Efficient finite element electromagnetic analysis of antennas and microwave devices: the FE-BI-FMM formulation and a posteriori error estimation for p adaptive analysis.

Matthys Michiel Botha

Dissertation presented for the
Degree of Doctor of Philosophy in Engineering
at the University of Stellenbosch.

Promoter: Prof. David B. Davidson

Stellenbosch, South Africa, September 2002

Declaration

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Summary

This document presents a Galerkin FE formulation for the full-wave, frequency domain, electromagnetic analysis of three dimensional structures relevant to microwave engineering, together with the investigation of two techniques to enhance the formulation's computational efficiency. The first technique considered is the fast multipole method (FMM) and the second technique is adaptive refinement of the discretization, based on a posteriori error estimation. Thus, the motivation for the work presented in this document is to increase the computational efficiency of the FE formulation considered.

The FE formulation considered is widely used within the microwave engineering, finite element community. Tetrahedral, rectilinear, curl-conforming, mixed- and full order, hierarchical vector elements are used. The formulation is extended to incorporate a cavity backed aperture employing the appropriate half-space Green function within a BI boundary condition, which represents a specific member of a large class of hybrid FE-BI formulations. The formulation is also extended to model coaxial ports via a Neumann boundary condition, using a priori knowledge of the dominant modal fields. Results are presented in support of the formulation and its extensions, including novel results on the coupling between microstrip patch antennas on a perforated substrate.

The FMM is investigated first, with the purpose of optimizing the non-local BI component of the cavity FE-BI formulation, in light of its coupling with the differential equation based, sparse FEM. The FMM results in a partly sparse factorization of the BI contribution to the system matrix. Error control schemes for the FMM are thoroughly reviewed and an additional, novel scheme is empirically devised.

The second technique investigated, which is more directly related to the FEM and larger in scope, is the use of a posteriori error estimation, in order to optimize the FE discretization through adaptive refinement. A overview of available a posteriori error estimation techniques in the general FE literature is given as well as a survey of available techniques that are specifically tailored to Maxwell's equations. Two known approaches within the applied mathematics literature are adapted to the FE formulation at hand, resulting in two novel, residual based error estimation procedures for this FE formulation — one explicit in nature and the other implicit. The two error estimators are then used to drive a single p adaptive analysis cycle of the FE formulation, experimentally demonstrating their effectiveness. A quasi-static condition is introduced and successfully used to enhance the adaptive algorithm's effectiveness, independently of the error estimation procedure employed. The novel error estimation schemes and adaptive results represent the main research contributions of this study.

Opsomming

Hierdie dokument beskryf 'n Galerkin eindige element (EE) formulering vir die volgolf, frekwensiegebied, elektromagnetiese analise van driedimensionele strukture relevant vir mikrogolfingenieurwese, saam met die ondersoek van twee tegnieke om die numeriese effektiwiteit van die formulering te verbeter. Die eerste tegniek wat ondersoek word, is die vinnige multipool metode (VMM) en die tweede is die aanpasbare verfyning van die EE diskretisering, gebaseer op a posteriori foutberaming. Dus, die motivering vir hierdie werk is om die numeriese effektiwiteit van die genoemde EE formulering te verbeter.

Die bogenoemde EE formulering word algemeen gebruik deur die mikrogolfingenieurswese, eindige element-gemeenskap. Tetrahedriese, reglynige, curl-ondersteunende, hierargiese vektorelemente van gemengde- en volledige ordes word gebruik. Die formulering word uitgebrei om holtes in 'n oneindige grondvlak te kan hanteer, deur gebruik te maak van die toepaslike Green funksie binne 'n grensintegraal (GI) grensvoorwaarde, wat 'n spesifieke lid is van 'n groot klas, hibriede, EE-GI formulerings. Die formulering word ook uitgebrei om koaksiale poorte to modelleer via 'n Neumann grensvoorwaarde, deur die gebruik van a priori kennis van die koaksiale, dominante modus-velde. Resultate word gelewer om die formulering, saam met die uitbreidings daarvan, te ondersteun, insluitende oorspronklike resultate in verband met die koppeling tussen mikrostrook plakantennes op 'n geperforeerde substraat.

Die VMM word eerste ondersoek, met die doel om die nie-lokale, GI komponent van die EE-GI formulering vir holtes te optimeer, weens die koppeling daarvan met die yl, differensiaalvergelyking-gebaseerde, eindige element-metode. Die VMM lei tot 'n gedeeltelik-yl faktorisering van die GI bydrae tot die algehele matriksvergelyking. Skemas om die VMM fout te beheer word deeglik ondersoek en 'n addisionele, oorspronklike skema word empiries ontwikkel.

Die tweede tegniek wat ondersoek word, wat meer direk verband hou met die eindige element-metode, en van groter omvang is, is die gebruik van a posteriori foutberaming om die EE diskretisasie te optimeer deur middel van aanpasbare verfyning. 'n Oorsig van beskikbare, a posteriori foutberamingstegnieke in die algemene EE literatuur word gegee, asook 'n opname van beskikbare tegnieke wat spesifiek gerig is op Maxwell se vergelykings. Twee bekende benaderings binne die toegepaste wiskunde-literatuur word aangepas by die bogenoemde EE formulering, wat lei tot twee oorspronklike residu-gebaseerde foutberamingstegnieke vir hierdie formulering — een van 'n eksplisiete aard en die ander implisiet. Die twee foutberamingstegnieke word gebruik om 'n enkel, p-aanpasbare analisesiklus aan te dryf, wat die effektiwiteit van die foutberamingstegnieke eksperimenteel demonstreer. 'n Kwasi-statiese vereiste word beskryf en suksesvol gebruik om die aanpasbare algoritme se effektiwiteit te verhoog, onafhanklik van die foutberamingstegniek wat gebruik word. Die oorspronklike foutberamingstegnieke en aanpasbare algoritme-resultate verteenwoordig die hoof navorsingsbydraes van hierdie studie.

Acknowledgements

Thanks to Professor David Davidson for his consistent support and encouragement, with whom I have had the pleasure of working together for the last four years. He has made an invaluable contribution to my professional development and has always been available for questions and advice — professionally, honestly, friendly and patiently. He reviewed the draft version of this document in record time.

Thanks to family and friends for their support and encouragement.

I gratefully acknowledge the financial support I received as a post graduate student, over the past four years, from the South African National Research Foundation (NRF) and the University of Stellenbosch. I would also like to thank EM Software & Systems for additional financial support.

This document was typeset with the LaTeX document preparation system [116, 114].

Contents

List of Figures

List of Tables

Conventions, terms and basic definitions

System of units

The SI system of units is used [171]. I.e. length in meters (m), mass in kilograms (kg), time in seconds (s) and electric current in ampère (A). The standard extensions in the table below will also be employed:

Quantity	Abbreviation	SI units
Electric potential in Volt	V	$kg \cdot m^2 \cdot A^{-1} \cdot s^{-3}$
Electric charge in Coulomb	C	$A \cdot s$
Electric resistance in Ohm	Ω	$kg \cdot m^2 \cdot A^{-2} \cdot s^{-3}$

Mathematical notations

- A scalar quantity is indicated in italic typeface (eg. a).
- Vectors are indicated by a bold typeface (eg. $\mathbf{x}$).
- Dyads are indicated by a normal or italic typeface with two over bars (eg. $\overline{\overline{C}}$).
- The absolute value of a vector is indicated by the same letter as the vector, but in italic typeface and no longer bold (eg. $|\mathbf{x}| = x$).
- Unit vectors are indicated by a hat over a letter in italic typeface (eg. $\hat{u}$).
- Vectors, in the context of matrix multiplication, are indicated in curly brackets (eg. $\{E\}$).
- Matrices are indicated in square brackets (eg. $[A]$).
- dl, dS and dV designate differential lines, surfaces and volumes and will indicate the dimension (1, 2 or 3) of the integral sign associated with them.

Phasor quantities

It can be assumed that all electromagnetic quantities are phasors throughout, unless stated otherwise. A positive time convention is used, thus, if $\mathbf{A}$ is a phasor, then its actual, time dependent value, $\mathbf{A}_t$, can be calculated as

$$\mathbf{A}_t = \mathrm{Re}\left\{\mathbf{A}e^{j\omega t}\right\}.$$

Terms and basic definitions

2D, 3D	Two- and three Dimensional.
BI	Boundary Integral.
BVP	Boundary Value Problem.
ERM	Element Residual Method.
FEM	Finite Element Method.
FMM	Fast Multipole Method.
RCS	Radar Cross Section.
TEM	Transverse ElectroMagnetic.
VBVP	Variational Boundary Value Problem.
$\mathcal{Z}$	The set of all integers: $0, \pm1, \pm2, \pm3, \ldots$.
$\mathcal{N}$	The set $\{a \in \mathcal{Z} \mid a \geq 0\}$.
$\mathcal{R}$	The set of all real numbers.
$\mathcal{C}$	The set of all complex numbers.
$\epsilon, \epsilon_r, \epsilon_0$	General-, free space- and relative permittivity. $\epsilon = \epsilon_r\epsilon_0$.
μ, μ_r, μ_0	General-, free space- and relative permeability. $\mu = \mu_r\mu_0$.
k_0	Free-space wavenumber, $k_0 = \frac{2\pi}{\lambda_0}$.
Z_0	Free-space wave impedance, $Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}$.
$G_0(\mathbf{r}, \mathbf{r}')$	Scalar, free space, Green function.
$\overline{\overline{\mathrm{G}}}_0(\mathbf{r}, \mathbf{r}')$	Dyadic, free space, Green function of electric type.
$\overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')$	Dyadic, half space, Green function of electric type.
Ω	Problem domain $\subset \mathcal{R}^3$.
Γ	Boundary of Ω. $\Gamma \equiv \partial\Omega$.
Γ_D	Homogeneous Dirichlet boundary of Ω.
Γ_N	Inhomogeneous Neumann boundary of Ω.
S_a	Cavity aperture.
S_c	Coaxial port aperture.
W	Infinite dimensional solution space.
W_h	Finite dimensional approximation of W.
$\mathbf{v}_h$	Approximate, finite dimensional representation of $\mathbf{v}$, $\mathbf{v}_h \in W_h$.
τ	The geometrical discretization.
N_τ	The number of elements in the mesh.
N_f	The number of faces in the mesh.

$\mathbf{N}_i$	Global basis functions, $i = 1, .., N_\tau$.
N_F	Global number of degrees of freedom.
K_i	Elemental volumes, $i = 1, .., N_\tau$.
W_i	Elemental version of W, $i = 1, .., N_\tau$.
$M_m(K)$	The space of vector functions of mixed polynomial order m, on elemental volume K.
$U_p(K)$	The approximation space on element K.
$V_{p,q}(K)$	The bubble space on element K.
V^e	Elemental volume, for the purpose of defining contributions to the system matrix equation.
S^s, S^t	Elemental faces, for the purpose of defining contributions to the system matrix equation.
S_u	Surface of a unit sphere, centered at the origin.
M	Number of FMM spherical quadrature points.
L_{tot}	Number of terms in the truncated FMM addition theorem.
$D_{\max}$	Maximum FMM group dimension.
$X_{\min}$	Minimum inter-group distance for using the FMM.
ε^L	Relative Green function error related to the FMM.
$\text{Ker}\,(A)$	The null-space (kernel) of the operator A.
$\text{span}\,A$	The function space spanned by the set of functions A.
$\dim A$	The dimensionality of the function space A.
$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{C},\alpha}$	The complex, vector inner product, defined as $\int_\alpha \mathbf{v} \cdot \mathbf{w}\, d\alpha$ on the domain α. Strictly speaking, this is not a true inner product [115].
$L^2(\alpha)$	The space of all square integrable functions on the domain α.
$\|\mathbf{v}\|_{L^2(\alpha)}$	The $L^2(\alpha)$-norm of $\mathbf{v}$ on the domain α, defined as $[\int_\alpha \mathbf{v} \cdot \mathbf{v}^* d\alpha]^{\frac{1}{2}}$. In the case of a scalar argument, the dot product reduces to normal multiplication.
$L^m(\alpha)$	The space of all functions on the domain α of which the m-th power is integrable over the domain α. $m \in \mathcal{N}$. This space has an associated L^m norm, defined as this integral, to the power $\frac{1}{m}$.
$C^m(\alpha)$	The space of scalar functions on the domain α that, together with their first m derivatives, are continuous [158].
$W^{m,p}(\alpha)$	The general Sobolev space of weak derivative order m and integration order p on the domain α [158], i.e. $W^{m,p}(\alpha) = \{u \in L^p(\alpha) \mid D^\beta u \in L^p(\alpha),\ \|\beta\| \leq m\}$. ($D^\beta u$ is a multi-index notation, indicating derivatives with respect to more than one variable, but with the total derivative order not exceeding m.)
$H^m(\alpha)$	The Sobolev space of weak derivative order m on the domain α. $H^m(\alpha) \equiv W^{m,2}(\alpha)$ [158].

$|v|_{H^m(\alpha)}$ Scalar Sobolev semi-norm of derivative order m on the domain α. $v \in H^m(\alpha)$. This norm is defined in terms of the previously mentioned multi-index notation, as follows [51]:

$$|v|_{H^m(\alpha)} = \left(\sum_{|\beta|=m} \|D^\beta v\|^2_{L^2(\alpha)} \right)^{\frac{1}{2}}.$$

$|\mathbf{v}|_{(H^m(\alpha))^3}$ Vector Sobolev semi-norm of derivative order m on the domain α. $\mathbf{v} \in (H^m(\alpha))^3$. This norm is defined in terms of the scalar Sobolev semi-norm, with $\mathbf{v} = v_1\hat{x} + v_2\hat{y} + v_3\hat{z}$, as follows [51]:

$$|\mathbf{v}|_{(H^m(\alpha))^3} = \left(\sum_{i=1}^{3} |v_i|^2_{H^m(\alpha)} \right)^{\frac{1}{2}}.$$

$H(\text{curl}, V)$ The space of all vector functions in the 3D volume V with square integrable curl. $H(\text{curl}, V) = \{\mathbf{a} \in (L^2(V))^3 \mid \nabla \times \mathbf{a} \in (L^2(V))^3\}$.

$\overline{P}_p(V)$ The space of polynomial functions with all terms of order p, on the 3D volume V. $p \in \mathcal{N}$.

$P_p(V)$ The space of all polynomial functions of order p on the 3D volume V. $P_p(V) = \overline{P}_0(V) \oplus \overline{P}_1(V) \oplus ... \oplus \overline{P}_{p-1}(V) \oplus \overline{P}_p(V)$ with $p \in \mathcal{N}$.

Chapter 1

Introduction

This document presents a FE formulation for the full-wave, frequency domain, electromagnetic analysis of structures relevant to microwave engineering, together with two techniques to enhance the formulation's computational efficiency: the FMM; and adaptive refinement of the discretization, based on a posteriori error estimation. Results are presented in support of all the techniques presented.

The basic FE formulation presented, using curl-conforming, mixed- and full order, hierarchical vector elements, is commonly employed within the microwave engineering, FE community. Here, the primary focus is on its extension to the analysis of cavity backed apertures in an infinite ground plane, which represents a specific member of a large class of commonly used, hybrid FE-BI formulations. Dominant mode, coaxial- and rectangular waveguide ports are also considered. Collectively, these boundary conditions cover most of the 'exact' boundary termination techniques currently employed in the FE modeling of microwave devices.

The motivation for this thesis can be formulated as a question. Given this FE formulation, how can the computational efficiency be improved? In answer to this question, two widely used approaches are investigated: Firstly, the FMM is investigated, with the purpose of optimizing the non-local BI component of the FE-BI formulation, in light of its coupling with the differential equation based, sparse FEM. Secondly, a posteriori error estimation techniques are investigated, established and extended in order to drive adaptive analysis cycles of this FE formulation.

Published work similar to that presented in this document is evaluated in the introductions to every chapter. The only significant approach – with respect to microwave engineering FEM formulations – that is absent here, is the use of absorbing boundary conditions [175], but this can be considered separately from the FMM and is worthy of exclusive investigation in the case of a posteriori error estimation. Other current areas of research within the microwave engineering FEM domain that are not touched upon here, are multigrid methods and other efficient techniques for solving the resulting system matrix equation [186].

Chapter 2 presents the general, Galerkin FEM used throughout this document. After establishing the VBVP, it is discretized with curl-conforming, mixed- and full order, hierarchical, vector elements. The elements are discussed in detail in Appendix C, which comprises an important part of the FEM foundation upon which the rest of this document builds.

Chapter 3 presents the specialization of the FEM presented in Chapter 2, to the analysis of cavities recessed in an infinite ground plane, resulting in the cavity backed aperture FE-BI

formulation. The chapter also introduces a dominant mode, coaxial port model, of which the theoretical background is presented in Appendix E. Numerous results are presented, illustrating the application of the formulations to practical problems, of which the most noteworthy and novel is a numerical investigation into the coupling between microstrip patch antennas on a perforated substrate.

Chapter 4 presents the application of the FMM to the BI component of the cavity FE-BI formulation presented in Chapter 3. The FMM results in a partly sparse factorization of the full BI matrix contribution to the system matrix equation — the first approach considered in this document as a means of improving the computational efficiency of the FE formulation used here. A significant part of the chapter is devoted to discussions on the control of the accuracy of the FMM versus the standard BI method. Techniques for controlling the accuracy are presented, with some of the mathematical details relegated to Appendix F.

A posteriori error estimators form the basis of adaptive improvements to the FE discretization — the second approach considered in this document as a means of improving the computational efficiency of the FE formulation used here. Chapter 5 starts by presenting a review of a posteriori error estimation techniques available in the general scientific literature, as well as those specifically aimed at the FEM for Maxwell's equations. In the rest of the chapter, two residual based error estimators are derived. One is based on explicit residual evaluations and the other is an implicit scheme, based on the solution of elemental VBVPs with the residuals as sources. Both techniques are novel to some degree and together they represent a core research contribution made within this document.

Chapter 6 presents p adaptive results (selective upgrading of the elemental approximation function spaces). The elements are marked for upgrading, based on the two error estimation techniques derived in Chapter 5. It is experimentally demonstrated that the two error estimators are effective in identifying those elements strongly in need of refinement. The estimators are only used for refinement and not to gauge the overall solution quality. A condition is introduced for indicating whether an element needs to be upgraded to full- or mixed polynomial order. This scheme is shown to be quite effective in the context of a single example. Rectangular waveguide problems are also considered for illustrating the effectiveness of the adaptive procedures, since the rectangular- and coaxial, dominant mode port formulations are practically identical. This chapter forms another core research contribution of this document, since it presents novel adaptive results based on the two error estimators presented in Chapter 5, for problems not previously considered in the literature in such an adaptive FE setting.

The research presented in this document should be especially relevant to developers of state-of-the-art FE codes for the engineering analysis of microwave devices, as well as researchers in this field.

Finally, regarding the software implementation of the formulations and techniques presented in this document: everything was implemented by the present author, based on an existing, non-adaptive code, capable of handling PEC boundary conditions and dominant mode, rectangular waveguide port boundary conditions, with hierarchical, curl-conforming, vector elements up to order QT/QN and with only a direct solver implemented. See Section 3.7 for further details.

Chapter 2

A general FE formulation for time-harmonic electromagnetic fields

2.1 Introduction

In this chapter, the foundation is laid for the rest of this document. A general, VBVP representation of Maxwell's equations in time-harmonic, phasor form is established in terms of the electric field, suitable for the analysis of 3D problems in microwave engineering, i.e. PEC- and isotropic, possibly lossy, regions of arbitrary shape, with homogeneous Dirichlet- and inhomogeneous Neumann boundary conditions. (Though inhomogeneous Dirichlet boundary conditions are not considered, since they are not relevant to any of the applications considered in this document, the extension is straight forward.) The Galerkin discretization of the VBVP, employing curl-conforming, vector elements, yielding a FEM, is described.

This specific variational formulation is used, since it is almost exclusively used within the microwave engineering community concerned with 3D, FE analysis and thus highly relevant. The electric field formulation (as opposed to the magnetic field formulation) is used, since it is a very common choice in the literature, as it allows for a much easier- and somewhat more efficient implementation of the commonly occurring PEC boundary condition and lastly, because it is necessary for the incorporation of a dominant mode, rectangular waveguide port model (used to generate some of the results in Chapter 6). Motivation for the choice of elements is provided in Appendix C, where the properties of the elements are extensively discussed.

The VBVP (or equivalent stationary functional) presented here, can be found in the major engineering texts on FE analysis for Maxwell's equations (see for example [99, 174, 167, 192]); although the derivations are sometimes a little unclear, the same end result is presented by all. Some references that treat the formulation in somewhat more detail than is customary in the microwave engineering community, are [132, 112]. An alternative, mixed formulation is presented in [188, 189, 152, 72] (2D) and in [75] (3D). It involves the addition of a Lagrange multiplier to the solution space in order to enforce Gauss' law explicitly, making the formulation stable as $\omega \to 0$, or equivalently, as the element size tends to zero. (See [167, Sec. 7.2.1.1] and [75] for discussions on the low frequency instability.) This is not an issue here, since only microwave applications are considered, using uniform meshes as far as possible. Finally, the elements used here take care of the spurious modes problem (see Appendix

C), therefore no alteration of the functional formulation to counteract this phenomenon is necessary [142].

The outline of this chapter is as follows: Section 2.2 derives a VBVP representation of the Maxwell equation BVP. A stationary functional representation, equivalent to the VBVP, is presented in Section 2.3. Section 2.4 discusses the discretizatin of the VBVP (or stationary functional), resulting in the FEM used throughout this document.

