Berechnung hochfrequenter 3-dimensionaler elektromagnetischer Felder mittels der Finite Elemente Methode auf Kantenbasis

vorgelegt von
Diplom-Ingenieur
Tobias Damm
aus Berlin

Vom Fachbereich 12 - Elektrotechnik -
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor-Ingenieur
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuß:
Vorsitzender: Prof. Dr.-Ing. Reinhold Orglmeister
Berichter: Prof. Dr.-Ing. Heino Henke
Berichter: Prof. Dr.-Ing. Manfred Kasper

Tag der wissenschaftlichen Aussprache: 11. Februar 2000

Berlin 2000
D83

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Theoretische Elektrotechnik und Nachrichtentechnik an der Technischen Universität unter der Betreuung von Herrn Prof. Dipl.-Ing. W. Schacht, dem ich an dieser Stelle meinen besonderen Dank aussprechen möchte. Herrn Prof. Dr.-Ing. M. Kasper von der TU-Hamburg, Prof. Dr.-Ing. H. Henke und meinen Kollegen danke ich für die anregenden Diskussionen und für das der Arbeit entgegengebrachte Interesse.

Zusammenfassung

In der vorliegenden Arbeit wird die Berechnung von 3-dimensionalen elektromagnetischen Feldern mittels der Finite Elemente Methode (FEM) unter der besonderen Verwendung von kantenbasierten Ansatzfunktionen entwickelt und verifiziert. Dafür ist im Rahmen dieser Arbeit das dreidimensionale FEM-Paket IFEP entwickelt worden.

Zuerst wird auf die Arbeitsweise der FEM eingegangen. Dabei stehen Methodik und Anschauung im Vordergrund. Die Probleme der Beschreibung und die Diskretisierung in Tetraeder des Berechnungsgebiets werden anschließend diskutiert. Dabei wird die Bewertung der Qualität von Diskretisierungen und ihre Verbesserungsmöglichkeiten behandelt. Anschließend werden die kantenbasierten Ansatzfunktionen mit der Ordnung $0,5$ im Tetraeder eingeführt. Sie sind Grundlage für alle Berechnungen und werden ausführlich hergeleitet und anschaulich dargestellt. Zusätzlich wird gezeigt, daß die Ansatzfuntionen im Element divergenzfrei und rotationskonform sind.

Anschließend wird die korrekte Anwendung der FEM mit Kantenelementen auf die vektorielle Helmholtzgleichung beschrieben. Dabei werden von der Problembeschreibung über die Einführung der Randbedingungen, der Kontinuitätsbedingung, und der Formulierung der Wellengleichung als Variationsproblem bis zum Aufstellen des zu lösenden linearen Gleichungssystems alle Schritte der FEM beschrieben, die in IFEP zur Anwendung kommen. Dann wird die Speicherung und das Lösen des Gleichungssystems behandelt. Dabei werden Themen wie die Reduktion der Bandbreite der Matrix bei Kantenelementen oder die Reduktion des Gleichungssystems durch Eleminierung innerer Freiheitsgrade untersucht. Besonders werden iterative Löser behandelt, wobei der Zusammenhang zwischen Feldlösung und Konvergenz an Beispielen untersucht wird. Es werden zwei Verfahren vorgestellt, um Streuparameter in Wellenleitern zu berechnen. Am Schluß werden die Ergebnisse mit anderen Veröffentlichungen, anderen Berechnungsmethoden und Messungen auf ihre Genauigkeit hin überprüft und diskutiert.

Inhaltsverzeichnis

1 Einleitung **7**

2 Die Finite-Elemente-Methode **11**

2.1 Problemstellung 11

2.2 Idee der finiten Elemente 12

2.3 Lösungsmethoden der FEM 14

2.3.1 Globale Ansatzfunktionen 14

2.3.1.1 Die Methode von Ritz 16

2.3.1.2 Die Methode von Galerkin 17

2.3.2 Lokale Ansatzfunktionen (FEM) 18

2.3.2.1 Lokale Ritzsche Finite-Elemente-Methode 19

2.3.2.2 Lokale Galerkinsche Finite-Elemente-Methode 20

2.3.3 Randbedingungen in der FEM 21

3 Diskretisierung bzw. Netzgenerierung **23**

3.1 Volumenbeschreibung 23

3.2 Netzgenerierung 24

3.2.1 Qualität eines Tetraedernetzes 24

3.2.2 Probleme der Netzgenerierung 25

3.2.2.1 Konvexe Gebietszerlegung 25

3.2.2.2 Schlechte Netzqualität 26

3.3 Netzverbesserung 26

3.3.1 Lokale Transformation 27

3.3.2 Knoten-Relaxation 27

3.3.3 Anwendung der Netzverbesserung 28

4 Die Kantenelemente **31**
4.1 Das lokale Koordinatensystem . . . 32
4.1.1 Baryzentrische Koordinaten im Dreieck . . . 32
4.1.1.1 Der Gradient einer Koordinate . . . 33
4.1.2 Baryzentrische Koordinaten im Tetraeder . . . 34
4.1.2.1 Der Gradient einer Koordinate . . . 34
4.2 Die kantenbasierten Ansatzfunktionen im Element . . . 35
4.2.1 Kantenbasierte Ansatzfunktion im Dreieck . . . 35
4.2.1.1 Der Ansatz der Vektorbasisfunktion . . . 36
4.2.1.2 Normierung der Vektorbasisfunktion . . . 37
4.2.2 Kantenbasierte Ansatzfunktion im Tetraeder . . . 40
4.2.2.1 Der Ansatz der Vektorbasisfunktion . . . 41
4.2.2.2 Normierung der Vektorbasisfunktion . . . 41
4.3 Die Divergenz im Element . . . 42
4.4 Die Rotation im Element . . . 42
4.5 Einordnung dieser Kantenelemente in der FEM . . . 43
4.5.1 Kantenelemente höherer Ordnung . . . 43
4.5.2 Konvergenzeigenschaften der Kantenelemente für die FEM . . . 44

5 Hochfrequente elektromagnetische Felder **45**
5.1 Problembeschreibung . . . 45
5.1.1 Die Wellengleichung im zeitharmonischen Fall . . . 45
5.1.2 Randbedingungen . . . 46
5.1.2.1 Elektrische Wand . . . 47
5.1.2.2 Magnetische Wand . . . 47
5.1.2.3 Absorbierende Randbedingungen . . . 48
5.1.3 Kontinuitätsbedingung . . . 52
5.2 Das Variationsproblem der Wellengleichung . . . 52
5.2.1 Beschränkungen bei der klassischen Ritz-Methode . . . 53
5.2.2 Ein modifiziertes Variationsverfahren . . . 54
5.2.3 Ein generelles Variationsverfahren der Wellengleichung . . . 55
5.3 Lokales Variationsproblem der Wellengleichung . . . 58
5.3.1 Das Volumenintegral . . . 59
5.3.2 Das Randintegral . . . 62
5.3.3 Das Stromintegral . . . 65
5.3.4 Das Elementgleichungssystem . . . 65
5.3.5 Aufstellen des Gesamtgleichungssystems . . . 66

6 Lösen des linearen Gleichungssystems 71
6.1 Datenstrukturen zur Matrixspeicherung . . . 71
6.1.1 Datenstrukturen für unsymmetrische Matrizen . . . 72
6.1.1.1 Datenstruktur vom CSR-Format . . . 73
6.1.2 Datenstrukturen für symmetrische Matrizen . . . 73
6.1.2.1 Datenstruktur des AWS-Formats . . . 73
6.2 Bandbreite der Matrix . . . 74
6.3 Optimale Numerierung der Kantenvariablen . . . 74
6.3.1 Der Algorithmus von Cuthill-McKee . . . 75
6.3.2 Probleme bei Kantenelementen . . . 75
6.3.3 Ein effizienter Algorithmus bei Kantenelementen . . . 76
6.4 Elimination von inneren Freiheitsgraden, Kondensation . . . 76
6.4.1 Kondensation . . . 77
6.5 Iterative Löser . . . 79
6.5.1 Bewertung der Löser . . . 81
6.5.1.1 Einfluß der Kondensation . . . 87
6.5.1.2 Einfluß von Resonanzen . . . 88

7 Auswertung der berechneten Felder 89
7.1 Visualisierung der Feldlösung . . . 89
7.1.1 Das elektrische Feld . . . 90
7.1.2 Das magnetische Feld . . . 90
7.1.3 Die Leistungsdichte . . . 90
7.1.4 Animation durch den Frequenzbereich (Wobbeln) . . . 91
7.2 Die Streuparameterberechnung . . . 91
7.2.1 Die einfache Anordnung (reflex) . . . 91
7.2.1.1 Die Reflexion . . . 92
7.2.1.2 Die Transmission . . . 94
7.2.2 Die Berechnung über die Leistung (power) . . . 94
7.2.2.1 Leistung durch ein Tor . . . 94
7.2.2.2 Die Streuparameter . . . 95

8 Anwendungen und Resultate 97
8.1 Resonatoren . . . 97
8.2 Rechteckwellenleiter . . . 98
8.2.1 Taper . . . 98
8.2.2 Wellenleiter mit dielektrischem Störkörper . . . 106
8.3 Hohlleiter-Bandpaß . . . 111

9 Zusammenfassung und Ausblick **117**

A Vektor- und Integralgleichungen **119**

A.1 Vektorgleichungen 119

A.2 Integralsätze 119

B Baryzentrische Koordinaten **121**

B.1 Dreieckskoordinaten 121

B.2 Tetraederkoordinaten 122

C Gauss-Legendre-Quadratur **125**

Verwendete Konstanten, Größen und Formeln

Konstanten im Vakuum:			
Magnetische Feldkonstante	μ_0	$=$	$4 \cdot \pi \cdot 10^{-7} \approx 1,257 \cdot 10^{-6} \ \left[\frac{Vs}{Am}\right]$
Elektrische Feldkonstante	ε_0	$\approx$	$8,854187817 \cdot 10^{-12} \ \left[\frac{As}{Vm}\right]$
Lichtgeschwindigkeit	c_0	$\approx$	$2,99792458 \cdot 10^{8} \ \left[\frac{m}{s}\right]$
Ausbreitungskonstante im Vakuum	k_0	$=$	$\omega\sqrt{\varepsilon_0\mu_0} \quad \left[\frac{1}{m}\right]$
Wellenwiderstand im freien Raum	Z_0	$=$	$\sqrt{\frac{\mu_0}{\varepsilon_0}} = 120 \cdot \pi\,\Omega \approx 376,7\,\Omega$
Mathematische Konstanten:			
	π	$\approx$	$3,141592653590$
Formel:			
Phasenmaß	β	$=$	$\frac{2\pi}{\lambda} = 2\pi f \cdot \sqrt{\mu_0\mu_r\varepsilon_0\varepsilon_r} = \frac{\omega}{v}$
Definitionen:			
Konjugiert Komplex	z^*	$=$	das Konjugiertkomplexe der Zahl z
Größen:			
Baryzentrische Koordinate i	ξ_i	$=$	L_i oder ζ_i (Bezeichnung in anderer Literatur)
Knoten i	P_i	$=$	Knotenpunkt i eines Elements
Flächennormale	$\widehat{n}$	$=$	Vektor der Länge 1 senkrecht auf der Fläche
Leitfähigkeit	κ	$=$	spezifische elektrische Leitfähigkeit $[\Omega^{-1}]$
Wellenlänge	λ	$=$	$\frac{2\cdot\pi}{\beta} \quad [m]$
Kreisfrequenz	ω	$=$	$2\,\pi\,f \quad [s^{-1}]$
Permeabilität	μ	$=$	$\mu_0\,\mu_r$
	μ_r	$=$	relative Permeabilität
Permitivität	ε	$=$	$\varepsilon_0\,\varepsilon_r$
	ε_r	$=$	relative Permitivität
Phasengeschwindigkeit	v	$=$	$\frac{\omega}{\beta} \quad \left[\frac{m}{s}\right]$

Kapitel 1

Einleitung

Heute ist der Begriff CAE (Computer Aided Engineering) in der Welt eines Ingenieurs keine Seltenheit mehr. Der computergestützte Arbeitsplatz eines Ingenieurs ist heute schon fast komplett in ein Computernetzwerk eingebunden. Wo früher der Ingenieur mit seinem Computer ein Insel-Dasein pflegte, ist er heute Teil eines großen, oft konzernübergreifenden, computergestützten Prozesses. So möchte man nach einem Entwurf an einem CAD[1]-Arbeitsplatz

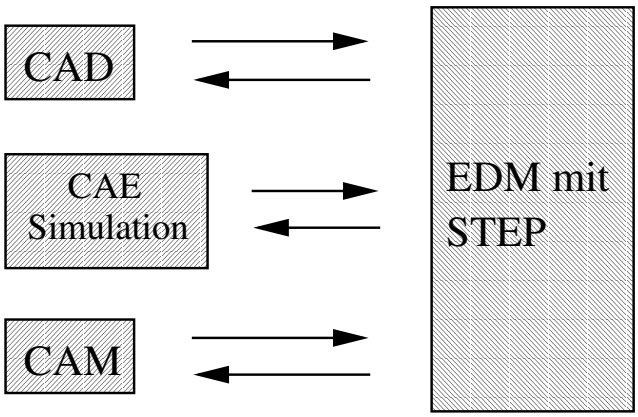


Abbildung 1.1: Computergestützter Entwicklungsprozeß

(s. Abb. 1.1) die Funktionsfähigkeit möglichst früh durch Simulationen verifizieren. Aus diesem Grund werden die Daten mit EDM[2] über den Industriestandard STEP[3] weiter zum CAE gegeben. Hier kann mittels Finite-Elemente-Methode oder anderer Simulationen der Entwurf überprüft oder auch verbessert werden. Nachdem die Simulation erfolgreich abgeschlossen wurde, kann der Entwurf über EDM weiter zum CAM[4] gegeben werden. Hier kann mit Rapid Prototyping ein Prototyp gefertigt werden oder auch gleich die Serienfertigung organisiert werden.

[1] engl. Computer Aided Design

[2] engl. Engineering Data Management

[3] STEP ist eine internationale Normung der ISO für Produktdaten, die vor ca. 13 Jahren begonnen wurde. Sie stellt heute wohl den umfassensden Standard für Produktdaten jeglicher Art dar. Der Abschnitt ISO 10303, der sich mit der Beschreibung von Formdaten beschäftigt, ist hier auch für FE-Daten erweitert worden. Der Vorteil dieser Norm liegt in der Möglichkeit sämtliche Produktdaten nach der gleichen Methode zu verarbeiten, zu speichern und zu transportieren. Damit ist es heute auch möglich, Projekte bei internationaler Zusammenarbeit auch mit FE-Daten leichter durchzuführen zu können [2] [3] [4] [5] [6].

[4] engl. Computer Aided Manufacturing

Nun ist es Aufgabe der Forschung, für diese Simulationen möglichst genaue und schnelle Werkzeuge zu schaffen, um den Entwicklungsprozeß möglichst kostengünstig und schnell zu gestalten. Seit vielen Jahren wird die Finite-Elemente-Methode (FEM) zur numerischen Berechnung elektromagnetischer Felder eingesetzt. Neben anderen Methoden, wie der Boundary Elemente Methode (BEM) und der Finite Differenzen Methode (FDM) wird die FEM wohl am häufigsten eingesetzt. Jede dieser Methoden hat je nach Lösungsansatz problemabhängige Vor- und Nachteile bei der Berechnung elektromagnetischer Felder. Dabei müssen abhängig vom Zeitverhalten der auftretenden elektromagnetischen Felder unterschiedliche physikalische Gesetzte simuliert werden.

In dieser Arbeit soll die numerische Berechnung 3-dimensionaler hochfrequenter elektromagnetischer Felder mittels der Finite-Elemente-Methode unter der besonderen Berücksichtigung von kantenbasierten Ansatzfunktionen behandelt werden.

Die FEM basiert auf der Diskretisierung des Berechnungsvolumens in kleine einfache Elemente. Für diese Elemente werden einfache Ansatzfunktionen gewählt. Mittels einer Optimierungsmethode werden dann die Koeffizienten der Ansatzfunktionen gesucht, die die herrschende Physik am besten beschreiben. Mit der Wahl der Ansatzfunktionen wird eine wichtige Entscheidung für die Lösbarkeit und Genauigkeit der Feldlösung getroffen.

Die ersten Ansatzfunktionen, die in der FEM angewendet wurden, waren knotenbasiert. Hier sind die unbekannten Koeffizienten Punkten im Element zugeordnet. Sie werden auch heute noch oft benutzt. Bei hochfrequenten elektromagnetischen Feldern führt ihre Verwendung jedoch zu fehlerhaften Ergebnissen, da sie die physikalische Lösung nicht richtig abbilden können.

Mit den hier angewendeten kantenbasierten Ansatzfunktion wird dagegen die Divergenzfreiheit des elektromagnetischen Feldes im Element erfüllt. Sie sind rotationskonform und erzeugen keine sogenannten *Geister*-Lösungen. Die Kontinuitätsbedingung an Materialübergängen wird nicht übererfüllt. Das heißt, daß hier wirklich nur die notwendige Kontinuitätsbedingung eingehalten wird und nicht mehr erzwungen wird. Zusätzlich lassen sich Singularitäten an elektrisch leitenden Kanten oder Ecken besser beschreiben.

Um die Effizienz und Zuverlässigkeit dieser Berechnungsmethode zu untersuchen, ist im Rahmen meiner Forschungsarbeit ein FEM-Programmpaket mit dem Namen IFEP entwickelt worden. Es ist in der graphischen Programmierumgebung GRAPE eingebettet [1] und in objektorientiertem C implementiert. Es ist modular aufgebaut und bietet die Möglichkeit, weitere Forschungsarbeiten hierauf aufzubauen.

Da an dieser Stelle nicht das gesamte Programmpaket beschrieben werden kann, sind in dieser Arbeit nur die wichtigsten Grundlagen und Ergebnisse zusammengefaßt. Im 2. Kapitel wird zunächst die Arbeitweise der Finite-Elemente-Methode beschrieben. Dabei steht die Methodik und die Anschauung im Vordergrund. Im 3. Kapitel wird die Diskretisierung des Berechnungsvolumens behandelt. Wobei Probleme der Geometriebeschreibung, der Netzgenerierung und der Netzverbesserung behandelt werden. Das 4. Kapitel beschäftigt sich mit den kantenbasierten Ansatzfunktionen in den zerlegten Elementen. Da sie Grundlage für alle Berechnungen und für das Verständnis der Arbeit sehr wichtig sind, werden sie hier ausführlich hergeleitet und anschaulich dargestellt. In Kapitel 5 steht die Berechnung der hochfrequenten elektromagnetischen Felder mit der FEM und den kantenbasierten Ansatzfunktion im Mittelpunkt. Hier wird der Rechenweg von den Maxwell'schen Gleichungen bis zum Gleichungssystem beschrieben. Das Lösen des linearen Gleichungssystems wird dann in Kapitel 6 untersucht. Dabei stehen hier iterative Löser im Vordergrund, die heute immer mehr zur Anwendung kommen.

Im 7. Kapitel wird gezeigt, welche Möglichkeiten in IFEP bestehen, um die berechneten Felder zu visualisieren und auszuwerten. Im 8. Kapitel sind an Beispielen mögliche Anwendungen dargestellt. Die erzielten Ergebnisse werden dabei mit anderen Berechnungen oder Messungen verglichen und diskutiert. In Kapitel 9 werden dann noch einmal alle wesentlichen Ergebnisse zusammengefaßt und Anregungen für mögliche zukünftige Forschungen und Entwicklungen gegeben.

Kapitel 2

Die Finite-Elemente-Methode

Für das Verständnis der später folgenden mathematischen Betrachtungen ist es sehr hilfreich, wenn zuvor eine Vorstellung von dem Problem und der Idee der Finite-Elemente-Methode aufgezeigt wird.

2.1 Problemstellung

In der Physik oder Technik ist ein System oft durch eine Differentialgleichung mit bestimmten Rand- oder Anfangsbedingungen beschrieben. Als Beispiel ist hier die Schwingung einer Saite genannt, die rechts und links befestigt ist:

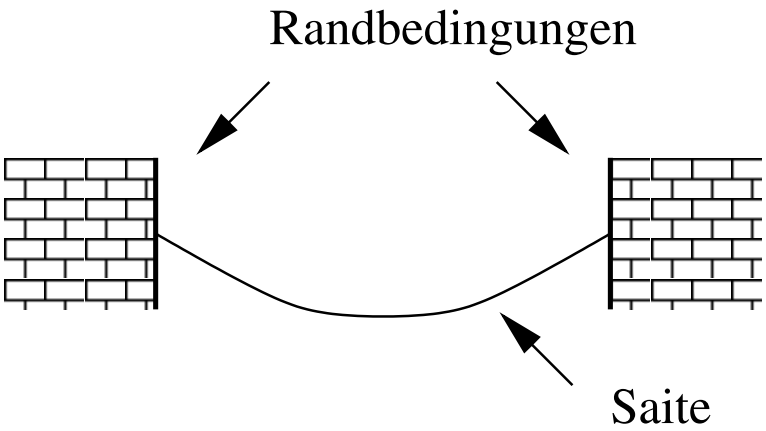


Abbildung 2.1: Schwingung einer Saite

Das Schwingungsverhalten ist bei kleinen Auslenkungen der Saite durch die eindimensionale Wellengleichung gegeben z.B.[7, S.475]:

$$\frac{\partial^2 u(t,x)}{\partial t^2} = a^2 \frac{\partial^2 u(t,x)}{\partial x^2} + f(t,x) \tag{2.1}$$

Dabei gibt $f(t,x)$ die Intensität der äußeren Störung an, a beschreibt die Eigenschaft des schwingenden Mediums und $u(t,x)$ beschreibt die Auslenkung an der Stelle x zur Zeit t und stellt die gesuchte Lösungsfunktion dar.

Unter der Annahme, daß die Saite frei schwingen kann und daß nach einer kleinen Störung bei $t = 0$ keine äußere Kraft mehr wirkt ($f(t > 0, x) = 0$), erhält man für $t > 0$:

$$\frac{\partial^2 u(t,x)}{\partial t^2} = a^2 \frac{\partial^2 u(t,x)}{\partial x^2} \tag{2.2}$$

Die Lösung dieses Randwertproblems kann leicht analytisch gefunden werden. **Es liegt auf der Hand, daß diese Lösung eine Überlagerung von Sinusfunktionen sein wird.** Dieses Wissen ist für die Wahl der Ansatzfunktion für $u(t, x)$ entscheidend. Hat man eine geeignete Ansatzfunktion für die Anfangs-Randwertaufgabe gefunden, so kann man eine Lösung analytisch bestimmen. Da es in der Regel bei partiellen Differentialgleichungen mit komplizierteren Randbedingungen nicht so einfach ist, einen geeigneten Lösungsansatz zu finden, mußte man andere Werkzeuge entwickeln.

2.2 Idee der finiten Elemente

Bei der Methode der **finiten Elemente** (engl. **F**inite **E**lement **M**ethod = kurz **FEM**) setzt man als Ansatzfunktion für die gesuchte Lösung $u(t, x)$ eine Näherung $u_{FEM}(t, x)$ an. Diese Näherung besteht aus dem Zerlegen des Gebiets in viele Teilgebiete, was dann die finiten Elemente der Länge h sind, und aus dem Wählen einer sehr einfachen Ansatzfunktion für jedes Teilgebiet (Element).

Man definiert nach [45, S.58] die Netzweite (oder Maschenweite) h bei m Elementen im Netz mit:

$$h = \max_{i=1,\ldots,m} Durchmesser(Element_i) \tag{2.3}$$

Wobei hier der Durchmesser die Länge eines Saitenstücks ist. Bei einem Dreieck oder Tetraeder ist es die längste Kante eines Elements.

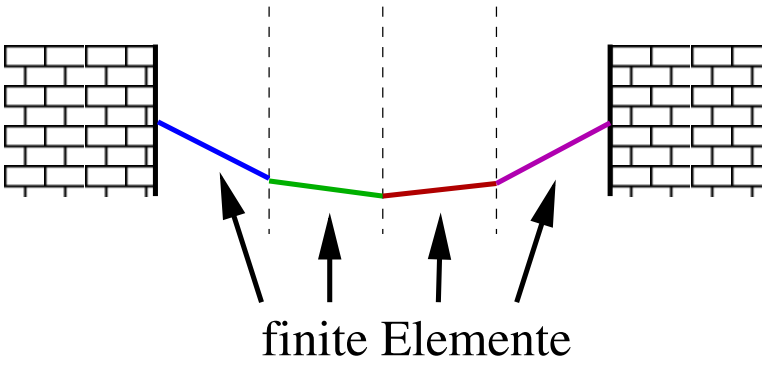


Abbildung 2.2: Qualitative Lösung $u_{FEM}(t, x)$ der Grundschwingung mit 4 finiten Elementen, die als Ansatzfunktion Geraden haben.

In Abbildung 2.2 sieht man qualitativ, wie eine Lösung aussehen könnte. Man sieht jedoch auch gleich die Grenzen dieser Methode. Die Näherung in Abbildung 2.2 für die Grundschwingung ist noch einigermaßen brauchbar. Für die nächste Oberschwingung ist die verwendete Gebietszerlegung schon sehr schlecht (siehe Abb. 2.3).

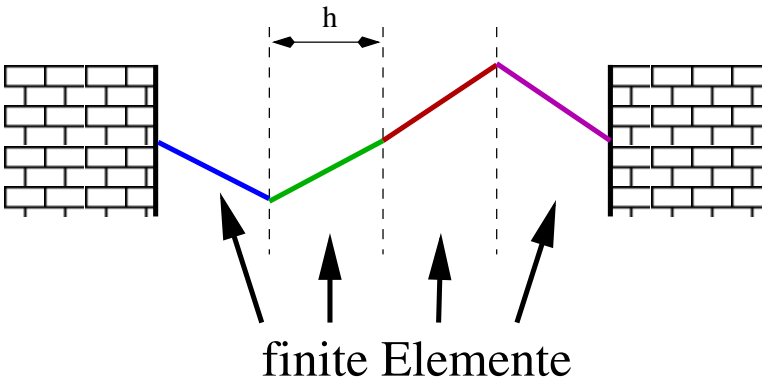


Abbildung 2.3: Erste Oberschwingung der Saite mit 4 finiten Elementen, die als Ansatzfunktionen Geraden haben.

Mit dieser Tatsache läßt sich das wichtige Kriterium (2.4) bei Wellenproblemen begründen:

$$Q_h = \frac{\lambda}{h} \tag{2.4}$$

Die Qualität der Diskretisierung Q_h ist direkt von den existierenden Wellenlängen λ abhängig. Mit kleinerem h, also mit mehr Elementen, wird die Qualität der Näherungslösung besser. Desweiteren muß Shannons[1] Abtasttheorem beachtet werden. D.h., die Bedingung

$$f_T > 2 \cdot f_g \tag{2.5}$$

zur Abtastung mit der Frequenz $f_T = \frac{1}{T}$ einer periodischer Funktionen, deren höchste darin enthaltene Frequenz f_g beträgt, muß mindestens eingehalten werden. Mit der Wellenlänge λ erhält man dann:

$$\frac{1}{h} > 2 \cdot \frac{1}{\lambda} \tag{2.6}$$

$$\lambda > 2 \cdot h \tag{2.7}$$

$$\frac{\lambda}{h} = Q_h > 2 \tag{2.8}$$

Das heißt, daß Q_h mindestens größer als 2 sein muß, möglichst aber noch größer. Das Abtasttheorem ist notwendig zum *Erkennen* einer Welle, hier ist jedoch der Verlauf der Lösungsfunktion von Interesse. Das bedeutet, daß Q_h noch größer sein muß. In die Genauigkeit der Näherungslösung geht jedoch neben h auch die Art der einfachen Ansatzfunktion ein. Je mehr es möglich ist, durch die einfache Ansatzfunktion die Lösung im Element zu nähern, um so genauer wird die Näherungslösung.

Hier ist als Ansatzfunktion in einem Element eine Gerade verwendet worden. Für die Beschreibung einer Geraden sind nur zwei Punkte notwendig, nämlich der Anfangspunkt und der Endpunkt. Man nennt Elemente, deren lokale Ansatzfunktion durch Punkte (auch Knoten genannt) beschrieben werden, auch Knotenelemente. Wenn sich zwei benachbarte Elemente

[1] Der Grundgedanke des Abtasttheorems geht bis zu J.L. Lagrange (1736-1813) zurück. Er zeigte, daß zur Darstellung einer periodischen Funktion durch eine trigonometrische Reihe mit je n *cos*− und *sin*−Gliedern die Kenntnis von $2n$ äquidistanten Funktionswerten einer Periode genügen. In der Nachrichtentechnik wurde das Abtasttheorem 1948 durch Claude E. Shannon bekannt, weshalb es heute auch Shannons Abtasttheorem genannt wird.

dann einen Punkt teilen, ist die genäherte Ansatzfunktion $u_{FEM}(t,x)$ **immer stetig**. Das ist bei einer schwingenden Saite auch richtig.

Bei den später behandelten elektromagnetischen Wellen ist diese Bedingung aber nicht immer gefordert. So springt zum Beispiel die Normalkomponente der elektrischen Feldstärke an Grenzflächen mit unterschiedlichem Dielektrikum. Mit der Wahl von Knotenelementen erzwingt man jedoch diese Stetigkeit, was folglich zu Fehlern führen muß. **Kantenelemente erlauben im Gegensatz dazu Unstetigkeiten in der Normalkomponente zwischen den Elementen.** Weiter soll dies hier nicht vertieft werden, da in Kapitel 4 auf Seite 31 das Thema genauer untersucht wird.

Im verbleibenden Teil des Kapitels sollen die Methoden zum Finden solcher Näherungslösungen, wie zum Beispiel $u_{FEM}(t,x)$, behandelt werden.

2.3 Lösungsmethoden der FEM

Es gibt für die FEM mehrere Verfahren die Lösungsfunktion zu finden, z.B. Kollokations-Methode, Untergebiets-Methode, Methode der kleinsten Quadrate, Galerkin-Methode und die Ritz-Methode [9, S.57] [24, S.12]. Dabei sind die Methoden von Ritz und Galerkin die heute am meisten verwendeten in der FEM, weshalb hier nur diese beiden vorgestellt werden. Im folgenden wird die Näherungslösung mit $\widetilde{\phi}$ bezeichnet und die exakte Lösung mit ϕ.

Allen Methoden gemein ist das grundsätzliche Vorgehen:

$\Rightarrow$ Aufstellen einer Ansatzfunktion $\widetilde{\phi}$, die von N unbekannten Koeffizienten abhängt.

$\Rightarrow$ Optimieren der Ansatzfunktion $\widetilde{\phi}$ durch Finden der besten Koeffizienten von $\widetilde{\phi}$ so, daß der Fehler $E = \phi - \widetilde{\phi}$ minimal wird, wobei ϕ die exakte Lösung ist.

Ausgehend von einem zusammenhängendem Gebiet $\Omega \subset \mathbb{R}^3$ mit einem umschließenden Rand Γ und einer Oberflächennormalen $\widehat{n}$, wie in Abb. 2.4, soll ein Randwertproblem gelöst werden.

In dem Gebiet Ω gelte eine Differentialgleichung:

$$\mathcal{L}\phi = f \tag{2.9}$$

Dabei ist $\mathcal{L}$ ein Differentialoperator, f der Anregungsterm und ϕ die gesuchte Lösungsfunktion. Auf dem Rand Γ gelten Dirichletsche, Neumannsche oder Cauchysche Randbedingungen. Diese Randbedingungen werden im Abschnitt 2.3.3 auf Seite 21 genauer erklärt.

2.3.1 Globale Ansatzfunktionen

Es wird hier davon ausgegangen, daß das Randwertproblem tatsächlich zu lösen ist.

Die folgenden Verfahren verwenden beide als Ansatzfunktion einen Satz von $N+1$ linear unabhängigen und dem Problem möglichst angepaßten Funktionen

$$\varphi_0(x,y,z),\ \varphi_1(x,y,z),\ \varphi_2(x,y,z),\ \ldots,\ \varphi_N(x,y,z) \tag{2.10}$$

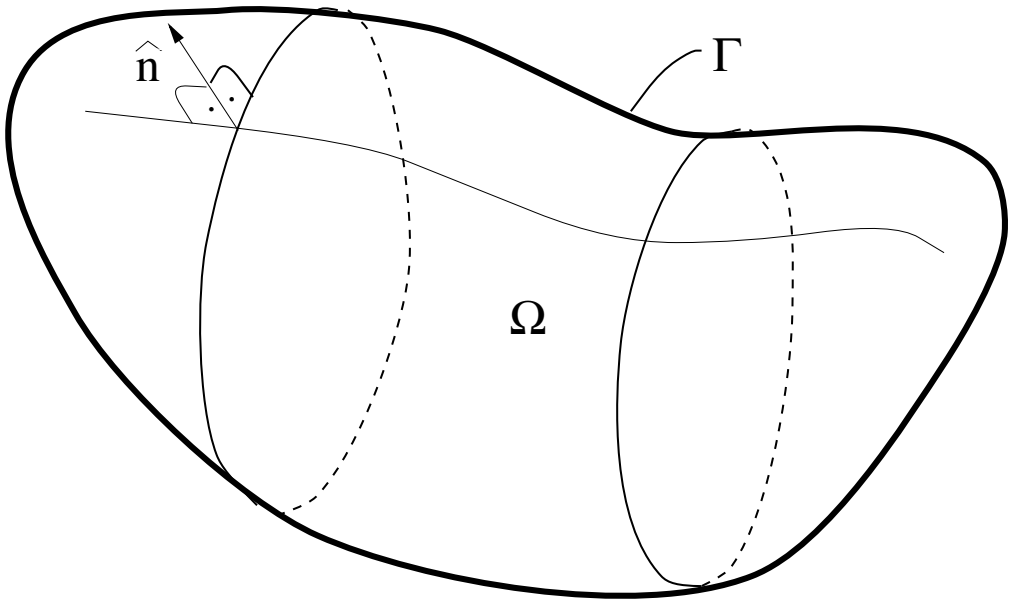


Abbildung 2.4: Einfach zusammenhängendes Berechnungsgebiet Ω mit Rand Γ und Oberflächennormale $\hat{n}$

Wobei die Funktionen im **gesamten** Gebiet Ω gelten, also **noch keine Gebietszerlegung erfolgt ist**. Aus diesem Satz von Funktion wird die gesuchte Funktion $\phi(x, y, z)$ durch folgende Linearkombination angesetzt:

$$\widetilde{\phi}(x, y, z) = \varphi_0(x, y, z) + \sum_{j=1}^{N} c_j \, \varphi_j(x, y, z) \tag{2.11}$$

Dabei muß die Funktion $\varphi_0(x, y, z)$ die inhomogenen Randbedingungen erfüllen. Die restlichen Funktionen sollen alle gegebenen homogenen Randbedingungen erfüllen, auf dem Rand mit Dirichletscher Randbedingung verschwinden und desweiteren auf Randstücken mit Cauchyscher Randbedingung diese erfüllen.

Mit dem Ansatz (2.11) werden bei beliebigen Koeffizienten c_j alle Randbedingungen erfüllt.

Der grundlegende Gedanke liegt nun darin, von dem Problem

⇒ *Eine Funktion im Raum aller möglichen Funktionen zu finden, die das Randwertproblem löst*

überzugehen zu dem Problem

⇒ *Bestimmen von N Gewichtungskoeffizienten c_j einer Linearkombination wie in (2.11) um das Randwertproblem zu lösen.*

Wie man die Gewichtungskoeffizienten c_j finden kann, wird in den beiden folgenden Abschnitten gezeigt.

2.3.1.1 Die Methode von Ritz

Die Methode von Ritz z.B. [23] [24, S.12] [22, S.41] [9, S.55], auch bekannt als Rayleigh-Ritz Verfahren, ist ein Variationsverfahren, wobei das Randwertproblem als Variationsproblem beschrieben wird. Das Funktional $F(\widetilde{\phi})$, das variiert werden soll, hat genau dort sein Minimum, wo die Funktion $\widetilde{\phi}$ die Differentialgleichung mit den gegebenen Randbedingungen am besten erfüllt.

Die Näherungslösung $\widetilde{\phi}(x,y,z)$ erhält man also durch Auffinden des Minimums des Funktionals.

Für das weitere Vorgehen wird zunächst das *innere Produkt* des Hilbertraums eingeführt, daß durch eckige Klammern angegeben wird:

$$\langle \phi, \psi \rangle = \int_\Omega \phi \psi^* d\Omega \;, \tag{2.12}$$

wobei der hochgestellte Stern das Konjugiertkomplexe bezeichnet. Setzt man voraus, das der Operator $\mathcal{L}$ linear, selbstadjungiert

$$\langle \mathcal{L}\phi, \psi \rangle = \langle \phi, \mathcal{L}\psi \rangle \tag{2.13}$$

und positiv definit ist

$$\langle \mathcal{L}\phi, \phi \rangle = \begin{cases} > 0 & \phi \neq 0 \\ = 0 & \phi = 0 \,, \end{cases} \tag{2.14}$$

dann kann die Lösung von (2.9) auf S.14 durch Minimieren des Funktionals

$$F(\widetilde{\phi}) = \frac{1}{2} \left\langle \mathcal{L}\widetilde{\phi}, \widetilde{\phi} \right\rangle - \frac{1}{2} \left\langle \widetilde{\phi}, f \right\rangle - \frac{1}{2} \left\langle f, \widetilde{\phi} \right\rangle \tag{2.15}$$

gefunden werden. Schreibt man die Ansatzfunktion (2.11) in vektorieller Form mit der Annahme, daß keine inhomogene Randbedingung (d.h.: $\varphi_0(x,y,z) = 0$) auftritt und der Operator $\mathcal{L}$ reellwertig[2] ist, dann wird

$$\widetilde{\phi} = \sum_{j=1}^{N} c_j \varphi_j = \{c\}^T \{\varphi\} = \{\varphi\}^T \{c\} \;. \tag{2.16}$$

Setzt man (2.16) in (2.15) ein, verwendet das inneren Produkt (2.12) und die Bedingung (2.13), daß der Operator $\mathcal{L}$ linear und selbstadjungiert ist, so erhält man für das Funktional

$$F(\widetilde{\phi}) = F(c_1, c_2, c_3, \ldots, c_N) \tag{2.17}$$

$$= \frac{1}{2} \{c\}^T \int_\Omega \{\varphi\} \mathcal{L} \{\varphi\}^T d\Omega \; \{c\} - \{c\}^T \int_\Omega \{\varphi\} \; f \, d\Omega \,. \tag{2.18}$$

Wie bei der allgemeinen Kurvendiskussion ist bei dem Minimum bzw. einem Maximum oder einem Sattelpunkt die 1. Ableitung Null. Man nennt eine Lösung an dieser Stelle einen

[2] Dies ist zunächst eine starke Einschränkung, später wird jedoch auch auf komplexwertige Operatoren eingegangen.

stationären Punkt, da sich das Funktional hier nicht wesentlich ändert. Bildet man nun von dem Funktional die partiellen Ableitungen[3] nach c_i und setzt diese Null, erhält man

$$\begin{aligned}\frac{\partial F(\widetilde{\phi})}{\partial c_i} &= \frac{1}{2}\int_\Omega \varphi_i \mathcal{L}\,\{\varphi\}^T\,d\Omega\;\{c\} + \frac{1}{2}\,\{c\}^T\int_\Omega \{\varphi\}\,\mathcal{L}\varphi_i\,d\Omega - \int_\Omega \varphi_i\,f\,d\Omega \\ &= \frac{1}{2}\sum_{j=1}^{N} c_j\int_\Omega \varphi_i\mathcal{L}\varphi_j + \varphi_j\mathcal{L}\varphi_i\,d\Omega - \int_\Omega \varphi_i\,f\,d\Omega \\ &\stackrel{!}{=} 0 \qquad \text{für } i = 1,2,3,\ldots,N\,. \end{aligned} \tag{2.19}$$

In Matrixschreibweise ist das ein lineares Gleichungssystem

$$[S]\,\{c\} = \{b\} \tag{2.20}$$

mit den Matrixelementen

$$S_{ij} = \frac{1}{2}\int_\Omega (\varphi_i\mathcal{L}\varphi_j + \varphi_j\mathcal{L}\varphi_i)\,d\Omega \tag{2.21}$$

und den Elementen in Vektor b

$$b_i = \int_\Omega \varphi_i\,f\,d\Omega\,. \tag{2.22}$$

Die gesuchten Koeffizienten c_i erhält man nun durch Lösen des Gleichungssystems (2.20). Damit ist die Näherungslösung $\widetilde{\phi}(x,y,z)$ bekannt.

An dieser Stelle sei noch erwähnt, das die Matrix $[S]$ symmetrisch ist, da der Operator $\mathcal{L}$ selbstadjungiert ist. Man kann dann (2.21) schreiben als:

$$S_{ij} = \int_\Omega \varphi_i\mathcal{L}\varphi_j\,d\Omega\,. \tag{2.23}$$

2.3.1.2 Die Methode von Galerkin

Die Methode von Galerkin, auch als *Methode der gewichteten Residuen* bekannt, ist universeller als die Methode von Ritz. Hier ist ein Extremalprinzip nicht notwendig, weshalb diese Methode auch auf einen bedeutend größeren Problemkreis anwendbar ist [22, S.45], [24, S.14], [9, S.55]. Als Ansatzfunktion $\widetilde{\phi}$ wird wieder (2.11) von Seite 15 verwendet. Setzt man den Ansatz (2.11) in die Differentialgleichung (2.9) von Seite 14 ein, so ist diese in den seltensten Fällen erfüllt. Man erhält vielmehr ein Residuum

$$r = \mathcal{L}\widetilde{\phi} - f \;\neq\; 0\,. \tag{2.24}$$

Die beste Näherung von $\widetilde{\phi}$ bekommt man, wenn r in allen Punkten im Gebiet Ω den kleinsten Wert annimmt. In der Methode der gewichteten Residuen verlangt man nun, daß das Integral des Residuums, gewichtet mit gewissen Gewichtsfunktionen w_i verschwindet

$$R_j = \int_\Omega w_j\,r\,d\Omega \stackrel{!}{=} 0\,. \tag{2.25}$$

[3] Man spricht auch von der ersten Variation des Funktionals $F(\widetilde{\phi})$ nach c_i und schreibt dann $\delta F(\widetilde{\phi})$.

Für den Ansatz $\widetilde{\phi}$ (2.11) benötig man N Bestimmungsgleichungen, da es N Koeffizienten c_j gibt. Mit N linear unabhängigen Gewichtsfunktionen w_j erhält man unter Verwendung von (2.25) genau diese N Bestimmungsgleichungen.

Bei der Methode von Galerkin werden als Gewichtsfunktionen w_j die Funktionen des Ansatzes φ_j verwendet

$$w_j = \varphi_j \qquad für \quad j = 1, 2, 3, \ldots, N\,. \tag{2.26}$$

Damit wird aus (2.25) :

$$R_j = \int_\Omega (\varphi_j \mathcal{L}\,\{\varphi\}^T\,\{c\} - \varphi_j\, f\, d\Omega \stackrel{!}{=} 0 \qquad für \quad j = 1, 2, 3, \ldots, N\,. \tag{2.27}$$

Als Ergebnis hat man wieder ein lineares Gleichungssystem wie (2.20), das zu lösen ist.

Falls der Operator $\mathcal{L}$ selbstadjungiert ist, ist die Matrix $[S]$ wieder symmetrisch[4]. Dann führt die Methode von Galerkin auf dieselben Bestimmungsgleichungen wie die Ritzsche Methode.

2.3.2 Lokale Ansatzfunktionen (FEM)

Bei den beiden bisher vorgestellten Methoden war es jedoch immer noch notwendig, die Funktionen

$$\varphi_0(x,y,z),\ \varphi_1(x,y,z),\ \varphi_2(x,y,z),\ \ldots,\varphi_N(x,y,z)$$

des Ansatzes (2.11) **problemangepaßt im gesamten Gebiet** Ω zu finden. Das ist bei beliebigen Berandungen gerade im 3-Dimensionalen ein sehr großes Problem, wenn nicht sogar unmöglich.

Deshalb unterteilt man das Gebiet Ω in viele *finite Elemente* und setzt hierfür **lokale** Funktionen $\tilde{\phi}^e$ an. Diese Funktionen sind nur innerhalb eines Elementes definiert und sonst identisch Null. Aufgrund der geringen Variation innerhalb eines Elementes kann man jetzt eine sehr einfache Funktion ansetzen. Dies ist die **Finite-Elemente-Methode.**

Als lokale Ansatzfunfunktionktion verwendet man meistens Polynome ersten[5] oder zweiten[6] Grades. Man kann auch Polynome höheren Grades verwenden, jedoch nimmt der Rechenaufwand dann schnell zu.

Bis auf besondere Ausnahmen sind die n lokalen Ansatzfunktionen in einem Element immer von der Struktur:

$$\tilde{\phi}^e = \sum_{j=1}^{n} \phi_j^e\, \varphi_j^e = \{\phi^e\}^T\,\{\varphi^e\} = \{\varphi^e\}^T\,\{\phi^e\}\ , \tag{2.28}$$

siehe Abbildung 2.5. Die Funktion $\tilde{\phi}^e$ setzt sich aus der Summe der Basisfunktionen φ_j^e multipliziert mit dem Stützwert ϕ_j^e zusammen. Dabei beschreibt jede Basisfunktion φ_j^e genau den Einfluß des j-ten Stützwertes ϕ_j^e auf die lokale Ansatzfunktion $\tilde{\phi}^e$. Als Stützstellen werden hauptsächlich Punkte (Knoten) verwendet. Bei vektoriellen Ansatzfunktion kann man jedoch auch Kanten verwenden.

[4] Die Symmetrie der Matrix ist ein wichtige Eigenschaft, da sie beim numerischen Lösen des Gleichungssystems große Vorteile mit sich bringt.

[5] Lineare Ansatzfunktion

[6] Quadratische Ansatz

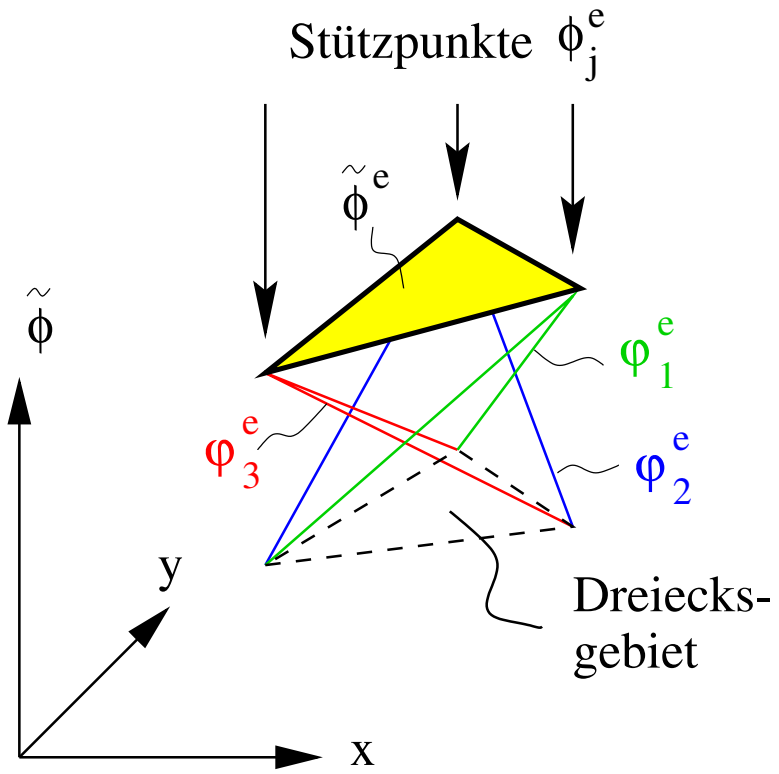


Abbildung 2.5: 2-Dimensionale knotenbasierte Ansatzfunktionen mit linearem Polynom-Grad (Ebene) bei einem Dreieck

2.3.2.1 Lokale Ritzsche Finite-Elemente-Methode

Das Funktional (2.15) von S.16 kann mit Hilfe der lokalen Ansätze (2.28) als

$$F(\tilde{\phi}) = \sum_{e=1}^{M} F^e(\tilde{\phi}^e) \tag{2.29}$$

geschrieben werden, dabei ist M die Anzahl der Elemente des gesamten Gebiets Ω. Für das lokale Funktional gilt dann

$$F^e(\tilde{\phi}^e) = \frac{1}{2}\int_{\Omega^e} \tilde{\phi}^e \mathcal{L}\tilde{\phi}^e \, d\Omega - \int_{\Omega^e} f\,\tilde{\phi}^e \, d\Omega\,. \tag{2.30}$$

Dies gilt unter der Annahme, daß das Problem reellwertig ist.

