.v \leq a_{\mathrm{FB}\,k}.l{-}1 \wedge \bigwedge_{k=1}^{n_q} 0 \leq \delta.v \leq b_{q\,k}.l{-}1 \wedge \bigwedge_{m=1}^{7} \bigwedge_{\mu=1}^{n_v^m} 0 \leq \delta.v \leq b_{v\,k}^{m}.l{-}1 \wedge \bigwedge_{k=1}^{n_e} 0 \leq \delta.v \leq b_{e\,k}.l{-}1 \wedge \\
&\bigwedge_{l=1}^{n_{aa}} \left[\bigwedge_{k=1}^{n_q} a_{\mathrm{FB}\,l}.typ = b_{q\,k}.typ \wedge \bigwedge_{k=1}^{n_e} a_{\mathrm{FB}\,l}.typ = b_{e\,k}.typ \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} a_{\mathrm{FB}\,l}.typ = b_{v\,k}^{\kappa}.typ \right] \wedge \\
&\delta.v \leq \mathrm{Max_i} \wedge \delta.l = -1 \wedge \delta.typ = \mathtt{int}\,,
\end{aligned} \tag{6.149}$$

wobei der Wert von $*$ der Übersicht halber nicht explizit angegeben wird. Zur Verdichtung des obigen Prädikats werden wir zeigen, dass I die Prädikate bzgl. der Datentypen und der Feldlänge impliziert. Aus Gleichung (6.143), Gleichung (6.144), Gleichung (6.33) und Gleichung (6.34) folgt

$$\mathsf{I} \Rightarrow \bigwedge_{k=1}^{n_{aa}} a_{\mathrm{FB}\,k}.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_q} b_{q\,k}.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_e} b_{e\,k}.typ = \texttt{float} \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} b_{v\,k}^\kappa.typ = \texttt{float}\ . \tag{6.150}$$

Durch Anwendung des Prädikatenkalküls folgt

$$\bigwedge_{l=1}^{n_{aa}} \left[\bigwedge_{k=1}^{n_q} a_{\mathrm{FB}\,l}.typ = b_{q\,k}.typ \wedge \bigwedge_{k=1}^{n_e} a_{\mathrm{FB}\,l}.typ = b_{e\,k}.typ \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} a_{\mathrm{FB}\,l}.typ = b_{v\,k}^\kappa.typ\,. \right] \tag{6.151}$$

und letzteres Prädikat kann aus Gleichung (6.149) abgetrennt werden. Ebenso kann aus Gleichung (6.143), Gleichung (6.144), Gleichung (6.33) und Gleichung (6.34)

$$\begin{aligned} \mathsf{I} \wedge b \Rightarrow \bigwedge_{k=1}^{n_{aa}} P-1 \le a_{\mathrm{FB}\,k}.l - 1 \wedge \bigwedge_{k=1}^{n_q} P-1 \le b_{q\,k}.l - 1 \wedge \bigwedge_{k=1}^{n_e} P-1 \le b_{e\,k}.l - 1 \wedge \\ \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} P-1 \le b_{v\,k}^\kappa.l - 1 \wedge P \le \mathrm{Max_i} \end{aligned} \tag{6.152}$$

gewonnen werden. Mit $\delta.v + 1 \le P \equiv \delta.v \le P - 1$ folgt

$$\begin{aligned} \mathsf{I} \wedge b \Rightarrow \bigwedge_{k=1}^{n_{aa}} 0 \le \delta.v \le a_{\mathrm{FB}\,k}.l - 1 \wedge \bigwedge_{k=1}^{n_q} 0 \le \delta.v \le b_{q\,k}.l - 1 \wedge \\ \bigwedge_{k=1}^{n_e} 0 \le \delta.v \le b_{e\,k}.l - 1 \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} 0 \le \delta.v \le b_{v\,k}^\kappa.l - 1 \wedge \delta.v + 1 \le \mathrm{Max_i}\,. \end{aligned} \tag{6.153}$$

Die Prädikate bzgl. der Länge und $\delta.v \le \mathrm{Max_i}$ können somit auch aus Gleichung (6.149) abgetrennt werden. Ferner gilt

$$\mathsf{I} \Rightarrow \delta\,.typ = \texttt{int} \wedge \delta\,.l = -1\ . \tag{6.154}$$

Wir zeigen nun, dass aus der Vorbedingung $\mathsf{P}_{\mathrm{aus}}$ die Beschränktheit der Summanden und der Summen folgt. Da die Matrix $\boldsymbol{A}$ eine spezielle Struktur hat, ergibt sich der (Soll-)Wert eines Ausgangssignals $a_{\mathrm{FB}k}(\delta)$ entweder durch

$$u_\mu = \sqrt{R_\mu}\,[a_\mu + b_\mu] \quad \text{oder durch} \quad i_\mu = \sqrt{G_\mu}\,[a_\mu - b_\mu]\ . \tag{6.155}$$

Zunächst weisen wir nach , dass jeder Summand beschränkt ist. Dazu berücksichtigen wir

$$\mathsf{I} \Rightarrow \mathsf{P}_{\mathrm{be}} \Rightarrow \bigwedge_{\delta=0}^{P-1} \left[\bigwedge_{\mu=1}^{n_q} |b_{q\,\mu}(\delta.v).v| 2\beta \le \mathrm{Max_f} \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{\mu=1}^{n_v^\kappa} |b_{v\,\mu}^\kappa(\delta.v).v| 2\beta \le \mathrm{Max_f} \wedge \bigwedge_{\mu=1}^{n_e} |b_{e\,\mu}(\delta.v).v| 2\beta \le \mathrm{Max_f} \right] \tag{6.156}$$

und durch Schwächung

$$\mathsf{I} \Rightarrow \mathsf{P}_{\mathrm{be}} \Rightarrow \bigwedge_{\delta=0}^{P-1} \left[\bigwedge_{\mu=1}^{n_q} |b_{q\,\mu}(\delta.v).v| \beta \le \mathrm{Max_f} \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{\mu=1}^{n_v^\kappa} |b_{v\,\mu}^\kappa(\delta.v).v| \beta \le \mathrm{Max_f} \wedge \bigwedge_{\mu=1}^{n_e} |b_{e\,\mu}(\delta.v).v| \beta \le \mathrm{Max_f} \right]\ .$$

(6.157)

Beachten wir die Definition von β nach Gleichung (6.38), so haben wir

$$\begin{aligned}\mathsf{I} \Rightarrow \mathsf{P}_{\mathrm{be}} \Rightarrow \bigwedge_{\delta=0}^{P-1} \Bigg[\bigwedge_{\mu=1}^{n_q} & \left[\sqrt{R_\mu}|b_{q\,\mu}(\delta.v).v| \le \mathrm{Max_f} \wedge \sqrt{G_\mu}|b_{q\,\mu}(\delta.v).v| \le \mathrm{Max_f}\right] \wedge \\ \bigwedge_{\kappa=1}^{7} \bigwedge_{\mu=1}^{n_v^\kappa} & \left[\sqrt{R_\mu}|b_{v\,\mu}^\kappa(\delta.v).v| \le \mathrm{Max_f} \wedge \sqrt{G_\mu}|b_{v\,\mu}^\kappa(\delta.v).v| \le \mathrm{Max_f}\right] \wedge \\ \bigwedge_{\mu=1}^{n_e} & \left[\sqrt{R_\mu}|b_{e\,\mu}(\delta.v).v| \le \mathrm{Max_f} \wedge \sqrt{G_\mu}|b_{e\,\mu}(\delta.v).v| \le \mathrm{Max_f}\right] \Bigg] ,\end{aligned} \tag{6.158}$$

d.h. die Tatsache, dass jede Welle a_μ gleich einer Welle b_ν ist und erhalten mittels Schwächung den gesuchten Nachweis der Beschränktheit jedes Summanden.

Wir wollen im Folgenden die Beschränktheit der Summen nachweisen, d. h.$|u_\mu| \le \mathrm{Max_f}$ und $|i_\mu| \le \mathrm{Max_f}$. Mit der Dreiecksungleichung erhalten wir aus Gleichung (6.155)

$$|u_\mu| \le \sqrt{R_\mu}\,[\,|a_\mu| + |b_\mu|\,] \quad , \quad |i_\mu| \le \sqrt{G_\mu}\,[\,|a_\mu| + |b_\mu|\,] \,. \tag{6.159}$$

Wenden wir darauf die Gleichung (6.158) an, so offenbart dies, dass die Beträge der Torspannungen und Torströme, die den Ausgangssignalen $a_{\mathrm{FB}k}(\delta)$ entsprechen, durch

$$\bigwedge_{\delta=0}^{P-1} \bigwedge_{\mu=1}^{n_g} \left[|a_{\mathrm{FB}k}(\delta)| \le \sqrt{R_\mu}\left[\frac{\mathrm{Max_f}}{2\beta} + \frac{\mathrm{Max_f}}{2\beta}\right] = \frac{\sqrt{R_\mu}}{\beta}\mathrm{Max_f} \wedge |a_{\mathrm{FB}k}(\delta)| \le \sqrt{G_\mu}\left[\frac{\mathrm{Max_f}}{2\beta} + \frac{\mathrm{Max_f}}{2\beta}\right] = \frac{\sqrt{G_\mu}}{\beta}\mathrm{Max_f}\right] \tag{6.160}$$

begrenzt sind. Aus der Definition von β folgt nun

$$\sqrt{R_\mu} \le \beta \iff \frac{\sqrt{R_\mu}}{\beta}\mathrm{Max_f} \le \mathrm{Max_f} \quad \wedge \quad \sqrt{G_\mu} \le \beta \iff \frac{\sqrt{G_\mu}}{\beta}\mathrm{Max_f} \le \mathrm{Max_f} \,. \tag{6.161}$$

Demnach gilt auch

$$\bigwedge_{\delta=0}^{P-1} \bigwedge_{\mu=1}^{n_g} \left[\frac{\sqrt{R_\mu}}{\beta}\mathrm{Max_f} \le \mathrm{Max_f} \,\wedge\, \frac{\sqrt{G_\mu}}{\beta}\mathrm{Max_f} \le \mathrm{Max_f} \right] \,. \tag{6.162}$$

Konjunktive Verknüpfung von Gleichung (6.160), Gleichung (6.162) und Beachtung von $0 \le \delta.v \le P-1$ ergibt

$$\bigwedge_{\mu=1}^{n_g} \left[|a_{\mathrm{FB}k}(\delta.v)| \le \mathrm{Max_f} \,\wedge\, |a_{\mathrm{FB}k}(\delta.v)| \le \mathrm{Max_f} \right] \,. \tag{6.163}$$

Es verbleibt somit von Gleichung (6.149)

$$\mathsf{I} \wedge b \wedge \bigwedge_{k=1}^{n_{aa}} \boldsymbol{a}_{k,\bullet}[\boldsymbol{b}_q^{\mathrm{T}}(\delta.v).v, \boldsymbol{b}_v^{1\mathrm{T}}(\delta.v).v, \ldots, \boldsymbol{b}_v^{7\mathrm{T}}(\delta.v).v, \boldsymbol{b}_e^{\mathrm{T}}(\delta.v).v]^{\mathrm{T}} = a_{\mathrm{FB}\,k}(\boldsymbol{\mu}(\delta.v)) \,. \tag{6.164}$$

Mit

$$\mathsf{I} \;\Rightarrow\; \boldsymbol{b}_q(\delta.v).v = \boldsymbol{b}_q(\boldsymbol{\mu}(\delta.v)) \,\wedge\, \boldsymbol{b}_e(\delta.v).v = \boldsymbol{b}_e(\boldsymbol{\mu}(\delta.v)) \,\wedge\, \bigwedge_{\kappa=1}^{7} \boldsymbol{b}_v^\kappa(\delta.v).v = \boldsymbol{b}_v^\kappa(\boldsymbol{\mu}(\delta.v)) \tag{6.165}$$

können wir in Gleichung (6.164) die prädikatenlogischen Variablen $\boldsymbol{b}(\delta.v).v$ durch die Sollwerte $\boldsymbol{b}(\boldsymbol{\mu}(\delta.v))$ ersetzen. Anschließend nutzen wir $\mathsf{I} \Rightarrow \mathsf{P}_{\mathrm{WDF}}$ und trennen die Skalarprodukte ab. Wir erhalten, wie behauptet, $\mathsf{I} \wedge b$.

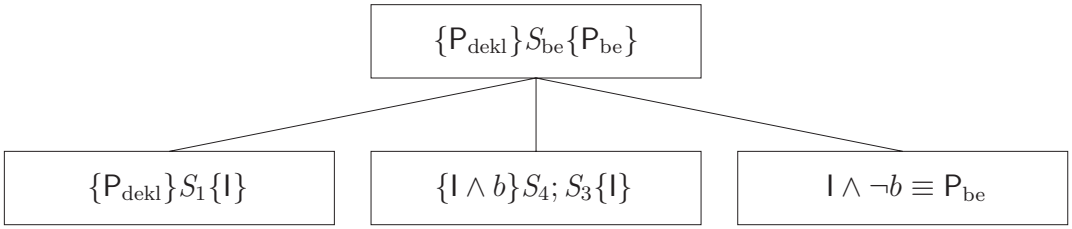


Bild 6.12: Berechnung der Wellengrößen der nichtdynamischen Elemente

6.6 Berechnung der Wellengrößen der nichtdynamischen Elemente

Wir behandeln nun die Anweisung S_{be}, die die Berechnung der Wellengrößen der nichtdynamischen Elemente $\boldsymbol{b}_e$ durchführt. Nach Gleichung (6.1) sind die Wellengrößen $\boldsymbol{b}_e$ für alle Gitterpunkte des Berechnungsgebietes zu ermitteln. Wir gehen davon aus, dass vor dem Aufruf des Funktionsbausteins zu jedem Gitterpunkt der Abtastschicht die Wellengrößen $\boldsymbol{b}_v$ und $\boldsymbol{b}_q$ bekannt sind. Die Nachbedingung von S_{be} lautet

$$\mathsf{P}_{\text{be}} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{dec}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}'_{\text{be}}\,. \tag{6.166}$$

Die Auswertung von Gleichung (6.1) erfolgt innerhalb einer for-Schleife mit der Laufvariablen $\delta.v$ von 0 bis $P-1$, sodass sich die Berechnung über alle Gitterpunkte der Abtastschicht erstreckt.

Die zusätzliche Nachbedingung P'_{be} besagt, dass die Werte der Variablen für die nichtdynamischen Elemente denen aus Gleichung (6.1) für den aktuellen Zeitpunkt μ_4 und für $\delta.v = 0, 1, \ldots, P-1$ entsprechen müssen. Anders formuliert lautet die Nachbedingung P_{be}

$$\mathsf{P}_{\text{be}} \equiv \mathsf{P}_{\text{dekl}} \wedge \bigwedge_{\delta=0}^{P-1}\bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge\ |b_{e\,k}(\delta).v|2\beta \leq \text{Max}_{\text{f}}]\ . \tag{6.167}$$

Zur Erfüllung der Bedingungen nutzen wir die folgende Anweisung in C-Code

Bedingung	Anweisung	Kommentar
$\{\mathsf{P}_{\text{dekl}}\}$		
	`for(delta=0 ; delta <` P `; delta=delta+1){`	
$\{\mathsf{I}\}$		Invariante
	S_{be1}	Berechnung von $\boldsymbol{b}_{e1}$
	$\vdots$	
	$S_{\text{be}n_e}$	Berechnung von $\boldsymbol{b}_{en_e}$
	`}`	
$\{\mathsf{P}_{\text{be}}\}$		

mit den Zuweisungsanweisungen

$$\begin{aligned} S_{\text{be1}} \equiv \texttt{b\_e\_1[delta]} =& l_{11} \texttt{*b\_q\_1[delta]} + \ldots + l_{1n_q} \texttt{*b\_q\_}n_q\texttt{[delta]} \\ &+ l_{1(n_q+1)} \texttt{*b\_v\_1\_1[delta]} + \ldots + l_{1(n_q+n_v)} \texttt{*b\_v\_6\_}n_v^6\texttt{[delta];} \end{aligned} \tag{6.168}$$

und für $k = 2 \ldots n_e$

$$\begin{aligned} S_{\text{be}k} \equiv \texttt{b\_e\_}k\texttt{[delta]} =& l_{k1} \texttt{*b\_q\_1[delta]} + \ldots + l_{kn_q} \texttt{*b\_q\_}n_q\texttt{[delta]} \\ &+ l_{k(n_q+1)} \texttt{*b\_v\_1\_1[delta]} + \ldots + l_{k(n_q+n_v)} \texttt{*b\_v\_6\_}n_v^6\texttt{[delta]} \\ &+ l_{k(n_q+n_v+1)} \texttt{*b\_e\_1[delta]} + \ldots + l_{k(n_q+n_v+k-1)} \texttt{*b\_e\_}k\text{–}1\texttt{[delta];}\,. \end{aligned} \tag{6.169}$$

Wir führen nun den Nachweis der Korrektheit von $\{\mathsf{P}_{\mathrm{dekl}}\}\, S_{\mathrm{be}}\, \{\mathsf{P}_{\mathrm{be}}\}$, vgl. dazu die Übersicht im Bild 6.12 mit der Schleifenbedingung $b \equiv \delta.v < P \equiv \delta.v \leq P-1$.

Satz 11 *Der Ausdruck*

$$\mathsf{I} \equiv 0 \leq \delta.v \leq P \wedge \mathsf{P}_{\mathrm{dekl}} \wedge \bigwedge_{\delta=0}^{\delta.v-1} \bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge \ |b_{e\,k}(\delta).v|2\beta \leq \mathrm{Max_f}] \tag{6.170}$$

mit

$$\mathsf{P}_{\mathrm{dekl}} \ \equiv \ \mathsf{P}_{\mathrm{WDF}} \ \wedge \ \mathsf{P}_{\mathrm{dec}} \ \wedge \ \mathsf{P}_{\mathrm{q}} \ \wedge \ \mathsf{P}_{\mathrm{v}} \tag{6.171}$$

ist eine gültige Invariante der Schleife.

Beweis 11

Gemäß Axiom (5.35) können wir den Nachweis auf das Zeigen der Initialisierung $\{\mathsf{P}_{\mathrm{dekl}}\}S_1\{\mathsf{I}\}$, des Schleifenendes $\mathsf{P}_{\mathrm{be}} \equiv \mathsf{I} \wedge \neg b$ und der Invarianz $\{\mathsf{I} \wedge b\}S_4; S_3\{\mathsf{I}\}$ zurückführen.

Die Vorbedingung von `delta=0;` ermittelt sich zu

$$0 \leq P \wedge \mathsf{P}_{\mathrm{dekl}} \wedge \bigwedge_{\delta=0}^{-1} \bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge \ |b_{e\,k}(\delta).v|2\beta \leq \mathrm{Max_f}] \wedge \delta\,.l = -1 \wedge \delta\,.typ = \texttt{int}\ . \tag{6.172}$$

Dies entspricht der Vorbedingung $\mathsf{P}_{\mathrm{dekl}}$, da $\mathsf{P}_{\mathrm{dekl}}$ die Bedingung $0 \leq P$ impliziert. Die Abbruchbedingung lautet $\neg b \equiv \delta.v \geq P$, sodass wir als Nachbedingung

$$\mathsf{I} \wedge \delta.v \geq P \equiv \delta.v = P \wedge \mathsf{P}_{\mathrm{dekl}} \wedge \bigwedge_{\delta=0}^{P-1} \bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge \ |b_{e\,k}(\delta).v|2\beta \leq \mathrm{Max_f}] \tag{6.173}$$

erhalten, was nach Schwächung um $\delta.v = P$, den Forderungen gemäß Gleichung (6.167) entspricht.