2.2 The VBVP formulation

This section establishes an equivalent, weak VBVP representation of the conventional BVP, describing the electric field on a finite volume in terms of Maxwell's equations. See [161, 158] for examples of similar procedures for scalar problems.

Start by defining the problem domain and its boundaries:

- $\Omega \subset \mathcal{R}^3$ represents the problem domain.
- Γ_D represents those parts of $\partial\Omega$ upon which a homogeneous Dirichlet boundary condition is imposed.
- Γ_N represents those parts of $\partial\Omega$ upon which an inhomogeneous Neumann boundary condition is imposed.
- $\Gamma \equiv \partial\Omega$ represents the total boundary. The following relations hold between Γ, Γ_D and Γ_N:

$$\Gamma = \Gamma_D \cup \Gamma_N \tag{2.1}$$
$$\emptyset = \Gamma_D \cap \Gamma_N. \tag{2.2}$$

- $\hat{n}$ is the unit, outward normal to Γ.

The BVP to be solved on Ω, in terms of the electric field, vector wave equation (equation (B.16)) and with the appropriate boundary conditions, is as follows:

$$\begin{cases} \nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E} - k_0^2 \epsilon_r \mathbf{E} = -jk_0 Z_0 \mathbf{J} & \text{on } \Omega \\ \hat{n} \times \mathbf{E} = 0 & \text{on } \Gamma_D \\ \hat{n} \times \nabla \times \mathbf{E} = \mathbf{N} & \text{on } \Gamma_N \end{cases}. \tag{2.3}$$

This is the strong version of the problem, because it exactly and explicitly specifies the conditions that must be satisfied by the solution.

The boundary conditions are 'appropriate', since by the uniqueness theorem they guarantee a unique solution to equation (2.3). The time-harmonic, uniqueness theorem [92] states that it is necessary and sufficient to state either the tangential electric field or tangential magnetic field over the whole boundary of a domain in order that the electromagnetic fields inside are uniquely defined. The Dirichlet boundary condition is a specification of the tangential electric field and the Neumann boundary condition is a specification of the tangential magnetic field (see Faraday's law, equation (B.5)).

Rewrite the differential equation in equation (2.3) in operator form as

$$L_H(\mathbf{E}) = \mathbf{F}. \tag{2.4}$$

Take the complex, vector inner product (as defined in [99]) of both sides of equation (2.4) with an arbitrary testing function $\mathbf{W}$, resulting in the following equivalent, strong representation of the problem:

$$\begin{cases} \langle L_H(\mathbf{E}), \mathbf{W} \rangle_{C,\Omega} = \langle \mathbf{F}, \mathbf{W} \rangle_{C,\Omega} & \\ \hat{n} \times \mathbf{E} = 0 & \text{on } \Gamma_D \\ \hat{n} \times \nabla \times \mathbf{E} = \mathbf{N} & \text{on } \Gamma_N \end{cases}. \tag{2.5}$$

The left side of the inner product equation in the system (2.5) can be rewritten, using the first vector Green's theorem (equation (A.10)), yielding

$$\begin{aligned} \langle L_H(\mathbf{E}), \mathbf{W} \rangle_{C,\Omega} &= \int_\Omega \left(\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E} \right) \cdot \mathbf{W} \, dV - k_0^2 \int_\Omega \epsilon_r \mathbf{E} \cdot \mathbf{W} \, dV \\ &= \int_\Omega \frac{1}{\mu_r} (\nabla \times \mathbf{E}) \cdot (\nabla \times \mathbf{W}) \, dV - k_0^2 \int_\Omega \epsilon_r \mathbf{E} \cdot \mathbf{W} \, dV \\ &\quad - \oint_\Gamma \frac{1}{\mu_r} (\mathbf{W} \times \nabla \times \mathbf{E}) \cdot \hat{n} \, dS. \end{aligned} \tag{2.6}$$

In view of identities (A.7) and (A.8), the kernel of the surface integral in equation (2.6) can be written in the following alternative forms:

$$\frac{1}{\mu_r} (\mathbf{W} \times \nabla \times \mathbf{E}) \cdot \hat{n} = -\frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}) \cdot \mathbf{W} = \frac{1}{\mu_r} \nabla \times \mathbf{E} \cdot (\hat{n} \times \mathbf{W}). \tag{2.7}$$

By separating the surface integral over Γ in equation (2.6) into its disjoint components (see equation (2.2)) and using the above equivalent forms of the kernel, the following result is obtained:

$$\begin{aligned} \langle L_H(\mathbf{E}), \mathbf{W} \rangle_{C,\Omega} &= \int_\Omega \frac{1}{\mu_r} (\nabla \times \mathbf{E}) \cdot (\nabla \times \mathbf{W}) \, dV - k_0^2 \epsilon_r \int_\Omega \mathbf{E} \cdot \mathbf{W} \, dV \\ &\quad + \int_{\Gamma_N} \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}) \cdot \mathbf{W} \, dS \\ &\quad - \int_{\Gamma_D} \frac{1}{\mu_r} \nabla \times \mathbf{E} \cdot (\hat{n} \times \mathbf{W}) \, dS. \end{aligned} \tag{2.8}$$

A weak representation of the original boundary value problem now follows by substituting the inhomogeneous Neumann boundary condition into (2.8) and substituting back into the system in (2.5), yielding the system

$$\begin{cases} \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{C,\Omega} - k_0^2 \langle \epsilon_r \mathbf{E}, \mathbf{W} \rangle_{C,\Omega} - \int_{\Gamma_D} \frac{1}{\mu_r} \nabla \times \mathbf{E} \cdot (\hat{n} \times \mathbf{W}) \, dS & \\ \quad = -\int_{\Gamma_N} \frac{1}{\mu_r} \mathbf{N} \cdot \mathbf{W} \, dS - jk_0 Z_0 \langle \mathbf{J}, \mathbf{W} \rangle_{C,\Omega} & \\ \hat{n} \times \mathbf{E} = 0 \text{ on } \Gamma_D & \end{cases}. \tag{2.9}$$

This system is written with the terms depending on both $\mathbf{W}$ and $\mathbf{E}$ on the left and the terms only depending on $\mathbf{W}$, on the right. In Section 2.3 it will be required that the left hand side be symmetric with respect to $\mathbf{W}$ and $\mathbf{E}$. This requirement leads to the conclusion that $\mathbf{W}$ must also satisfy the homogeneous Dirichlet boundary condition in order that the unsymmetric surface integral over Γ_D falls away. Note that the inhomogeneous Neumann boundary condition is not explicitly present any more, since it is implicitly incorporated in the variational formulation of equation (2.9).

The final form of the VBVP is

$$\begin{cases} B(\mathbf{E}, \mathbf{W}) = L(\mathbf{W}) \ \ \forall \ \mathbf{W} \in W \\ \mathbf{E} \in W \end{cases} \tag{2.10}$$

with the solution- and testing vector function space defined as

$$W = \{\mathbf{a} \in H(\text{curl}, \Omega) \mid \hat{n} \times \mathbf{a} = 0 \text{ on } \Gamma_D\} \tag{2.11}$$

and with the bilinear and linear forms defined as

$$B(\mathbf{E}, \mathbf{W}) = \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{C,\Omega} - k_0^2 \langle \epsilon_r \mathbf{E}, \mathbf{W} \rangle_{C,\Omega} \tag{2.12}$$

$$L(\mathbf{W}) = -\int_{\Gamma_N} \frac{1}{\mu_r} \mathbf{N} \cdot \mathbf{W} \, dS - jk_0 Z_0 \langle \mathbf{J}, \mathbf{W} \rangle_{C,\Omega} \, . \tag{2.13}$$

Note that the Dirichlet boundary condition is sometimes referred to as *essential*, since it must be explicitly enforced via a restriction on the space W, while the Neumann boundary condition is referred to as *natural*, since it is implicitly incorporated within the variational formulation of the problem [174, 99, 167].

Equation (2.10) is a *weak* form of the original BVP, since the derivative conditions required of the solution has been relaxed and the associated strong form is satisfied in a distributional manner [158, 161]. The weak form allows a larger space of approximate solutions, making the problem easier to solve in an approximating manner [167].

2.3 Establishing the stationary functional formulation

In Section 2.2, the BVP was cast into an equivalent, weak integral form, or VBVP. In this section the VBVP is used to establish a functional that also yields a weak solution to the original BVP upon rendering the functional stationary. This stationary approach is sometimes referred to as the Rayleigh-Ritz method [99, 167] and is equivalent to the VBVP in this case. In the case of non-self-adjoint boundary value problems, the stationary functional representation of the problem takes a different form, as discussed in [47, 99].

The first variation of a functional, generalized to the vector valued function case, is defined as follows [166]:

$$\delta F(\mathbf{v}) = \lim_{\theta \to 0} \left[\frac{F(\mathbf{v} + \theta \mathbf{h}) - F(\mathbf{v})}{\theta} \right] . \tag{2.14}$$

In words, the above definition states that the first variation is equal to the coefficient of the first order θ term in $F(\mathbf{v}+\theta\mathbf{h})-F(\mathbf{v})$. Clearly the variation $\mathbf{h}$ must satisfy the homogeneous form of the essential boundary conditions imposed on $\mathbf{v}$ for the variational vector function $\mathbf{v}+\theta\mathbf{h}$ to be a valid argument.

Rewrite the integral equation in the VBVP of equation (2.10) as

$$B(\mathbf{E},\mathbf{W}) - L(\mathbf{W}) = 0. \tag{2.15}$$

The aim of this section is to establish a functional that, upon rendering it stationary, will result in equation (2.15). To this end observe that bilinear form $B(\cdot,\cdot)$ is linear in both arguments separately and would therefore be 'quadratic' in one argument if the two arguments were set equal. Also observe that the linear form $L(\cdot)$ is linear in its one and only argument. Recalling the analytic integration rules (A.20) and (A.21), equation (2.15) is 'integrated' to obtain the functional $F(\mathbf{E}) = \frac{1}{2}B(\mathbf{E},\mathbf{E}) - L(\mathbf{E})$, of which the first variation can be calculated as

$$\begin{aligned}
\delta F(\mathbf{E}) &= \lim_{\theta\to 0}\left[\frac{F(\mathbf{E}+\theta\delta\mathbf{E}) - F(\mathbf{E})}{\theta}\right] \\
&= \lim_{\theta\to 0}\left[\frac{\frac{1}{2}\theta B(\mathbf{E},\delta\mathbf{E}) + \frac{1}{2}\theta B(\delta\mathbf{E},\mathbf{E}) + \frac{1}{2}\theta^2 B(\delta\mathbf{E},\delta\mathbf{E}) - \theta L(\delta\mathbf{E})}{\theta}\right] \\
&= B(\mathbf{E},\delta\mathbf{E}) - L(\delta\mathbf{E})
\end{aligned} \tag{2.16}$$

where the symmetry of $B(\cdot,\cdot)$ was used, as enforced in Section 2.2. As stated earlier in this section, $\delta\mathbf{E} \in W$ must hold, thus replacing $\delta\mathbf{E}$ with $\mathbf{W} \in W$ in equation (2.16) and setting $\delta F(\mathbf{E}) = 0$, equation (2.15) is obtained, as required.

Clearly $F(\mathbf{E})$ is the correct functional, as it is also stated in [161, 158, 47, 99]. The stationary functional representation of the original boundary value problem (2.3) therefore follows as

$$\begin{cases} F(\mathbf{E}) = \dfrac{1}{2}B(\mathbf{E},\mathbf{E}) - L(\mathbf{E}) \\ \delta F(\mathbf{E}) = 0 \\ \mathbf{E} \in W \end{cases} \tag{2.17}$$

2.4 The discrete problem

In order to solve the variational boundary value problem (or equivalently, the stationary functional), one has to approximate the infinite dimensional space that the unknown function and weighting function belong to, by an appropriate subspace of finite dimension. This leads to an approximate solution. This section describes how to obtain the discrete representation through finite element discretization, leading to a discrete version of the VBVP (or equivalently, the stationary functional).

2.4.1 General discretization procedure

Approximating the unknown function and the weighting function with the same finite dimensional space, is called the Galerkin method [158], which is a type of weighted residual

method [158, 161, 167] and is employed here.

To construct the approximate function spaces, the problem domain, Ω, is discretized into a set, τ, of elemental volumes, K, such that

$$\Omega = \bigcup_{i=1}^{N_\tau} K_i \tag{2.18}$$

$$K_m \cap K_n = \partial K_m \cap \partial K_n \quad \{m \neq n\}, \tag{2.19}$$

where N_τ is the number of elements in the set. Thus, the union of the elements constitute the problem domain exactly and elements are non-overlapping. The set τ is called the *mesh*.

On every elemental volume a number of basis functions are defined. The unknown function is approximated as the sum of all these basis functions, each multiplied by a coefficient to be determined. Every basis function is also used as a weighting function. Coefficients are set equal where continuity must hold (assuming the basis functions, without weighting, are continuous in the desired sense).

This procedure of approximating the solution and weighting spaces in order to obtain a finite, discrete system is called the *FEM*.

2.4.2 The vector wave equation, Galerkin FEM

The space W is discretized with the tetrahedral, rectilinear, curl-conforming, mixed- and/or complete order, vector elements described in Appendix C. These elements are ideally suited to discretizing W, as discussed in Appendix C in detail. A further constraint is placed on the mesh, as follows:

$$K_m \cap K_n \in \{\emptyset;\ n_{mn};\ e_{mn};\ f_{mn}\} \quad \{m \neq n\}, \tag{2.20}$$

where n_{mn} is an elemental node in both elements, e_{mn} is an elemental edge in both elements and f_{mn} is an elemental face in both elements. Thus the mesh is *regular* [71] or *proper* [4], since every node is a vertex of all its neighbouring elements. Every elemental face is either shared, along with its three vertices, by two elements, or it is part of the boundary Γ.

Within every element, the space W is approximated by the basis functions presented in Table C.4, modeling the elemental vector function spaces $U_p(K)$, with $0.5 \leq p \leq 2$ (see equation (C.12)) . The field within element K is approximated as follows (equation (C.1) repeated):

$$\mathbf{E}_h^K = \sum_{i=1}^{N_F^K} E_i^K \mathbf{N}_i^K. \tag{2.21}$$

If element K has a face in Γ_D, then those terms associated with the tangential components on this face and its edges, are discarded. This is the unconstrained approximation [71], since tangential continuity between elements is not yet enforced. Tangential continuity must be enforced, for the approximation to be a subspace of W.

Tangential continuity is enforced by equating the (curl-conforming) degrees of freedom associated with shared edges and shared faces. Thus the spatial domain of basis functions associated with shared edges and faces are the union of all elements sharing that edge or

face, since those elemental functions are all scaled by the same coefficient. Enforcing tangential continuity leads to the constrained approximation [71], defined over the whole problem domain Ω, as follows:

$$\mathbf{E}_h = \sum_{i=1}^{N_F} E_i \mathbf{N}_i, \tag{2.22}$$

where N_F is the total, global number of degrees of freedom. The $\mathbf{N}_i$ are the constrained basis functions and the E_i are the degrees of freedom associated with every $\mathbf{N}_i$.

Substituting the constrained approximation into the VBVP (2.10) and using the constrained basis functions as weighting functions (spanning the discrete solution space W_h), yields the approximate, discrete, VBVP:

$$\begin{cases} B(\mathbf{E}_h, \mathbf{W}_h) = L(\mathbf{W}_h) \quad \forall\ \mathbf{W}_h \in W_h \\ \mathbf{E}_h \in W_h \end{cases}. \tag{2.23}$$

The discrete variational boundary value problem represents a set of linear equations and can be written in matrix form, yielding

$$[A]\{E\} = \{b\}, \tag{2.24}$$

with

$$[A] = [K] \tag{2.25}$$

$$\{b\} = \{b_J\} + \{b_N\}. \tag{2.26}$$

The elemental contributions to the system matrix equation are as follows:

$$[K^e] = \int_{V^e} \left[\frac{1}{\mu_r^e} \{\nabla \times \mathbf{N}^e\} \cdot \{\nabla \times \mathbf{N}^e\}^T - k_0^2 \epsilon_r^e \{\mathbf{N}^e\} \cdot \{\mathbf{N}^e\}^T \right] dV \tag{2.27}$$

$$\{b_J^e\} = -jk_0 Z_0 \int_{V^e} \mathbf{J} \cdot \{\mathbf{N}^e\}\, dV \tag{2.28}$$

$$\{b_N^s\} = -\int_{S^s} \frac{1}{\mu_r^s} \mathbf{N} \cdot \{\mathbf{N}^s\}\, dS, \tag{2.29}$$

where the superscripts e and s indicate association with a specific element and association with a specific triangular, elemental face in the Neumann boundary, respectively. $\{\mathbf{N}^e\}$ and $\{\mathbf{N}^s\}$ represent the column vectors of vector basis functions associated with element e and those associated with the element to which boundary face s belongs to, respectively.

These elemental contributions are added to their corresponding positions in the system matrix equation via the association between elemental, local degrees of freedom and global degrees of freedom. These associations are established by the constrained approximation, such that shared degrees of freedom in adjacent elements will both associate with the same global degree of freedom.

The system matrix dimension is equal to the dimension of the approximate solution space, i.e.

$$N_F = \dim W_h. \tag{2.30}$$

This includes the reduction in degrees of freedom caused by enforcing the homogeneous Dirichlet boundary condition as well as the reduction in degrees of freedom caused by the change from the unconstrained approximation to the constrained approximation.

Finally, if the alternative, stationary functional route is followed, one arrives at

$$F(\{E\}) = \frac{1}{2}\{E\}^T[A]\{E\} - \{E\}^T\{b\}, \tag{2.31}$$

which, upon rendering it stationary by setting

$$\frac{\partial F}{\partial E_i} = 0 \; ; \quad i = 1, ..., N_F, \tag{2.32}$$

also yields equation (2.24).

2.5 Conclusion

This chapter derived a Galerkin, FEM for general, 3D, time-harmonic, microwave problems in terms of the electric field, which will be build upon throughout this document. No specific research contributions were made here, although the formulation as presented here is not available in its entirety in any single microwave engineering FE text.

In the rest of this document, the FE formulation presented in this chapter will be extended (through specification of the inhomogeneous Neumann boundary condition) to modeling cavity backed apertures in an infinite ground plane (Chapter 3) and to the modeling of dominant mode ports (Chapter 6 and Appendix E). A posteriori error estimators will be derived for these extended formulations (Chapter 5).

Chapter 3

The cavity FE-BI formulation: theory, implementation and results

3.1 Introduction

This chapter extends the FE formulation presented in Chapter 2 to the analysis of a cavity recessed in an infinite PEC half space ($z \leq 0$) with vacuum above, as shown in Figure 3.1. The cavity aperture is defined by the boundary between the cavity and the vacuum half space. Since the FEM will be applied on the inside of the cavity, it can be inhomogeneously filled. The cavity aperture is incorporated into the formulation via an MoM type of boundary integral. The formulation presented can be applied unchanged to multiple cavities. This chapter also rigorously defines a dominant mode, coaxial port model by reference to Appendix E. By way of example, results obtained with the cavity FE-BI formulation as well as the coaxial port model are presented — including a numerical study of the coupling between microstrip patches on a perforated substrate.

The FE-BI formulation is a very efficient and elegant way to model the cavity geometry, since the smallest possible volume is discretized. The whole vacuum half space is incorporated as a BI boundary condition. This combines the FEM's ability to model inhomogeneous regions with the MoM's ability to model infinite geometries. The main drawback of the formulation is that the system matrix is no longer completely sparse (as is generally the case with the FEM), because the BI boundary condition is non-local. An alternative formulation to the BI would be to terminate the mesh with an absorbing boundary condition somewhere sufficiently far above the cavity aperture [175]. However, this introduces substantially more degrees of freedom and some uncertainty with respect to accuracy, since it is not an exact treatment. Also, the FMM can be used to increase the computational efficiency of the BI component, as shown in Chapter 4. The main aim of this chapter therefore is to establish the cavity backed aperture formulation, in order to build upon it in the rest of this document.

Hybrid FE formulations to take infinite geometries into account are widely used in microwave engineering. See [175] for an extensive review. The FE-BI formulation used here, is presented in [100, 102, 101, 99]. Similar FE-BI formulations in free space are available; some also couple to purely MoM, PEC objects, see [99, 172, 173]. In the cavity backed aperture FE-BI formulation, there is only one integral equation that can be used, either in electric or magnetic field form, but as indicated in the references concerning the free space FE-BI

formulation, various choices are possible in that case.

The outline of this chapter is as follows: Section 3.2 derives a boundary integral equation to describe the electric field above the aperture in terms of the field in the aperture. Section 3.3 uses the BI representation as a Neumann boundary condition in the general VBVP, yielding the cavity backed aperture FE-BI formulation. The discretization and evaluation of the elemental matrices are discussed in Section 3.4. Sections 3.5 and 3.6 are concerned with the introduction of the excitation and the extraction of observable quantities from the solution. The dominant mode, coaxial port model is presented here. After implementation issues are addressed in Section 3.7, results are finally presented in Section 3.8.

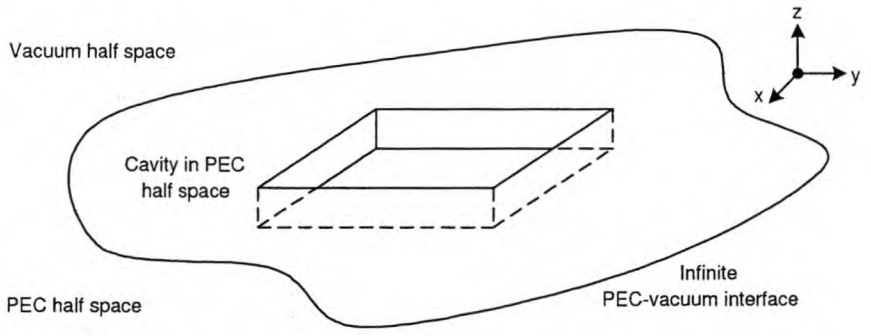


Figure 3.1: Generic cavity backed aperture geometry.