Setzt man nun die lokalen Ansatzfunktionen (2.28) in das lokale Funktional (2.30) ein, so erhält man

$$F^e(\tilde{\phi}^e) = \frac{1}{2}\{\phi^e\}^T \int_{\Omega^e} \{\varphi^e\}\,\mathcal{L}\,\{\varphi^e\}^T \, d\Omega \, \{\phi^e\} - \{\phi^e\}^T \int_{\Omega^e} f\,\{\varphi^e\}\, d\Omega\,. \tag{2.31}$$

In Matizenschreibweise ergibt das

$$F^e(\tilde{\phi}^e) = \frac{1}{2}\{\phi^e\}^T [K^e] \{\phi^e\} - \{\phi^e\}^T \{b^e\} \tag{2.32}$$

mit der $n \times n$ Matrix $[K^e]$ und dessen Matrixeinträgen

$$K^e_{ji} = \int_{\Omega^e} \varphi^e_j \mathcal{L} \varphi^e_i \, d\Omega \tag{2.33}$$

und dem $n \times 1$ Zeilenvektor $\{b^e\}$

$$b^e_j = \int_{\Omega^e} f \, \varphi^e_j \, d\Omega \, . \tag{2.34}$$

Setzt man nun das lokale Funktional (2.32) in das globale Funktional (2.29) ein, so erhält man für das gesamte Gebiet Ω

$$F(\tilde{\phi}) = \sum_{e=1}^{M} \left(\frac{1}{2} \{\phi^e\}^T [K^e] \{\phi^e\} - \{\phi^e\}^T \{b^e\} \right) . \tag{2.35}$$

Führt man die Summation über alle Elemente durch und vergibt eine globale Knotennumerierung, so lautet (2.35)

$$F = \frac{1}{2} \{\phi\}^T [K] \{\phi\} - \{\phi\}^T \{b\} \, . \tag{2.36}$$

Dabei werden Beiträge von mehreren Elementen zu einem Knoten aufaddiert, wodurch jeder Knoten auch wirklich nur einmal vorkommt.

So ist der Vektor $\{\phi\}$ ein $N \times 1$ Spaltenvektor, wobei jedes Element den Stützwert des globalen Knotens beschreibt. N ist die Anzahl der globalen Knoten. Die Matrix $[K]$ ist eine symmetrische $N \times N$ Matrix und durch (2.33) bestimmt. Der Vektor $\{b\}$ ist auch ein $N \times 1$ Spaltenvektor und beschreibt die Anregung des Systems. Seine Werte sind durch (2.34) gegeben.

Um das Funktional zu minimieren, wird wieder nach den unbekannten Koeffizienten (hier ϕ_j) partiell abgeleitet und gleich Null gesetzt. Daraus ergeben sich wieder N Bestimmungsgleichungen, die in Matrixschreibweise auf das lineare Gleichungssystem führen

$$[K] \{\phi\} = \{b\} \, . \tag{2.37}$$

Nach dem Einarbeiten der Randbedingungen in (2.37) ist mit numerischem Lösen von (2.37) die Näherungslösung bekannt.

2.3.2.2 Lokale Galerkinsche Finite-Elemente-Methode

Wie in Abschnitt 2.3.1.2 schon beschrieben, geht die Methode von Galerkin von dem Residuum der Differentialgleichung aus. Für das Residuum des e-ten Elements ergibt sich mit der Gewichtfunktion $w_j = \varphi^e_j$ und (2.27)

$$R^e_j = \int_{\Omega^e} \varphi^e_j (\mathcal{L} \tilde{\phi}^e - f) \, d\Omega \quad mit \, j = 1, 2, \ldots, n \, . \tag{2.38}$$

Setzt man nun die lokale Ansatzfunktion (2.28) in die lokalen Residuen (2.38) ein, so erhält man

$$R^e_j = \int_{\Omega^e} \varphi^e_j \mathcal{L} \{\varphi^e\}^T \, d\Omega \, \{\phi^e\} - \int_{\Omega^e} f \, \varphi^e_j \, d\Omega \stackrel{!}{=} 0 \quad mit \, j = 1, 2, \ldots, n \tag{2.39}$$

und in Matrixschreibweise

$$\{R^e\} = [K^e]\,\{\phi^e\} - \{b^e\} \stackrel{!}{=} \{0\} \tag{2.40}$$

mit $\{R^e\} = [R_1^e, R_2^e, R_3^e, \ldots, R_n^e]^T$. Die Matrixelemente K_{ji}^e und die Vektorelemente b_j^e werden entsprechend (2.33) bzw. (2.34) berechnet. Summiert man (2.40) wieder über alle Elemente des Gebiets Ω und geht auf globale Knotennumerierung über, so erhält man wieder ein Gleichungssystem der Form:

$$[K]\,\{\phi\} = \{b\}\ . \tag{2.41}$$

Nach dem Einarbeiten der Randbedingungen in (2.41) ist mit numerischem Lösen von (2.41) die Näherungslösung bekannt.

2.3.3 Randbedingungen in der FEM

Die unterschiedlichen Randbedingungen auf dem Rand Γ des Gebiets Ω müssen **vor** dem numerischen Lösen des linearen Gleichungssystems von (2.37) bzw. (2.41) in das Gleichungssystem eingearbeitet werden.

Dirichletsche Randbedingung: Sie wird auch Randbedingung erster Art genannt und im Folgenden auf dem Randgebiet S_3 angenommen. Hier handelt es sich um ein *Zwangsbedingung.*

$$\phi_j = Wert$$

Für eine Stützstelle ϕ_j soll ein ganz bestimmter Wert erzwungen werden. Man erreicht das, indem man in der Zeile j alle K_{ji} mit $i \neq j$ auf Null setzt, $K_{jj} = 1$ setzt und b_j setzt man auf den Wert, den die Stützstelle $\phi_j$erhalten soll. Im Gleichungssystem existiert dann für ϕ_j nur noch die Gleichung:

$$1 \cdot \phi_j = Wert$$

Wird das Gleichungssystem gelöst, hat ϕ_j den Wert. Durch diesen Eingriff ist die Matrix in ihrer Struktur aber verändert worden. Sie hat die symmetrische Struktur verloren. Man kann die Matrix jedoch wieder in eine symmetrische Form bringen. Dafür ist die Spalte j in K bis auf K_{jj} auch auf Null zu bringen [24, S.44].
Die Matrixeinträge in der Spalte j werden bei der Matrix-mal-Vektor-Operation gerade mit dem Wert ϕ_j multipliziert und zu b_i addiert. Da ϕ_j aber ja schon bekannt ist, kann man diese Multiplikation schon jetzt durchführen, das Produkt von b_i abziehen und den Matrixeintrag nullsetzen. Damit bleibt das Gleichungssystem erhalten und die Matrix K ist wieder symmetrisch, falls sie es vorher auch schon war. Da man diese Randbedingung in das Gleichungssystem explizit einarbeiten muß, wird sie als *essentielle Randbedingung* bezeichnet, wegen ihrer geometrischen Bedeutung aber auch als *geometrische Randbedingung.*

Neumannsche Randbedingung: Sie wird auch Randbedingung zweiter Art genannt [24, S.34]. Sie stellt einen Sonderfall (mit $\gamma = 0$ s. Gl. (2.44) auf S.22) der Cauchyschen Randbedingung dar

$$\alpha \frac{\partial \phi}{\partial \hat{n}} = q \tag{2.42}$$

mit der nach außen gerichteten Normalen $\hat{n}$ der Oberfläche Γ vom Gebiet Ω. Die Parameter α und q seien bekannt. Für $q = 0$ spricht man von der homogenen und bei $q \neq 0$ von der inhomogenen Neumannschen Randbedingung. Die homogene Neumannsche Randbedingung ist eine *natürliche* Randbedingung. Sie wird beim Minimieren des Funktionals automatisch miterfüllt. Für weiterführende Literatur sei hier [24, S.45] empfohlen. Hier sei nur angemerkt, daß dieser Anteil im Funktional durch partielle Integration entsteht und $\alpha \frac{\partial \phi_j}{\partial \hat{n}}\Big|_\Gamma$ zu b_j dazu addiert werden muß. Da

$$\alpha \frac{\partial \phi_j}{\partial \hat{n}}\bigg|_\Gamma = q \tag{2.43}$$

ist, braucht nur q_j dazuaddiert zu werden. In dem Sonderfall $q_j = 0$ wird b_j nicht verändert, wobei alles beim Alten bleibt.

Cauchysche Randbedingung: Sie wird auch Randbedingung dritter Art genannt [24, S.34],

$$\alpha \frac{\partial \phi}{\partial \hat{n}} + \gamma \phi = q \tag{2.44}$$

mit der nach außen gerichteten Normalen $\hat{n}$ der Oberfläche Γ vom Gebiet Ω. Die Parameter α, γ und q seien bekannt. Für $q = 0$ spricht man von der homogenen (später auf dem Randgebiet S_2 angenommen) und bei $q \neq 0$ von der inhomogenen (später auf dem Randgebiet S_1 angenommen) Cauchyschen Randbedingung. In dem Sonderfall $\gamma = 0$ stellt sie wieder die Neumannsche Randbedingung dar.
Schreibt man (2.44) in der Form

$$\alpha \frac{\partial \phi_j}{\partial \hat{n}}\bigg|_\Gamma = q_j - \gamma \phi_j \,, \tag{2.45}$$

so kann der Term $q_j - \gamma \phi_j$ in $[K]$ und $\{b\}$ absorbiert werden. Für weiterführende Literatur sei hier wieder [24, S.45-47] empfohlen.

Kapitel 3

Diskretisierung bzw. Netzgenerierung

Allgemein unterscheidet man zwischen *strukturierten* und *unstrukturierten* Netzen bzw. Diskretisierungen. Wobei mit strukturierten Netzen *regelmäßige* Unterteilungen des Gebiets gemeint sind und mit unstrukturierten Netzen eine *unregelmäßige* Unterteilung. Die FEM hat den Vorteil, daß sie mit unstrukturierten Netzen komplizierte Geometrien gut approximieren kann. Gleichzeitig entsteht dadurch aber auch die Notwendigkeit, ein unstrukturiertes Netz für das gerade benötigte Problem zu konstruieren [12].

So wie man jedes ebene Polygon in Dreiecke zerlegen kann, so ist es im 3-dimensionalen Raum auch möglich, jeden Polyeder in Tetraeder zerlegen. Dies ist jedoch keine einfache Aufgabe. Aus diesem Grund wurden alle in dieser Arbeit verwendeten Netze mit dem Netzgenerator GEOMPACK von Barry Joe [16] generiert. Da für den Netzgenerator ein vollkommen eigenes Format der Volumenbeschreibung verwendet wird, wurde eine Studienarbeit zum Thema Schnittstelle CAD-Netzgenerator bearbeitet [18]. Weil der Netzgenerator jedoch sehr schwierig zu bedienen ist und selten zu Netzen ausreichender Qualität für die FEM führte, wurde desweiteren eine Studienarbeit zum Thema Netzverbesserung bearbeitet [20].

3.1 Volumenbeschreibung

Mit dem Begriff Volumenbeschreibung verbindet sich die eindeutige Beschreibung eines Gebietes und nicht nur einer Oberfläche. Es ist also wichtig, daß man feststellen kann, ob ein Punkt *im* Gebiet liegt oder *außerhalb*.

Man kann grundsätzlich zwischen zwei Arten unterscheiden, wie man ein Volumen beschreiben kann. Entweder man verwendet analytische Funktionen, die das Gebiet begrenzen. Hier werden krummlinige Grenzflächen immer exakt beschrieben, solange eine Funktion zur Verfügung steht, die dieses krummlinige Verhalten besitzt. Als andere Möglichkeit kann man die Oberfläche diskretisieren. Dann ist man von der Wahl der Funktionen unabhängig und kann beliebig Oberflächen verwenden. Dementsprechend gibt es Netzgeneratoren, die analytische Funktionen oder eben Polyeder zur Volumenbeschreibung verwenden. GEOMPACK verwendet Polyeder.

Die heute wohl natürlichste Art ein Berechnungsgebiet zu beschreiben, ist die Eingabe in ein CAD-Programm. In den meisten CAD-Systemen wird heute die interne Darstellung mit analytischen Funktionen realisiert. Man spricht hier von einer *vollparametrischen* Darstellung. Erst beim Export werden die Daten diskretisiert.

Aus diesem Grund wurde eine Schnittstelle in IFEP entwickelt, um die Geometriedaten aus einem DXF-File von Autocad R12 zu importieren [18].

Dabei stellte sich heraus, daß in dem DXF-Format von R12 zum letzten Mal noch 3D-Volumeninformationen enthalten waren. In den nachfolgenden DXF-Standards bis zu R14 befinden sich nur noch 2D-Objekte, sodaß es in diesen Versionen nicht mehr möglich ist, Informationen über Polyeder aus dem Format zu gewinnen.

Im DXF-Format R12 sind alle 3D-Zeichenbefehle als einzelnes 3D-Objekt gespeichert. Um aus diesen vielen 3D-Objekten aber *ein* Gebiet zu machen, müssen diese wieder verbunden werden.

Bei jedem Teilpolyeder müssen die Flächennormalen nach außen zeigen. Liegt an der Fläche ein Nachbarpolyeder an, muß, da jede Fläche nur eimal vorkommt, die Flächennormale mit dem anderen Polyeder abgestimmt werden. Man sieht schnell, daß dies keine triviale Aufgabe ist. Aus diesem Grund sind im Rahmen dieser Arbeit neben anderen komplizierten dynamischen Programmierungen auch umfangreiche Visualisierungen (s. z.B. Abb. 3.1) dieser Daten in IFEP entstanden.

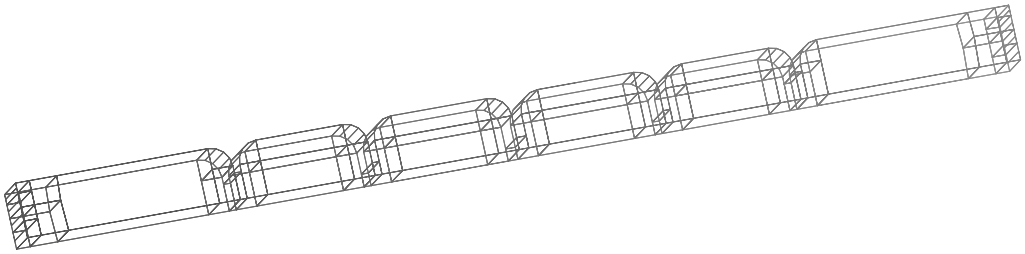


Abbildung 3.1: Volumenbeschreibung des Hohlleiter-Bandpasses von Abschnitt 8.3

Alle Routinen für diesen Aufgabenbereich sind in IFEP in Form eines neuen Projekts namens GEOP3D eingebunden. Dort können die einzelnen 3D-Objekte visualisiert und dann verbunden werden. Anschließend wird die gesamte Volumenbeschreibung in einem „*.g3”-File abgelegt und man kann mit dem Netzgenerator ein Tetraeder-Netz erzeugen.

3.2 Netzgenerierung

Für einen Überblick über die verschieden Methoden zur Netzgenerierung sei auf [12] verwiesen. Der Netzgenerator GEOMPACK benötigt für seinen Algorithmus konvexe Gebiete. Deshalb werden die vorgegebenen Polyeder zunächst in konvexe Gebiete zerlegt, um sie anschließend zu vernetzen. Dabei werden mit Hilfe eines Zufallsgenerators Knoten in dem konvexen Gebiet generiert, die dann nach verschiedenen Kriterien zu Tetraedern verbunden werden.

Zunächst wird erst einmal die Qualität eines Tetraedernetzes eingeführt und anschließend wird auf die Probleme eingegangen, die bei der Netzgenerierung auftauchen können. Wie ein bereits generiertes Netz nachträglich noch verbessert werden kann, wird im folgenden Abschnitt dann gezeigt.

3.2.1 Qualität eines Tetraedernetzes

Für die Finite-Elemente-Methode wäre es eigentlich am besten, wenn alle Tetraeder gleichlange Kanten hätten. Dies ist aber nicht möglich, deshalb besteht ein Tetraeder-Netz immer

aus mehr oder weniger guten Tetraedern. Um die einzelnen Tetraeder in ihrer Qualität unterscheiden zu können, existiert unter anderem das „Innenkugel-Umkugelkriterium" mit ϱ für den Radius der Innenkugel und R für den Radius der Umkugel:

$$Q = 3\frac{\varrho}{R} \tag{3.1}$$

Die Werte von Q variieren im Intervall $[0, 1]$. Ein Tetraeder mit gleichlangen Kanten hat ein $Q = 1$, während ein degeneriertes Tetraeder mit keinem Volumen den Wert $Q = 0$ besitzt. Tetraeder mit $\varrho/R < 0.01$ (d.h. $Q < 0.03$) werden als Sliver Elemente bezeichnet [13].

Neben diesem Kriterium gibt es noch andere, wie z.B. das Delaunay-, das Babuska- und das MMSA-Kriterium. Für einen Überblick sei hier [20] empfohlen.

Um nun die Qualität über das gesamte Netz zu bestimmen, existieren die beiden Formulierungen
-gemittelte Qualität (mean)

$$Q_{mean} = \frac{1}{N}\sum_{i=1}^{N} Q_i \tag{3.2}$$

-„parallel geschaltete" Qualität (joint)

$$Q_{joint} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{Q_i}}, \tag{3.3}$$

wobei Q_i die Qualität des Tetraeders i und N die Anzahl aller Tetraeder im Netz ist [14] [15]. Die gemittelte Qualität gibt eine gute Aussage über das gesamte Netz an. Die Joint Qualität bewertet schlechte Tetraeder dagegen höher. Bei beiden Kriterien stellt 0 den schlechteste Wert, d.h. die schlechteste Netzqualität, dar und der Wert 1 den besten Qualität.

3.2.2 Probleme der Netzgenerierung

Nachdem man das Problem der Erstellung einer Eingabedatei für GEOMPACK gelöst war, hatte der Netzgenerator zwei grundlegende Probleme.

3.2.2.1 Konvexe Gebietszerlegung

Das erste besteht in der Notwendigkeit, die Polyeder in konvexe Polyeder zu zerlegen. Wie in Abbildung 3.2 zu sehen ist, kann das zu einer schwierigen Aufgabe werden. Es war für GEOMPACK nicht möglich, dieses Polyeder in konvexe Gebiete zu zerlegen. Dazu muß man verstehen, wie der Algorithmus in GEOMPACK arbeitet. Es wird eine Kante genommen und versucht, eine Schnittebene mit der Kante als Achse zu finden, um das Gebiet in konvexe Gebiete zu zerschneiden. Daß diese Aufgabe bei einer Geometrie wie dem Kopf kaum zu lösen ist, leuchtet schnell ein. Für solche Geometrien eignen sich besser Advancing-Front-Netzgeneratoren. Sie bauen vom Randgebiet das Netz Schicht für Schicht auf, haben jedoch bei der Vernetzung am Schluß die Schwierigkeit, gute Tetraeder zu generieren.

Bei Geometrien, die mehr blockartig sind, oder bei Polyedern, die schon konvex sind, hatte der Netzgenerator keine Schwierigkeiten, ein Netz zu erzeugen. Wie man in Abbildung 3.1 sehen kann, sind dort alle Polyeder konvex, um dem Netzgenerator die Zerlegung abzunehmen. Dabei war an den Enden des Gebiets eine höhere Diskretisierung erwünscht. Aus diesem Grunde wurden diese Gebiete zusätzlich zerlegt, um dort mehr Tetraeder zu erzwingen.

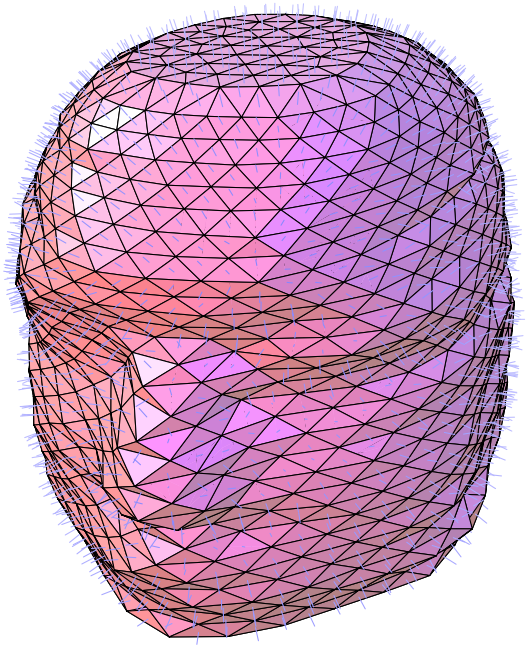


Abbildung 3.2: Polyeder eines Kopfes, der von GEOMPACK nicht in konvexe Gebiete zerlegt werden konnte.

3.2.2.2 Schlechte Netzqualität

Ein zweites Problem ist die Qualität der generierten Tetraeder. Sind z.B. in den Polyedern sehr kleine Raumwinkel vorhanden, kann der Netzgenerator dort niemals gute Tetraeder generieren. Aber selbst wenn der Netzgenerator gute Polyeder vernetzen soll, kann es zu schlechten Tetraedern kommen. Dies bewirkt eine schlechte Konvergenz des iterativen Lösers.

Bei Netzen geringer Größe (ca. 15000 Tetraeder) ist bei GEOMPACK meistens ein Netz guter Qualität generiert worden. Sollen jedoch mehr Tetraeder (ca. 50000 -150000 Tetraeder) generiert werden, entstehen viele schlechte Elemente. Die einzige Methode, doch noch gute Netze zu erhalten, war das Verändern der gewünschten Tetraederanzahl. So konnte es vorkommen, daß man 60 verschieden Netze generieren mußte, um ein Netz mit ausreichend guter Qualität zu erhalten. Aus diesem Grund werden im nächsten Abschnitt die verwendeten Methoden zur Netzverbesserung beschrieben.

3.3 Netzverbesserung

Für die Netzverbesserung stehen eine Vielzahl von Verfahren zur Auswahl. So kann man zum Beispiel durch Einfügen von einem Knoten die längste Kante eines Tetraeders halbieren. Diese Methode hat jedoch auch den Nebeneffekt, daß das Netz dadurch verfeinert wird. Durch einen Punkt kann man z.B. leicht 7 neue Kanten im Netz erzeugen. Wenn das oft der Fall ist, wird das Netz schnell erheblich größer. Aus diesem Grund wurde von dieser Methode Abstand genommen und Methoden angewendet, die die Größe des Netzes nur unwesentlich verändern.

3.3.1 Lokale Transformation

Eine andere Methode ist die der *lokalen Transformationen*. Dabei werden keine Knoten verändert, sondern nur die Kanten, die die Knoten verbinden, werden vertauscht, eingefügt oder entfernt. Dadurch entstehen entweder neue Tetraeder, oder sie werden vertauscht oder gar eliminiert. Der Begriff der *lokalen Transformationen* erklärt sich daher, daß die Transformationen nur Auswirkungen auf einen kleinen lokalen Bereich haben. Außerhalb der Hülle, die alle an der Transformation beteiligten Elemente enthält, ändert sich nichts. Aus diesem Grund sind diese Transformationen überschaubar. Man kann die Berechnung der Qualitätsverbesserung lokal vornehmen und die Entscheidung, ob eine Transformation sinnvoll ist, mit geringem Aufwand treffen.

Es gibt prinzipiell beliebig viele Transformationen. Üblicherweise hält man die Anzahl der möglichen Transformationen jedoch gering, da man alle Möglichkeiten prüfen muß, bevor man die Entscheidung zur Durchführung treffen kann.

In IFEP werden nur die folgenden Transformationen eingesetzt:

⇒ zwei Tetraeder werden zu zwei Tetraedern

⇒ zwei Tetraeder werden zu drei Tetraedern

⇒ drei Tetraeder werden zu zwei Tetraedern

Obwohl so nur wenige Transformationen verwendet werden, müssen pro Tetraeder immer noch 22 Möglichkeiten geprüft werden. Anschließend wird die Transformation durchgeführt, bei der sich die Joint-Qualität am stärksten verbessert. Tritt eine Verschlechterung ein, wird keine Transformation durchgeführt. In IFEP wird diese Methode mit *transform* bezeichnet. Eine genaue Beschreibung ist in [20] nachzulesen.

3.3.2 Knoten-Relaxation

Diese Methode beruht auf der Verschiebung von Knoten in der Art, daß sich die Qualität der umliegenden Tetraeder verbessert. In IFEP wird von den umliegenden Knoten der Schwerpunkt gebildet

$$\vec{r}_j = \frac{1}{N} \sum_{i \neq j}^{N_j} \vec{r}_i \tag{3.4}$$

und als möglicher Verschiebungspunkt vorgeschlagen. Man nennt diese Art der Verschiebung *Laplace*-Smoothing [14]. Anschließend muß geprüft werden, ob dies ein erlaubter Punkt ist.

Man kann dies einfach durch eine Volumenberechnung der beteiligten Tetraeder prüfen. Wenn die Volumen aller Tetraeder weiterhin positiv sind, wird durch die Verschiebung kein Netzfehler verursacht.

Weiter muß geprüft werden, ob sich das Netz durch die Verschiebung verbessert hat. In IFEP werden hier zwei Kriterien verwendet:

⇒ Die Joint-Qualität (siehe (3.3)) der beteiligten Tetraeder darf sich durch die Verschiebung nicht verschlechtern (in IFEP *relax1*).

⇒ Das schlechteste Tetraeder darf durch die Verschiebung nicht schlechter werden (in IFEP *relax2*).

Diese Verschiebungen dürfen natürlich nicht überall im Netz durchgeführt werden. Es gibt Knoten, die auf keinen Fall nach dieser Regel verschoben werden dürfen. Dazu gehören:

⇒ Knoten, die auf einer Trennfläche zwischen zwei Medien liegen (interface nodes).

⇒ Knoten, die auf einer Randfläche liegen (surface nodes).

Man könnte diese Knoten zwar in einer Randfläche oder Trennfläche verschieben. Das ist jedoch sehr aufwendig, da diese Flächen nicht in einer parametrischen Form vorliegen. Aus diesem Grund wird auf die Relaxation dieser Punkte vollständig verzichtet.

3.3.3 Anwendung der Netzverbesserung

Nach [29, S. 156] werden Transformation und Relaxation kombiniert eingesetzt. Man führt sie abwechselnd durch, bis sich keine merkbare Verbesserung mehr einstellt. Dabei ist es nicht möglich, vorauszusagen, welche Reihenfolge zu dem besten Ergebnis führt.

Als Beispiel ist hier auf Seite 29 in Tabelle 3.1 die Netzverbesserung beim TEAM Problem 19 (siehe Abb. 8.1 auf Seite 98) angegeben. Dabei beschreibt *smooth1* einen Durchgang der lokalen Transformation *transform* und anschließendem Relaxieren mit *relax1*. *smooth2* besteht aus einem Durchgang *transform* und anschließendem Relaxieren mit *relax2*. Dabei sieht man in der Spalte GEOMPACK die Qualität, mit der das Netz vom Netzgenerator erzeugt wurde. Von links nach recht sieht man dann, wie sich die Qualität durch die verschiedenen Aufrufe Schritt für Schritt verändert.

Man sieht am Anfang die stärksten Verbesserungen im Netz, besonders bei den schlechten Tetraedern. Beim zweiten Aufruf von *relax2* und *smooth2* gibt es nur noch Verbesserungen bei sehr guten Elementen. Bis hier wurde die Relaxation nur durchgeführt, wenn das schlechteste Element auch verbessert wurde. Das erklärt, warum bei keinem Schritt eine Verschlechterung eingetreten ist. Bei *relax1* und *smooth1* wird die Relaxation durchgeführt, wenn sich die Joint-Qualität der beteiligten Tetraeder gemeinsam verbessert. Man sieht hier, daß sich zwar die Joint-Qualität und die Mittlere-Qualität des gesamten Netzes noch verbessert, dafür entstehen aber wieder mehr schlechtere Tetraeder.

	GEOMPACK	relax2	smooth2	relax2	smooth2	relax1	smooth1
Mittlere Qualität	0,8187	0,8357	0,8402	0,8415	0,8418	0,8547	0,8561
Joint-Qualität	0,8037	0,8253	0,8304	0,8317	0,8321	0,8442	0,8458
schlechteste Tetraeder	0,0638	0,1913	0,3459	0,3526	0,3526	0,3590	0,3589
beste Tetraeder	0,9981	0,9979	0,9980	0,9980	0,9980	0,9977	0,9980
Qualitätshäufigkeit im Netz							
0,0-0,1	3	0	0	0	0	0	0
0,1-0,2	5	1	0	0	0	0	0
0,2-0,3	8	1	0	0	0	0	0
0,3-0,4	36	7	4	4	4	7	9
0,4-0,5	125	50	36	36	35	62	54
0,5-0,6	596	317	259	252	254	275	273
0,6-0,7	2168	1545	1416	1389	1379	1275	1240
0,7-0,8	6370	5999	5827	5758	5758	4322	4196
0,8-0,9	12058	11809	11653	11528	11487	10924	10870
0,9-1,0	4909	6549	7071	7299	7349	9401	9624

Tabelle 3.1: Netzverbesserung an einem Netz für das Beispiel TEAM Problem 19 von S. 98

Kapitel 4

Die Kantenelemente

Die finiten Elemente hatten ursprünglich als unbekannte Parameter die Werte eines skalaren Feldes immer auf den Knoten des Elements. Diese Elemente werden als *Knotenelemente* bezeichnet. Als man begann, vektorielle Probleme zu berechnen, war es naheliegend, das vektorielle Feld in seine skalaren Bestandteile zu zerlegen und wieder skalare Elemente zu verwenden. Jedoch befanden sich nun drei Unbekannte auf jedem Knoten. Sie wurden und werden auch heute noch oft verwendet.

Vektoren können jedoch nicht immer als Tripel von Zahlen aufgefaßt werden. Sie haben mathematische und physikalische Eigenschaften, die sich nicht immer leicht in Komponenten eines Koordinatensystems ausdrücken lassen. Durch das Teilen des Vektors in die Komponenten des Koordinatensystems werden diese Eigenschaften verletzt. Bei elektromagnetischen Feldern zum Beispiel gibt es oft eine Randbedingung, die nur die tangentiale Komponente an den Elementgrenzen berücksichtigt. Bei Knotenelementen muß diese einfache physikalische Randbedingung in lineare Beziehungen zwischen den kartesischen Koordinaten transformiert werden. Zusätzlich muß an Knoten, an denen sich die tangentiale Richtung ändert, ein Mittelwert gebildet werden. Das ist nicht unmöglich, jedoch schwierig und unpraktisch [33].

Eine Alternative dazu stellen die **Kantenelemente** (engl. *edge elements*) dar. Hier werden als unbekannte Parameter vektorielle Größen verwendet, die den Kanten des Elements zugeordnet sind. Zuerst eingeführt wurde der Begriff von Hassler Whitney[1] im Jahr 1957 [30], noch bevor der Begriff „Methode der Finiten Elemente" im Jahr 1960 von Clough geprägt wurde [29, S. 39]. Die Bedeutung der Kantenelemente für die Berechnung elektromagnetischer Felder wurde lange nicht erkannt. Erst 1980 mit der Arbeit von J. C. Nédélec [31] wurden diese Elemente für die elektromagnetische Feldberechnung diskutiert. Die hier verwendeten Elemente werden als „Whitney-Elemente der Ordnung 1" bezeichnet [32, S. 76]. Sie besitzen ein *konstantes tangentiales/lineares normales* Feld entlang der Elementkanten und ein lineares Feld auf den Elementflächen und im Element. Man bezeichnet sie deshalb auch als „Komplett mit der Ordnung 0.5" [37] und wegen ihrer gemischten Ordnung auch als „Mixed Elements" . Nédélec zeigte, daß es ein nicht notwendiger Vorteil ist, wenn man polynom-komplette finite tangentiale Vektorelemente in der FEM verwendet [31].

Für eine tiefergehende Untersuchung dieser Elemente werden sie im folgenden hergeleitet.

[1] Autor vieler grundlegender Arbeiten in der Differentialgeometrie

4.1 Das lokale Koordinatensystem

Da die Ansatzfunktionen im Tetraeder in baryzentrischen Koordinaten beschrieben werden, ist zum Verständnis eine genaue Kenntnis dieser Koordinaten Vorraussetzung. Der Anschauung halber wird dieses Koordinatensystem erst für Dreiecke, und dann davon abgeleitet, für Tetraeder eingeführt.

4.1.1 Baryzentrische Koordinaten im Dreieck

Die baryzentrischen Koordinaten werden auch lokale[2] Koordinaten genannt und gehören zu den Verhältniskoordinaten [7, 2.6.5][9, S. 572].

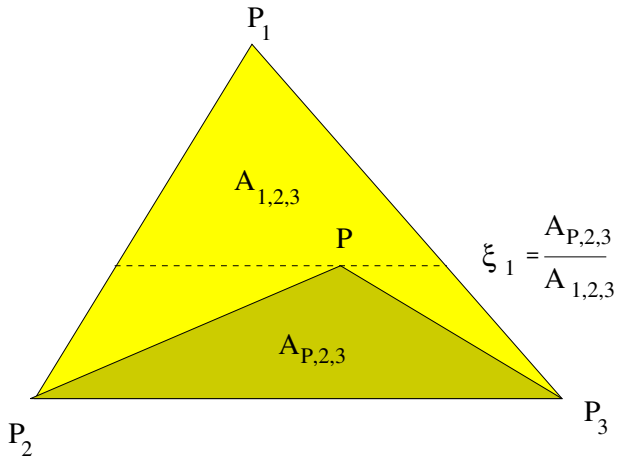


Abbildung 4.1: Baryzentrische Koordinate ξ_1 im Dreieck

Dabei beschreibt die Koordinate[3] ξ_1 die Punkte P auf der gestrichelten Linie, die im Dreieck durch das Verhältnis der Teilfläche $A_{p,2,3}$ zur Gesamtfläche $A_{1,2,3}$ des Dreiecks beschrieben sind (siehe Abb. 4.1). Da die Fläche eines Dreiecks nur von der Grundlinie und der Höhe abhängt, ist ξ_1 auf Parallelen zur Grundlinie konstant.

So definieren sich die baryzentrischen Koordinaten im Dreieck wie folgt

$$\xi_1 = \frac{A_{p,2,3}}{A_{1,2,3}}, \tag{4.1}$$

$$\xi_2 = \frac{A_{1,p,3}}{A_{1,2,3}}, \tag{4.2}$$

$$\xi_3 = \frac{A_{1,2,p}}{A_{1,2,3}}. \tag{4.3}$$

Da für ein zweidimensionales Element nun drei Koordinaten existieren, müssen die Koordina-

[2] Bezüglich des Elementes

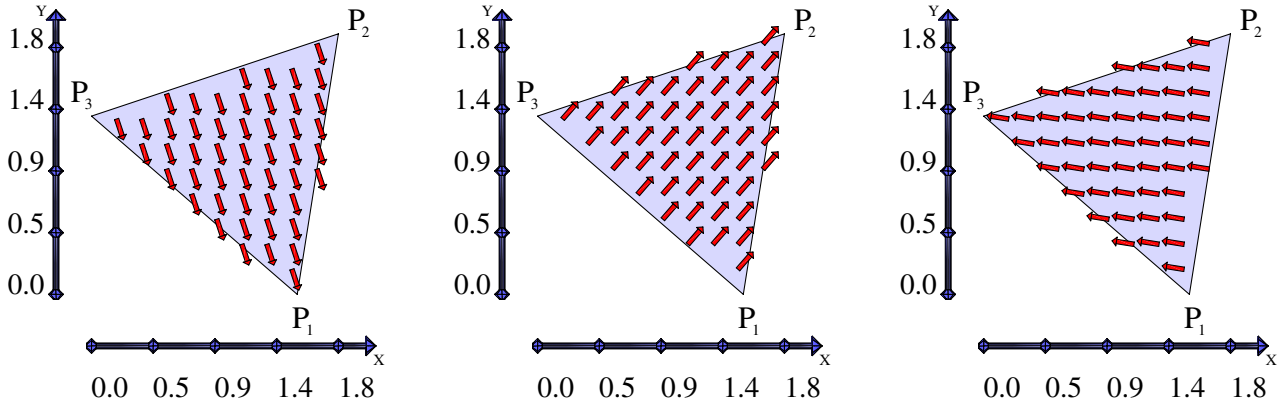
[3] In dieser Arbeit werden die baryzentrischen Koordinaten mit ξ_i bezeichnet. Es gibt in der Literatur aber auch die Bezeichnung mit L_i oder ζ_i.

ten linear abhängig sein. Deshalb[4] gilt

$$\xi_1 + \xi_2 + \xi_3 = 1\,. \tag{4.4}$$

4.1.1.1 Der Gradient einer Koordinate

Die Koordinaten ξ_1, ξ_2 und ξ_3 sind lineare Funktionen, deshalb ist jeder Gradient $\vec{\nabla}\xi_i$ im Dreieck ein Vektor mit konstantem Betrag und fester Richtung (siehe Abb. 4.2).



(a) Gradient von ξ_1 (b) Gradient von ξ_2 (c) Gradient von ξ_3

Abbildung 4.2: Gradienten $\vec{\nabla}\xi_i$ im Dreieck

Sie stehen senkrecht zur Grundlinie der Teilfläche und zeigen in das Element. Damit ist die Richtung fest gegeben. Die Länge des Gradienten in kartesischen Koordinaten ist für die Visualisierung und die Stetigkeit der berechneten Felder sehr wichtig. Deshalb wird an dieser Stelle auch die Herleitung des Betrages von $\vec{\nabla}\xi_i$ durchgeführt.

Für das Wegintegral von $\xi_1 = 0$ nach $\xi_1 = 1$ muß in baryzentrischen Koordinaten gelten

$$\int_{\xi_i=0}^{1} \vec{\nabla}\xi_i \cdot \overrightarrow{ds} = 1\,. \tag{4.5}$$

In kartesischen Koordinaten muß dann analog gelten

$$\int_{P_{\xi_i=0}}^{P_i} \vec{\nabla}\xi_i \cdot \overrightarrow{ds} \;=\; 1\,. \tag{4.6}$$

Sind $\vec{\nabla}\xi_i$ und $\overrightarrow{ds}$ entlang des Weges parallel und ist h_i = *Abstand von P_i zur Grundlinie* (Höhe), dann gilt

[4] Man kann das leicht mit (4.1)(4.2)(4.3) überprüfen: *Die Summe der drei Teildreiecke ergibt immer die Gesamtfläche des Dreiecks.*

$$\int_0^{h_i} \left|\vec{\nabla}\xi_i\right| ds = 1 \tag{4.7}$$

$$\left|\vec{\nabla}\xi_i\right| h_i = 1\,. \tag{4.8}$$

Damit ist der Betrag des Gradienten in kartesischen Koordinaten wie folgt definiert:

$$\left|\vec{\nabla}\xi_i\right| = \frac{1}{h_i}\,. \tag{4.9}$$

4.1.2 Baryzentrische Koordinaten im Tetraeder

Wie im Dreieck sind im Tetraeder die baryzentrischen Koordinaten über ein Verhältnis definiert. Im Unterschied zum Dreieck wird beim Tetraeder statt der Fläche das Volumen $V_{p,2,3,4}$ des Teiltetraeders betrachtet.

So definieren sich die baryzentrischen Koordinaten im Tetraeder wie folgt

$$\xi_1 = \frac{V_{P,2,3,4}}{V_{1,2,3,4}}\,, \tag{4.10}$$

$$\xi_2 = \frac{V_{1,P,3,4}}{V_{1,2,3,4}}\,, \tag{4.11}$$

$$\xi_3 = \frac{V_{1,2,P,4}}{V_{1,2,3,4}}\,, \tag{4.12}$$

$$\xi_4 = \frac{V_{1,2,3,P}}{V_{1,2,3,4}}\,. \tag{4.13}$$

Das Volumen eines Tetraeders ist nur von der Grundfläche und der Höhe abhängig. So sind auf den parallelen Ebenen zur Grundfläche die Koordinaten wieder konstant.

Für das dreidimensionale Element existieren vier Koordinaten. Sie sind wie beim Dreieck linear abhängig. Deshalb gilt[5] beim Tetraeder

$$\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1\,. \tag{4.14}$$

4.1.2.1 Der Gradient einer Koordinate

Die Koordinaten ξ_1, ξ_2, ξ_3 und ξ_4 sind lineare Funktionen, deshalb ist jeder Gradient $\vec{\nabla}\xi_i$ im Tetraeder ein Vektor mit konstantem Betrag und fester Richtung. Er steht senkrecht zur Grundfläche des Teiltetraeders und zeigt in das Element. Damit ist die Richtung fest gegeben. Die Betrag des Gradienten in kartesischen Koordinaten ist für die Visualisierung und Stetigkeit

[5] Man kann das leicht mit (4.10)(4.11)(4.12)(4.13) überprüfen: *Die Summe der vier Teiltetraeder ergibt wieder das Gesamtvolumen des Tetraeders.*

der berechneten Felder sehr wichtig. Deshalb wird an dieser Stelle auch die Herleitung des Betrages von $\vec{\nabla}\xi_i$ durchgeführt.

Für das Wegintegral von $\xi_1 = 0$ nach $\xi_1 = 1$ muß wieder gelten

$$\int_{\xi_i=0}^{1} \vec{\nabla}\xi_i \cdot \overrightarrow{ds} = 1\,. \tag{4.15}$$

In kartesischen Koordinaten muß dann analog gelten

$$\int_{P_{\xi_i=0}}^{P_i} \vec{\nabla}\xi_i \cdot \overrightarrow{ds} \;=\; 1\,. \tag{4.16}$$

Sind $\vec{\nabla}\xi_i$ und $\overrightarrow{ds}$ entlang des Weges parallel und ist $h_i = \textit{der Abstand von } P_i \textit{ zur Grundfläche}$ (im Gegensatz zum Dreieck mit der Grundlinie), dann gilt

$$\int_0^{h_i} \left|\vec{\nabla}\xi_i\right| \overrightarrow{ds} \;=\; 1 \tag{4.17}$$

$$\left|\vec{\nabla}\xi_i\right| h_i \;=\; 1\,. \tag{4.18}$$

Damit ist der Betrag der Gradienten in kartesischen Koordinaten für Tetraeder wie folgt definiert

$$\left|\vec{\nabla}\xi_i\right| = \frac{1}{h_i}\,. \tag{4.19}$$

4.2 Die kantenbasierten Ansatzfunktionen im Element

Im Gegensatz zu knotenbasierten Formfunktionen [9] sind bei kantenbasierten Formfunktionen die gesuchten Koeffizienten nicht den Knoten (Punkten) P_1, P_2, P_3, ..., sondern den Kanten zugeordnet. Die knotenbasierten Formfunktionen approximieren immer eine skalare Funktion in dem Element. Wenn man damit vektorielle Funktionen approximieren will, benötigt man für jede Komponente des Vektorfeldes eine eigene Formfunktion (z.B. im $\mathbb{R}^2 \Rightarrow 2$ und im $\mathbb{R}^3 \Rightarrow 3$). Kantenbasierte Formfunktionen approximieren immer vektorielle Größen. Die gesuchten Koeffizienten sind zwar wieder skalare Größen, jedoch beschreiben sie durch die Gewichtung der vektoriellen Formfunktionen wieder eine vektorielle Größe in dem Element.

4.2.1 Kantenbasierte Ansatzfunktion im Dreieck

Bei den hier verwendeten Formfunktionen beschreibt ein Koeffizient den Betrag der tangentialen Vektorgröße auf der Kante. Damit die Richtung der tangentialen Vektorgröße eindeutig zugeordnet ist, wird eine feste Orientierung für jede Kante festgelegt (siehe Abb. 4.3, S. 36).

Durch das Vorzeichen des Koeffizienten ist somit auch die Richtung auf jeder Kante eindeutig zugeordnet:

$$Kante\,1 = \overrightarrow{P_1P_2}\,, \tag{4.20}$$
$$Kante\,2 = \overrightarrow{P_2P_3}\,, \tag{4.21}$$
$$Kante\,3 = \overrightarrow{P_3P_1}\,. \tag{4.22}$$

Für die Approximation im Element werden nun drei *tangential unabhängige* Vektorbasisfunk-

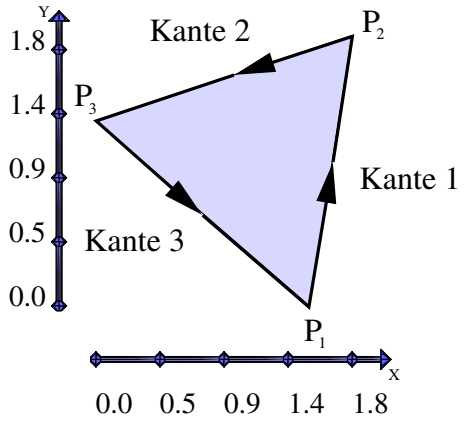


Abbildung 4.3: Kantenzuordnung im Dreieck

tionen benötigt. Das bedeutet, daß jede Funktion *nur zur eigenen Kante* einen *tangentialen* Beitrag leisten darf. Dafür gibt es zwei Möglichkeiten:

1. Die Vektorbasisfunktion *verschwindet* auf den anderen Kanten *ganz*.
2. Die Vektorbasisfunktion liefert *nur* Beiträge zur *normalen Komponente* auf den anderen Kanten.

Es liegt nahe, für die 2. Möglichkeit die Gradienten $\vec{\nabla}\xi_i$ der baryzentrischen Koordinaten zu verwenden, da sie immer senkrecht zur Kante gegenüber P_i stehen (siehe Abb. 4.2, S.33). Auf der anderen Kante muß der Gradient jedoch verschwinden, was durch die Multiplikation mit einem ξ_i erreicht werden kann, das auf der Kante gerade verschwindet.

4.2.1.1 Der Ansatz der Vektorbasisfunktion

Betrachtet man ein beliebiges Dreieck und bezeichnet die Knoten und Kanten gemäß Abb. 4.3, so kann man für die Kante 1 eine Vektorfunktion

$$\vec{W}_{12} = \xi_1\,\vec{\nabla}\xi_2 - \xi_2\,\vec{\nabla}\xi_1 \tag{4.23}$$

ansetzen. Dieser Ansatz wurde zuerst von Whitney [30] eingeführt. Er beschreibt ein konstantes tangentiales/lineares normales Feld entlang einer Elementkante und ein lineares Feld im Inneren.

4.2.1.2 Normierung der Vektorbasisfunktion

Wie in (4.9) auf S. 34 schon beschrieben, ist der Gradient $\vec{\nabla}\xi_i$ aber stark von der Form des Dreiecks abhängig. Da die Kante und damit auch der Koeffizient der Kante später auch von Nachbarelementen verwendet werden soll, muß die *Tangentialkomponente* unabhängig vom Rest des Elements sein. Anders ausgedrückt, die tangentiale Komponente muß zwischen zwei Elementen stetig sein. Aus diesem Grund muß die Vektorbasisfunktion normiert werden.

Sei $\overrightarrow{e_1}$ der Einheitsvektor entlang der Kante 1 von Knoten 1 zu Knoten 2 und l_1 die Länge dieser Kante. Die baryzentrische Koordinate ξ_1 ist eine lineare Funktion von $\xi_1 = 1$ bei Knoten P_1 bis zu $\xi_1 = 0$ bei Knoten P_2 und ξ_2 ist eine lineare Funktion von $\xi_2 = 0$ bei Knoten P_1 bis zu $\xi_2 = 1$ bei Knoten P_2 (s. Abb. 4.4, S.37).