Für den Nachweis von $\{\mathsf{I} \wedge b\}S_4; S_3\{\mathsf{I}\}$ bestimmen wir zunächst die durch $\{\mathsf{Z}\}S_3\{\mathsf{I}\}$ definierte Zwischenbedingung Z und bereiten diese durch Nutzen von Gleichung (6.170) zu

$$\begin{aligned}
\mathsf{Z} \ \equiv \ & 0 \leq \delta.v+1 \leq P \ \wedge \ \mathsf{P}_{\mathrm{dekl}} \wedge \bigwedge_{\delta=0}^{\delta.v} \bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge \ |b_{e\,k}(\delta).v|2\beta \leq \mathrm{Max_f}] \ \wedge \\
& \delta.v \leq \mathrm{Max_i} \ \wedge \ \delta\,.typ = \texttt{int} \ \wedge \ \delta\,.l = -1 \\
\equiv \ & 0 \leq \delta.v+1 \leq P \wedge \mathsf{P}_{\mathrm{dekl}} \wedge \bigwedge_{\delta=0}^{\delta.v-1} \bigwedge_{k=1}^{n_e} [b_{ek}(\delta).v = b_{ek}(\boldsymbol{\mu}(\delta)) \ \wedge \ |b_{e\,k}(\delta).v|2\beta \leq \mathrm{Max_f}] \wedge \\
& \bigwedge_{k=1}^{n_e} [b_{ek}(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,k}(\delta.v).v|2\beta \leq \mathrm{Max_f}] \wedge \delta.v \leq \mathrm{Max_i} \wedge \delta\,.typ = \texttt{int} \wedge \delta\,.l = -1
\end{aligned} \tag{6.174}$$

auf. Wir stärken die Vorbedingung zudem durch $0 \leq \delta.v$ und erhalten nach einigen Vereinfachungen

$$\mathsf{Z}' \equiv \mathsf{Z} \wedge 0 \leq \delta.v \equiv \mathsf{I} \wedge b \wedge \bigwedge_{k=1}^{n_e} [b_{ek}(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,k}(\delta.v).v|2\beta \leq \mathrm{Max_f}]\ . \tag{6.175}$$

Es verbleibt $\{\mathsf{I} \wedge b\} S_4 \{\mathsf{Z}'\}$ zu zeigen. Zunächst behandeln wir die Anweisungen $S_{\mathrm{be}k}$, $k = 1, \ldots, n_e$. Den Beweis haben wir mit Axiom (5.7) erbracht, wenn wir mit

$$\begin{aligned}
\mathsf{V}_{\mathrm{Sbe}k} &\equiv \mathsf{I} \wedge b \wedge \bigwedge_{\kappa=1}^{k-1} [b_{e\kappa}(\delta.v).v = b_{e\kappa}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,\kappa}(\delta.v).v|2\beta \leq \mathrm{Max_f}] \\
\mathsf{P}_{\mathrm{Sbe}k} &\equiv \mathsf{I} \wedge b \wedge \bigwedge_{\kappa=1}^{k} [b_{e\kappa}(\delta.v).v = b_{e\kappa}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,\kappa}(\delta.v).v|2\beta \leq \mathrm{Max_f}] \\
&\equiv \mathsf{V}_{\mathrm{Sbe}\kappa} \wedge b_{ek}(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,k}(\delta.v).v|2\beta \leq \mathrm{Max_f}\,.
\end{aligned} \tag{6.176}$$

die Korrektheit von $\mathsf{I} \wedge b \equiv \mathsf{V}_{\mathrm{Sbe}1}$, $\mathsf{Z}' \equiv \mathsf{P}_{\mathrm{Sbe}n_e}$ und $\{\mathsf{V}_{\mathrm{Sbe}k}\} S_{\mathrm{be}k} \{\mathsf{P}_{\mathrm{Sbe}k}\}$ nachweisen. Das entspricht dem Zeigen der Richtigkeit von

$$\{\mathsf{V}_{\mathrm{Sbe}1}\} \underbrace{S_{be1}; \ldots S_{ben_e}}_{S_4} \{\mathsf{P}_{\mathrm{Sbe}n_e}\}\,. \tag{6.177}$$

Die Gültigkeit von $\mathsf{I} \wedge b \equiv \mathsf{V}_{\mathrm{Sbe}1}$ ist offensichtlich. Weiterhin gilt

$$\mathsf{P}_{\mathrm{Sbe}n_e} \equiv \mathsf{I} \wedge b \wedge \bigwedge_{k=1}^{n_e} [b_{ek}(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)) \wedge |b_{e\,k}(\delta.v).v|2\beta \leq \mathrm{Max_f}] \tag{6.178}$$

und ist Z' wie zu zeigen war. Es verbleibt $\{\mathsf{V}_{\mathrm{Sbe}k}\} S_{\mathrm{be}k} \{\mathsf{P}_{\mathrm{Sbe}k}\}$ zu zeigen. Die Vorbedingung von $S_{\mathrm{be}k}$ ergibt sich zu

$$\begin{aligned}
&\mathsf{V}_{\mathrm{be}k} \wedge b_{ek}(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)) \begin{bmatrix} b_{ek}(\delta.v).v \\ \boldsymbol{l}_{q\,k,\bullet}\, \boldsymbol{b}_q(\delta.v).v + \boldsymbol{l}_{v\,k,\bullet}\, \boldsymbol{b}_v(\delta.v).v + \boldsymbol{l}_{e\,k,\bullet}\, \boldsymbol{b}_e(\delta.v).v \end{bmatrix} \wedge \\
&|b_{e\,k}(\delta.v).v|2\beta \leq \mathrm{Max_f} \begin{bmatrix} b_{ek}(\delta.v).v \\ \boldsymbol{l}_{q\,k,\bullet}\, \boldsymbol{b}_q(\delta.v).v + \boldsymbol{l}_{v\,k,\bullet}\, \boldsymbol{b}_v(\delta.v).v + \boldsymbol{l}_{e\,k,\bullet}\, \boldsymbol{b}_e(\delta.v).v \end{bmatrix} \wedge \\
&\bigwedge_{m=0}^{n_q} \left| \sum_{\mu=1}^{m} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\
&\bigwedge_{l=1}^{7} \bigwedge_{m=1}^{n_v^l} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{l} \sum_{\mu=1}^{m} l_{v\,k*} b^{\kappa}_{v\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\
&\bigwedge_{m=1}^{n_e} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{7} \sum_{\mu=1}^{n_v^\kappa} l_{v\,k*} b^{\kappa}_{v\,\mu}(\delta.v).v + \sum_{\mu=1}^{m} l_{e\,k*} b_{e\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\
&\bigwedge_{\mu=1}^{n_q} \left| l_{q\,k\mu} b_{q\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{\mu=1}^{n_v^\kappa} \left| l_{v\,k*} b^{\kappa}_{v\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \bigwedge_{\mu=1}^{n_e} \left| l_{e\,k\mu} b_{e\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\
&\bigwedge_{k'=1}^{n_q} 0 \leq \delta.v \leq b_{q\,k'}.l - 1 \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k'=1}^{n_v^\kappa} 0 \leq \delta.v \leq b^{\kappa}_{v\,k'}.l - 1 \wedge \bigwedge_{k'=1}^{n_e} 0 \leq \delta.v \leq b_{e\,k'}.l - 1 \wedge \\
&\bigwedge_{k'=1}^{n_q} b_{e\,k}.typ = b_{q\,k'}.typ \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k'=1}^{n_v^\kappa} b_{e\,k}.typ = b^{\kappa}_{v\,k'}.typ \wedge \bigwedge_{k'=1}^{k-1} b_{e\,k}.typ = b_{e\,k'}.typ \wedge \\
&\delta.v + 1 \leq \mathrm{Max_i} \wedge \delta.l = -1 \wedge \delta.typ = \texttt{int}\,.
\end{aligned} \tag{6.179}$$

Zur Verdichtung des obigen Prädikats werden wir zeigen, dass $\mathsf{I} \wedge b$ die Prädikate bzgl. der Datentypen und der Feldlänge impliziert. Mit Gleichung (6.170), Gleichung (6.171), Gleichung (6.33) und Gleichung

(6.34) erhalten wir zum einen

$$\mathsf{I} \wedge b \Rightarrow \bigwedge_{k=1}^{n_q} b_{q\,k}.typ = \texttt{float} \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} b_{v\,k}^\kappa.typ = \texttt{float} \wedge \bigwedge_{k=1}^{n_e} b_{e\,k}.typ = \texttt{float}\,, \tag{6.180}$$

was somit aus Gleichung (6.179) abgetrennt werden kann und zum anderen

$$\mathsf{I} \wedge b \Rightarrow \bigwedge_{k=1}^{n_q} P-1 \leq b_{q\,k}.l-1 \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} P-1 \leq b_{v\,k}^\kappa.l-1 \wedge \bigwedge_{k=1}^{n_e} P-1 \leq b_{e\,k}.l-1 \wedge P \leq \mathrm{Max_i}\,. \tag{6.181}$$

Daraus gewinnen wir mit $\mathsf{I} \wedge b \Rightarrow 0 \leq \delta.v \leq P-1$

$$\begin{aligned}\mathsf{I} \wedge b \Rightarrow\ & \bigwedge_{k=1}^{n_q} 0 \leq \delta.v \leq b_{q\,k}.l-1 \wedge \bigwedge_{\kappa=1}^{7} \bigwedge_{k=1}^{n_v^\kappa} 0 \leq \delta.v \leq b_{v\,k}^\kappa.l-1 \wedge \bigwedge_{k=1}^{n_e} 0 \leq \delta.v \leq b_{e\,k}.l-1 \wedge \\ & \delta.v \leq \mathrm{Max_i} \wedge \delta\,.typ = \texttt{int} \wedge \delta\,.l = -1\,,\end{aligned} \tag{6.182}$$

was wiederum aus Gleichung (6.179) abgetrennt wird. Wir zeigen nun, dass die Beschränktheit der Teilskalarprodukte in Gleichung (6.179) durch $\mathsf{P}_{\mathrm{dekl}} \Rightarrow \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{max}}$ impliziert wird. Zum Beweis betrachten wir die Ergebnisse von Gleichung (C.11), Gleichung (C.15) und Gleichung (C.35) (für die PL-Variablen der Programmvariablen), die wiederum durch I impliziert werden. Wir haben somit

$$\begin{aligned}\bigwedge_{\delta.v=0}^{P-1} \bigwedge_{k=1}^{n_e} \Bigg[\ & \bigwedge_{m=0}^{n_q} \left| \sum_{\mu=1}^{m} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v \right|^2 \beta^2\, 4 \leq {\mathrm{Max_f}}^2 \wedge \\ & \bigwedge_{l=1}^{7} \bigwedge_{m=1}^{n_v^l} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{l} \sum_{\mu=1}^{m} l_{v\,k*} b_{v\,\mu}^\kappa(\delta.v).v \right|^2 \beta^2\, 4 \leq {\mathrm{Max_f}}^2 \wedge \\ & \bigwedge_{m=1}^{n_e} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{7} \sum_{\mu=1}^{n_v^\kappa} l_{v\,k*} b_{v\,\mu}^\kappa(\delta.v).v + \sum_{\mu=1}^{m} l_{e\,k*} b_{e\,\mu}(\delta.v).v \right|^2 \beta^2\, 4 \leq {\mathrm{Max_f}}^2 \Bigg]\,.\end{aligned} \tag{6.183}$$

Wir radizieren jeweils auf beiden Seiten, beachten, dass β nicht kleiner als 1 ist, berücksichtigen Gleichung (A.5) und erhalten dann

$$\begin{aligned}\bigwedge_{\delta.v=0}^{P-1} \bigwedge_{k=1}^{n_e} \Bigg[\ & \bigwedge_{m=0}^{n_q} \left| \sum_{\mu=1}^{m} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\ & \bigwedge_{l=1}^{7} \bigwedge_{m=1}^{n_v^l} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{l} \sum_{\mu=1}^{m} l_{v\,k*} b_{v\,\mu}^\kappa(\delta.v).v \right| \leq \mathrm{Max_f} \wedge \\ & \bigwedge_{m=1}^{n_e} \left| \sum_{\mu=1}^{n_q} l_{q\,k\mu} b_{q\,\mu}(\delta.v).v + \sum_{\kappa=1}^{7} \sum_{\mu=1}^{n_v^\kappa} l_{v\,k*} b_{v\,\mu}^\kappa(\delta.v).v + \sum_{\mu=1}^{m} l_{e\,k*} b_{e\,\mu}(\delta.v).v \right| \leq \mathrm{Max_f} \Bigg]\,,\end{aligned} \tag{6.184}$$

sodass dieser Teil und

$$|\boldsymbol{l}_{q\,k,\bullet}\, \boldsymbol{b}_q(\delta.v).v + \boldsymbol{l}_{v\,k,\bullet}\, \boldsymbol{b}_v(\delta.v).v + \boldsymbol{l}_{e\,k,\bullet}\, \boldsymbol{b}_e(\delta.v).v| 2\beta \leq \mathrm{Max_f} \tag{6.185}$$

ebenfalls von Gleichung (6.179) abgetrennt werden kann. Aufgrund der von I implizierten unitären Beschränktheit der Matrix $\boldsymbol{L}$ ist jedes Element der Matrix dem Betrage nach nicht größer als eins. Folglich können auch Forderungen an die Beschränktheit der Summanden des Skalarprodukts aus Gleichung (6.179) abgetrennt werden. Die Vorbedingung ist somit äquivalent zu

$$\mathsf{V}_{\mathrm{be}k} \wedge \boldsymbol{l}_{q\,k,\bullet}\,\boldsymbol{b}_q\,(\delta.v).v + \boldsymbol{l}_{v\,k,\bullet}\,\boldsymbol{b}_v\,(\delta.v).v + \boldsymbol{l}_{e\,k,\bullet}\,\boldsymbol{b}_e\,(\delta.v).v = b_{ek}(\boldsymbol{\mu}(\delta.v)\;. \tag{6.186}$$

Aufgrund der strikten unteren Dreiecksgestalt von $\boldsymbol{L}$ hat das dritte Skalarprodukt i. Allg. nur $k-1$ Summanden

$$\boldsymbol{l}_{e\,k,\bullet}\,\boldsymbol{b}_e\,(\delta.v).v = \sum_{\nu=1}^{n_e} \boldsymbol{l}_{e\,k\nu}\,\boldsymbol{b}_{e\,\nu}(\delta.v).v = \sum_{\nu=1}^{k-1} \boldsymbol{l}_{e\,k\nu}\,\boldsymbol{b}_{e\,\nu}(\delta.v).v\;. \tag{6.187}$$

Mit

$$\mathsf{V}_{\mathrm{Sbe}k} \Rightarrow \bigwedge_{k=1}^{n_q} b_{q\,k}(\delta.v).v = b_{q\,k}(\boldsymbol{\mu}(\delta.v)) \wedge \bigwedge_{\kappa=1}^{7}\bigwedge_{k=1}^{n_v^\kappa} b_{v\,k}^{\kappa}(\delta.v).v = b_{v\,k}^{\kappa}(\boldsymbol{\mu}(\delta.v)) \wedge \bigwedge_{\nu=1}^{k-1} b_{e\,\nu}(\delta.v).v = b_{e\,\nu}(\boldsymbol{\mu}(\delta.v)) \tag{6.188}$$

kann die Vorbedingung durch

$$\mathsf{V}_{\mathrm{Sbe}k} \wedge \boldsymbol{l}_{q\,k,\bullet}\,\boldsymbol{b}_q\,(\boldsymbol{\mu}(\delta.v)) + \boldsymbol{l}_{v\,k,\bullet}\,\boldsymbol{b}_v\,(\boldsymbol{\mu}(\delta.v)) + \boldsymbol{l}_{e\,k,\bullet}\,\boldsymbol{b}_e\,(\boldsymbol{\mu}(\delta.v)) = b_{e\,k}(\boldsymbol{\mu}(\delta.v)) \tag{6.189}$$

oder vereinfacht durch

$$\mathsf{V}_{\mathrm{Sbe}k} \wedge \boldsymbol{l}_{q\,k,\bullet}\,\boldsymbol{b}_q\,(\boldsymbol{\mu}(\delta.v)) + \boldsymbol{l}_{v\,k,\bullet}\,\boldsymbol{b}_v\,(\boldsymbol{\mu}(\delta.v)) + \sum_{\nu=1}^{k-1} \boldsymbol{l}_{e\,k\nu}\,\boldsymbol{b}_{e\,\nu}(\boldsymbol{\mu}(\delta.v)) = b_{e\,k}(\boldsymbol{\mu}(\delta.v)) \tag{6.190}$$

dargestellt werden. Mit $\mathsf{I} \Rightarrow \mathsf{P}_{\mathrm{WDF}}$ und der Abtrennung der wahren Aussage verbleibt letztendlich von der Vorbedingung Gleichung (6.179), wie behauptet, nur $\mathsf{V}_{\mathrm{Sbe}k}$.

6.7 Variablendeklarationen

In den vorangegangenen Kapiteln hatten wir festgestellt, dass S_{dekl} die Nachbedingung

$$\mathsf{P}_{\mathrm{dekl}} \equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{dec}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \tag{6.191}$$

haben muss. Durch die Anweisung S_{dekl} erfolgt die Deklaration der Variablen. Für unseren Algorithmus unterscheiden wie zwei Typen von Variablen. Zum einen haben wir die Variablen, die schon vor dem Aufruf des Funktionsbausteins deklariert sind und beim Programmaufruf übergeben werden. Das sind die Ein- und Ausgangssignale und die Verzögererwerte, d.h. die Variablen, die den Vektoren $\boldsymbol{b}_v, \boldsymbol{b}_q$ und $\boldsymbol{a}_{\mathrm{FB}}$ entsprechen. Deren Deklariertheit bildet einen Teil der Vorbedingung V. Zum anderen haben wir die lokalen Variablen, d.h. Variablen, die innerhalb des Funktionsbausteins deklariert werden. Hierzu gehören diejenigen Variablen, die nicht für die nächste Abtastschicht verwendet werden, d.h. die Variablen, die den Vektoren $\boldsymbol{b}_e, \boldsymbol{a}_v^1, \ldots, \boldsymbol{a}_v^6$ entsprechen und die Laufvariablen der for-Schleifen. Wir fassen nun die Forderungen an die gesamte Vorbedingung $\{\mathsf{V}\}$ des Algorithmus zusammen. Die Variablen der Verzögerer $\boldsymbol{b}_v$, der Quellen $\boldsymbol{b}_q$ und der Ausgangssignale $\boldsymbol{a}_{\mathrm{FB}}$ müssen deklariert sein, die Feldlänge P haben und vom Typ `float` sein. Zudem müssen die Werte der Variablen für die Verzögerer und die Quellen zum aktuellen

Zeitpunkt μ_4 und zu jedem Ortspunkt mit den tatsächlichen Werten übereinstimmen. Ferner muss ein gültiges MDWDF vorliegen. Die Vorbedingung V lautet daher

$$\mathsf{V} \equiv \mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge \mathsf{P}_{\mathrm{decaFB}} \wedge \mathsf{P}_{\mathrm{be}} \wedge \mathsf{P}_{\mathrm{name}} \tag{6.192}$$

mit

$$\mathsf{P}_{\mathrm{name}} \equiv \mathsf{P}_{\mathrm{avname}} \wedge \mathsf{P}_{\mathrm{bename}} \wedge \mathsf{P}_{\mathrm{uvname}} \tag{6.193}$$

und

$$\begin{aligned}
\mathsf{P}_{\mathrm{avname}} &\equiv \bigwedge_{\kappa=1}^{6} \bigwedge_{k=1}^{n_v^\kappa} [a_{v\,k}^{\kappa}.name \notin \mathcal{N} \wedge a_{v\,k}^{\kappa}.name \in \mathcal{B}] \ , \\
\mathsf{P}_{\mathrm{bename}} &\equiv \bigwedge_{k=1}^{n_e} [b_{e\,k}.name \notin \mathcal{N} \wedge b_{e\,k}.name \in \mathcal{B}] \ , \\
\mathsf{P}_{\mathrm{uvname}} &\equiv \delta.name \notin \mathcal{N} \wedge \delta.name \in \mathcal{B} \wedge \bigwedge_{\kappa=1}^{3} [\mu_\kappa.name \notin \mathcal{N} \wedge \mu_\kappa.name \in \mathcal{B}] \ .
\end{aligned} \tag{6.194}$$