3.2 The boundary integral equation

In this section a boundary integral representation of the electric field in the vacuum half space ($z > 0$) is established. The version of the equivalence principle to be used in the derivation is stated first. Then, starting with the electric field vector wave equation and using the half space dyadic Green function of electric type, the desired representation of the electric field is established.

3.2.1 The equivalence principle

The equivalence principle [92, 162] is used to transform the original geometry into one within which the half space dyadic Green function applies. Consider the infinite volumes

$$V_\infty = \{\mathbf{r} \mid z > 0\} \tag{3.1}$$
$$V_{-\infty} = \{\mathbf{r} \mid z \leq 0\}. \tag{3.2}$$

The equivalence principle states that all sources residing within $V_{-\infty}$ can be removed and replaced with impressed magnetic- and electric surface current sources on the interface between V_∞ and $V_{-\infty}$, i.e. $z = 0$, such that the fields will remain unchanged within V_∞, while null fields will be established within $V_{-\infty}$. Should the null field region be filled with a PEC, then the impressed surface electric currents can be disregarded, since they cannot radiate

under such conditions [162]. The impressed magnetic and electric surface current sources are calculated as follows:

$$\begin{aligned} \mathbf{J}_s &= \hat{z} \times \mathbf{H}|_{z=0} \\ &= \frac{j}{\omega\mu} \hat{z} \times \nabla \times \mathbf{E}|_{z=0} \qquad (3.3) \\ \mathbf{M}_s &= -\hat{z} \times \mathbf{E}|_{z=0}. \qquad (3.4) \end{aligned}$$

3.2.2 A boundary integral representation of the electric field above the PEC half space

The electric field in the vacuum half space ($\mu_r = 1$, $\epsilon_r = 1$, $z > 0$) satisfies the vector wave equation:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k_0^2 \mathbf{E}(\mathbf{r}) = -jk_0 Z_0 \mathbf{J}^{\text{inc}}(\mathbf{r}), \qquad (3.5)$$

where $\mathbf{J}^{\text{inc}}(\mathbf{r})$ represent impressed current sources located inside the vacuum half space.

Integrating the dot product of equation (3.5) and the dyadic Green function of electric type $\overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')$ (see Section B.5.2), over the vacuum half space, V_∞, yields

$$\begin{aligned} \int_{V_\infty} \left\{ [\nabla \times \nabla \times \mathbf{E}(\mathbf{r})] \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') - k_0^2 \mathbf{E}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \right\} dV \\ = -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\text{inc}}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV. \qquad (3.6) \end{aligned}$$

This can be cast into a different form by applying the second vector-dyadic Green's theorem (see equation (A.19)) to the first term on the left and subsequently applying the vector-dyadic identity (A.15) to the kernel of the resulting surface integral:

$$\begin{aligned} &\int_{V_\infty} \mathbf{E}(\mathbf{r}) \cdot \left[\nabla \times \nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') - k_0^2 \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \right] dV \\ &= -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\text{inc}}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV \\ &\quad - \oint_{S_\infty} \left\{ [\hat{n} \times \mathbf{E}(\mathbf{r})] \cdot [\nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')] + [\hat{n} \times \nabla \times \mathbf{E}(\mathbf{r})] \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \right\} dS. \qquad (3.7) \end{aligned}$$

Here $\hat{n}$ points away from the volume and S_∞ represents the $z = 0$ plane together with a half sphere surface of infinite radius, centered at $z = 0$ and forming part of the $z > 0$ half space.

Now replace the relevant part of the LHS volume integral's kernel by the RHS of equation (B.35), i.e. $\int_{V_\infty} \mathbf{E}(\mathbf{r}) \cdot \overline{\overline{\mathbf{I}}} \delta(\mathbf{r} - \mathbf{r}')\, dV = \mathbf{E}(\mathbf{r}')$, yielding

$$\begin{aligned} \mathbf{E}(\mathbf{r}') &= -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\text{inc}}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV \\ &\quad - \oint_{S_\infty} \left\{ [\hat{n} \times \mathbf{E}(\mathbf{r})] \cdot [\nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')] + [\hat{n} \times \nabla \times \mathbf{E}(\mathbf{r})] \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \right\} dS. \qquad (3.8) \end{aligned}$$

Up to this point, the Green function was not suited to the geometry, since it does not take into account the cavity. Now apply the equivalence principle. The cavity volume is replaced by PEC and equivalent magnetic surface currents are impressed upon the $z = 0$

interface. Comparing equations (3.3) and (3.4) with equation (3.8), one observes that the equivalent, impressed, surface currents are already present in the surface integral of equation (3.8), except that the electric current term must be disregarded on the $z = 0$ surface, as required by the equivalence principle. It is gratifying to observe that the second term in the surface integral can be transformed via application of the identity (A.15), such that $[\hat{n} \times \nabla \times \mathbf{E}(\mathbf{r})] \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') = -[\nabla \times \mathbf{E}(\mathbf{r})] \cdot [\hat{n} \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')]$. This clearly shows that the term is identically zero at $z = 0$, since $\hat{n} \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')|_{z=0} = 0$, thus corresponding with the equivalence principle.

Now that it has been established that the second term in the surface integral of equation (3.8) is zero on the $z = 0$ surface, it can also be observed that $\overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')$ adheres to the Sommerfeld radiation condition as $\mathbf{r} \to \infty|_{z>0}$, therefore this term can be discarded completely, yielding

$$\begin{aligned} \mathbf{E}(\mathbf{r}') &= -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\mathrm{inc}}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV - \int_{S_{z=0}} \left\{ \mathbf{M}_s(\mathbf{r}) \cdot [\nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')] \right\} dS \\ &\quad - \int_{S_\infty - S_{z=0}} \left\{ [\hat{n} \times \mathbf{E}(\mathbf{r})] \cdot [\nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')] \right\} dS. \end{aligned} \tag{3.9}$$

By specifying that all current sources be located at finite distances from the origin one enforces that $\mathbf{E}(\mathbf{r})$ also satisfies the Sommerfeld radiation condition. By further observing that $\mathbf{M}_s(\mathbf{r}) = \hat{n} \times \mathbf{E}(\mathbf{r}) = 0$ everywhere in the $z = 0$ plane, except in the cavity aperture, the conclusion can be drawn that the kernel of the first surface integral in equation (3.9) only has a non-zero value on the cavity aperture, S_a, and that the kernel of the second surface integral vanishes on its domain of integration, leading to

$$\mathbf{E}(\mathbf{r}') = -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\mathrm{inc}}(\mathbf{r}) \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV - \int_{S_a} \left\{ \mathbf{M}_s(\mathbf{r}) \cdot \nabla \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \right\} dS. \tag{3.10}$$

Interchanging the coordinate designations yields

$$\mathbf{E}(\mathbf{r}) = -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\mathrm{inc}}(\mathbf{r}') \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r})\, dV' - \int_{S_a} \left\{ \mathbf{M}_s(\mathbf{r}') \cdot \nabla' \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r}) \right\} dS'. \tag{3.11}$$

Substitute the definition of $\overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')$ (see equation (B.36)), resulting in

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= -jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\mathrm{inc}}(\mathbf{r}') \cdot \overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r})\, dV' \\ &\quad + jk_0 Z_0 \int_{V_\infty} \mathbf{J}^{\mathrm{inc}}(\mathbf{r}') \cdot \left[\overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r}_i) - 2\hat{z}\hat{z} G_0(\mathbf{r}', \mathbf{r}_i) \right] dV' \\ &\quad - \int_{S_a} \left\{ \mathbf{M}_s(\mathbf{r}') \cdot \nabla' \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r}) \right\} dS'. \end{aligned} \tag{3.12}$$

Noting that the first term represents the contribution of free space, impressed current sources located in the $z \geq 0$ half space and that the second term represents their images, the last equation is rewritten as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\mathrm{inc}}(\mathbf{r}) + \mathbf{E}^{\mathrm{ref}}(\mathbf{r}) - \int_{S_a} \left\{ \mathbf{M}_s(\mathbf{r}') \cdot \nabla' \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r}) \right\} dS'. \tag{3.13}$$

As a result of equation (A.17) and the definition of $\overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r})$, it can be stated that

$$\nabla' \times \overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r}) = \nabla' \times \left(\overline{\overline{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r}) - \hat{x}\hat{x} G_{0i}(\mathbf{r}', \mathbf{r}) - \hat{y}\hat{y} G_{0i}(\mathbf{r}', \mathbf{r}) + \hat{z}\hat{z} G_{0i}(\mathbf{r}', \mathbf{r}) \right). \tag{3.14}$$

Since $\mathbf{M}_s(\mathbf{r}')$ only possesses $\hat{x}$ and $\hat{y}$ components and since the surface integral over S_a in equation (3.13) is carried out in the $z' = 0$ plane, the result of equation (B.57) can be used to rewrite equation (3.13) as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\text{inc}}(\mathbf{r}) + \mathbf{E}^{\text{ref}}(\mathbf{r}) - 2\int_{S_a} \left\{\mathbf{M}_s(\mathbf{r}') \cdot \nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right\} dS'. \tag{3.15}$$

Clearly, the procedure leading to equation (3.15) started by replacing the PEC half space and the cavity with a uniform PEC half space and a surface magnetic current distribution at the position of the aperture. The PEC half space is then removed by providing $\mathbf{E}^{\text{inc}}$ and $\mathbf{M}_s(\mathbf{r}')$ with their appropriate images.

3.3 Incorporating the boundary integral in the functional

A relation in the form of the inhomogeneous Neumann boundary condition is first constructed using the boundary integral equation (3.15). This relation is then imposed within the VBVP of equation (2.10), leading to redefinition of the bilinear and linear forms B and L. This is an outward-looking approach in the terminology of [143] and is therefore free of the interior resonance problem [175, 143, 99].

3.3.1 The inhomogeneous Neumann boundary condition

In order to obtain an equation of the form $\hat{n} \times \nabla \times \mathbf{E}|_{S_a}$, start by taking the curl of equation (3.15), yielding

$$\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) &= -jk_0 Z_0 \mathbf{H}^{\text{inc}}(\mathbf{r}) - jk_0 Z_0 \mathbf{H}^{\text{ref}}(\mathbf{r}) \\ &\quad -2\int_{S_a} \nabla \times \left\{\mathbf{M}_s(\mathbf{r}') \cdot \nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right\} dS', \end{aligned} \tag{3.16}$$

where the fact was used, that $\nabla \times \mathbf{E} = -jk_0 Z_0 \mathbf{H}$ in free space. Now use equation (A.14) to obtain

$$\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) &= -jk_0 Z_0 \mathbf{H}^{\text{inc}}(\mathbf{r}) - jk_0 Z_0 \mathbf{H}^{\text{ref}}(\mathbf{r}) \\ &\quad -2\int_{S_a} \nabla \times \left\{\left[\nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right]^T \cdot \mathbf{M}_s(\mathbf{r}')\right\} dS'. \end{aligned} \tag{3.17}$$

From equation (B.33) it is clear that $\left[\nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right]^T = -\nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})$; additionally noting that $\nabla' \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r}) = -\nabla \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})$, the following equation is obtained:

$$\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) &= -jk_0 Z_0 \mathbf{H}^{\text{inc}}(\mathbf{r}) - jk_0 Z_0 \mathbf{H}^{\text{ref}}(\mathbf{r}) \\ &\quad -2\int_{S_a} \nabla \times \left\{\left[\nabla \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right] \cdot \mathbf{M}_s(\mathbf{r}')\right\} dS' \\ &= -jk_0 Z_0 \mathbf{H}^{\text{inc}}(\mathbf{r}) - jk_0 Z_0 \mathbf{H}^{\text{ref}}(\mathbf{r}) \\ &\quad -2\int_{S_a} \left\{\nabla \times \nabla \times \bar{\bar{\mathrm{I}}} G_0(\mathbf{r}', \mathbf{r})\right\} \cdot \mathbf{M}_s(\mathbf{r}')\, dS', \end{aligned} \tag{3.18}$$

the last equation being obtained through the use of equation (A.18). Now substitute the result of equation (B.34), yielding

$$\nabla \times \mathbf{E}(\mathbf{r}) = -jk_0 Z_0 \mathbf{H}^{\text{inc}}(\mathbf{r}) - jk_0 Z_0 \mathbf{H}^{\text{ref}}(\mathbf{r}) - 2k_0^2 \int_{S_a} \overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r}) \cdot \mathbf{M}_s(\mathbf{r}')\, dS'. \tag{3.19}$$

$\hat{n} = \hat{z}$ on S_a, since it is defined as the outward pointing normal with respect to the domain of the variational boundary value problem. Also noting that on a PEC surface $\mathbf{E}^{\text{inc}} = -\mathbf{E}^{\text{ref}}$ and thus $\mathbf{H}^{\text{inc}} = \mathbf{H}^{\text{ref}}$, the inhomogeneous Neumann boundary condition on the cavity aperture can now be stated as

$$\hat{z} \times \nabla \times \mathbf{E}(\mathbf{r})|_{S_a} = -2jk_0 Z_0 \hat{z} \times \mathbf{H}^{\text{inc}}(\mathbf{r}) - 2k_0^2 \hat{z} \times \int_{S_a} \overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r}) \cdot \mathbf{M}_s(\mathbf{r}')\, dS'. \tag{3.20}$$

This inhomogeneous Neumann boundary condition is also called a boundary condition of the third kind [99] or Cauchy type [174].

3.3.2 The FE-BI formulation

According to Maxwell's equations, $\hat{n} \times \frac{1}{\mu_r} \nabla \times \mathbf{E}$ is continuous across any interface where a surface current is not present, such as the parts of the cavity aperture where a homogeneous Dirichlet boundary condition does not apply.

Now substitute the RHS of equation (3.20) in place of $\frac{1}{\mu_r}\mathbf{N}$ in the general VBVP (equation (2.10)), to obtain

$$\begin{cases} \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{C,\Omega} - k_0^2 \langle \epsilon_r \mathbf{E}, \mathbf{W} \rangle_{C,\Omega} + jk_0 Z_0 \langle \mathbf{J}, \mathbf{W} \rangle_{C,\Omega} \\ \quad = 2k_0^2 \int_{S_a} \hat{z} \times \left[\int_{S_a} \overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r}) \cdot \mathbf{M}_s(\mathbf{r}')\, dS' \right] \cdot \mathbf{W}\, dS \\ \quad +2jk_0 Z_0 \int_{S_a} \left[\hat{z} \times \mathbf{H}^{\text{inc}}(\mathbf{r}) \right] \cdot \mathbf{W}\, dS \quad \forall\, \mathbf{W} \in W \\ \quad \mathbf{E} \in W \end{cases} \tag{3.21}$$

(Remember that $\hat{n}$ now represents the outward pointing normal with respect to the variational boundary value problem's domain.)

Substituting the definition of $\overline{\overline{\mathrm{G}}}_e(\mathbf{r}', \mathbf{r})$ into the above variational boundary value problem, Jin [99] shows that the resulting problem can be rewritten with the gradient operators symmetrically transferred to $\hat{z} \times \mathbf{W}$ and $\hat{z} \times \mathbf{E}(\mathbf{r}')$. This is done to reduce the order of the singularity, from R^3 in $\overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r})$ (see equations (B.31) and (B.32)), to order R^1. [194] notes that the order of a surface integral singularity must be $< R^2$ in order that the integral converge *uniformly*, which in turn implies that the techniques discussed in Section D.3 can used. Otherwise one has to resort to the *principal value integral* approach [146, 194], which is much more cumbersome.

This procedure decribed above yields the final form of the VBVP representation of the cavity BVP (a repetition of equation (2.10)):

$$\begin{cases} B(\mathbf{E}, \mathbf{W}) = L(\mathbf{W}) \quad \forall\, \mathbf{W} \in W \\ \mathbf{E} \in W \end{cases}, \tag{3.22}$$

with the bilinear and linear forms now redefined as

$$\begin{aligned} B(\mathbf{E},\mathbf{W}) &= \left\langle \frac{1}{\mu_r}\nabla\times\mathbf{E}, \nabla\times\mathbf{W}\right\rangle_{C,\Omega} - k_0^2\left\langle \epsilon_r\mathbf{E},\mathbf{W}\right\rangle_{C,\Omega} \\ &\quad - 2k_0^2\int_{S_a}[\hat{z}\times\mathbf{W}]\cdot\left\{\int_{S_a}[\hat{z}\times\mathbf{E}(\mathbf{r}')]G_0(\mathbf{r},\mathbf{r}')\,dS'\right\}dS \\ &\quad + 2\int_{S_a}(\nabla\cdot[\hat{z}\times\mathbf{W}])\left\{\int_{S_a}(\nabla'\cdot[\hat{z}\times\mathbf{E}(\mathbf{r}')])\,G_0(\mathbf{r},\mathbf{r}')\,dS'\right\}dS \end{aligned} \tag{3.23}$$

$$L(\mathbf{W}) = -jk_0Z_0\left\langle \mathbf{J},\mathbf{W}\right\rangle_{C,\Omega} - 2jk_0Z_0\int_{S_a}[\hat{z}\times\mathbf{W}]\cdot\mathbf{H}^{\text{inc}}(\mathbf{r})\,dS, \tag{3.24}$$

where $\mathbf{M}_s(\mathbf{r}')$ was replaced with its definition from equation (3.4).

The above bilinear and linear forms of the cavity variational boundary value problem can be substituted into equation (2.17) to yield an equivalent stationary functional formulation for the cavity boundary value problem, since the above bilinear form is symmetric.

3.4 Discretization and the resulting matrix equation

The electric field in the cavity is approximated with the tetrahedral elements described in Appendix C (see [124] for an implementation with the lowest order, curl-conforming, brick elements). The aperture therefore consists of a triangular mesh and its field distribution is that of the volume, evaluated in the $z = 0$ plane.

The discretization process takes place exactly as described in Section 2.4. Discretizing the cavity bilinear and linear forms (equations (3.23) and (3.24)) results in the following elemental contributions to the system matrix equation:

$$[K^e] = \int_{V^e}\left[\frac{1}{\mu_r^e}\{\nabla\times\mathbf{N}^e\}\cdot\{\nabla\times\mathbf{N}^e\}^T - k_0^2\epsilon_r^e\{\mathbf{N}^e\}\cdot\{\mathbf{N}^e\}^T\right]dV \tag{3.25}$$

$$\begin{aligned} [P^{st}] &= 2\int_{S^s}\{\nabla\cdot\hat{z}\times\mathbf{N}^s(\mathbf{r})\}\left\{\int_{S^t}\{\nabla'\cdot\hat{z}\times\mathbf{N}^t(\mathbf{r}')\}^T G_0(\mathbf{r},\mathbf{r}')\,dS'\right\}dS \\ &\quad - 2k_0^2\int_{S^s}\{\hat{z}\times\mathbf{N}^s(\mathbf{r})\}\cdot\left\{\int_{S^t}\{\hat{z}\times\mathbf{N}^t(\mathbf{r}')\}^T G_0(\mathbf{r},\mathbf{r}')\,dS'\right\}dS \end{aligned} \tag{3.26}$$

$$\{b_I^s\} = -2jk_0Z_0\int_{S^s}\{\hat{z}\times\mathbf{N}^s\}\cdot\mathbf{H}^{\text{inc}}\,dS \tag{3.27}$$

$$\{b_J^e\} = -jk_0Z_0\int_{V^e}\mathbf{J}\cdot\{\mathbf{N}^e\}\,dV, \tag{3.28}$$

where the superscripts e and st indicate association with a specific element and association with a specific pair of triangular, elemental faces in the cavity aperture, respectively. $\{\mathbf{N}^e\}$, $\{\mathbf{N}^s\}$ and $\{\mathbf{N}^t\}$ represent the column vectors of vector basis functions associated with element e and those associated with the elements to which boundary faces s and t belong to, respectively.

The resulting system matrix equation is as follows:

$$[A]\{E\} = \{b\} \tag{3.29}$$

with

$$[A] = [K] + [P] \tag{3.30}$$

$$\{b\} = \{b_I\} + \{b_J\}. \tag{3.31}$$

Since the bilinear form is symmetric, the system matrix is symmetric.

3.5 Introducing the excitation

This section discusses the different ways in which the excitation can be introduced into the VBVP formulation. Among other things, a dominant mode, coaxial port formulation is presented.

3.5.1 External incident field excitation

An incident field excitation is incorporated via the contribution of $\{b_I\}$, defined by equation (3.27), to the excitation vector $\{b\}$ in the system matrix equation (3.29). Only the TEM, plane wave case is considered, since it is needed for RCS evaluation, but there is no fundamental or practical reason not to implement incident spherical or cylindrical waves, for example. The calculation of $\mathbf{H}^{\text{inc}}$ for the TEM, plane wave case is discussed in Section B.6.

3.5.2 Internal current probe excitation

An arbitrary, impressed current distribution inside the FEM volume can be introduced into the discretized functional via the contribution of $\{b_J\}$, defined by equation (3.28), to the excitation vector $\{b\}$ in the system matrix equation (3.29).