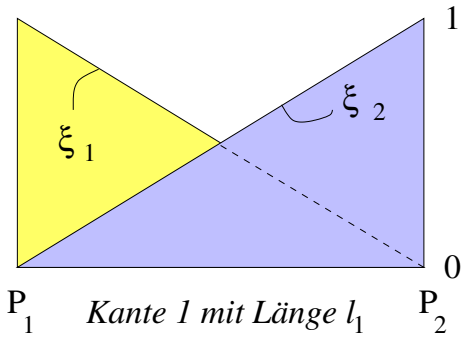


Abbildung 4.4: Die Funktionen ξ_1 und ξ_2 entlang der Kante von P_1 nach P_2

Die tangentiale Komponente von $\vec{W}_{12}$ ist dann im Knoten P_1

$$\vec{e}_1 \cdot \vec{W}_{12}\Big|_{P_1} = \vec{e}_1 \cdot \vec{\nabla}\xi_2 \tag{4.24}$$

$$= \frac{1}{l_1} \tag{4.25}$$

und im Knoten P_2

$$\vec{e}_1 \cdot \vec{W}_{12}\Big|_{P_2} = -\vec{e}_1 \cdot \vec{\nabla}\xi_1 \tag{4.26}$$

$$= \frac{1}{l_1}. \tag{4.27}$$

Auf der Kante 1 gilt dann (s. Abb. 4.4):

$$\vec{e}_1 \cdot \vec{W}_{12} = \frac{\xi_1 + \xi_2}{l_1} \tag{4.28}$$

$$= \frac{1}{l_1}. \tag{4.29}$$

Normiert man $\vec{W}_{12}$ mit l_1, ergibt sich für die tangentiale Komponente auf der Kante immer ein Wert von 1.

So ergibt sich für jede Kante i eine Vektorfunktion N_i (s. Abb. 4.5, Abb. 4.6 S.38 und Abb. 4.7 S.39)

$$\begin{aligned} \vec{N}_1 &= \vec{W}_{12}\, l_1 = (\xi_1\, \vec{\nabla}\xi_2 - \xi_2\, \vec{\nabla}\xi_1)\, l_1\,, \\ \vec{N}_2 &= \vec{W}_{23}\, l_2 = (\xi_2\, \vec{\nabla}\xi_3 - \xi_3\, \vec{\nabla}\xi_2)\, l_2\,, \\ \vec{N}_3 &= \vec{W}_{31}\, l_3 = (\xi_3\, \vec{\nabla}\xi_1 - \xi_1\, \vec{\nabla}\xi_3)\, l_3\,. \end{aligned} \tag{4.30}$$

Jede der Vektorbasisfunktionen $\vec{N}_i$ stellt eine Rotation *um den gegenüberliegenden Punkt* der Kante i dar.

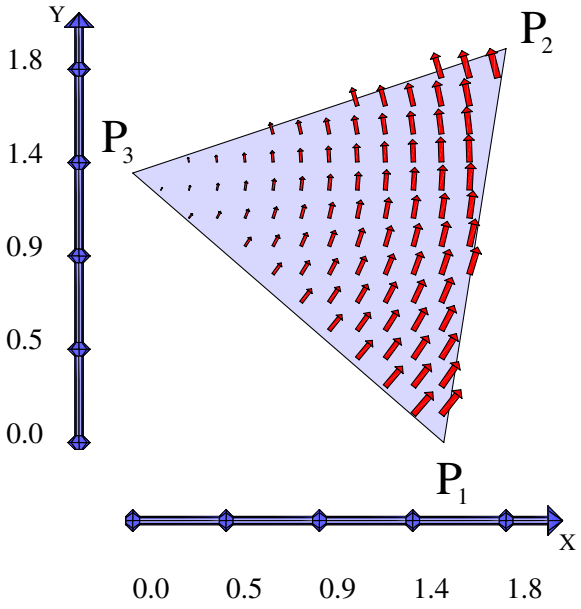


Abbildung 4.5: Vektorfunktion $\vec{N}_1$

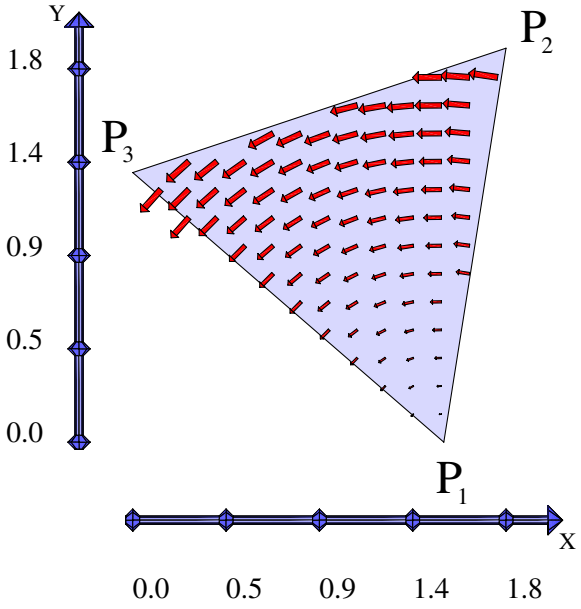


Abbildung 4.6: Vektorfunktion $\vec{N}_2$

Damit kann das Vektorfeld $\vec{E}^e$ im Element e ausgedrückt werden mit

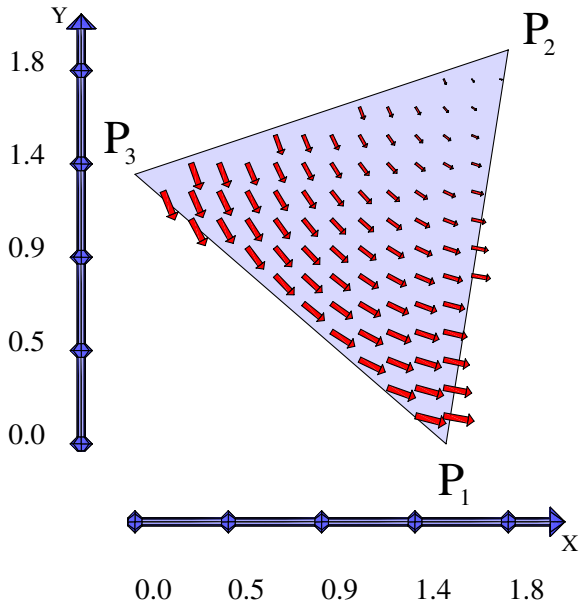


Abbildung 4.7: Vektorfunktion $\vec{N}_3$

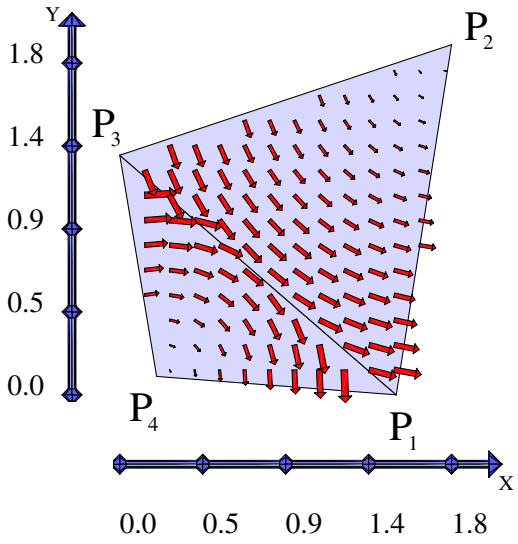


Abbildung 4.8: Stetigkeit der Tangentialen beim Elementübergang

$$\vec{E}^e = \sum_{i=1}^{3} \vec{N}_i^e E_i^e \,. \tag{4.31}$$

Dabei beschreibt E_i^e die tangentiale Feldkomponente entlang der i-ten Kante. Als Beispiel sieht man in Abbildung 4.9 auf S.40 ein Vektorfeld $\vec{E}^e$, wobei alle E_i^e's den Betrag 1 besitzen und somit eine Rotation beschreiben.

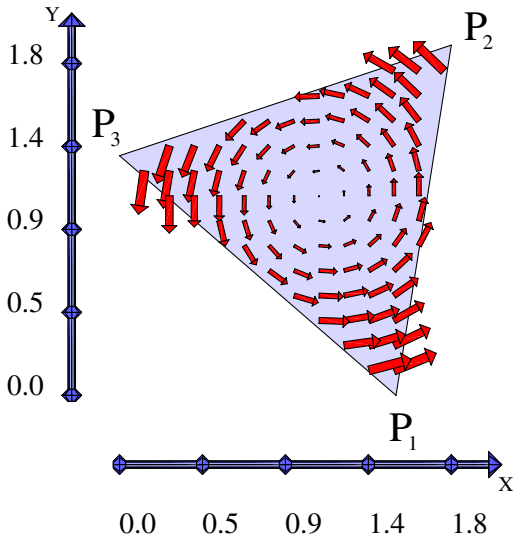


Abbildung 4.9: Beispiel für ein Vektorfeld $\vec{E}^e$ im Dreieck

4.2.2 Kantenbasierte Ansatzfunktion im Tetraeder

Wie auch beim Dreieck beschreibt beim Tetraeder ein Koeffizient den Betrag der tangentialen Vektorgrösse auf einer Kante. Damit die Richtung der tangentialen Vektorgrösse eindeutig zugeordnet ist, wird wieder eine feste Orientierung für jede Kante festgelegt (siehe Abb. 4.10, S. 41) .

$$Kante\,1 = \overrightarrow{P_1P_2} \tag{4.32}$$
$$Kante\,2 = \overrightarrow{P_1P_3} \tag{4.33}$$
$$Kante\,3 = \overrightarrow{P_1P_4} \tag{4.34}$$
$$Kante\,4 = \overrightarrow{P_2P_3} \tag{4.35}$$
$$Kante\,5 = \overrightarrow{P_4P_2} \tag{4.36}$$
$$Kante\,6 = \overrightarrow{P_3P_4} \tag{4.37}$$

Für die Approximation des Vektorfeld $\vec{E}^e$ im Element e sind beim Tetraeder sechs *tangential unabhängige* Vektorbasisfunktionen $\vec{N}_i^e$ nötig:

$$\vec{E}^e = \sum_{i=1}^{6} E_i^e \, \vec{N}_i^e \,. \tag{4.38}$$

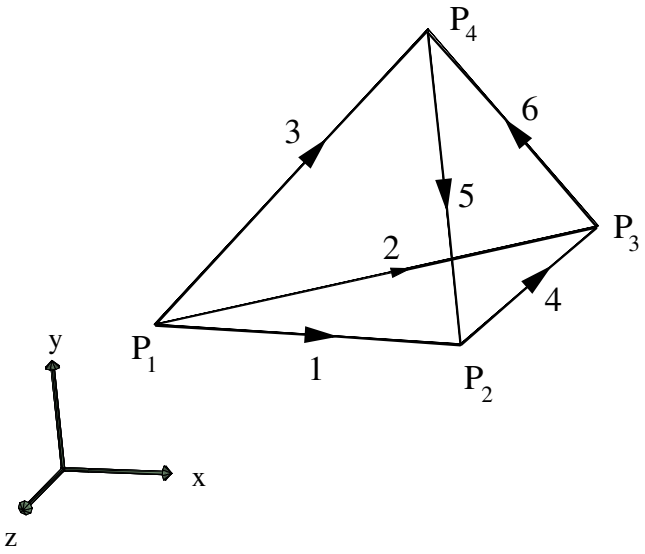


Abbildung 4.10: Kantenzuordnung im Tetraeder

Durch das Vorzeichen der Koeffizienten E_i^e kann auch die Richtung auf jeder Kante frei zugeordnet werden.

Die Vektorbasisfunktionen $\vec{N}_i^e$ müssen wie beim Dreieck mindestens eine der Bedingungen 1 oder 2 von S. 36 erfüllen:

1. Die Vektorbasisfunktion *verschwindet* auf den anderen Kanten *ganz*.
2. Die Vektorbasisfunktion liefert *nur* Beiträge zur *normalen Komponente* auf den anderen Kanten.

4.2.2.1 Der Ansatz der Vektorbasisfunktion

Um die Bedingungen zu erfüllen, kann für das Tetraeder derselbe Ansatz wie beim Dreieck (4.23 s. S. 36) verwendet werden, hier z.B. für die Kante 1:

$$\vec{W}_{12} = \xi_1 \vec{\nabla}\xi_2 - \xi_2 \vec{\nabla}\xi_1 \,. \tag{4.39}$$

4.2.2.2 Normierung der Vektorbasisfunktion

Der Ansatz (4.39) muß analog wie in 4.2.1.2 auf S. 37 mit der Länge l_i der Kante i normiert werden.

So ergibt sich für jede Kante i eine Vektorbasisfunktion $\vec{N}_i$, also

$$\vec{N}_1 = \vec{W}_{12}\, l_1 = (\xi_1 \vec{\nabla}\xi_2 - \xi_2 \vec{\nabla}\xi_1)\, l_1 \,, \tag{4.40}$$
$$\vec{N}_2 = \vec{W}_{13}\, l_2 = (\xi_1 \vec{\nabla}\xi_3 - \xi_3 \vec{\nabla}\xi_1)\, l_2 \,, \tag{4.41}$$
$$\vec{N}_3 = \vec{W}_{14}\, l_3 = (\xi_1 \vec{\nabla}\xi_4 - \xi_4 \vec{\nabla}\xi_1)\, l_3 \,, \tag{4.42}$$
$$\vec{N}_4 = \vec{W}_{23}\, l_4 = (\xi_2 \vec{\nabla}\xi_3 - \xi_3 \vec{\nabla}\xi_2)\, l_4 \,, \tag{4.43}$$
$$\vec{N}_5 = \vec{W}_{42}\, l_5 = (\xi_4 \vec{\nabla}\xi_2 - \xi_2 \vec{\nabla}\xi_4)\, l_5 \,, \tag{4.44}$$
$$\vec{N}_6 = \vec{W}_{34}\, l_6 = (\xi_3 \vec{\nabla}\xi_4 - \xi_4 \vec{\nabla}\xi_3)\, l_6 \tag{4.45}$$

und allgemein ausgedrückt, mit Kantenanfang ($Index: 1$) und Kantenende ($Index: 2$):

$$\vec{N}_i = \vec{W}_{12}\, l_i = (\xi_{i_1}\, \vec{\nabla}\xi_{i_2} - \xi_{i_2}\, \vec{\nabla}\xi_{i_1})\, l_i. \tag{4.46}$$

Jede der Vektorbasisfunktionen $\vec{N}_i$ stellt eine Rotation *um die gegenüberliegende Kante*[6] der Kante i dar.

4.3 Die Divergenz im Element

Wendet man den Divergenzoperator auf die einzelnen Ansatzfunktionen an, ergibt sich folgendes

$$\begin{aligned} div\, \vec{W}_{12} &= \vec{\nabla} \cdot \vec{W}_{12} && (4.47)\\ &= \vec{\nabla} \cdot (\xi_1 \vec{\nabla}\xi_2 - \xi_2 \vec{\nabla}\xi_1) \\ &= \vec{\nabla} \cdot (\xi_1 \vec{\nabla}\xi_2) - \vec{\nabla} \cdot (\xi_2 \vec{\nabla}\xi_1) \\ &= \vec{\nabla}\xi_1 \cdot \vec{\nabla}\xi_2 - \vec{\nabla}\xi_2 \cdot \vec{\nabla}\xi_1 \\ &= 0\,. && (4.48) \end{aligned}$$

Allgemein gilt dann

$$div\, \vec{W}_{ij} = 0\,. \tag{4.49}$$

4.4 Die Rotation im Element

Wendet man die Rotation auf die einzelnen Ansatzfunktionen an, ergibt sich

$$\begin{aligned} rot\, \vec{W}_{12} &= \vec{\nabla} \times \vec{W}_{12} && (4.50)\\ &= \vec{\nabla} \times (\xi_1 \vec{\nabla}\xi_2 - \xi_2 \vec{\nabla}\xi_1) \\ &= \vec{\nabla} \times (\xi_1 \vec{\nabla}\xi_2) - \vec{\nabla} \times (\xi_2 \vec{\nabla}\xi_1) \\ &= (\vec{\nabla}\xi_1 \times \vec{\nabla}\xi_2) - (\vec{\nabla}\xi_2 \times \vec{\nabla}\xi_1) \\ &= 2\,(\vec{\nabla}\xi_1 \times \vec{\nabla}\xi_2) && (4.51) \end{aligned}$$

und damit für $\vec{N}_1$:

$$rot\, \vec{N}_1 = 2 l_1\, (\vec{\nabla}\xi_1 \times \vec{\nabla}\xi_2)\,. \tag{4.52}$$

Allgemein gilt

[6] Die Vektorbasisfunktion $\vec{N}_i$ ist der Kante i zugeordnet. Diese Kante hat genau **eine** gegenüberliegende Kante, mit der sie keinen gemeinsammen Knoten teilt. Und genau um diese gegenüberliegende Kante drehen sich die Vektoren der Funktion $\vec{N}_i$.

$$rot\,\vec{W}_{ij} = 2\,(\vec{\nabla}\xi_i \times \vec{\nabla}\xi_j) \tag{4.53}$$

und

$$rot\,\vec{N}_i = 2l_i\,(\vec{\nabla}\xi_{i_1} \times \vec{\nabla}\xi_{i_2})\,. \tag{4.54}$$

Da der Gradient einer baryzentrischen Koordinate $\vec{\nabla}\xi_i$ immer konstant ist, ist auch das Kreuzprodukt in (4.53) immer ein konstanter Vektor. Somit ist die Rotation in einem Element immer eine konstanter Vektor.

4.5 Einordnung dieser Kantenelemente in der FEM

Man unterscheidet bei Kantenelementen zwei verschiedene Arten. Das Kriterium ist dabei der Operator der Diffentialgleichung, die gelöst werden soll. Auf der einen Seite gibt es die divergenzkonformen und auf der anderen die rotationskonformen Kantenelemente [36].

Divergenzkonforme Kantenelemente besitzen Stetigkeit in der *normalen* Komponente beim Übergang zwischen benachbarten Elementen. Bei diesen Ansatzfunktionen verschwindet die Divergenz in einem Element nicht (im Gegensatz zu GL (4.49)). Diese Ansatzfunktionen entsprechen dem Operator der EFIE (engl. Electric Field Integral Equation).

Rotationskonforme Kantenelemente besitzen Stetigkeit in der *tangentialen* Komponente bei Übergang zwischen benachbarten Elementen. Die Divergenz der Ansatzfunktionen verschwindet in einem Element (siehe (4.49)). Diese Ansatzfunktionen entsprechen z.B. dem Operator der vektoriellen Helmholzgleichung. Die in dieser Arbeit vorgestellten Kantenelemente sind die einfachsten dieser Art und sind am meisten verbreitet. Sie wurden zuerst von Nédélec [31] vorgestellt und werden nach Nédélec als *„first-order" edge elements* bezeichnet. Es gibt aber auch die Bezeichnung *„Whitney 1 elements".*

4.5.1 Kantenelemente höherer Ordnung

Wie bei Knotenelementen ist es auch bei Kantenelementen möglich, Ansatzfunktionen höherer Ordnung zu verwenden, um bei gleicher Diskretisierung eine höhere Genauigkeit der Lösung zu erreichen. Dabei sind bisher zwei verschiedene Wege beschritten:

1. Durch zusätzliche Basisfunktionen werden die einfachen Kantenelemente aus Abschnitt 4.2.2 additiv erweitert und man erhält dadurch eine höhere Genauigkeit im Element. Es handelt sich dabei um einen hierarchischen Ansatz. Einen Überblick über die verschiedenen Elemente erhält man schnell in [37]. Als Beispiel sind hier Ansatzfunktionen von J. P. Webb und B. Forghani tabellarisch angegeben [35]. Man spricht in diesem Zusammenhang auch von Kantenelementen, die komplett mit der Ordnung n sind. Die Bedeutung der Bezeichnung „Mixed Elements", womit Elemente mit gemischten Polynomordnung (hier $m = 0,5$ und $m = 1,5$) bezeichnet werden, wird hier verständlich.

m	Kantenordnung	Elementordnung	Anzahl der BasisFunktionen
$0,5$	konst. tangential	linear	6
$1,0$	linear tangential	linear	$12 = 6 + 6$
$1,5$	linear tangential	quadratisch	$20 = 6 + 6 + 8$
$2,0$	quadrat. tangential	quadratisch	$30 = 6 + 6 + 8 + 10$

Tabelle 4.1: Hierarchische Ansatzfunktionen von Kantenelementen nach [35]

Kantenelemente mit ganzer Polynomordnung m (hier $m = 1,0$ und $m = 2,0$) bezeichnet man auch als „konsistent" mit der Ordnung m (z.B.[34]). An Materialübergängen wo die Normalkomponente springen kann, können jedoch nur Ansatzfunktionen mit der Ordung $m = 0,5$ verwendet werden.

2. Die andere Möglichkeit besteht darin die einfachen Kantenelemente mit der Ordnung $m = 0,5$ von Nédélec [31] zu verwenden und durch Multiplikation mit Lagrangeschen Interpolationspolynomen den Freiheitsgrad den Ansatzfunktionen zu erhöhen. Diese Methode ist in [36] von Graglia, Wilton, Peterson gezeigt und überprüft worden. Sie hat den großen Vorteil, daß die Normalkomponente am Elementübergang weiterhin frei bleibt, was zum Erfüllen der Kontinuitätsbedingung an den Materialgrenzen notwendig ist.

4.5.2 Konvergenzeigenschaften der Kantenelemente für die FEM

Die Konvergenzeigenschaft wird durch das lokalen Fehlermaß $O(h^n)$ bezeichnet. Es beschreibt das Verhalten der Genauigkeit der Lösung, wenn man die Diskretisierung des Netzes (durch h ausgedrückt[7]) erhöht.

Es gibt hier leider nur wenig Literatur bei Kantenelementen. Die einzige mir bekannte Veröffentlichung stammt von Gerrit Mur [34]. Er vergleicht dort die numerische Effizienz zwischen Knoten- und Kantenelementen. Kantenelemente mit dem Polynomgrad $0,5$ (also die Whitney 1-Elemente) konvergieren demnach mit $O(h)$. Mit dem Polynomgrad $1,0$ (also konsistente Elemente) erreicht man dagegen $O(h^2)$, was auch mit linearen Knotenelementen erreicht wird.

[7] siehe Definition (2.3) auf S. 12

Kapitel 5

Hochfrequente elektromagnetische Felder

Technische Bauelemente werden in der Hochfrequenz gewöhnlich im Frequenzbereich beschrieben. Deshalb wird die weitere Betrachtung auf den zeitharmonischen Fall eingeschränkt. Hierbei läßt sich das elektromagnetische Feld mit der Kreisfrequenz $\omega > 0$ für jede Zeit $t \in \mathbb{R}$ darstellen als

$$\vec{E}(\vec{r},t) = \vec{E}(\vec{r})\, e^{j\omega t}\,, \tag{5.1}$$
$$\vec{H}(\vec{r},t) = \vec{H}(\vec{r})\, e^{j\omega t}\,. \tag{5.2}$$

$\vec{E}(\vec{r})$ und $\vec{H}(\vec{r})$ sind dabei komplexe Amplituden am Ort $\vec{r} = (x, y, z)$.

5.1 Problembeschreibung

Als nächstes wird das Randwertproblem mit seiner Differentialgleichung und den Randbedingungen beschrieben.

5.1.1 Die Wellengleichung im zeitharmonischen Fall

Ausgehend von einem Raum, der keine Ladungen enthält, und mit der Beschränkung auf ein lineares isotropes Medium gelten folgende Maxwell'sche Gleichungen z.B. [11, S.476] [10, S.728]:

$$rot\vec{H} = \kappa\,\vec{E} + \varepsilon\,\frac{\partial \vec{E}}{\partial t}\,, \tag{5.3}$$
$$rot\vec{E} = -\mu\,\frac{\partial \vec{H}}{\partial t}\,, \tag{5.4}$$
$$div\vec{H} = 0\,, \tag{5.5}$$
$$div\vec{E} = 0\,. \tag{5.6}$$

Angesichts des Ansatzes (5.1) und (5.2) gilt dann

$$rot\vec{H} = (\kappa + j\omega\varepsilon)\,\vec{E}\,, \tag{5.7}$$
$$rot\vec{E} = -j\omega\mu\,\vec{H}\,, \tag{5.8}$$
$$div\vec{H} = 0\,, \tag{5.9}$$
$$div\vec{E} = 0\,. \tag{5.10}$$

Gleichung (5.9) und (5.10) werden durch die Ansatzfunktionen der Kantenelemente lokal erfüllt (siehe Gleichung (4.49) auf Seite:42).

Multipliziert man die Gleichung (5.8) mit $\frac{1}{\mu}$ und bildet anschließend die Rotation, folgt

$$rot\left(\frac{1}{\mu}rot\vec{E}\right) = -j\omega\, rot\vec{H}\,. \tag{5.11}$$

Multipliziert man dann Gleichung (5.7) mit $-j\omega$ und setzt Gleichung (5.11) in Gleichung (5.7) ein, so erhält man

$$rot\left(\frac{1}{\mu}rot\vec{E}\right) = -\left(j\omega\kappa - \omega^2\varepsilon\right)\,\vec{E}\,. \tag{5.12}$$

Bringt man $-j\omega\kappa\vec{E}$ allein auf die rechte Seite

$$rot\left(\frac{1}{\mu}rot\vec{E}\right) - \omega^2\varepsilon\,\vec{E} = -j\omega\kappa\vec{E}\,. \tag{5.13}$$

Multipliziert man (5.13) mit μ_0 und führt man die Ausbreitungskonstante im freien Raum $k_0^2 = \omega^2\mu_0\varepsilon_0$, den Wellenwiederstand im freien Raum $Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$ und die Stromdichte $\vec{J} = \kappa\vec{E}$ ein, erhält man die vektorielle Wellengleichung für das elektrische Feld in der Form

$$rot\left(\frac{1}{\mu_r}rot\vec{E}\right) - k_0^2\varepsilon_r\vec{E} = -jk_0Z_0\vec{J}\,. \tag{5.14}$$

Mit gleichen Überlegungen läßt sich die vektorielle Wellengleichung für die magnetische Feldstärke herleiten

$$rot\left(\frac{1}{\varepsilon_r}rot\vec{H}\right) - k_0^2\mu_r\vec{H} = rot\left(\frac{1}{\varepsilon_r}\vec{J}\right)\,. \tag{5.15}$$

5.1.2 Randbedingungen

Die Wellengleichung ist für den gesamten Raum gültig. Da das diskretisierte Gebiet aber immer endlich sein muß, sind an den Rändern bestimmte Bedingungen (Randbedingungen) einzuhalten. Diese Randbedingungen sind ein Teil des zu lösenden Problems und erfordern besondere Aufmerksamkeit.

In der Regel stehen dem Anwender eines FEM-Programms nur eine begrenzte Anzahl von Randbedingungen zu Verfügung. So ist man bei der Beschreibung des Problems bemüht, das zu diskretisierende Problem immer innerhalb der zur Verfügung stehenden Randbedingungen zu definieren. Als nächstes sollen die gegebenen physikalischen Randbedingungen bei der Wellengleichung im zeitharmonischen Fall vorgestellt werden.

5.1.2.1 Elektrische Wand

Bei dieser Randbedingung (engl. **P**erfect **E**lectric **C**onductor, kurz: **PEC**) handelt es sich um eine Fläche mit elektrisch ideal leitender Eigenschaft ($\kappa \to \infty$). Das ist zum Beispiel Metall, unter der Annahme, daß die eventuell auftretenden Wandströme verlustfrei fließen können. Da diese Ströme Potentialunterschiede sofort verlustfrei ausgleichen, existieren auf der Oberfläche der Wand keine tangentialen elektrischen Feldstärken:

$$\hat{n} \times \vec{E} = 0 \tag{5.16}$$

bzw. in $\vec{H}$ ausgedrückt

$$\hat{n} \times rot\vec{H} = 0\,. \tag{5.17}$$

Ist die elektrische Wand verlustbehaftet, treten Wandströme auf, die nicht sofort verschwunden sind. Man kann dann die verlustbehaftete elektrische Wand durch eine widerstandsbehaftete Randbedingung (engl. **I**mpedance **B**oundary **C**ondition, kurz: **IBC**) nähern [25][24, S. 7]. Somit gilt dann die inhomogene Randbedingung

$$\widehat{n} \times \vec{E} = \vec{P} = -Z \cdot H_{tan} \tag{5.18}$$

bzw. in $\vec{H}$ ausgedrückt

$$\widehat{n} \times rot\overrightarrow{H} = \vec{P} = -Z \cdot H_{tan}\,. \tag{5.19}$$

Die elektrische Wand in IFEP Bei der Beschreibung des Problems in $\vec{E}$ (5.14)(5.16) ist die Diskretisierung recht einfach.

Die tangentialen elektrischen Feldstärken sind bei Kantenelementen direkt als unbekannte Variablen entlang den Kanten vorhanden. Liegt diese Randbedingung an dem Rand des diskretisierten Gebiets, so liegen auch die Kanten der Oberfläche direkt in der elektrischen Wand. Somit kann man die unbekannten Variablen dieser Kanten direkt auf Null setzen. Man spricht hier von einer *Dirichletschen Randbedingung*, da die unbekannten Variablen des FEM-Problems direkt gesetzt werden. Sie werden auch *Zwangsbedingungen* genannt [22, S.25].

Bei der Beschreibung des Problems in $\vec{H}$ (5.15)(5.17) ist die Diskretisierung auch recht einfach.

Man spricht hier von einer *natürlichen Randbedingung* [22, S.25] . Diese Randbedingung steckt implizit im Variationsintegral. Die Variablen auf diesem Rand werden also durch das Lösen des Problems gleich mitgelöst. Im Gegensatz zur Dirichletschen Randbedingung sind es aber noch Unbekannte, die erst bestimmt werden müssen.

Die verlustbehaftete elektrische Wand (IBC) ist in IFEP nicht implementiert. Es ist jedoch möglich, diese Randbedingung im Bedarfsfall noch zu implementieren [25].

5.1.2.2 Magnetische Wand

Bei dieser Randbedingung handelt es sich um eine Fläche mit magnetisch ideal leitender Eigenschaft ($\mu_r \to \infty$). Diese Randbedingung ist physikalisch unmöglich und wird hauptsächlich zum Beschreiben von Symmetrie-Ebenen verwendet. Die magnetische Feldstärke steht

senkrecht auf diesem Rand. Auf der Oberfläche der Wand muß die tangentiale magnetische Feldstärke dementsprechend Null sein:

$$\widehat{n} \times \vec{H} = 0 \tag{5.20}$$

bzw. in $\vec{E}$ ausgedrückt:

$$\widehat{n} \times rot\vec{E} = 0 \tag{5.21}$$

Die magnetischen Wand in IFEP Bei der Beschreibung des Problems in $\vec{H}$ (5.15)(5.20) ist die Diskretisierung der magnetischen Wand wieder eine Dirichletsche Randbedingung.

Die tangentialen magnetischen Feldstärken sind dann direkt als unbekannte Variablen auf den Kanten vorhanden. Liegt diese Randbedingung auf dem Rand des diskretisierten Gebiets, so liegen auch die Kanten der Oberfläche direkt in der magnetischen Wand. Somit kann man die unbekannten Variablen dieser Kanten direkt auf Null setzen.

Bei der Beschreibung des Problems in $\vec{E}$ (5.14)(5.21) ist die Diskretisierung des Randes eine natürliche Randbedingung und wird durch das Variationsintegral wieder implizit gelöst.

5.1.2.3 Absorbierende Randbedingungen

Neben einem Abschluß des Gebiets, (elektrische Wand) oder der Ausnutzung von Symmetrieen (magnetische Wand), gibt es auch **offene** Randflächen, an denen z.B. eine bestimmte Strahlung hindurch kann. Die Randbedingungen bewirken, daß die Welle von der Randbedingung absorbiert wird.

Hier soll eine absorbierende Randbedingung (engl. **A**bsorbing **B**oundary **C**ondition, kurz: **ABC**) für einen Rechteckwellenleiter wie in Abbildung 5.1 hergeleitet werden.

Der Wellenleiter ist an S_1 offen und wird mit dem TE_{10}-Mode angeregt. Dann folgt ein gerader Rechteckhohlleiter (Wände als PEC). Danach befindet sich ein Bereich, der durch geometrische oder materielle Veränderungen die Wellenausbreitung stört und wieder auf ein Rechteckhohlleiter führt, der an S_2 wieder den TE_{10}-Mode absorbiert.

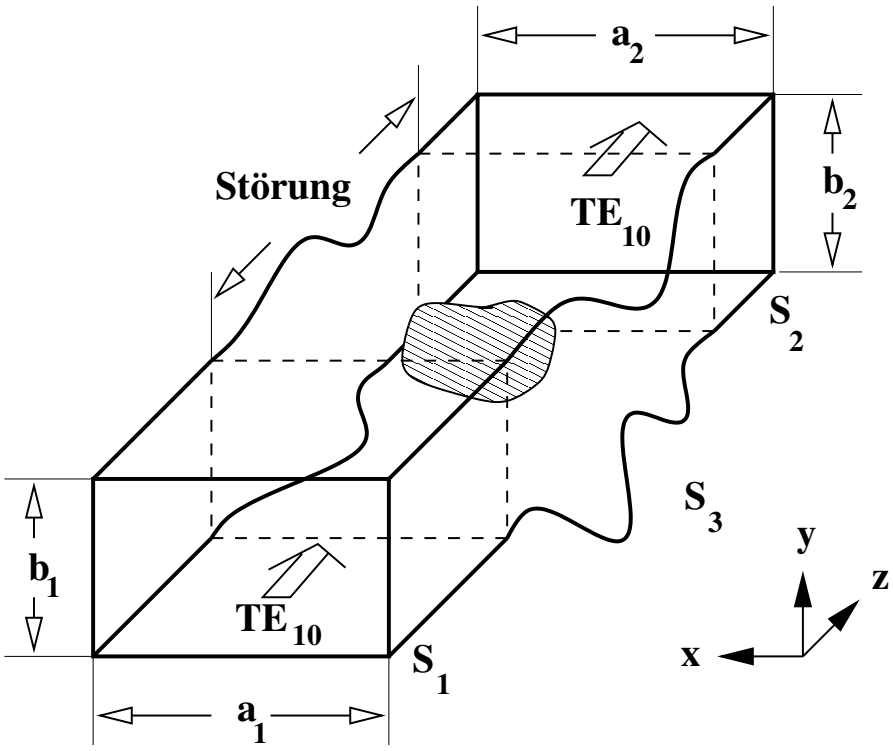


Abbildung 5.1: Wellenleiter mit TE_{10}-Mode Anregung

Damit der Wellenleiter aber auch so betrieben werden kann, müssen einige Bedingungen erfüllt werden:

1. In den Rechteckhohlleiterstücken ist nur der dominante Mode TE_{10} ungedämpft ausbreitungsfähig [10, S. 800ff].
2. Die Störstelle ist von den Toren S_1 und S_2 so weit entfernt, daß durch die Störung angeregte höhere Moden an den Toren soweit abgeklungen sind, daß sie vernachlässigt werden können.
3. Der anregende TE_{10}-Mode breitet sich an S_1 und S_2 nur in positiver z-Richtung aus.

ABC am Tor S_1 Die Störung bewirkt eine Reflexion der Wellen. Deshalb gilt für die elektrische Feldstärke $\vec{E}(x,y,z)$ am Tor S_1

$$\vec{E}^{S_1}(x,y,z) = \vec{E}^{ein}(x,y,z) + \vec{E}^{ref}(x,y,z) \tag{5.22}$$

$$= E_0\,\vec{e}_{10_1}(x,y)\,e^{-jk_{S_1}z} + R\,E_0\,\vec{e}_{10_1}(x,y)\,e^{jk_{S_1}z}\,. \tag{5.23}$$

Dabei steht *ein* für den Anteil der *einfallenden* Welle (die Anregung) und *ref* für den Anteil der *reflektierten* Welle. E_0 beschreibt die Amplitude der einfallenden Welle. R ist der Reflexionskoeffizient. $\vec{e}_{10_1}$ und k_{S_1} sind beim TE_{10}-Mode bei S_1 gegeben durch

$$\vec{e}_{10_1}(x,y) = \sin\left(\frac{\pi x}{a_1}\right)\vec{e}_y\,, \tag{5.24}$$

$$k_{S_1} = \sqrt{k_0^2 - \left(\frac{\pi}{a_1}\right)^2}\,. \tag{5.25}$$

Wobei a_1 die Weite des Rechteckhohlleiters bei S_1 ist (s.Abb 5.1).

Am Tor S_1 muß das elektrische Feld $\vec{E}^{S_1}(x,y,z)$ die eindimensionale Strahlungsbedingung für eine sich in z-Richtung ausbreitende ebene Wellen im freien Raum erfüllen:

$$\frac{\partial}{\partial z}\vec{E} = jk_{S_1}\vec{E}\,. \tag{5.26}$$

Wenn $\vec{E}^{S_1}(x,y,z)$ Gleichung (5.26) erfüllt, muß auch $\vec{E}^{ein}(x,y,z)$ und $\vec{E}^{ref}(x,y,z)$ die Bedingung erfüllen. Dann gilt für die reflektierte Welle

$$\frac{\partial}{\partial z}\vec{E}^{ref} = jk_{S_1}\vec{E}^{ref} \tag{5.27}$$

und mit Gleichung (5.22) auch

$$\frac{\partial}{\partial z}\left(\vec{E}^{S_1} - \vec{E}^{ein}\right) = jk_{S_1}\left(\vec{E}^{S_1} - \vec{E}^{ein}\right)\,. \tag{5.28}$$

Formt man (5.28) nach $\vec{E}^{S_1}$ und $\vec{E}^{ein}$ um, erhält man folgende Bedingung an Tor S_1

$$\frac{\partial}{\partial z}\vec{E}^{S_1} + jk_{S_1}\vec{E}^{S_1} = -2jk_{S_1}\vec{E}^{ein}\,. \tag{5.29}$$

Mit der Vektorgleichung (A.4) von Seite 119 und der Gewißheit, daß $\vec{E}^{ein}$ nur tangentiale Komponenten besitzt, kann man (5.29) schreiben als

$$\hat{n}\times\left(\vec{\nabla}\times\vec{E}^{S_1}\right) + jk_{S_1}\hat{n}\times\left(\hat{n}\times\vec{E}^{S_1}\right) = -2jk_{S_1}\vec{E}^{ein}\,. \tag{5.30}$$

Mit

$$\gamma_1 = \frac{jk_{S_1}}{\mu_r}, \qquad \vec{U}^{ein} = -\frac{2jk_{S_1}}{\mu_r}\vec{E}^{ein} \tag{5.31}$$

erhält man die Form (2.44) einer Cauchyschen Randbedingung:

$$\frac{1}{\mu_r}\hat{n} \times \left(rot\vec{E}\right) + \gamma_1 \hat{n} \times \left(\hat{n} \times \vec{E}\right) = \vec{U}^{ein} \tag{5.32}$$

Diese inhomogene Cauchysche Randbedingung ist also bei der Variation auf S_1 mit zu erfüllen.

ABC am Tor S_2 Durch die Reflexion der Welle an der Störung erreicht nur ein bestimmter Teil der Wellen das Tor S_2. Deshalb gilt für die elektrische Feldstärke $\vec{E}(x,y,z)$ an dem Tor S_2

$$\begin{aligned} \vec{E}^{S_2}(x,y,z) &= \vec{E}^{trans}(x,y,z) && (5.33)\\ &= T\,\vec{E}^{ein}(x,y,z) && (5.34)\\ &= T\,E_0\,\vec{e}_{10_2}(x,y)\,e^{-jk_{S_2}z}\,. && (5.35) \end{aligned}$$

Dabei steht *trans* für den Anteil der *transmittierten* Welle und T ist der Transmissionskoeffizient[1]. $\vec{e}_{10_2}$ und k_{S_2} sind beim TE_{10}-Mode durch die Geometrie bei S_2 gegeben durch

$$\begin{aligned} \vec{e}_{10_2}(x,y) &= \sin\left(\frac{\pi x}{a_2}\right)\vec{e}_y\,, && (5.36)\\ k_{S_2} &= \sqrt{k_0^2 - \left(\frac{\pi}{a_2}\right)^2}\,. && (5.37) \end{aligned}$$

Wobei a_2 die Weite des Rechteckhohlleiters bei S_2 ist (s.Abb 5.1).

Am Tor S_2 muß das elektrische Feld $\vec{E}^{S_2}(x,y,z)$ wieder die eindimensionale Strahlungsbedingung für eine sich in z-Richtung ausbreitende ebene Wellen im freien Raum (5.26) erfüllen.

Mit (5.33) ergibt sich dann die Bedingung

$$\begin{aligned} \hat{n} \times \left(\vec{\nabla} \times \vec{E}^{S_2}\right) = \vec{e}_z \times \left(\vec{\nabla} \times \vec{E}^{S_2}\right) &= jk_{S_2}\vec{E}^{trans}_{tan} && (5.38)\\ \vec{e}_z \times \left(\vec{\nabla} \times \vec{E}^{S_2}\right) - jk_{S_2}\vec{E}^{trans}_{tan} &= 0\,. && (5.39) \end{aligned}$$

Mit der Vektorgleichung (A.4) kann man (5.39) schreiben als

$$\hat{n} \times \left(\vec{\nabla} \times \vec{E}^{S_2}\right) + jk_{S_2}\hat{n} \times \left(\hat{n} \times \vec{E}^{S_2}\right) = 0\,. \tag{5.40}$$

[1] Unter der Annahme, daß Tor S_1 auch die selben Maße besitzt wie Tor S_2 stimmt das. Sonst berechnet sich der Transmissionsfaktor über die Leistung (siehe Abschnitt 7.2.2 auf S. 94)

Mit

$$\gamma_2 = \frac{jk_{S_2}}{\mu_r} \tag{5.41}$$

erhält man die Form (2.44) einer Cauchyschen Randbedingungen

$$\frac{1}{\mu_r}\hat{n} \times \left(rot\vec{E}\right) + \gamma_2 \hat{n} \times \left(\hat{n} \times \vec{E}\right) = 0\,. \tag{5.42}$$

Diese homogene Cauchysche Randbedingung ist also bei der Variation auf S_2 mit zu erfüllen.

Für das magnetischen Feld sei hier nur der Vollständigkeit halber die allgemeine Form der Randbedingung angegeben

$$\frac{1}{\varepsilon_r}\hat{n} \times \left(rot\vec{H}\right) + \gamma_H \hat{n} \times \left(\hat{n} \times \vec{H}\right) = \vec{V}\,. \tag{5.43}$$

Dabei ist γ_H ein bekannter Parameter und $\vec{V}$ ein bekannter Vektor.

5.1.3 Kontinuitätsbedingung

An der Grenzfläche zwischen zwei unterschiedlichen Materialien sind die folgenden Kontinuitätsbedingungen einzuhalten. Die tangentiale Stetigkeit der elektrischen Feldstärke $\vec{E}$

$$\hat{n} \times \vec{E}^{+} \;=\; \hat{n} \times \vec{E}^{-} \tag{5.44}$$

bzw. in $\vec{H}$ ausgedrückt

$$\frac{1}{\varepsilon_r^{+}}\,\hat{n} \times \left(\vec{\nabla} \times \vec{H}^{+}\right) = \frac{1}{\varepsilon_r^{-}}\,\hat{n} \times \left(\vec{\nabla} \times \vec{H}^{-}\right) \tag{5.45}$$

und die tangentiale Stetigkeit der magnetischen Feldstärke $\vec{H}$

$$\hat{n} \times \vec{H}^{+} = \hat{n} \times \vec{H}^{-} \tag{5.46}$$

bzw. in $\vec{E}$ ausgedrückt

$$\frac{1}{\mu_r^{+}}\,\hat{n} \times \left(\vec{\nabla} \times \vec{E}^{+}\right) \;=\; \frac{1}{\mu_r^{-}}\,\hat{n} \times \left(\vec{\nabla} \times \vec{E}^{-}\right)\,. \tag{5.47}$$

5.2 Das Variationsproblem der Wellengleichung

In Kapitel 2.3.1.1 wurde die Ritzsche Methode schon beschrieben. Bei den Wellengleichungen (5.14) und (5.15) handelt es sich jedoch um vektorielle Gleichungen. Die Ritzsche Methode ist auch auf vektorielle Probleme anwendbar. Dabei ergeben sich jedoch für elektromagnetische Probleme Grenzen, die im folgenden Abschnitt verdeutlicht werden sollen. Anschließend wird ein generelles Variationsverfahren vorgestellt, mit dem die meisten Randwertprobleme in der elektromagnetischen Feldberechnung gelöst werden können [24, S. 189].

5.2.1 Beschränkungen bei der klassischen Ritz-Methode

Dazu führt man das vektorielle innere Produkt ein

$$\left\langle \vec{a}, \vec{b} \right\rangle = \iiint_{\Omega} \vec{a} \cdot \vec{b}^{*} \, d\Omega \,. \tag{5.48}$$

Die vektorielle Wellengleichung für das elektrische Feld (5.14) folgt mit dem Operator $\mathcal{L}$ in der Nabla-Schreibweise

$$\mathcal{L} = \vec{\nabla} \times \left(\frac{1}{\mu_r} \vec{\nabla} \times \right) - k_0^2 \varepsilon_r \tag{5.49}$$

folgendermaßen

$$\mathcal{L}\vec{E} = f \,. \tag{5.50}$$

Für das Innere Produkt[2] des Ritzschen Funktionals (2.15) gilt dann

$$\left\langle \mathcal{L}\vec{E}, \vec{F} \right\rangle = \iiint_{\Omega} \vec{F}^{*} \cdot \left[\vec{\nabla} \times \left(\frac{1}{\mu_r} \vec{\nabla} \times \vec{E} \right) - k_0^2 \varepsilon_r \vec{E} \right] d\Omega \,. \tag{5.51}$$

Wendet man den zweiten vektoriellen Greenschen Satz (A.8) an, ergibt sich

$$\begin{aligned} \left\langle \mathcal{L}\vec{E}, \vec{F} \right\rangle \;=\; & \iiint_{\Omega} \vec{E} \cdot \left[\vec{\nabla} \times \left(\frac{1}{\mu_r} \vec{\nabla} \times \vec{F}^{*} \right) - k_0^2 \varepsilon_r \vec{F}^{*} \right] d\Omega \\ & + \oiint_{S} \frac{1}{\mu_r} \left[\vec{E} \times \left(\vec{\nabla} \times \vec{F}^{*} \right) - \vec{F}^{*} \times \left(\vec{\nabla} \times \vec{E} \right) \right] \cdot \hat{n} \, dS \,. \end{aligned} \tag{5.52}$$

Mit dem Entwicklungssatz (A.2) und (A.3) und dem Spatprodukt (A.1) gilt

$$\left[\vec{E} \times \left(\vec{\nabla} \times \vec{F}^{*} \right) \right] \cdot \hat{n} = \left(\hat{n} \times \vec{E} \right) \left(\vec{\nabla} \times \vec{F}^{*} \right) = -\vec{E} \left[\hat{n} \times \left(\vec{\nabla} \times \vec{F}^{*} \right) \right] . \tag{5.53}$$

Damit ist klar, wenn $\vec{E}$ und $\vec{F}$ die homogene Dirichletsche Randbedingung (5.16)

$$\hat{n} \times \vec{E} = 0 \quad auf \quad S_3 \tag{5.54}$$

und die homogene Cauchysche Randbedingung dritter Art,

$$\frac{1}{\mu_r} \hat{n} \times \left(\vec{\nabla} \times \vec{E} \right) + \gamma_E \, \hat{n} \times \left(\hat{n} \times \vec{E} \right) = 0 \quad auf \quad S_2 \tag{5.55}$$

mit $S = S_2 + S_3$ erfüllen, dann verschwindet das Oberflächenintegral in (5.52) unter der Bedingung, daß μ_r und γ_E reell sind. Sind ε_r und μ_r reell, dann kann man (5.52) schreiben als:

$$\left\langle \mathcal{L}\vec{E}, \vec{F} \right\rangle = \left\langle \vec{E}, \mathcal{L}\vec{F} \right\rangle \tag{5.56}$$

[2]Im Gegensatz zu (2.15) ist hier $\vec{F}$ statt $\vec{E}$ für das eine $\vec{E}$ im inneren Produkt geschrieben worden, um für die weitere Rechnung die einzelnen Teile des inneren Produkt besser verfolgen zu konnen. Also es gilt $\vec{F} = \vec{E}$.

und damit ist $\mathcal{L}$ selbstadjungiert. Die Bedingung, damit $\mathcal{L}$ selbstadjungiert ist, ist also zum einen, daß ε_r, $\mu_r$und γ_E **reellwertig** sind und zum anderen, daß die **homogenen** Randbedingungen erfüllt werden.