Die durch S_{dekl} zu erfüllenden Aufgaben können in der zusätzlichen Nachbedingung

$$\mathsf{P}'_{\mathrm{dekl}} \equiv \mathsf{P}_{\mathrm{decbe}} \wedge \mathsf{P}_{\mathrm{decav}} \wedge \mathsf{P}_{\mathrm{decuv}} \tag{6.195}$$

zusammengefasst werden. Zur Erfüllung der Bedingungen werden wir den (hier kommentierten) C-Code

Bedingung	Anweisung	Kommentar
$\{\mathsf{V}\} \equiv$	$\mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge$ $\mathsf{P}_{\mathrm{decaFB}} \wedge \mathsf{P}_{\mathrm{uvname}} \wedge \mathsf{P}_{\mathrm{bename}} \wedge \mathsf{P}_{\mathrm{avname}}$	
	`int delta;`	Deklarationen
	`int mu1;`	der
	`int mu2;`	unabhängigen
	`int mu3;`	Variablen
$\{\mathsf{V}_{\mathrm{deklbe}}\} \equiv$	$\mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge$ $\mathsf{P}_{\mathrm{decaFB}} \wedge \mathsf{P}_{\mathrm{decuv}} \wedge \mathsf{P}_{\mathrm{bename}} \wedge \mathsf{P}_{\mathrm{avname}}$	
	`float b_e_1[` P `];`	Variablendeklarationen
	$\vdots$	für
	`float b_e_`n_e`[` P `];`	den Vektor $\boldsymbol{b}_e$
$\{\mathsf{P}_{\mathrm{deklbe}}\} \equiv$	$\mathsf{P}_{\mathrm{WDF}} \wedge \mathsf{P}_{\mathrm{q}} \wedge \mathsf{P}_{\mathrm{v}} \wedge \mathsf{P}_{\mathrm{decbq}} \wedge \mathsf{P}_{\mathrm{decbv}} \wedge$ $\mathsf{P}_{\mathrm{decaFB}} \wedge \mathsf{P}_{\mathrm{decuv}} \wedge \mathsf{P}_{\mathrm{decbe}} \wedge \mathsf{P}_{\mathrm{avname}}$	
	`float a_v_1_1[` P `];`	Variablendeklarationen
	$\vdots$	für
	`float a_v_1_`n_v^1`[` P `];`	den Vektor $\boldsymbol{a}_v^1$
	`float a_v_2_1[` P `];`	Variablendeklarationen
	$\vdots$	für
	`float a_v_2_`n_v^2`[` P `];`	den Vektor $\boldsymbol{a}_v^2$
	`float a_v_3_1[` P `];`	Variablendeklarationen
	$\vdots$	für

	`float a_v_3_`n_v^3`[ `P` ];`	den Vektor $\boldsymbol{a}_v^3$
	`float a_v_4_1[ `P` ];`	Variablendeklarationen
	$\vdots$	für
	`float a_v_4_`n_v^4`[ `P` ];`	den Vektor $\boldsymbol{a}_v^4$
	`float a_v_5_1[ `P` ];`	Variablendeklarationen
	$\vdots$	für
	`float a_v_5_`n_v^5`[ `P` ];`	den Vektor $\boldsymbol{a}_v^5$
	`float a_v_6_1[ `P` ];`	Variablendeklarationen
	$\vdots$	für
	`float a_v_6_`n_v^6`[ `P` ];`	den Vektor $\boldsymbol{a}_v^6$
$\{\mathsf{P}_{\text{dekl}}\} \equiv$	$\mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}_{\text{decbq}} \wedge \mathsf{P}_{\text{decbv}} \wedge$ $\mathsf{P}_{\text{decaFB}} \wedge \mathsf{P}_{\text{decuv}} \wedge \mathsf{P}_{\text{decbe}} \wedge \mathsf{P}_{\text{decav}}$	

verwenden. Im Folgenden werden wir die Korrektheit des Codes nachweisen. Anwendung von Axiom (5.15) auf die Deklarationsanweisungen für $\boldsymbol{a}_v$ liefert die Vorbedingung von `float a_v_1_1[P];`

$$\mathsf{P}_{\text{dekl}} \begin{bmatrix} a_{v\,1}^{1}.typ \\ \texttt{float} \end{bmatrix} \cdots \begin{bmatrix} a_{v\,n_v^6}^{6}.typ \\ \texttt{float} \end{bmatrix} \begin{bmatrix} a_{v\,1}^{1}.l \\ P \end{bmatrix} \cdots \begin{bmatrix} a_{v\,n_v^6}^{6}.l \\ P \end{bmatrix} \wedge \texttt{float} \in \mathcal{D} \wedge \mathsf{P}_{\text{avname}} \wedge P > 0\,. \tag{6.196}$$

Die Ersetzungen, die an P_{dekl} vorzunehmen sind, betreffen nur $\mathsf{P}_{\text{decav}}$. Der Datentyp `float` existiert, so dass $\texttt{float} \in \mathcal{D}$ eine wahre Aussage ist. Weiterhin berücksichtigen wir $\mathsf{P}_{\text{dekl}} \Rightarrow P > 0$. Daher wird aus Gleichung (6.196)

$$\begin{aligned} &\mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}_{\text{decbq}} \wedge \mathsf{P}_{\text{decbv}} \wedge \mathsf{P}_{\text{decaFB}} \wedge \mathsf{P}_{\text{decuv}} \wedge \mathsf{P}_{\text{decbe}} \wedge \mathsf{P}_{\text{avname}} \wedge \\ &\bigwedge_{\kappa=1}^{6} \bigwedge_{l=1}^{n_v^\kappa} [\,\texttt{float} = \texttt{float} \wedge P \leq P\,]\,. \end{aligned} \tag{6.197}$$

Die wahren Aussagen entfallen aufgrund des Prädikatenkalküls, so dass Gleichung (6.197), wie behauptet, $\mathsf{P}_{\text{deklbe}}$ darstellt.

Die Vorbedingung zur Anweisung `float b_e_1[P];` ergibt sich durch Anwendung von Axiom (5.15) zu

$$\mathsf{P}_{\text{deklbe}} \begin{bmatrix} b_{e\,1}.typ \\ \texttt{float} \end{bmatrix} \cdots \begin{bmatrix} b_{e\,n_e}.typ \\ \texttt{float} \end{bmatrix} \begin{bmatrix} b_{e\,1}.l \\ P \end{bmatrix} \cdots \begin{bmatrix} b_{e\,n_e}.l \\ P \end{bmatrix} \wedge \texttt{float} \in \mathcal{D} \wedge \mathsf{P}_{\text{bename}} \wedge P > 0\,. \tag{6.198}$$

Mit gleicher Argumentation wie zuvor erhalten wir aus Gleichung (6.198)

$$\begin{aligned} &\mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}_{\text{decbq}} \wedge \mathsf{P}_{\text{decbv}} \wedge \mathsf{P}_{\text{decaFB}} \wedge \mathsf{P}_{\text{decuv}} \wedge \\ &\bigwedge_{l=1}^{n_e} [\,\texttt{float} = \texttt{float} \wedge P \leq P \leq \text{Max}_{\text{i}}\,] \wedge \mathsf{P}_{\text{avname}} \wedge \mathsf{P}_{\text{bename}}\,. \end{aligned} \tag{6.199}$$

Die wahren Aussagen entfallen wiederum auf Grund des Prädikatenkalküls, und es verbleibt von Gleichung (6.199), wie behauptet, $\mathsf{V}_{\text{declbe}}$.

Die eigentliche Vorbedingung V des Algorithmus bestimmt sich durch Anwendung von Axiom (5.12) zu

$$\begin{aligned} &\mathsf{V}_{\text{deklbe}} \begin{bmatrix} \delta\,.typ \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_1\,.typ \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_2\,.typ \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_3\,.typ \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \delta\,.l \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_1\,.l \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_2\,.l \\ \texttt{int} \end{bmatrix} \begin{bmatrix} \mu_3\,.l \\ \texttt{int} \end{bmatrix} \\ &\wedge \texttt{int} \in \mathcal{D} \wedge \mathsf{P}_{\text{uvname}} \wedge P > 0\,. \end{aligned} \tag{6.200}$$

Der Datentyp `int` existiert, so dass $\texttt{int} \in \mathcal{D}$ eine wahre Aussage ist. Die Ersetzungen, die an $\mathsf{V}_{\text{deklbe}}$ vorzunehmen sind, betreffen nur $\mathsf{P}_{\text{decuv}}$, so dass Gleichung (6.200) ausführlich

$$\mathsf{V} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}_{\text{decbq}} \wedge \mathsf{P}_{\text{decbv}} \wedge \mathsf{P}_{\text{decaFB}} \wedge \mathsf{P}_{\text{uvname}} \wedge \mathsf{P}_{\text{bename}} \wedge \mathsf{P}_{\text{avname}} \wedge \texttt{int} = \texttt{int} \wedge -1 = -1 \tag{6.201}$$

ist. Die wahren Aussagen entfallen wiederum auf Grund des Prädikatenkalküls und liefern die Vorbedingung für den Algorithmus, wie behauptet, zu

$$\mathsf{V} \equiv \mathsf{P}_{\text{WDF}} \wedge \mathsf{P}_{\text{q}} \wedge \mathsf{P}_{\text{v}} \wedge \mathsf{P}_{\text{decbq}} \wedge \mathsf{P}_{\text{decbv}} \wedge \mathsf{P}_{\text{decaFB}} \wedge \mathsf{P}_{\text{uvname}} \wedge \mathsf{P}_{\text{bename}} \wedge \mathsf{P}_{\text{avname}} \,. \tag{6.202}$$

6.8 Zusammenfassung

Wir wollen nun die wesentlichen Ergebnisse dieses Kapitels zusammenfassen und bewerten. Es wurde ein Algorithmus durch Verknüpfen von Elementaranweisungen der Programmiersprache C angegeben, der ein lineares, konstantes MDWDF für ein quaderförmiges Berechnungsgebiet simuliert. Die Anweisungssequenz, die eine Abtastschicht berechnet, ist dabei so geartet, dass sie einen, gemäß Kapitel 3.5 spezifizierten, Funktionsbaustein des Programmpaketes SPACE darstellt. Weiterhin erfolgte die Niederlegung der Anforderungen mittels einer formalen Spezifikation. Durch konsequente Anwendung der aus der Literatur bekannten und im Kapitel 5 dargelegten Ergebnisse wurde aus der Nachbedingung und dem Algorithmus eine geeignete Vorbedingung ermittelt. Die Tatsache, dass auch das Nicht-Überschreiten des Rechenbereiches durch eine geeignete Vorbedingung sichergestellt werden kann, hebt dieses Verfahren gegenüber anderen hervor. Mit diesem formalen Korrektheitsbeweis ist nunmehr das Gebiet der numerischen Integration partieller Differentialgleichungen auch für sicherheitskritische Anwendungen erschlossen worden.

Kapitel 7

Zusammenfassung

Die vorliegende Arbeit befasst sich vornehmlich mit zwei bisher nicht behandelten Aufgabenstellungen aus dem Bereich der mehrdimensionalen Wellendigitalfilter. Zum einen mit der systematischen Erzeugung von Simulations-Algorithmen, ausgehend von einem System linearer, konstanter partieller Differentialgleichungen. Den zweiten bildet der Nachweis der Korrektheit des verwendeten Programmcodes, der die Simulations-Algorithmen implementiert.

Im Kapitel 2 wurden zunächst die wichtigsten für die vorliegende Arbeit notwendigen Grundlagen der mehrdimensionalen Wellendigitalfilter wiederholt. Dabei wurde auf eine Darstellungsform Wert gelegt, die den Aufgaben der folgenden Kapitel Rechnung trägt.

Im Kapitel 3 wurde das Programmpaket SPACE beschrieben und die Einbindung von mehrdimensionalen Wellendigitalfiltern erörtert. Dabei zeigte sich, dass die vorhandenen Schnittstellen der einzelnen Softwaremodule, des zur Behandlung eindimensionaler Probleme vorgesehenen Programmpaketes SPACE, ungeeignet für unser Ziel sind. Dem Problem sind wir durch Erweiterung der Signale des Programmpaketes SPACE auf vektorielle Signale (Felder) entgegengetreten. Diese erforderliche Abänderung der vorhandenen Schnittstelle und die daraus resultierende Möglichkeit der Einbindung der mehrdimensionalen Wellendigitalfilter sind ebenfalls im Kapitel 3 niedergelegt.

Im Kapitel 4 wurde zunächst durch Passivitätsbetrachtungen die Klasse der PDGLn eingeschränkt, die behandelt werden. Für diese Klasse PDGLn wurde ein Verfahren zur automatischen Synthese von Referenzschaltung und Wellendigitalfilter entwickelt, welches sich somit auch zur automatischen Codeerzeugung für eine Simulation des zu Grunde liegenden Systems mittels mehrdimensionaler Wellendigitalfilter eignet. Aspekte wie die Genauigkeit des eingesetzten numerischen Integrationsverfahrens und die Effizienz der Implementierung rückten hierbei in den Hintergrund. Vielmehr war es das Ziel ein numerisches Integrationsverfahren, welches die Möglichkeit der formalen Verifikation bietet, zu erhalten.

Die aus der Literatur bekannten, anerkannten und für die vorliegende Arbeit notwendigen Grundlagen der Software-Verifikation mittels formaler Methoden wurden im Kapitel 5 rekapituliert. Die Definitionen der notwendigen Programmanweisungen finden sich ebenfalls in diesem Kapitel.

Im Kapitel 6 wurde aus den zu Grunde liegenden Anforderungen eine formale Spezifikation entwickelt. Im Anschluss daran wurde ein Programmcode zur Simulation des mehrdimensionalen Wellendigitalfilters festgelegt. Dabei wurde der Entwurf des Codes so durchgeführt, dass dieser sich in das Programmpaket SPACE einbinden lässt. Im Anschluss daran erfolgte für diesen vorliegenden Programmcode der Nachweis der Korrektheit mittels formaler Methoden. Es zeigt sich, dass auch bei anspruchsvolleren Problemstellungen der digitalen Signalverarbeitung eine Beweisführung der Korrektheit relativ einfach durchzuführen ist. Im Zusammenhang mit der Beschränktheit der verwendeten Signale hat die Passivität des eingangs betrachteten Systems ein wesentlichen Anteil daran. Viele der für dieses Kapitel notwendigen Berechnungen und Abschätzungen in Bezug auf die Beschränktheit der verwendeten Signale, insbesondere im Falle von endlicher Signalwortlänge, wurden in den Anhang ausgelagert.

Durch die vorliegende Arbeit ergeben sich i. W. zwei neue Ergebnisse zur numerischen Integration von

PDGLn mithilfe mehrdimensionaler Wellendigitalfilter. Zum einen kommt der effizienten Codeerzeugung in der Praxis eine erhebliche Bedeutung zu, der durch das entwickelte automatische Syntheseverfahren ermöglicht wird. Auf Basis dieses Verfahrens wird zur Zeit am Lehrstuhl für Nachrichtentechnik ein Werkzeug zur Erzeugung von C-Code entwickelt [BV03]. Nach Vorliegen dieses Simulations-Werkzeugs wird sich der Anwenderkreis des numerischen Integrationsverfahrens nach der Wellendigitalmethode erheblich erweitern. Zum anderen eröffnet der geführte formale Korrektheitsbeweis die Möglichkeit der Anwendung des Integrationsverfahrens in sicherheitskritischen Systemen.

In weiterführenden Untersuchungen könnte das vorgestellte Syntheseverfahren auf quasilineare, symmetrisch hyperbolische PDGL erweitert werden. Hierbei wird man zunächst nicht den allgemeinen Fall behandeln, sondern Einschränkungen an die PDGL machen, insbesondere in Bezug auf die Abhängigkeiten der Koeffizienten von der Zeit. Ferner könnte geklärt werden, ob die Führung eines Korrektheitsbeweises für ein Werkzeug zur automatischen Codeerzeugung möglich ist. Weiterhin wäre eine Skalierung des Algorithmus, in Teilalgorithmen entsprechend der Teilberechnungsgebiete inklusive Kommunikation zwischen den Rechnern sinnvoll. In diesem Zusammenhang empfehlen sich auch Untersuchungen zur Behandlung von beliebig gearteten Berechnungsgebieten mit dem vorliegenden Verfahren. Die automatische Codeerzeugung für den nichtlinearen Fall stellt eine weitere interessante Aufgabenstellung dar. Als besondere Herausforderung dürfte sich in diesem Fall der Nachweis der Korrektheit der zu implementierenden Algorithmen erweisen, die der Lösung der algebraischen, nichtlinearen Gleichungen dienen.

Anhang A

Die Operatoren min und max

Die folgenden Sätze erscheinen trivial, ermöglichen aber ermüdungsfreies Arbeiten. α, β sollen reelle Zahlen sein und x, x_μ, $\mathrm{Max_f}$ sollen positiv sein.

$$\min\left\{x, \frac{1}{x}\right\} = \frac{1}{\max\{x, \frac{1}{x}\}} \tag{A.1}$$

$$\min_\mu\left\{\frac{1}{x_\mu}\right\} = \frac{1}{\max\limits_\mu\{x_\mu\}} \tag{A.2}$$

$$\min_\mu\left\{\min\left\{x_\mu, \frac{1}{x_\mu}\right\}\right\} = \frac{1}{\max\limits_\mu\left\{\max\left\{x_\mu, \frac{1}{x_\mu}\right\}\right\}} \tag{A.3}$$

$$\max\left\{x, \frac{1}{x}\right\} \geq 1 \ , \ \min\left\{x, \frac{1}{x}\right\} \leq 1 \tag{A.4}$$

$$\begin{aligned} [\quad \alpha \cdot x < \mathrm{Max_f} \ \wedge \ 1 \leq \alpha \quad] \quad &\Longrightarrow \quad [\quad x < \mathrm{Max_f} \quad] \\ [\quad x < \beta \cdot \mathrm{Max_f} \ \wedge \ \beta \leq 1 \quad] \quad &\Longrightarrow \quad [\quad x < \mathrm{Max_f} \quad] \end{aligned} \tag{A.5}$$

Anhang B

Unitär beschränkte Matrizen

Wir wollen eine $m \times n$ Matrix $\boldsymbol{S}$ als unitär beschränkt (oder subunitär) bezeichnen, wenn sie

$$\|\boldsymbol{S}\boldsymbol{a}\|^2 \leq \|\boldsymbol{a}\|^2 \; \forall \; \boldsymbol{a} \in \mathbb{C}^n \tag{B.1}$$

genügt. Es ergeben sich für die hermitesch konjugierte Matrix $\boldsymbol{S}^{\mathrm{H}}$ und die um die Spalte ν reduzierte Matrix ${}^{-\nu}\boldsymbol{S}$ die folgenden Eigenschaften

$$\begin{array}{ccccc}
\boxed{\|\boldsymbol{S}\boldsymbol{a}\|^2 \leq \|\boldsymbol{a}\|^2 \; \forall \; \boldsymbol{a}} & \Leftrightarrow & \boxed{\boldsymbol{D} = \mathbf{1} - \boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} \geq 0} & \Leftrightarrow & \boxed{\tilde{\boldsymbol{D}} = \mathbf{1} - \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} \geq 0} \; \Leftrightarrow \; \boxed{\left\|\boldsymbol{S}^{\mathrm{H}}\boldsymbol{b}\right\|^2 \leq \|\boldsymbol{b}\|^2 \; \forall \; \boldsymbol{b}} \\
\Downarrow & & \Downarrow & & \Downarrow \qquad\qquad\qquad \Downarrow \\
\boxed{\left\|{}^{-\nu}\boldsymbol{S}_{-\nu}\boldsymbol{a}\right\|^2 \leq \left\|{}_{-\nu}\boldsymbol{a}\right\|^2 \; \forall \; {}_{-\nu}\boldsymbol{a}} & \Leftrightarrow & \boxed{{}^{-\nu}_{-\nu}\boldsymbol{D} \geq 0} & & \boxed{{}^{-\nu}_{-\nu}\tilde{\boldsymbol{D}} \geq 0} \; \Leftrightarrow \; \boxed{\left\|{}^{-\nu}\boldsymbol{S}^{\mathrm{H}}_{-\nu}\boldsymbol{b}\right\|^2 \leq \left\|{}_{-\nu}\boldsymbol{b}\right\|^2 \; \forall \; {}_{-\nu}\boldsymbol{b}\,.}
\end{array} \tag{B.2}$$

Anhang C

Beschränktheit der Wellengrößen

Im Kapitel 6 benötigten wir einige Sätze im Zusammenhang mit der Beschränktheit der Programmvariablen. Diese Sätze werden in diesem Anhang nachgeliefert. Um eine übersichtliche Darstellung zu gewähren, ersetzen wir die Programmvariablen durch ihre PL-Variablen. Das Postfix $.v$ entfällt somit regelmäßig.