An impressed current probe excitation is introduced by modeling the probe as a tube of constant, axially directed, surface current density, thus the assumption is made that all current flows on the probe conductor's surface. This assumption is valid since the skin depth becomes very small at microwave frequencies [97]. The probe diameter and length should be kept electrically small, otherwise the approximate assumption that all currents on the probe can be modeled as being in phase and unidirectional, will become invalid.

3.5.3 Dominant mode, coaxial port, boundary excitation

Coaxial ports are incorporated via an inhomogeneous Neumann boundary condition, as described in detail in Appendix E. The coaxial port formulation depends on the assumption that the fields in the coaxial aperture are the sum of purely TEM, dominant mode, incident and reflected parts. The degrees of freedom associated with the coaxial apertures contribute to the system matrix, since they are not part of the homogeneous, Dirichlet boundary.

Two coaxial port formulations are presented in Appendix E. The first is a general, dominant mode approach, based on previous work in [99, 159]. The second formulation [85] is shown to be a special case of the first. It requires that CT/LN elements be used at the port aperture and that the port aperture be meshed as a hexagon, comprising of 6 equilateral triangles. Both formulations were implemented.

3.5.4 Internal voltage source excitation

An internal voltage source can be implemented between two points via an inhomogeneous Dirichlet boundary condition. For example, this can be achieved along an element edge, employing only the Whitney elements. The Whitney function coefficient is calculated from the desired potential difference between the two edge nodes, as follows:

$$E_i = \frac{V_{n_1} - V_{n_2}}{\int_{n_1}^{n_2} \mathbf{N}_i \cdot \hat{l}\, dl}, \tag{3.32}$$

where $\mathbf{N}_i$ represent the Whitney basis function associated with the edge. All other degrees of freedom associated with basis functions that contribute to the tangential field representation at the edge, must be set equal to zero (discarded).

This excitation technique is discussed in [192], but was not implemented because an expression for the feed current is not readily available, which is essential for feed point, input impedance calculation.

3.6 Calculation of observable quantities

After solving the system matrix equation, it still remains to calculate the antenna observables one would possibly like to measure in the physical equivalent of the simulated problem. Calculation of the electric and magnetic fields are discussed first, since most of the other calculations in this section depend on these quantities.

Note that the h subscript again indicates quantities resulting from the discrete, approximate solution to the VBVP; as was the case in Section 2.4.

3.6.1 Calculation of electric and magnetic fields

Fields within the cavity

Since the electric field within the cavity is directly solved by the FEM, its calculation is very straight forward:

$$\mathbf{E}_h(\mathbf{r})|_{z\leq 0} = \{\mathbf{N}^e\}^T\{E^e\} \tag{3.33}$$

where e represents the element within which the observation point lies. $\{\mathbf{N}^e\}$ is the column vector of basis functions defined on K^e. $\{E^e\}$ is the column vector of degrees of freedom associated with the elemental basis functions on K^e.

The magnetic field is calculated by the application of Faraday's law (equation (B.5)) to the above equation, yielding

$$\begin{aligned}\mathbf{H}_h(\mathbf{r})|_{z\leq 0} &= -\frac{\nabla \times \mathbf{E}_h(\mathbf{r})|_{z\leq 0}}{j\omega\mu} \\ &= \frac{j}{\omega\mu^e}\{\nabla \times \mathbf{N}^e\}^T\{E^e\}. \end{aligned} \tag{3.34}$$

Fields above the PEC half space

The electric field above the PEC half space is calculated from the FEM representation of the electric field within the cavity aperture ($z = 0$) by using equation (3.15), as follows:

$$\mathbf{E}_h(\mathbf{r})|_{z>0} = \mathbf{E}^{\text{inc}}(\mathbf{r}) + \mathbf{E}^{\text{ref}}(\mathbf{r}) + 2\int_{S_a}\{\hat{z} \times \mathbf{E}_h(\mathbf{r}')|_{z'=0}\} \cdot \left\{\nabla' \times \bar{\bar{I}}G_0(\mathbf{r}', \mathbf{r})\right\} dS'. \tag{3.35}$$

The magnetic field above the PEC half space is also calculated from the FEM representation of the electric field within the cavity aperture ($z = 0$) by using equation (3.19) and Faraday's law, yielding

$$\mathbf{H}_h(\mathbf{r})|_{z>0} = \mathbf{H}^{\text{inc}}(\mathbf{r}) + \mathbf{H}^{\text{ref}}(\mathbf{r}) + \frac{2jk_0}{Z_0}\int_{S_a}\{\hat{z} \times \mathbf{E}_h(\mathbf{r}')|_{z'=0}\} \cdot \bar{\bar{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r})\, dS'. \tag{3.36}$$

3.6.2 Monostatic RCS

Use equation (3.36) to define the scattered cavity magnetic field as

$$\begin{aligned}\mathbf{H}_h^{\text{sc}}(\mathbf{r})|_{z>0} &= \mathbf{H}_h(\mathbf{r}) - \mathbf{H}^{\text{inc}}(\mathbf{r}) - \mathbf{H}^{\text{ref}}(\mathbf{r}) \\ &= \frac{2jk_0}{Z_0}\int_{S_a}\{\hat{z} \times \mathbf{E}_h(\mathbf{r}')|_{z'=0}\} \cdot \bar{\bar{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r})\, dS'. \end{aligned} \tag{3.37}$$

The monostatic Radar Cross Section (RCS) is defined in terms of the scattered cavity magnetic field resulting from an incident, TEM, plane wave as well as the incident wave itself (see [99] for example). The co-polarized and cross-polarized components of the monostatic RCS are

$$\sigma_{\text{co}}^{\text{mono}} \equiv \lim_{r\to\infty} 4\pi r^2 \frac{|\hat{\eta} \cdot \mathbf{H}^{\text{sc}}(\mathbf{r})|^2}{|\mathbf{H}^{\text{inc}}(\mathbf{r})|^2} \tag{3.38}$$

$$\sigma_{\text{cross}}^{\text{mono}} \equiv \lim_{r\to\infty} 4\pi r^2 \frac{|[\hat{\eta} \times \hat{r}] \cdot \mathbf{H}^{\text{sc}}(\mathbf{r})|^2}{|\mathbf{H}^{\text{inc}}(\mathbf{r})|^2} \tag{3.39}$$

where $\hat{\eta}$ is the polarization unit vector of $\mathbf{H}^{\text{inc}}(\mathbf{r})$. Taking the limit translates in practice to evaluation of the field quantities in the far field. Calculation of the incident field value is discussed in Appendix B.6.

Note that the RCS, as defined above, is that of the cavity alone, otherwise the scattered field would be dominated by the reflection of the plane wave off the infinite ground plane.

3.6.3 Input impedance at a current probe feed

In order to calculate the impedance that is being driven by the current probe described in Section 3.5.2, one needs the probe current and voltage. The probe current is already known since it is impressed. The voltage is calculated by using the definition of potential [97, 154].

Care must be taken when calculating the impedance: if the probe current is defined to flow from point a to point b, that is I_{ab}, then the line integral must be performed from b to a, that is $V_b - V_a$, as follows:

$$V_b - V_a = \int_b^a \mathbf{E}_h \cdot d\hat{l}. \tag{3.40}$$

These conventions correctly define the load current to flow into the positive terminal of the load.

In the case of multiple current probe sources, z- and/or s-parameters can be calculated by considering the correct sets of excitations, as described in Section E.2. Note that arbitrary characteristic impedances must be assigned to the current probe sources in order to calculate s-parameters.

3.6.4 s- and z-parameters at coaxial ports

The calculation of s- and z-parameters for both types of coaxial port models considered, is described in Appendix E.

3.7 Implementation issues

The cavity FE-BI implementation forms part of FEMFEKO, a research FEM kernel for microwave engineering analysis in 3D, developed at the University of Stellenbosch [65]. It is written in the FORTRAN90 programming language. This section gives a general outline of the cavity FE-BI formulation implementation and discusses specific issues encountered in the course of the computer code development. All code used for the purposes of this document was developed by the present author, except the routines reading the input file and the routines calculating the elemental matrices $[K^e]$, defined in equation (3.25).

3.7.1 Input and processing of the problem geometry

The problem geometry is defined and tetrahedrally meshed within a commercial meshing program FEMAP [183], which outputs a mesh file that is then converted, using WinFEFO [76], into a file format that can be read by the FEM kernel. To this file is appended all necessary analysis specifications such as frequency, type of solver, polynomial order of elements, excitation locations, PEC structure locations, material parameters etc..

The kernel starts by reading the analysis parameters and the nodal data of every element into memory. Next, all the edges and faces in the mesh are identified and numbered globally, therefore many edge and face numbers are shared by more than one element. The edges and faces where a homogeneous Dirichlet boundary condition must be imposed (cavity wall and

PEC structures) are flagged. The edges and faces that form part of a port are flagged. The edges and faces that belong to the cavity aperture are flagged. All the edges and faces that do not form part of a PEC surface are flagged as free.

Every free edge and free face will have a number of basis functions associated with it, depending on the polynomial order of the elements that share that edge or face. For every basis function that every free edge and free face support, a degree of freedom is assigned to that edge or face, therefore an edge or face can have more than one degree of freedom associated with it. In this way, the total number of degrees of freedom, N_F, is counted as well. This is the necessary basis needed for setting up the system matrix equation.

3.7.2 Setting up and solving the system matrix equation

The system matrix has two principal contributions: the sparse FE contribution of equation (3.25) and the fully populated BI contribution of equation (3.26).

First, storage is allocated for the system matrix. Single precision storage is used throughout (8 bytes per complex number). In the case of a direct solution this amounts to allocating storage for a square, complex matrix with dimension equal to the total number of degrees of freedom. An iterative solution can benefit from storing the FE contribution in a sparse format, but the BI matrix always needs to be stored fully, except when the FMM is used (see Chapter 4). The sparse format used is CRS (compressed row storage) [165, 192]. Note that a direct solution based on sparse storage is also possible, but the computational efficiency of such a procedure will depend on the sparsity of the LU decomposition, which can vary.

The elemental and facial contributions to the system matrix $[A]$ and excitation vector $\{b\}$ are calculated as follows:

- Cycling through all the elements, their elemental FE contributions are calculated using existing routines. These routines employ analytical- or numerical evaluation of the integral in equation (3.25), as described in Section D.2.
- Calculate the facial BI interaction in equation (3.26) for every pair of faces that were flagged as lying in the cavity aperture. The integration scheme is presented in Section D.3. In the case of the self-term, a special integration scheme is used, which has the advantage that it can easily be extended to arbitrary polynomial order basis functions. The non-self-terms are evaluated with triangular surface quadrature rules.
- Cycling through all the coaxial port faces that should be handled with the general port formulation, their system matrix- and excitation vector contributions, as defined by equations (E.33) and (E.34), are calculated using triangular surface quadrature (see Section D.1.1).
- Cycling through all the coaxial port faces that should be handled with the Whitney element port formulation, their system matrix- and excitation vector contributions, as defined by equations (E.53) and (E.54), are calculated.
- The contribution of an incident field at the aperture to the excitation vector is calculated by cycling through all the aperture faces and evaluating equation (3.27) with triangular surface quadrature (see Sections B.6 and D.1.1).

- The contributions of probe current sources to the excitation vector are calculated using simple trapezoidal quadrature.

All these elemental and facial contributions are assembled into the system matrix and excitation vector by using the associations between degree of freedom numbers and the edges and faces, as made earlier.

The direct solution of the system matrix equation $[A]\{E\} = \{b\}$ is accomplished by using the appropriate *LAPACK* routines for *LU* decomposition and back substitution [12] (full matrix storage only). The following iterative solution algorithms are used: Conjugate gradient (CG), Bi-conjugate gradient (BiCG) or Quasi-minimal residual (QMR), with simple diagonal preconditioning [165, 25, 192]. Generally, BiCG was found to be the most efficient, but sometimes there is a problem with convergence and one needs to switch over to another scheme.

3.7.3 Extracting the observable quantities

After the solution of the system matrix equation, the various observable quantities are calculated as follows:

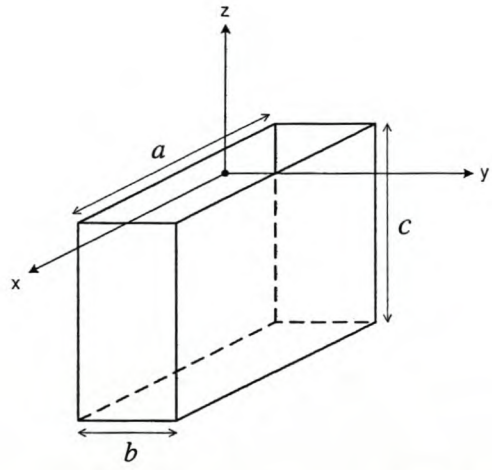
- Calculating the electric (or magnetic) field within the cavity (the FE region), as shown in Section 3.6.1, amounts to a straight-forward evaluation of the basis functions (or their curl) within the element where the field point is located.
- Calculating the electric (or magnetic) field above the cavity (the BI region), as shown in Section 3.6.1, amounts to a surface integral over the cavity aperture. This integral is carried out by using triangular surface quadrature on every aperture face (see Section D.1.1). If the requested field point lies within the aperture, then it is actually evaluated at very small distance (typically $\frac{\lambda_0}{100}$) above the aperture, in order to avoid the singularity. In such a case, the quadrature rule used for the aperture face immediately below the field point, is set to a very high degree of accuracy.
- The RCS is calculated by using the field calculation procedure for points above the aperture (described above), together with the equations in Section 3.6.2. If the RCS is required over an angular range and an iterative solution is used, then the system matrix equation solution has to be repeated at every angle of incidence, because $\{b\}$ changes at every angle. This is not a problem in the case of a direct solution, because the *LU* decomposition of $[A]$ is not affected by the value of $\{b\}$.
- The voltage of an impressed current source is calculated by evaluating the relevant equation in Section 3.6.3 with a trapezoidal quadrature rule. The kernel is a function of $\mathbf{E}_h$ within the FE region, of which the evaluation is discussed above.
- Port parameter calculation for the general, dominant mode coaxial port formulation, rest upon the evaluation of equation (E.40). The integral is evaluated with triangular surface quadrature (see Section D.1.1).
- Port parameter calculation for the Whitney element, dominant mode coaxial port formulation, rest upon the evaluation of equation (E.57), which is straight-forward.

3.8 Results

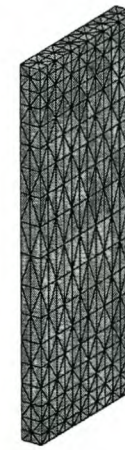
This section presents some results to illustrate the application of the cavity FE-BI- and coaxial port formulations.

RCS of cavities

Monostatic RCS results with CT/LN elements, presented in [37, 36]. Figures 3.2 and 3.4 present the empty- and lossy cavity geometries considered, along with their discretizations. Results for the pattern cut $0 \leq \phi \leq 90°$ with $\theta = 40°$ are presented in Figures 3.3 and 3.5. The definition of the plotted value, normalized RCS, is $\frac{\sigma}{\lambda^2}$. The results are compared with data from [160], where a similar code was developed, also using tetrahedral, CT/LN elements. E-polarization refers to the co-polarized RCS, with $\eta = 0°$. H-polarization refers to the co-polarized RCS, with $\eta = 90°$. X-polarization refers to the cross-polarized RCS, with $\eta = 0°$ (or $\eta = 90°$).



(a) $a = 0.7\lambda$. $b = 0.1\lambda$. $c = 1.73\lambda$.



(b) 1467 elements. $\frac{\lambda_0}{10}$ average edge length.

Figure 3.2: Geometry and mesh of the empty cavity.

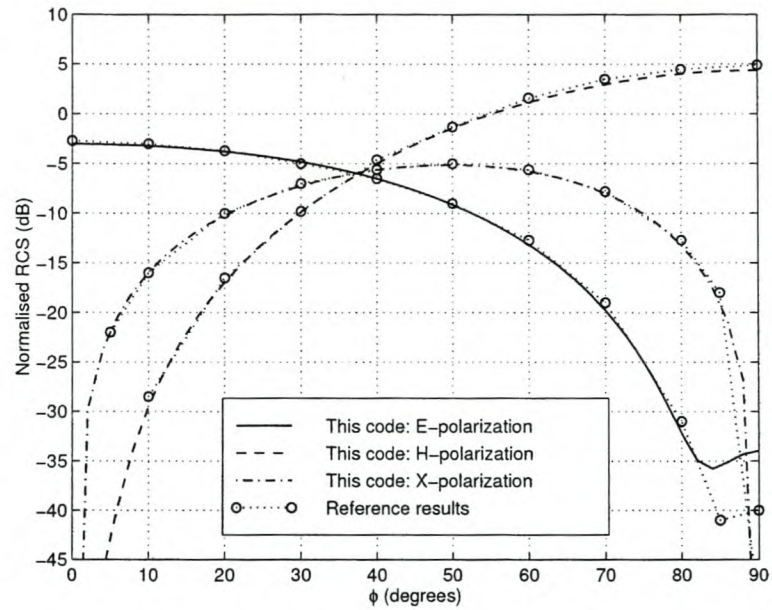
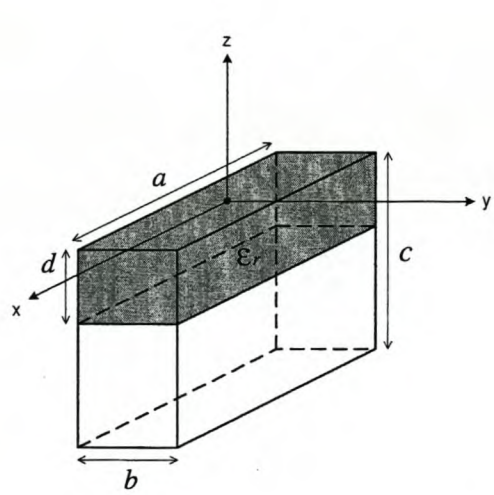
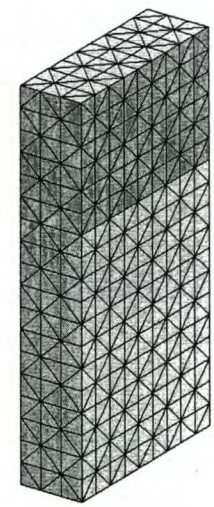


Figure 3.3: Normalized RCS of the empty cavity problem, shown in Figure 3.2. Compared with results presented in [160].



(a) $a = 0.3\lambda$. $b = 0.1\lambda$. $c = 0.6\lambda$. $d = 0.2\lambda$. $\epsilon_r = 2 - j2$.



(b) 1787 elements. $\frac{\lambda_0}{22}$ average edge length.

Figure 3.4: Geometry and mesh of the lossy cavity.

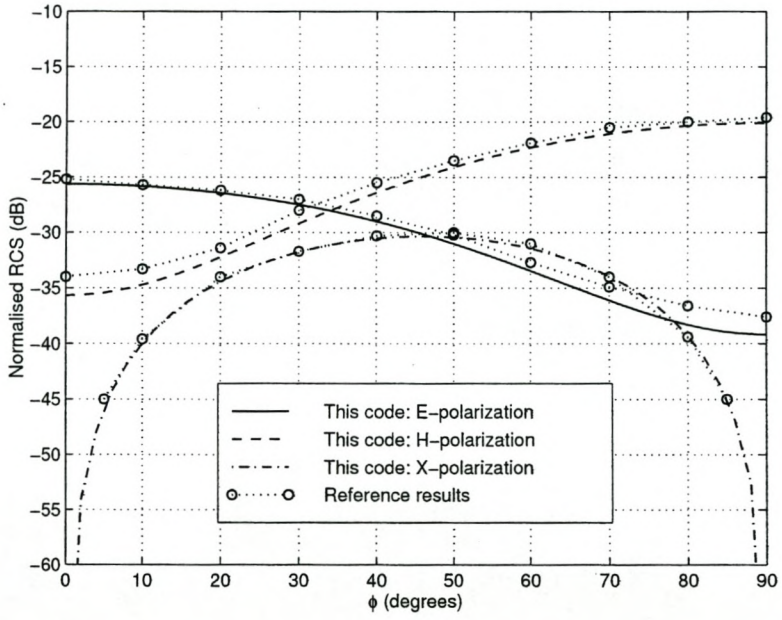
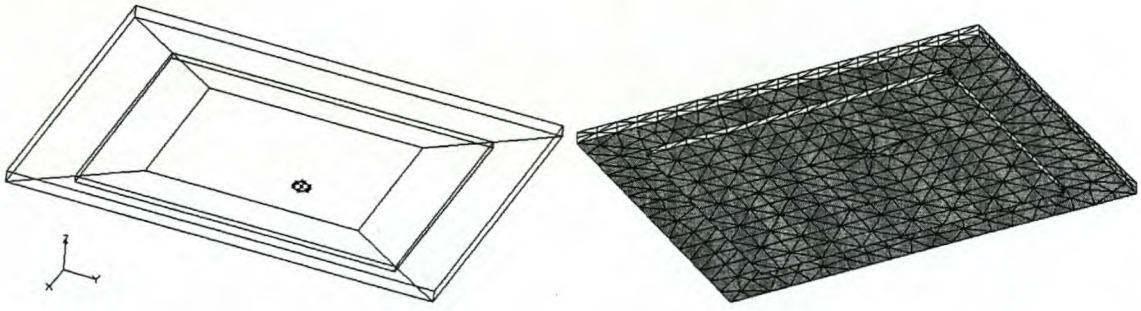


Figure 3.5: Normalized RCS of the lossy cavity problem, shown in Figure 3.4. Compared with results presented in [160].