Dies stellt bei elektromagnetischen Problemen eine große Einschränkung dar. Wenn verlustbehaftete Medien berechnet werden sollen, sind ε_r und μ_r komplexwertig. Eine weitere große Einschränkung sind die homogenen Randbedingungen. Die Randbedingung (5.54) könnte inhomogen sein, d.h. es könnten elektrische Wandströme $\vec{P}$ auftreten. Oder die Randbedingung (5.55) wäre inhomogen, wenn z.B. anregende Felder $\vec{U}$ auf einem Randgebiet auftreten (siehe S_1). In diesen Fällen wären die Randbedingungen inhomogen und das Problem ist mit dieser Methode nicht mehr lösbar [24, S.181ff].

5.2.2 Ein modifiziertes Variationsverfahren

Man kann durch eine Modifikation der Ansatzfunktion (hier $\vec{E}(x,y,z)$) das Problem mit inhomogenen Randbedingungen wieder zu einem mit einem selbstadjungierten Operator $\mathcal{L}$ überführen. Dazu führt man eine neue Ansatzfunktion

$$\vec{E}' = \vec{E} - \vec{u} \tag{5.57}$$

ein. Dabei ist $\vec{u}$ eine Funktion, die die **inhomogenen** Randbedingungen **erfüllt**. Deshalb brauchen mit Ansatz (5.57) wieder nur die homogenen Randbedingungen erfüllt zu werden und das Problem wird wieder selbstadjungiert, aber **immer noch unter der Bedingung, daß ε_r, $\mu_r$und γ_e reellwertig sind**. Wendet man nun die klassische Ritzsche Methode an und verwendet $\vec{E} = \vec{E}' + \vec{u}$ im Funktional (2.15) von Seite 16, so kann man die Differentialgleichung von $\vec{E}'$ schreiben als:

$$\mathcal{L}\vec{E}' = \vec{f}' \tag{5.58}$$

mit $\vec{f}' = \vec{f} - \mathcal{L}\vec{u}$. Für das Funktional ergibt sich dann:

$$F(\vec{E}') = \frac{1}{2}\left\langle \mathcal{L}\vec{E}', \vec{E}' \right\rangle - \frac{1}{2}\left\langle \vec{E}', \vec{f}' \right\rangle - \frac{1}{2}\left\langle \vec{f}', \vec{E}' \right\rangle \tag{5.59}$$

oder

$$\begin{aligned} F(\vec{E}) \;=\; & \frac{1}{2}\left\langle \mathcal{L}\left(\vec{E}-\vec{u}\right), \left(\vec{E}-\vec{u}\right) \right\rangle - \frac{1}{2}\left\langle \left(\vec{E}-\vec{u}\right), \left(\vec{f}-\mathcal{L}\vec{u}\right) \right\rangle \\ & -\frac{1}{2}\left\langle \left(\vec{f}-\mathcal{L}\vec{u}\right), \left(\vec{E}-\vec{u}\right) \right\rangle \end{aligned} \tag{5.60}$$

Da das Funktional nur über Terme variiert wird, die von $\vec{E}$ abhängen, können Terme, die $\vec{E}$ nicht enthalten, entfallen. Damit ergibt sich für das Funktional:

$$F(\vec{E}) = \frac{1}{2}\left\langle \mathcal{L}\vec{E}, \vec{E} \right\rangle - \frac{1}{2}\left\langle \mathcal{L}\vec{E}, \vec{u} \right\rangle + \frac{1}{2}\left\langle \vec{E}, \mathcal{L}\vec{u} \right\rangle - \frac{1}{2}\left\langle \vec{E}, \vec{f} \right\rangle - \frac{1}{2}\left\langle \vec{f}, \vec{E} \right\rangle \tag{5.61}$$

Unter Verwendung des ersten vektoriellen Greenschen Satz (A.7) und der inhomogenen Randbedingungen (5.18) auf S_3 und (5.32) auf S_1 mit $S = S_1 + S_3$ erhält man nach einiger Um-

rechnung [24, S. 186] das Funktional:

$$\begin{aligned} F(\vec{E}) &= \frac{1}{2}\iiint\limits_V \left[\frac{1}{\mu_r}\left(\vec{\nabla}\times\vec{E}\right)\cdot\left(\vec{\nabla}\times\vec{E}\right)^* - k_0^2\varepsilon_r\vec{E}\cdot\vec{E}^*\right] dV \\ &+ \frac{1}{2}\iint\limits_{S_2}\left[\gamma_E\left(\hat{n}\times\vec{E}\right)\cdot\left(\hat{n}\times\vec{E}^*\right) + \vec{E}\cdot\vec{U}^* + \vec{E}^*\cdot\vec{U}\right] dS \\ &+ \frac{jk_0Z_0}{2}\iiint\limits_V \left(\vec{E}^*\cdot\vec{J} - \vec{E}\cdot\vec{J}^*\right) dV \end{aligned} \tag{5.62}$$

Trotzdem bleibt noch die Einschränkung, daß ε_r, $\mu_r$und γ_E reellwertig sein müssen.

5.2.3 Ein generelles Variationsverfahren der Wellengleichung

Nach [24, S. 188] gibt es jedoch ein generelles Variationsverfahren. Wie im Abschnitt 5.2.1 gezeigt wurde, ist die Forderung nach einem reellwertigen Operator $\mathcal{L}$ eine direkte Folge der Definition des inneren Produkts (5.48) bzw. im skalaren Fall (2.12). Aus diesem Grund verwendet man nicht das innere Produkt (5.48) bzw. (2.12) des Hilbertraums, sondern das symmetrische Produkt

$$\langle\phi,\psi\rangle = \iiint\limits_\Omega \phi\cdot\psi\, d\Omega \tag{5.63}$$

bzw vektoriell

$$\left\langle\vec{a},\vec{b}\right\rangle = \iiint\limits_\Omega \vec{a}\cdot\vec{b}\, d\Omega\,. \tag{5.64}$$

Um der Frage nachzugehen, ob das eingeführte Variationsprinzip noch Gültigkeit besitzt, soll zunächst die erste Variation des Funktionals (2.15) auf S. 16 betrachtet werden:

$$\delta F(\vec{E}) = \frac{1}{2}\left\langle\mathcal{L}\delta\vec{E},\vec{E}\right\rangle + \frac{1}{2}\left\langle\mathcal{L}\vec{E},\delta\vec{E}\right\rangle - \frac{1}{2}\left\langle\delta\vec{E},\vec{f}\right\rangle - \frac{1}{2}\left\langle\vec{E},\delta\vec{f}\right\rangle \tag{5.65}$$

Da $\mathcal{L}$ auch mit dem symmetrischen Produkt selbstadjungiert ist, kann man (5.65) schreiben als:

$$\delta F(\vec{E}) = \frac{1}{2}\left\langle\delta\vec{E},\mathcal{L}\vec{E} - f\right\rangle.$$

Will man den stationären Punkt der ersten Variation $\delta F = 0$ erzwingen, folgt:

$$\left\langle\delta\vec{E},\mathcal{L}\vec{E} - f\right\rangle \stackrel{!}{=} 0$$

Man sieht, daß $\vec{E}$ die Diffentialgleichung (5.50) von S. 53 erfüllen muß, da $\delta\vec{E}$ eine willkürliche Variation ist.

Damit ist das Variationsprinzip weiterhin gültig, obwohl das symmetrische Produkt nicht mehr die Bedingung des Hilbertraums nach einem Skalarprodukt mit einem reellen Ergebnis

erfüllt. Es handelt sich bei dem neu eingeführten Funktional um eine komplexe Qualität für ein komlexes Problem. Die Frage nach einem Minimum, Maximum oder Sattelpunkt ist hier nicht mehr von Bedeutung, da bei diesem Funktional nicht mehr Qualitäten beurteilt werden, die einer physikalischen Bedeutung wie Energie, Arbeit oder Leistung entsprechen. Die Forderung nach einem „stationären Punkt der ersten Variation δF" ist das benötigte Kriterium der Variationsrechnung [24, S.189].

Da die inhomogenen Randbedingungen weiterhin berücksichtigt werden sollen, verwendet man statt (2.15) das Funktional (5.61) des modifizierte Variationsverfahren mit dem symmetrischen Produkt (5.64):

$$F(\vec{E}) = \frac{1}{2}\left\langle \mathcal{L}\vec{E}, \vec{E} \right\rangle - \frac{1}{2}\left\langle \mathcal{L}\vec{E}, \vec{u} \right\rangle + \frac{1}{2}\left\langle \vec{E}, \mathcal{L}\vec{u} \right\rangle - \left\langle \vec{E}, \vec{f} \right\rangle \tag{5.66}$$

Für die vektorielle Wellengleichung (5.14) mit dem Operator (5.49) und $\vec{f}$ als

$$\vec{f} = -jk_0 Z_0 \vec{J} \tag{5.67}$$

ergibt sich

$$\begin{aligned} F(\vec{E}) \;=\; & \frac{1}{2}\iiint_V \left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\right]\cdot\vec{E}\,dV \\ & - \frac{1}{2}\iiint_V \left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\right]\cdot\vec{u}\,dV \\ & + \frac{1}{2}\iiint_V \vec{E}\cdot\left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{u}\right) - k_0^2\varepsilon_r\vec{u}\right]\,dV \\ & - \iiint_V \vec{E}\cdot\left(-jk_0 Z_0\vec{J}\right)\,dV \end{aligned} \tag{5.68}$$

Wendet man den ersten vektoriellen Greenschen Satz auf die zweite und dritte Zeile von (5.68) an, folgt:

$$\begin{aligned} F(\vec{E}) \;=\; & \frac{1}{2}\iiint_V \left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\right]\cdot\vec{E}\,dV \\ & + \oiint_S \frac{1}{\mu_r}\left[\vec{u}\times\left(\vec{\nabla}\times\vec{E}\right) - \vec{E}\times\left(\vec{\nabla}\times\vec{u}\right)\right]\cdot\hat{n}\,dS \\ & + jk_0 Z_0 \iiint_V \vec{E}\cdot\vec{J}\,dV \end{aligned} \tag{5.69}$$

Da $\vec{E}$ und $\vec{u}$ die inhomogenen Randbedingungen (5.18) auf S_3 und (5.32) auf S_1 mit $S = S_1 + S_3$ erfüllen, kann man in (5.69) das Oberflächenintegral zerlegen und mit Verwendung von (5.53)

schreiben als:

$$\begin{aligned} F(\vec{E}) &= \frac{1}{2}\iiint\limits_V \left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\right]\cdot\vec{E}\,dV \\ &+ \iint\limits_{S_3}\frac{1}{\mu_r}\left[\vec{P}\cdot\left(\vec{\nabla}\times\vec{E}\right) - \vec{P}\cdot\left(\vec{\nabla}\times\vec{u}\right)\right]dS \\ &+ \iint\limits_{S_1}\left[\vec{E}\cdot\vec{U} - \vec{u}\cdot\vec{U}\right]dS \\ &+ jk_0Z_0\iiint\limits_V \vec{E}\cdot\vec{J}\,dV \end{aligned} \tag{5.70}$$

Streicht man die Terme, die von $\vec{E}$ unabhängig sind, folgt:

$$\begin{aligned} F(\vec{E}) &= \frac{1}{2}\iiint\limits_V \left[\vec{\nabla}\times\left(\frac{1}{\mu_r}\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\right]\cdot\vec{E}\,dV \\ &+ \iint\limits_{S_3}\frac{1}{\mu_r}\left[\vec{P}\cdot\left(\vec{\nabla}\times\vec{E}\right)\right]dS \\ &+ \iint\limits_{S_1}\vec{E}\cdot\vec{U}\,dS \\ &+ jk_0Z_0\iiint\limits_V \vec{E}\cdot\vec{J}\,dV \end{aligned} \tag{5.71}$$

Wendet man den ersten vektoriellen Greenschen Satz jetzt auf die erste Zeile von (5.71) an:

$$\begin{aligned} F(\vec{E}) &= \frac{1}{2}\iiint\limits_V \left[\frac{1}{\mu_r}\left(\vec{\nabla}\times\vec{E}\right)\cdot\left(\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\cdot\vec{E}\right]dV \\ &- \oiint\limits_S \frac{1}{\mu_r}\vec{E}\times\left(\vec{\nabla}\times\vec{E}\right)\hat{n}\cdot dS \\ &+ \iint\limits_{S_3}\frac{1}{\mu_r}\left[\vec{P}\cdot\left(\vec{\nabla}\times\vec{E}\right)\right]dS \\ &+ \iint\limits_{S_1}\vec{E}\cdot\vec{U}\,dS \\ &+ jk_0Z_0\iiint\limits_V \vec{E}\cdot\vec{J}\,dV \end{aligned} \tag{5.72}$$

Zerlegt man das Oberflächenintegral in (5.72) wieder mit den inhomogenen Randbedingungen (5.18) auf S_1 und (5.32) auf S_2 und verwendet (5.53), ergibt sich mit den Entwicklungssätzen (A.2),(A.3):

$$
\begin{aligned}
F(\vec{E}) \;=\; & \frac{1}{2}\iiint_V \left[\frac{1}{\mu_r}\left(\vec{\nabla}\times\vec{E}\right)\cdot\left(\vec{\nabla}\times\vec{E}\right) - k_0^2\varepsilon_r\vec{E}\cdot\vec{E}\right] dV \\
& + \iint_{S_1} \left[\frac{\gamma_E}{2}\left(\hat{n}\times\vec{E}\right)\cdot\left(\hat{n}\times\vec{E}\right) + \vec{E}\cdot\vec{U}\right] dS \\
& + jk_0Z_0 \iiint_V \vec{E}\cdot\vec{J}\,dV
\end{aligned} \tag{5.73}
$$

Dieses Funktional (5.73) ist nun auch für *komplexes* ε_r, $\mu_r$und γ_E und *inhomogene* Randbedingungen gültig [24, S. 190]. Für die vektorielle Wellengleichung mit der magnetischen Feldstärke $\vec{H}$ (5.15) kann man mit demselben Rechenweg das äquivalente Funktional herleiten:

$$
\begin{aligned}
F(\vec{H}) \;=\; & \frac{1}{2}\iiint_V \left[\frac{1}{\varepsilon_r}\left(\vec{\nabla}\times\vec{H}\right)\cdot\left(\vec{\nabla}\times\vec{H}\right) - k_0^2\mu_r\vec{H}\cdot\vec{H}\right] dV \\
& + \iint_{S_1} \left[\frac{\gamma_H}{2}\left(\hat{n}\times\vec{H}\right)\cdot\left(\hat{n}\times\vec{H}\right) + \vec{H}\cdot\vec{V}\right] dS \\
& - \iiint_V \vec{H}\cdot\left(\vec{\nabla}\times\frac{1}{\varepsilon_r}\vec{J}\right) dV
\end{aligned} \tag{5.74}
$$

Es ist hier der Vollständigkeit halber noch angegeben. Im Folgenden wird nur noch die Formulierung mit der elektrischen Feldstärke betrachtet, da sie auch im Programm IFEP verwendet wird.

5.3 Lokales Variationsproblem der Wellengleichung

Im globalen Variationsproblem mit dem Funktional (5.73) ist nun, wie in Abschnitt 2.3.2 gezeigt, die globale Ansatzfunktion durch lokale Ansatzfunktionen zu ersetzen. Dabei sollen jetzt die lokalen vektoriellen Ansatzfunktionen für Tetraeder von Abschnitt 4.2.2 verwendet werden. In diesem Fall wird die lokale Ansatzfunktion (2.28) mit den Kantenelementen (4.38) zu:

$$
\vec{E}^e = \sum_{i=1}^{6} E_i^e\,\vec{N}_i^e = \{E^e\}^T\cdot\left\{\vec{N}^e\right\} = \left\{\vec{N}^e\right\}^T\cdot\{E^e\} \tag{5.75}
$$

Setzt man den lokalen Ansatz (5.75) in das Funktional (5.73) ein, so erhält man das Funktional für ein Tetraederelement:

$$F^e(\vec{E}^e) = \frac{1}{2}\iiint_V \left[\frac{1}{\mu_r^e}\left(\vec{\nabla}\times\vec{E}^e\right)\cdot\left(\vec{\nabla}\times\vec{E}^e\right) - k_0^2\varepsilon_r^e\vec{E}^e\cdot\vec{E}^e\right] dV \tag{5.76}$$

$$+\iint_{S_1}\left[\frac{\gamma_E}{2}\left(\hat{n}\times\vec{E}^e\right)\cdot\left(\hat{n}\times\vec{E}^e\right) + \vec{E}^e\cdot\vec{U}\right] dS \tag{5.77}$$

$$+ jk_0Z_0\iiint_V \vec{E}^e\cdot\vec{J}\,dV \tag{5.78}$$

Als nächstes werden die drei Anteile des Funktionals getrennt betrachtet. Im Vordergrund steht nun das Berechnen der Integrale. Dabei wird (5.76) als Volumenintegral F_V^e, (5.77) als Randintegral F_R^e und (5.78) als Stromintegral F_J^e bezeichnet.

5.3.1 Das Volumenintegral

Setzt man in (5.76) die lokalen Ansatzfunktionen ein, kann man die Summanden ausmultiplizieren. Die Koeffizenten E_i^e für die tangentiale elektrische Feldstärke sind *für die Integration konstant* und können aus dem Integral herausgezogen werden. Damit ergibt sich für das Volumenintegral:

$$F_V^e = \frac{1}{2}\left(\{E^e\}^T[A^e]\{E^e\} - k_0^2\{E^e\}^T[B^e]\{E^e\}\right) \tag{5.79}$$

mit den Matrixelementen

$$A_{ij}^e = \iiint_{V^e}\frac{1}{\mu_r^e}\left\{\vec{\nabla}\times\vec{N}_i^e\right\}\left\{\vec{\nabla}\times\vec{N}_j^e\right\} dV \tag{5.80}$$

der 6×6 Matrix A^e und den Matrixelementen

$$B_{ij}^e = \iiint_{V^e}\varepsilon_r^e\,\vec{N}_i^e\cdot\vec{N}_j^e\,dV \tag{5.81}$$

der 6×6 Matrix B^e.

Um das Integral (5.80) berechnen zu können, braucht man den Ausdruck $\vec{\nabla}\times\vec{N}_i^e$ in kartesischen Koordinaten. Setzt man (B.14) in (4.54) von S.43 ein und führt das Kreuzprodukt durch, ergibt sich:

$$\vec{\nabla}\times\vec{N}_i^e = 2l_i^e\left(\vec{\nabla}\xi_{i_1}^e\times\vec{\nabla}\xi_{i_2}^e\right)$$

$$= \frac{2l_i^e}{(6V^e)^2}\left[\left(c_{i_1}^e d_{i_2}^e - d_{i_1}^e c_{i_2}^e\right)\vec{e}_{x+}\left(d_{i_1}^e b_{i_2}^e - b_{i_1}^e d_{i_2}^e\right)\vec{e}_{y+}\left(b_{i_1}^e c_{i_2}^e - c_{i_1}^e b_{i_2}^e\right)\vec{e}_z\right] \tag{5.82}$$

Man sieht hier deutlich, daß $\vec{\nabla} \times \vec{N}_i^e$ ein konstanter Vektor im Element ist. Damit ergibt sich für das Skalarprodukt ein konstanter Wert im Element:

$$\begin{aligned}\left\{\vec{\nabla} \times \vec{N}_i^e\right\}\left\{\vec{\nabla} \times \vec{N}_j^e\right\} &= \frac{4 l_i^e l_j^e}{\left(6V^e\right)^4}\left[\left(c_{i_1}^e d_{i_2}^e - d_{i_1}^e c_{i_2}^e\right)\left(c_{j_1}^e d_{j_2}^e - d_{j_1}^e c_{j_2}^e\right) + \right. \\ &\quad \left(d_{i_1}^e b_{i_2}^e - b_{i_1}^e d_{i_2}^e\right)\left(d_{j_1}^e b_{j_2}^e - b_{j_1}^e d_{j_2}^e\right) + \\ &\quad \left.\left(b_{i_1}^e c_{i_2}^e - c_{i_1}^e b_{i_2}^e\right)\left(b_{j_1}^e c_{j_2}^e - c_{j_1}^e b_{j_2}^e\right)\right] \end{aligned} \tag{5.83}$$

Da (5.83) im Integral (5.80) konstant ist, kann man es vor das Integral ziehen und es folgt:

$$\begin{aligned} A_{ij}^e &= \frac{4 l_i^e l_j^e V^e}{\left(\mu_r^e 6V^e\right)^4}\left[\left(c_{i_1}^e d_{i_2}^e - d_{i_1}^e c_{i_2}^e\right)\left(c_{j_1}^e d_{j_2}^e - d_{j_1}^e c_{j_2}^e\right) + \right. \\ &\quad \left(d_{i_1}^e b_{i_2}^e - b_{i_1}^e d_{i_2}^e\right)\left(d_{j_1}^e b_{j_2}^e - b_{j_1}^e d_{j_2}^e\right) + \\ &\quad \left.\left(b_{i_1}^e c_{i_2}^e - c_{i_1}^e b_{i_2}^e\right)\left(b_{j_1}^e c_{j_2}^e - c_{j_1}^e b_{j_2}^e\right)\right] \end{aligned} \tag{5.84}$$

Somit ist die Matrix A im Element e exakt bekannt.

Bei dem Integral (5.81) geht man folgendermaßen vor: Setzt man den Gradienten in kartesischen Koordinaten (B.14) in (4.46) ein und führt das Skalarprodukt $\vec{N}_i^e \cdot \vec{N}_j^e$ durch, ergibt sich:

$$\vec{N}_i^e \cdot \vec{N}_j^e = \frac{l_i^e l_j^e}{(6V)^2}\left[\xi_{i_1}^e \xi_{j_1}^e f_{i_2 j_2} - \xi_{i_1}^e \xi_{j_2}^e f_{i_2 j_1} - \xi_{i_2}^e \xi_{j_1}^e f_{i_1 j_2} + \xi_{i_2}^e \xi_{j_2}^e f_{i_1 j_1}\right] \tag{5.85}$$

wobei $f_{ij} = b_i^e b_j^e + c_i^e c_j^e + d_i^e d_j^e$ ist. Setzt man (5.85) in das Integral (5.81) ein und verwendet

die Gleichung (B.16) für die Integration, erhält man:

$$
\begin{aligned}
B^e_{11} &= \frac{\varepsilon^e_r \left(l^e_1\right)^2}{360V^e}\left(f_{22} - f_{12} + f_{11}\right) \\
B^e_{12} = B^e_{21} &= \frac{\varepsilon^e_r l^e_1 l^e_2}{720V^e}\left(2f_{23} - f_{21} - f_{13} + f_{11}\right) \\
B^e_{13} = B^e_{31} &= \frac{\varepsilon^e_r l^e_1 l^e_3}{720V^e}\left(2f_{24} - f_{21} - f_{14} + f_{11}\right) \\
B^e_{14} = B^e_{41} &= \frac{\varepsilon^e_r l^e_1 l^e_4}{720V^e}\left(f_{23} - f_{22} - 2f_{13} + f_{12}\right) \\
B^e_{15} = B^e_{51} &= \frac{\varepsilon^e_r l^e_1 l^e_5}{720V^e}\left(f_{22} - f_{24} - f_{12} + 2f_{14}\right) \\
B^e_{16} = B^e_{61} &= \frac{\varepsilon^e_r l^e_1 l^e_6}{720V^e}\left(f_{24} - f_{23} - f_{14} + f_{13}\right) \\
B^e_{22} &= \frac{\varepsilon^e_r \left(l^e_2\right)^2}{360V^e}\left(f_{33} - f_{13} + f_{11}\right) \\
B^e_{23} = B^e_{32} &= \frac{\varepsilon^e_r l^e_2 l^e_3}{720V^e}\left(2f_{34} - f_{13} - f_{14} + f_{11}\right) \\
B^e_{24} = B^e_{42} &= \frac{\varepsilon^e_r l^e_2 l^e_4}{720V^e}\left(f_{33} - f_{23} - f_{13} + 2f_{12}\right) \\
B^e_{25} = B^e_{52} &= \frac{\varepsilon^e_r l^e_2 l^e_5}{720V^e}\left(f_{23} - f_{34} - f_{12} + f_{14}\right) \\
B^e_{26} = B^e_{62} &= \frac{\varepsilon^e_r l^e_2 l^e_6}{720V^e}\left(f_{34} - f_{33} - 2f_{14} + f_{34}\right) \\
B^e_{33} &= \frac{\varepsilon^e_r \left(l^e_3\right)^2}{360V^e}\left(f_{44} - f_{14} + f_{11}\right) \\
B^e_{34} = B^e_{43} &= \frac{\varepsilon^e_r l^e_3 l^e_4}{720V^e}\left(f_{34} - f_{24} - f_{13} + f_{12}\right) \\
B^e_{35} = B^e_{53} &= \frac{\varepsilon^e_r l^e_3 l^e_5}{720V^e}\left(f_{24} - f_{44} - 2f_{12} + f_{14}\right) \\
B^e_{36} = B^e_{63} &= \frac{\varepsilon^e_r l^e_3 l^e_6}{720V^e}\left(f_{44} - f_{34} - f_{14} + 2f_{13}\right) \\
B^e_{44} &= \frac{\varepsilon^e_r \left(l^e_4\right)^2}{360V^e}\left(f_{33} - f_{23} + f_{22}\right) \\
B^e_{45} = B^e_{54} &= \frac{\varepsilon^e_r l^e_4 l^e_5}{720V^e}\left(f_{23} - 2f_{34} - f_{22} + f_{24}\right) \\
B^e_{46} = B^e_{64} &= \frac{\varepsilon^e_r l^e_4 l^e_6}{720V^e}\left(f_{34} - f_{33} - 2f_{24} + f_{23}\right)
\end{aligned}
$$

$$
\begin{aligned}
B^e_{55} &= \frac{\varepsilon^e_r \left(l^e_5\right)^2}{360V^e}\left(f_{22} - f_{24} + f_{44}\right) \\
B^e_{56} = B^e_{65} &= \frac{\varepsilon^e_r l^e_5 l^e_6}{720V^e}\left(f_{24} - 2f_{24} - f_{44} + f_{34}\right) \\
B^e_{66} &= \frac{\varepsilon^e_r \left(l^e_6\right)^2}{360V^e}\left(f_{44} - f_{34} + f_{33}\right)
\end{aligned}
$$

Damit kann das Volumenintegral (5.76) vollständig in jedem Element e berechnet werden.

5.3.2 Das Randintegral

Mit dem Randintegral F_R^e (5.77) werden die inhomogenen Cauchyschen Randbedingungen auf S_1 (5.32) erfüllt. Zerlegt man das Integral (5.77) in zwei Anteile,

$$\iint\limits_{S_1} \frac{\gamma_E}{2} \left(\hat{n} \times \vec{E}^e\right) \cdot \left(\hat{n} \times \vec{E}^e\right) \, dS + \iint\limits_{S_1} \vec{E}^e \cdot \vec{U} \, dS \tag{5.86}$$

so kann man im ersten Integral noch die ggf. auftretenden homogenen Cauchyschen Randbedingungen (5.55) auf dem Rand S_2 mit einarbeiten:

$$\int\limits_{S_1+S_2} \int \frac{\gamma_E}{2} \left(\hat{n} \times \vec{E}^e\right) \cdot \left(\hat{n} \times \vec{E}^e\right) \, dS + \iint\limits_{S_1} \vec{E}^e \cdot \vec{U} \, dS \tag{5.87}$$

Mit der Diskretisierung des Volumens ist auch die Oberfläche diskretisiert worden. In diesem Fall mit Dreiecken. In diesen Dreiecken sind nun die beiden Integrale von (5.87) zu berechnen.

An dieser Stelle wird oft die Frage aufgeworfen, ob für die Dreiecke die vektoriellen Ansatzfunktionen (4.30) der 2D-Elemente verwendet werden oder die des Tetraeders. Die Antwort wird durch die beiden Kriterien schnell deutlich:

Konformität von E_i: Nicht nur die Diskretisierung, sondern auch die Koeffizienten E_i^e müssen konform sein. Alle Integralanteile sind Funktionen von E_i^e und müssen später zusammengefaßt werden, damit eine Lösung für alle globalen E_i gefunden wird.

Konformität von $\vec{N}_i$: Die Koeffizienten E_i stehen aber im direkten Zusammenhang mit den Vektoransatzfunktion $\vec{N}_i$, deshalb sind an dieser Stelle auch die vektoriellen 3D-Ansatzfunktion (4.40) bis (4.45) zu verwenden.

Die 3D-Ansatzfunktionen besitzen jedoch eine Normalkomponente. Sie wird durch die Kanten im Volumen hervorgerufen. Da das Volumen aber nicht mehr zum Rand gehört, dürfen sie zum Randintegral nicht beitragen. Aus diesem Grund wird die tangentiale Komponente vom Feld des Tetraeders im Dreieck als Ansatz verwendet:

$$\hat{n} \times \vec{E}^S = \sum_{i=1}^{6} \vec{S}_i^S E_i^S = \left\{E^S\right\}^T \left\{\vec{S}^S\right\} = \left\{\vec{S}^S\right\} \left\{E^S\right\} \tag{5.88}$$

Mit $\vec{S}_i^S = \hat{n} \times \vec{N}_i$, wobei $\vec{N}_i$ die Ansatzfunktionen $\vec{N}_i^e$ des Tetraeders sind, unter der Voraussetzung, daß die Kanten i zum Oberflächendreieck S gehören. Sonst ist die Funktion $\vec{N}_i = \vec{0}$. Mit diesem Ansatz wird die Konformität von E_i und $\vec{N}_i$ erfüllt. Mit dieser Wahl wird

$$E_i^s \equiv E_i^e \tag{5.89}$$

und damit

$$\left\{E^S\right\} \equiv \left\{E^e\right\} \tag{5.90}$$

Setzt man den Ansatz (5.88) in das Randintegral (5.77) ein, ergibt sich:

$$F^S_{R_{(S_1+S_2)}} = \frac{1}{2}\{E^e\}^T\left[C^S\right]\{E^e\} \tag{5.91}$$

$$F^S_{R_{S_1}} = -\{E^e\}^T\left\{c^S\right\} \tag{5.92}$$

$$F^S_R = F^S_{R_{(S_1+S_2)}} + F^S_{R_{S_1}} \tag{5.93}$$

mit den Matrixelementen

$$C^S_{ij} = \iint\limits_{S^S} \gamma\, \vec{S}^S_i \cdot \vec{S}^S_j \, dS \tag{5.94}$$

der 6×6 Matrix C^S und den Vektorelementen

$$c^S_i = \iint\limits_{S^S} \vec{S}^S_i \cdot \left(\vec{U} \times \hat{n}\right) dS \tag{5.95}$$

des Vektors c^S mit 6 Komponenten, wobei 3 immer[3] identisch Null sind.

Im Programm IFEP besteht an dieser Stelle die Einschränkung, daß die Normale $\hat{n}$ immer in z-Richtung zeigt. Damit vereinfacht sich viel.

Mit den Transformationen (B.10) bis (B.13) von Seite 123 erhält man mit $\hat{n} = -\vec{e}_z$

$$\vec{S}_i = -\vec{e}_z \times \vec{N}_i = \frac{l_i}{6V}\left[\left(\xi_{i_1}c_{i_2} - \xi_{i_2}c_{i_1}\right)\vec{e}_{x-}\left(\xi_{i_1}b_{i_2} - \xi_{i_2}b_{i_1}\right)\vec{e}_y\right] \tag{5.96}$$

und es folgt damit:

$$\vec{S}_i \cdot \vec{S}_j = \frac{l_i l_j}{(6V)^2}\left[\xi_{i_1}\xi_{j_1}\, f_{i_1 j_2} - \xi_{i_1}\xi_{j_2}\, f_{i_2 j_1} - \xi_{i_2}\xi_{j_1}\, f_{i_1 j_2} + \xi_{i_2}\xi_{j_2}\, f_{i_2 j_2}\right] \tag{5.97}$$

wobei $f_{ij} = b_i b_j + c_i c_j$ ist. Die Koeffizienten b_i bzw. b_j und c_i bzw. c_j berechnen sich entsprechend den Transformationen (B.10) bis (B.13). Die Werte für i_1 bzw. i_2 und j_1 bzw. j_2 entnimmt man den Kantenzuordnungen (4.32) bis (4.37) auf Seite 40. Die Integration über

[3] Unter der Bedingung, daß nur ein Dreieck des Tetraeders zum Rand S gehört.

das Dreieck wird dann mit (B.7) durchgeführt und man erhält für C^e_{ij} :

$$C^S_{11} = \frac{\gamma A \left(l^e_1\right)^2}{216V} \left(f_{22} - f_{12} + f_{11}\right) \tag{5.98}$$

$$C^S_{12} = C^S_{21} = \frac{\gamma A l^e_1 l^e_2}{432V^e} \left(2f_{23} - f_{21} - f_{13} + f_{11}\right) \tag{5.99}$$

$$C^S_{13} = C^S_{31} = \frac{\gamma A l^e_1 l^e_3}{432V^e} \left(2f_{24} - f_{21} - f_{14} + f_{11}\right) \tag{5.100}$$

$$C^S_{14} = C^S_{41} = \frac{\gamma A l^e_1 l^e_4}{432V^e} \left(f_{23} - f_{22} - 2f_{13} + f_{12}\right) \tag{5.101}$$

$$C^S_{15} = C^S_{51} = \frac{\gamma A l^e_1 l^e_5}{432V^e} \left(f_{22} - f_{24} - f_{12} + 2f_{14}\right) \tag{5.102}$$

$$C^S_{16} = C^S_{61} = \frac{\gamma A l^e_1 l^e_6}{432V^e} \left(f_{24} - f_{23} - f_{14} + f_{13}\right) \tag{5.103}$$

$$C^S_{22} = \frac{\gamma A \left(l^e_2\right)^2}{216V^e} \left(f_{33} - f_{13} + f_{11}\right) \tag{5.104}$$

$$C^S_{23} = C^S_{32} = \frac{\gamma A l^e_2 l^e_3}{432V^e} \left(2f_{34} - f_{13} - f_{14} + f_{11}\right) \tag{5.105}$$

$$C^S_{24} = C^S_{42} = \frac{\gamma A l^e_2 l^e_4}{432V^e} \left(f_{33} - f_{23} - f_{13} + 2f_{12}\right) \tag{5.106}$$

$$C^S_{25} = C^S_{52} = \frac{\gamma A l^e_2 l^e_5}{432V^e} \left(f_{23} - f_{34} - f_{12} + f_{14}\right) \tag{5.107}$$

$$C^S_{26} = C^S_{62} = \frac{\gamma A l^e_2 l^e_6}{432V^e} \left(f_{34} - f_{33} - 2f_{14} + f_{13}\right) \tag{5.108}$$

$$C^S_{33} = \frac{\gamma A \left(l^e_3\right)^2}{216V^e} \left(f_{44} - f_{14} + f_{11}\right) \tag{5.109}$$

$$C^S_{34} = C^S_{43} = \frac{\gamma A l^e_3 l^e_4}{432V^e} \left(f_{34} - f_{24} - f_{13} + f_{12}\right) \tag{5.110}$$

$$C^S_{35} = C^S_{53} = \frac{\gamma A l^e_3 l^e_5}{432V^e} \left(f_{24} - f_{44} - 2f_{12} + f_{14}\right) \tag{5.111}$$

$$C^S_{36} = C^S_{63} = \frac{\gamma A l^e_3 l^e_6}{432V^e} \left(f_{44} - f_{34} - f_{14} + 2f_{13}\right) \tag{5.112}$$

$$C^S_{44} = \frac{\gamma A \left(l^e_4\right)^2}{216V^e} \left(f_{33} - f_{23} + f_{22}\right) \tag{5.113}$$

$$C^S_{45} = C^S_{54} = \frac{\gamma A l^e_4 l^e_5}{432V^e} \left(f_{23} - 2f_{34} - f_{22} + f_{24}\right) \tag{5.114}$$

$$C^S_{46} = C^S_{64} = \frac{\gamma A l^e_4 l^e_6}{432V^e} \left(f_{34} - f_{33} - 2f_{24} + f_{23}\right) \tag{5.115}$$

$$C^S_{55} = \frac{\gamma A \left(l^e_5\right)^2}{216V^e} \left(f_{22} - f_{24} + f_{44}\right) \tag{5.116}$$

$$C^S_{56} = C^S_{65} = \frac{\gamma A l^e_5 l^e_6}{432V^e} \left(f_{24} - 2f_{24} - f_{44} + f_{34}\right) \tag{5.117}$$

$$C^S_{66} = \frac{\gamma A \left(l^e_6\right)^2}{216V^e} \left(f_{44} - f_{34} + f_{33}\right) \tag{5.118}$$

Aber nur unter der Bedingung, daß i und j Kanten des Dreiecks sind, da sonst ja S_i oder S_j Null ist. Es gibt also immer 9 Nicht-Null-Einträge in C^S. Damit ist (5.91) bis auf den Faktor

γ bekannt, der dann von dem speziellen Problem abhängt (siehe (5.31) auf S.51 oder (5.41) auf S.52).

Für die Einträge c_i^S wird $\vec{S}_i$ und $\vec{U} \times \hat{n}$ benötigt. $\vec{S}_i$ kann wieder mit (5.96) berechnet werden. Die Vektorfunktion $\vec{U}$ ist gegeben, aber problemabhängig. Deshalb wird das Integral (5.95) mit Hilfe der Gauss-Legendre-Quadratur C.1 von Seite 125 numerisch berechnet.

5.3.3 Das Stromintegral

Da für die Stromdichte nach dem Ohmschen Gesetz gilt:

$$\vec{J} = \kappa \cdot \vec{E} \tag{5.119}$$

müßte es im Berechnungsgebiet Ω eine elektrische Leitfähigkeit ungleich Null geben. In Wellenleitern oder Resonatoren befinden sich aber entweder Materialien mit sehr kleiner ($\kappa \approx 0$) oder mit *unendlicher* ($\kappa \approx \infty$) elektrischer Leitfähigkeit. Für ($\kappa \approx 0$) verschwindet das Stromintegral. Im Fall ($\kappa \approx \infty$) stellt der Rand dieses Teilgebiets eine elektrische Wand dar und kann durch eine Dirichletsche Randbedingung abgeschlossen werden. Aus diesen Gründen wird das Stromintegral (5.78) in IFEP nicht berücksichtigt. Es ist jedoch möglich, dies später im Programm noch einzuführen.

5.3.4 Das Elementgleichungssystem

Setzt man die Anteile (5.76),(5.77) und (5.78)[4] des Funktionals $F^e(E^e)$ wieder zusammen, erhält man:

$$\begin{aligned} F^e(E^e) &= F_V^e + F_R^e + F_S^e && (5.120)\\ &= \frac{1}{2}\left(\{E^e\}^T [A^e] \{E^e\} - k_0^2 \{E^e\}^T [B^e] \{E^e\}\right) \\ &\quad + \frac{1}{2} \{E^e\}^T \left[C^S\right] \{E^e\} - \{E^e\}^T \{c^S\} \\ &\quad + 0 && (5.121) \end{aligned}$$

Um das Funktional $F^e(E^e)$ zu minimieren, wird es nach den Koeffizienten E_i^e partiell abgeleitet:

$$\begin{aligned} \delta F^e(\vec{E}^e) &= [A^e] \{E^e\} - k_0^2 [B^e] \{E^e\} + \left[C^S\right] \{E^e\} - \{c^S\} && (5.122)\\ &= \left([A^e] - k_0^2 [B^e] + \left[C^S\right]\right) \{E^e\} - \{c^S\} && (5.123) \end{aligned}$$

und dann gleich Null gesetzt:

$$\begin{aligned} \left([A^e] - k_0^2 [B^e] + \left[C^S\right]\right) \{E^e\} - \{c^S\} &\stackrel{!}{=} \{0\} && (5.124)\\ \left([A^e] - k_0^2 [B^e] + \left[C^S\right]\right) \{E^e\} &= \{c^S\} && (5.125) \end{aligned}$$

Mit

$$[K^e] = [A^e] - k_0^2 [B^e] + \left[C^S\right] \tag{5.126}$$

[4] Dieser Anteil ist hier Null. Wenn er berücksichtigt werden sollte, müßte er an dieser Stelle weiter betrachtet werden.

erhält man die Form $Ax = b$ eines linearen Gleichungssystem für ein Element e:

$$[K^e]\{E^e\} = \{c^S\} \tag{5.127}$$

$$\begin{bmatrix} k^e_{11} & k^e_{12} & k^e_{13} & k^e_{14} & k^e_{15} & k^e_{16} \\ k^e_{21} & k^e_{22} & k^e_{23} & k^e_{24} & k^e_{25} & k^e_{26} \\ k^e_{31} & k^e_{32} & k^e_{33} & k^e_{34} & k^e_{35} & k^e_{36} \\ k^e_{41} & k^e_{42} & k^e_{43} & k^e_{44} & k^e_{45} & k^e_{46} \\ k^e_{51} & k^e_{52} & k^e_{53} & k^e_{54} & k^e_{55} & k^e_{56} \\ k^e_{61} & k^e_{62} & k^e_{63} & k^e_{64} & k^e_{65} & k^e_{66} \end{bmatrix} \begin{Bmatrix} E^e_1 \\ E^e_2 \\ E^e_3 \\ E^e_4 \\ E^e_5 \\ E^e_6 \end{Bmatrix} = \begin{Bmatrix} c^S_1 \\ 0 \\ 0 \\ c^S_4 \\ 0 \\ c^S_6 \end{Bmatrix} \tag{5.128}$$

5.3.5 Aufstellen des Gesamtgleichungssystems

Mit (5.128) hat man nun für jedes Element ein eigenes Gleichungssystem. Man möchte jedoch alle Gleichungssysteme zusammenfassen zu einem Gesamtgleichungssystem. Jedes Element kann eine eigene Materialkonstante ε_r und μ_r haben. Dann sind an den Elementgrenzen die Kontinuitätsbedingungen (5.44) bis (5.47) von Seite 52 einzuhalten.

An dieser Stelle kommt ein besonderer Vorteil der Kantenelemente zum tragen. Die Unbekannten E^e_i repräsentieren gerade **nur** die tangentiale elektrische Feldstärke und nicht auch die normale elektrische Feldstärke, wie bei Knotenelementen. Wenn sich nun zwei Tetraeder eine Kante teilen, haben dann beide die Kontinuitätsbedingung (5.44) automatisch erfüllt.

Wie alle Elementgleichungen zu einem Gesamtgleichungssystem zusammengefaßt werden können, kann man in Abb. 5.2 sehen.

Was bei den einzelnen Schritten besonders zu beachten ist, kann in den Abbildungen 5.3 bis 5.6 genauer betrachtet werden.

Da die Matrix nur dünn besetzt sein wird, ist das Vorgehen in Abb. 5.4 sehr verschwenderisch mit dem Speicher. In IFEP wird an dieser Stelle eine Matrix mit volldynamischer Speicherverwaltung verwendet, die nur Nicht-Null-Einträge in der unteren Dreiecksmatrix mit der Diagonalen abspeichert.

Was beim Einarbeiten des Elementgleichungssystems in die Gesamtmatrix passiert, kann man in Gleichung (5.129) sehen. Hier ist ein Netz aus zwei Tetraedern verwendet worden, wobei die beiden Tetraeder sich genau ein Dreieck teilen. Es sind die Kanten E^g_4 bis E^g_6 die von beiden Tetraedern benutzt werden. Dementsprechend liefern auch beide Elementematrizen $[K^1]$ und $[K^2]$ Beiträge zu den Einträgen (siehe Runde Klammern in der Matrix). Im Element 2 ist die Kante 3 anders orientiert als bei den globalen Kanten. Aus diesem Grund sind in der lokalen Zeile und Spalte von k^2_{33} die Vorzeichen vertauscht. Bei k^2_{33} selbst zweimal, wodurch der Eintrag wieder positiv wird. Im Spaltenvektor $\{c^2\}$ erhält auch der Eintrag c^2_3 einen Vorzeichenwechsel.

Grundsätzlicher Ablauf: Aufstellen des Gesamtgleichungssystems

Vorbereiten des Tetraedernetzes
Speicher für Gesamtgleichungssystem holen
Schleife über alle Tetraeder
Aufstellen des Elementgleichungssystems für das Tetraeder unter Berücksichtigung der lokalen Kantennumerierung
Orientieren der Matrixeinträge nach der globalen Kantenrichtung
Einarbeiten des Elementgleichungssystems ins Gesamtgleichungssystem, unter Berücksichtigung der globalen Numerierung
Einarbeiten der Dirichletschen Randbedingungen

Abbildung 5.2: Aufstellen des Gesamtgleichungssystems

Schritt: Vorbereiten des Tetraedernetzes

Orientieren aller Tetraeder in der Form, daß alle ein positives Volumen haben
Neu-Numerierung der Kanten bzw. Knoten, um später eine geringe Bandbreite der Matrix zu erhalten
Aufbauen einer Kantenliste aus den Knoten der Tetraeder mit der Bedingung, daß jede Kante nur einmal vorkommt
Aufbauen einer Datenstruktur, die die Tetraeder durch die Kantenliste beschreibt

Abbildung 5.3: Vorbereiten des Netzes

Schritt: Speicher für Gesamtgleichungssystem holen

Bestimmung der Größe des Gleichungssystems durch die Anzahl der Kanten im Netz
Speicher für eine $n \times n$ Matrix holen
Speicher für E_i^g-Vektor der Länge n holen
Speicher für RHS c_i^g-Vektor der Länge n holen

Abbildung 5.4: Speicher für das Gesamtgleichungssystem holen

Schritt: Orientieren der Elementmatrixeinträge nach der globalen Kantenrichtung

Wenn (globale Richtung $\neq$ lokale Richtung) dann	sonst
Vorzeichen aller Matrixeinträge in Zeile wechseln	nichts
Vorzeichen aller Matrixeinträge in Spalte wechseln	

Abbildung 5.5: Orientieren der Elementmatrixeinträge nach der globalen Kantenrichtung

Schritt: Einarbeiten des Elementgleichungssystems ins Gesamtgleichungssystem unter Berücksichtigung der globalen Numerierung

Für k_{ij}^e die globale Zeile i^g von i und die globale Spalte j^g von j bestimmen
$k_{i^g j^g}^g \longleftarrow k_{i^g j^g}^g + k_{ij}^e$

Abbildung 5.6: Einarbeiten des Elementgleichungssystems in die Gesamtmatrix

$$\begin{bmatrix} k_{11}^1 & k_{12}^1 & k_{13}^1 & k_{14}^1 & k_{15}^1 & k_{16}^1 & 0 & 0 & 0 \\ k_{21}^1 & k_{22}^1 & k_{23}^1 & k_{24}^1 & k_{25}^1 & k_{26}^1 & 0 & 0 & 0 \\ k_{31}^1 & k_{32}^1 & k_{33}^1 & k_{34}^1 & k_{35}^1 & k_{36}^1 & 0 & 0 & 0 \\ k_{41}^1 & k_{42}^1 & k_{43}^1 & \left(k_{44}^1 + k_{11}^2\right) & \left(k_{45}^1 + k_{12}^2\right) & \left(k_{46}^1 - k_{13}^2\right) & k_{14}^2 & k_{15}^2 & k_{16}^2 \\ k_{51}^1 & k_{52}^1 & k_{53}^1 & \left(k_{54}^1 + k_{21}^2\right) & \left(k_{55}^1 + k_{22}^2\right) & \left(k_{56}^1 - k_{23}^2\right) & k_{24}^2 & k_{25}^2 & k_{26}^2 \\ k_{61}^1 & k_{62}^1 & k_{63}^1 & \left(k_{64}^1 - k_{31}^2\right) & \left(k_{65}^1 - k_{32}^2\right) & \left(k_{66}^1 + k_{33}^2\right) & -k_{34}^2 & -k_{35}^2 & -k_{36}^2 \\ 0 & 0 & 0 & k_{41}^2 & k_{42}^2 & -k_{43}^2 & k_{44}^2 & k_{45}^2 & k_{46}^2 \\ 0 & 0 & 0 & k_{51}^2 & k_{52}^2 & -k_{53}^2 & k_{54}^2 & k_{55}^2 & k_{56}^2 \\ 0 & 0 & 0 & k_{61}^2 & k_{62}^2 & -k_{63}^2 & k_{64}^2 & k_{65}^2 & k_{66}^2 \end{bmatrix} \begin{Bmatrix} E_1^g \\ E_2^g \\ E_3^g \\ E_4^g \\ E_5^g \\ E_6^g \\ E_7^g \\ E_8^g \\ E_9^g \end{Bmatrix} = \begin{Bmatrix} c_1^1 \\ 0 \\ 0 \\ c_3^1 \\ 0 \\ c_6^1 - c_3^2 \\ 0 \\ c_5^2 \\ c_6^2 \end{Bmatrix} \tag{5.129}$$

Sind alle Elementgleichungssystem eingearbeitet, müssen nur noch die Dirichletschen Randbedingungen erzwungen werden. Als Beispiel soll auf Kante 7 der tangentialen Wert E_7^g auf $10,5$ gezwungen werden. Dafür wird k_{77} auf dem Wert 1 gesetzt, **alle** anderen Einträge in der Zeile auf Null und c_7 auf den gewünschten Wert, hier $10,5$ (siehe Gleichung (5.130)). Dadurch ist die Matrix $[K]$ aber unsymmetrisch geworden.