Ziel dieses Kapitels ist es, unter der Voraussetzung der Beschränktheit von $\|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2$, die Beschränktheit der Wellen $b_{e\mu}$ und der zur Berechnung von $b_{e\mu}$ benutzten Zwischengrößen nachzuweisen. Zunächst wird eine Relation zwischen den oben angegebenen Größen für die Leistungswellen hergeleitet, um daraus eine Beziehung für die Spannungswellen zu bestimmen. Der Beweis ist dem in [Rumm98] für unitäre Matrizen $\boldsymbol{L}'$ geführten ähnlich, wobei wir im Folgenden $\boldsymbol{L}$ anstatt $\boldsymbol{L}'$ schreiben wollen. In unserem Fall ist $\boldsymbol{L}$ eine unitär beschränkte Matrix, was nun gezeigt wird. Dazu bereiten wir zunächst

$$\boldsymbol{L}\,\boldsymbol{L}^{\mathrm{H}} = \boldsymbol{P}_b \boldsymbol{S} \left[\, \boldsymbol{P}_{eq} \;\; \boldsymbol{P}_{ev} \;\; (\boldsymbol{P}_{ee}\,\boldsymbol{P}_b^{\mathrm{T}}) \,\right] \begin{bmatrix} \boldsymbol{P}_{eq}^{\mathrm{T}} \\ \boldsymbol{P}_{ev}^{\mathrm{T}} \\ \boldsymbol{P}_b\,\boldsymbol{P}_{ee}^{\mathrm{T}} \end{bmatrix} \boldsymbol{S}^{\mathrm{H}}\,\boldsymbol{P}_b^{\mathrm{T}} \tag{C.1}$$

auf. Mit $\boldsymbol{P}_b^{\mathrm{T}}\boldsymbol{P}_b = \boldsymbol{1}$ und $\boldsymbol{P}_e\boldsymbol{P}_e^{\mathrm{T}} = \boldsymbol{1}$ erhalten wir

$$\boldsymbol{L}\,\boldsymbol{L}^{\mathrm{H}} = \boldsymbol{P}_b\,\boldsymbol{S}\,\boldsymbol{S}^{\mathrm{H}}\,\boldsymbol{P}_b^{\mathrm{T}}\;. \tag{C.2}$$

Aus der Passivität folgt die unitäre Beschränktheit von $\boldsymbol{S}$. Wegen Gleichung (B.2) ist auch $\boldsymbol{S}^{\mathrm{H}}$ unitär beschränkt, d. h. es gilt

$$\boldsymbol{x}^{\mathrm{H}} \left[\, \boldsymbol{1} - \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} \,\right] \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x}\;. \tag{C.3}$$

Aufgrund der Unitarität von $\boldsymbol{P}_b$ ist dies äquivalent zu

$$\boldsymbol{x}^{\mathrm{H}}\,\boldsymbol{P}_b \left[\, \boldsymbol{1} - \boldsymbol{S}\,\boldsymbol{S}^{\mathrm{H}} \,\right] \boldsymbol{P}_b^{\mathrm{T}}\,\boldsymbol{x} = \boldsymbol{x}^{\mathrm{H}} \left[\, \boldsymbol{1} - \boldsymbol{P}_b\,\boldsymbol{S}\,\boldsymbol{S}^{\mathrm{H}}\,\boldsymbol{P}_b^{\mathrm{T}} \,\right] \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x}\;. \tag{C.4}$$

Nutzen von Gleichung (C.2) liefert zunächst

$$\boldsymbol{x}^{\mathrm{H}} \left[\, \boldsymbol{1} - \boldsymbol{L}\,\boldsymbol{L}^{\mathrm{H}} \,\right] \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \tag{C.5}$$

und bestätigt mit Gleichung (B.2) die behauptete Aussage.

Die Koordinate μ aus dem Spaltenvektor $\boldsymbol{b}_e = \boldsymbol{L}_q\,\boldsymbol{b}_q + \boldsymbol{L}_v\,\boldsymbol{b}_v + \boldsymbol{L}_e\,\boldsymbol{b}_e$ berechnet sich zu

$$b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \sum_{\nu=1}^{n_e} l_{e\,\mu\nu}\; b_{e\nu}\;. \tag{C.6}$$

Wir definieren nun den Vektor

$$ {}^m\boldsymbol{b}_e = \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{en_e} \end{bmatrix} \tag{C.7} $$

dessen Koordinate μ aus den ersten $n_q + n_v + m$ Summanden von $b_{e\mu}$ bestehen soll. Dabei unterscheiden wir für m drei Fälle. Zunächst soll sich die Summation nur auf die gewichteten Quellenwellen beschränken, d. h.

$$ {}^m b_{e\mu} = \sum_{\nu=1}^{m+n_q+n_v} l_{q\,\mu\nu}\; b_{q\nu}\,, \tag{C.8} $$

wobei hier $m = 1 - n_q - n_v, 2 - n_q - n_v, \ldots, -1 - n_v, -n_v$ gilt. Da die Matrix $\boldsymbol{L}$ unitär beschränkt ist, gilt dies auch für deren Untermatrizen, d. h.

$$ |{}^m b_{e\mu}|^2 = |\sum_{\nu=1}^{m+n_q+n_v} l_{q\,\mu\nu}\; b_{q\nu}|^2 \leq \sum_{\nu=1}^{m+n_q+n_v} |b_{q\nu}|^2\,. \tag{C.9} $$

Berücksichtigen wir nun

$$ \sum_{\nu=1}^{m+n_q+n_v} |b_{q\nu}|^2 \leq \|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2 < \frac{{\mathrm{Max_f}}^2}{4\beta^2}, \tag{C.10} $$

so gilt tatsächlich

$$ |{}^m b_{e\mu}|^2 < \frac{{\mathrm{Max_f}}^2}{4\beta^2}\,. \tag{C.11} $$

Im zweiten Fall erfolgt die Summation von allen gewichteten Quellenwellen und über gewichtete Verzögererwellen, d. h.

$$ {}^m b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \sum_{\nu=1}^{m+n_v} l_{v\,\mu\nu}\; b_{v\nu}\,, \tag{C.12} $$

wobei hier $m = 1 - n_v, 2 - n_v, \ldots, -1, 0$ gilt. Da die Matrix $\boldsymbol{L}$ unitär beschränkt ist, gilt dies auch für deren Untermatrizen, d. h.

$$ |{}^m b_{e\mu}|^2 = |\boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \sum_{\nu=1}^{m+n_v} l_{v\,\mu\nu}\; b_{v\nu}|^2 \leq \|\boldsymbol{b}_q\|^2 + \sum_{\nu=1}^{m+n_v} |b_{v\nu}|^2\,. \tag{C.13} $$

Berücksichtigen wir nun

$$ \|\boldsymbol{b}_q\|^2 + \sum_{\nu=1}^{m+n_v} |b_{q\nu}|^2 \leq \|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2 < \frac{{\mathrm{Max_f}}^2}{4\beta^2}, \tag{C.14} $$

so gilt tatsächlich auch im zweiten Fall

$$ |{}^m b_{e\mu}|^2 < \frac{{\mathrm{Max_f}}^2}{4\beta^2}\,. \tag{C.15} $$

Der dritte Fall ist der komplizierteste. Für ihn gilt

$$ {}^{m}b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \sum_{\nu=1}^{m} l_{e\,\mu\nu}\; b_{e\nu}\,, \tag{C.16} $$

wobei m hier die Werte $1, 2, \ldots, n_e$ annehmen darf.

Wegen der strikten unteren Dreiecksgestalt von $\boldsymbol{L}_e$ gilt $l_{e\,\mu\nu} = 0$ für $\nu \geq \mu$. Nutzen wir dies für die Berechnung von $b_{e\mu}$ aus, so lautet Gleichung (C.6)

$$ b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \sum_{\nu=1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu}\,. \tag{C.17} $$

Sinnvoll ist es hier, zwei Fälle zu unterscheiden

$$ b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \sum_{\nu=1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \begin{cases} \displaystyle\sum_{\nu=1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu} & \text{für } m \geq \mu-1 \\ \displaystyle\sum_{\nu=1}^{m} l_{e\,\mu\nu}\; b_{e\nu} + \sum_{\nu=m+1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu} & \text{für } m < \mu-1\,. \end{cases} \tag{C.18} $$

Für die Teilsummen gilt entsprechend

$$ {}^{m}b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \begin{cases} \displaystyle\sum_{\nu=1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu} & \text{für } m \geq \mu-1 \\ \displaystyle\sum_{\nu=1}^{m} l_{e\,\mu\nu}\; b_{e\nu} & \text{für } m < \mu-1 \text{ (und für } m \geq \mu-1)\,. \end{cases} \tag{C.19} $$

Die Differenz aus Gleichung (C.18) und Gleichung (C.19) lautet

$$ b_{e\mu} - {}^{m}b_{e\mu} = \begin{cases} 0 & \text{für } m \geq \mu-1 \\ \displaystyle\sum_{\nu=m+1}^{\mu-1} l_{e\,\mu\nu}\; b_{e\nu} & \text{für } m < \mu-1\,. \end{cases} \tag{C.20} $$

Offenbar stimmen die ersten $m+1$ Koordinaten der Vektoren $\boldsymbol{b}_e$ und ${}^{m}\boldsymbol{b}_e$ überein, d. h.

$$ \begin{bmatrix} b_{e1} \\ \vdots \\ b_{em+1} \end{bmatrix} = \begin{bmatrix} {}^{m}b_{e1} \\ \vdots \\ {}^{m}b_{em+1} \end{bmatrix} \quad \text{bzw.} \quad [\,\mathbf{1}_{m+1} \quad \mathbf{0}\,]\boldsymbol{b}_e = [\,\mathbf{1}_{m+1} \quad \mathbf{0}\,]\,{}^{m}\boldsymbol{b}_e\,. \tag{C.21} $$

Umbenennung der Indizes von μ nach ν liefert ${}^{m}b_{e\nu} = b_{e\nu}$ für $\nu \leq m+1$. Somit können wir in Gleichung (C.16) jede Welle $b_{e\nu}$ durch ${}^{m}b_{e\nu}$ ersetzen d. h.

$$ {}^{m}b_{e\mu} = \boldsymbol{l}_{q\,\mu,\bullet}\boldsymbol{b}_q + \boldsymbol{l}_{v\,\mu,\bullet}\boldsymbol{b}_v + \sum_{\nu=1}^{m} l_{e\,\mu\nu}\; {}^{m}b_{e\nu}\,. \tag{C.22} $$

Zu beachten ist, dass diese Gleichung für alle μ und m gilt, da die Voraussetzung $\nu \leq m+1$ sich nur auf den lokalen Summenindex ν bezog. Auf den ersten Blick mag diese Formel verwundern, da auf der

rechten Gleichungsseite nun auch die Teilsummen auftreten. Dies liegt daran, dass zur Berechnung von ${}^m b_{e\mu}$ die Wellen b_{e1} bis b_{em} benötigt werden. Diese Wellen sind aber mit den Teilsummen ${}^m b_{e1}$ und ${}^m b_{em}$ identisch. Somit hängt ${}^m b_{e\mu}$ auch nur von ${}^m b_{e1}$ bis ${}^m b_{em}$ ab.

Gleichung (C.22) kann auch in Matrixschreibweise ermittelt werden. Ausgehend von Gleichung (C.16) für $\mu = 1, \ldots, n_e$

$$ {}^m\boldsymbol{b}_e = \boldsymbol{L}_q\,\boldsymbol{b}_q + \boldsymbol{L}_v\,\boldsymbol{b}_v + \boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{n_e-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{b}_e \tag{C.23} $$

wenden wir die aus Gleichung (C.21) unmittelbar folgende Gleichung $[\,\mathbf{1}_m \quad \mathbf{0}_m^{n_e-m}\,]\boldsymbol{b}_e = [\,\mathbf{1}_m \quad \mathbf{0}_m^{n_e-m}\,]\,{}^m\boldsymbol{b}_e$ an und erhalten

$$ {}^m\boldsymbol{b}_e = \boldsymbol{L}_q\,\boldsymbol{b}_q + \boldsymbol{L}_v\,\boldsymbol{b}_v + \boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{n_e-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} {}^m\boldsymbol{b}_e\,. \tag{C.24} $$

Von dieser Gleichung betrachten wir nun nur die ersten μ Zeilen, d. h.

$$ [\,\mathbf{1}_\mu \quad \mathbf{0}_\mu^{n_e-\mu}\,]\,{}^m\boldsymbol{b}_e = [\,\mathbf{1}_\mu \quad \mathbf{0}_\mu^{n_e-\mu}\,]\,[\boldsymbol{L}_q\,\boldsymbol{b}_q + \boldsymbol{L}_v\,\boldsymbol{b}_v] + [\,\mathbf{1}_\mu \quad \mathbf{0}_\mu^{n_e-\mu}\,]\,\boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{n_e-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} {}^m\boldsymbol{b}_e\,. \tag{C.25} $$

Wir wollen nun zeigen, dass

$$ |{}^m b_{e\mu}|^2 \le \|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2 \tag{C.26} $$

gilt. Dazu betrachten wir zunächst den Fall $m \le \mu - 1 \Leftrightarrow m + 1 \le \mu$. Für $m \le \mu - 1$ gilt offenbar

$$ \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{n_e-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} {}^m\boldsymbol{b}_e = \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{n_e-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{em} \\ \vdots \\ {}^m b_{en_e} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{\mu-1-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{em} \\ \vdots \\ {}^m b_{e\,\mu-1} \end{bmatrix}, \tag{C.27} $$

so dass aus Gleichung (C.25)

$$ \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{e\mu} \end{bmatrix} = [\,\mathbf{1}_\mu \quad \mathbf{0}\,][\boldsymbol{L}_q\boldsymbol{b}_q + \boldsymbol{L}_v\boldsymbol{b}_v] + [\,\mathbf{1}_\mu \quad \mathbf{0}\,]\boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{\mu-1-m} \\ \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{em} \\ \vdots \\ {}^m b_{e\,\mu-1} \end{bmatrix} \tag{C.28} $$

wird. Um diese Gleichung übersichtlicher schreiben zu können, führen wir die Vektoren

$$ {}^{m,\mu}\boldsymbol{b}_e = \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{e\,\mu} \end{bmatrix} \quad , \quad {}^{m,\mu-1}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ {}^{m,\mu-1}\boldsymbol{b}_e \end{bmatrix} \tag{C.29} $$

und die Matrix ${}^{\mu,m}\boldsymbol{L}$ als die linke obere Untermatrix von $\boldsymbol{L}$ mit μ Zeilen und $n_q + n_v + \mu - 1$ Spalten, d. h.

$$ {}^{\mu,m}\boldsymbol{L} = [\,\mathbf{1}_\mu \quad \mathbf{0}] \left[\boldsymbol{L}_q \quad \boldsymbol{L}_v \quad \boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{\mu-1-m} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right] \tag{C.30} $$

ein. Damit erhalten wir Gleichung (C.28) zunächst in der Form

$$ {}^{m,\mu}\boldsymbol{b}_e = [\,\mathbf{1}_\mu \quad \mathbf{0}\,][\boldsymbol{L}_q\boldsymbol{b}_q + \boldsymbol{L}_v\boldsymbol{b}_v] + [\,\mathbf{1}_\mu \quad \mathbf{0}\,]\boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{\mu-1-m} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} {}^{m,\mu-1}\boldsymbol{b}_e \tag{C.31} $$

und weiter

$$ {}^{m,\mu}\boldsymbol{b}_e = [\,\mathbf{1}_\mu \quad \mathbf{0}] \left[\boldsymbol{L}_q \quad \boldsymbol{L}_v \quad \boldsymbol{L}_e \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m^{\mu-1-m} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right] {}^{m,\mu-1}\boldsymbol{b} = {}^{\mu,m}\boldsymbol{L}\ {}^{m,\mu-1}\boldsymbol{b}\,. \tag{C.32} $$

Wegen der unitären Beschränktheit von $\boldsymbol{L}$ ist nach Gleichung (B.2) auch die Teilmatrix ${}^{\mu,m}\boldsymbol{L}$ unitär beschränkt, d. h.

$$ \left\| {}^{\mu,m}\boldsymbol{L}\ {}^{m,\mu-1}\boldsymbol{b}\right\|^2 \le \left\|{}^{m,\mu-1}\boldsymbol{b}\right\|^2 \quad \forall\ {}^{m,\mu-1}\boldsymbol{b}\,. \tag{C.33} $$

Setzen wir ${}^{m,\mu}\boldsymbol{b}_e = {}^{\mu,m}\boldsymbol{L}\ {}^{m,\mu-1}\boldsymbol{b}$ aus Gleichung (C.32) ein, so haben wir

$$ \sum_{\nu=1}^{\mu} |{}^m b_{e\nu}|^2 \le \|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2 + \sum_{\nu=1}^{\mu-1} |{}^m b_{e\nu}|^2\,, \tag{C.34} $$

also das gewünschte Ergebnis

$$ |{}^m b_{e\mu}|^2 \le \|\boldsymbol{b}_q\|^2 + \|\boldsymbol{b}_v\|^2\ ,\ |{}^m b_{e\mu}|^2 < \frac{{\mathrm{Max_f}}^2}{4\beta^2}\ , \tag{C.35} $$

die gesuchte Obergrenze von $|{}^m b_{e\mu}|$, welche für μ, m mit $m+1 \le \mu$ gültig ist .

Wir zeigen nun, dass sie auch für $m+1 > \mu$ gültig ist. Gleichung (C.22) lautet in Matrixform

$$ \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{e\mu} \end{bmatrix} = [\,\mathbf{1}_\mu \quad \mathbf{0}\,]\boldsymbol{L} \begin{bmatrix} \mathbf{1}_{n_q+n_v+\mu-1} & \mathbf{0} \\ & \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{b} = [\,\mathbf{1}_\mu \quad \mathbf{0}\,]\boldsymbol{L} \begin{bmatrix} \mathbf{1}_{n_q+n_v+\mu-1} \\ \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ {}^m b_{e1} \\ \vdots \\ {}^m b_{e\,\mu-1} \end{bmatrix}\,. \tag{C.36} $$

Die unitäre Beschränktheit der ersten μ Zeilen und der ersten $n_q + n_v + \mu - 1$ Spalten von $\boldsymbol{L}$ liefert

$$ \left\| \begin{bmatrix} {}^m b_{e1} \\ \vdots \\ {}^m b_{e\mu} \end{bmatrix} \right\|^2 = \left\| [\,\mathbf{1}_\mu \quad \mathbf{0}\,]\boldsymbol{L} \begin{bmatrix} \mathbf{1}_{n_q+n_v+\mu-1} \\ \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ {}^m b_{e1} \\ \vdots \\ {}^m b_{e\,\mu-1} \end{bmatrix} \right\|^2 \le \left\| \begin{bmatrix} \boldsymbol{b}_q \\ \boldsymbol{b}_v \\ {}^m b_{e1} \\ \vdots \\ {}^m b_{e\,\mu-1} \end{bmatrix} \right\|^2\,, \tag{C.37} $$

was wieder Gleichung (C.35) aber diesmal gültig für $m+1 \le \mu$, darstellt.