Microstrip patch antenna within a stepped cavity

Input impedance results for a coaxially fed, microstrip patch antenna within a stepped cavity, were presented in [37, 36, 40]. (This structure was also considered in [124].) Two different meshes were used in order to apply all three possible excitation methods: excitation with a current probe or with the Whitney element, dominant mode coaxial port formulation or with the general, dominant mode coaxial port formulation. Figures 3.6 and 3.7 show the geometries of the two meshes for solving the same problem. The superstrate properties are $60\text{mm} \times 80\text{mm} \times 2.27\text{mm}$ with $\epsilon_r = 1$. The substrate properties are $45\text{mm} \times 60\text{mm} \times 1.135\text{mm}$ with $\epsilon_r = 2.48$. The patch dimensions are $30\text{mm} \times 40\text{mm}$ with feed location at $(x, y) = (5.0\text{mm}, 5.3692\text{mm})$.

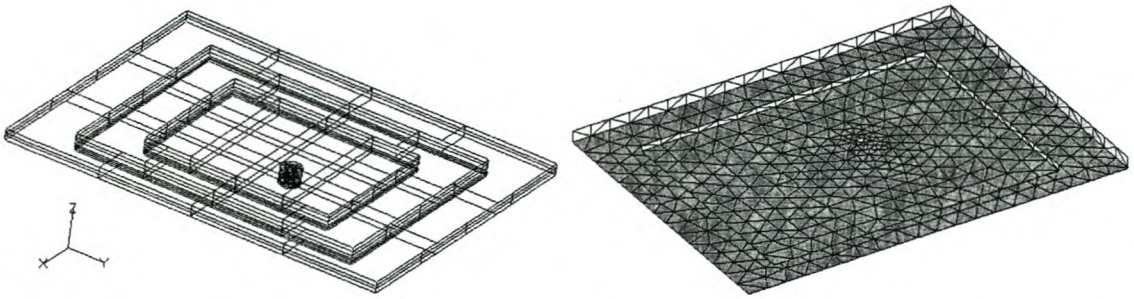
Figures 3.8 and 3.9 present the S_{11} results. In the CT/LN case, where all three feeding schemes can be applied, they produce near identical results. This is a consequence of the very thin substrate ($\frac{\lambda}{56}$), which makes the assumption that the currents on the probe are all in phase, a very accurate one. Figure 3.8 also shows a probe feed result associated with a mesh not shown here. It demonstrates the improvement afforded by decreasing the element size (to 4658 elements with $\frac{\lambda_0}{29.4}$ average edge length), while keeping the element polynomial order the same. In Figure 3.9, observe the effect of increasing the polynomial order of the elements, on the results. The fact that the LT/LN result is so accurate can be attributed to chance more than anything else. One can expect erratic results from the LT/LN elements, because the polynomial order to which they approximate the field and its curl, is in strong imbalance (see the section on mixed- and full order elements in Appendix C). Other than the LT/LN result, all the others improve in sequence of increasing polynomial order, as one would expect. The QT/QN results is well within manufacturing tolerance from the measurement. The numbers of degrees of freedom for the various solutions are listed in Table 3.1. For this problem, further results will be presented in Section 6.3.



(a) Coaxial radii: $a = \frac{b}{100}$ and $b = 1.475\text{mm}$.

(b) 2971 elements. $\frac{\lambda_0}{24.2}$ average edge length (3 GHz).

Figure 3.6: Geometry and mesh of a stepped cavity, microstrip patch antenna for excitation with a current probe or the Whitney element, coaxial port formulation.



(a) Coaxial radii: $a = 1\text{mm}$ and $b = 2\text{mm}$.

(b) 3433 elements. $\frac{\lambda_0}{26.6}$ average edge length (3 GHz) .

Figure 3.7: Geometry and mesh of a stepped cavity, microstrip patch antenna for excitation with the general coaxial port formulation.

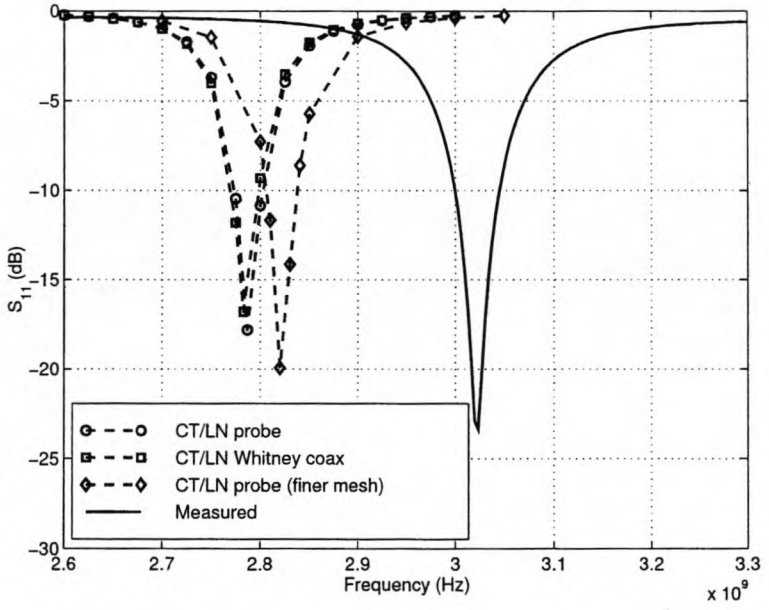


Figure 3.8: Stepped cavity patch antenna s_{11} results with the mesh shown in Figure 3.6. Comparison between a probe feed and the Whitney element, coaxial port formulation. $Z^c = 50\,\Omega$.

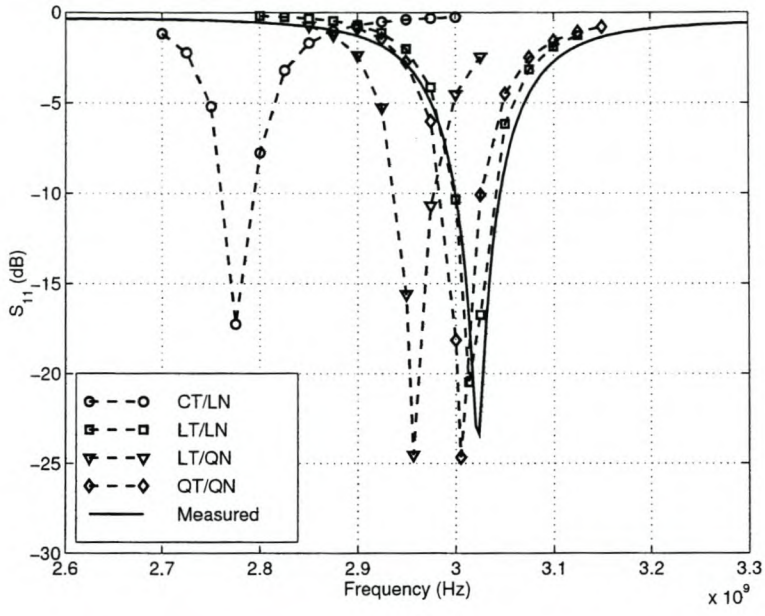


Figure 3.9: Stepped cavity patch antenna s_{11} results with the mesh shown in Figure 3.7. General coaxial port formulation. Solutions with elements of varying polynomial order. $Z^c = 50\,\Omega$.

Table 3.1: Numbers of degrees of freedom for the uniform element polynomial order solutions of the stepped cavity patch antenna problem.

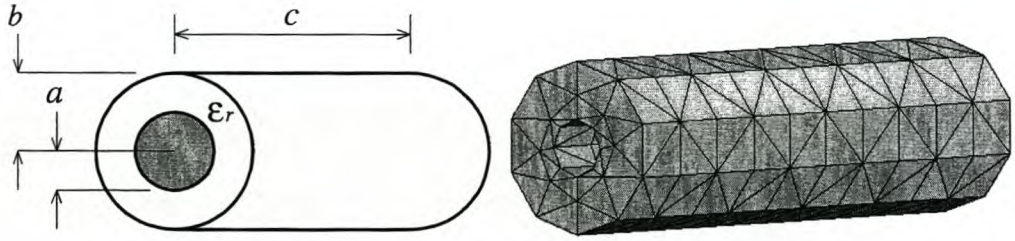
Results	Geometry	Excitation type	Element order	D.o.f.
Figure 3.8	Figure 3.6	Current probe	CT/LN	2957
·	·	Whitney port	CT/LN	2962
·	Not shown	Current probe	CT/LN	4707
Figure 3.9	Figure 3.7	General port	CT/LN	3465
·	·	General port	LT/LN	6930
·	·	General port	LT/QN	19970
·	·	General port	QT/QN	29955

Coaxial line segment

Consider the problem of a coaxial line segment, as shown in Figure 3.10. Tables 3.2 and 3.3 list s-parameter results, considering the structure as a 1-port (shorted) and a 2-port (through), respectively. The general, dominant mode, coaxial port formulation of Section E.3 is used. The results improve as the element polynomial order is increased, indicating that the coaxial geometry is accurately represented by the mesh and that the port formulation functions correctly.

This problem cannot be solved through use of the Whitney element, dominant mode, coaxial port formulation of Section E.4. Supposing the Whitney element formulation is used, then it requires the coaxial line that terminates at the port, to be modeled with an infinitely thin center conductor (a single edge, $Z^c \to \infty$), since the port aperture is required to be meshed as 6 equilateral triangles. On the other hand, the Whitney element formulation incorporates the non-zero center conductor radius (finite Z^c) through the formulation and not through the mesh geometry (the opposite is true for the general model), leading to an unavoidable impedance mismatch at the port face.

To use the Whitney element formulation (as well as the general coaxial port formulation), the assumption that a coaxial, TEM mode distribution exists in the port aperture must be valid. This can be insured in two ways. The first possibility is to model a coaxial line segment between the port and the structure under consideration, such that any higher order, non-propagating modes will have attenuated sufficiently at the port surface. The second possibility is to make the port aperture very small. Either one of these techniques can be used with the general formulation, but in the case of the Whitney element formulation, the latter must always apply and the former never, severely restricting its use.



(a) $a = 0.025$m. $b = 0.075$m. $c = 0.5$m. $\epsilon_r = 1$. $Z^c = 65.871\,\Omega$.

(b) 589 elements. 0.052m average edge length.

Figure 3.10: Geometry and mesh of a coaxial line segment.

Table 3.2: s-parameter of the coaxial line segment shown in Figure 3.10, with one end shorted.

Element order	D.o.f.	$f = 300$MHz		$f = 400$MHz	
		$\|S_{11}\|$	$\angle S_{11}$	$\|S_{11}\|$	$\angle S_{11}$
CT/LN	417	0.99959	173.82°	0.90459	56.224°
LT/LN	834	0.99998	−168.01°	0.99918	81.037°
LT/QN	2830	0.99999	178.23°	0.99900	58.710°
QT/QN	4245	0.99990	−178.13°	0.99910	62.954°
Exact	–	1.00000	180.00°	1.00000	60.000°

Table 3.3: s-parameters of the coaxial line segment shown in Figure 3.10, with two ports.

Element order	D.o.f.	$f = 300$MHz		
		S_{11}	$\|S_{21}\|$	$\angle S_{21}$
CT/LN	459	$-0.058947 + j0.013300$	0.93961	177.52°
LT/LN	918	$-0.0026573 + j0.016822$	0.99918	−173.04°
LT/QN	2982	$-0.00056800 + j0.0053777$	0.99932	179.42°
QT/QN	4473	$-0.00078667 + j0.0033287$	0.99930	−178.88°
Exact	–	$0.0 + j0.0$	1.00000	180.00°

Element order	D.o.f.	$f = 400$MHz		
		S_{11}	$\|S_{21}\|$	$\angle S_{21}$
CT/LN	459	$-0.072045 - j0.0064747$	0.93955	116.23°
LT/LN	918	$-0.019537 - j0.015528$	0.99898	129.58°
LT/QN	2982	$-0.0046557 - j0.0021571$	0.99929	119.21°
QT/QN	4473	$0.0010820 + j0.0012005$	0.99929	121.52°
Exact	–	$0.0 + j0.0$	1.00000	120.00°

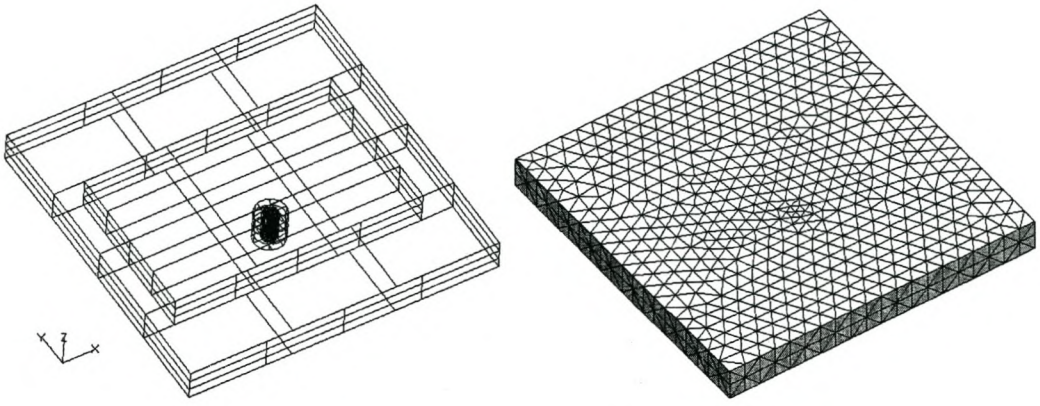
Comparison between microstrip patch excitation schemes

This section presents an experimental, numerical study of the effect that the different feeding schemes have on the simulated input impedance of a microstrip patch antenna on a very thick substrate, recessed in an infinite ground plane. The antenna geometry and mesh is shown in Figure 3.11. The cavity dimensions are $x \times y \times z = 1\text{m} \times 1\text{m} \times 0.1\text{m}$, with $\epsilon_r = \mu_r = 1$. The patch is centered with dimensions $x \times y = 0.8\text{m} \times 0.5\text{m}$, with feed location between the cavity wall and the patch at $(x, y) = (0.0\text{m}, -0.1\text{m})$. The substrate thickness is $\frac{\lambda}{10}$ at the intended resonant frequency (300MHz) of the patch itself.

On this single mesh, the current probe-, Whitney coaxial port- and general coaxial port excitation schemes are compared, using CT/LN elements throughout. The current probe is employed as a single vertical line source. The Whitney coaxial port (see Appendix E) is employed on an aperture with radii $b = 0.025\text{m}$, $a = \frac{b}{200}$. The general coaxial port is employed on an aperture with radii $b = 0.05\text{m}$, $a = 0.025\text{m}$. In all three cases, the port aperture center is connected to the patch with a PEC conductor modeled as infinitely thin. A reference result was generated by considering the same mesh, but with an added coaxial cable with radii $b = 0.05\text{m}$, $a = 0.025\text{m}$ and length 0.4m, extending downwards, away from the port. At this extended position, the general port formulation becomes 'exact', as discussed in conjunction with the coaxial line segment problem. For comparison, the reference results were de-embedded. They represent the most accurate CT/LN solution possible on this mesh. (Unavoidably, dispersion error along the line segment degrades the reference results slightly).

Obviously, results obtained on this mesh with higher polynomial order elements (and with the added coaxial line segment), will be more accurate in comparison with the true solution, but the aim here is to highlight the effects of using the different feeding schemes, based on the same discretization of the solution space.

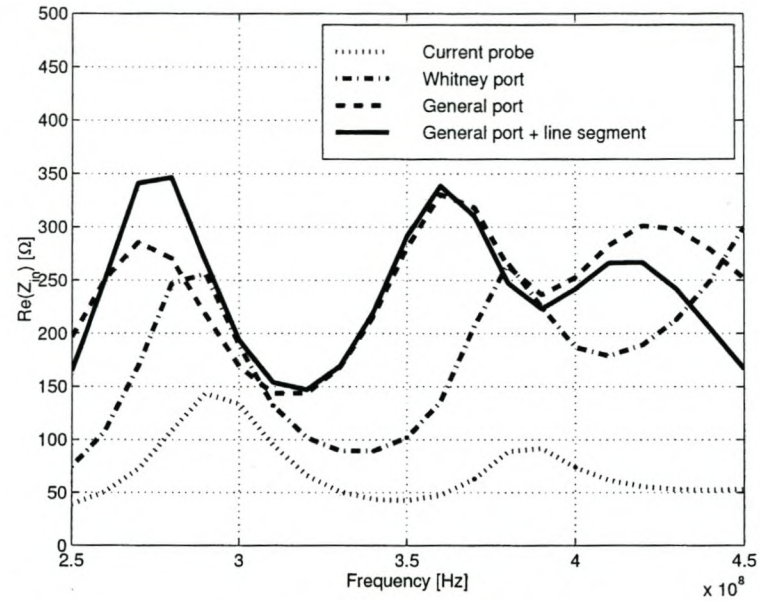
Figure 3.12 presents the three input impedance results, compared with the reference result. The probe current source is completely ineffective, since its length is too long to support the argument for constant phase along its length. The Whitney element port improves substantially upon the probe current source, but is still somewhat inferior to the general port formulation (of which it is a special case). The general port results compare most favourably, but are still in error, relative to the reference results — such a comparison firstly demonstrates the presence of higher order modes at the coaxial aperture in the cavity wall and secondly demonstrates the adverse effect that these modes have on the dominant mode, port formulation. Clearly, an exact feed formulation is needed when dealing with thick substrates (and feeds in general), as opposed to the thin substrate case, where all three feeding schemes gave the same results (see the previous section on a patch antenna in a stepped cavity).



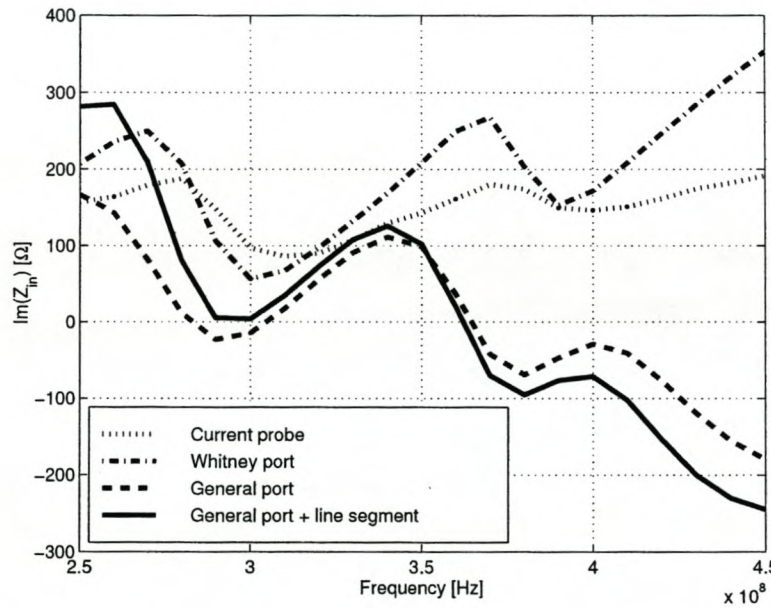
(a) Geometry.

(b) 7426 elements. 0.0518m average edge length.

Figure 3.11: Geometry and mesh of the patch on a thick substrate; used to compare excitation schemes.



(a) Real component of Z_{in}.



(b) Imaginary component of Z_{in}.

Figure 3.12: Input impedance vs. frequency of the patch on a thick substrate. A comparison between various excitation formulations.

Coupling between microstrip patch antennas on a perforated substrate

This section is based on [38]. Consider microstrip patch antennas on a substrate embedded in a conducting plane. The substrate around the patches is perforated, as shown in Figure 3.13. Such structures are geometrically similar to photonic bandgap (PBG) materials (such as used in [86, 52] as patch substrates), but are viewed differently: here the substrate perforation is used as a convenient manufacturing technique to produce regions of varying substrate permittivity, as in [56, 55], of which this section can be viewed as an extension. The possibility of using this permittivity variation to shield two patches from each other by interfering with surface wave propagation and thus reducing mutual coupling is investigated numerically.

Concerning the antenna design, note the following desirable substrate characteristics:

- Increased thickness, generally leading to an increase in bandwidth [91].
- High permittivity, to minimize the size of the radiating structures.

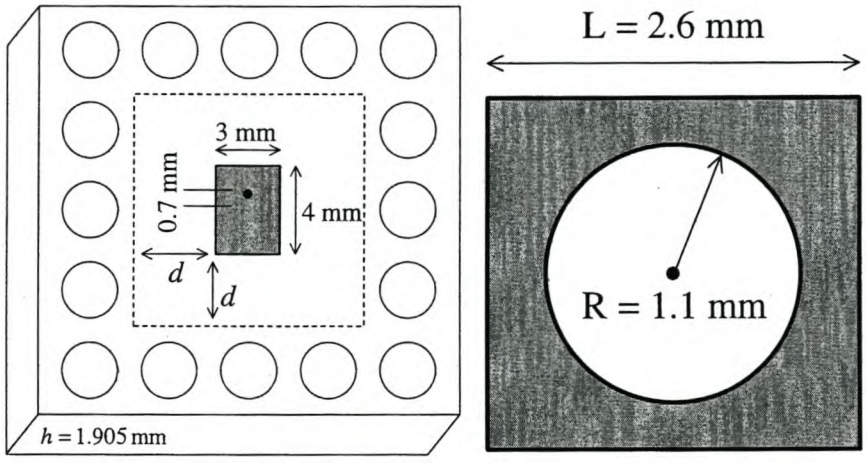
For these reasons the high permittivity, thick substrate case is of interest. Unfortunately, these characteristics also lead to more pronounced surface waves [91], causing unwanted coupling between array elements or with substrate edges, motivating this investigation. A substrate with $h = 1.905$mm and $\epsilon_r = 10.2$ was chosen with a 3mm $\times$ 4mm patch. The feed is 0.7mm offset from the patch centre, in the larger dimension's direction. This closely corresponds to the patch geometry used in [56], where the effects of substrate perforation on a single patch is investigated with the FDTD. The distance between the two patches is set to 17mm $\approx 0.5\lambda_0$ here. The minimum value of the parameter d (see Figure 3.13), such that the influence on S_{11} is negligible, was found to be 3.5mm in [56]. With the array spacing restriction this value results in too little influence on mutual coupling, therefore this criterion is abandoned here. The only criterion is to reduce mutual coupling (thus to lower Z_{21}). 56% of the substrate was replaced by holes, resulting in a volumetric average permittivity of 5.03 in the perforated regions. To ensure that the volumetric average is applicable the perforation cell must be designed such that all dimensions are below $0.5\lambda_s$, according to [56]. Figure 3.13 shows the perforation cell geometry.