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} & k_{15} & k_{16} & 0 & 0 & 0 \\ k_{21} & k_{22} & k_{23} & k_{24} & k_{25} & k_{26} & 0 & 0 & 0 \\ k_{31} & k_{32} & k_{33} & k_{34} & k_{35} & k_{36} & 0 & 0 & 0 \\ k_{41} & k_{42} & k_{43} & k_{44} & k_{45} & k_{46} & k_{47} & k_{48} & k_{49} \\ k_{51} & k_{52} & k_{53} & k_{54} & k_{55} & k_{56} & k_{57} & k_{58} & k_{59} \\ k_{61} & k_{62} & k_{63} & k_{64} & k_{65} & k_{66} & k_{67} & k_{68} & k_{69} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & k_{84} & k_{85} & k_{86} & k_{87} & k_{88} & k_{89} \\ 0 & 0 & 0 & k_{94} & k_{95} & k_{96} & k_{97} & k_{98} & k_{99} \end{bmatrix} \begin{Bmatrix} E_1^g \\ E_2^g \\ E_3^g \\ E_4^g \\ E_5^g \\ E_6^g \\ E_7^g \\ E_8^g \\ E_9^g \end{Bmatrix} = \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ 10,5 \\ c_8 \\ c_9 \end{Bmatrix} \tag{5.130}$$

Man kann die Matrix jedoch wieder symmetrisch machen. Da E_7^g jetzt bekannt ist, kann man

die Einträge k_{i7} mit $i \neq 7$ mit $E_7^g = 10,5$ multiplizieren und in den Vektor $\{c\}$ einarbeiten:

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} & k_{15} & k_{16} & 0 & 0 & 0 \\ k_{21} & k_{22} & k_{23} & k_{24} & k_{25} & k_{26} & 0 & 0 & 0 \\ k_{31} & k_{32} & k_{33} & k_{34} & k_{35} & k_{36} & 0 & 0 & 0 \\ k_{41} & k_{42} & k_{43} & k_{44} & k_{45} & k_{46} & 0 & k_{48} & k_{49} \\ k_{51} & k_{52} & k_{53} & k_{54} & k_{55} & k_{56} & 0 & k_{58} & k_{59} \\ k_{61} & k_{62} & k_{63} & k_{64} & k_{65} & k_{66} & 0 & k_{68} & k_{69} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & k_{84} & k_{85} & k_{86} & 0 & k_{88} & k_{89} \\ 0 & 0 & 0 & k_{94} & k_{95} & k_{96} & 0 & k_{98} & k_{99} \end{bmatrix} \begin{Bmatrix} E_1^g \\ E_2^g \\ E_3^g \\ E_4^g \\ E_5^g \\ E_6^g \\ E_7^g \\ E_8^g \\ E_9^g \end{Bmatrix} = \begin{Bmatrix} c_1 - 0 \cdot 10,5 \\ c_2 - 0 \cdot 10,5 \\ c_3 - 0 \cdot 10,5 \\ c_4 - k_{47} \cdot 10,5 \\ c_5 - k_{57} \cdot 10,5 \\ c_6 - k_{67} \cdot 10,5 \\ 10,5 \\ c_8 - k_{87} \cdot 10,5 \\ c_9 - k_{97} \cdot 10,5 \end{Bmatrix} \tag{5.131}$$

Nun hat man ein Gleichungssystem, dessen Lösung unter Berücksichtigung der gegebenen Randbedingungen die Lösung für das gesuchte Randwertproblem ist.

Kapitel 6

Lösen des linearen Gleichungssystems

Es gibt sehr verschiedene Verfahren, um lineare Gleichungssystem zu lösen [22] [43] [44] [45]. Dabei hat man immer zu beachten: Die Verfahren sollen nicht viel Speicher verbrauchen, aber trotzdem schnell sein.

Grundsätzlich unterscheidet man zwischen den **direkten Lösern** und den **iterativen Lösern.** Direkte Löser beruhen auf dem Eliminationsverfahren von Gauß beziehungsweise Cholesky. In der FEM werden direkte Gleichungslöser mit Erfolg angewendet. Allerdings führen diese Gleichungslöser oft zu sehr großen Zwischenmatrizen und damit zu extrem hohen Anforderungen an Haupt- und Massenspeicherkapazitäten sowie an die Rechenleistung der eingesetzten Computer.

Bei der FEM sind die Matrizen in der Regel groß, jedoch nur dünn besetzt. Da bei direkten Lösern während der Berechnung neue Matrixeinträg (sogenannte Fill-Ins) entstehen, ist hier die Speicherverwaltung sehr kompliziert. Zusätzlich entstehen beim Lösen Rundungsfehler, die mit der Ordnung des Gleichungssystems zunehmen.

Aus diesen Gründen sind die in IFEP vorhandenen direkten Löser[1] nur für sehr kleine Probleme anwendbar. Ihre Weiterentwicklung ist in IFEP nicht weiter verfolgt worden. Im Folgenden werden nur noch iterative Löser betrachtet.

Vor der Behandlung der iterativer Löser selbst, werden zunächst einige Maßnahmen vorgestellt, die das Lösen erst ermöglichen oder zumindest erleichtern.

6.1 Datenstrukturen zur Matrixspeicherung

Die Matrix A ist symmetrisch und dünn besetzt.

Beispiel: Bei einer $n \times n$-Matrix mit $n = 68793$ entstehen bei voller Besetzung der Matrix

$$n \cdot n = 4,732 \cdot 10^9$$

Matrixelemente. Mit ABC´s oder komplexen Materialkonstanten ist die Matrix auch komplexwertig. Dann folgt ein Speicherbedarf:

$$4,732 \cdot 10^9 \cdot 16 \cdot Byte = 72,211\, GByte$$

[1] Siehe Tabelle 6.2

nur für die Matrix. Soviel Speicher steht selten zur Verfügung. Da diese Matrix nicht voll besetzt ist, d.h. es gibt konkret nur nnz[2] $= 767739$ Nicht-Null-Einträge, ist der eigentliche Bedarf wesentlich geringer:

$$767739 \cdot 16 \cdot Byte = 11,996\, MByte$$

Nutzt man noch die Symmetrie aus und speichert nur die untere Dreiecksmatrix mit der Diagonalen, so ergibt sich ein Bedarf von:

$$\begin{aligned}\left(\frac{(nnz-n)}{2}+n\right)\cdot 16\cdot Byte &= \\ 418293\cdot 16\cdot Byte &= 6,38\, MByte\end{aligned}$$

Dieser Speicherplatz kann heute in jedem PC bereitgestellt werden.

Das Hauptziel der folgenden Datenstrukturen ist es, nur die Nicht-Null-Einträge nnz zu speichern. Trotzdem soll man effizient alle nötigen Matrix-Operationen durchführen können. Da hier viele verschiedene Arten existieren, werden hier nur die wichtigsten diskutiert und weiterführende Literatur empfohlen [22] [43] [44] [45] [46].

Alle hier vorgestellten Formate sind **statisch,** d.h. die Matrix muß fertig assembliert sein. Dies ist in IFEP mit einer volldynamischen Datenstruktur realisiert, in der natürlich nur die Nicht-Null-Einträge gespeichert werden. Zusätzlich wird auch nur die untere Dreiecksmatrix mit Diagonale (also die Symmetrie voll ausgenutzt) aufgebaut. So wird die Gesamtmatrix nie voll aufgebaut. Dies ist die beste Lösung, die eine optimale Nutzung des Arbeitsspeichers ermöglicht. Jedoch ist die Implementierung mit schnellen Rechenzeiten sehr anspruchsvoll, da beim Symmetrisieren nach dem Einarbeiten der Dirichletschen Randbedingungen eigentlich die Gesamtmatrix benötigt wird. Durch die dynamische Datenstruktur mit Zeilenlisten kann man nicht jederzeit auf jedes Element zugreifen, ohne die Rechenzeit schnell in die Höhe zu treiben. In IFEP ist hier eine schnelle Realisierung entwickelt worden, so daß die Dimension der Matrix und die Anzahl der Nicht-Null-Einträge feststeht, wenn die Matrix über eine temporäre Datei[3] in das statische Format umgespeichert wird.

6.1.1 Datenstrukturen für unsymmetrische Matrizen

Eine weit verbreitet Art, eine dünn besetzte Matrix zu speichern, ist das **CSR**-Format (engl. ***C****ompressed* ***S****parse* ***R****ow)* und das **CSC**-Format (engl. ***C****ompressed* ***S****parse* ***C****olumn)* [45, 84ff]. Sie unterscheiden sich nur durch die zeilenweise (bei CSR) beziehungsweise spaltenweise (bei CSC) Speicherung.

Der Vorteil der CSR- bzw. CSC-Datenstruktur ist, daß sich sowohl symmetrische wie auch unsymmetrische Matrizen damit speichern lassen. Man kann eine Matrix auch leicht transponieren, indem man z.B. eine CSR-Matrix als CSC-Matrix verwendet oder umgekehrt. Nachteil ist aber der komplizierte Zugriff auf die Diagonalelemente. Eine Abhilfe schafft da das **MSR**-Format (*engl.* ***M****odified* ***S****parse* ***R****ow*). Hier wird die Diagonale als erstes in das Array AA gespeichert, zusätzlich werden die Arrays JA und IA zu einem zusammengefaßt. Die genaue Beschreibung ist in [45, S.85] und [46, S.78] nachzulesen. Der Nachteil von diesem Format liegt darin, daß man die Diagonale **immer** separat behandeln muß und komplizierter auf die Elemente zugreifen muß.

[2] nnz steht für **N**umber of **N**on **Z**ero, also die Anzahl der nicht Nullelemente in der Matrix A

[3] Dies verhindert einen kurzzeitigen hohen Speicherbedarf beim Umspeichern der Matrix.

6.1.1.1 Datenstruktur vom CSR-Format

Dieses Format besteht aus 3 statischen Arrays. Einem Array aus den Nicht-Null-Matrixeinträgen und zwei integer Arrays zur Speicherung der Position der Matrixeinträge . Sei eine unsymmetrische Matrix A gegeben:

$$A = \begin{bmatrix} a_{11} & & & a_{14} & \\ a_{21} & a_{22} & & & \\ & & a_{33} & & \\ & a_{42} & & a_{44} & \\ a_{51} & & a_{53} & & a_{55} \end{bmatrix} \tag{6.1}$$

dann ergeben sich für die CSR-Speicherung folgende Daten (s. Abb. 6.1):

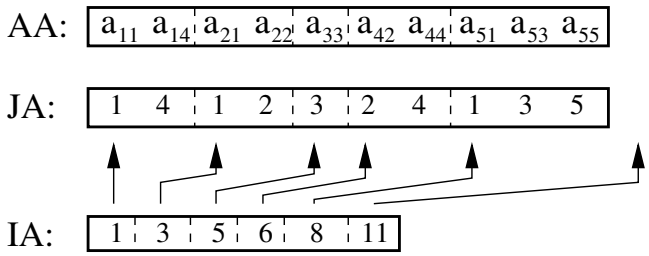


Abbildung 6.1: CSR-Speicherformat einer unsymmetrischen Matrix

Das Daten-Array AA beinhalten die Matrixelemente, und zwar Zeile für Zeile. Die Elemente können dabei reell oder komplex sein. Im Array JA ist die Spalte eines jeden Eintrags festgehalten. Mit dem Array IA wird die Position des ersten Elements einer Zeile in den Arrays AA und JA festgehalten.

6.1.2 Datenstrukturen für symmetrische Matrizen

Die CSR-, CSC- und MSR-Formate bieten auch alle die Möglichkeit, nur den symmetrischen Teil einer Matrix abzuspeichern. Jedoch bleiben die genannten Nachteile. Ein Format, mit dem man so einfach arbeiten kann wie mit dem CSR- bzw. CSC-Format, und mit dem man auf die Diagonalelemente direkt zugreifen kann, ist das **AWS**-Format [22, S.269]. Es unterscheidet sich nur geringfügig vom CSR-Format. Dabei speichert man nur Nicht-Null-Einträge in und unterhalb der Diagonalen wie im CSR-Format. Im Array IA speichert man nun statt des Zeileanfangs den **Index jeden Zeilenendes.** Damit ist der direkte Zugriff auf die Diagonale möglich und der Zugriff auf alle Matrixeinträge bleibt trotzdem einfach. Durch eine Index-Verschiebung[4] im Array IA läßt es sich auch einfach ins CSR- bzw. CSC-Format umwandeln und somit auch leicht transponieren.

6.1.2.1 Datenstruktur des AWS-Formats

Dieses Format besteht wie das CSR-Format aus 3 statischen Arrays. Ein Array aus den Nicht-Null-Matrixeinträgen und zwei integer Arrays zur Speicherung der Position der Matrixeinträge.

[4] Dabei ist auf die Länge des Arrays IA zu achten. Im CSR-Format ist es um eins länger als im AWS-Format.

Sei eine symmetrische Matrix A gegeben:

$$A = \begin{bmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ & & a_{33} & & \\ & a_{42} & & a_{44} & \\ a_{51} & & a_{53} & & a_{55} \end{bmatrix} \tag{6.2}$$

dann ergeben sich für die CSR-Speicherung folgende Daten (s. Abb. 6.2):

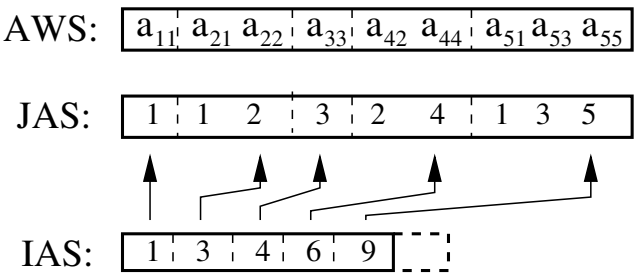


Abbildung 6.2: AWS-Speicherformat einer symmetrischen Matrix

Das Daten-Array AWS beinhaltet die Matrixelemente, und zwar Zeile für Zeile. Die Elemente können dabei reell oder komplex sein. Im Array JAS ist die Spalte eines jeden Eintrags von AWS festgehalten. Mit dem Array IAS wird die Position **des letzten Elements** einer Zeile in den Arrays AWS und JAS festgehalten.

6.2 Bandbreite der Matrix

Unter der Bandbreite m einer Matrix A versteht man die kleinste Zahl m, so daß

$$a_{ij} = 0 \; für\, alle\, i,\, j\, mit \;\; |i - j| > m$$

gilt [22, S.160].

6.3 Optimale Numerierung der Kantenvariablen

Die Numerierung der Kanten und damit auch der Kantenvariablen hat entscheidenden Einfluß auf die Besetzungsstruktur der Matrix A [22, S.165]. Die i-te Zeile der Matrix, welche der i-ten Kantenvariablen zugeordnet ist, enthält außer dem Diagonalelement nur dann einen von Null verschiedenen Betrag a_{ij} bzw. a_{ji}, wenn die Kanten i und j von *einem gemeinsamen Element* verwendet werden. Das heißt, daß *die Elemente* durch das Element-Gleichungssystem die Einträge $a_{ij} = a_{ji} \neq 0$ mit $i \neq j$ verursachen.

Beispiel: Wird die Kante 1 und die Kante 1345 von einem Element verwendet, entsteht in der Matrix A bei $a_{1,1345} = a_{1345,1}$ eine Wert ungleich Null. Wenn es zum Beispiel nur 1345 Kanten gibt, hätte die Matrix A eine Bandbreite $m = 1345$.

Daraus folgt, daß für eine geringe Bandbreite der Matrix jedes Element immer eine möglichst kleine Differerenz in der Kantennumerierung haben sollte.

Nun gibt es für die Numerierung von Knotenvariablen Algorithmen, die die Bandbreite der Matrix gut reduzieren z.B. [27][28]. Bei Kantenelementen ist das Problem jedoch nicht so leicht zu lösen, da zwei Numerierungen im Netz existieren: Die Numerierung der Knoten und darauf aufbauend die Numerierung der Kanten. Das Problem ist nicht, nur die Kanten neu zu numerieren, sondern daß dadurch die Datenstuktur der Kantennumerierung auf der Knotennumerierung inkonsistent wird. Zunächst wird der Algorithmus von Cuthill-McKee beschrieben, dann werden die Probleme bei Kantenelementen diskutiert und anschließend der hier verwendete Algorithmus vorgestellt.

6.3.1 Der Algorithmus von Cuthill-McKee

Der von Cuthill und McKee vorgeschlagene Algorithmus für Knoten basiert auf graphentheoretischen Überlegungen. Hier soll auf die Graphentheorie nicht näher eingegangen werden, da der Algorithmus auch so verständlich ist.

1. Schritt: Man sucht einen Knoten, der möglichst nur von einem Tetraeder verwendet wird. Dieser Knoten ist der **Startknoten** und bekommt die Nummer 1.

2. Schritt: Man bestimmt alle benachbarten Knoten (Schicht-1) zu dem Startknoten. Anschließend bestimmt man, wieviele neue Nachbarknoten diese Schicht-1 Knoten hat und numeriert, die mit den wenigsten Nachbarn zuerst, fortlaufend durch. Da es mehrere Knoten mit der gleichen Anzahl von Nachbarn gibt, ist diese Numerierung innerhalb dieser Nachbarn willkürlich. Alle Knoten der Schicht-1 haben die *Distanz 1* zum Startknoten. Sie bilden die **Schicht-1.**

3. Schritt: Zu den Knoten der Schicht-1 bestimmt man mit aufsteigenden Nummern die noch nicht vergebenen Nachbarknoten. Sie haben die *Distanz 2* zum Startknoten. Sie bilden die **Schicht-2**. Genau wie im 2. Schritt werden sie dann durchnumeriert.

4. Schritt: Man wiederholt den 3. Schritt, bis alle Knoten neu numeriert worden sind.

Je kleiner die Schichten bleiben, um so kleiner bleibt die Bandbreite der Matrix.

6.3.2 Probleme bei Kantenelementen

Knoten sind im kartesischen Koordinatensystem eindeutig bestimmt. Kanten sind durch zwei Knoten beschrieben. Mit der zusätzlichen Bedingung, daß die Knotennummer am Kantenende 1 immer kleiner ist als am Kantenende 2, kann eine Kante auch eindeutig zugeordnet werden.

Das Problem besteht nun im schnellen Suchen einer bestimmten Kante (ihr Index) aus allen Kanten, wenn nur die Kantenenden gegeben sind. Soll mit linearer Ordnung gesucht werden, muß die Kantennumerierung auch eine systematische Struktur in den Knoten haben. Eine brauchbare Struktur ist das aufsteigende Sortieren der Kanten nach Kantenende 1 und bei gleichem Kantenende 1 mit aufsteigenden Sortieren nach Kantenende 2. In dieser Struktur kann man mit den Kantenenden jede Kante schnell finden.

Mit dieser Struktur ist bei gegebener Knotennumerierung die Kantennumerierung schon fest vorgegeben. Wenn man nun den Algorithmus von Cuthill-McKee nur auf die Kanten überträgt, muß die Struktur in den Kantenenden zerstört werden. Deshalb wird hier ein Algorithmus vorgestellt, der die Besonderheit der Kantenelemente berücksichtigt.

6.3.3 Ein effizienter Algorithmus bei Kantenelementen

Die Distanz bei Knoten wie bei Kantenelementen ist durch die Elemente selbst gegeben. Man kann diese Verwandschaft nun gezielt verwenden, indem man wie folgt vorgeht:

1. Schritt: Anwenden des Algorithmus von Cuthill-McKee auf die Knoten. Gleichzeitig baut man eine Datenstruktur auf, mit der man aus den alten Knotennummern die Neuen für jeden Knoten erhalten kann.

2. Schritt: Umsetzen aller Elemente von der alten Knotennumerierung auf die Neue.

3. Schritt: Aufbau der Kantennumerierung nach der Struktur

$\Rightarrow$ (Knotennummer von Kantenende 1) $<$ (Knotennummer von Kantenende 2)

$\Rightarrow$ Kantennumerierung aufsteigend sortiert nach Knotennummer von Kantenende 1

$\Rightarrow$ Wenn Knotennummer vom Kantenende 1 gleich, dann aufsteigend sortiert nach Knotennummer von Kantenende 2.

Als Ergebnis erhält man eine geringe Bandbreite der Matrix A, siehe Beispiel Abb. 6.3 b) auf Seite 77 mit einer Bandbreite von $m = 746$ bei einer $n \times n$-Matrix mit $n = 12318$ und nnz[5]$= 96416$. Zusätzlich existiert eine Struktur in der Knotennumerierung der Kantenenden, die eine schnelle Suche auch in sehr großen Datenmengen erlauben.

6.4 Elimination von inneren Freiheitsgraden, Kondensation

Ein Teil des Lösungsvectors x ist direkt von den anderen Lösungswerten abhängig.

$$x_1 = f(x_2) \tag{6.3}$$

Diese Unbekannten treten gewöhnlich am Rand des Gebietes auf. Unter Verwendung dieses Wissens kann man das Gleichungssystem noch Reduzieren. Im Gleichungssystem sind noch Kantenvariablen vorhanden, die nur vom eigenen Element abhängig sind. Man bezeichnet diese Kantenvariablen als *innere Freiheitsgrade* [22, Seite 186]. Aufgrund der Extremalprinzipien lassen sich die Werte der *inneren* Kantenvariablen durch die Werte der *äußeren* Kantenvariablen darstellen und deshalb in dem Gleichungssystem eliminieren, was man Kondensation

[5] nnz steht für **N**umber of **N**on **Z**ero, also die Anzahl der nicht Nullelemente in der Matrix A

nennt. Ihre Häufigkeit ist abhängig von der Anzahl und Nachbarn der Kanten auf dem Rand des Gebiets Ω und der Anzahl und Nachbarn der Kanten mit Dirichletscher Randbedingung.

Nach der Elimination der *inneren* Kantenvariablen bleibt ein reduziertes Gleichungssystem der *äußeren* Kantenvariablen übrig. Man nennt die resultierende Matrix *kondensierte Steifigkeitsmatrix* und die Rechte-Hand-Seite *kondensierten Elementvektor.*

6.4.1 Kondensation

Geht man von folgendem Gleichungssystem

$$A\,x = b \tag{6.4}$$

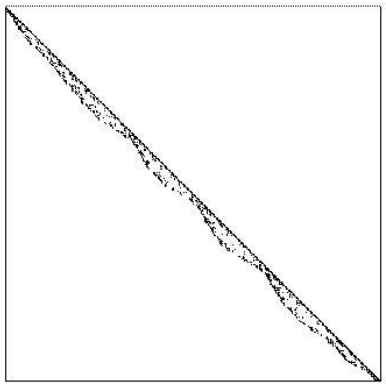
aus und liegt die Matrix A

$$A = L + D + L^T \tag{6.5}$$

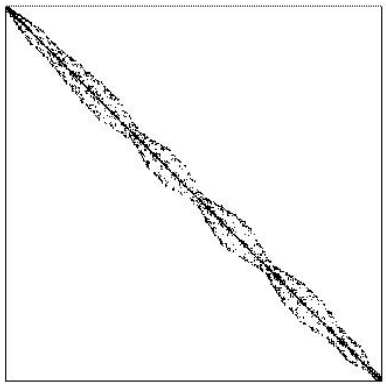
nur in symmetrischer Form A_{Sym}

$$A^{Sym} = L + D \tag{6.6}$$

vor, also nur aus unterer Dreiecksmatrix L mit Diagonalen D bestehend. So soll aus Gründen der Speichereffizenz statt der Matrix A nur noch A^{Sym} verwendet werden, wie in Abbildung 6.3 a) zu sehen ist. Mit dieser Matrix wird durch die Suche nach den unabhängigen Kantenva-



(a) Matrix A^{Sym}



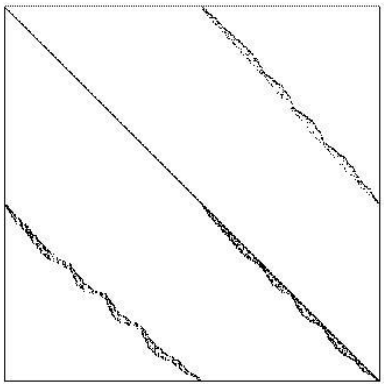
(b) Gesamte Matrix A

Abbildung 6.3: Matrix A einer FEM-Berechnung mit $n = 12318$

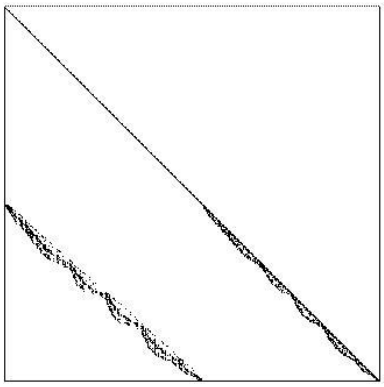
riablen (Algorithmus: *greedy independent set ordering* [45, S. 79]) eine Vertauschungsmatrix P ermittelt, so daß man durch Zeilen- und Spalten-Vertauschung $\widetilde{A}$ erhält.

$$\widetilde{A} = P\,A\,P^T \tag{6.7}$$

Wird die Vertauschung nun auf A^{Sym}angewendet (siehe Abb. 6.4 a), S. 78), entstehen in der oberen Dreiecksmatrix Einträge, die aus der unteren Dreiecksmatrix stammen. Dementsprechend fehlen in der unteren Dreiecksmatrix noch Einträge, die aus der oberen Dreiecksmatrix



(a) $\tilde{A}$ von Matrix A^{Sym}



(b) $\tilde{A}$ nach Transponieren des Anteils

Abbildung 6.4: Die vertauschte Matrix von A^{Sym}

stammen. Da A symmetrisch ist, sind auch beide Anteile symmetrisch. Durch Transponieren dieses Anteil kann $\widetilde{A}^{Sym}$ bestimmt werden, ohne daß die gesamte Matrix benötigt wird.

Vertauscht man auch x und b, so erhält man mit

$$\widetilde{x} = P\,x \tag{6.8}$$
$$\widetilde{b} = P\,b \tag{6.9}$$

ein neues Gleichungssystem

$$\widetilde{A}\,\widetilde{x} = \widetilde{b} \tag{6.10}$$
$$\begin{pmatrix} \widetilde{D} & \widetilde{E}^T \\ \widetilde{E} & \widetilde{H} \end{pmatrix} \begin{pmatrix} \widetilde{x_1} \\ \widetilde{x_2} \end{pmatrix} = \begin{pmatrix} \widetilde{b_1} \\ \widetilde{b_2} \end{pmatrix} \tag{6.11}$$

Betrachtet man (6.11) als System von zwei Gleichungen mit zwei Unbekannten:

$$\widetilde{D}\,\widetilde{x_1} + \widetilde{E}^T\,\widetilde{x_2} = \widetilde{b_1} \tag{6.12}$$
$$\widetilde{E}\,\widetilde{x_1} + \widetilde{H}\,\widetilde{x_2} = \widetilde{b_2} \tag{6.13}$$

Kann man durch Auflösen von (6.12) nach $\widetilde{x_1}$:

$$\widetilde{x_1} = (\widetilde{b_1} - \widetilde{E}^T\,\widetilde{x_2})\,\widetilde{D}^{-1} \tag{6.14}$$

erhalten. Gleichung (6.14) stellt nun genau die Beziehung (6.3) dar. Mit Einsetzen von (6.14) in (6.13) erhält man dann:

$$\widetilde{E}\,(\widetilde{b_1} - \widetilde{E}^T\,\widetilde{x_2})\,\widetilde{D}^{-1} + \widetilde{H}\,\widetilde{x_2} = \widetilde{b_2} \tag{6.15}$$
$$(\widetilde{H} - \widetilde{E}\,\widetilde{D}^{-1}\widetilde{E}^T)\widetilde{x_2} = \widetilde{b_2} - \widetilde{E}\,\widetilde{D}^{-1}\widetilde{b_1} \tag{6.16}$$
$$A^{reduz}x^{reduz} = b^{reduz} \tag{6.17}$$

Gleichung (6.16) ist wieder ein Gleichungssystem der Form $A\,x = b$, was numerisch zu lösen ist. Das Ergebnis $\widetilde{x}_2$ ist dann in (6.14) einzusetzen und man erhält den gesamten Vektor $\widetilde{x}$. Mit

$$x = P^{-1}\,\widetilde{x} \tag{6.18}$$

wird der Vektor $\widetilde{x}$ wieder zurück getauscht und man hat das Ergebnis des ursprünglichen Gleichungssystems (6.4) von S.77.

Das Verfahren der Kondensation hat als primäres Ziel, die *inneren* Kantenvariablen zu eliminieren und die Gesamtzahl der Unbekannten zu reduzieren. Es hat jedoch auch Nachteile, so daß man den Einsatz der Kondensation immer erst abwägen muß.

Nachteil 1: Das Matrix A^{reduz} hat bei allen Berechnungen immer mehr Nicht-Null-Einträge in der Matrix erhalten als es in der Matrix A waren. Das bedeutet, daß für eine Matrix-mal-Vektor-Operation immer ein höherer Rechenaufwand entsteht.

Nachteil 2: Für das Reduzieren des Gleichungsystems wird zusätzlicher Speicher und Rechenzeit benötigt, der bei größeren Systemen auch entsprechend mitwächst.

Abwägungskriterien:

⇒ Der Nachteil 1 ist nicht von Bedeutung, wenn dann der iterative Löser viel schneller konvergiert. Dann resultieren trotzdem weniger Rechenoperationen. Das ist jedoch nicht zu beobachten gewesen.

⇒ Je größer ein Netz wird, um so geringer wird in der Regel das Verhältnis der Oberflächen-Kanten zur Gesamt-Kanten-Anzahl und damit auch der Anteil der *inneren* Kantenvariablen im Verhältnis zur Ordnung des Gesamtgleichungssystems. Dann wird Nachteil 2 immer stärker. Bleibt der Anteil der Oberflächen-Kanten zur Gesamt-Kanten-Anzahl groß, so kann die Anzahl der Unbekannten im reduzierten Gleichungssystem bei Kondensation stark reduziert werden.

In der Praxis werden in der Regel mit einem Netz viele Berechnung bei unterschiedlichen Frequenzen durchgeführt. Die Effektivität der Kondensation ändert sich dabei nur unwesentlich. So ist es zu empfehlen, bei einer Frequenz verschiedene Solver mit und ohne Kondensation zu testen und für die vielen Berechnungen den schnellsten Weg zu wählen. Leider gibt es wegen der Vielfalt der Probleme hier keinen eindeutigen Weg, jedoch kann man mit den hier vorgestellten Abwägungskriterien schnell zu einer guten Entscheidung gelangen. Welche Wirkung die Kondensation bei den in dieser Arbeit berechneten Problemen hatte, ist in Tabelle 6.1 zu sehen. Man sieht, daß die Anzahl der *inneren* Knoten nur zwischen $26 - 31\,\%$ liegt. Die Anzahl der Nicht-Null-Einträge *nnz* erhöht sich jedoch um $65 - 60\,\%$. Dadurch steigt der Aufwand für eine Matrix-mal-Vektor-Operation auch um ca. $65 - 60\,\%$. Damit der iterative Lösungsprozeß nicht langsamer wird, muß er um $65 - 60\,\%$ schneller konvergieren. Zusätzlich muß auch noch der Aufwand für die Kondensation selbst berücksichtigt werden.

6.5 Iterative Löser

Welche Löser in IFEP zur Verfügung stehen, ist in Tabelle 6.2 zu sehen. Um auch Löser mit voller Matrixspeicherung (solv-1 und solv-2) vergleichen zu können, wurde ein sehr kleines Problem gewählt. Es handelt sich dabei um einen ungestörten Rechteckresonator.

FEM Problem	n der Matrix A	nnz	*innere* Knoten	n der Matrix A^{reduz}	nnz
Team 19 Abb. 8.1, S. 98	33 787	454 385	8 842	24 945	751 365
Taper Abb. 8.5, S. 102	65 335	844 117	18 601	46 734	1 377 786
Wellenleiter Abb. 8.9, S. 106	61 780	802 262	17 817	43 963	1 329 677
Bandpass Abb. 8.15, S. 112	74 812	921 452	23 209	51 603	1 477 095

Tabelle 6.1: Auswirkung der Kondensation bei den in dieser Arbeit vorgestellten Beispielen ($nnz \equiv$ Anzahl der Nicht-Null-Einträge der ganzen Matrix)

Aufruf:	Algorithmus	Speicherform von A	Speicher f. $Ax = b$ in MByte	Zahlentyp
solv-1-send	Direkt LU-Decomp.	volle Matrix	12,54	double
solv-2-send	linbcg	volle Matrix	12.54	double
solv-3-send	Direkt LU-Decomp.	Bandmatrix	6,94	double
solv-4-send	linbcg	MSR	0,1072	double
solv-5-send	linbcg	sym Bandmatrix	2,326	double
solv-6-send	linbcg	AWS	0,10028	COMPLEX
solv-7-send	linbcg m. Kondensation	AWS	0,10028	COMPLEX
solv-8-send	Lanczos	AWS	0,10028	COMPLEX
solv-9-send	Lanczos m. Kondensation	AWS	0,10028	COMPLEX
solv-10-send	Lanczos	AWS	76,208 KB	double
solv-11-send	miccg	AWS	76,208 KB	double

Tabelle 6.2: Löser die in IFEP zur Verfügung stehen

6.5.1 Bewertung der Löser

Der verwendete Zahlentyp ist von sehr großer Bedeutung, weil er die Anwendbarkeit auf bestimmte Problemstellungen einschränkt.

Eine magnetische Wand bewirkt zum Beispiel, daß auf der Diagonalen Werte <0 entstehen. Beim Lanczos-Löser (solv-10-send) wird zum Beispiel der *Trick von Eisenstadt* [44] verwendet. Es handelt sich dabei um eine Preconditionierung des Gleichungssystems, bei dem durch die Wurzel der Diagonalen geteilt wird. Im Reellen ist das nicht mehr definiert und der Löser kann im reellen nicht verwendet werden.

Wird eine ABC oder komplexwertige Materialkonstante verwendet, führt das immer auf ein komplexes Gleichungssystem.

Für die praktische Anwendung sind also von diesen Lösern nur die komplexen Löser von Bedeutung und sollen genauer auf ihre Konvergenz untersucht werden. Die Algorithmen sind dann:

1. Ein iterativer Bi-Conjugierte Gradienten-Löser (mit *linbcg* bezeichnet) [46, S.86].
2. Ein iterativer Lanczos-Type Conjugierte Gradienten-Löser mit Drei-Term-Rekursion und SSOR Vorkonditionierung für symmetrische indefinite Lineare Gleichungssysteme (mit *Lanczos* bezeichnet) [47] [48] [45, S.174ff].

Dabei steht die Frage im Vordergrund: *Welchen Einfluß hat die Anzahl der in der Feldlösung vorhandenen Wellenlängen auf die Konvergenz der iterativen Löser?*

Dazu werden zwei Beispiele betrachtet:

Kurzer Wellenleiter: Am linken Ende ist der Rechteck-Hohlleiter offen (ABC) und wird mit einem TE_{10}-Mode bei $f = 4,4\,GHz$ angeregt. Am rechten Ende ist er auch offen, hier wird der TE_{10}-Mode absorbiert. Die Länge ist so bemessen, daß sich nur ca. $\frac{1}{4}$ der Wellenlänge im Berechnungsgebiet befindet (siehe Abb. 6.5 auf S. 82). Die Netz besitzt 61221 Kanten, die halbe Bandbreite der Matrix A beträgt 2863. Die gesamte Matrix A besteht aus 888029 Nicht-Null-Einträgen, wobei 4752 Kanten erzwungen sind. Zum Speichern der Matrix werden ca. 10 MByte benötigt. Die Qualität der Netzes ist in Tabelle 6.3 auf S. 82 zu sehen.

Langer Wellenleiter: Dieser Rechteck-Hohlleiter ist wie der kurze Wellenleiter angeregt, jedoch ist die Länge so bemessen, daß sich ca. 2 Wellenlängen im Berechnungsgebiet befindet (siehe Abb. 6.6 auf S. 83). Die Qualität der Netzes ist in Tabelle 8.4auf S. 107 zu sehen. Es besitzt 61780 Kanten, die halbe Bandbreite der Matrix A beträgt 1428. Die gesamte Matrix A besteht aus 802262 Nicht-Null-Einträgen, wobei 9846 Kanten erzwungen sind. Zum Speichern der Matrix werden ca. 9,4 MByte benötigt.

Bei diesen Lösungsverfahren wird die Lösung als Grenzwert einer Folge von Näherungen bestimmt.

Ziel eines iterativen Lösers ist es immer, einen Lösungsvektor x_{test} zu finden, der eine Lösung des linearen Gleichungssystems (6.4) ist. Da jedoch immer nur Näherungen ausprobiert werden, entsteht dabei immer ein Residuum

$$r = A\,x_{test} - b\,. \tag{6.19}$$

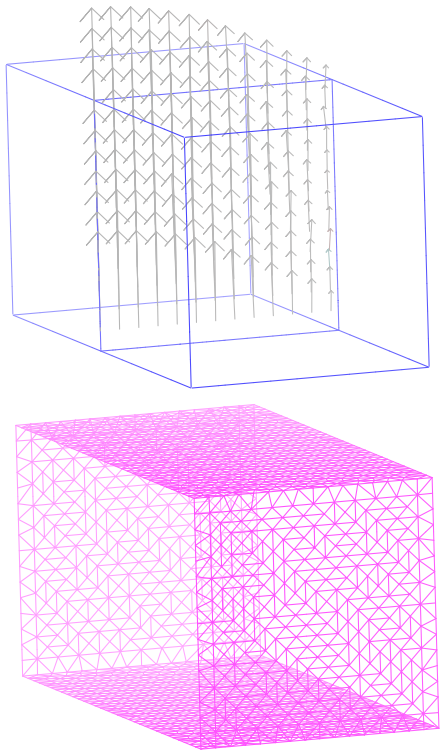


Abbildung 6.5: Kurzer Rechteckwellenleiter

Mittlere Qualität	0.864489
Joint-Qualität	0.857738
schlechteste Tetraeder	0.345961
beste Tetraeder	0.998683
Qualitätshäufigkeit im Netz	
0,0-0,1	0
0,1-0,2	0
0,2-0,3	0
0,3-0,4	5
0,4-0,5	23
0,5-0,6	150
0,6-0,7	1263
0,7-0,8	6893
0,8-0,9	23330
0,9-1,0	17396
kleinstes h	$0,8552314 \cdot 10^{-3}\,m$
größtes h	$3,068177 \cdot 10^{-3}\,m$

Tabelle 6.3: Qualität des Netzes vom kurzen Wellenleiter

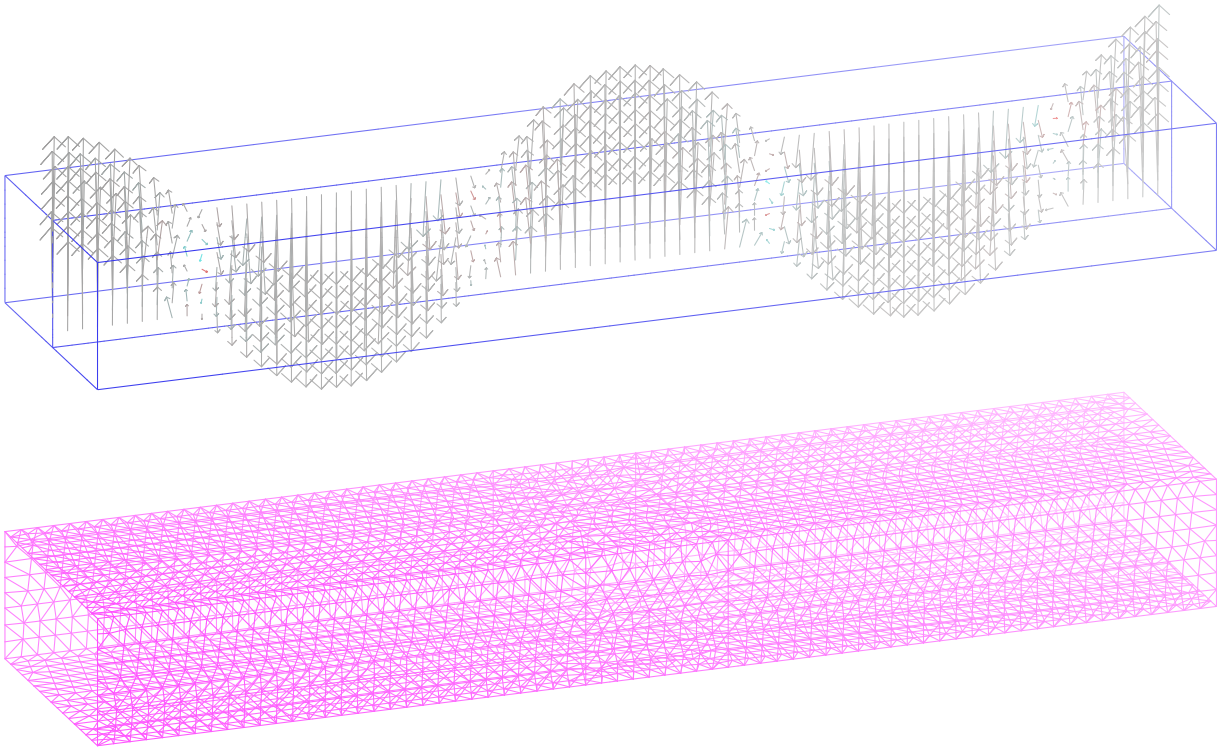


Abbildung 6.6: Langer Rechteckwellenleiter

Dieses Residuum soll durch den Löser minimiert werden. Da r ein Vektor mit der Dimension n ist, verwendet man für den *Abstand*[6] r der Lösung eine Norm, hier die Euklidsche Norm[7] z.B.[45, S.7] im $\mathbb{R}^n$ und $\mathbb{C}^n$:

$$\|r\| = \sqrt{\sum_{i=1}^{n} r_i \cdot r_i^*}\,. \tag{6.20}$$

An der Norm kann man schon gut erkennen, wieweit bei einem Problem eine *gute* oder *schlechte* Konvergenz zu erwarten ist. Die Kanten sind mit dem in Abschnitt 6.3.1 von S.75 beschriebenen Verfahren numeriert worden. Dabei sind die Kanten, vereinfacht gesprochen, in der Geometrie von links nach rechts durchnumeriert worden.

Nun existieren aber nur Anregungen im Bereich der ersten ABC, also nur bei den ersten Kanten des Gleichungssystems. Alle anderen Kantenwerte ergeben sich aus der Lösung im Anregungsgebiet. Da die Matrix A eine dünne Bandstruktur hat und die Unbekannten im Vektor x über die gemeinsammen Kanten der Tetraeder in der Matrix gekoppelt sind, bewertet die Norm (6.20) das Residuum **örtlich.** Schaut man sich das Konvergenzverhalten in Abb. 6.7 an, erkennt man beim langen Wellenleiter, daß in den ersten 300 Iterationen keine Konvergenz zu sehen ist. Das liegt an der Welligkeit der Lösung (siehe Abb.6.6) und der Kopplung der Unbekannten zueinander. Ein kleiner Änderung im Anregungsbereich (hier bei den ersten Kanten) bewirkt beim langen Wellenleiter einen großen Fehler im restlichen Gebiet (also den Kanten die ausschließlich über ihre Kopplung mit den ersten Kanten eindeutig bestimmt sind). Ich bezeichne diese Berechnungszeit als **Ausbilden der Grundwelle im Berechnungsgebiet**.

[6] Abstand bezieht sich auf die Differenz zwischen $A\,x_{test}$ und b im $\mathbb{R}^n$ bzw. im $\mathbb{C}^n$

[7] auch mit L_2-Norm bezeichnet

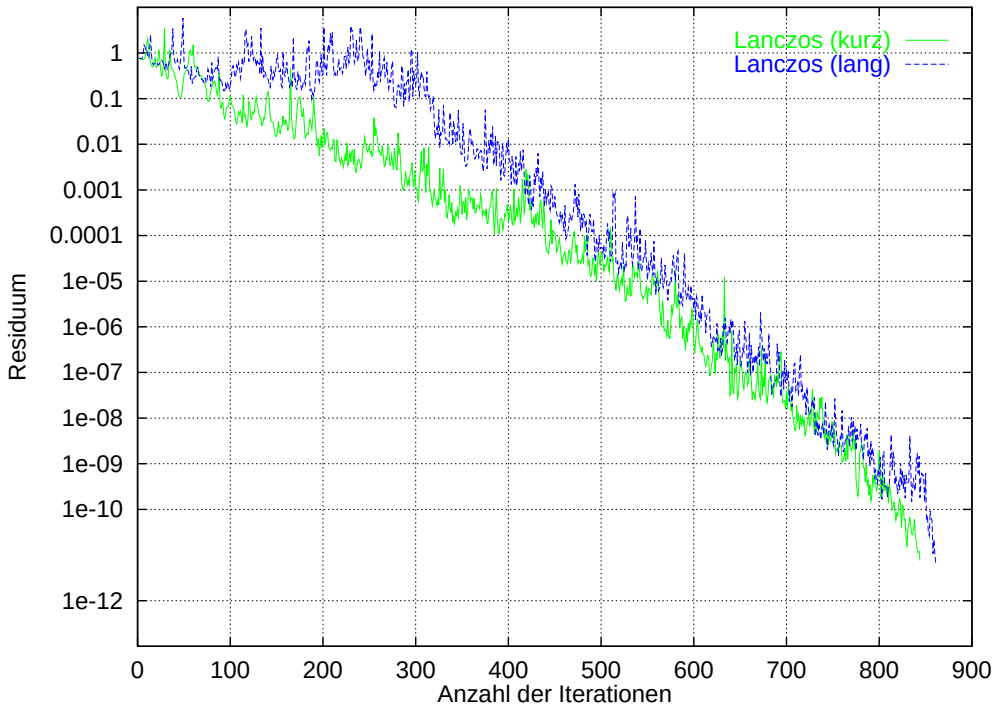


Abbildung 6.7: Konvergenzverhalten vom Lanczos-Löser beim Wellenleiter *kurz* und *lang*

Beim kurzen Wellenleiter ist die Lösung nicht wellig (siehe Abb. 6.5), deshalb kann der Löser sofort die Lösung der Grundwelle kontinuierlich verbessern, was sich sofort durch ein kontinuierliches Konvergieren der Lösung darstellt. Es bleibt die Frage, ob dieser Effekt ein spezielles Problem dieses Lösers ist.