Um dieses Resultat auf die Spannungswellen zu übertragen, nutzen wir die Beziehung $\boldsymbol{b} = \frac{1}{2}\,\mathbf{diag}\left\{\boldsymbol{G}_q^{1/2}, \boldsymbol{G}_v^{1/2}, \boldsymbol{G}_e^{1/2}\right\}\boldsymbol{b}'$ und ersetzen die Leistungswellen in Gleichung (C.35)

$$ |{}^m b'_{e\mu}|^2 \le \left\|(R_{e\mu}\boldsymbol{G}_q)^{1/2}\boldsymbol{b}'_q\right\|^2 + \left\|(R_{e\mu}\boldsymbol{G}_v)^{1/2}\boldsymbol{b}'_v\right\|^2. \tag{C.38} $$

Anhang D

Auswirkungen endlicher Rechengenauigkeit

Anhang C ging von unendlicher Rechengenauigkeit aus (ideale Rechenoperationen). Bekannt ist, dass passive Bauelemente auch bei endlicher Rechengenauigkeit durch Auswahl eines geeigneten Rundungsverfahrens noch passiv sind, [FM75], [Fett78], [Fett90]. In diesem Abschnitt werden praxisgerechte Implementierungen der Arithmetik für verschiedene Zielsysteme vorgestellt. Die zentrale Frage ist, wie die Subunitarität der Streumatrizen nichtreaktiver Bauelemente unter realen Bedingungen gewährleistet werden kann. Untersuchungen zu diesem Thema, bei Benutzung von Festkomma-Arithmetik, finden sich in [Meer79]. In dieser Arbeit steht die Gleitkomma-Arithmetik im Mittelpunkt und wird in diesem Anhang ausschließlich betrachtet. Dazu werden nur WDF-Realisierungen eines (im Idealfall energieneutralen) Zweitors untersucht, da jede unitäre Streumatrix auf eine Realisierung zurückgeführt werden kann, die aus eben diesen Zweitor-Streumatrizen besteht, vgl. Kapitel 2.8.1. Die Untersuchungen werden für die gängigen Rundungsarten des Betragsschneidens und des Gitterpunktschneidens durchgeführt. Es werden für beide Rundungsarten Realisierungen angegeben, die unter realen Bedingungen passiv sind.

D.1 Zahlendarstellungen und Rundungsfehler

Darstellung von Gleitkommazahlen

Wir legen für diese Arbeit die Darstellung einer Gleitkommazahl (Floating-Point-Number) nach dem Standard IEEE 754-1985 zu Grunde. Die Menge aller von null verschiedenen Gleitkommazahlen wird durch

$$\tilde{z}_i = (-1)^{s_i}[1 + f_i 2^{-p+1}]\, 2^{e_i - v}\ ,\ f_i = \sum_{\nu=1}^{p-1} b_{i\nu} 2^{\nu-1}\ ,\ b_{i\nu} \in \{0,1\} \tag{D.1}$$

gebildet. Hierin ist das Zahlensystem durch

- die Anzahl der Bits für den Betrag der Mantisse $p-1$,
- die Anzahl der Bits für den Exponent q,
- die Gesamtzahl der Bits $p+q$ und
- der Verschiebung $v = 2^{q-1} - 1$

bestimmt. Für Zahlen einfacher Länge (Datentyp `float` in der Programmiersprache C) gilt $p = 24$ und $q = 8$. Eine konkrete Zahl $\neq 0$ ist dann zum einen bestimmt durch die Mantisse $(-1)^{s_i}[1 + f_i 2^{-p+1}]$ bestehend aus dem Vorzeichenbit $s_i \in \{0, 1\}$ und dem informationstragenden Teil des Betrages (der so genannten Charakteristik) f_i und zum anderen durch den Exponenten $2^{e_i - v}$ mit $0 < e_i < 2v+1 = 2^q - 1$. Die Null wird durch $e_i = 0 \wedge f_i = 0$ dargestellt. Der Maximalwert von f_i ist

$$f_{\max} = \sum_{\nu=1}^{p-1} 2^{\nu-1} = \sum_{\nu=0}^{p-2} 2^{\nu} = \frac{1 - 2^{p-1}}{1 - 2} = 2^{p-1} - 1 \; . \tag{D.2}$$

f_i ist somit eine ganze Zahl aus dem Intervall $0 \leq f_i \leq f_{\max}$.

Für die folgenden Überlegungen bzgl. der Rundung nehmen wir an, dass keine Überlaufeffekte auftreten. Ist z das Ergebnis einer idealen Rechenoperation und nicht durch eine Gleitkommazahl des verwendeten Zahlensystems darstellbar, so entsteht bei der realen Rechenoperation der Rundungsfehler $z - \tilde{z}$. Gebräuchliche Rundungsarten sind

- Rundung in Richtung 0 (Betragsschneiden), d. h. $\tilde{z} = \mathrm{sgn}(z)\lceil |z| \rceil$, wobei sich hier die Gauß-Klammern auf das Zahlenraster und nicht auf die ganzen Zahlen beziehen,
- Rundung mit minimalem Betrag des Fehlers, d. h. $|z - \tilde{z}| \Rightarrow \min$,
- Rundung in Richtung ∞, d. h. $\tilde{z} = \lfloor z \rfloor$,
- Rundung in Richtung $-\infty$, d. h. $\tilde{z} = \lceil z \rceil$.

Wir werden nur die ersten beiden betrachten.

Maximaler Fehler bei der Addition und Subtraktion zweier Zahlen

Wir betrachten das Ergebnis der idealen Addition zweier Gleitkommazahlen $z_1 = \tilde{z}_1$ und $z_2 = \tilde{z}_2$

$$z_3 = z_1 + z_2 \tag{D.3}$$

und der realen Addition

$$\tilde{z}_3 = z_1 \oplus z_2 \; . \tag{D.4}$$

Zunächst untersuchen wir das Betragsschneiden. O.E.d.A. nehmen wir $z_1 > z_2$ und $s_1 = s_2 = 0$ an. Der maximale betragliche relative Fehler tritt genau dann auf, wenn z_2 gleich der größten Zahl ist, die keinen Beitrag zu $\tilde{z}_3$ liefert, d. h. $\tilde{z}_3 = z_1$ und z_3 die kleinstmögliche Mantisse hat. Dann gilt offenbar $z_3 - \tilde{z}_3 = z_1 + z_2 - z_1 = z_2 = \tilde{z}_2$. Die oben gemachten Überlegungen besagen, dass die Mantissen von z_1 durch $f_1 = 0$ und von z_2 durch $f_2 = f_{\max}$ bestimmt sind. Damit z_2 keinen Beitrag zu $\tilde{z}_3$ liefert, müssen die Exponenten von z_1 und z_2 gerade um p differieren. Bei Festkommazahlen tritt bei der Addition überhaupt kein Fehler auf, da die Exponenten gleich sind. Der maximale Betrag des relativen Fehlers bezogen auf z_3 ist

$$\max\left\{\left|\frac{z_3 - \tilde{z}_3}{z_3}\right|\right\} = \frac{z_2}{z_3} = \frac{1 + f_{\max} \cdot 2^{-p+1}}{1 + 0 \cdot 2^{-p+1}} \frac{1}{2^p} = \frac{2 - 2^{-p+1}}{2^p} = \frac{1 - 2^{-p}}{2^{p-1}} < \frac{1}{2^{p-1}} = 2^{1-p} \tag{D.5}$$

und bezogen auf $\tilde{z}_3$

$$\max\left\{\left|\frac{z_3 - \tilde{z}_3}{\tilde{z}_3}\right|\right\} = \frac{z_2}{\tilde{z}_3} = \frac{z_2}{z_3 - z_2} = \frac{z_2/z_3}{1 - z_2/z_3} = \frac{2 - 2^{-p+1}}{2^p - 2 + 2^{-p+1}} = \frac{1 - 2^{-p}}{2^{p-1} - 1 + 2^{-p}} \; . \tag{D.6}$$

Beispiel : Sei $p = 4$ und es gelten die folgenden Zahlenwerte

$$\begin{aligned} z_1 &= (1 + 0 \cdot 2^{-1} + 0 \cdot 2^{-2} + 0 \cdot 2^{-3}) \cdot 2^4 = (1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1) = 16 \\ z_2 &= (1 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3}) \cdot 2^0 = 1,875 \\ z_3 &= (1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3}) \\ \tilde{z}_3 &= (1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1) = z_1 \ . \end{aligned} \tag{D.7}$$

Der relative Fehler ist

$$\frac{z_3 - \tilde{z}_3}{z_3} = \frac{z_2}{z_3} = \frac{1,875}{16} = \frac{1 - 2^{-4}}{2^{4-1}} \approx 0,1172 \ . \tag{D.8}$$

Zum Vergleich der abgeschätzte relative Fehler

$$\frac{z_3 - \tilde{z}_3}{z_3} < 2^{-3} = 0,125 \ . \tag{D.9}$$

Im Bild D.1 ist die fehlerhafte Addition dargestellt.

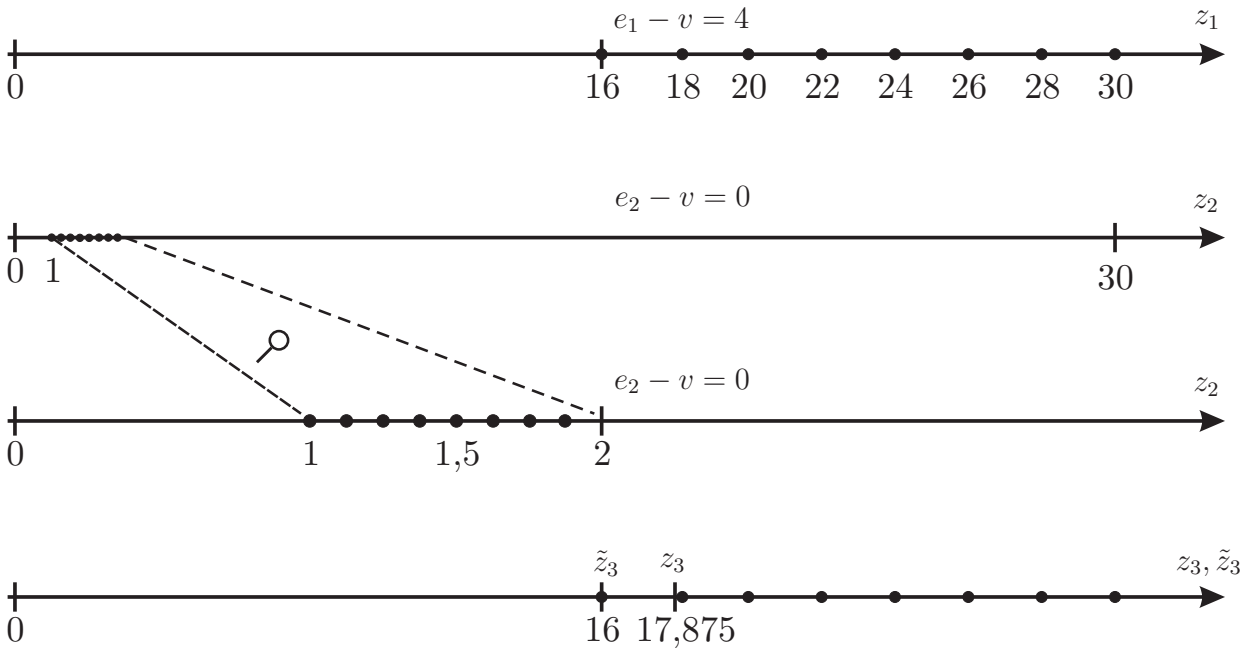


Bild D.1: Zahlenraster zu dem Beispiel der fehlerhaften Addition.

Im Fall der Rundung zum nächsten Gitterpunkt halbiert sich der relative Fehler bezogen auf z_3 zu

$$\max\left\{\left|\frac{z_3 - \tilde{z}_3}{z_3}\right|\right\} = \frac{1 - 2^{-p}}{2^p} < 2^{-p} \ . \tag{D.10}$$

Die unterschiedlichen Vorzeichen des Fehlers $z_3 - \tilde{z}_3$ und von z_3 legen eine Ober- und eine Untergrenze fest, welche im Folgenden unter Zuhilfenahme der Abkürzung

$$\varepsilon = \begin{cases} 2^{1-p} & \text{für Betragsschneiden,} \\ 2^{-p} & \text{für Gitterpunktschneiden.} \end{cases} \tag{D.11}$$

ermittelt werden soll. Beide Schneideoperationen bewirken, dass die Vorzeichen von $\tilde{z}_3$ und z_3 gleich

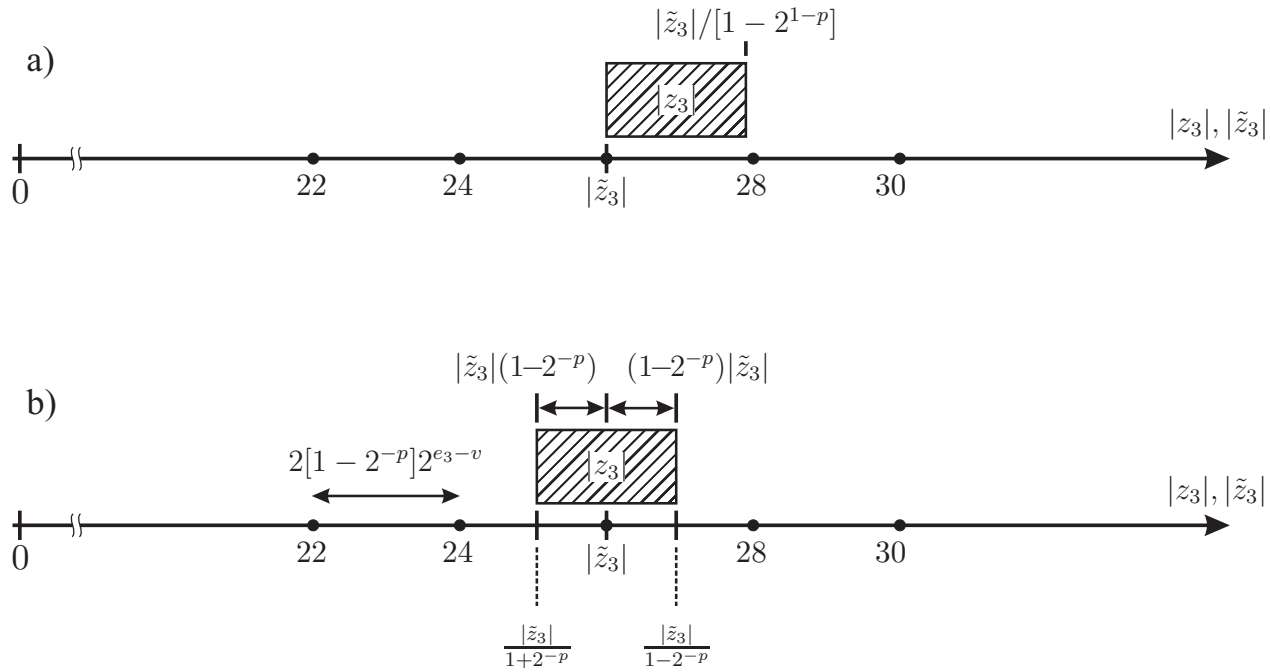


Bild D.2: Fehler bei der Addition a) Betragsschneiden b) Gitterpunktschneiden

sind, d. h. $\tilde{z}_3 z_3 \geq 0$. Dann ergibt sich

$$\frac{|z_3 - \tilde{z}_3|}{|z_3|} < \varepsilon \;\Leftrightarrow\; \pm[\,|z_3| - |\tilde{z}_3|\,] < \varepsilon|z_3| \;\Leftrightarrow\; |z_3| - |\tilde{z}_3| \overset{<}{>} \pm\varepsilon|z_3| \;\Leftrightarrow\; |z_3|[1 \mp \varepsilon] \overset{<}{>} |\tilde{z}_3| \,. \tag{D.12}$$

Dieses Ergebnis können wir zu

$$[1-\varepsilon] < \frac{|\tilde{z}_3|}{|z_3|} < [1+\varepsilon] \quad\Longleftrightarrow\quad \frac{1}{1+\varepsilon} < \frac{|z_3|}{|\tilde{z}_3|} < \frac{1}{1-\varepsilon} \tag{D.13}$$

zusammenfassen. Beim Betragsschneiden gilt zusätzlich die schärfere Bedingung $\frac{|\tilde{z}_3|}{|z_3|} \leq 1$ und folglich

$$1 - 2^{1-p} < \frac{|\tilde{z}_3|}{|z_3|} \leq 1 \quad\Longleftrightarrow\quad 1 \leq \frac{|z_3|}{|\tilde{z}_3|} < \frac{1}{1-2^{1-p}} \,. \tag{D.14}$$

Für das Gitterpunktschneiden hingegen gilt

$$[1-2^{-p}] < \frac{|\tilde{z}_3|}{|z_3|} < [1+2^{-p}] \quad\Longleftrightarrow\quad \frac{1}{1+2^{-p}} < \frac{|z_3|}{|\tilde{z}_3|} < \frac{1}{1-2^{-p}} \,. \tag{D.15}$$

Im Bild D.2 sind die Fehler in beiden Fällen dargestellt.

Maximaler Fehler bei der Multiplikation zweier Zahlen

Das Ergebnis der idealen Multiplikation zweier Gleitkommazahlen $z_1 = \tilde{z}_1$ und $z_2 = \tilde{z}_2$ ist

$$z_3 = z_1 \cdot z_2 \tag{D.16}$$

und bei realer Multiplikation

$$\tilde{z}_3 = z_1 \odot z_2 \,. \tag{D.17}$$

Wir betrachten den Fall $f_1 = f_2 = f_{\max}$ (lauter Einsen in der Binärdarstellung der Mantisse), da dies die maximale Mantisse bei z_3 liefert. O.E.d.A. nehmen wir $z_1, z_2 > 0$ an,

$$\begin{aligned}
z_3 &= z_1\, z_2 = [1 + f_{\max} 2^{1-p}]^2\, 2^{e_1+e_2-2v} = [2 - 2^{1-p}]^2\, 2^{e_1+e_2-2v} = [1 - 2^{-p}]^2\, 2^{2+e_1+e_2-2v} \\
&= [1 - 2 \cdot 2^{-p} + 2^{-2p}]\, 2^{2+e_1+e_2-2v} = [1 - 2^{-p+1} + 2^{-2p}]\, 2^{2+e_1+e_2-2v} \\
&= [2^{1-p} \sum_{\nu=0}^{p-2} 2^{\nu} + 2^{-2p}]\, 2^{2+e_1+e_2-2v} = [\sum_{\nu=0}^{p-2} 2^{1+\nu-p} + 2^{-2p}]\, 2^{2+e_1+e_2-2v} \\
&= [1 + \sum_{\nu=0}^{p-3} 2^{2+\nu-p} + 2^{1-2p}]\, 2^{1+e_1+e_2-2v} = [1 + 2^{1-2p} \sum_{\nu=1}^{2p-1} b_{3\nu} 2^{\nu-1}]\, 2^{1+e_1+e_2-2v} \\
&= [1 + 2^{1-p} \sum_{\nu=1}^{2p-1} b_{3\nu} 2^{\nu-1-p}]\, 2^{1+e_1+e_2-2v} = [1 + 2^{1-p} \sum_{\nu=1}^{p} b_{3\nu} 2^{\nu-1-p} + 2^{1-p} \sum_{\nu=1}^{p-1} b_{3\nu+p} 2^{\nu-1}]\, 2^{1+e_1+e_2-2v}.
\end{aligned} \tag{D.18}$$

Es gilt $e_3 - v = 1 + e_1 + e_2 - 2v$. Die Länge der Charakteristik ist offenbar von $p - 1$ auf $2p - 1$ angestiegen. Presst man eine Zahl mit Mantissenlänge $2p$ auf einen Gitterpunkt eines Zahlensystems mit Mantissenlänge p, so ist der Betrag des dadurch entstehenden maximalen Fehlers dann gegeben, wenn $b_{3\nu} = 1$ für $\nu = 1, 2, \ldots p$ gilt. Der maximale Fehler in der Mantisse ist also bei Rundung mittels Betragsschneiden

$$z_3 - \tilde{z}_3 = 2^{1-2p} \sum_{\nu=1}^{p} 2^{\nu-1}\, 2^{1+e_1+e_2-2v} = 2^{2-2p+e_1+e_2-2v} \sum_{\nu=0}^{p-1} 2^{\nu} = 2^{2-p+e_1+e_2-2v}(1 - 2^{-p})\ . \tag{D.19}$$

Der größte relative Fehler bezogen auf ein z_3 entsteht, wenn z_3 die kleinste Mantisse hat, zu

$$\frac{z_3 - \tilde{z}_3}{z_3} < \frac{2^{2-p+e_1+e_2-2v}(1 - 2^{-p})}{2^{1+e_1+e_2-2v}} = 2^{1-p}(1 - 2^{-p}) < 2^{1-p}\ . \tag{D.20}$$

Der relative Fehler ist also nicht größer als das letzte Bit der Mantisse von z_3, d. h. $2^{1-p}\, 2^{e_3-v}$.