The cavity FE-BI formulation is used for the analysis, to which it is well suited. CT/LN elements are used. Concerning the excitation, an enforced current probe between the patch and the cavity wall could be used to emulate a coaxial feed, but this becomes inaccurate if the substrate is electrically thick, due to feed pin inductance, therefore the Whitney element coaxial port formulation of Section E.4 is used. All results were obtained with $0.033\lambda_0$ discretization.

This paragraph discusses the results for a single patch. Figure 3.14 shows the analysis model. The s-parameter versus frequency, for different values of d, is shown in Figure 3.15. Clearly the smaller d is, the greater the influence on the impedance. A plot is also shown where the perforation was replaced by a volumetric average material, showing the validity of the assumption. Investigation of the directivity of a single patch on a $\lambda_0 \times \lambda_0$ substrate revealed that a large broadside dip in directivity can virtually be eliminated by bringing the perforation close enough to the patch, presumably as a result of less surface wave scattering from the substrate edges. This was also reported in [56].

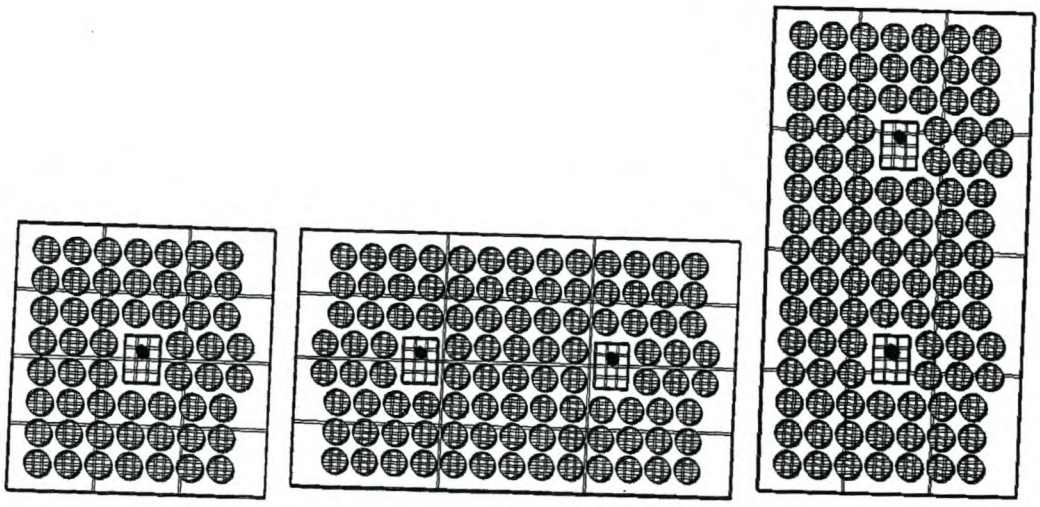
This paragraph discusses the results for arrays of two patches. Figure 3.14 shows the horizontally- and vertically spaced analysis models. Figures 3.16 and 3.17 show the s-parameters and $|Z_{21}|$, both versus frequency, for the horizontally spaced case with varying d. The same data is presented in Figures 3.18 and 3.19, for the vertically spaced case. Through substrate perforation, the magnitude of the transfer impedance at resonance was reduced by 40% in the horizontally spaced case and by 35% in the vertically spaced case (both with $d = 0.5\,\text{mm}$).

This numerical investigation has shown that the possibility does exist to reduce mutual coupling between microstrip patch array elements significantly through the use of substrate perforation, without adverse effect, except for tuning. Further study in order to establish more concrete design guidelines as well as the investigation of PBG substrates for microstrip patch arrays is definitely warranted.



(a) Antenna element. (b) Perforation cell.

Figure 3.13: Geometry of the microstrip patch antenna with substrate perforation.



(a) Single patch. 22mm × 22.8mm substrate.

(b) Horizontally spaced patches. 39mm × 22.8mm substrate.

(c) Vertically spaced patches. 22mm × 41mm substrate.

Geometry	Number of elements	Average edge length	CT/LN d.o.f.
Single	7969	1.075mm	9001
Horiz.	14874	1.049mm	16899
Vert.	14972	1.050mm	16984

(d) Analysis data.

Figure 3.14: Geometry models and analysis data of the perforated substrate microstrip patch antennas analyzed.

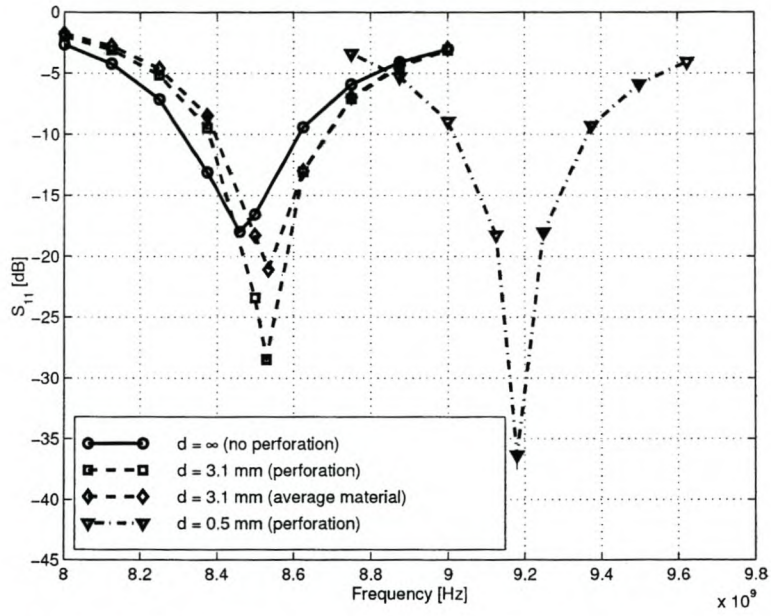


Figure 3.15: s-parameter versus frequency with varying d, for the single patch on a perforated substrate. $Z^c = 50\,\Omega$.

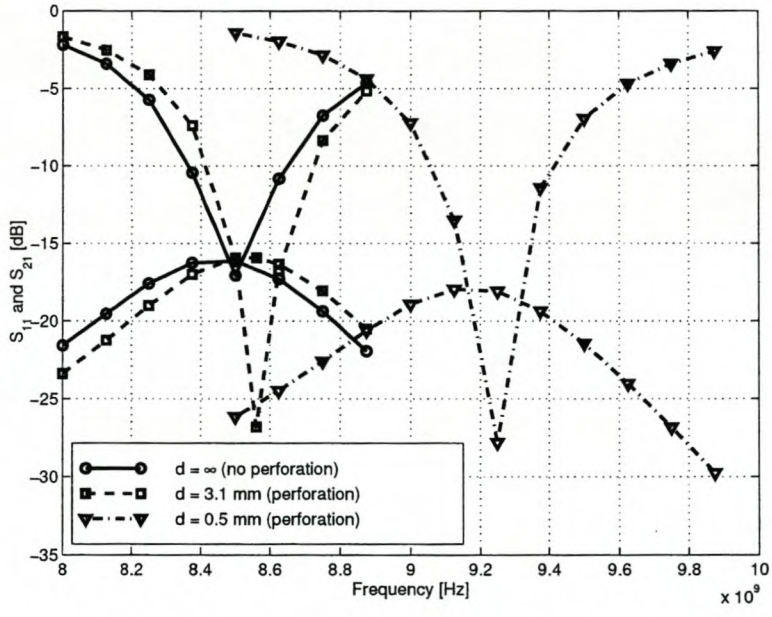
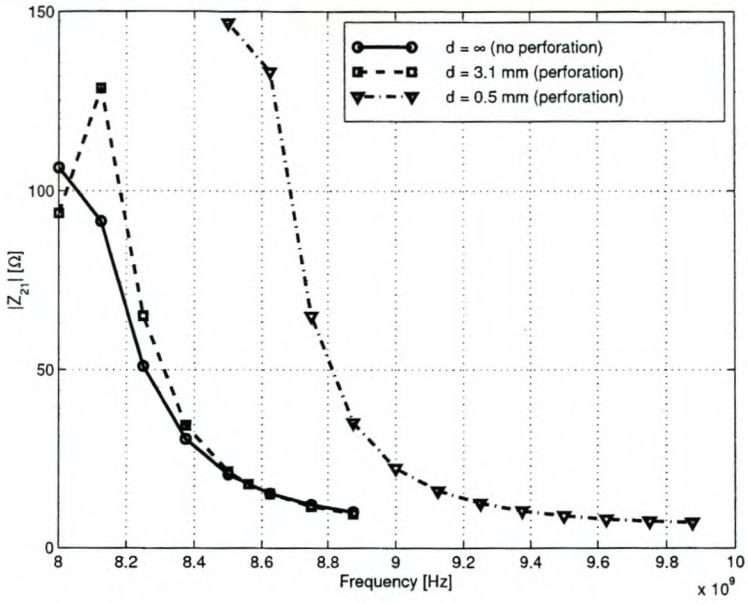


Figure 3.16: s-parameters versus frequency with varying d, for the horizontally spaced pair of patches on a perforated substrate. $Z^c = 50\,\Omega$.



(a) $|Z_{21}|$ versus frequency.

d	f_R	$\lvert Z_{21}\rvert$
∞ mm	8.50 GHz	20.7Ω
3.1 mm	8.56 GHz	18.0Ω
0.5 mm	9.25 GHz	12.6Ω

(b) $|Z_{21}|$ at resonance.

Figure 3.17: $|Z_{21}|$ data with varying d, for the horizontally spaced pair of patches on a perforated substrate.

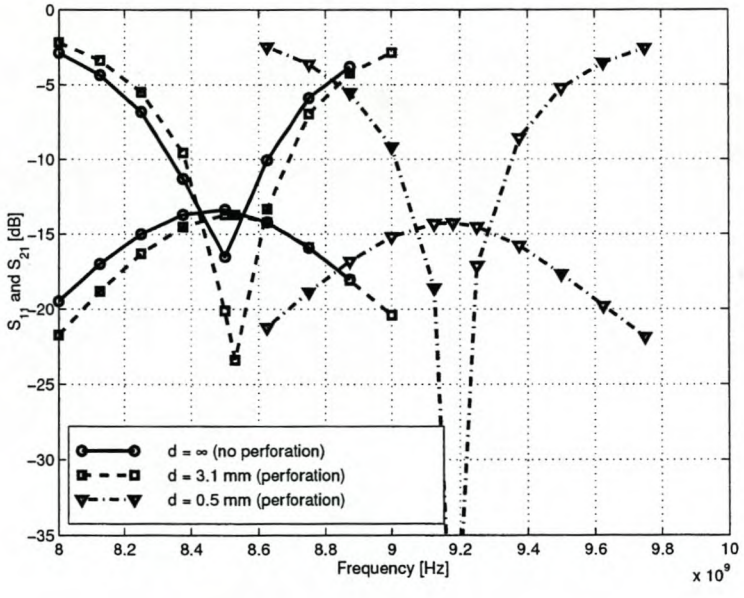
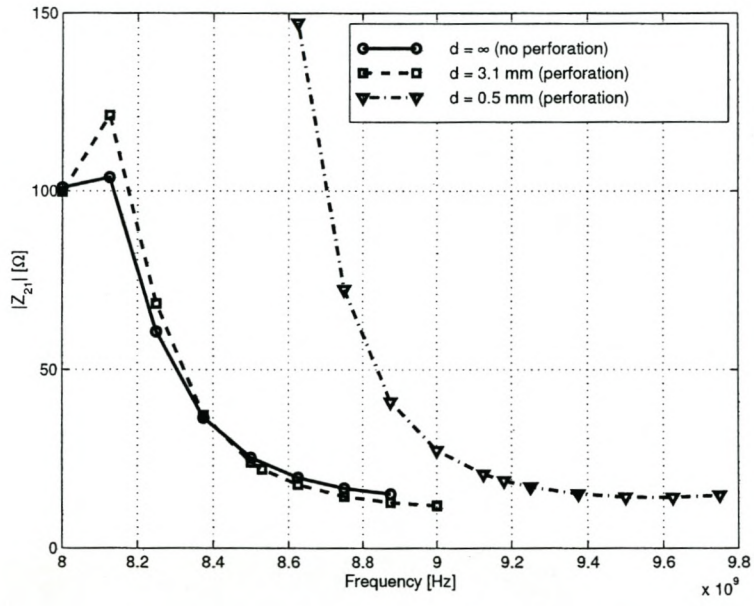


Figure 3.18: s-parameters versus frequency with varying d, for the vertically spaced pair of patches on a perforated substrate. $Z^c = 50\,\Omega$.



(a) $|Z_{21}|$ versus frequency.

d	f_R	$\lvert Z_{21}\rvert$
∞ mm	8.46 GHz	29.0Ω
3.1 mm	8.53 GHz	22.2Ω
0.5 mm	9.18 GHz	18.9Ω

(b) $|Z_{21}|$ at resonance.

Figure 3.19: $|Z_{21}|$ data with varying d, for the vertically spaced pair of patches on a perforated substrate.

3.9 Conclusion

This chapter presented a cavity backed aperture FE-BI formulation. Excitation, post-processing and implementation issues were extensively discussed, including a dominant mode, coaxial port model. Various results were presented, demonstrating the use and effectiveness of the formulations. It was shown that increases in elemental polynomial orders lead to substantially improved results.

The main contributions of this chapter have been the following: A BI self-term evaluation scheme for the cavity FE-BI formulation that can easily take into account boundary elements of arbitrary polynomial order was presented. A dominant mode coaxial port model was rederived and another existing coaxial port model was shown to be a special case thereof. Novel results for the coaxial port model was presented, adding some insight into using the technique. Novel higher order element results were presented. Novel simulated results on the coupling between microstrip patch antennas on a perforated substrate was presented.

In the rest of this document, further extensions will be made, and techniques applied, to the formulations presented in this chapter. In Chapter 4, the FMM will be applied to the BI component of the FE-BI formulation presented in this chapter, in order to increase its computational efficiency. In Chapters 5 and 6, a posteriori error estimators will be constructed for the formulations presented in this chapter and will be used to drive adaptive refinements to the discretization.

Chapter 4

The cavity FE-BI-FMM formulation: theory, implementation and results

4.1 Introduction

This chapter presents the application of the Fast Multipole Method (FMM) to the BI component of the cavity FE-BI formulation presented in Chapter 3. In this chapter, N_F will refer to the number of BI unknowns, rather than the number of FE-BI unknowns.

Solution schemes for electromagnetic, time-harmonic radiation/scattering problems, based on the discretization of related integral equations, are grouped together under the term *Method of Moments* (MoM) [194]. The MoM leads to fully populated system matrices, since the integral relations between basis functions are of a non-local nature due to the presence of the Green function. These matrices require $\mathcal{O}\left(N_F^2\right)$ storage as well as $\mathcal{O}\left(N_F^2\right)$ time for matrix setup and for a matrix-vector product. The reduction of these orders of complexity would lead to improved computational efficiency, translating into the solution of larger problems than before. The FMM algorithm reduces these complexities.

In its standard form, the FMM reduces the complexities of the above operations to $\mathcal{O}\left(N_F^{1.5}\right)$, as shown in Section 4.4 and also mentioned in [53, 61, 179, 123, 30, 121, 163]. In its multilevel form, it is possible to reduce these complexities further to $\mathcal{O}\left(N_F(\log N_F)^2\right)$ [62, 89, 122] or $\mathcal{O}\left(N_F \log N_F\right)$ [179, 177, 122], depending on the accuracy of the implementation. An efficient implementation will reduce the constant multipliers of these complexities [89, 63].

Generally speaking, the FMM algorithm can be defined as a technique to factorize a translation operator, such that the arguments are confined to separate factors. A translation operator is a function describing the interaction between two spatial coordinates, usually the Green function associated with the governing PDE and geometry of the problem at hand. In the FMM literature, this factorization procedure is referred to as *diagonalising the translation operator*. The general approach is to replace the Green function with an equivalent addition theorem (series expansion) and then rewrite the terms of the series in the desired, factorized form. When modeling electromagnetic fields, the governing PDE is the Laplace equation in the case of static problems and the Helmholtz equation in the case of time-harmonic, radiation/scattering problems. See [78, 50, 113, 53, 164] for the 3D, homogeneous medium, Helmholtz equation case (with the zero-th order, spherical Hankel function as Green function) and [121, 122, 163] for the 2D, homogeneous medium, Helmholtz

equation case (with the zero-th order, Hankel function as Green function). Note that [163] is historically the first formulation of the FMM algorithm for a time-harmonic problem (2D acoustic scattering).

The translation operator's diagonalization together with a positional grouping of the basis functions comprise the FMM algorithm. These two steps allow the source and observation coordinates to be separated to the extent that the conventional, full system matrix can be factorized into sparse components, hence the improvement in computational efficiency. Only the interactions between groups still generate a full matrix and need to be stored as such. The factorization process can be carried out to an arbitrary level by using a hierarchical grouping structure, the only limitation being the amount of boundary elements; this results in a multilevel, FMM algorithm. References describing the FMM algorithm as a whole for time-harmonic, electromagnetic radiation/scattering problems are [121, 30, 192, 77] (2D, standard FMM), [122] (2D, multilevel FMM), [53, 61, 178, 123] (3D, standard FMM) and [179, 89] (3D, multilevel FMM).

In the literature, many applications of the FMM in microwave engineering are documented. All are fundamentally based on the addition theorems for the Green functions of the 2D and 3D, scalar Helmholz equations in homogeneous media. For example, see [49, 202, 83] and the references therein.

The BI contribution to the cavity FE-BI system matrix is a typical MoM matrix, based on the scalar, 3D, free space, Green function, therefore it is amenable to application of the 3D, Helmholz FMM. The 3D, cavity FE-BI formulation of Chapter 3 has a 2D counterpart. The application of the FMM to the cavity FE-BI formulation, as well as results, are presented in [31, 30, 192] for the 2D case and in [123, 38] for the 3D case.

Alternative fast methods that can also be applied to the BI component of the cavity FE-BI formulation, are the following: The BI can be written as a FFT if the aperture discretization is regular in the correct sense [192, 59]. This technique is generalized by the adaptive integral method (AIM) [192], which maps an arbitrary aperture discretization to a regular grid of points, on which the FFT can again be used. The impedance matrix localization (IML) method is compared with the FMM in [44]. The IML method is based on a sparsifying preconditioner for the MoM matrix and seems not to be as generally applicable as the FMM. [48] gives a general review of fast methods in computational electromagnetics.

An outline of this chapter is as follows: Section 4.2 presents the application of the single level, 3D, FMM to the BI component of the cavity FE-BI formulation presented in Chapter 3. Sources of error introduced by the FMM and techniques to control them are discussed in Section 4.3. The performance of the FMM versus the standard BI formulation is investigated in Section 4.4. Section 4.5 discusses implementation issues. Section 4.6 presents numerical results. In the conclusion, some possible extensions to the FMM as presented here, are mentioned.

4.2 Application of the FMM to the BI component of the cavity FE-BI formulation

The application of the FMM to the BI component of the cavity FE-BI formulation can be viewed as consisting of two parts –

- substitution of the addition theorem (series expansion) of the 3D, free space, scalar Green function (zero-th order, spherical Hankel function) into the double surface integrals of equation (3.23),
- which allows one to separate the source and observation coordinates to the extent that the only full matrix stored is the group interactions, which is accomplished by appropriately grouping the aperture basis functions together.

4.2.1 Series expansion of the scalar, 3D, free space, Green function

This section is based on [53]. From equations (A.4), (A.5) and (A.6) and the identity $e^{j\phi} = \cos\phi + j\sin\phi$, the following addition theorem of the scalar, 3D, free space, Green function can be constructed (see equation (B.19) for the Green function's definition):

$$\frac{e^{-jk_0|\mathbf{X}+\mathbf{d}|}}{4\pi|\mathbf{X}+\mathbf{d}|} = -\frac{jk_0}{4\pi}\sum_{l=0}^{\infty}(-1)^l(2l+1)\,j_l(k_0d)\,h_l^{(2)}(k_0X)\,P_l(\hat{d}\cdot\hat{X}) \quad \{d<X\} \tag{4.1}$$

where $j_l(z)$, $h_l^{(2)}(z)$ and $P_l(z)$ are the spherical Bessel function of the first kind, spherical Hankel function of the second kind and the Legendre polynomial, respectively – all of the lth order, as defined in [1]. The above addition theorem is also provided in [53, 164].