Um die Frage genauer nach zu untersuchen, sind die selben Problemstellungen mit dem linbcg-Löser berechnet worden. Man sieht in Abb 6.8 deutlich, daß dieser Effekt wieder auftritt. Der Unterschied zwischen beiden Lösern besteht einmal in der Anzahl der Iterationen, die gebraucht werden bis sich die Grundwelle ausgebildet hat: Lanczos $\Rightarrow\sim$ 300 Iterationen und linbcg $\Rightarrow\sim$ 2100 Iterationen. Der Effekt des **Ausbilden der Grundwelle** ist also unabhängig vom verwendeten iterativen Löser.

Zum anderen ist aber auch die Konvergenzgeschwindigkeit (Steigung im Falle von Konvergenz) zwischen beiden Lösern unterschiedlich. Beim linbcg-Löser werden für den kurzen Wellenleiter 2504 Iterationen benötigt, wogegen der Lanczos-Löser nur 844 Iterationen braucht (siehe Abb. 6.9).

Beim langen Wellenleiter ist der Lanczos-Lösern in der Konvergenzgeschwindigkeit im Falle der Konvergenz sogar besser (siehe Abb. 6.7). Der linbcg-Löser konvergiert dagegen etwas langsamer als beim kurzen Wellenleiter (siehe Abb. 6.8).

Bleibt noch die Frage, was passiert, wenn sich mehrere Wellenlängen in dem Berechnungsgebiet befinden. Zu diesem Zweck wird hier das Konvergenzverhalten von dem Berechnungsbeispiel Taper betrachtet. Dieses Beispiel hat $5\frac{3}{4}$Wellenlängen in dem Rechteckhohleiter. Alle drei Konvergenzkurven sind in Abbildung 6.11 zu sehen. Man sieht deutlich, daß sich durch die höhere Anzahl der Wellenlängen im Taper das Ausbilden der Grundwelle schwieriger wird und mehr Iterationen benötigt werden. Zusätzlich ist im Fall der Konvergenz eine geringere Konvergenzgeschwindigkeit zu beobachten.

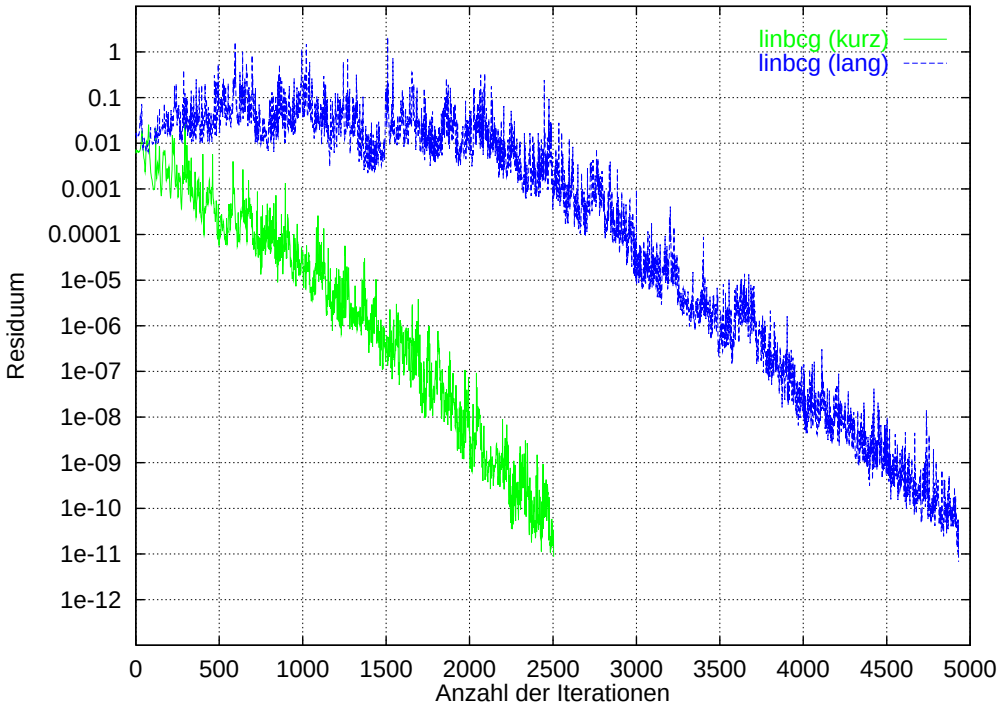


Abbildung 6.8: Konvergenzverhalten vom linbcg-Löser bei Wellenleiter *kurz* und *lang*

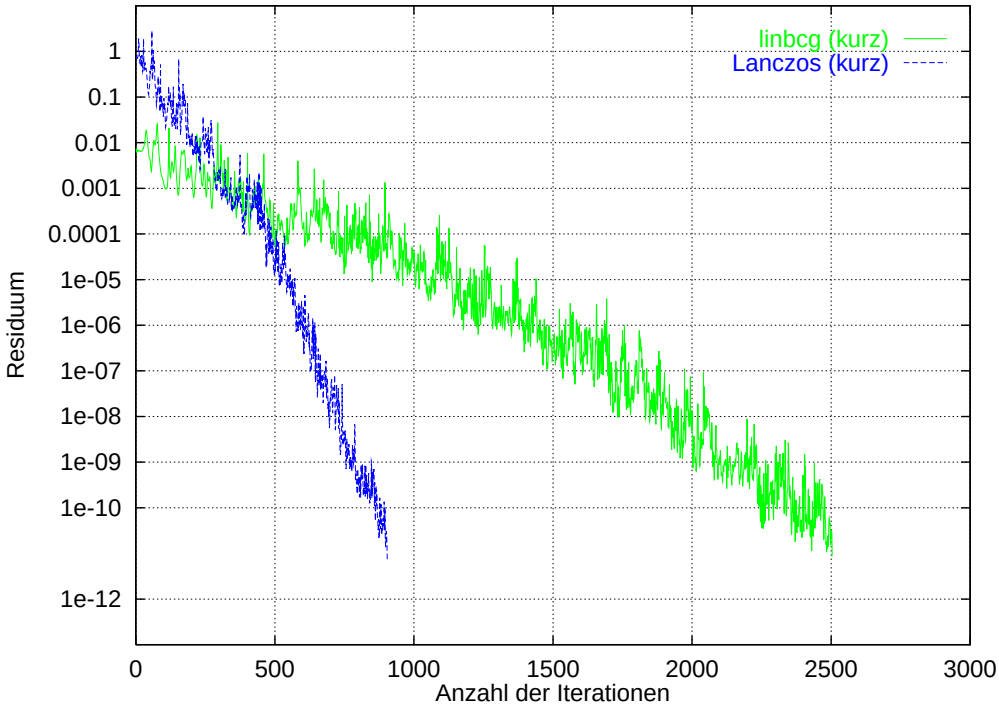


Abbildung 6.9: Konvergenzverhalten vom Lanczos- und linbcg-Löser beim kurzen Wellenleiter

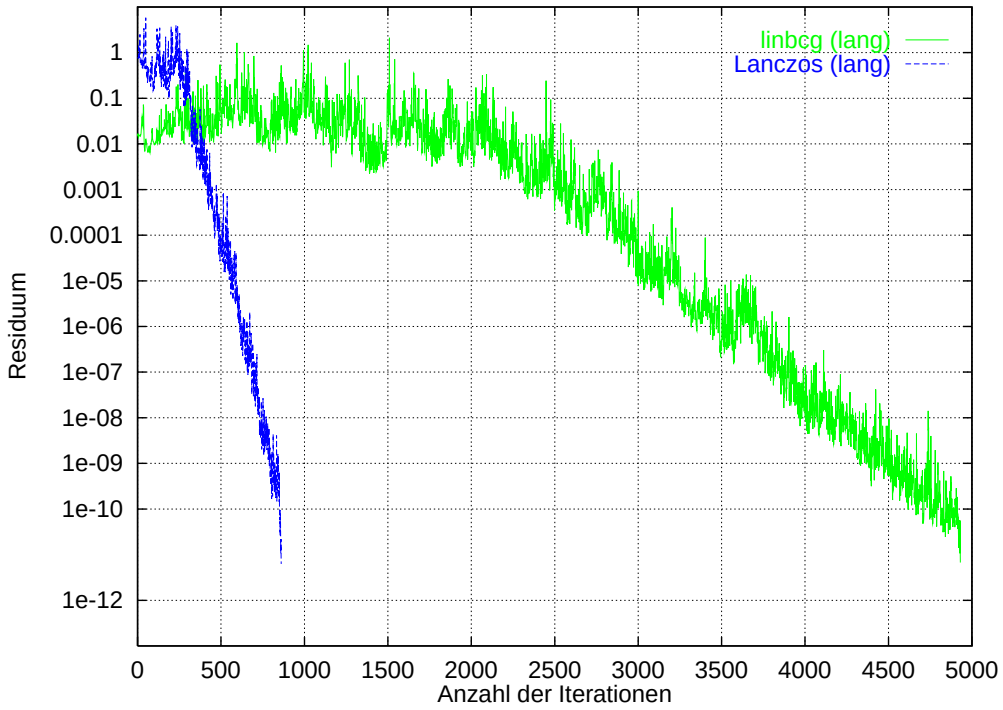


Abbildung 6.10: Konvergenzverhalten vom Lanczos- und linbcg-Löser beim langen Wellenleiter

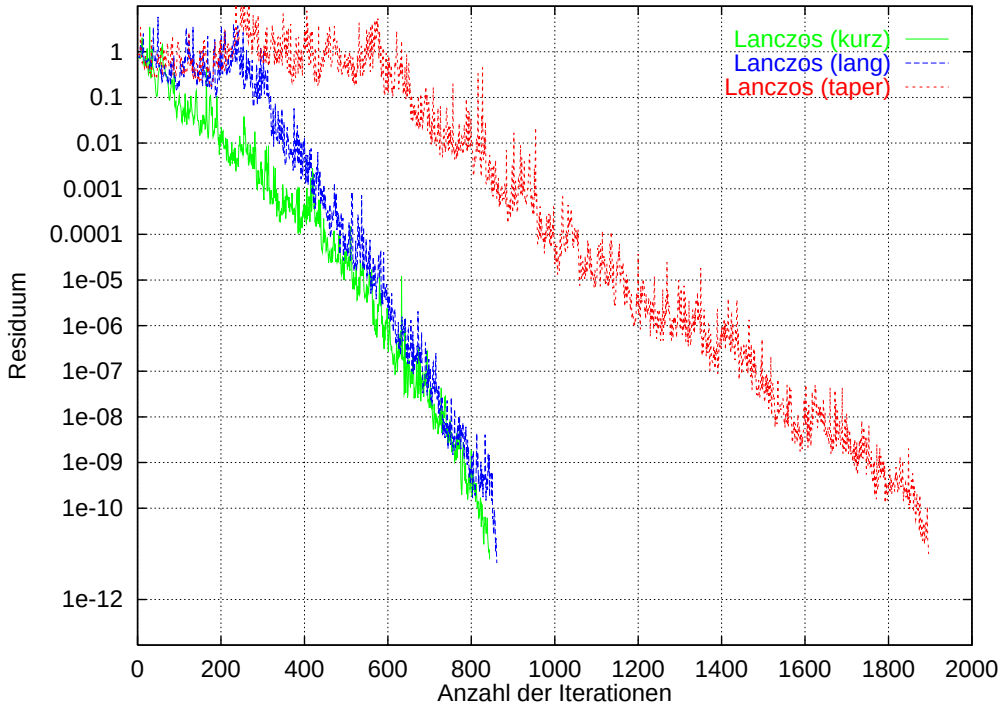


Abbildung 6.11: Konvergenz bei Problemen mit 1/4- , 2-, 5 3/4-fachem der Wellenlänge mit Lanczos-Löser

6.5.1.1 Einfluß der Kondensation

Die in Abschnitt 6.4 beschriebene Kondensation ist auf den langen Wellenleiter angewendet worden. Das dabei beobachtete Konvergenzverhalten ist in Abbildung 6.12 zu sehen. Man erkennt hier deutlich, daß die Kondensation in der Konvergenz keinerlei Gewinn bringt. Dieses Ergebnis erhält man auch für die anderen Beispiele und ist bei anderen Problemen immer wieder bestätigt worden.

Da für die Kondensation selbst jedoch viel Speicher und Rechenzeit benötigt wird, und – das ist noch viel schwerwiegender – die kondensierte Matrix voller besetzt wird, hat sich die Kondensation als Hilfsmittel zum Lösen des Gleichungssystems als nicht brauchbar herausgestellt.

Die Begründung ist leicht nachzuvollziehen. Selbst wenn man die zusätzliche Rechenzeit für die Kondensation einmal vernachlässigt und man von der gleichen Anzahl von Iterationen wie in Abb. 6.12 aus geht, dann hängt es allein davon ab, wie aufwendig eine Iteration ist. Bei einem iterativem Löser wird pro Iteration mindestens eine Matrix-mal-Vektor Operation durchgeführt. Wobei nur die Rechenoperationen durchgeführt werden, bei denen die Matrix A auch wirklich Werte ungleich Null besitzt. Der Rechenaufwand für die Matrix-mal-Vektor Operation ist also direkt von nnz der Matrix A abhängig. Nun besitzt die unkondensierte Matrix 802262 und die kondensierte Matrix 1329677 Nicht-Null-Einträge. Daraus ergibt sich, daß ein Iteration mit Kondensation ca. $1,66$ mal so aufwendig ist. Damit ist bei gleicher Anzahl von Iterationen die Kondensation nicht brauchbar.

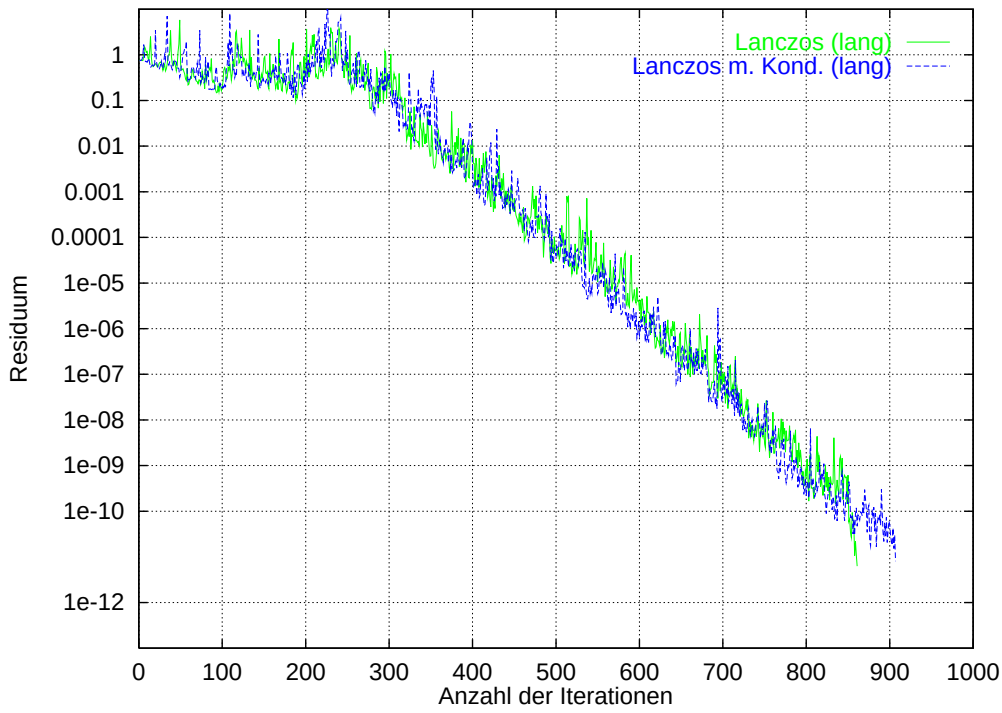


Abbildung 6.12: Konvergenz vom Lanczos-Löser mit und ohne Kondensation beim langen Wellenleiter

6.5.1.2 Einfluß von Resonanzen

Bei den im vorigen Abschnitt gezeigten Problemen traten keine Resonanzen in der Feldlösung auf. Wird die Feldstärke am Ort der Resonanz nicht durch eine Randbedingung künstlich begrenzt, müßte sie dort eigentlich unendlich groß werden. Ein iterativer Löser findet in diesem Bereich immer einen Fehler und erhöht dort die Feldstärke weiter, bis die Fehlerschranke des Residuums erreicht ist. Durch die Feinheit des Netzes wird die maximale Amplitude begrenzt, deshalb konvergiert der Löser trotzdem. Die Konvergenzgeschwindigkeit sinkt jedoch, siehe Abbildung 6.13. Durch die geringe Kopplung des Resonators mit dem Wellenleiter sieht man hier ein zweites Mal das „Ausbilden der Grundwelle", diesmal im Resonator.

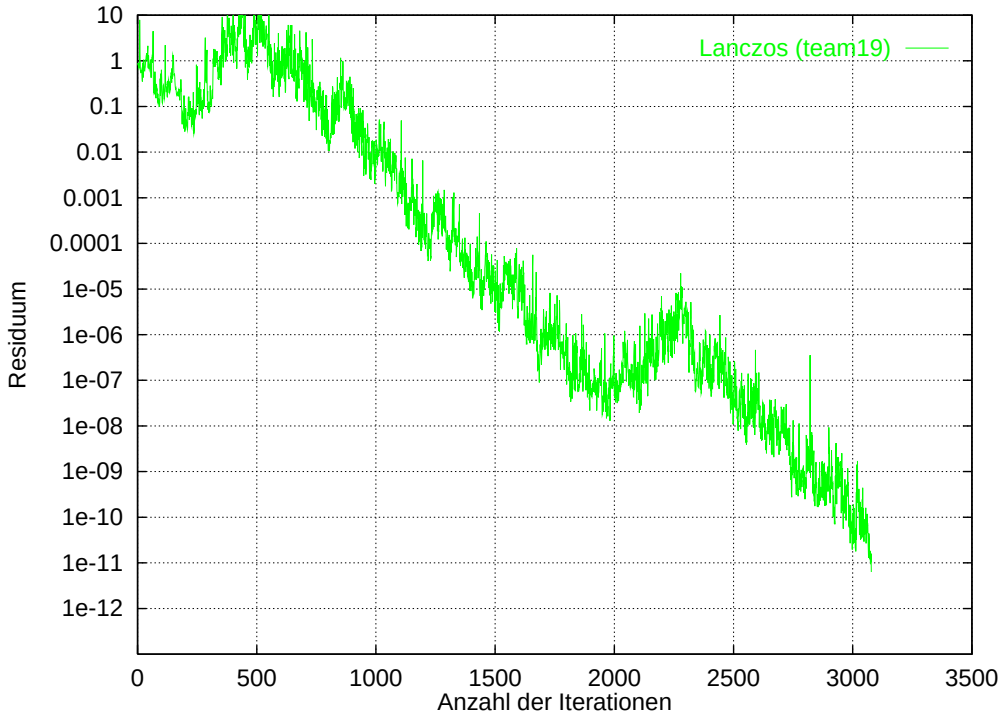


Abbildung 6.13: Konvergenzverhalten des Lanczos-Löser beim TEAM Problem 19 von Abschnitt 8.1 bei der Resonanzfrequenz $f = 2,543\,GHz$

Kapitel 7

Auswertung der berechneten Felder

Man kann die berechneten Felder auf viele verschiedene Weisen auswerten. Hier wird auf die Visualisierung der Feldlösung und die aus der Feldlösung berechenbaren Streuparameter eingegangen. Es hängt ganz davon ab, was bei dem Entwurf optimiert werden soll. Meist sind es jedoch konkrete Parameter, die optimiert werden sollen. Die Visualisierung dient mehr als qualitative Kontrolle, bei der der Ingenieur mit seiner Erfahrung für Feldbilder schnell und effektiv die Richtigkeit der Lösung abschätzen kann und schnell ein Bild von den physikalischen Effekten bekommt.

7.1 Visualisierung der Feldlösung

Durch die Sprünge in den Normalkomponenten der elektrischen Feldstärke an den Elementgrenzen wird schnell deutlich, daß die Visualisierung nicht trivial ist.

Es gibt eine einfache Methode, die Felder darzustellen. Man stellt an den Knoten den Mittelwert aller tangentialen Feldstärken dar, die auf den angrenzenden Kanten auftreten und interpoliert zwischen allen Knoten linear. Die Methode bewirkt jedoch eine starke Glättung des Feldverlaufs und damit eine Verfälschung der Lösung. An Grenzflächen mit Sprung in der Normalkomponente wäre dieser nicht mehr zu erkennen. Die Kantenelemente haben ja als Besonderheit, daß sie eine „aufgeweichte" Stetigkeit besitzten. Damit eine richtige Abschätzung zwischen der numerischen Lösung und dem physikalisch Erwarteten vorgenommen werden kann, **muß die wirkliche numerische Feldlösung visualisiert werden**.

In IFEP werden alle Berechnungsobjekte mit Kantenelementen durch eine Klasse namens FED3D spezifiziert. Ein Objekt (Instance) dieser Klasse besteht aus dem Tetraedernetz, der Kantenbeschreibung, den Randbedingungen, der Frequenz, für die die Feldlösung berechnet wurde, und der E-Feldlösung. Das Objekt besitzt durch die FEM-Berechnung die bekannten Werte auf den Kanten und die genaue Ansatzfunktion. Damit steht immer die exakte numerische Lösung im gesamten Berechungsgebiet zur Verfügung.

Für alle Objekte der Klasse FED3D existieren nun ganz unterschiedliche Visualisierungs-Methoden, bei denen für die Auswertung die speziellen Ansatzfunktionen in den Tetraedern verwendet werden. Dadurch kann man im Gegensatz zu vielen anderen FEM-Programmen auch wirklich die Lösung betrachten, die berechnet wurde, und nicht eine Mittelung, Glättung oder Interpolation.

Bei vielen Parametern der Visualisierungen besteht die Möglichkeit, sich eine Parametervariation visualisieren zu lassen. So kann man zum Beispiel Schnittebenen durch das Volumen fahren lassen oder sich alle Ebenen gleichen Feldstärkebetrags (z.B. level-disp) von der kleinsten bis zur größten Feldstärke in einer Animation anschauen. Oder man bewegt einfach seinen Blickpunkt durch das Volumen.

7.1.1 Das elektrische Feld

Für das elektrische Feld stehen für Objekte der Klasse FED3D die folgenden Display-Methoden zur Verfügung:

Display-Methode	Wirkung
clip-efeld-disp	Pfeildarstellung in beliebiger Schnittebene des Realteils von $\vec{E}$
clip-iefeld-disp	Pfeildarstellung in beliebiger Schnittebene des Imaginärteils von $\vec{E}$
level-disp	Darstellung der Flächen gleichen Betrags des Realteils von $\vec{E}$
ilevel-disp	Darstellung der Flächen gleichen Betrags des Imaginärteils von $\vec{E}$
edge-disp	Darstellung der Kanten mit Färbung proportional zum Betrag des Realteils von $\vec{E}$

7.1.2 Das magnetische Feld

Das magnetische Feld wird aus der Rotation des E-Feldes berechnet[1]. Dabei ist zu beachten, daß dadurch die Funktion $\vec{H}$ in einem Elementen konstant sind. Was sich aus der Rotation der Ansatzfunktionen ergibt (siehe auch Abschnitt 4.4 auf S.42).

Für das magnetische Feld stehen die folgenden Display-Methoden zur Verfügung:

Display-Methode	Wirkung
clip-hfeld-disp	Pfeildarstellung in beliebiger Schnittebene des Realteils von $\vec{H}$
clip-ihfeld-disp	Pfeildarstellung in beliebiger Schnittebene des Imaginärteils von $\vec{H}$
level-disp	Darstellung der Flächen gleichen Betrags des Realteils von $\vec{H}$
ilevel-disp	Darstellung der Flächen gleichen Betrags des Imaginärteils von $\vec{H}$

7.1.3 Die Leistungsdichte

Die Berechnung des Poynting-Vektor $\vec{S}$ wird im Abschnitt 7.2.2 auf S.94 gezeigt.

Für das Leistungsdichte-Feld stehen die folgenden Display-Methoden zur Verfügung:

Display-Methode	Wirkung
clip-Power-disp	Pfeildarstellung in beliebiger Schnittebene der gesamten Leistungsdichte $\vec{P}$
clip-pwirk-disp	Pfeildarstellung in beliebiger Schnittebene des Realteils von $\vec{P}$
clip-pblind-disp	Pfeildarstellung in beliebiger Schnittebene des Imaginärteils von $\vec{P}$

[1] siehe auch Gleichung (95) auf S. 7.29

7.1.4 Animation durch den Frequenzbereich (Wobbeln)

In IFEP besteht die Möglichkeit, ein Problem in einem bestimmten Frequenzbereich zu berechnen und sich die Visualisierungen über den Frequenzbereich animieren zu lassen. So kann man zum Beispiel visuell sehen, bei welcher Frequenz es zu einer Resonanz kommt.

7.2 Die Streuparameterberechnung

Man verwendet zur Schaltungsbeschreibung bei Frequenzen oberhalb $100\,MHz$ die sogenannten Streuparameter [42], die bei Anpassung gemessen werden. Dabei werden die vom Netzwerk gestreuten (reflektierten) Wellen in Bezug zu den einfallenden Wellen gesetzt. Die sich ergebende Streumatrix wird zur Beschreibung von aktiven wie passiven Bauelementen benutzt.

In IFEP finden zwei verschiedene Methoden zur Streuparameterberechnung ihre Anwendung. Die erste ist [24, S.266] entnommen. Diese Methode ist bei Transmissionsfaktorberechnungen nur sehr eingeschränkt brauchbar und wird hier für Referenzzwecke verwendet. Die zweite Methode ist aus der Definition der Streuparameter hergeleitet (z.B. [41]) und viel flexibler in den Anwendungsmöglichkeiten.

7.2.1 Die einfache Anordnung (reflex)

Bei dieser Berechnungsmethode wird immer von einer Anordnung ausgegangen, wie sie in Abbildung 5.1 auf S. 49 zu sehen ist, für die Transmissionsfaktor-Berechnung jedoch mit der Einschränkung, daß das Einfalls-Tor S_1 genau so groß ist wie das Ausfalls-Tor S_2. Das heißt, daß $a_1 = a_2$ und $b_1 = b_2$ sein muß. Des weiteren darf sich die Ausbreitungsrichtung des TE_{10}-Modes zwischen den beiden Toren nicht ändern.

Grundlage dieser Berechnung ist die analytische $\vec{E}$-Feld-Lösung der einfallenden Wellen an Tor S_1 und die Kenntnis über die analytische $\vec{E}$-Feld-Lösung der ausfallenden Wellen, wenn sich die Wellen im ungestörten Fall um den Abstand zwischen S_1 und S_2 weiter bewegt hätte.

Die Wellenzahl $k_{z_{10}}$ und der Vektor $\overrightarrow{e_{10}}$ sind

$$k_{z_{10}} = \sqrt{k_0^2 - \left(\frac{\pi}{a}\right)^2}, \tag{7.1}$$

$$\vec{e}_{10} = \sin\left(\frac{\pi x}{a}\right)\overrightarrow{e_y}\,. \tag{7.2}$$

Für den Reflexionsfaktor R gilt für die Fläche S_1 bei $z = z_1$

$$\begin{aligned}\vec{E}(x,y,z_1) &= \vec{E}^{inc}(x,y,z) + \vec{E}^{ref}(x,y,z_1)\\ &= E_0\vec{e}_{10}(x,y)e^{-jk_{z_{10}}z_1} + R\,E_0\vec{e}_{10}(x,y)e^{jk_{z_{10}}z_1}\end{aligned} \tag{7.3}$$

und für den Transmissionfaktor T gilt auf der Fläche S_2 bei $z = z_2$

$$\begin{aligned}\vec{E}(x,y,z_2) &= \vec{E}^{trans}(x,y,z_2)\\ &= T\,E_0\vec{e}_{10}(x,y)e^{-jk_{z_{10}}z_2}\,.\end{aligned} \tag{7.4}$$

So gilt für *jeden* Punkt auf der Fläche S_1

$$R = \left.\frac{\vec{E}(x,y,z) - \vec{E}_0\vec{e}_{10}(x,y)e^{-jk_{z_{10}}z}}{E_0\vec{e}_{10}(x,y)e^{jk_{z_{10}}z}}\right|_{z=z_1} \tag{7.5}$$

und für *jeden* Punkt auf der Fläche S_2

$$T = \left.\frac{E(x,y,z)}{E_0 e_{10}(x,y)e^{-jk_{z_{10}}z}}\right|_{z=z_2}\,. \tag{7.6}$$

Nun berechnet man den Reflexions- oder Transmissionsfaktor nicht punktweise. Eine bessere Genauigkeit kann man durch Ausnutzen der Orthogonalität der Wellenleiter-Moden erhalten. Da der dominante Mode zu allen anderen Moden orthogonal ist, kann man (7.3) und (7.4) mit $\vec{e}_{10}$ skalar multiplizieren[2] und anschließend über die Fläche integrieren.

7.2.1.1 Die Reflexion

Für die Reflexion gilt dann

$$\begin{aligned}\int_{S_1}\vec{E}(x,y,z)\cdot\vec{e}_{10}\,dxdy &= \int_{S_1}E_0\,\vec{e}_{10}(x,y)e^{-jk_{z_{10}}z}\cdot\vec{e}_{10}\,dxdy\\ &\quad+\int_{S_1}R\,E_0\,\vec{e}_{10}(x,y)e^{jk_{z_{10}}z}\cdot\vec{e}_{10}\,dxdy\,.\end{aligned} \tag{7.7}$$

Setzt man nun (7.2) ein, erhält man

$$\int_{S_1}\vec{E}(x,y,z)\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy = \int_{S_1}E_0\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,e^{-jk_{z_{10}}z}\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy \tag{7.8}$$

$$+\int_{S_1}R\,E_0\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,e^{jk_{z_{10}}z}\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy \tag{7.9}$$

Mit $\vec{e}_y\cdot\vec{e}_y = 1$ folgt dann

$$\int_{S_1}\vec{E}(x,y,z)\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy = E_0e^{-jk_{z_{10}}z}\int_{S_1}\sin^2\left(\frac{\pi x}{a}\right)\,dxdy \tag{7.10}$$

$$+R\,E_0e^{jk_{z_{10}}z}\int_{S_1}\sin^2\left(\frac{\pi x}{a}\right)\,dxdy \tag{7.11}$$

[2] Gilt nur wenn ε und μ konstant in der Fläche sind.

beim Rechteckhohlleiter gilt dann für die Fläche S_1:

$$\int_{S_1} \vec{E}(x,y,z_1) \cdot \sin\left(\frac{\pi x}{a}\right) \vec{e}_y \, dxdy = E_0 e^{-jk_{z_{10}} z_1} \int_0^b \int_0^a \sin^2\left(\frac{\pi x}{a}\right) dxdy \quad (7.12)$$

$$+R\, E_0 e^{jk_{z_{10}} z_1} \int_0^b \int_0^a \sin^2\left(\frac{\pi x}{a}\right) dxdy \quad (7.13)$$

Löst man nun das Integral über $\sin^2\left(\frac{\pi x}{a}\right)$

$$\int_{S_1} \vec{E}(x,y,z_1) \cdot \sin\left(\frac{\pi x}{a}\right) \vec{e}_y \, dxdy = E_0 e^{-jk_{z_{10}} z_1} \int_0^b \left[\frac{1}{2}x - \frac{a}{4\pi}\sin\left(\frac{2\pi x}{a}\right)\right]_0^a dy \quad (7.14)$$

$$+R\, E_0 e^{jk_{z_{10}} z_1} \int_0^b \left[\frac{1}{2}x - \frac{a}{4\pi}\sin\left(\frac{2\pi x}{a}\right)\right]_0^a dy \quad (7.15)$$

erhält man

$$\int_{S_1} \vec{E}(x,y,z_1) \cdot \sin\left(\frac{\pi x}{a}\right) \vec{e}_y \, dxdy = E_0 e^{-jk_{z_{10}} z_1} \int_0^b \frac{1}{2} a \, dy \quad (7.16)$$

$$+R\, E_0 e^{jk_{z_{10}} z_1} \int_0^b \frac{1}{2} a \, dy\,. \quad (7.17)$$

Löst man das Integral von 0 bis b

$$\int_{S_1} \vec{E}(x,y,z_1) \cdot \sin\left(\frac{\pi x}{a}\right) \vec{e}_y \, dxdy = \frac{E_0 ab}{2} e^{-jk_{z_{10}} z_1} \quad (7.18)$$

$$+R\, \frac{E_0 ab}{2} e^{jk_{z_{10}} z_1} \quad (7.19)$$

und formt die Gleichung nach R um, erhält man

$$\frac{2}{E_0 ab} \int_{S_1} \vec{E}(x,y,z_1) \cdot \sin\left(\frac{\pi x}{a}\right) \vec{e}_y \, dxdy \quad -e^{-jk_{z_{10}} z_1} = R\, e^{jk_{z_{10}} z_1}\,. \quad (7.20)$$

Multipliziert man die Gleichung mit $e^{-jk_{z_{10}} z_1}$, erhält man schließlich

$$R = \frac{2e^{-jk_{z_{10}} z_1}}{abE_0} \int_{S_1} \vec{E}(x,y,z_1) \vec{e}_{10}(x,y) \, dxdy - e^{-2jk_{z_{10}} z_1}\,. \quad (7.21)$$

Durch die Ausnutzung der Orthogonalität und den analytischen Ansatz ist diese Berechnungsmethode für den Reflexionsfaktor in IFEP die genaueste. Hier ist die Beschaffenheit des nachfolgenden Hohlleiters weitgehend egal, jedoch müssen die höheren Moden der reflektierten Welle am Tor S_1 abgeklungen sein.

7.2.1.2 Die Transmission

Für die Transmission gilt dann

$$\int_{S_2} \vec{E}(x,y,z_2)\cdot\vec{e}_{10}\,dxdy \;=\; \int_{S_2} T\,E_0\vec{e}_{10}(x,y)e^{-jk_{z_{10}}z_2}\cdot\vec{e}_{10}\,dxdy\,.$$

Setzt man nun (7.2),von S.91 ein, erhält man

$$\begin{aligned}\int_{S_2} \vec{E}(x,y,z_2)\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy &= \int_{S_2} T\,E_0\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,e^{-jk_{z_{10}}z_2} && (7.22)\\ &\quad\cdot\sin\left(\frac{\pi x}{a}\right)\vec{e}_y\,dxdy && (7.23)\\ &= T\,\frac{E_0ab}{2}e^{-jk_{z_{10}}z_2}\,. && (7.24)\end{aligned}$$

Löst man die Gleichungen nach T auf, erhält man

$$T \;=\; \frac{2e^{jk_{z_{10}}z_2}}{abE_0}\int_{S_2}\vec{E}(x,y,z_2)\cdot\vec{e}_{10}(x,y)\,dxdy \qquad (7.25)$$

Durch die Ausnutzung der Orthogonalität und den analytischen Ansatz ist diese Berechnungsmethode für den Transmissionsfaktor in IFEP die genaueste. Wegen der Einschränkung der möglichen Anordnungen, wie: z_2 muß bekannt sein, gleiches a und b sowie gleiche Ausbreitungsrichtung an S_1 und S_2, ist diese Methode leider nur sehr selten anwendbar. Auch hier müssen die höheren Moden am Tor S_2 abgeklungen sein.

7.2.2 Die Berechnung über die Leistung (power)

Es ist auch möglich, die Streuparameterbeträge aus den Leistungsbilanzen an den Toren zu berechnen. Für diese Berechnungsmethode ist es jedoch nötig, daß man den Leistungsfluß durch die Tore berechnen kann. Im ersten Abschnitt wird beschrieben, wie in IFEP der Leistungfluß durch ein Tor berechnet wird. Anschließend wird gezeigt, wie die Streuparameter aus den Leistungen an den Toren berechnet werden.

7.2.2.1 Leistung durch ein Tor

Mit der Welle ist ein Leistungsfluß in Ausbreitungsrichtung verbunden. Allgemein berechnet man den Momentanwert der Leistungsflußdichte (Poynting-Vektor) $\vec{S}$ aus

$$\overrightarrow{S} = \overrightarrow{E}\times\overrightarrow{H}\,. \qquad (7.26)$$

Für komplexe Größen ist der zeitliche Mittelwert S_m

$$\overrightarrow{S_m} = \frac{1}{2} Re\left(\overrightarrow{E} \times \overrightarrow{H}^*\right). \tag{7.27}$$

Die Amplitude des zeitlich gemittelten Poynting-Vektors (7.27) stellt dann den Leistungsfluß pro Quadratmeter in der Welle dar.

Die Leistung $\widetilde{P}$ durch eine Fläche A ist dann durch

$$\widetilde{P} = \int_A \overrightarrow{S_m} \cdot d\vec{A} = \int_A \frac{1}{2} Re\left(\overrightarrow{E} \times \overrightarrow{H}^*\right) \cdot d\vec{A} \tag{7.28}$$

gegeben.

Nun liegt aus der numerischen Berechnung für die Welle nur das $\overrightarrow{E}$-Feld vor. Das $\overrightarrow{H}$-Feld muß über die Maxwellschen Gleichungen z.B. [10, Seite 728] mit

$$\overrightarrow{H} = -\frac{1}{j\omega\mu} rot\, \overrightarrow{E} \tag{7.29}$$

berechnet werden.

7.2.2.2 Die Streuparameter

Die Streuparameter sind bei einem Zweitor wie folgt über die Leistung definiert z.B. [41, S. 127] :

$$\begin{aligned} |S_{11}|^2 &= \frac{Leistung\, reflektiert\, an\, S_1}{Leistung\, einfallend\, an\, S_1} && (7.30)\\ &= |r_1|^2 && (7.31)\\ &= \frac{P_{in_{ref}} - P_{in}}{P_{in_{ref}}} && (7.32) \end{aligned}$$

$$\begin{aligned} |S_{22}|^2 &= \frac{Leistung\, reflektiert\, an\, S_2}{Leistung\, einfallend\, an\, S_2} && (7.33)\\ &= |r_2|^2 && (7.34)\\ &= \frac{P_{out_{ref}} - P_{out}}{P_{out_{ref}}} && (7.35) \end{aligned}$$

$$\begin{aligned} |S_{21}|^2 &= \frac{Leistung\, geliefert\, an\, S_2}{Leistung\, verfügbar\, an\, S_1} && (7.36)\\ &= |t_2|^2 && (7.37)\\ &= \frac{P_{out}}{P_{in_{ref}}} && (7.38) \end{aligned}$$

$$|S_{12}|^2 = \frac{Leistung\, geliefert\, an\, S_1}{Leistung\, verfügbar\, an\, S_2} \tag{7.39}$$

$$= |t_1|^2 \tag{7.40}$$

$$= \frac{P_{in}}{P_{out_{ref}}}\,. \tag{7.41}$$

Der Nachteil bei dieser Berechnungsmethode besteht in der zusätzlichen Berechnung eines ungestörten Wellenleiters. Diese Referenzberechnung dient dazu, die Leistung $P_{in_{ref}}$ an Tor 1 zu bestimmen, die bei einem ungestörten Wellenleiter mit genau diesem Querschnitt bei dieser Frequenz durch das Tor dringen würde. Hierdurch entsteht eine zusätzliche Fehlerquelle. Hat die Referenzberechnung einen Fehler, wirkt er sich direkt auf die Streuparameter aus.

Dabei hat sich herausgestellt, daß sich die Genauigkeit erhöht, wenn:

⇒ die Tore durch möglichst viele Dreiecksflächen diskretisiert sind und

⇒ das Netz möglichst sehr unregelmäßig ist.

Das erste Kriterium ist sicher klar. Die zweite Forderung nach einem möglichst zufälligen Netz dagegen nicht. In [38] wurde gezeigt, daß strukturierte Netze Phasenfehler besitzen, die sich durch die Regelmäßigkeit des Netzes nicht aufheben können. Bei unregelmäßigen Netzen kann es jedoch positive wie negative Phasenfehler geben. Diese können sich dann aufheben und bewirken eine höheren Genauigkeit.

In IFEP führten Leistungsberechungen bei Referenzberechnung mit regelmäßigen Netzen immer zu einem bleibenden Fehler zwischen Eingangsleistung zu Ausgangsleistung unabhängig von der Genauigkeit der Diskretisierung an den Toren. Aus diesem Grund wird für die Referenzberechnung in IFEP immer automatisch ein unstrukturiertes Netz mit guter Qualität generiert. Für eine Frequenzberechnung wird auf einem Linux PC mit 200 MHz Pentium eine Rechenzeit von ca. 60 sec benötigt.

Im nächsten Kapitel wird dieser Fehler auch an Beispielen näher diskutiert.

Kapitel 8

Anwendungen und Resultate

Das Programm IFEP bietet noch Möglichkeiten für andere Anwendungen, hier sollen jedoch die wichtigsten vorgestellt werden und in Ihrer Qualität bewertet werden. Dabei werden das Suchen von Resonanzfrequenzen in geschlossenen wie offenen Resonatoren sowie die Berechnung von Transmission und Reflexion in Wellenleiterstukturen betrachtet. Bei allen Beispielen werden die Qualität der Tetraeder, sowie die Netzgröße h und die Qualität der Diskretisierung Q_h angegeben.

8.1 Resonatoren

Man kann zwei Arten von Resonatoren unterscheiden: Einmal *geschlossene* Resonatoren, sie haben keine Öffnung zu anderen Gebieten und besitzen damit auch keine Anregung eines anderen Feldes. Oder *offene* Resonatoren, sie werden durch eine Öffnung zu einem anderen Berechnungsgebiet angeregt. Im Falle der Resonanzfrequenz entsteht dann im Verhältnis zur Anregung eine sehr große Feldstärke im Resonator.

Bei geschlossenen Resonatoren besteht die Schwierigkeit in der Anregung bei der Suche der Resonancefrequenz. Man muß auf einer geeigneten Kante einen Wert ungleich Null erzwingen. Dabei ist es wichtig, daß sich die Kante an einem Ort befindet, an dem man eine geringe Feldstärke erwartet. Das kann zum Beispiel nahe an einer PEC-Wand sein. Dann wird sich im Falle einer Resonanz im Resonator eine sehr hohen Feldstärke einstellen und man kann die Resonanz erkennen. Liegt die erzwungene Kante am Ort, wo das resonante Feld sowieso ein Maximum hat, wird durch die erzwungene Kante die Feldlösung künstlich begrenzt und man kann die Resonanz vielleicht gar nicht sehen, da das Maximum ja erzwungen wird.

Hat man eine geeignete Anregung gewählt, berechnet man im zu untersuchenden Frequenzbereich viele Frequenzen und erkennt dann in der Animation über den Frequenzbereich, wo Resonanzen auftreten.

Das nun folgende Beispiel ist ein offener Resonator und stammt von dem internationalem TEAM Workshop für Berechnung von elektromagnetischen Feldern. Dort hat das Beispiel die Bezeichnung „Team Problem 19" [49].

In Abbildung 8.1 ist die untere Hälfte des Aufbaus zu sehen. Die Irisweite beträgt bei der hier betrachteten Berechnung $l = 15\,mm$. Die Struktur wird vorne an dem Rechteckhohlleiter mit einem TE_{10}-Mode angeregt und man erhält im Falle der Resonanzfrequenz eine große Feldstärke im angekoppelten Zylinder.

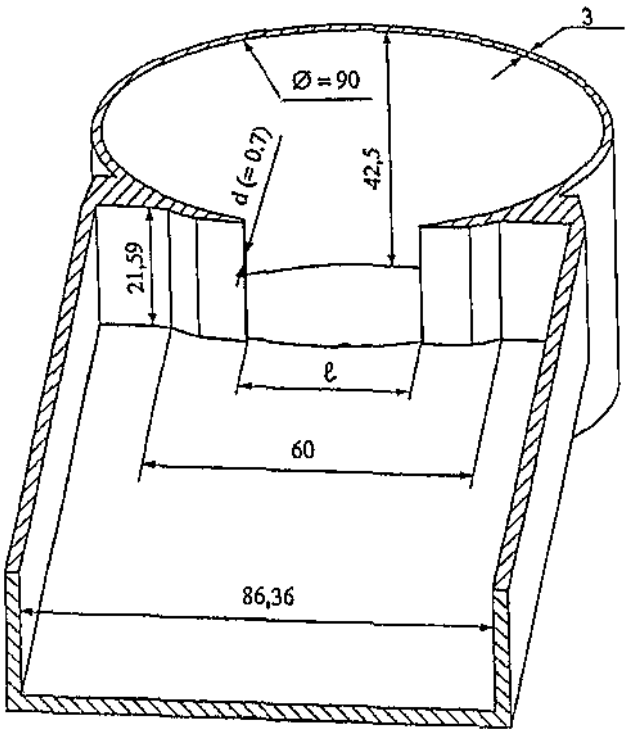


Abbildung 8.1: Technische Abmessungen des TEAM Problems 19 (alle Maße sind mm)

Das verwendete Netz ist in Abbildung 8.2 zu sehen. Es sind beide vorhandenen Symmetrieebenen mit einer magnetischen Wand und einer elektrischen Wand (PEC) ausgenutzt worden. Somit ist das Berechnungsgebiet nur 1/4 so groß. Das Netz besteht aus 5488 Knoten, 26266 Tetraedern und 33787 Kanten (davon sind 4303 Kanten erzwungen). Die Qualität des Netzes sieht man in Tabelle 8.1.

Im Netz gibt es keine Sliver (plattgedrückte Tetraeder), was man in Tabelle 8.1 an der Häufigkeit im Bereich von $0,0 - 0,3$ erkennen kann.

Der durchsuchte Frequenzbereich begann bei $f_a = 2400\,MHz$ und endete bei $f_e = 2600\,MHz$. Dabei ergeben sich Wellenlängen von $\lambda_a = 0,124913\,m$ und $\lambda_e = 0,115305\,m$. Die Qualität der Diskretisierung Q_h ist dann im schlechtesten Fall $Q_h = 10,053$. Das heißt, daß eine Wellenlänge im schlechtesten Fall mit ca. 10 Kanten diskretisiert ist.

8.2 Rechteckwellenleiter

8.2.1 Taper

Dieses Beispiel ist im Rahmen des Forschungsprojekts: *Planare lineare Millimeter-Beschleunigungs-strukturen* als Input/Output-Koppeleinrichtung von Herrn R. Merte entworfen worden [52]. Dabei setzte Herrn R. Merte das Finite-Differenzen-Programm GdfidL ein [54].