Bei Rundung zum nächsten Gitterpunkt ist hingegen der Fehler nicht größer als die Hälfte des letzten Bits von z_3, d. h. $2^{-p} 2^{e_3-v}$. Die relativen Fehler bezogen auf z_3 sind dann bei beliebigen Zahlen durch

- bei der Rundung mittels Betragsschneidens durch

$$\left| \frac{z_3 - \tilde{z}_3}{z_3} \right| < 2^{-p+1} \tag{D.21}$$

- und bei Rundung zum nächsten Gitterpunkt durch

$$\left| \frac{z_3 - \tilde{z}_3}{z_3} \right| < 2^{-p}\ . \tag{D.22}$$

beschränkt. Die maximalen Fehler, die bei der Addition und der Multiplikation entstehen sind zwar nicht exakt gleich, aber die angegebenen oberen Abschätzungen Gleichung (D.5) bzw. Gleichung (D.10) und Gleichung (D.21) bzw. Gleichung (D.22) sind bei beiden Rechenoperationen gleich. Da wir ein z_3 mit kleinster Mantisse angenommen haben, gilt die Abschätzung auch für den relativen Fehler bezogen auf $\tilde{z}_3$, d. h.

- bei der Rundung mittels Betragsschneidens

$$\left|\frac{z_3 - \tilde{z}_3}{\tilde{z}_3}\right| < 2^{-p+1} \tag{D.23}$$

- und bei Rundung zum nächsten Gitterpunkt

$$\left|\frac{z_3 - \tilde{z}_3}{\tilde{z}_3}\right| < 2^{-p} \; . \tag{D.24}$$

Beispiel Erneut gilt $p = 4$ und die Zahlenwerte lauten

$$\begin{aligned} z_1 &= (1 + 0 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3}) \cdot 2^3 = 2^3 + 2^1 + 2^0 = 11 \\ z_2 &= (1 + 1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3}) \cdot 2^3 = 2^3 + 2^2 + 2^0 = 13 \; . \end{aligned} \tag{D.25}$$

Die Berechnung liefert

$$\begin{array}{lcccccccc} z_1 \cdot 2^3 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ z_1 \cdot 2^2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ z_1 \cdot 2^0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ & -- & -- & -- & -- & -- & -- & -- & -- \\ z_3 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array} \tag{D.26}$$

oder in Dezimaldarstellung $z_3 = 143$. Bei Abrundung haben wir $\tilde{z}_3 = 2^7 = 128$ und bei Aufrundung $\tilde{z}_3 = 2^7 + 2^4 = 144$. Die relativen Fehler $15/143 = 0,\overline{104895}$ bzw. $1/143 = 0,\overline{006993}$ sind, wie behauptet, durch die Obergrenzen $2^{-p+1} = 0,125$ bzw. $2^{-p} = 0,0625$ begrenzt.

Die unterschiedlichen Vorzeichen des Fehlers von z_3 legen die Ober- und die Untergrenze fest. Bereits ermittelt wurde

$$\left|\frac{z_3 - \tilde{z}_3}{z_3}\right| < \varepsilon \Longrightarrow \begin{cases} 1 - 2^{1-p} < \frac{|\tilde{z}_3|}{|z_3|} \leq 1 & \text{für Betragsschneiden,} \\ [1 - 2^{-p}] < \frac{|\tilde{z}_3|}{|z_3|} < [1 + 2^{-p}] & \text{für Gitterpunktschneiden.} \end{cases} \tag{D.27}$$

D.2 Idealer und realer Zweitor-Parallel-Adaptor

Wir werden nun Untersuchungen zu einem Zweitor-Parallel-Adaptor durchführen, der energieneutral ist und im realen Fall passiv sein soll. Die Berechnungsgleichungen im idealen Fall sind

$$\boldsymbol{b} = [\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} - \mathbf{1}_2]\boldsymbol{a} \quad , \quad \boldsymbol{\Gamma} = [\gamma_1 \, , \, \gamma_2]^{\mathrm{T}} \quad , \quad \gamma_1 \, , \, \gamma_2 > 0 \quad , \quad \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} = 2 \; . \tag{D.28}$$

Die Bestimmung von b_1 und b_2 soll durch

$$a_0 = \gamma_1 a_1 + \gamma_2 a_2 \quad , \quad b_1 = \gamma_1 a_0 - a_1 \quad , \quad b_2 = \gamma_2 a_0 - a_2 \tag{D.29}$$

erfolgen. Das Bild D.3 zeigt das Signalfluss-Diagramm des idealen und Bild D.4 das des realen Zweitor-Adaptors. Im realen Fall kann die reale Operation (Multiplikation/Addition) durch die ideale Operation und einem Reformatierungsbaustein dargestellt werden. Die Untersuchungen in Bezug auf den Fehler betreffen 4 Punkte :

1. die ersten zwei Multiplikationen,

2. die erste Addition,

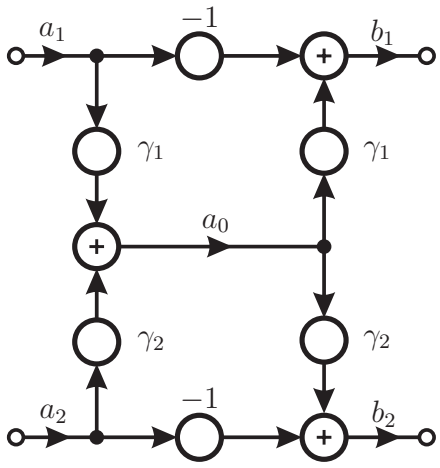


Bild D.3: Signalfluss-Diagramm eines idealen Zweitor-Parallel-Adaptors.

3. die zweiten zwei Multiplikationen und

4. die zweiten zwei Additionen.

Dabei wird jeweils die Untersuchung für die Rundungsart a) Betragsschneiden als auch für b) Gitterpunktschneiden durchgeführt.

Im weiteren Verlauf werden wir die Passivität an manchen Stellen dadurch garantieren, dass anstelle der unitären Matrix $\boldsymbol{S}$ die unitär beschränkte Matrix $\boldsymbol{S'} = \boldsymbol{\Gamma'}_b\boldsymbol{\Gamma'}_a^{\mathrm{T}} - \mathbf{1}_2$ verwendet wird, wobei sich die Vektoren $\boldsymbol{\Gamma'}_a$ und $\boldsymbol{\Gamma'}_b$ durch Kürzung des Vektors $\boldsymbol{\Gamma}$ ergeben, d. h. es gilt $\boldsymbol{\Gamma'}_a\|\boldsymbol{\Gamma}\| = \boldsymbol{\Gamma}\|\boldsymbol{\Gamma'}_a\|$, $\boldsymbol{\Gamma'}_b\|\boldsymbol{\Gamma}\| = \boldsymbol{\Gamma}\|\boldsymbol{\Gamma'}_b\|$, $\|\boldsymbol{\Gamma'}_a\| \leq \|\boldsymbol{\Gamma}\|$ und $\|\boldsymbol{\Gamma'}_b\| \leq \|\boldsymbol{\Gamma}\|$. Wir müssen nun noch zeigen, dass die Eigenwerte die für die unitäre Beschränktheit notwendigen und hinreichenden Bedingungen $|\lambda'_1| \leq 1 \wedge |\lambda'_2| \leq 1$ erfüllen. Dazu berechnen wir explizit die Eigenwerte

$$\begin{aligned}\det[\mathbf{1}_2\lambda - \boldsymbol{S'}] &= \det[\mathbf{1}_2\lambda - \boldsymbol{\Gamma'}_b\boldsymbol{\Gamma'}_a^{\mathrm{T}} + \mathbf{1}_2] = \det[\mathbf{1}_2(\lambda+1) - \boldsymbol{\Gamma'}_b\boldsymbol{\Gamma'}_a^{\mathrm{T}}] = (\lambda+1)^2\det[\mathbf{1}_2 - \boldsymbol{\Gamma'}_b\boldsymbol{\Gamma'}_a^{\mathrm{T}}/(\lambda+1)] \\ &= (\lambda+1)^2[1 - \boldsymbol{\Gamma'}_a^{\mathrm{T}}\boldsymbol{\Gamma'}_b/(\lambda+1)] = (\lambda+1)[\lambda+1-\boldsymbol{\Gamma'}_a^{\mathrm{T}}\boldsymbol{\Gamma'}_b] = (\lambda+1)[\lambda+1-\|\boldsymbol{\Gamma'}_a\|\|\boldsymbol{\Gamma'}_b\|]\,.\end{aligned} \tag{D.30}$$

Beide Eigenwerte

$$\lambda'_1 = -1 \quad \text{und} \quad \lambda'_2 = \|\boldsymbol{\Gamma'}_a\|\|\boldsymbol{\Gamma'}_b\| - 1 \tag{D.31}$$

sind unimodular beschränkt. Für λ'_2 sieht man das wie folgt

$$\|\boldsymbol{\Gamma'}_a\|\|\boldsymbol{\Gamma'}_b\| \leq \|\boldsymbol{\Gamma}\|^2 = 2 \Leftrightarrow \lambda'_2 = \|\boldsymbol{\Gamma'}_a\|\|\boldsymbol{\Gamma'}_b\| - 1 \leq 1\,. \tag{D.32}$$

Da der Betrag einer Zahl nicht negativ sein kann, ist λ'_2 natürlich auch nicht kleiner als -1.

1. Die erste Multiplikation

Die reale Multiplikation $\gamma_1 \odot a_1$ entspricht (für feste Faktoren γ_1 und a_1) der idealen Multiplikation von a_1 mit einem modifiziertem $\tilde{\gamma}_1$, d. h.

$$\tilde{z}_3 = \gamma_1 \odot a_1 = \tilde{\gamma}_1 a_1\,. \tag{D.33}$$

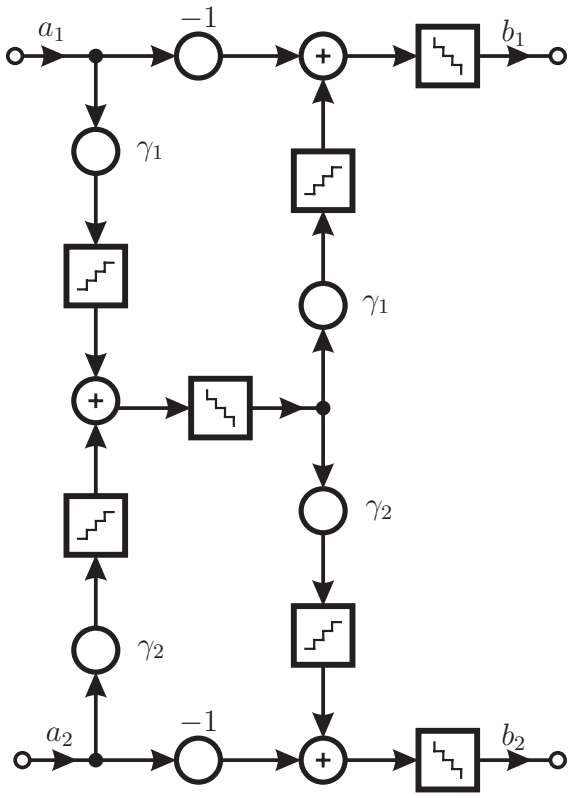


Bild D.4: Signalfluss-Diagramm eines realen Zweitor-Parallel-Adaptors.

Entsprechendes gilt für $\gamma_2 \odot a_2$. In Matrixform lautet dies $\tilde{\boldsymbol{b}} = [\boldsymbol{\Gamma}\tilde{\boldsymbol{\Gamma}}^{\mathrm{T}} - \mathbf{1}_2]\boldsymbol{a} = \tilde{\boldsymbol{S}}\boldsymbol{a}$. Aus Passivitätsgründen muss $\tilde{\boldsymbol{S}}$ unitär beschränkt sein. Die Passivität ist gemäß Gleichung (D.32) genau dann gegeben, wenn $||\tilde{\boldsymbol{\Gamma}}|| \leq ||\boldsymbol{\Gamma}||$ gilt. Nehmen wir an, $\gamma_2 \odot a_2$ sei fehlerfrei, so folgt $\tilde{\gamma}_1 \leq \gamma_1$ und somit

$$\tilde{\gamma}_1 \, |a_1| \leq \gamma_1 \, |a_1| \; . \tag{D.34}$$

Da die Multipliziererkoeffizienten positiv sein sollen, können sie in die Betragsoperation gezogen werden.

$$|\tilde{\gamma}_1 \, a_1| \leq |\gamma_1 \, a_1| \; . \tag{D.35}$$

Mit Gleichung (D.33) haben wir

$$|\gamma_1 \odot a_1| \leq |\gamma_1 \, a_1| \; . \tag{D.36}$$

Die Passivität fordert also, dass das Ergebnis der realen Multiplikation nicht größer als das der idealen Multiplikation ist.

a) Rundung mittels Betragsschneiden

Bei Verwendung des Betragsschneidens ergibt sich das Ergebnis der realen Multiplikation durch Abrundung des Betrages aus der idealen Rechnung. Ohne weitere Eingriffe ist Gleichung (D.36) erfüllt und die Passivität sichergestellt.

b) Gitterpunktschneiden

Bei Verwendung des Gitterpunktschneidens existieren hingegen Fälle, die Gleichung (D.36) verletzen. Abhilfe schafft für diese Rundungsart ein Stauchen der Multipliziererkoeffizienten. Nach Gleichung (D.27) ist mit $z_3 = \gamma_1 a_1$ und $\tilde{z}_3 = \gamma_1 \odot a_1$ der Fehlerbereich nur durch

$$[1 - 2^{-p}] < \left| \frac{\gamma_1 \odot a_1}{\gamma_1 a_1} \right| < [1 + 2^{-p}] \tag{D.37}$$

eingeschränkt, was Gleichung (D.36) zur oberen Schranke hin, wie eingangs behauptet, verletzen kann. Um dem zu begegnen, wird in der realen Multiplikation der Koeffizient γ_1 durch $\gamma_1' = \gamma_1/[1+2^{-p}]$ ersetzt. Dann haben wir

$$[1 - 2^{-p}] < \left| \frac{\gamma_1' \odot a_1}{\gamma_1' a_1} \right| < [1 + 2^{-p}] \, . \tag{D.38}$$

Setzen wir $\gamma_1' = \gamma_1/[1 + 2^{-p}]$ ein, so folgt

$$\left| \frac{[1 + 2^{-p}]\gamma_1' \odot a_1}{\gamma_1 a_1} \right| < [1 + 2^{-p}] \Longleftrightarrow |\gamma_1' \odot a_1| < |\gamma_1 a_1| \, . \tag{D.39}$$

Offenbar wird nun Gleichung (D.36) immer erfüllt.

2. Die erste Addition

a) Rundung mittels Betragsschneiden

Wir nehmen nun an, dass die ersten beiden Multiplikationen exakt ausgeführt wurden. Anstatt der idealen Addition $a_0 = \gamma_1 a_1 + \gamma_2 a_2 = \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{a}$ haben wir im realen Fall $\tilde{a}_0 = \gamma_1 a_1 \oplus \gamma_2 a_2$. Wir werden zeigen, dass ein Betragsschneiden von $\tilde{a}_0$ die Passivität des realen Zweitor-Paralleladaptors sicherstellt. Dies ist ein bemerkenswertes Ergebnis (welches übrigens auch für n-Tor-Serien- und Paralleladaptoren gültig ist), da a_0 im Gegensatz zu a_1, a_2, b_1, b_2 nicht mehr als eine Leistungsgröße interpretiert werden kann.

Setzen wir zudem voraus, dass die weiteren Operationen auch ideal durchgeführt werden, so berechnen sich die Ausgangswellen im realen Fall durch

$$\tilde{\boldsymbol{b}} = \boldsymbol{\Gamma} \tilde{a}_0 - \boldsymbol{a} \, . \tag{D.40}$$

Die „reflektierte Leistung“ des Adaptors ist das Normquadrat von $\tilde{\boldsymbol{b}}$ und lautet

$$\|\tilde{\boldsymbol{b}}\|^2 = \|\tilde{a}_0 \, \boldsymbol{\Gamma}\|^2 - 2 \, \tilde{a}_0 \, \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{a} + \|\boldsymbol{a}\|^2 = 2 \, \tilde{a}_0 [\tilde{a}_0 - a_0] + \|\boldsymbol{a}\|^2 \Leftrightarrow \|\tilde{\boldsymbol{b}}\|^2 - \|\boldsymbol{a}\|^2 = 2 \, \tilde{a}_0 [\tilde{a}_0 - a_0] \, . \tag{D.41}$$

Die Passivitätbedingung lautet in dem vorliegenden Fall

$$||\tilde{\boldsymbol{b}}||^2 \leq \|\boldsymbol{a}\|^2 \Leftrightarrow ||\tilde{\boldsymbol{b}}||^2 - \|\boldsymbol{a}\|^2 \leq 0 \, . \tag{D.42}$$

Offenbar ist $\tilde{a}_0 [a_0 - \tilde{a}_0] \geq 0$ notwendig und hinreichend für die Passivität. Diese Beziehung wird durch ein Betragsschneiden von a_0 nach $\tilde{a}_0$ erfüllt, wie im Folgenden gezeigt wird. Das Betragsschneiden impliziert das Nichtändern des Vorzeichens, d. h. $|\tilde{a}_0| \operatorname{sgn}(a_0) = |\tilde{a}_0| \operatorname{sgn}(\tilde{a}_0)$. Mit $\tilde{a}_0 = |\tilde{a}_0| \operatorname{sgn}(\tilde{a}_0)$ erhalten wir dann

$$\tilde{a}_0 [a_0 - \tilde{a}_0] = |\tilde{a}_0| [\operatorname{sgn}(\tilde{a}_0) \, a_0 - \operatorname{sgn}(\tilde{a}_0) \, \tilde{a}_0] = |\tilde{a}_0| [\operatorname{sgn}(a_0) \, a_0 - \operatorname{sgn}(\tilde{a}_0) \, \tilde{a}_0] = |\tilde{a}_0| [|a_0| - |\tilde{a}_0|] \geq 0 \Leftarrow |a_0| \geq |\tilde{a}_0| \tag{D.43}$$