This equation can be rewritten by substitution of the following spectral domain representation of $j_l(k_0d)\,P_l(\hat{d}\cdot\hat{X})$ (see [53, 178, 123]):

$$j_l(k_0d)\,P_l(\hat{d}\cdot\hat{X}) = \frac{1}{4\pi}(j)^l\oint_{S_u} e^{-j\mathbf{k_0}\cdot\mathbf{d}}\,P_l(\hat{k}_0\cdot\hat{X})\,d\hat{k}_0 \tag{4.2}$$

where S_u denotes the surface of a unit sphere, centered at the origin.

The substitution results in

$$\frac{e^{-jk_0|\mathbf{X}+\mathbf{d}|}}{4\pi|\mathbf{X}+\mathbf{d}|} = -\frac{jk_0}{(4\pi)^2}\sum_{l=0}^{\infty}(-j)^l(2l+1)\,h_l^{(2)}(k_0X)\oint_{S_u} e^{-j\mathbf{k_0}\cdot\mathbf{d}}\,P_l(\hat{k}_0\cdot\hat{X})\,d\hat{k}_0 \quad \{d<X\}. \tag{4.3}$$

The reason for this substitution is to separate the factors containing $\mathbf{d}$ and $\mathbf{X}$ in order to decrease the computational effort in evaluating the function when only one of $\mathbf{d}$ or $\mathbf{X}$ is varied.

$|\mathbf{X}+\mathbf{d}|$ cannot simply be replaced with $|\mathbf{r}-\mathbf{r}'|$ (in order to obtain the true scalar, 3D, free space, Green function), because $d < X$ must always apply for the addition theorem to be valid. The next section discusses the solution of this substitution problem by way of introducing a grouping strategy.

4.2.2 Grouping the BI basis functions together

This section is based on [53]. A description similar to this section can be found in [123].

Consider the cavity BI interaction between two basis functions as defined by equation (3.26). Let the basis functions have global numbers a and b, i.e. $\mathbf{N}_a$ and $\mathbf{N}_b$. If the unaccented coordinates refer to integration over the domain of $\mathbf{N}_a$ and the accented coordinates refer to integration over the domain of $\mathbf{N}_b$, then the contribution to the system matrix is as follows:

$$\begin{aligned} P(a,b) &= 2\int_{S_a}\{\nabla\cdot\hat{z}\times\mathbf{N}_a(\mathbf{r})\}\left\{\int_{S_b}\{\nabla'\cdot\hat{z}\times\mathbf{N}_b(\mathbf{r}')\}\frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}dS'\right\}dS \\ &\quad - 2k_0^2\int_{S_a}\{\hat{z}\times\mathbf{N}_a(\mathbf{r})\}\cdot\left\{\int_{S_b}\{\hat{z}\times\mathbf{N}_b(\mathbf{r}')\}\frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}dS'\right\}dS. \end{aligned} \tag{4.4}$$

In order to substitute equation (4.3) into (4.4), $\mathbf{r}-\mathbf{r}'$ must be related to $\mathbf{X}+\mathbf{d}$. The aperture area is divided into smaller non-overlapping areas such that their union is at least equal to the aperture area. All aperture basis functions that are located within a certain smaller area is said to form one group. Every group has a center coordinate $\mathbf{g}$. Define the group index vector that associates every basis function with the group it belongs to:

$$g(i) = (\text{Group that } \mathbf{N}_i \text{ belongs to}); \quad i = 1,..,N_F. \tag{4.5}$$

$\mathbf{r}-\mathbf{r}'$ can now be rewritten as follows:

$$\mathbf{r}-\mathbf{r}' = (\mathbf{r}-\mathbf{g}_{g(a)}) - (\mathbf{r}'-\mathbf{g}_{g(b)}) + \mathbf{X}_{g(a)g(b)} \tag{4.6}$$

where $\mathbf{g}_{g(a)}$ and $\mathbf{g}_{g(b)}$ represent the centers of the groups that $\mathbf{N}_a$ and $\mathbf{N}_b$ belong to; $\mathbf{X}_{g(a)g(b)} = \mathbf{g}_{g(a)} - \mathbf{g}_{g(b)}$.

From the above equation, the following replacements can be made within the addition theorem:

$$\mathbf{d} \leftarrow (\mathbf{r}-\mathbf{g}_{g(a)}) - (\mathbf{r}'-\mathbf{g}_{g(b)}) \tag{4.7}$$

$$\mathbf{X} \leftarrow \mathbf{X}_{g(a)g(b)}. \tag{4.8}$$

Rewrite the scalar, 3D, free space, Green function by making the above substitutions in equation (4.3), yielding

$$\begin{aligned} G_0(\mathbf{r},\mathbf{r}') &\equiv \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \\ &\approx -\frac{jk_0}{(4\pi)^2}\sum_{l=0}^{L}\Big[(-j)^l(2l+1)\,h_l^{(2)}(k_0X_{g(a)g(b)}) \\ &\quad \oint_{S_u} e^{-j\mathbf{k}_0\cdot\{(\mathbf{r}-\mathbf{g}_{g(a)})-(\mathbf{r}'-\mathbf{g}_{g(b)})\}}\,P_l(\hat{k}_0\cdot\hat{X}_{g(a)g(b)})\,d\hat{k}_0\Big] \end{aligned} \tag{4.9}$$

where the approximate equality results from truncating the addition theorem to L_{tot} terms, i.e.

$$L_{\text{tot}} \equiv L+1. \tag{4.10}$$

This truncation is an obvious, practical necessity and in Section 4.3 it is shown how the error of the FMM versus the standard BI can be controlled via an appropriate choice of L_{tot}.

Substituting equation (4.9) into equation (4.4) yields

$$\begin{aligned}
&P(a,b) \approx \\
&-\frac{2jk_0}{(4\pi)^2}\oint_{S_u}\left[\int_{S_a}\{\nabla\cdot\hat{z}\times\mathbf{N}_a(\mathbf{r})\}e^{-j\mathbf{k}_0\cdot(\mathbf{r}-\mathbf{g}_{g(a)})}dS\int_{S_b}\{\nabla'\cdot\hat{z}\times\mathbf{N}_b(\mathbf{r}')\}e^{j\mathbf{k}_0\cdot(\mathbf{r}'-\mathbf{g}_{g(b)})}dS'\right.\\
&\left.\sum_{l=0}^{L}(-j)^l(2l+1)\,h_l^{(2)}(k_0X_{g(a)g(b)})\,P_l(\hat{k}_0\cdot\hat{X}_{g(a)g(b)})\right]d\hat{k}_0\\
&+\frac{2jk_0^3}{(4\pi)^2}\oint_{S_u}\left[\int_{S_a}\{\hat{z}\times\mathbf{N}_a(\mathbf{r})\}e^{-j\mathbf{k}_0\cdot(\mathbf{r}-\mathbf{g}_{g(a)})}dS\cdot\int_{S_b}\{\hat{z}\times\mathbf{N}_b(\mathbf{r}')\}e^{j\mathbf{k}_0\cdot(\mathbf{r}'-\mathbf{g}_{g(b)})}dS'\right.\\
&\left.\sum_{l=0}^{L}(-j)^l(2l+1)\,h_l^{(2)}(k_0X_{g(a)g(b)})\,P_l(\hat{k}_0\cdot\hat{X}_{g(a)g(b)})\right]d\hat{k}_0 \qquad (4.11)
\end{aligned}$$

where the order of integration and summation has been interchanged.

The standard BI interaction of equation (4.4) has been rewritten in terms of separate elemental integrals and a group interaction term, together forming the kernel of the closed surface integral with respect to $\hat{k}_0$. Since this integral (with respect to $\hat{k}_0$) will in practice be carried out by some quadrature scheme, it is now replaced with a quadrature rule summation. $w_{k_0}^m$ and $\mathbf{k}_0^m$ represent the mth weight and kernel evaluation value of the quadrature scheme, with $m = 1,..,M$. Thus equation (4.11) can be rewritten as

$$\begin{aligned}
P(a,b) \approx -\frac{jk_0}{(4\pi)^2}\sum_{m=1}^{M}w_{k_0}^m\Big[&2\,V_1^m(a,g(a))\,T^m(g(a),g(b))\,V_1^m(b,g(b))^*\\
&-2k_0^2\,\mathbf{V}_2^m(a,g(a))\cdot T^m(g(a),g(b))\,\mathbf{V}_2^m(b,g(b))^*\Big] \qquad (4.12)
\end{aligned}$$

with

$$V_1^m(p,q) = \begin{cases}\displaystyle\int_{S_p}\{\nabla\cdot\hat{z}\times\mathbf{N}_p(\mathbf{r})\}e^{-j\mathbf{k}_0^m\cdot(\mathbf{r}-\mathbf{g}_q)}dS & \text{if } q = g(p)\\ 0 & \text{otherwise}\end{cases} \qquad (4.13)$$

$$\mathbf{V}_2^m(p,q) = \begin{cases}\displaystyle\int_{S_p}\{\hat{z}\times\mathbf{N}_p(\mathbf{r})\}e^{-j\mathbf{k}_0^m\cdot(\mathbf{r}-\mathbf{g}_q)}dS & \text{if } q = g(p)\\ 0 & \text{otherwise}\end{cases} \qquad (4.14)$$

$$T^m(p,q) = \begin{cases}\displaystyle\sum_{l=0}^{L}(-j)^l(2l+1)\,h_l^{(2)}(k_0X_{pq})\,P_l(\hat{k}_0^m\cdot\hat{X}_{pq}) & \text{if groups } p \text{ and } q \text{ are not near neighbours}\\ 0 & \text{otherwise}\end{cases} \qquad (4.15)$$

where p and q are arbitrary indices. As a consequence of the constraint on equation (4.3), the interactions of basis functions belonging to neighbouring groups, or the same group, cannot be calculated with equation (4.12), thus the group interaction term in equation (4.15) is set to zero in such cases. Such groups are called near neighbours. Section 4.2.4 discusses the concept of near neighbours in more detail, but the above definition will suffice here.

Careful consideration of equation (4.12) reveals that the whole BI contribution to the cavity FE-BI system matrix can be factorized. For every cavity aperture basis function the two

surface integral factors must be calculated and for every far interacting pair of groups the group interaction term must be calculated; and all three of these data sets have to be calculated for every $\mathbf{k}_0^m$, $m = 1, .., M$.

Since the interaction of basis function belonging to groups that are near neighbours must be calculated normally, using equation (3.26), the following quantity is needed to write out the factorization:

$$P'(p,q) = \begin{cases} P(p,q) & \text{if groups } g(p) \text{ and } g(q) \text{ are near neighbours} \\ 0 & \text{otherwise} \end{cases} . \tag{4.16}$$

The factorization of the BI matrix now follows as

$$[P] = [P'] - \frac{jk_0}{(4\pi)^2} \sum_{m=1}^{M} w_{k_0}^m \Big[2\,[V_1^m][T^m][V_1^m]^\dagger - 2k_0^2\,[\mathbf{V}_2^m] \cdot [T^m][\mathbf{V}_2^m]^\dagger \Big] \tag{4.17}$$

where $[\cdot\cdot]^\dagger$ represents the complex conjugate transpose of the matrix $[\cdot\cdot]$. The entries of $[P']$, $[V_1^m]$, $[\mathbf{V}_2^m]$ and $[T^m]$ are defined by equations (4.16), (4.13), (4.14) and (4.15). The dimensions of these matrices are presented in Table 4.1. Note that $[P']$ is sparse, since it only contains contributions of near interacting basis functions. Also note that $[V_1^m]$ and $[\mathbf{V}_2^m]$ only have one nonzero element per row, therefore they are sparse too.

Table 4.1: Dimensions and storage requirements of the FMM factorization terms. N_F represents the number BI degrees of freedom, and N_g represents the number of groups.

Matrix Name	Dimensions	Storage
$[P']$	$N_F \times N_F$	sparse
$[V_1^m]$	$N_F \times N_g$	sparse
$[\mathbf{V}_2^m]$	$N_F \times N_g$	sparse
$[T^m]$	$N_g \times N_g$	full

4.2.3 The grouping scheme

The cavity aperture is divided into squares, all with the same dimensions: $D \times D$. A group consists of all the basis functions located within a certain square. The group center of the ith group, $\mathbf{g}_i$, is defined as the geometric center of the ith group's corresponding square. The groups are numbered from left to right (increasing x coordinate of the centers) and bottom to top (increasing y coordinate of the centers). Figure 4.1 shows this scheme.

All basis functions with components tangential to the cavity BI surface are either edge-based or face based (see Appendix C). A face is said to be located within a certain square if its center point (average values of the coordinates of its vertices) is located within that square. An edge is said to be located within a certain square if its center point is located within that square (average values of the coordinates of its vertices). A point in turn, is said to be located within a square if the following inequalities are satisfied:

$$\begin{aligned} 0 \leq (x - x_{ll}) < D \\ 0 \leq (y - y_{ll}) < D \end{aligned} \tag{4.18}$$

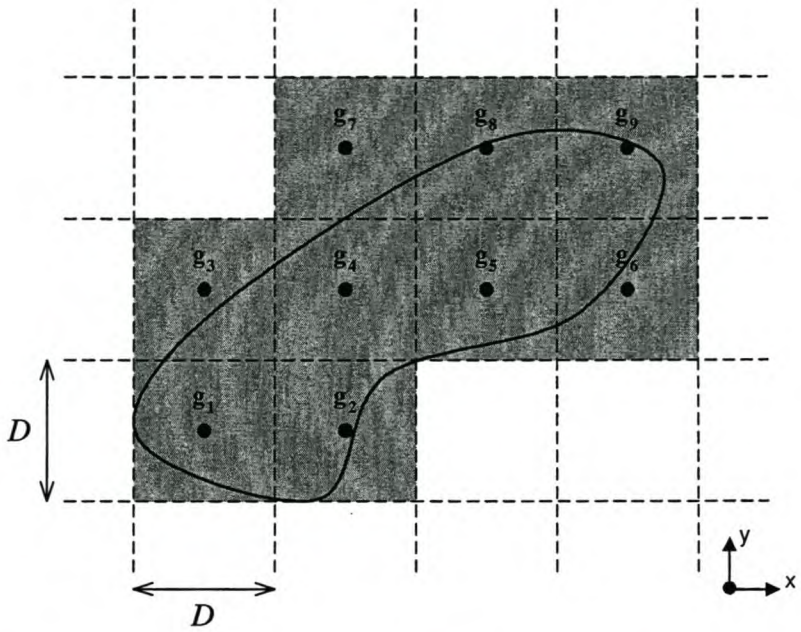


Figure 4.1: FMM grouping scheme for an arbitrary aperture.

where (x, y) are the coordinates of the point and (x_{ll}, y_{ll}) are the coordinates of the lower left corner of the square.

This grouping scheme is presented in [39] and is similar to one described in [179] with regard to the edge basis functions.

Define the maximum group dimension as

$$D_{\max} \equiv \sqrt{2}D. \tag{4.19}$$

Note that the group dimension influences the accuracy and efficiency of the FFM (see Sections 4.3 and 4.4).

4.2.4 The far interaction criterion

The addition theorem for the scalar, 3D, free space, Green function is only valid if $d < X$ (see equation (4.1)). From the grouping scheme it is apparent that $d_{\max} = D_{\max}$, therefore

$$X > D_{\max} \tag{4.20}$$

must at least apply, otherwise the interaction is near and must be calculated conventionally.

The specification of a minimum inter-group distance is called a far interaction criterion. The last equation is called a far interaction criterion and represents the theoretical limit.

This limit is not a practical far interaction criterion, because the number of spherical quadrature points is of $\mathcal{O}(L_{\text{tot}}^2)$ and L_{tot} increases as the far interaction criterion decreases (see Section 4.3.1). Since the FMM's computational efficiency is clearly of $\mathcal{O}(M)$, it follows that

the efficiency will decrease as the limit is approached. In practice, larger minimum values are specified:

$$X \geq X_{\min} \quad \text{with} \quad \{X_{\min} > D_{\max}\}. \tag{4.21}$$

Section 4.3 investigates the effect of the far interaction criterion on the accuracy of the FMM.

Finally, note that the criterion for determining whether an interaction between two basis functions is near or far is based on the absolute distance between the centers of the groups they belong to, therefore all interactions between basis functions from two specific groups fall into the same category (either near or far).

4.3 Errors introduced by the FMM

The FMM intrudes two additional sources of error, relative to the standard BI formulation. This section discusses these errors and how to control them.

4.3.1 Addition theorem truncation error

The choice of L_{tot}, the number of terms at which the addition theorem is truncated, is crucial to the level of accuracy with which the FMM approximates the free space Green function. Also, it is important to use the minimum number of orders that will yield the desired accuracy, otherwise computational efficiency will be compromised as discussed in Section 4.4.

Define the relative Green function truncation error (as in Appendix F) as

$$\varepsilon^L = \left| \frac{G_0^L - G_0}{G_0} \right|, \tag{4.22}$$

with G_0^L signifying the series in equation (4.1), truncated at L_{tot} terms and G_0 signifying the true Green function value.

The relative error reaches a maximum value when X is at a minimum and d is at a maximum [62, 8, 113] and when $\hat{X}$, $\hat{d}$ are collinear, i.e. $\hat{d} \cdot \hat{X} = 1$ [113, 176, 177]. These observations have also been confirmed experimentally, see for example Figure F.1, which shows plots of ε^L versus L (equation (4.22)) for various values $-1 \leq \hat{d} \cdot \hat{X} \leq 1$ with constant X and d. The relative error is not a function of λ_0, since all geometric quantities in the addition theorem are multiplied by $k_0 = \frac{2\pi}{\lambda_0}$ (i.e. normalized with respect to the wavelength).

In the literature, the truncation error has been studied from a theoretical point of view, producing bounds on the error and/or the number of terms needed [62, 113, 8, 176, 61]. It has also been studied from a more practical/experimental point of view [69, 38, 177, 118]. Efforts has been made to establish a formula to (approximately) calculate the number of terms needed (L_{tot}), given a specified relative error (ε^L) [38, 176, 53, 177]. Such a formula can be used within an FMM code (especially in the multilevel case, when different groups sizes and minimum inter-group distances are used at every level) or when studying the effect of the FMM parameters on its computational efficiency.

In [62, 113, 8], bounds on the absolute, unnormalised error are derived under various conditions. [62] shows that

$$L_{\text{tot}} = C_1 + C_2 D_{\max} + C_2' \log D_{\max} + C_3 \log\left(\left|G_0 \varepsilon^L\right|^{-1}\right), \tag{4.23}$$

but without providing explicit values for the constants. [8] provides an explicit, computable bound on the error, but notes that the bound should be used in an iterative manner to obtain the minimum L_{tot} from a specified ε^L — which means that one can just as well use equation (4.22) directly in the same manner.

In general, an approximate formula for the value of L_{tot} should be a function of the following variables:

- $d = D_{\max}$. The maximum value that d reaches, therefore the maximum group dimension.
- $X = X_{\min}$. The minimum value that X reaches, as specified by the far interaction criterion.
- ε^L. The desired maximum relative error resulting from the truncation.

[53] presents the following, widely used (see [69, 123] for example) formula:

$$\begin{gathered} L_{\text{tot}} = 1 + k_0 D_{\max} + 5 \ln(k_0 D_{\max} + \pi) \\ \text{for} \\ \varepsilon^L = 10^{-6}; \quad X_{\min} = D_{\max} + \tfrac{5}{k_0} \ln(k_0 D_{\max} + \pi) \end{gathered} . \tag{4.24}$$

[177] extends the above formula to different error levels and states that their formula is valid for asymptotically large inter-group distances:

$$\begin{array}{lllll} L_{\text{tot}} = 1 + k_0 D_{\max} + \ln(k_0 D_{\max} + \pi) & \text{for} & \varepsilon^L = 10^{-1}; & X_{\min} \to \infty \\ L_{\text{tot}} = 1 + k_0 D_{\max} + 3 \ln(k_0 D_{\max} + \pi) & \text{for} & \varepsilon^L = 10^{-3}; & X_{\min} \to \infty \\ L_{\text{tot}} = 1 + k_0 D_{\max} + 5 \ln(k_0 D_{\max} + \pi) & \text{for} & \varepsilon^L = 10^{-6}; & X_{\min} \to \infty \end{array} . \tag{4.25}$$

[176] derives the following formula by considering the most significant, leading term in the error series:

$$\begin{gathered} L_{\text{tot}} = 1 + k_0 D_{\max} + 1.8 \left[\log\left(\tfrac{1}{\varepsilon^L}\right)\right]^{\frac{2}{3}} (k_0 D_{\max})^{\frac{1}{3}} \\ \text{for} \\ 10^{-10} \leq \varepsilon^L \leq 10^{-1}; \quad X_{\min} \to \infty \end{gathered} . \tag{4.26}$$

All the above formulas have some restrictions and none incorporate all three parameters explicitly and generally. In Section F.2 (based on [39]) a completely empirical approach to is used to derive the following, general formula:

$$L_{\text{tot}}(D_{\max}) = k e^{-\alpha D_{\max}} + m D_{\max} + c \tag{4.27}$$

with

$$\begin{aligned}
\varepsilon_{10}^{L} &= \log_{10}\varepsilon^{L} \\
x' &= X_{\min}/D_{\max} \\
m &= 7.826 \\
c &= -1.92\varepsilon_{10}^{L} - 0.138 \\
\alpha &= [1.85\arctan(2.8(x'-1.74)) + 2.455]\cdot 0.8^{-\varepsilon_{10}^{L}} \\
k &= \left[37.2\, e^{-2.42\sqrt{x'-1}} - 1.4\right]\cdot 1.3^{-\varepsilon_{10}^{L}}
\end{aligned}$$

and tested for

$$10^{-6} \leq \varepsilon^{L} \leq 10^{-2}; \quad D_{\max} < X_{\min} < \infty; \quad 0 \leq D_{\max} \leq 4.$$

Finally, [69] presents FMM 'plans', derived via numerical experimentation. Each plan contains three constants of the form $(X_{\min}, D_{\max}, L_{\text{tot}})$. Associated with every plan is a specified average error level that it should yield. No formula is provided.