Die Besonderheit bei dieser Struktur ist, daß das Eingangs- und Ausgangs-Tor unterschiedlich groß sind. Damit ist es bei dieser Struktur nicht möglich, die Methode *reflex* mit der einfachen Anordnung (siehe Abschnitt 7.2.1 S.91) für die Transmissionsfaktor-Berechnung zu verwenden. Weil sich b an S_1 und S_2 unterscheidet und z_2 am Ausgang nicht bekannt ist, muß

Mittlere Qualität	0.841493
Joint-Qualität	0.831747
schlechteste Tetraeder	0.352636
beste Tetraeder	0.998019
Qualitätshäufigkeit im Netz	
0,0-0,1	0
0,1-0,2	0
0,2-0,3	0
0,3-0,4	4
0,4-0,5	36
0,5-0,6	252
0,6-0,7	1389
0,7-0,8	5758
0,8-0,9	11528
0,9-1,0	7299
kleinstes h	$7,0 \cdot 10^{-4}\, m$
größtes h	$1,147 \cdot 10^{-2}\, m$

Tabelle 8.1: Qualität des Netzes vom Problem TEAM 19

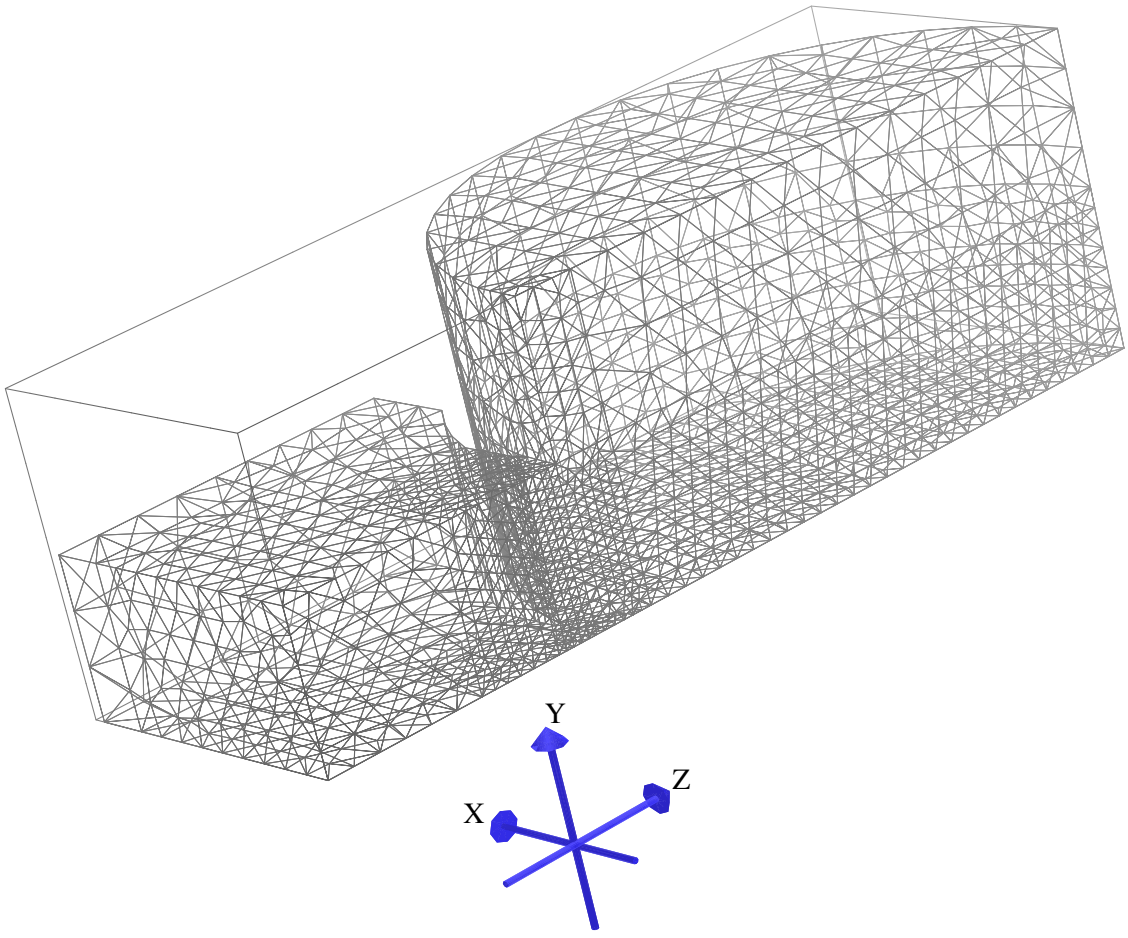


Abbildung 8.2: Randgebiet mit Dreiecken vom TEAM-Problem 19

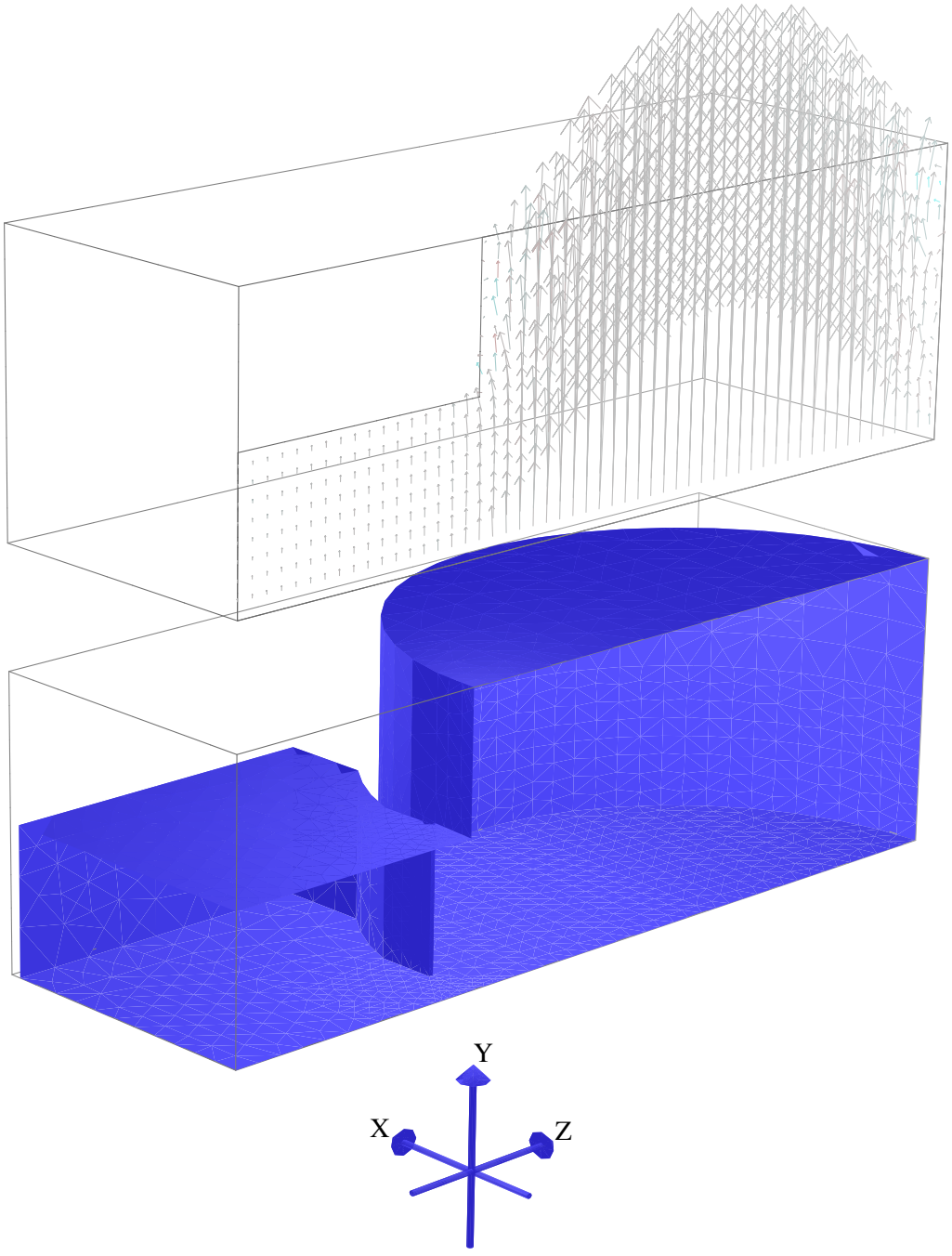


Abbildung 8.3: Elektrisches Feld bei Resonanz im Beispiel TEAM Workshop 19

Numerische Methode	FEM mit Kanten-elementen IFEP	FEM mit Knoten-elementen [50]	Finite Integrations-Methode [51]	Meßwerte [49]
Unbekannte	29484	keine Angabe	ca 240000	–
f_{res} in GHz	$2,542667$	$2,5543$	$2,5448$	$f_1 = 2,54454$ $f_2 = 2,54538$
Δf in MHz bezügl. $\frac{(f_1+f_2)}{2}$	$2,29$	$9,34$	$0,16$	$0,42$
Fehler in % bezügl. $\frac{(f_1+f_2)}{2}$	$0,09$	$0,37$	$0,006$	$0,017$

Tabelle 8.2: Vergleich der berechneten Resonanzfrequenz mit [50], [51] und mit zwei verschiedenen Messungen f_1 und f_2 in [49]

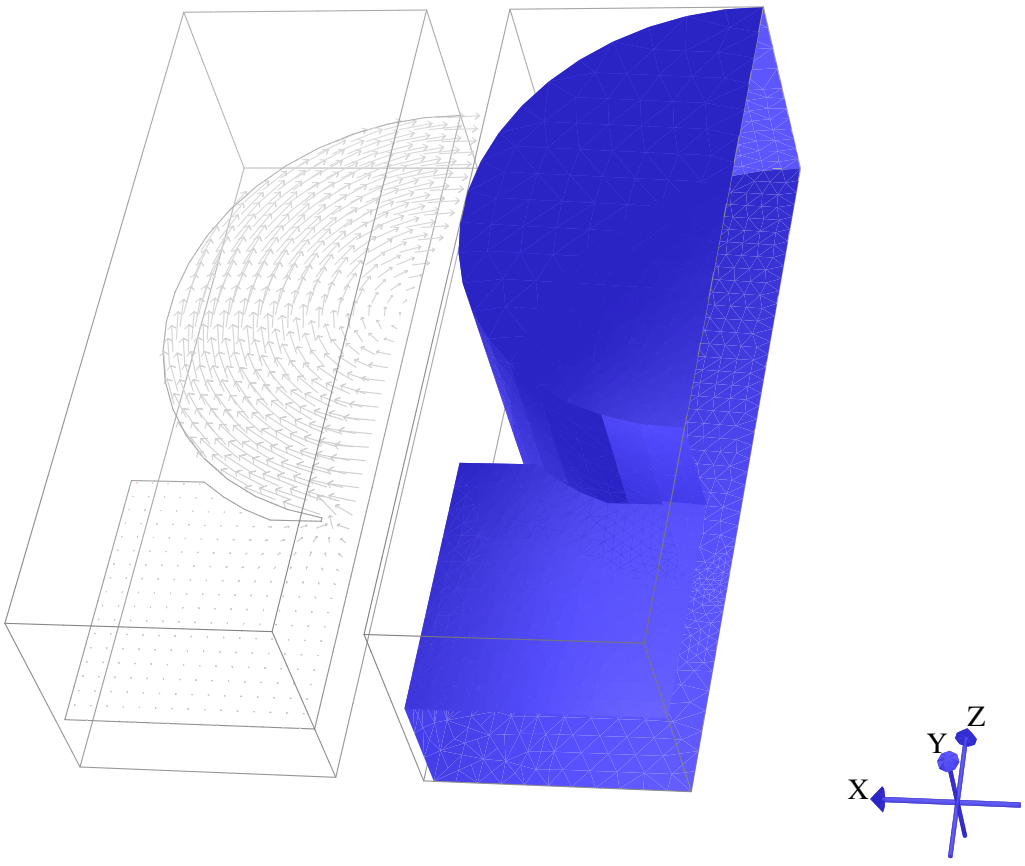


Abbildung 8.4: Magnetisches Feld bei Resonanz im Beispiel TEAM Workshop 19

der Transmisionsfaktor über die Leistung, die durch die Tore tritt, berechnet werden (siehe Abschnitt 7.2.2 S. 94). Man sieht in Abbildung 8.8 auf S. 104 deutlich, wie sich am Ausgang wegen der geringeren Fläche eine größeres E_0 als am Eingang einstellt, weil die eingespeiste Leistung eine geringere Fläche durchdringen muß.

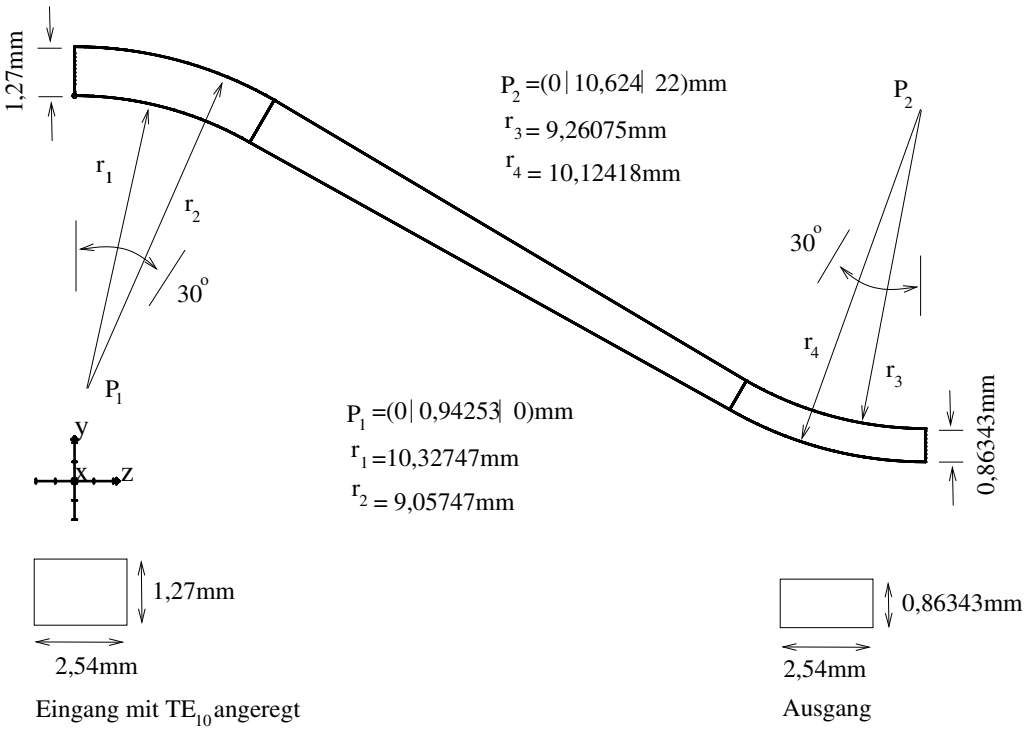


Abbildung 8.5: Technische Abmessungen des Tapers

Das verwendete Netz ist in Abbildung 8.6 zu sehen. Es sind die in der yz-Ebene vorhandenen Symmetrieebenen bei $x = 1,27mm$ mit einer magnetischen Wand ausgenutzt worden. Somit ist das Berechnungsgebiet nur 1/2 so groß. Das verbleibende Netz besteht aus 9925 Knoten, 48350 Tetraedern, 61780 Kanten (davon sind 9846 Kanten erzwungen) und hat folgende Qualität (siehe Tabelle 8.3).

Mit der Arbeitsfrequenz $f = 91,392\,GHz$ und der sich damit ergebenden Wellenlänge $\lambda = 0,00328\,m$ ergibt sich für die Qualität der Diskretisierung Q_h, dann im schlechtesten Fall $Q_h = 6,55$. Das heißt, daß eine Wellenlänge im schlechtesten Fall mit ca. 6 Kanten diskretisiert ist. Im besten Fall ist eine Wellenlänge mit 73 Kanten diskretisiert.

In Abbildung 8.7 ist elektrische Feld in der yz-Ebene bei $x = 1.27mm$ zu sehen. Die halbe Struktur wird durch das GEOP3D-Objekt[1] dargestellt, was die Geometrie für die Netzgenerierung beschreibt.

Bei der Arbeitsfrequenz ergab sich über die Leistung (*power)* ein Reflexionsfaktor $|S_{11}| = 0.0655$ und ein Transmissionsfaktor $|S_{21}| = 0.99647$ mit einem Leistungsbilanzfehler von $1 \cong 0.996936$, was 0.31% Fehler bedeutet. In [52] ist der Leistungsbilanzfehler $1 \cong 1.003$, was einen Fehler von 0.3% bedeutet. Mit der Methode *reflex* ist bei dieser Anordnung nur der Reflexionsfaktor berechenbar. Er beträgt $|S_{11}| = 0.007633$.

[1] Genauso wie es für Kanten FEM-Berechnungen die Klasse FED3D gibt, existiert für die Geometriebeschreibung als Input für den Netzgenerator ein Klasse namens GEOP3D. Siehe auch [18][16]. Für diese Klasse stehen in IFEP auch diverse Methoden zur Verfügung.

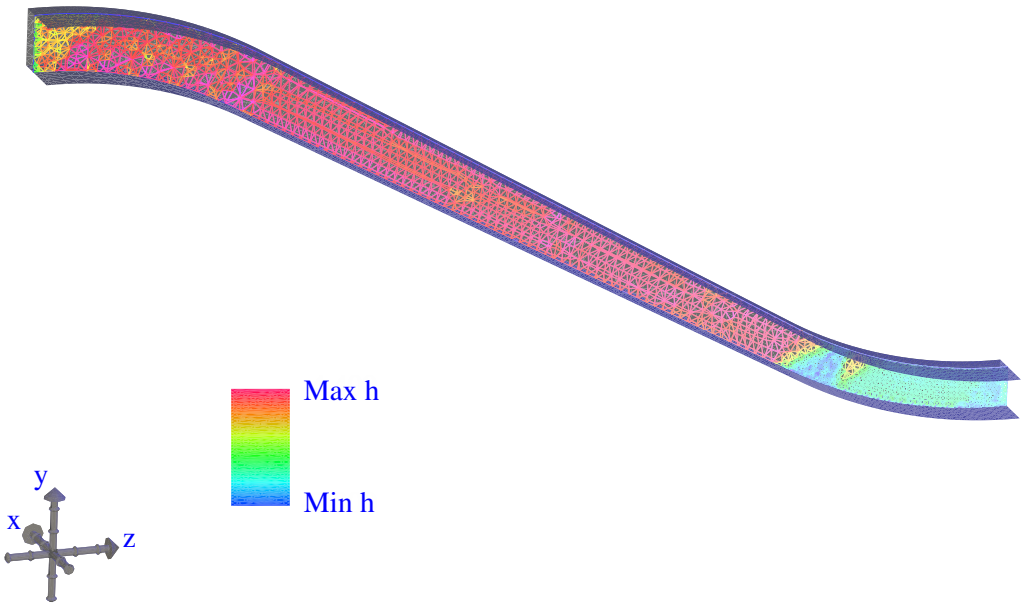


Abbildung 8.6: Qualitative Darstellung der Feinheit des Netzes in einer Schnittebene beim Taper (Balken: unten fein - oben grob)

Mittlere Qualität	0.855663
Joint-Qualität	0.844079
schlechteste Tetraeder	0.367740
beste Tetraeder	0.998287
Qualitätshäufigkeit im Netz	
0,0-0,1	0
0,1-0,2	0
0,2-0,3	0
0,3-0,4	16
0,4-0,5	159
0,5-0,6	664
0,6-0,7	2406
0,7-0,8	8429
0,8-0,9	19161
0,9-1,0	19273
kleinstes h	$4,488 \cdot 10^{-5}\,m$
größtes h	$5,011 \cdot 10^{-4}\,m$

Tabelle 8.3: Qualität des Netzes vom Taper

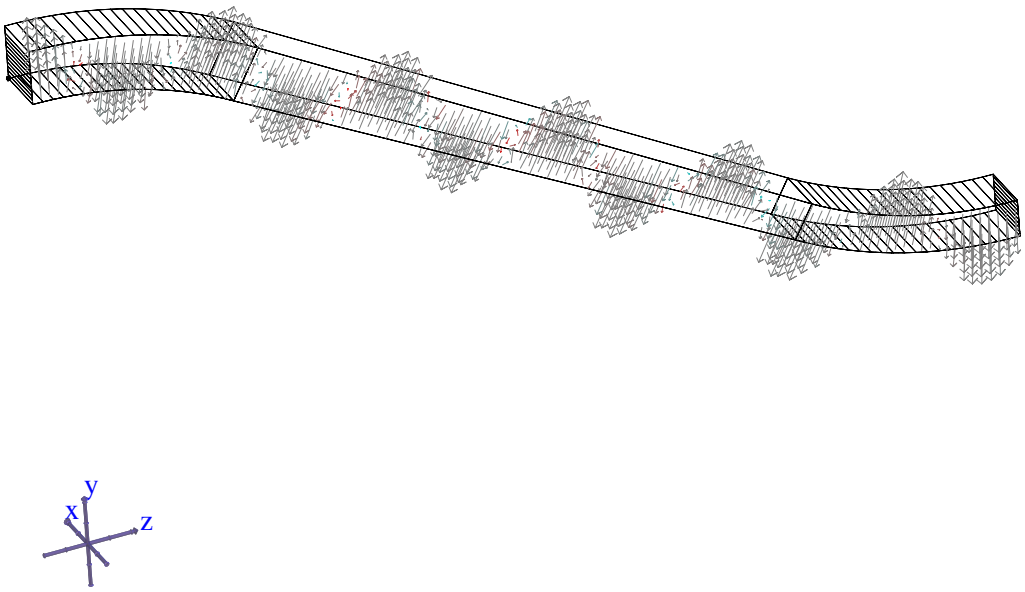


Abbildung 8.7: Elektrisches Feld in der Symmetrie-Ebene des Tapers

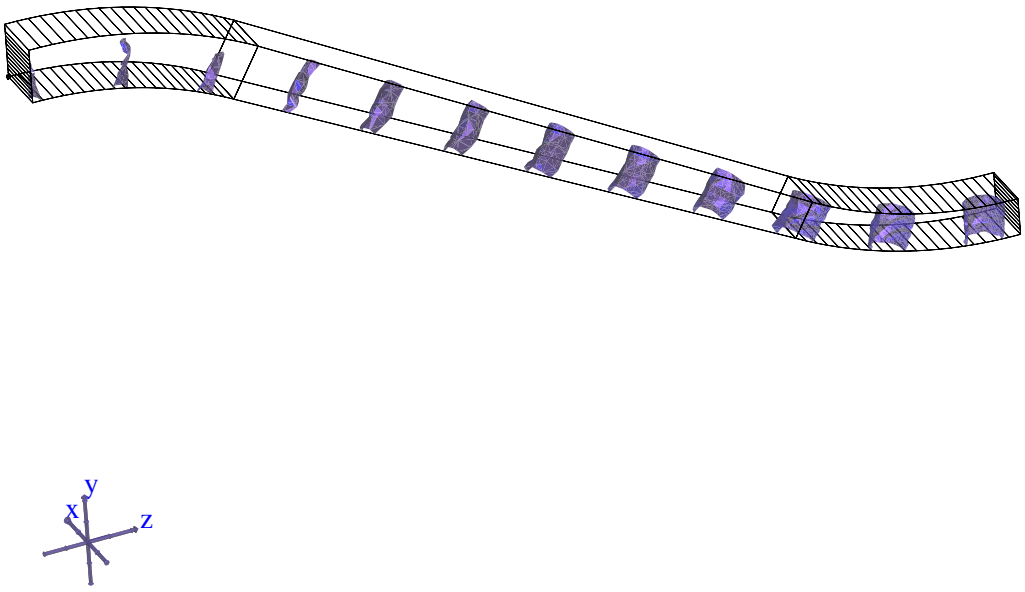


Abbildung 8.8: Level-Darstellung des Betrags der elektrischen Feldstärke beim Taper

Im Stehwellenverhältnis $VSWR$ (engl. voltage standing wave ratio) ausgedrückt, erhält man für den mit der *power*-Methode berechneten Reflexionsfaktor

$$VSWR = \frac{1 + |S_{11}|}{1 - |S_{11}|} = 1,14$$

und über die *reflex*-Methode

$$VSWR = \frac{1 + |S_{11}|}{1 - |S_{11}|} = 1,015\,.$$

Der Wert der *reflex*-Methode besitzt nur eine sehr geringe Abweichung von dem in [52] angegeben Wert von $VSWR = 1.02$. Der Wert der *power*-Methode besitzt dagegen eine größere Abweichung, die auf eine zu hohe Feldlösung, das heißt $P_{in_{ref}}$ ist zu groß, der Referenzberechnung für den ungestörten Wellenleiter schließen läßt.

8.2.2 Wellenleiter mit dielektrischem Störkörper

Bei diesem Beispiel von [55] und [56] befindet sich ein dielektrischer Störkörper in der Mitte des Wellenleiters. An der Grenzfläche des Störkörpers muß die Kontinuitätsbedingung (siehe Abschnitt 5.1.3 auf S.52) erfüllt werden, die mit den Kantenelementen sehr gut abgebildet werden kann.

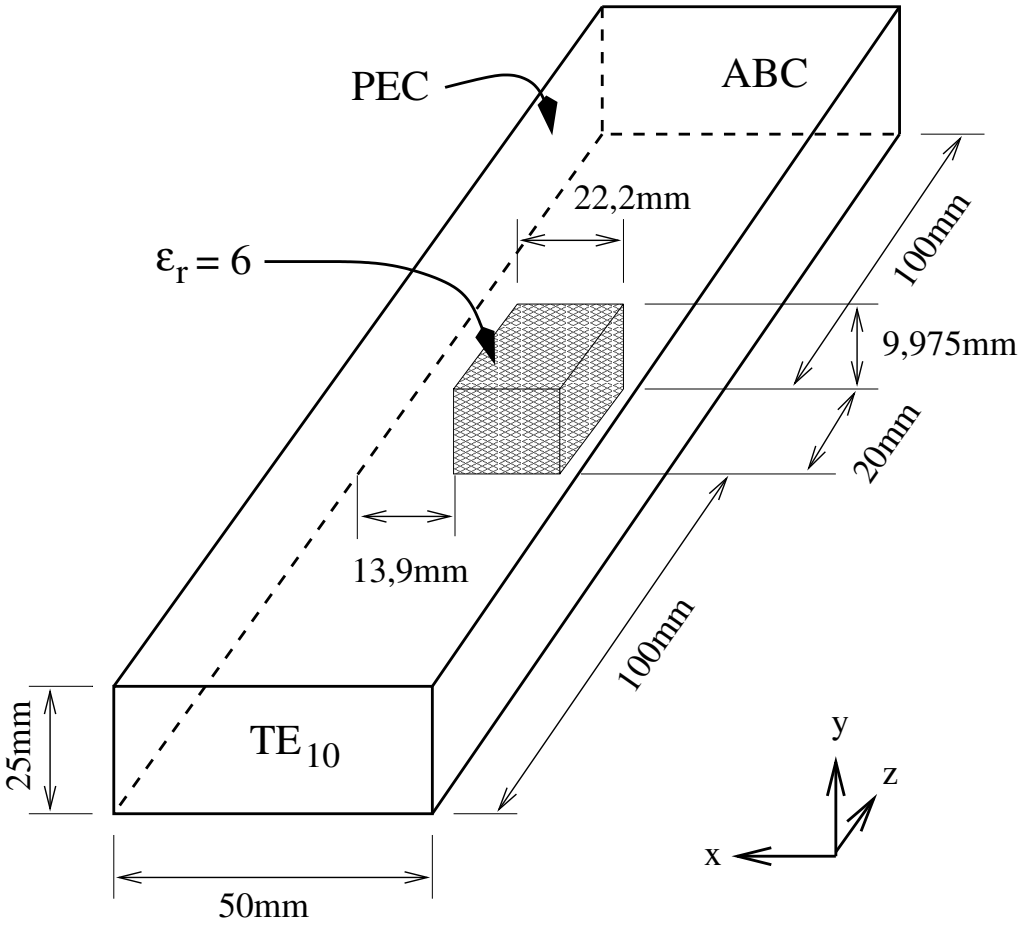


Abbildung 8.9: Technische Abmessungen des Rechteckwellenleiters mit dielektrischem Störkörper

Zu Testzwecken ist hier darauf verzichtet worden, die Symmetrie-Ebene auszunutzen. Im Netz gibt es keine Sliver (plattgedrückte Tetraeder), was man in Tabelle 8.4 an der Häufigkeit im Bereich von $0,0 - 0,3$ erkennen kann.

Der Frequenzbereich, in dem berechnet wurde, erstreckt sich von $f_a = 4,0\,GHz$ bis $f_e = 4,8\,GHz$. Dabei ergeben sich Wellenlängen von $\lambda_a = 0,074948\,m$ und $\lambda_e = 0,062457\,m$. Die Qualität der Diskretisierung Q_h ist dann im schlechtesten Fall $Q_h = 10,2$. Das heißt, daß eine Wellenlänge im schlechtesten Fall mit mehr als 10 Kanten diskretisiert ist.

Da hier beide Methoden zur Berechnung der Streuparameter angewendet werden können, sind in den Abbildungen 8.10 und 8.11 beide gegenübergestellt worden. So bezeichnet (Power) die Berechnung über die Leistung und (reflex) die Berechnung mit der einfachen Anordnung.

Dabei fällt in Abb. 8.10 auf, daß der Reflexionsfaktor von (Power) bei kleinen Werten stark von (reflex) abweicht. Das liegt daran, daß die Leistung der Referenzberechnung $P_{in_{ref}}$ kleiner ist als die Leistung P_{in}, die in den Wellenleiter einfällt. In Formel (7.32) auf S. 95 wird $|S_{11}|^2$dann

Mittlere Qualität	0.845305
Joint-Qualität	0.834452
schlechteste Tetraeder	0,245750
beste Tetraeder	0,999015
Qualitätshäufigkeit im Netz	
0,0-0,1	0
0,1-0,2	0
0,2-0,3	0
0,3-0,4	4
0,4-0,5	32
0,5-0,6	165
0,6-0,7	1348
0,7-0,8	7751
0,8-0,9	20219
0,9-1,0	18831
kleinstes h	$1,556 \cdot 10^{-3}\,m$
größtes h	$6,132 \cdot 10^{-3}\,m$

Tabelle 8.4: Qualität des Netzes vom Rechteckwellenleiter mit dielektrischem Störkörper

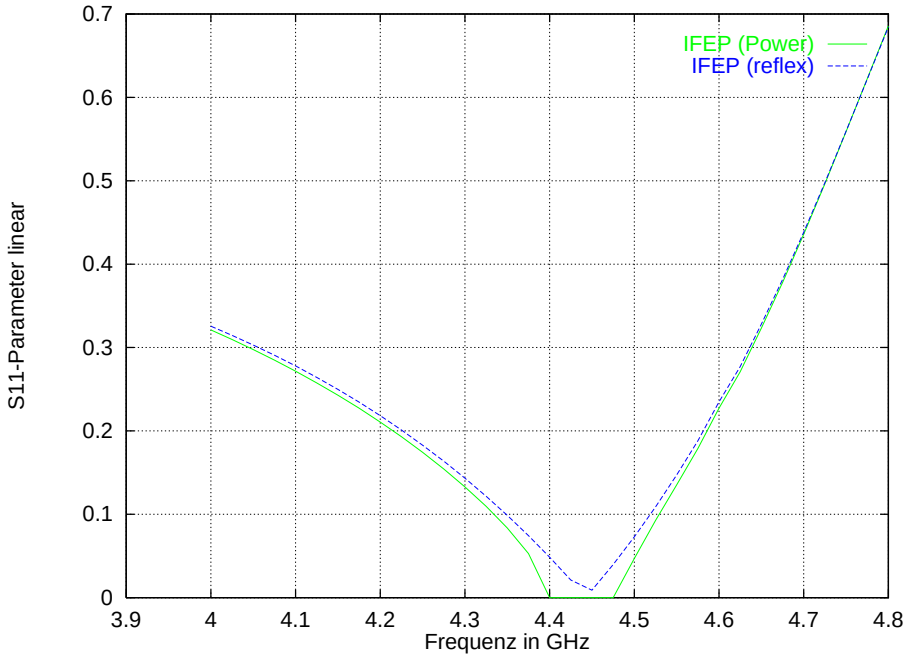


Abbildung 8.10: Vergleich beider Berechnungsmethoden für $|S_{11}|$ beim Rechteckwellenleiter mit dielektrischem Störkörper

kleiner als Null, was nicht sein kann. In Abb. 8.11 sieht man in diesem Frequenzbereich einen Transmissionsfaktor bei (Power), der größer als 1 ist. Das bedeutet, daß man einen Energiegewinn in einem passiven Bauteil hätte, was auch nicht sein kann. Hier ist die Leistung $P_{in_{ref}}$ kleiner als die abgegebene Leistung P_{out} des Wellenleiters. Was mit Formel (7.38) auf S. 95 zu einem Wert größer Null führt.

Die Frage, die sich hier stellt, lautet: *Ist die Leistung in der Referenzrechnung $P_{in_{ref}}$ zu klein oder die Leistung im Wellenleiter P_{in} sowie P_{out} zu groß?*

Dafür wurde eine sehr genaue Referenzrechnung (mit Index g bezeichnet) bei $f^g = 4,4\,GHz$ durchgeführt. Die Geometrie ist in Abb. 6.5 auf S. 82 schon dargestellt worden. Das Netz hat keine Sliver-Elemente, wie alle in dieser Arbeit angegeben Netze. Die Wellenlänge beträgt bei dieser Frequenz $\lambda^g = 0.068135\,m$. Mit $h^g_{max} = 3,068177 \cdot 10^{-3}$ ergibt sich für die Qualität der Diskretisierung $Q^g_h = 22,2$. Das heißt, daß eine Wellenlänge im schlechtesten Fall mit mehr als 22 Kanten diskretisiert wird. Der TE_{10}-Mode wird an der ABC mit einem $E_0 = 10\,V$ angeregt. In der Feldlösung der genauen Referenzrechnung stellt sich ein Wert für $E^g_0 = 10,0093\,V$ ein.

In Tabelle 8.5 sind alle Leistung gegenübergestellt:

Berechnung	P_{in}	P_{out}	Fehler in %
genaue Referenzrechnung	$6,06324 \cdot 10^{-5}$	$6,06312 \cdot 10^{-5}$	$2,0354 \cdot 10^{-3}$
normale Referenzrechnung (Power)	$6,08336 \cdot 10^{-5}$	$6,03889 \cdot 10^{-5}$	$7,31 \cdot 10^{-1}$
Wellenleiter mit Störkörper	$6,08622 \cdot 10^{-5}$	$6,09661 \cdot 10^{-5}$	$1,71 \cdot 10^{-1}$

Tabelle 8.5: Leistungsvergleich mit der genauen Referenzrechnung

Da alle Berechnungen verlustfrei sind und keine aktiven Quellen enthalten, muß die Leistung, die an Tor 1 eintritt, genauso groß sein, wie die Leistung, die den Wellenleiter an Tor 2 wieder verläßt. Daß das in den Berechnungen nicht so ist, sieht man an dem Fehler in %. Man sieht, daß die genaue Berechnung zwei Größenordnung besser ist. Die Leistung ist auf jeden Fall aber kleiner als die Leistung im Wellenleiter mit Störkörper. Damit ist eindeutig, daß die Leistung im Wellenleiter zu groß ist. In diesem Frequenzbereich treten beim Dielektrikum hohe Feldstärken durch Resonanzen auf. Dies führt scheinbar dazu, daß ingesamt die Feldlösung in der Amplitude angehoben wird.

In Abbildung 8.12 auf S. 109 ist der Leistungsbilanzfehler der Streuparameter für beide Methoden zu sehen. Dabei wurde der Bereich mit Transmissionsfaktor > 1 in der Kurve von (*power*) ausgespart.

Der Vergleich der berechneten Reflexionsfaktoren (*reflex)* auf S. 110 in Abbildung 8.13 mit den Ergebnissen von Ise, Inoue und Koshiba [57] (ebenfalls Kantenelemente, Anzahl der Freiheitsgrade unbekannt) zeigt eine sehr gute Übereinstimmung. Die Ergebnisse von Christ und Hartnagel [55] (Finite Differenzen, Anzahl der Freiheitsgrade unbekannt) haben dagegen einen Frequenzversatz zu den Werten von IFEP und [57].

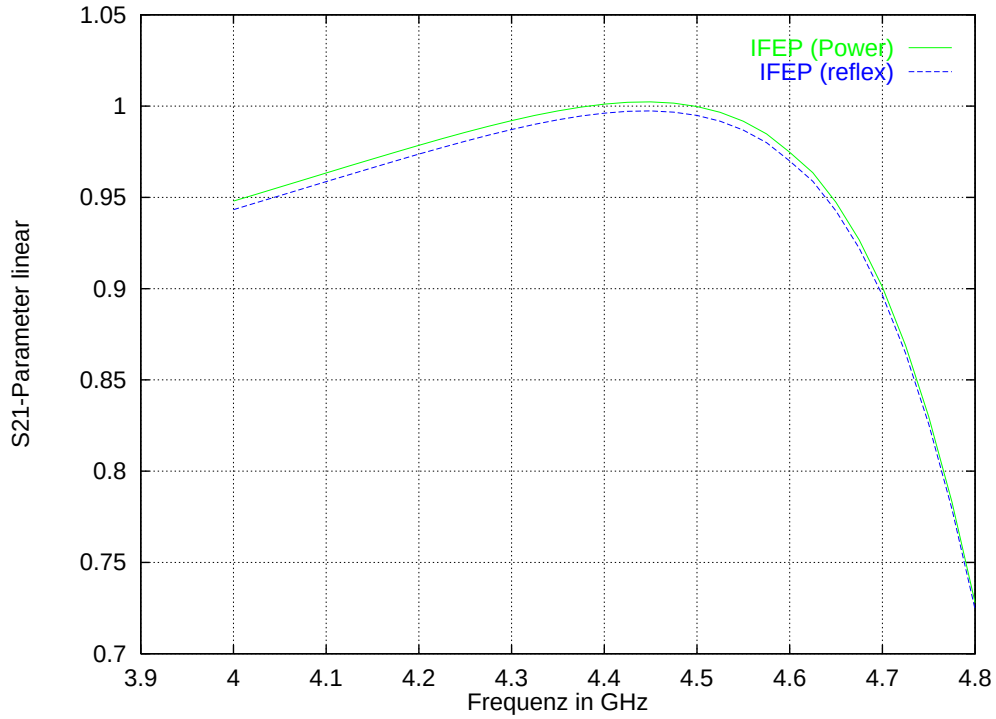


Abbildung 8.11: Vergleich beider Berechnungsmethoden für $|S_{21}|$ beim Rechteckwellenleiter mit dielektrischem Störkörper

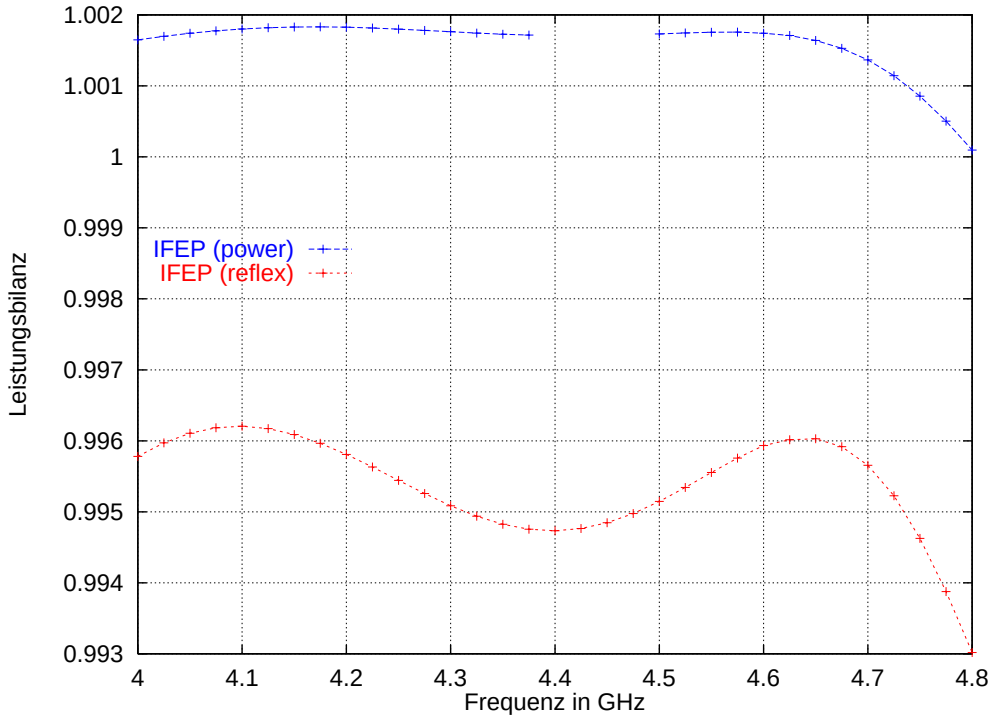


Abbildung 8.12: Vergleich beider Leistungsbilanz-Fehler aus den Streuparametern beim Rechteckwellenleiter mit dielektrischem Störkörper

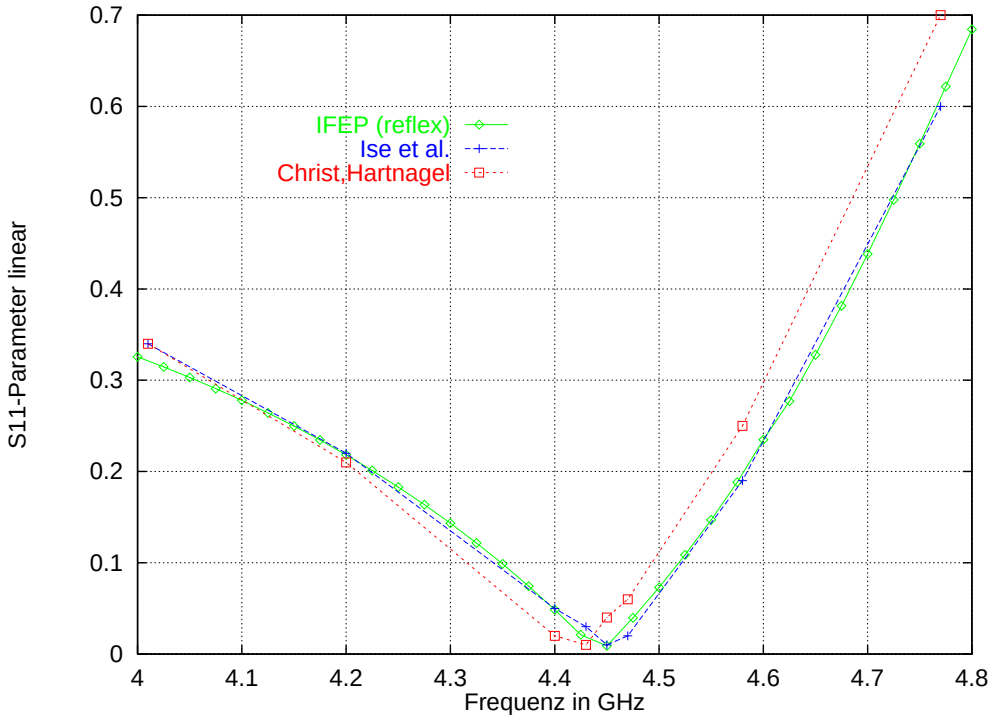


Abbildung 8.13: Vergleich der mit (reflex) berechneten $|S_{11}|$-Werte mit [55] und [57]

8.3 Hohlleiter-Bandpaß

Dieses Beispiel ist für eine Meßeinheit im Rahmen des Linearbeschleuniger Projekts *TESLA* von Herrn S. Sabah entworfen worden [53]. Der Bandpaß soll für die Signalverarbeitung bei einem HF-Empfänger eingesetzt werden, um den Strahlversatz des Elektronenstrahls im Beschleuniger zu messen.

Da dieser Bandpaß auch gebaut und gemessen wurde, steht hier die Vergleichsmöglichkeit mit gemessenen Werten zur Verfügung. In Abbildung 8.14 ist der geöffnete Bandpaß zu sehen. Er ist aus Aluminium gefertigt und wird durch kleine Antennen in der Abdeckplatte betrieben. Die numerische Simulation unterscheidet sich hier. Dabei wurde in der Simulation von einem nach links und rechts unendlich ausgedehnten Hohlleiter R140 nach 153-IEC[2] ausgegangen, der denselben Querschnitt besitzt wie der Bandpaß, der auf der linken Seite mit einem TE_{10}-Mode angeregt und auf der rechten Seite wieder absorbiert wurde (s. Abb. 8.15). Im Unterschied zur Simulation tritt dann bei der Messung eine zusätzlich Dämpfung auf. Sie ist durch nichtideales Material, einen noch verbleibenden Spalt an der Abdeckplatte (abstrahlende Spaltantenne) und die verlustbehaftete Antennenkopplung verursacht.

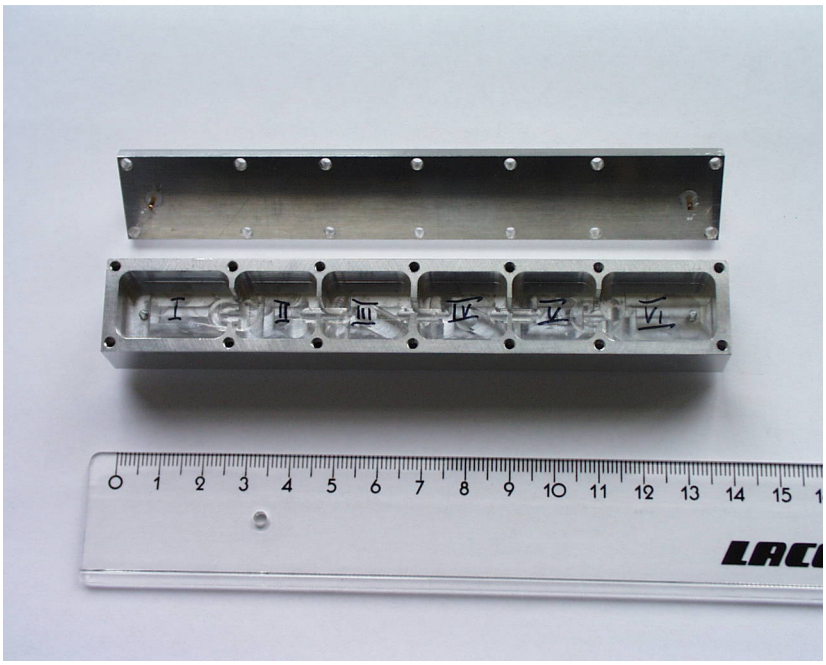


Abbildung 8.14: Realisierter Hohlleiter-Bandpaß von Herrn Sabah Sabah

In dem für die Simulation verwendeten Netz gibt es keine Sliver (plattgedrückte Tetraeder), was man in Tabelle 8.6 an der Häufigkeit im Bereich von $0,0 - 0,2$ erkennen kann.

Der Frequenzbereich, in dem gerechnet wurde, erstreckt sich von $f_a = 11,9\,GHz$ bis $f_e = 12,05\,GHz$. Dabei ergeben sich Wellenlängen von $\lambda_a = 0,025193\,m$ und $\lambda_e = 0,024879\,m$. Dann ergibt sich für die Qualität der Diskretisierung Q_h im schlechtesten Fall $Q_h = 9,32$. Das heißt, daß eine Wellenlänge im schlechtesten Fall mit mehr als 9 Kanten und im besten Fall mit 116 Kanten diskretisiert ist. Für den Frequenzbereich wurden insgesamt 51 Frequenzpunkte berechnet. Dabei brach der Löser bei einer Genauigkeit von $< 1,0 \cdot 10^{-6}$ den Iterationsprozeß ab. Dafür brauchte ein Linux PC mit 200MHz Pentium 26 Std, 53 Min., d.h. ca 32 Min pro

[2]IEC = International Electronical Commission

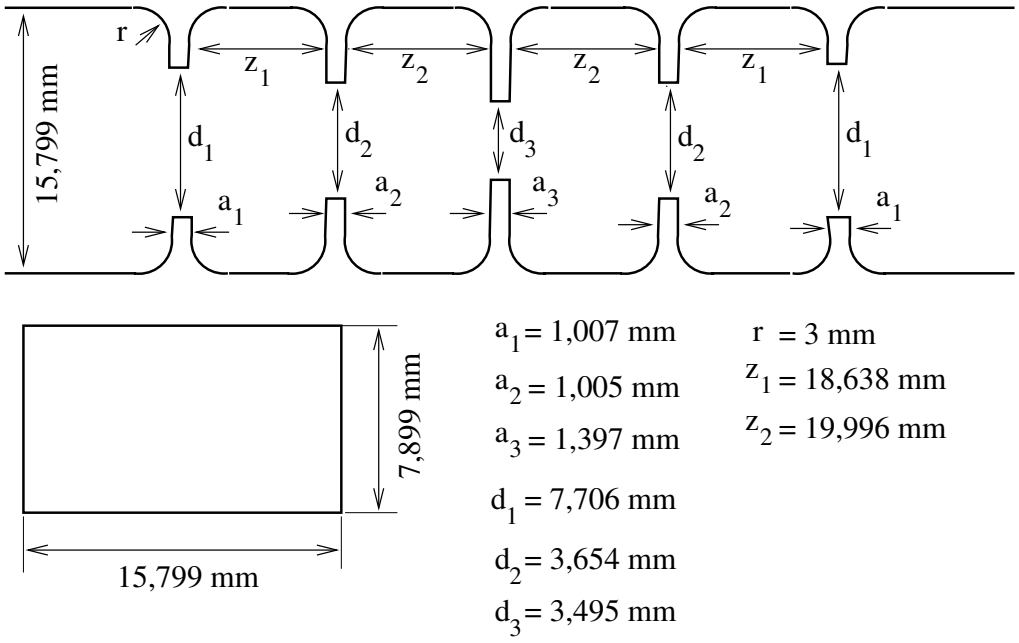


Abbildung 8.15: Technische Maße des Bandpaß-Filters im Ku-Band (12-18 GHz)

Frequenzpunkt. Für jeden Frequenzpunkt wurden beide Methoden (*reflex* und *power*) zur Streuparameterberechung angewendet. Da der geringe Unterschied der Kurven jedoch kaum in den folgenden Grafiken zu erkennen wäre, werden nur die Ergebnisse von *power* dargestellt.