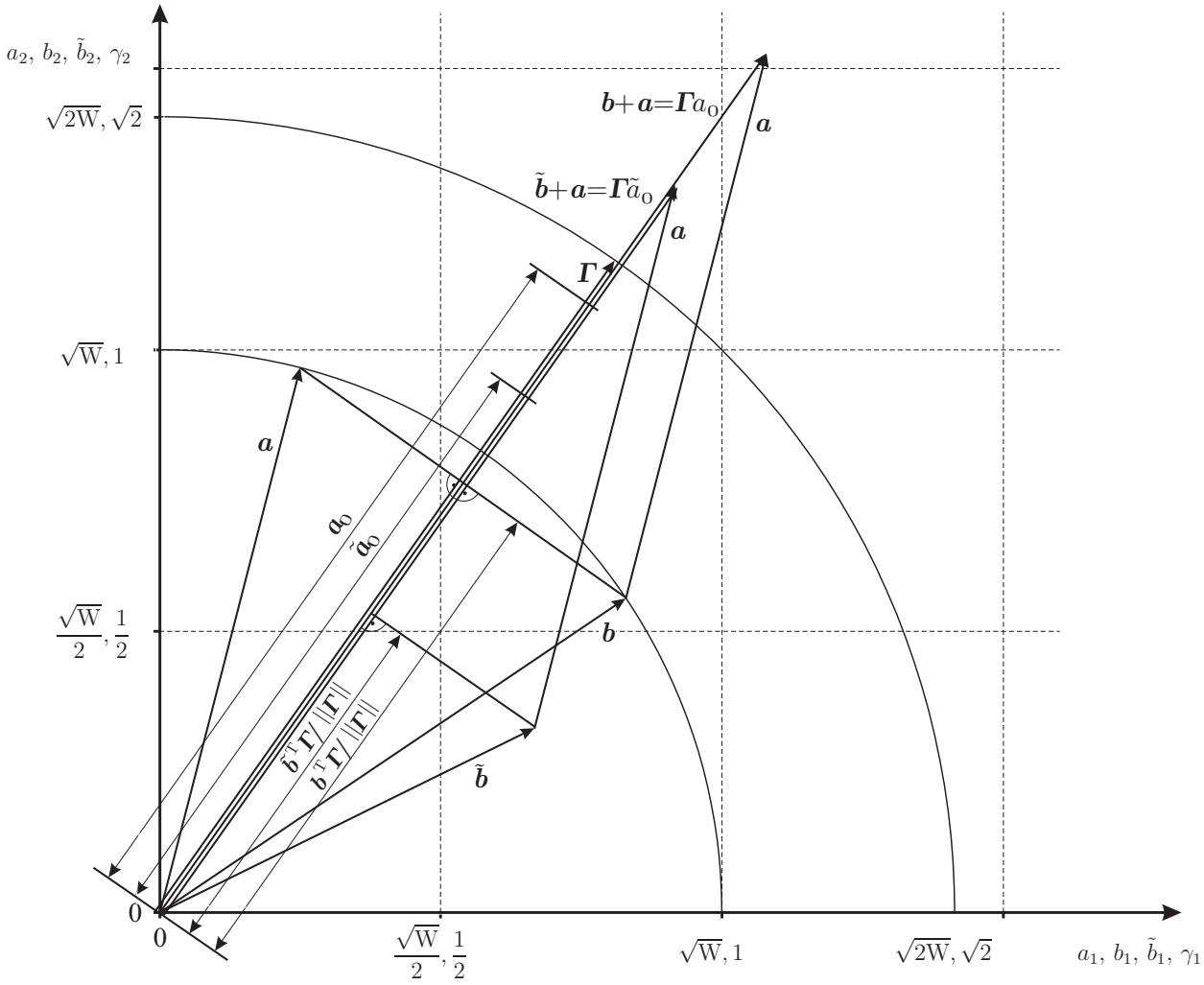


Bild D.5: Wellengrößen eines realen und eines idealen Zweitor-Paralleladaptors

Der Vorgang des Betragsschneidens und der Berechnung von $\boldsymbol{b}$ und $\tilde{\boldsymbol{b}}$ ist im Bild D.5 für $\|\boldsymbol{a}\|^2 = 1\mathrm{W}$ gezeichnet.

b) Gitterpunktschneiden

Wie wir im vorherigen Abschnitt festgestellt haben, ist $\tilde{a}_0[a_0 - \tilde{a}_0] \geq 0$ notwendig und hinreichend für die Passivität. Durch das Gitterpunktschneiden kann natürlich die Situation $\tilde{a}_0 > a_0 > 0$ auftreten, die die obige Bedingung verletzt.

Wir werden dieses Problem erneut dadurch lösen, dass wir statt der ersten Multipliziererkoeffizienten $\boldsymbol{\Gamma}$ die Multipliziererkoeffizienten $\boldsymbol{\Gamma}' = \alpha\boldsymbol{\Gamma}$ (mit den Koordinaten $\gamma_1' = \gamma_1\alpha$ und $\gamma_2' = \gamma_2\alpha$) mit $1 > \alpha > 0$ verwenden. Der Wert $\tilde{a}_0' = \tilde{z}_3$ der realen Addition berechnet sich nun durch

$$\tilde{a}_0' = \tilde{z}_3 = \gamma_1' a_1 \oplus \gamma_2' a_2 \ , \tag{D.44}$$

mit den angenommenen Summanden $z_1 = \gamma_1' a_1 > \gamma_2' a_2 = z_2 > 0$. Somit ergibt sich der Ausgangsvektor zu

$$\tilde{\boldsymbol{b}}' = \boldsymbol{\Gamma}\tilde{a}_0' - \boldsymbol{a} \ . \tag{D.45}$$

Nun stellt sich die Frage, wie groß die Stauchung sein muss. Die Antwort ergibt sich aus der Passivitätsbedingung. Demnach darf für beliebige Eingangsvektoren $\boldsymbol{a}$ die Ausgangsleistung $\|\tilde{\boldsymbol{b}}'\|^2$ nicht größer als

die Eingangsleistung $\|\boldsymbol{a}\|^2$ sein. Hinreichend und notwendig dafür ist $|\tilde{a}_0'| \leq |a_0|$. Wir müssen nun den fehlerhaften Wert $\tilde{a}_0'$ durch a_0, p und α ausdrücken und die Bedingung nach α auflösen. Der Fehler wird durch die Ungleichung D.15 mit

$$z_3 = a_0' = \gamma_1' a_1 + \gamma_2' a_2 = \alpha\gamma_1 a_1 + \alpha\gamma_2 a_2 = \alpha \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{a} = \alpha a_0 \tag{D.46}$$

zu

$$[1 - 2^{-p}] < \frac{|\tilde{a}_0'|}{|a_0'|} < [1 + 2^{-p}] \tag{D.47}$$

eingegrenzt. In Bezug auf die Passivität ist die Aufrundung des Betrages, d. h. die Obergrenze

$$|\tilde{a}_0'| < |a_0'|[1 + 2^{-p}] \tag{D.48}$$

entscheidend. Setzen wir hier $a_0' = \alpha\, a_0$ ein, so haben wir

$$|\tilde{a}_0'| < \alpha|a_0|[1 + 2^{-p}] \; . \tag{D.49}$$

Im schlimmsten Fall darf $|\tilde{a}_0'|$ gerade so groß wie $|a_0|$ sein, d. h. $|a_0| \geq |\tilde{a}_0'|$. Aus der in Gleichung (D.49) bestimmten Obergrenze von $|\tilde{a}_0'|$ erhalten wir den gesuchten Wert für α

$$|a_0| \geq \alpha|a_0|[1 + 2^{-p}] \quad \Leftrightarrow \quad \alpha \leq \frac{1}{[1 + 2^{-p}]} \approx [1 - 2^{-p}] \; . \tag{D.50}$$

Somit wurde der Nachweis der Passivität bei Stauchung der Multiplizierer und Anwendung des Gitterpunktschneidens für die erste Addition erbracht.

3. Die zweite Multiplikation

Wir gehen hier von einem korrekt berechnetem a_0 aus und betrachten die ideale Multiplikation in $b_1 = \gamma_1 a_0 - a_1$. Das Ergebnis der realen Multiplikation $\gamma_1 \odot a_0$ ist gleich der idealen Multiplikation mit einem modifiziertem γ_1, d. h.

$$\tilde{z}_3 = \gamma_1 \odot a_0 = \tilde{\gamma_1}\, a_0 \; . \tag{D.51}$$

Gemäß Gleichung (D.32) ist für die Passivität

$$\tilde{\gamma_1} \leq \gamma_1 \tag{D.52}$$

notwendig. Somit auch

$$\tilde{\gamma_1}|a_0| \leq \gamma_1|a_0| \tag{D.53}$$

und mit $\gamma_1, \tilde{\gamma_1} > 0$

$$|\tilde{\gamma_1} a_0| \leq |\gamma_1 a_0| \; . \tag{D.54}$$

Verwenden von Gleichung (D.51) liefert dann

$$|\gamma_1 \odot a_0| \leq |\gamma_1 a_0| \; . \tag{D.55}$$

Dies bedeutet, dass das Ergebnis der realen Multiplikation dem Betrage nach kleiner als das Ergebnis der idealen Multiplikation sein muss, um die Passivität zu gewährleisten.

a) Rundung mittels Betragsschneiden

Bei Verwendung des Betragsschneidens ergibt sich das Ergebnis der realen Multiplikation durch Abrundung des Betrages einer idealen Rechnung. Ohne weitere Eingriffe ist Gleichung (D.55) erfüllt und die Passivität sichergestellt.

b) Gitterpunktschneiden

Bei Verwendung des Gitterpunktschneidens existieren hingegen Fälle, die Gleichung (D.55) verletzen. Abhilfe schafft für diese Rundungsart ein Stauchen der Koeffizienten. Nach Gleichung (D.27) ist der relative Fehler bei der Multiplikation mit $z_3 = \gamma_1 a_0$ und $\tilde{z}_3 = \tilde{\gamma}_1 a_0 = \gamma_1 \odot a_0$ durch

$$1 - 2^{-p} < \frac{|\gamma_1 \odot a_0|}{|\gamma_1 a_0|} < 1 + 2^{-p} \tag{D.56}$$

eingeschränkt. Das bedeutet, dass der Fall $|\gamma_1 \odot a_0| > |\gamma_1 a_0|$ auftreten kann und Gleichung (D.55), wie eingangs behauptet, verletzt. Um dem zu begegnen, wird in der realen Multiplikation der Koeffizienten γ_1 durch $\gamma_1' = \gamma_1/[1 + 2^{-p}]$ ersetzt. Dann haben wir aus Gleichung (D.27) die Fehlerschranke

$$1 - 2^{-p} < \frac{|\gamma_1' \odot a_0|}{|\gamma_1' a_0|} < 1 + 2^{-p} \,, \tag{D.57}$$

d. h.

$$\frac{|[1 + 2^{-p}]\gamma_1' \odot a_0|}{|\gamma_1 a_0|} < 1 + 2^{-p} \iff |\gamma_1' \odot a_0| < |\gamma_1 a_0| \,. \tag{D.58}$$

Offenbar wird nun Gleichung (D.55) immer erfüllt.

4. Die zweite Addition

Hier wird der Rundungsfehler berücksichtigt, der bei der Berechnung von b_1 und b_2 auftritt. Die idealen Additionen lauten $b_1 = \gamma_1 a_0 + (-a_1)$ und $b_2 = \gamma_2 a_0 + (-a_2)$, die realen Additionen $\tilde{b}_1 = \gamma_1 a_0 \oplus (-a_1)$ und $\tilde{b}_2 = \gamma_2 a_0 \oplus (-a_2)$. Für die Passivität muss $||\tilde{\boldsymbol{b}}|| \leq ||\boldsymbol{a}|| = ||\boldsymbol{b}||$ gelten, was durch $|\tilde{b}_1| \leq |b_1|$ und $|\tilde{b}_2| \leq |b_2|$ gewährleistet ist.

a) Rundung mittels Betragsschneiden

Im Falle der Anwendung des Betragsschneidens gilt $|\tilde{b}_1| \leq |b_1|$ und $|\tilde{b}_2| \leq |b_2|$. Die Passivität an dieser Additionsstelle ist nicht gefährdet, da b_1 und b_2 Energiegrößen sind.

b) Gitterpunktschneiden

Im Falle von Gitterpunktschneiden ist dies aber nicht sichergestellt. Im Gegenteil, wir werden zeigen, dass das System selbst dann aktiv ist, wenn wir den Vektor $\boldsymbol{\Gamma}$ stauchen.

Der Fehler bei der Berechnung von $b_1 = -a_1 + \gamma_1 a_0$ durch $\tilde{b}_1 = -a_1 \oplus \gamma_1 a_0$ ist gemäß Gleichung (D.15) durch

$$1 - 2^{-p} < \frac{|\tilde{b}_1|}{|b_1|} < 1 + 2^{-p} \tag{D.59}$$

eingeschränkt. Somit ist auch

$$|b_1| < |\tilde{b}_1| \iff -|\tilde{b_1}|^2 < -|b_1|^2 \tag{D.60}$$

möglich. Die aufgenommene Leistung berechnet sich dann durch

$$\tilde{p} = \|\boldsymbol{a}\|^2 - ||\tilde{\boldsymbol{b}}||^2 = \|\boldsymbol{a}\|^2 - |\tilde{b}_1|^2 - |b_2|^2 < \|\boldsymbol{a}\|^2 - |b_1|^2 - |b_2|^2 = \|\boldsymbol{a}\|^2 - \|\boldsymbol{b}\|^2 = 0 \tag{D.61}$$

und ist negativ. Das System ist folglich aktiv.

Die Streumatrix $\boldsymbol{S}$ wird nun durch $\boldsymbol{S}' = \boldsymbol{\Gamma}'\boldsymbol{\Gamma}'^{\mathrm{T}} - \mathbf{1}_2$ ersetzt, wobei $\alpha\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'$ und $\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} > \boldsymbol{\Gamma}'^{\mathrm{T}}\boldsymbol{\Gamma}'$ gilt. Wir legen die Bezeichnung $\tilde{\boldsymbol{b}'} = \boldsymbol{S}' \odot \boldsymbol{a}$ für die reale Matrixmultiplikation (hier nur mit fehlerhafter Addition bei den Skalarprodukten) fest. Wie wir eben gesehen haben, kann sich $\boldsymbol{b}'$ maximal durch die zweite Addition auf $\boldsymbol{b}[1 + 2^{1-p}]$ vergrößern. Dieser Fehler bei der realen Multiplikation mit der Matrix $\boldsymbol{S}'$ entspricht dem der idealen Multiplikation mit der Matrix $\boldsymbol{S}'[1 + 2^{1-p}]$, d. h.

$$\tilde{\boldsymbol{b}}' = \boldsymbol{S}'[1 + 2^{1-p}]\boldsymbol{a} = \boldsymbol{S}' \odot \boldsymbol{a}\ . \tag{D.62}$$

Das Ergebnis der realen Multiplikation soll aber gerade der Passivitätsbedingung

$$||\tilde{\boldsymbol{b}}'||^2 = \|\boldsymbol{S}' \odot \boldsymbol{a}\|^2 \leq \|\boldsymbol{a}\|^2 \tag{D.63}$$

genügen. Das heißt, der neue Vektor $\boldsymbol{\Gamma}'$ muss nun so geartet sein, dass $\boldsymbol{S}'[1 + 2^{1-p}]$ unitär ist. Die Eigenwerte dürfen dem Betrag nach nicht größer als eins sein, d. h.

$$\begin{aligned}
&\det[\mathbf{1}_2\lambda - [1 + 2^{1-p}]\boldsymbol{S}'] = \det[\mathbf{1}_2\lambda - [1 + 2^{1-p}][\boldsymbol{\Gamma}'\boldsymbol{\Gamma}'^{\mathrm{T}} - \mathbf{1}_2]] = \det[\mathbf{1}_2(\lambda + 1 + 2^{1-p}) - [1 + 2^{1-p}]\boldsymbol{\Gamma}'\boldsymbol{\Gamma}'^{\mathrm{T}}] \\
&= (\lambda + 1 + 2^{1-p})^2 \det[\mathbf{1}_2 - [1 + 2^{1-p}]\boldsymbol{\Gamma}'\boldsymbol{\Gamma}'^{\mathrm{T}}/(\lambda + 1 + 2^{1-p})] \\
&= (\lambda + 1 + 2^{1-p})[(\lambda + 1 + 2^{1-p}) - [1 + 2^{1-p}]\boldsymbol{\Gamma}'^{\mathrm{T}}\boldsymbol{\Gamma}']\ .
\end{aligned} \tag{D.64}$$

Offenbar ist ein Eigenwert $\lambda = -1 - 2^{1-p}$ und somit vom Betrag her größer als eins, ganz gleich wie wir $\boldsymbol{\Gamma}'$ wählen.

Als Ergebnis halten wir zum einen fest, dass wir für das im Bild D.4 dargestellte Signalflussdiagramm bei Verwendung des Betragsschneidens eine passive Realisierung gefunden haben. Hingegen waren wir bei Verwendung der Gitterpunkt-Rundung nicht in der Lage, ein Signalflussdiagramm passiv zu realisieren. Abschließend sei erwähnt, dass sich die gefundenen Ergebnisse auf Serienadaptoren verallgemeinern lassen.

D.3 Direkte Umsetzung einer Zweitor-Adaptor-Streumatrix

Leider sind die meisten C-Compiler (aufgrund des Zielsystems) auf Gitterpunktschneiden eingerichtet. Wie in D.2 nachgewiesen wurde, führt dies dazu, dass die Implementierung nicht mehr passiv ist. Ein Betragsschneiden müsste daher softwareseitig implementiert werden. Dies ist zwar möglich, stellt sich aber als zu rechenaufwendig heraus und ist für die Praxis völlig ungeeignet. Als Alternative bietet sich die direkte Umsetzung der Streumatrix mit gestauchten Koeffizienten an. Die Berechnung erfolgt im idealen Fall durch

$$b_1 = S_{11}a_1 + S_{12}a_2 \quad ; \quad b_2 = S_{21}a_1 + S_{22}a_2 \tag{D.65}$$

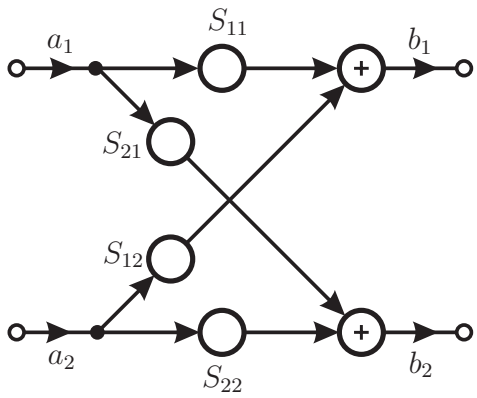


Bild D.6: Signalflussdiagramm eines realen Zweitor-Paralleladaptors bei direkter Umsetzung.

und im realen Fall durch

$$\tilde{b}_1 = S_{11} \odot a_1 \oplus S_{12} \odot a_2 \quad ; \quad \tilde{b}_2 = S_{21} \odot a_1 \oplus S_{22} \odot a_2 \; , \tag{D.66}$$

wobei auch hier Punkt- vor Strichrechnung gilt. Im Bild D.6 ist das Signalflussdiagramm eines realen Zweitor-Paralleladaptors bei direkter Umsetzung dargestellt.
Die Berechnung erfolgt im realen Fall mit gestauchten Koeffizienten $|S'_{\mu\nu}| < |S_{\mu\nu}|$ für $\mu, \nu = 1, 2$ durch

$$\tilde{b}'_1 = S'_{11} \odot a_1 \oplus S'_{12} \odot a_2 \quad ; \quad \tilde{b}'_2 = S'_{21} \odot a_1 \oplus S'_{22} \odot a_2 \; . \tag{D.67}$$

Wir müssen zunächst eine Stauchung aufgrund der fehlerhaften Multiplikation vornehmen. Der Fehler bei der Multiplikation kann gemäß Gleichung (D.27) durch

$$[1 - 2^{-p}] < \frac{|S'_{11} \odot a_1|}{|S'_{11} a_1|} < [1 + 2^{-p}] \tag{D.68}$$

eingegrenzt werden. Nehmen wir den schlimmsten Fall an, dass alle 4 Multiplikationen eine Erhöhung der Leistung bewirken, so gilt für die Norm der Ausgangswellen $||\tilde{\boldsymbol{b}}'|| < (1 + 2^{-p})||\boldsymbol{b}'||$. Setzen wir $\boldsymbol{b}' = \boldsymbol{S}'\boldsymbol{a}$ ein, so ergibt sich

$$||\tilde{\boldsymbol{b}}'|| < (1 + 2^{-p})||\boldsymbol{S}'\boldsymbol{a}|| \; . \tag{D.69}$$

Für die Passivität muss $||\tilde{\boldsymbol{b}}'|| \leq ||\boldsymbol{a}||$ gelten. Diese Bedingung wird von $\boldsymbol{S}' = \boldsymbol{S}/(1 + 2^{-p})$ erfüllt, wie durch Einsetzen in Gleichung (D.69) nachprüfbar ist

$$||\tilde{\boldsymbol{b}}'|| < (1 + 2^{-p})||\boldsymbol{S}'\boldsymbol{a}|| = (1 + 2^{-p})||\boldsymbol{S}\boldsymbol{a}/(1 + 2^{-p})|| = ||\boldsymbol{S}\boldsymbol{a}|| = ||\boldsymbol{a}|| \; . \tag{D.70}$$