Mittlere Qualität	0.845305
Joint-Qualität	0.834452
schlechteste Tetraeder	0,245750
beste Tetraeder	0,999015
Qualitätshäufigkeit im Netz	
0,0-0,1	0
0,1-0,2	0
0,2-0,3	3
0,3-0,4	16
0,4-0,5	92
0,5-0,6	604
0,6-0,7	3281
0,7-0,8	11281
0,8-0,9	23456
0,9-1,0	17449
kleinstes h	$2,141 \cdot 10^{-4}\,m$
größtes h	$2,668 \cdot 10^{-3}\,m$

Tabelle 8.6: Qualität des Netzes vom Bandpaß-Filter

Beim Vergleich des Reflexionsfaktor S_{11} mit den gemessenen Werten am Netzwerkanalysator 8722C von HP sieht man in Abb. 8.16, daß die Bandbreite bei der Messung größer ist als bei IFEP, jedoch herrscht bei der Mittenfrequenz eine sehr gute Übereinstimmung. Hier hat die Berechnung mit GdfidL einen Frequenzversatz gegenüber der Messung und IFEP. In Abb. 8.17 sieht man am Transmissionsfaktor S_{21} noch besser die sehr gute Übereinstimmung zwischen IFEP und Messung. Auch der Frequenzversatz der Berechnung mit GdfidL[54] ist deutlich zu sehen.

Gdfidl rechnet mit FD-TD[3] im Zeitbereich. Anschließend wird das Ergebnis in den Frequenzbereich transformiert. Anscheinend bewirken die numerischen Fehler im Zeitbereich einen Versatz im Frequenzbereich. Dieser Frequenzversatz ist bei den Ergebnissen von Christ und Hartnagel [55] in Abb. 8.13 auf S. 110 auch zu beobachten.

IFEP rechnet direkt im Frequenzbereich, was die Genauigkeit im Frequenzgang erklären kann. Jedoch bleibt die Frage, warum im Durchlaßbereich der Reflexionsfaktor S_{11} nicht zu $\approx$ Null wird und der Transmissionsfaktor nicht zu ≈ 1. Bei diesen Frequenzen treten in den Kammern des Bandpasses Resonanzen auf. Es scheint, als ob diese starken Felder, mit einer vielfachen Amplitude der Felder an den Toren, die Ergebnisse hier verfälschen. Hohe Amplituden resultieren im Lösungsprozeß auch in großen Fehlern. Dadurch werden Fehler in Gebieten mit geringerer Amplitude auch geringer gewichtet. Die Streuparameterberechnung verwendet jedoch nur Feldlösungen an den Toren, die dann auch größere Fehler haben können.

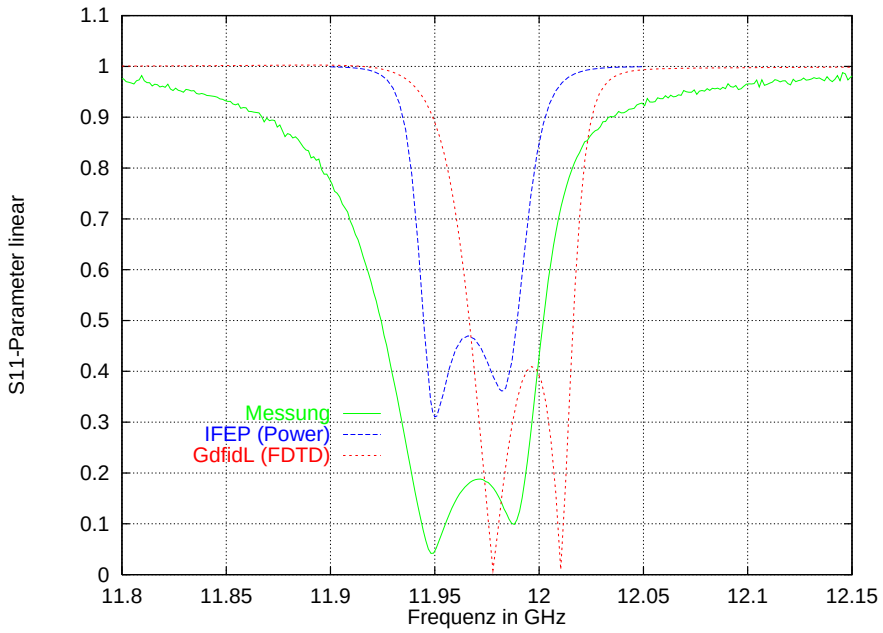


Abbildung 8.16: Vergleich der S_{11}-Parameter beim Bandpaßfilter

Um den Einfluß der Verluste in den Messung abschätzen zu können, wird die Leistungsbilanz der Messung betrachtet. So müßte im verlustfreien Fall

$$|S_{11}|^2 + |S_{21}|^2 = 1 \tag{8.1}$$

gelten. Real sind die Verluste jedoch wesentlich größer, siehe Abbildung 8.18.

Da Transmissionfaktor und Reflexionsfaktoren getrennt gemessen werden, ist es wichtig zu verstehen, wie bei jeder Messung Verluste einfließen können.

[3] Finite Difference-Time Domain

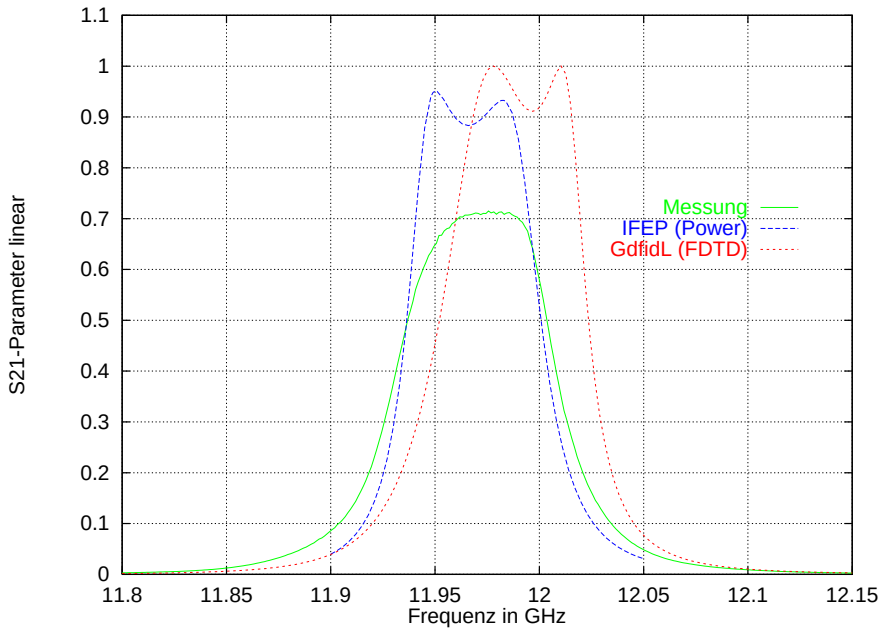


Abbildung 8.17: Vergleich der S_{21}-Parameter beim Bandpaß-Filter

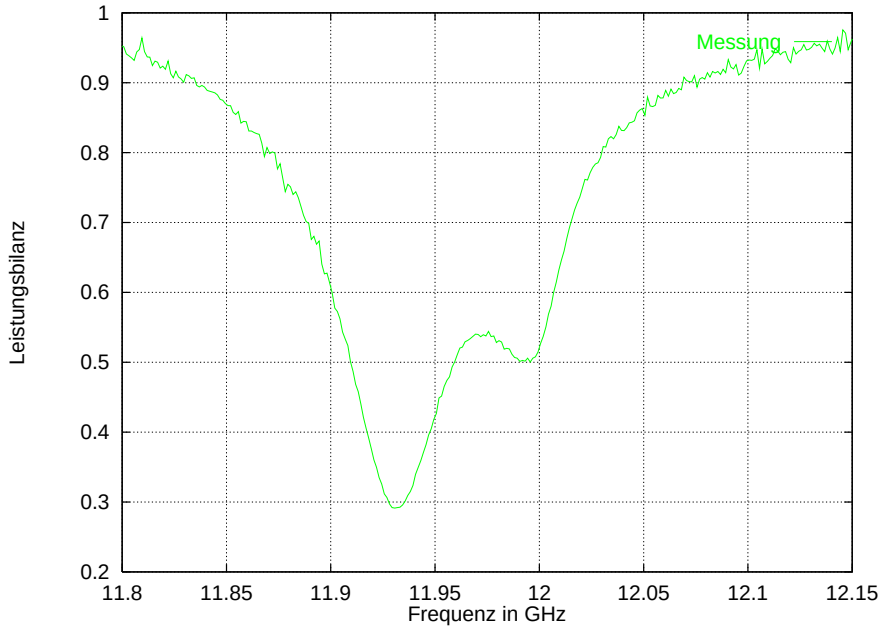


Abbildung 8.18: Leistungsbilanz nach Gl. (8.1) bei der Messung des Bandpaß-Filters

Bei der Transmissionfaktor-Messung wird am Tor 1 eingespeist und an Tor 2 gemessen. Die Verluste sind hier proportional zu den Amplituden der transmittierten Wellen an Tor 2. Aus diesem Grund haben die Verluste in Abbildung 8.17 den größten Einfluß auf die Transmission bei großen Transmissionswerten, also im Durchlaßbereich. Im Durchlaßbereich dämpfen die Verluste jedoch die transmittierte Leistung. Dies erklärt die maximale Transmission von ca. 0,7 statt 1 im Durchlaßbereich. Damit hat man eine Erklärung der Verluste, die durch die Transmissionsfaktor-Messung hervorgerufen werden.

Bei der Reflexionsfaktor-Messung wird nur an Tor 1 gemessen und die gemessenen Werte mit den Messungen verglichen, die beim Kalibrieren des Meßgeräts ohne Meßobjekt aufgenommen wurden. Bei der Messung haben Verluste also immer Einfluß, wenn sich an Tor 1 Felder einstellen, die gedämpft sein können. Das erklärt, warum auch Verluste in der Messung auftreten, auch wenn keine Transmission vorhanden ist (s. Abb. 8.18, 8.16 und 8.17). Die Verluste treten also immer an Tor 1 auf, wenn mit einer einfallenden Welle gespeist wird. Da das Feld durch eine Antenne eingekoppelt wird, treten die Feldverluste immer dann auf, wenn die Antenne ein Feld im Filter anregt.

Zum Vergleich ist die Leistungsbilanz (8.1) beim Bandpass für beide Berechnungsmethoden *power* und *reflex* in Abbildung 8.19 zu sehen.

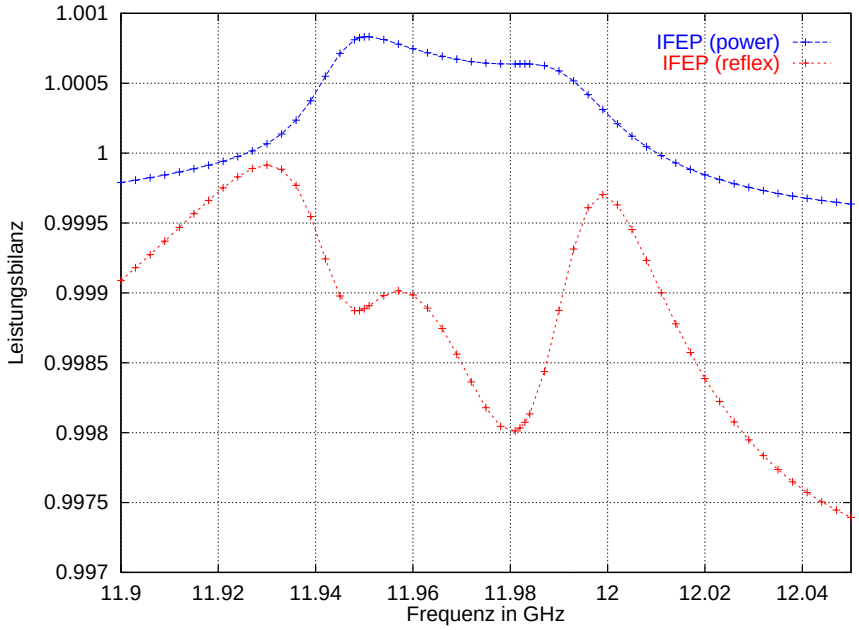


Abbildung 8.19: Vergleich der zwei Methoden zur Streuparameterberechnung an der Leistungsbilanz

Kapitel 9

Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit ist das dreidimensionale FEM-Paket IFEP entwickelt worden, um die vektorielle Helmholtzgleichung im Frequenzbereich zu lösen. Für die Diskretisierung sind Tetraeder mit kantenbasierten Ansatzfunktionen mit der Ordnung $0,5$ verwendet worden.

Dabei stand die Frage im Mittelpunkt, mit welcher Genauigkeit man mit dieser Methode Resonanzfrequenzen und Streuparameter in Mikrowellenstrukturen berechnen kann. Zum Lösen des Gleichungssystems wurden ausschließlich iterative Löser verwendet, die auf ihr Konvergenzverhalten bei unterschiedlichen Problemen untersucht worden sind.

Nach einer Einführung in die Finite-Elemente-Methode in Kapitel 3, wurden die wichtigsten Anforderungen an die Diskretisierung eingeführt, so daß die verwendeten Zerlegungen der Berechnungsgebiete die Anforderungen der Finite-Elemente-Methode voll erfüllen konnten.

Kapitel 4 beschäftigte sich ausführlich mit der Herleitung der kantenbasierten Ansatzfunktionen. Es wurde die Divergenzfreiheit des elektromagnetischen Feldes im Element gezeigt. Desweiteren wurde gezeigt, daß die Ansatzfunktionen im Element rotationskonform sind.

Die konkrete Anwendung der FEM mit Kantenelementen auf die vektorielle Wellengleichung wurde in Kapitel 5 beschrieben. Dabei wurde von der Problembeschreibung über die Einführung der Randbedingungen, der Koninuitätbedingung, und der Formulierung der Wellengleichung als Variationsproblem bis zum Aufstellen des zu lösenden linearen Gleichungssystems alle Schritte der FEM beschrieben, die in IFEP zur Anwendung kommen.

Kapitel 6 behandelte die Aspekte zum Lösen des linearen Gleichungssystems, wobei das effiziente Speichermodell im AWS-Format verwendet wurde. Die Bandbreite der Matrix und ihre Reduktion wurde unter der besonderen Bedingung von Kantenelementen behandelt. Zusätzlich wurde gezeigt, wie das lineare Gleichungssystem durch Elimination der inneren Freiheitsgrade (Kondensation) reduziert werden kann. Anschließend sind die verwendeten iterativen Löser gegenübergestellt worden. Wobei die Gleichungslöser im komplexen wegen ihrer Anwendbarkeit auf alle in dieser Arbeit vorkommenden Probleme genauer untersucht wurden.

Dabei stellte sich heraus, daß das Verfahren der Kondensation den Rechnenaufwand zum Lösen des Gleichungssystem nicht reduziert, sondern trotz des reduzierten Gleichungssystems durch eine voller besetzte Matrix den Aufwand nur erhöht.

Weiter wurde gezeigt, daß das Konvergenzverhalten der iterativen Löser von der Feldlösung abghängig ist. Die verwendeten Löser konvergierten immer erst, wenn die Grundwelle sich im Berechnungsgebiet ausgebreitet hat. Dieser Effekt wurde unabhängig vom angewendeten

iterativem Verfahren nachgewiesen und mit der örtlichen Verteilung der Residuen und ihrer Koppelung im Berechnungsgebiet begründet.

Für den schnellsten hier untersuchten iterativen Löser im komplexen wurde der Lanczos-Löser befunden, der optimiert, im Quellcode und mit volldynamischer Speicherverwaltung in IFEP zur Verfügung steht.

In Kapitel 7 wurden zwei Verfahren vorgestellt, um die Streuparameter bei Wellenleitern zu berechnen. Das allgemeiner anwendbare Verfahren *power*, was die Beträge der Streuparameter über die Leistung an den Toren berechnet, wurde durch das einfachere Verfahren *reflex* in seiner Genauigkeit überprüft. Obwohl das *reflex*-Verfahren nicht so allgemein anwendbar ist, dient es trotzdem durch den analytischen Ansatz der Felderregung und der Verwendung der Othogonalitätseigenschaft als gute Referenz.

Der Verifizierung der Genauigkeit des FEM-Programms IFEP war das Kapitel 8 gewidmet.

So wurde am TEAM Workshop Beispiel 19 gezeigt, daß man mit IFEP eine Resonanzfrequenz mit einer Genauigkeit von $0,09\%$ Abweichung von der gemessenen Frequenz finden kann, wobei die Messung selbst eine Varianz von $0,033\%$ aufwies.

Beim Rechteckwellenleiter Taper ergab sich in der Berechnung des Stehwellenverhältnisses $VSWR$ mit *reflex* eine Abweichung von $0,5\%$ und mit *power* eine Abweichung von $11,8\%$ von dem in [52] veröffentlichem Wert.

Die Berechnung des Rechteckwellenleiters mit dielektrischem Störkörper zeigte, daß die Berechnung mit *power* bei kleinen Reflexionsfaktoren zu größeren Fehlern neigt, was auch schon beim Taper zu beobachten war. Besonders interessant war bei dieser Berechnung der Vergleich mit anderen numerischen Berechnungen aus der Literatur. So hatte das Ergebnis von [55], was mit der FDTD-Methode berechnet worden war, einen Frequenzversatz der Streuparameter. Der Vergleich mit der Berechnung von [57], die auch auf der FEM mit Kantenelementen beruht, zeigte dagegen nur sehr geringe Differenzen.

Die Berechnungen des letzten Beispiels, des Bandpasses, bestätigten diese Beobachtung auch. Durch den Vergleich mit der Messung des gebauten Filters konnte eindeutig gezeigt werden, daß der Frequenzversatz der FDTD zugeordnet werden muß und nicht der FEM mit Kantenelementen.

Damit steht ein FEM-Paket zur Verfügung, mit dem schon einige Problemklassen mit guter Genauigkeit und vertretbarem Rechenaufwand berechnet werden können. Zusätzlich stellt es eine Grundlage für weitere Forschungsarbeiten dar, auf der aufgebaut werden kann. So könnten noch schnellere iterativer Löser entwickelt werden oder höhere Ansatzfunktionen implementiert werden und auf ihre Konvergenzverhalten untersucht werden. Auch sind Weiterentwicklungen im Bereich der absorbierenden Randbedingungen sehr gut vorstellbar.

Anhang A

Vektor- und Integralgleichungen

A.1 Vektorgleichungen

Das Spatprodukt:

$$\left(\vec{a}\times\vec{b}\right)\cdot\vec{c}=\left(\vec{b}\times\vec{c}\right)\cdot\vec{a}=(\vec{c}\times\vec{a})\cdot\vec{b} \tag{A.1}$$

Der Entwicklungssatz:

$$\left(\vec{a}\times\vec{b}\right)\times\vec{c} = (\vec{a}\cdot\vec{c})\cdot\vec{b}-\left(\vec{b}\cdot\vec{c}\right)\cdot\vec{a} \tag{A.2}$$

$$\left(\vec{a}\times\vec{b}\right)\cdot\left(\vec{c}\times\vec{d}\right) = (\vec{a}\cdot\vec{c})\left(\vec{b}\cdot\vec{d}\right)-\left(\vec{a}\cdot\vec{d}\right)\left(\vec{b}\cdot\vec{c}\right) \tag{A.3}$$

Bei Tangentialität mit (A.2):

$$\begin{aligned}\hat{n}\times(\hat{n}\times\vec{a}) &= -\left[(\hat{n}\cdot\hat{n})\ \vec{a}-(\vec{a}\cdot\hat{n})\ \hat{n}\right]\\ &= -\left(\vec{a}-\vec{a}_{\hat{n}}\right)\\ &= -\vec{a}_t\end{aligned} \tag{A.4}$$

A.2 Integralsätze

Erster skalarer Greenscher Satz:

$$\iiint\limits_V\left[a\vec{\nabla}\cdot\left(u\vec{\nabla}b\right)+u\left(\vec{\nabla}a\right)\left(\vec{\nabla}b\right)\right]\,dV=\oiint\limits_S a\,u\frac{\partial b}{\partial n}\,dS \tag{A.5}$$

Zweiter skalarer Greenscher Satz:

$$\iiint\limits_V\left[a\vec{\nabla}\cdot\left(u\vec{\nabla}b\right)-b\vec{\nabla}\cdot\left(u\vec{\nabla}a\right)\right]\,dV=\oiint\limits_S u\left(a\frac{\partial b}{\partial n}-b\frac{\partial a}{\partial n}\right)\,dS \tag{A.6}$$

Erster vektorieller Greenscher Satz:

$$\iiint\limits_V \left[u\left(\vec{\nabla}\times\vec{a}\right)\cdot\left(\vec{\nabla}\times\vec{b}\right)-\vec{a}\cdot\left(\vec{\nabla}\times\left(u\vec{\nabla}\times\vec{b}\right)\right)\right]\,dV$$
$$=\oiint\limits_S u\left(\vec{a}\times\left(\vec{\nabla}\times\vec{b}\right)\right)\cdot\hat{n}\,dS \tag{A.7}$$

Zweiter vektorieller Greenscher Satz:

$$\iiint\limits_V \left[\vec{b}\cdot\left(\vec{\nabla}\times\left(u\vec{\nabla}\times\vec{a}\right)\right)-\vec{a}\cdot\left(\vec{\nabla}\times\left(u\vec{\nabla}\times\vec{b}\right)\right)\right]\,dV$$
$$=\oiint\limits_S u\left(\vec{a}\times\left(\vec{\nabla}\times\vec{b}\right)-\vec{b}\times\left(\vec{\nabla}\times\vec{a}\right)\right)\cdot\hat{n}\,dS \tag{A.8}$$

Anhang B

Baryzentrische Koordinaten

B.1 Dreieckskoordinaten

Summe der baryzentrischen Koordinaten

$$\xi_1 + \xi_2 + \xi_3 = 1 \tag{B.1}$$

Dreiecksfläche

$$A = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} \tag{B.2}$$

Koordinatentransformation: Punkt kartesisch: $(x, y) \rightarrow$ Punkt baryzentrisch: (ξ_1, ξ_2, ξ_3)
Nach (4.1) Seite 32 gilt:

$$\xi_1 = \frac{A_{P,2,3}}{A_{1,2,3}}$$

mit (B.2) ergibt sich:

$$\xi_1 = \frac{\frac{1}{2} \begin{vmatrix} 1 & x_P & y_P \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}}{A} \tag{B.3}$$

Bezeichnet man die Unterdeterminanten mit

$$a_1 = \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}$$

$$b_1 = - \begin{vmatrix} 1 & y_2 \\ 1 & y_3 \end{vmatrix}$$

$$c_1 = \begin{vmatrix} 1 & x_2 \\ 1 & x_3 \end{vmatrix}$$

kann man (B.3) schreiben als

$$\xi_1 = \frac{a_1 + b_1 x + c_1 y}{2A} \tag{B.4}$$

Analog kann mit Zeilenvertauschen[1] für die restlichen Koordinaten schreiben:

$$\xi_2 = \frac{a_2 + b_2 x + c_2 y}{2A} \tag{B.5}$$

$$\xi_3 = \frac{a_3 + b_3 x + c_3 y}{2A} \tag{B.6}$$

Integral über das Produkt von baryzentrische Koordinaten mit den beliebigen Exponenten k, l und m:

$$\iint_A (\xi_1)^k (\xi_2)^l (\xi_3)^m \, dA = 2A \frac{k!\, l!\, m!}{(k+l+m+2)!} \tag{B.7}$$

B.2 Tetraederkoordinaten

Summe der baryzentrischen Koordinaten

$$\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1 \tag{B.8}$$

Tetraedervolumen

$$V = \frac{1}{6} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix} \tag{B.9}$$

Koordinatentransformation: Punkt kartesisch: $(x, y, z) \rightarrow$ Punkt baryzentrisch: $(\xi_1, \xi_2, \xi_3, \xi_4)$
Nach (4.10) Seite 34 gilt:

$$\xi_1 = \frac{V_{P,2,3,4}}{V_{1,2,3,4}}$$

Mit (B.9) ergibt sich dann:

$$\xi_1 = \frac{\frac{1}{6} \begin{vmatrix} 1 & x_P & y_P & z_P \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}}{V}$$

[1] Achtung: Zeilen- oder Spaltenvertauschen bewirkt bei einer Determinante einen Vorzeichenwechsel.

Bezeichnet man die Unterdeterminanten mit

$$a_1 = \begin{vmatrix} x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{vmatrix}$$

$$b_1 = -\begin{vmatrix} 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \\ 1 & y_4 & z_4 \end{vmatrix}$$

$$c_1 = \begin{vmatrix} 1 & x_2 & z_2 \\ 1 & x_3 & z_3 \\ 1 & x_4 & z_4 \end{vmatrix}$$

$$d_1 = -\begin{vmatrix} 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{vmatrix}$$

so kann man ξ_1 schreiben als

$$\xi_1 = \frac{a_1 + b_1 x + c_1 y + d_1 z}{6V} \tag{B.10}$$

Analog kann mit Zeilenvertauschen[2] für die restlichen Koordinaten schreiben:

$$\xi_2 = \frac{a_2 + b_2 x + c_2 y + d_2 z}{6V} \tag{B.11}$$

$$\xi_3 = \frac{a_3 + b_3 x + c_3 y + d_3 z}{6V} \tag{B.12}$$

$$\xi_4 = \frac{a_4 + b_4 x + c_4 y + d_4 z}{6V} \tag{B.13}$$

Dementsprechend gilt für die Gradienten in kartesischen Koordinaten:

$$\vec{\nabla}\xi_i = \frac{b_i}{6V}\vec{e}_x + \frac{c_i}{6V}\vec{e}_y + \frac{d_i}{6V}\vec{e}_z \tag{B.14}$$

Gradient von Koordinatentransformation: Punkt baryzentrisch: $(\xi_1, \xi_2, \xi_3, \xi_4) \rightarrow$ Punkt kartesisch: (x, y, z)

$$x = \sum_{i=1}^{4} \xi_i x_i \qquad y = \sum_{i=1}^{4} \xi_i y_i \qquad z = \sum_{i=1}^{4} \xi_i z_i \tag{B.15}$$

Integral über das Produkt von baryzentrische Koordinaten mit den beliebigen Exponenten k, l, m und n :

$$\iiint_V (\xi_1)^k (\xi_2)^l (\xi_3)^m (\xi_4)^n \, dV = 6V \frac{k!\, l!\, m!\, n!}{(k+l+m+n+3)!} \tag{B.16}$$

[2]Achtung: Zeilen- oder Spaltenvertauschen bewirkt bei einer Determinante einen Vorzeichenwechsel.

Anhang C

Gauss-Legendre-Quadratur

Hier handelt es sich um die numerische Integration einer Funktion f im Dreieck mit der Fläche A. Dabei wird davon ausgegangen, daß die Funktion f von einem bestimmten Polynomgrad ist. Nach [21] gilt dann:

$$\iint_A f(\xi_1, \xi_2, \xi_3)\, dA = A \sum_{i=1}^{N} w_i\, f(\xi_{1_i}, \xi_{2_i}, \xi_{3_i}) \tag{C.1}$$

Mit N Stützstellen i der Funktion f an den Stellen $(\xi_{1_i}, \xi_{2_i}, \xi_{3_i})$ und der Gewichtung w_i der Stützstellen. Je nach Polynomgrad der Funktion f ergeben sich folgende Stützstellen i und Gewichtungsfaktoren w_i:

Polynomgrad	N	i	ξ_{1_i}	ξ_{2_i}	ξ_{3_i}	w_i
1	1	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
2	3	1	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{3}$
		2	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$
		3	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{3}$
3	4	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{27}{48}$
		2	$\frac{11}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{25}{48}$
		3	$\frac{2}{15}$	$\frac{11}{15}$	$\frac{2}{15}$	$\frac{25}{48}$
		4	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{11}{15}$	$\frac{25}{48}$

Tabelle C.1: Gauss-Legendre-Quadraturpunkte und Gewichtungsfaktoren

Index

$O(h^2)$, 44
Z_0, 46
$\nabla\xi_i$, 33, 34, 123
π, 5
$\vec{J}$, 46
h, 44
k_0, 46
nnz, 76

Ausbreitungskonstante im freien Raum, 46

Bandbreite von A, 74
Baryzentrische Koordinaten, 5, 32

CAD, 7
CAE, 7
CAM, 7
Compressed Sparse Column, 72
Compressed Sparse Row, 72
CSC, 72
CSR, 72

Dirichletsche Randbedingung, 21, 47
Divergenz, 42
divergenzkonform, 43
DXF-Datei, 24

EDM, 7
Elektrische Feldkonstante, 5
essentielle Randbedingung, 21

FED3D, 89
Fehlermaß, lokales, 44
Flächennormale, 5
Flächennormalen, 24
Frequenzbereich, 45
Funktional, 58

GEOMPACK, 23
GEOP3D, 102
Gradient von ξ_i, 33, 34, 123

h, 12

homogene Cauchy RB, 22
homogene Cauchysche Randbedingung, 52
homogene Neumann RB, 22

IBC, 47
IEC, 111
IFEP, 89
Impedance Boundary Condition, 47
inhomogene Cauchy RB, 22, 51
inhomogene Neumann RB, 22

Kantenenden, 76
Kantennumerierung, 74
Kantenzuordnung im Dreieck, 36
Kantenzuordnung im Tetraeder, 40
Klasse, 89
Knotenelemente, 13, 31
Knotennumerierung, 76
komplexes Funktional, 58
kondensierte Steifigkeitsmatrix, 77
konsistent, 44
Kreisfrequenz, 45
Ku-Band, 112

Leistung, 95
Lichtgeschwindigkeit, 5
lokale Ansatzfunktionen, 18

Magnetische Feldkonstante, 5
Magnetische Feldstärke, 95
magnetische Wand, 47
Matrixaufbau, 72
Maxwellsche Gleichungen, 45
Mixed Elements, 31, 43
Modified Sparse Row, 72
MSR, 72

natürliche Randbedingung, 22, 47
Neumannsche Randbedingung, 22
numerische Integration, 125

Objekt, 89

PEC, 47
Perfect Electric Conductor, 47
polynom-komplett, 31
Polynomordnung, 43
Power, 106
Poynting Vektor, 94

Quauss-Legendre-Quadratur, 125

Randbedingungen, 21
Reduktion des Gleichungssystems, 77
reflex, 106
Ritzsche Verfahren, 16
Rotation, 42
rotationskonform, 43

selbstadjungiert, 16–18, 54
Speicherbedarf, 71
STEP, 7
Stetigkeit, 14
Stromdichte, 46

Verhältniskoordinaten, 32
Verlustbehaftet elek. Wand, 47
Verschiebungsstromdichte, 65
Volumenbeschreibung, 23

Wellenleiter, 48
Wellenwiderstand, 5
Wellenwiderstand im freien Raum, 46
Wellenzahl, 91
Whitney, 31

zeitharmonisch, 45
Zwangsbedingung, 21, 47

Abbildungsverzeichnis

1.1 Computergestützter Entwicklungsprozeß 7

2.1 Schwingung einer Saite 11
2.2 Qualitative Lösung $u_{FEM}(t,x)$ der Grundschwingung mit 4 finiten Elementen, die als Ansatzfu
2.3 Erste Oberschwingung der Saite mit 4 finiten Elementen, die als Ansatzfunktionen Geraden ha
2.4 Einfach zusammenhängendes Berechnungsgebiet Ω mit Rand Γ und Oberflächennormale "705E
2.5 2-Dimensionale knotenbasierte Ansatzfunktionen mit linearem Polynom-Grad (Ebene) bei eine

3.1 Volumenbeschreibung des Hohlleiter-Bandpasses von Abschnitt 8.3 24
3.2 Polyeder eines Kopfes, der von GEOMPACK nicht in konvexe Gebiete zerlegt werden konnte.

4.1 Baryzentrische Koordinate ξ_1 im Dreieck 32
4.2 Gradienten "017$E\nabla\xi_i$ im Dreieck 33
4.3 Kantenzuordnung im Dreieck 36
4.4 Die Funktionen ξ_1 und ξ_2 entlang der Kante von P_1 nach P_2 37
4.5 Vektorfunktion "017EN_1 38
4.6 Vektorfunktion "017EN_2 38
4.7 Vektorfunktion "017EN_3 39
4.8 Stetigkeit der Tangentialen beim Elementübergang 39
4.9 Beispiel für ein Vektorfeld "017EE^e im Dreieck 40
4.10 Kantenzuordnung im Tetraeder 41

5.1 Wellenleiter mit TE_{10}-Mode Anregung 49
5.2 Aufstellen des Gesamtgleichungssystems 67
5.3 Vorbereiten des Netzes 67
5.4 Speicher für das Gesamtgleichungssystem holen 68
5.5 Orientieren der Elementmatrixeinträge nach der globalen Kantenrichtung . . . 68
5.6 Einarbeiten des Elementgleichungssystems in die Gesamtmatrix 68

6.1 CSR-Speicherformat einer unsymmetrischen Matrix 73

6.2 AWS-Speicherformat einer symmetrischen Matrix 74
6.3 Matrix A einer FEM-Berechnung mit $n = 12318$ 77
6.4 Die vertauschte Matrix von A^{Sym} 78
6.5 Kurzer Rechteckwellenleiter 82
6.6 Langer Rechteckwellenleiter 83
6.7 Konvergenzverhalten vom Lanczos-Löser beim Wellenleiter *kurz* und *lang* . . . 84
6.8 Konvergenzverhalten vom linbcg-Löser bei Wellenleiter *kurz* und *lang* 85
6.9 Konvergenzverhalten vom Lanczos- und linbcg-Löser beim kurzen Wellenleiter 85
6.10 Konvergenzverhalten vom Lanczos- und linbcg-Löser beim langen Wellenleiter 86
6.11 Konvergenz bei Problemen mit 1/4- , 2-, 5 3/4-fachem der Wellenlänge mit Lanczos-Löser 86
6.12 Konvergenz vom Lanczos-Löser mit und ohne Kondensation beim langen Wellenleiter 87
6.13 Konvergenzverhalten des Lanczos-Löser beim TEAM Problem 19 von Abschnitt 8.1 bei der Resonanzfr

8.1 Technische Abmessungen des TEAM Problems 19 (alle Maße sind mm) 98
8.2 Randgebiet mit Dreiecken vom TEAM-Problem 19 99
8.3 Elektrisches Feld bei Resonanz im Beispiel TEAM Workshop 19 100
8.4 Magnetisches Feld bei Resonanz im Beispiel TEAM Workshop 19 101
8.5 Technische Abmessungen des Tapers 102
8.6 Qualitative Darstellung der Feinheit des Netzes in einer Schnittebene beim Taper (Balken: unten fein -
8.7 Elektrisches Feld in der Symmetrie-Ebene des Tapers 104
8.8 Level-Darstellung des Betrags der elektrischen Feldstärke beim Taper 104
8.9 Technische Abmessungen des Rechteckwellenleiters mit dielektrischem Störkörper 106
8.10 Vergleich beider Berechnungsmethoden für $|S_{11}|$ beim Rechteckwellenleiter mit dielektrischem Störkörp
8.11 Vergleich beider Berechnungsmethoden für $|S_{21}|$ beim Rechteckwellenleiter mit dielektrischem Störkörp
8.12 Vergleich beider Leistungsbilanz-Fehler aus den Streuparametern beim Rechteckwellenleiter mit dielekt
8.13 Vergleich der mit (reflex) berechneten $|S_{11}|$-Werte mit [55] und [57] 110
8.14 Realisierter Hohlleiter-Bandpaß von Herrn Sabah Sabah 111
8.15 Technische Maße des Bandpaß-Filters im Ku-Band (12-18 GHz) 112
8.16 Vergleich der S_{11}-Parameter beim Bandpaßfilter 114
8.17 Vergleich der S_{21}-Parameter beim Bandpaß-Filter 115
8.18 Leistungsbilanz nach Gl. (8.1) bei der Messung des Bandpaß-Filters 115
8.19 Vergleich der zwei Methoden zur Streuparameterberechnung an der Leistungsbilanz116

Literaturverzeichnis

[1] GRAPE Manual,
Version 5.3, Sonderforschungsbereich 256, Institut für Angewandte Mathematik, Uni-Bonn, Deutschland

[2] D.Thomas, J. van Maanen, M.Mead,
Specification for Exchange of Product Analysis Data,
Esprit Project 322 „CAD*I", Berlin, Springer-Verlag, 1989

[3] TGL 44640,
Datenschnittstelle für Finite-Elemente-Modelle,
Amt für Standardisierung, Berlin, und Institut für Festkörpermechanik, Magdeburg, 1989

[4] ISO 10303, Part 11,
EXPRESS Language Reference Manual,
International Standardization Organisation, 1994

[5] ISO 10303, Part 41-46,
Integrated Generic Ressorces,
International Standardization Organisation, 1995

[6] ISO 10303, Part 101-105,
Integrated Application Ressorces,
International Standardization Organisation, 1994

[7] N.Bronstein, K.A.Semendjajew:
Taschenbuch der Mathematik,
Verlag Harri Deutsch, 22. Auflage, 1985, ISBN 3-87144-492-8

[8] N.Bronstein, K.A.Semendjajew:
Taschenbuch der Mathematik, Ergänzende Kapitel,
BSB B.G. Teubner Verlagsgesellschaft, 4. Auflage, Leipzig, 1997

[9] David S. Burnett:
Finite Element Analysis,
Addison-Wesley, 5. Auflage, 1991, ISBN 0-201-10806-2

[10] Károly Simony:
Theoretische Elektrotechnik,
Barth Verlag, 10. Auflage, 1993, ISBN 3-335-00375-6

[11] Günther Lehner:
Elektromagnetische Feldtheorie für Ingenieure und Physiker,
Springer Verlag, 2.Auflage, Berlin, ISBN 3-540-56873-5

[12] K. Ho-Le:
Finite element mesh generation Methods: a review and classification,
computer-aided-design, Vol. 20, No. 1, S.27-38 Jan./Feb. 1988

[13] J. C. Cavendish, D. A. Field, W. H. Frey:
An approach to automatic three-dimensional finite element generation,
Int. J. Numer. Methods Eng. 21, S. 329-347, 1985

[14] N. A. Golias, T. D. Tsiboukis:
An Approach to Refining Three-Dimensional Tetrahedral Meshes Based on Delaunay Transformations,
Int. J. Numer. Meth. Eng., Vol. 37, S. 793-812, 1994

[15] N. A. Golias, T. D. Tsiboukis, E. E. Kriezis:
Automatic finite element analysis: Application to the shielding by a spherical shell,
Archiv f. Elektrotechnik 77, S. 85-93, Springer-Verlag 1994

[16] Barry Joe:
GEOMPACK Users' Guide,
Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1

[17] Thorsten Kowalski:
Entwicklung eines FEM-Programms mit Kantenelementen in der Hochsprache C,
TU-Berlin, Diplomarbeit im Fachgebiet Theoretische Elektrotechnik, 1997

[18] Klaus Schneider:
Konstruktion von Geometrien unter AUTOCAD R12 für die Netzgenerierung bei FEM-Berechnungen,
TU-Berlin, Studienarbeit im Fachgebiet Theoretische Elektrotechnik, 1997

[19] Klaus Schneider:
Entwicklung einer parallelisierten Version des FEM-Programms IFEP,
TU-Berlin, Diplomarbeit im Fachgebiet Theoretische Elektrotechnik, 1997

[20] Christian-Alexander Bunge:
Entwicklung und Implementation eines Verfahrens zur Verbesserung der Tetraedernetz-Eigenschaften für die Finite-Elemente-Berechnung im Programm IFEP,
TU-Berlin, Studienarbeit im Fachgebiet Theoretische Elektrotechnik, 1997

[21] G.R. Cowper:
Gaussian Quadrature Formulas for Triangles,
Int. J. Numer. Meth. Eng., Vol. 7, S. 405-408, 1973

[22] Hans Rudolf Schwarz:
Methode der finiten Elemente,
3. neubearb. Aufl., Stuttgart, Teubner, 1991, ISBN 3-519-22349-X

[23] W. Ritz:
Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik, J.f. reine und angew. Math. 135, 1909, S.1-61

[24] Jiaming Jin:
The Finite Element Methode in Electromagnetics,
New York, John Wiley & Sons, 1993, ISBN 0-471-58627-7

[25] T. B. A. Senior:
Impedance boundary conditions for imperfectly conducting surface,
Appl. Sci. Res. B, Vol. 8, S. 418-436, 1960

[26] Andrew F. Peterson:
Absorbing boundary conditions for the vector wave equation,
Microwave Opt. Tech. Lett., Vol. 1 No. 2, S. 62-64, April 1988

[27] E. Cuthill:
Several strategies for reducing the band width of matrices.
In: D. J. Rose, R.A.Willoughby (ed.): Sparse matrices and their applications, Plenum, New York, S. 157-166, 1972

[28] E. Cuthill, J. McKee:
Reducing the Bandwidth of sparce symmetric matrices,
In: Proc. ACM Nat. Conf., New York, S. 157-172, 1969

[29] Arnulf Kost:
Numerische Methoden in der Berechnung elektromagnetischer Felder,
Berlin, Springer Verlag, 1994, ISBN 3-540-55005-4

[30] Hassler Whitney:
Geometric Integration Theory,
Princton, New Jersey: Princton University Press, 1957

[31] J. C. Nédélec:
Mixed Finite Elements in $\mathbb{R}^3$, Numerische Mathematik 35, S. 315-341, Springer Verlag, 1980

[32] Alain Bossavit:
A Rationale for "Edge-Elements" in 3-D Fields Computations,
IEEE Transactions on Magnetics, Vol. 24, No. 1, S. 74-79, January 1988

[33] J. P. Webb:
Edge Elements and What They can do for You,
IEEE Transactions on Magnetics, Vol. 29, No. 2, S. 1460-1465, March 1993

[34] Gerrit Mur:
Edge Elements, their Advantages and their Disadvantages,
IEEE Transactions on Magnetics, Vol. 30, No. 5, S. 3552-3557, March 1993

[35] J.P.Webb, P Forghani:
Hierarchal Scalar and Vector Tetrahedra,
IEEE Transactions on Magnetics, Vol. 29, No. 2, S. 1495-1498, March 1993

[36] R. Graglia, D. Wilton, and A. F. Peterson:
Higher Order Interpolatory Vector Bases for Computational Electromagnetics,
IEEE Trans. Antennas Propagat., Vol. 45, No. 3, S.329-342, March 1997

[37] Lars S. Andersen and John L. Volakis:
Hierarchical Tangential Vector Finite Elements for Tetrahedra,
IEEE Microwave and Guided Wave Letters, Vol. 8, No. 3, S.127-129, March 1998

[38] Jo-Yu Wu and Robert Lee:
The Advantages of Triangular and Tetrahedral Edge Elements for Electromagnetic Modelling with the Finite-Element Method,
IEEE Trans. Antennas Propagat., Vol. 45 No. 9, S.1431-1437, September 1997

[39] Meinke, Gundlach:
Taschenbuch der Hochfrequenztechnik, Band 1: Grundlagen,
Springer-Verlag, Berlin, 1992, ISBN 3-540-54714-2

[40] Meinke, Gundlach:
Taschenbuch der Hochfrequenztechnik, Band 2: Komponenten,
Springer-Verlag, Berlin, 1992, ISBN 3-540-54715-0

[41] Erich Pehl:
Mikrowellentechnik, Band 1: Wellen und Leitungsbausteine,
Hüthig Verlag, Heidelberg, 1988, ISBN 3-7785-1611-6

[42] Martin Paul:
Schaltungsanalyse mit S-Parametern,
Dr. Alfred Hüthig Verlag Heidelberg, 1977

[43] Wolfgang Hackbusch:
Iterative Lösung großer schwachbesetzter Gleichungssysteme,
Teubner Studienbücher, Stuttgart, 1991, ISBN 3-519-02372-5

[44] Are Magnus Bruaset:
A survey of preconditioned iterative methods,
Longman Scientific & Technical, Essex, 1995, ISBN 0-582-27654-3

[45] Yousef Saad:
Iterative Methods for Sparse Linear Systems,
PWS Publishing Company, Boston, 1995, ISBN 0-534-94776-X

[46] William H. Press:
Numerical recipes in C: The Art of scientific computing,
Cambridge University Press, Cambridge, 1992, ISBN 0-521-43108-5

[47] P. Deuflhard:
A Study of Lanczos-Type Iterations for Symmetric Indefinite Linear Systems,
Konrad-Zuse-Zentrum für Informationstechnik Berlin, Preprint SC 93-6, März 1993.

[48] P. Deuflhard, A. Hohmann:
Numerische Mathematik. Eine algorithmisch orientierte Einführung.
Kap. 6: Drei-Term-Rekursionen. Verlag de Gruyter Berlin, New York, 1991.

[49] A.Bossavit, J.F. Lamaudière, B. Maestrali:
Team Workshop Problem 19, International Team Workshop, 1992

[50] Xiaoming Xu, Hai Zhou,
Solving TEAM Workshop Problem 19 using Nodal Elements and Magnetic Vector Potential,
Proceedings of the Fifth International TEAM Worshop, S. 46, Berlin, Germany, 1995, ISBN 3-7983-1665-1

[51] R. Ehmann, M. Burkhardt, H. Wolter, T. Weiland:
Elektromagnetic Simulation of a Loaded Cavity,
Proceedings of the Fifth International TEAM Worshop, S. 42, Berlin, Germany, 1995, ISBN 3-7983-1665-1

[52] Rolf Merte:,
Improved Shape of an Input/Output-Waveguide with Integrated Taper for W-Band Muffin-Tin WBAND-003,
TET-Note 97-18, TU-Berlin, Institut für Theoretische Elektrotechnik und Nachrichtentechnik, Sek. EN-2, Einsteinufer 17, D-10587 Berlin, Germany, 1997

[53] Sabah Sabah:,
Entwurf und Aufbau eines Hohraumresonators mit HF-Elektronik zur Messung und Auswertung der transversalen Strahlposition im Linearbeschleuniger, Dissertation, TU-Berlin, Germany, 2000

[54] W.Bruns:
GdfidL: A finite difference program for arbitrarily small perturbations in rectangular geometries,
IEEE Trans. Magn. vol. 32, No. 3, May 1996

[55] A. Christ, H.L. Hartnagel:
Three-Dimensional Finite-Difference Method for the Analysis of Microwave-Device Embedding,
IEEE Trans. Microwave Theory Tech., vol. MMT-35, No. 8, S. 688-696, Aug. 1987

[56] K. Ise, K. Inoue, and M. Koshiba:
Three-Dimensional Finite-Element Solution of Dielectric Scattering Obstacles in a Rectangular Waveguide,
IEEE Trans. Microwave Theory Tech., vol. MMT-38, S. 1352-1359, Sept. 1990

[57] K. Ise, K. Inoue, and M. Koshiba:
Three-Dimensional Finite-Element Method with Edge Elements for Electromagnetic Waveguide Discontinuities,
IEEE Trans. Microwave Theory Tech., vol. MMT-39, No. 8, S. 1289-1295, Aug. 1991