Desweiteren müssen wir eine Stauchung der Matrixelemente aufgrund der fehlerhaften Addition durchführen, da durch die reale Addition eine Erhöhung der Leistung auftreten kann. Mit

$$\tilde{b}'_1 = S'_{11} a_1 \oplus S'_{12} a_2 \tag{D.71}$$

gilt die Eingrenzung nach Gleichung (D.15)

$$[1 - 2^{-p}] < \frac{|S'_{11} a_1 \oplus S'_{12} a_2|}{|S'_{11} a_1 + S'_{11} a_2|} < [1 + 2^{-p}] \; . \tag{D.72}$$

Der zur Erhöhung der Ausgangsleistung führende Fall ist durch $|S'_{11} a_1 \oplus S'_{12} a_2| < |[S'_{11} a_1 + S'_{11} a_2][1 + 2^{-p}]|$ beschränkt. Bewirken beide Additionen eine Erhöhung der Leistung, so gilt

$$||\tilde{\boldsymbol{b}}'|| < (1 + 2^{-p})||\boldsymbol{S}'\boldsymbol{a}|| \; . \tag{D.73}$$

Für die Passivität muss $||\tilde{\boldsymbol{b}}'|| \leq ||\boldsymbol{a}||$ gelten. Dies ist dann erfüllt, wenn sich die modifizierte Streumatrix zu $\boldsymbol{S}' = \boldsymbol{S}/[1+2^{-p}]$ berechnet, denn auch hier ergibt sich durch Einsetzen

$$||\tilde{\boldsymbol{b}}'|| < (1+2^{-p})||\boldsymbol{S}'\boldsymbol{a}|| = (1+2^{-p})||\boldsymbol{S}\boldsymbol{a}/(1+2^{-p})|| = ||\boldsymbol{S}\boldsymbol{a}|| = ||\boldsymbol{a}|| \; . \tag{D.74}$$

Fassen wir die Änderungen wegen fehlerbehafteter Multiplikationen und Additionen zusammen, so berechnet sich insgesamt die modifizierte Streumatrix zu

$$\boldsymbol{S}' = \frac{\boldsymbol{S}}{[1+2^{-p}]^2} \; . \tag{D.75}$$

Zum Leistungsvergleich berechnen wir mit $\boldsymbol{b}' = \boldsymbol{S}'\boldsymbol{a} = \boldsymbol{S}\boldsymbol{a}/[1+2^{-p}]^2$ die Dämpfung in dB

$$\begin{aligned} A_{\mathrm{dB}} &= 10\lg\left(\frac{\|\boldsymbol{a}\|^2}{\|\boldsymbol{b}'\|^2}\right) = 20\lg\left(\frac{[1+2^{-p}]^2\|\boldsymbol{a}\|}{\|\boldsymbol{S}\boldsymbol{a}\|}\right) = 20\log_{10}\left([1+2^{-p}]^2\right) \approx 20\log_{10}\left(\frac{1+2^{-p}}{1-2^{-p}}\right) \\ &= \frac{20}{\ln 10}\ln\left(\frac{1+2^{-p}}{1-2^{-p}}\right) \approx \frac{20}{\ln 10}(2\cdot 2^{-p}) = \frac{10}{\ln 10}2^{2-p} \approx 2^{4-p} \; . \end{aligned} \tag{D.76}$$

Für $p = 24$ ergibt sich $A_{\mathrm{dB}} \approx 2^{-20}\mathrm{dB} \approx 1000^{-2}\mathrm{dB} = 1\mu\mathrm{dB}$.

Wir wollen festhalten, dass sich die Ergebnisse auch hier auf Serienadaptoren verallgemeinern lassen. Zum Abschluss fassen wir das Vorgehen zur Bestimmung der Multipliziererkoeffizienten bei direkter Implementierung nach Bild D.6 zusammen

1. Berechnung von $\boldsymbol{S}$
2. Stauchung der Streumatrix zu $\boldsymbol{S}' = \boldsymbol{S}/[(1+2^{-p})]^2]$
3. Quantisierung der Elemente von $\boldsymbol{S}'$ durch Betragsschneiden.

Literaturverzeichnis

[Babe95] R. L. Baber: *Praktische Anwendbarkeit mathematisch rigoroser Methoden zum Sicherstellen der Programmkorrektheit*, Walter de Gruyter Verlag, Berlin, 1995.

[Bele68] V. Belevitch: *Classical network theory*, Holden-Day, Inc., 1968.

[BF87] S. Basu, A. Fettweis: "New Results on Stable Multidimensional Polynominals-Part II: Discrete Case", *IEEE Transactions on Circuits and Systems, Bd. 34, Nr. 11, S. 1264-1274*, 1987.

[Bilb01] S. D. Bilbao: "Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations", Dissertation an der Stanford University, 2001.

[Bilb04] S. D. Bilbao: *Wave and Scattering Methods for the Numerical Simulation*, John Wiley and Sons. Chichester, UK, 2004, 2004.

[Boeh76] B. Boehm: "Software engineering", *IEEE Transactions on Electronic Computers* **25**(12), 1976.

[Boeh81] B. Boehm: *Software Engineering Economics*, Prentice-Hall, 1981.

[Boeh86] B. Boehm: "A spiral model of program development and enhancement", *Software Engineering Notes* **11**(4), 1986, S. 14–24.

[Bose79] N. K. Bose (Hrsg.): *Multidimensional Systems: Theory and Application*, IEEE Press Selected Reprint Series, New York: IEEE Press, 1979.

[Bose82] N. K. Bose: *Applied Multidimensional System Theory*, Van Nordstrand Reinhold Company, 1982.

[Bose01] N. K. Bose: "Multidimensional Signal Processing", Vorlesung an der Ruhr-Universität Bochum, 2001.

[Brey96] U. Breymann: *C++ Eine Einführung*, Carl Hanser Verlag, 1996.

[Bun96] "IT-Phasenmodell (Vorgehensmodell V-Modell 97)", *Bundesanzeiger, Nr.125, Seite 7722ff.*, 1996.

[BV03] S. Buchholz, M. Vollmer: "Ein Werkzeug zur automatische Codegenerierung zur numerischen Integration linearer, konstanter, partieller Differentialgleichungen mit Hilfe von Wellendigitalfiltern", Interner Bericht des Lehrstuhls für Nachrichtentechnik der Ruhr-Universität Bochum, 2003.

[Dahl63] G. Dahlquist: "A special stability problem for linear multistep methods", *BIT 3, S. 27-43*, 1963.

[Dijk68] E. W. Dijkstra: "A Constructive Approach to the Problem of Program Correctness", *BIT Nord. Tidskr. Inform. Bd. 8, S. 174-186*, 1968.

[FB87] A. Fettweis, S. Basu: "New Results on Stable Multidimensional Polynominals-Part I: Continuous Case", *IEEE Transactions on Circuits and Systems, Bd. 34, Nr. 10, S. 1221-1232*, 1987.

[Feld95] T. Felderhoff: "Digitale Simulation nichtlinearer Systeme mit Methoden der Netzwerktheorie", Dissertation an der Universität-Gesamthochschule Paderborn, 1995.

[Fett70] A. Fettweis: "Entwurf von Digitalfiltern in Anlehnung an Verfahren der klassischen Netzwerktheorie", *NTZ-Fachtagung 15./16.10.1970 Stuttgart*, 1970.

[Fett78] A. Fettweis: "Suppression of parasitic oscillations in multidimensional wave digital filters", *IEEE Transactions on Circuits and Systems, Bd. 25, Nr. 12, S. 1060-1066*, 1978.

[Fett79] A. Fettweis: "Principles of multidimensional wave digital filtering". in J. K. Aggarwal (Hrsg.), *Digital Signal Processing*, S. 261–282. Western Periodicals, Point Lobos Press, North Hollywood, CA, USA, 1979.

[Fett86] A. Fettweis: "Wave digital filters: Theory and Practice", *(invited paper) Proceedings of the IEEE (The Institute of Electrical and Electronics Engineers), Bd. 74, Nr. 2, S. 270-327*, Februar 1986.

[Fett90] A. Fettweis: "On assessing robustness of recursive digital filters", *European Transactions on Telecommunications, Bd. 1, Nr. 2, S. 103-109*, 1990.

[Fett92] A. Fettweis: "Discrete passive modelling of physical systems described by PDEs", *European Signal Proceeding Conference (EUSIPCO), Bd. 1, S. 55-62 Brüssel, Belgien, 24.08. - 27.08.*, 1992.

[Fett98] A. Fettweis: "Numerische Integration nach dem Wellendigitalfilterprinzip", Vorlesung an der Ruhr-Universität Bochum, 1998.

[Fett99] A. Fettweis: "Numerische Integration nach dem Wellendigitalfilterprinzip", Vorlesung an der Ruhr-Universität Bochum, 1999.

[FH92] A. Fettweis, G. Hemetsberger: *Grundlagen der Theorie elektrischer Schaltungen*, Universitätsverlag Dr. N. Brockmeyer, 1992.

[Fisc84] H. D. Fischer: "Wave digital filters for numerical integration", *ntz-Archiv, Bd. 6, Nr. 2, S. 37-40*, 1984.

[Fisc99] H. D. Fischer: "Numerische Optimierung", Vorlesung an der Ruhr-Universität Bochum, 1999.

[Fisc00a] H. D. Fischer: "Ergänzungen zur Berechnung elektrischer Netze im stationären Zustand", Vorlesung an der Ruhr-Universität Bochum, 2000.

[Fisc00b] H. D. Fischer: "Persönliche Mitteilung", 2000.

[Fisc03] H. D. Fischer: "Technische Zuverlässigkeit", Vorlesung an der Ruhr-Universität Bochum, 2003.

[Floy67] R. W. Floyd: "Assigning Meanings to Programs", *Proceedings of the Symposium of Applied Mathematics, Bd. 19, S. 19-32, American Mathematical Society, Providence, RI*, 1967.

[FM75] A. Fettweis, K. Meerkötter: "Suppression of parasitic oscillations in wave digital filters", *IEEE Transactions on Circuits and Systems, Bd. 22, Nr. 3, S. 239-246*, 1975.

[FN90a] A. Fettweis, G. Nitsche: "Massively parallel algorithms for numerical integration of partial differential equations", *International Workshop on Algorithms and Parallel VLSI Architectures, Summaries of Contributions, S. 475-484, Pont-à-Mousson, France*, 1990.

[FN90b] A. Fettweis, G. Nitsche: "Numerical integration of partial differential equations by means of multidimensional wave digital filters", *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '90), Bd. 2, S. 954-957,New Orleans, LA, USA, 30 May - 2 June*, 1990.

[FN91a] A. Fettweis, G. Nitsche: "Numerical integration of partial differential equations using principles of multidimensional wave digital filters", *Journal of VLSI Signal Processing, Bd. 3, S. 7-24, Kluwer Academic Publishers, Boston, MA, USA*, 1991.

[FN91b] A. Fettweis, G. Nitsche: "Transformation approach to numerically integrating PDEs by means of WDF principles", *Multidimensional Systems and Signal Processing, Bd. 2, Nr. 2, S. 127-159*, 1991.

[Fränken] D. Fränken: "Passive Systeme zur Verarbeitung komplexer zeitdiskreter Signale", Dissertation an der Universität-Gesamthochschule Paderborn, 1997.

[Frie54] K. O. Friedrichs: "Symmetric hyperbolic differential equations", *Comm. Pure Appl. Math., Bd. 7, S. 354-392*, 1954.

[Frie95] M. Fries: "Numerical integration of Euler flow by means of multidimensional wave digital principles", Dissertation an der Ruhr-Universität Bochum, 1995.

[Frie00] M. Fries: "Persönliche Mitteilung", 2000.

[FS99] A. Fettweis, G. A. Seraji: "New results in numerically integrating PDEs by the wave digital approach", *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '99), Bd. 5, S. 17-20, Orlando, FL, USA, 30 May - 2 June*, 1999.

[HA28] D. Hilbert, W. Ackermann: *Grundzüge der theoretischen Logik*, Springer-Verlag, 1928.

[Heme95] G. Hemetsberger: "Numerische Integration hyperbolischer partieller Differentialgleichungen unter Verwendung mehrdimensionaler Wellendigitalfilter", Dissertation an der Ruhr-Universität Bochum, 1995.

[Hero99] H. Herold: *C-Kompaktreferenz*, ADDISON-WESLEY LONGMAN, 1999.

[HNW93] E. Hairer, S. P. Nørsett, G. Wanner: *Solving Ordinary Differential Equations I*, Springer-Verlag, 1993.

[Hoar69] C. Hoare: "An Axiomatic Basis for Computer Programming", *CACM, Bd. 12, Nr. 10, S. 576-580+583*, 1969.

[Huan81] Huang: *Two-Dimensional Digital Signal Processing I*, Springer-Verlag, 1981.

[Inte99] International Electrotechnical Commission: “Programming languages – C”, *ISO/IEC 9899, International Electrotechnical Commission,Genf*, 1999.

[Juss00] J. Jussen: “Untersuchungen zur automatischen Codegenerierung einer steuerbaren, eindimensionalen Wellendigitalstruktur”, Studienarbeit an der Ruhr-Universität Bochum, 2000.

[Koga69] T. Koga: “Synthesis of passive n-ports with prescribed positive real matrices of several variables”, *IEEE Transactions on Circuit Theory, Bd. 15, S. 2-23*, 1969.

[Krau97] H. Krauß: “Simulation linearer dissipativer Wellenausbreitungsvorgänge mit mehrdimensionalen digitalen Filtern”, Dissertation an der Friedrich Alexander Universität Erlangen-Nürnberg, 1997.

[Kumm88] A. Kummert: “Beiträge zur Synthese mehrdimensionaler Reaktanzmehrtore”, Dissertation an der Ruhr-Universität Bochum, 1988.

[KWU 98] KWU NLLZ ST: “Systemübersicht zum digitalen Leitsystem Teleperm XS”, 1998.

[KWU 99] KWU NLLZ ST: “Entwicklungsunterlagen : Design FB ModuleDesign”, 1999.

[LF90] X. Liu, A. Fettweis: “Multidimensional Digital Filtering by Using Parallel Algorithms Based on Diagonal Processing”, *Multidimensional Systems and Signal Processing, S. 51-66*, 1990.

[Linn84] G. Linnenberg: “Über die diskrete Verarbeitung mehrdimensionaler Signale unter Verwendung mehrdimensionaler Wellendigitalfilter”, Dissertation an der Ruhr-Universität Bochum, 1984.

[Lisc91] J. Lischetzki: “Untersuchungen zur Genauigkeit von Wellendigitalfilterverfahren zur numerischen Integration partieller Differentialgleichungen”, Diplomarbeit an der Ruhr-Universität Bochum, 1991.

[Luhm04] K. Luhmann: “Die numerische Lösung der Neutronendiffusionsgleichungen in zwei Energiegruppen mit dem Wellendigital-Konzept”, Dissertation an der Ruhr-Universität Bochum, 2004.

[McCa62] J. McCarthy: “Towards a Mathematical Science of Computation”, *Proceedings of the IFIP Congress, S.21-28, North-Holland, Amsterdam*, 1962.

[McCa63] J. McCarthy: “A Basis of Mathematical Theory of Computations”, *Computer programming and formal systems, S.33-70, North-Holland, Amsterdam*, 1963.

[Meer79] K. Meerkötter: “Beiträge zur Theorie der Wellendigitalfilter”, Dissertation an der Ruhr-Universität Bochum, 1979.

[Meye88] B. Meyer: *Objektorientierte Softwareentwicklung*, Hanser, 1988.

[MF92] K. Meerkötter, T. Felderhoff: “Simulation of nonlinear transmission lines by wave digital filter principles”. in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS ’92), Bd. 2, S. 875-878, San Diego, CA, USA*, 1992.

[MF96] K. Meerkötter, D. Fränken: “Digital Realization of Connection Networks by Voltage-Wave Two-Port Adaptors”. in *Archiv für Elektronik und Übertragungstechnik 50*, 1996.

[Nits93] G. Nitsche: "Numerische Lösung partieller Differentialgleichungen mit Hilfe von Wellendigitalfiltern", Dissertation an der Ruhr-Universität Bochum, 1993.

[Novi73] P. S. Novikov: *Mathematische Logik*, Friedrich Vieweg und Sohn Verlag, 1973.

[Ochs01a] K. Ochs: "Passive Integration Methods : Fundamental Theory", *Archiv für Elektronik und Übertragungstechnik, Bd. 55, Nr. 3, S. 153-163*, 2001.

[Ochs01b] K. Ochs: "Passive Integrationsmethoden", Dissertation an der Universität-Gesamthochschule Paderborn, 2001.

[Ochs02] K. Ochs: "Persönliche Mitteilung", 2002.

[Pott98] R. Pott: "Numerische Integration von Neutronendiffusionsgleichungen unter Verwendung mehrdimensionaler Wellendigitalfilter auf parallelen Rechnerarchitekturen", Dissertation an der Ruhr-Universität Bochum, 1998.

[Rao69] T. Rao: "Minimal synthesis of two-variable reactance matrices", *Bell Systems Technical Journal, S. 163-199*, 1969.

[Royc87] W. Royce: "Managing the development of large software systems: Concepts and techniques", 1987, S. 328–338. Nachdruck aus dem Jahre 1970.

[Rumm98] E. Rummert: "Methodik eines formalen Korrektheitsbeweises bei graphisch spezifizierter Software", Dissertation an der Ruhr-Universität Bochum, 1998.

[Sera00] G. A. Seraji: "Randwertfragen bei der numerischen Integration mit Hilfe der Wellendigitalmethode", Dissertation an der Ruhr-Universität Bochum, 2000.

[Shek74] J. Shekel: "The junction matrix in the analysis of scattering networks", *IEEE Transactions on Circuits and Systems, Bd. 21, Nr. 1, S. 21-22*, 1974.

[Smid75] D. Smidt: *Reaktortechnik*, G. Braun,Karlsruhe, 1975.

[Stra80] G. Strang: *Linear Algebra and Its Applications*, Harcourt Brace Jonanovich, 1980.

[Tell54] B. D. H. Tellegen: *Theorie der Electrische Netwerken*, Nordhoff, 1954.

[Voll02] M. Vollmer: "Automatische Codegenerierung zur numerischen Integration linearer, konstanter, partieller Differentialgleichungen mit Hilfe von Wellendigitalfiltern", Interner Bericht des Lehrstuhls für Nachrichtentechnik der Ruhr-Universität Bochum, 2002.

[Voll04a] M. Vollmer: "An approach to automatic generation of wave digital structures from PDEs", *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '04), Vancouver, Canada, 23 May - 26 May*, 2004.

[Voll04b] M. Vollmer: "Numerical Integration of PDEs for Safety Critical Applications implemented by I & C Systems", *The 23rd International Conference on Computer Safety, Reliability and Security, 21-24 September 2004, Potsdam, Germany*, 2004.

[Voll04c] M. Vollmer: "Reference circuits for nonenergic multi-ports without directed loops in the wave domain", *zur Veröffentlichung eingereicht*, 2004.

[Waed94] K. Waedt: "Fehlervermeidende Codegenerierung für verteilte, responsive Systeme", Dissertation an der Friedrich Alexander Universität Erlangen-Nürnberg, 1994.

[Waed00] K. Waedt: “Persönliche Mitteilung”, 2000.

Lebenslauf

Persönliche Daten

Name:	Michael Vollmer
Geburtsdatum:	14. Dezember 1971
Geburtsort:	Hamm
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

Schulbildung

08/1978 – 07/1982	Grundschule in Hamm
08/1982 – 06/1988	Realschule in Hamm
08/1991 – 07/1992	Fachoberschule in Hamm

Berufsausbildung

09/1988 – 06/1991	Ausbildung zum Energieelektroniker bei der Deutschen Bundesbahn in Hamm

Wehrdienst

10/1992 – 09/1993	Grundwehrdienst in Hamm

Studium

10/1993 – 05/1998	Elektrotechnik mit der Vertiefungsrichtung Informationstechnik in Paderborn

Berufstätigkeit

09/1998 – 09/2004	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Nachrichtentechnik der Ruhr–Universität Bochum