Figure 4.2 compares the various approximations of L_{tot} for the far interaction criterion of equation (4.24), i.e. $X_{\min} = D_{\max} + \frac{5}{k_0}\ln(k_0 D_{\max} + \pi)$. In the lower range of $D_{\max}$, all the approximations fare nearly equally well, but in the higher range, equation (4.26) is clearly superior.

Figure 4.3 compares the various approximations of L_{tot} for the far interaction criterion $X_{\min} = 1.4D_{\max} + \frac{0.2\lambda^2}{D_{\max}}$. In the lower range of $D_{\max}$, equation (4.27) is far superior, but in the higher range equation (4.26) again results in the best approximation.

The far interaction criterion used in Figure 4.2 can be considered as 'close' to $X_{\min} \to \infty$, resulting in the reasonable performance of all the approximations. On the other hand, the far interaction criterion used in Figure 4.3 is quite close to $D_{\max}$ and has a strong influence on L_{tot} in the lower range, therefore it is important that the approximation formula takes finite $X_{\min}$ into account in the lower range, hence the clear superior performance of equation (4.27), since it is the only one that incorporates the effect of finite $X_{\min}$.

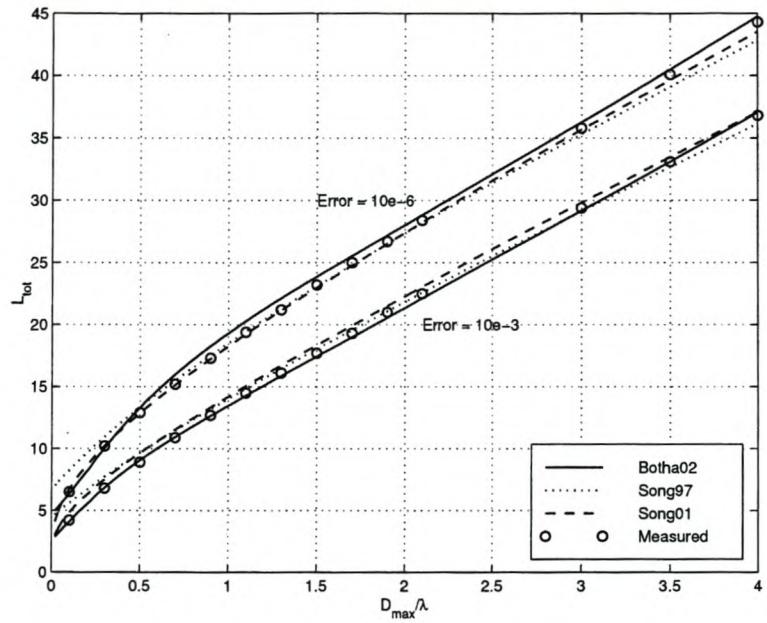
In light of these results together with other experimental experience, the following, improved scheme is proposed:

$$L_{\text{tot}} = \begin{cases} \text{RHS of equation (4.27)} & \text{if } 0 < \dfrac{D_{\max}}{\lambda} < 5 \\ \text{RHS of equation (4.26)} & \text{if } 5 \leq \dfrac{D_{\max}}{\lambda} < \infty \end{cases} . \tag{4.28}$$

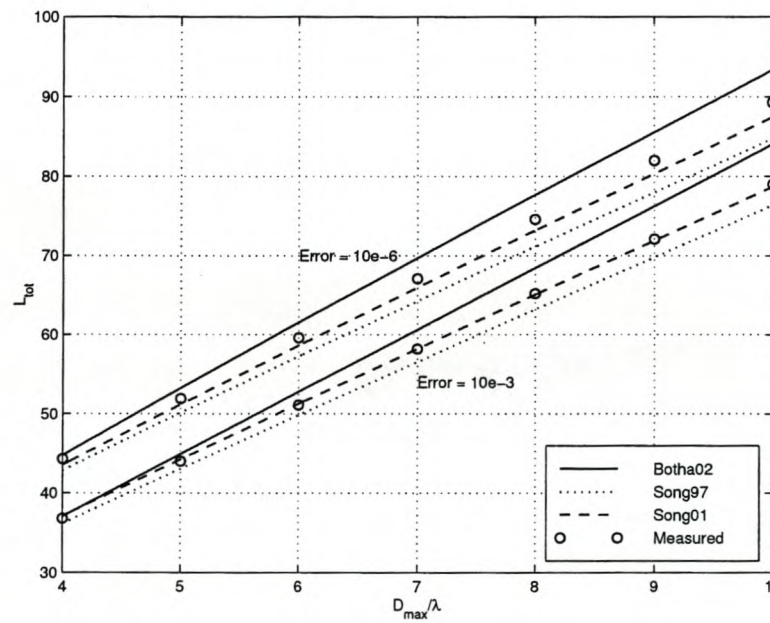
Finally, as stated in [113, 53, 69, 62], if the order of the spherical Hankel function in the addition theorem exceeds its argument, it oscillates wildly, causing inaccurate results with finite computer precision. Therefore, any choice of L_{tot} must also satisfy the following condition:

$$L_{\text{tot}} \leq k_0 X_{\min} + 1. \tag{4.29}$$

This implies that not all combinations $(D_{\max}, X_{\min}, \varepsilon)$ are valid.

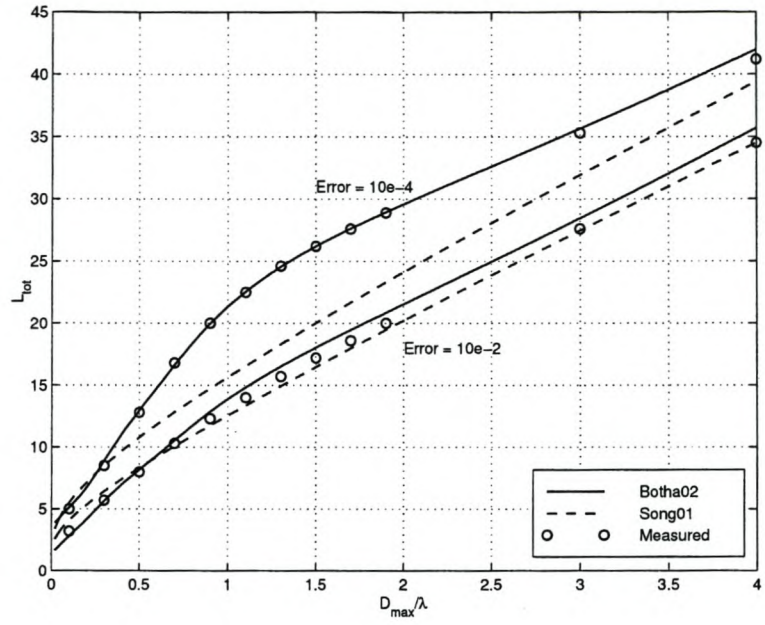


(a) $0 < \frac{D_{\max}}{\lambda} \leq 4$

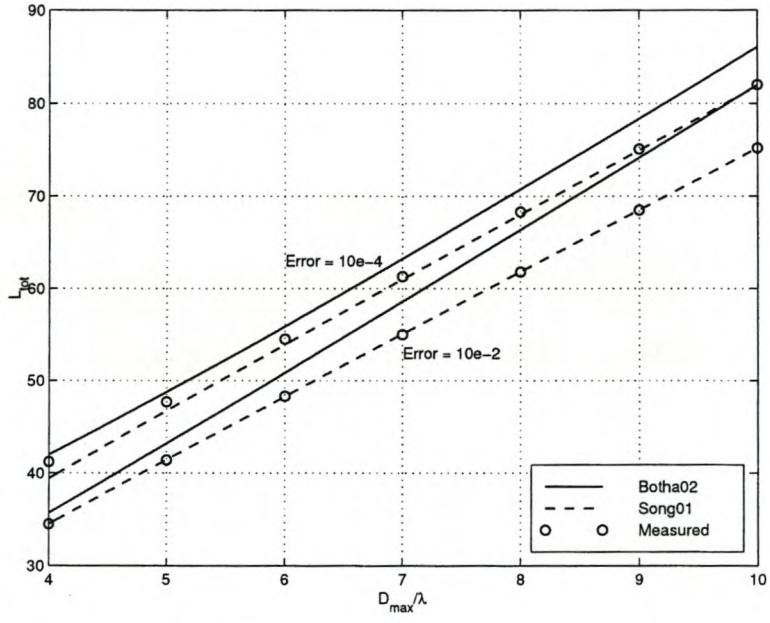


(b) $4 \leq \frac{D_{\max}}{\lambda} \leq 10$

Figure 4.2: Measured number of FMM addition theorem orders vs. normalized group size for error levels $\varepsilon^L = 10^{-3}$ and $\varepsilon^L = 10^{-6}$. Comparison between the analytic approximations presented in Section 4.3.1. $X_{\min} = D_{\max} + \frac{5}{k_0}\ln(k_0 D_{\max} + \pi)$.



(a) $0 < \frac{D_{\max}}{\lambda} \leq 4$



(b) $4 \leq \frac{D_{\max}}{\lambda} \leq 10$

Figure 4.3: Measured number of FMM addition theorem orders vs. normalized group size for error levels $\varepsilon^L = 10^{-2}$ and $\varepsilon^L = 10^{-4}$. Comparison between the analytic approximations presented in Section 4.3.1. $X_{\min} = 1.4D_{\max} + \frac{0.2\lambda^2}{D_{\max}}$.

4.3.2 Spherical surface quadrature error

Replacing the spherical surface integral in equation (4.9) with a quadrature rule is a critical step in the FMM algorithm. This quadrature rule is another source of error introduced by the FMM in relation to the standard BI formulation. Various references comment on the choice of this rule and come to more or less the same conclusion. This section reviews these references and notes the scheme that was adopted.

The kernel of the FMM, spherical surface integral (see equation (4.12)) has the following properties, relevant to the choice of an appropriate quadrature scheme:

- The two integration variables are the spherical coordinates θ and ϕ.
- A symmetry plane of the kernel exists at the $\theta = \frac{\pi}{2}$ plane (i.e. $z = 0$).
- $T(p, q)$ has rotational symmetry around the $\hat{X}_{pq}$-axis (see equation (4.15)).
- $\mathbf{X}$ always lies in the $\theta = \frac{\pi}{2}$ plane (see equation (4.15)).
- The kernel is smoothly periodic in ϕ at any constant θ.

There are two main types of spherical surface quadrature schemes: product rules and rules specifically tailored for this surface. The second class (see for example [128, 182]) is discarded for the following reasons: the available rules are spaced too far apart in terms of accuracy (i.e. one is likely to use a rule of excessive accuracy), a rule of arbitrary order cannot be constructed in a automated manner and such rules of arbitrary order cannot necessarily exploit the $\theta = \frac{\pi}{2}$ symmetry plane of the kernel.

In the literature, a quadrature rule of product type is always suggested [62, 53, 113, 89, 118]. A product rule for a two dimensional integral can be defined as follows: If $Q_1\ (x_1^i, w_1^i, i = 1, .., N_1)$ and $Q_2\ (x_2^j, w_2^j, j = 1, .., N_2)$ are quadrature rules for integrals of one variable over the intervals X_1 and X_2 respectively, then the integral

$$\int_{X_2} \int_{X_1} f(x_1, x_2)\, dx_1 dx_2 \tag{4.30}$$

can be approximated with the product rule $Q_3\ (\mathbf{x}_3^k, w_3^k, k = 1, .., N_3)$, with

$$k = (i-1)N_2 + j \quad \forall \begin{cases} i = 1, .., N_1 \\ j = 1, .., N_2 \end{cases} \tag{4.31}$$

$$\begin{aligned} \mathbf{x}_3^k &= (x_1^i, x_2^j) \\ w_3^k &= w_1^i w_2^j \end{aligned} . \tag{4.32}$$

The previously mentioned references all suggest the following combination:

- N_θ point, cosinusoidally transformed, Gauss-Legendre rule on $\theta\ [0, \pi]$,
- N_ϕ point, trapezoidal rule on $\phi\ [0, 2\pi]$.

By considering combinations of various 1D rules applied to equation (4.3), it was verified that this combination indeed produces the least points for a given accuracy. (Rules considered: Gauss-Legendre and Gauss-Chebyshev rules linearly or cosinusoidally transformed to an angular coordinate domain [68] and a trapezoidal rule.) This integration scheme can take advantage of the $\theta = \frac{\pi}{2}$ kernel symmetry plane, because the Gauss-Legendre rule is symmetric with respect to $\theta = \frac{\pi}{2}$, thus only the points in the range $[0, \frac{\pi}{2}]$ have to be evaluated and the weights of those in the range $[0, \frac{\pi}{2})$ have to be multiplied by 2. Note that the rotational symmetry of every $T(p,q)$ with respect to its $\hat{X}_{pq}$ axis cannot be exploited, because the same set of quadrature points has to be used for all $T(p,q)$. The adopted scheme does not favour any particular $\hat{X}_{pq}$ direction, since all $\hat{X}_{pq}$ are in the $\theta = \frac{\pi}{2}$ plane.

It can be proved [67] that the trapezoidal rule T ($w^0 = w^N = \frac{\pi}{N}, w^k = \frac{2\pi}{N}, \phi^k = \frac{2\pi k}{N}, k = 0, .., N$) is exact for the $2N$ smoothly periodic functions 1, $\sin\phi$, $\cos\phi$,.., $\sin(N-1)\phi$, $\cos(N-1)\phi$, $\sin N\phi$. The kernel is periodic in ϕ, therefore it is not surprising that this rule is so successful.

It remains to find the numbers of quadrature points to be used in the θ and ϕ directions. [62] shows that equation (4.23), together with perfect integration of the first

$$L_{\text{int}} = D_1 + D_2 D_{\max} + D_2' \log D_{\max} + D_3 \log\left(|G_0\varepsilon|^{-1}\right) \tag{4.33}$$

spherical harmonics ($P_l^m(\cos\theta)e^{jm\phi}$ $\{-l \leq m \leq l\}$, see [164, 113, 1]) will ensure an absolute Green function error $\leq G_0\varepsilon$. This implies that $N_\theta = \lceil 0.5 L_{\text{int}} \rceil$ and $N_\phi = L_{\text{int}}$ is needed. From equations (4.23) and (4.33) it follows that L_{int} should be a linear function of L_{tot}. [53, 62, 113, 89] suggest $L_{\text{int}} = 2L_{\text{tot}}$. [118] experiments with increasing L_{int} beyond $2L_{\text{tot}}$, showing improvement in accuracy in some cases, though the truncation error obviously remains present. [113] states that the truncation error will dominate the total error when $L_{\text{int}} = 2L_{\text{tot}}$ is used, which is confirmed in Section F.3. Therefore, the following scheme is adopted:

$$N_\theta = L_{\text{tot}} \tag{4.34}$$

$$N_\phi = 2L_{\text{tot}} \tag{4.35}$$

$$M = 2L_{\text{tot}}\lceil 0.5 L_{\text{tot}} \rceil. \tag{4.36}$$

(The 0.5 factor results from exploiting the kernel's symmetry with respect to $\theta = \frac{\pi}{2}$.)

4.4 Performance analysis of the FMM versus the standard BI

The parameters used in this section are defined in Table 4.2. The algorithm complexities are considered first, followed by discussions on memory requirements and the optimal choice of the number of groups.

4.4.1 Complexity analysis of the FMM

Complexity analysis is an assessment of the amount of mathematical operations ($+$, $-$, $\times$ or $\div$) involved in calculation of a certain entity. This is carried out for the system matrix setup and a single matrix-vector product.

Table 4.2: Parameters used in the performance analysis of the FMM algorithm vs. the standard BI formulation.

N	Number of degrees of freedom in the aperture
G	Number of FMM groups
M	Number of spherical quadrature points
L	Number of addition theorem terms used
a	Number of operations to integrate $\hat{z} \times \mathbf{N}$ on one aperture facet ($\mathbf{N}$ represents a specific aperture basis function)
b	Number of operations to calculate one element of $[T^m]$
n	Average number of near neighbours of a group
d	Group dimension, normalized with respect to λ
u	Number of degrees of freedom per λ^2 area

The operation counts of the standard BI formulation are as follows:

$$\begin{array}{lll} [P] \text{ setup} & \simeq a^2N^2 & \text{operations} \\ [P]\{E\} \text{ product} & \simeq 2N^2 & \text{operations} \end{array}$$

where the factor 2 results from 1 addition and 1 multiplication per matrix element.

The operation counts of the FMM matrices are as follows:

$$\begin{array}{lll} [P'] \text{ setup} & \simeq \frac{na^2N^2}{G} & \text{operations} \\ [V_1^m] \text{ and } [\mathbf{V}_2^m] \text{ setup } \forall\, m & \simeq aNM & \text{operations} \\ [T^m] \text{ setup } \forall\, m & \simeq bG^2M & \text{operations} \\ [P']\{E\} \text{ product} & \simeq \frac{2nN^2}{G} & \text{operations} \\ [V_1^m]^\dagger\{E\} \text{ and } [\mathbf{V}_2^m]^\dagger\{E\} \text{ products } \forall\, m & \simeq 4 \cdot 2NM & \text{operations} \\ [T_m]\{\cdot\} \text{ products } \forall\, m & \simeq 4 \cdot 2G^2M & \text{operations} \\ [V_1^m]\{\cdot\} \text{ and } [\mathbf{V}_2^m]\{\cdot\} \text{ products } \forall\, m & \simeq 4NM & \text{operations} \end{array}$$

where the factor 4 results from $[V_1^m]$ having scalar valued entries and $[\mathbf{V}_2^m]$ having vector valued entries.

Before adding these values to yield the total setup- and matrix-vector operation counts, the dependence of M on the other parameters must be established. From Section 4.3.2 it follows that $M \approx L^2$. From Section 4.3.1 it follows that $L \approx 8\sqrt{2}d$. The combination of these two relations yields $M \approx 128d^2$. Further note that d^2, the group area in square wavelengths, can be written as $d^2 \approx \frac{N}{uG}$. This finally leads to

$$M \approx \frac{128N}{uG}. \tag{4.37}$$

The totals now follow as

$$\begin{array}{lll} \text{Setup} & \simeq \frac{na^2N^2}{G} + \frac{128aN^2}{uG} + \frac{128bGN}{u} & \text{operations} \\ \text{Matrix-vector product} & \simeq \frac{2nN^2}{G} + \frac{1536N^2}{uG} + \frac{1024GN}{u} & \text{operations} \end{array}$$

These approximations are not comprehensive, but they are sufficient to indicate an appropriate choice of G and they allow the conclusion that given a large enough N, the FMM will always result in less operations for the matrix setup and the matrix-vector product.

4.4.2 Computer memory requirements

The standard BI- and FMM matrices have the following computer memory requirements:

$$\begin{array}{lll} [P] & = 8N^2 & \text{bytes} \\ [P'] & = \frac{8nN^2}{G} & \text{bytes} \\ [V_1^m]\ \forall\, m & = 8NM & \text{bytes} \\ [\mathbf{V}_2^m]\ \forall\, m & = 24NM & \text{bytes} \\ [T^m]\ \forall\, m & = 8G^2M & \text{bytes} \end{array}$$

leading to a total FMM memory requirement of

$$\text{FMM total} = \frac{8nN^2}{G} + \frac{4096N^2}{uG} + \frac{1024GN}{u} \quad \text{bytes}$$

compared to the total standard BI memory requirement of $8N^2$ bytes. Single precision storage is assumed, that is 4 bytes per real number and 8 bytes per complex number (all matrices are complex valued).

From these results the same conclusion can be drawn as before, which is that the FMM will always use less computer memory, given a large enough value of N.

4.4.3 Choosing the number of groups

G can be chosen to minimize the matrix setup operation count, the matrix-vector product operation count or the memory requirement of the program. These three values of G are calculated by differentiating the relevant quantities with respect to G and setting the results equal to 0. This results in the following equations:

$$\text{Setup operation count:} \quad G = \sqrt{\frac{una^2 + 128a}{128b}} \cdot \sqrt{N} \tag{4.38}$$

$$\text{Matrix-Vector product operation count:} \quad G = \sqrt{\frac{un + 768}{512}} \cdot \sqrt{N} \tag{4.39}$$

$$\text{Memory:} \quad G = \sqrt{\frac{un + 512}{128}} \cdot \sqrt{N}. \tag{4.40}$$

All three equations indicate that the optimal choice is $G = C\sqrt{N}$ (where C is some positive constant). Substituting this choice back into the equations defining the three quantities, it follows that the FMM reduces all three to $\mathcal{O}\left(N_F^{1.5}\right)$, relative to $\mathcal{O}\left(N_F^2\right)$ in the standard BI case.

To approximate C, note that it is difficult to estimate the values of a and b, but that n and u can be estimated accurately. The fact that the constant is difficult to estimate in the

case of minimized setup operations is not so important, because in practice the other two specifications will always take preference when considering large problems.

u is dependent on the discretization size and on the spectral order of the elements. n is dependent on $X_{\min}$. To give an idea of the value of C for minimized matrix-vector product operations, consider the following two examples on a regular grid:

$$\begin{array}{l} \text{CT/LN elements; disc.} = \frac{\lambda}{10};\ n = 10 \ \Rightarrow\ C = 2.7 \\ \text{LT/QN elements; disc.} = \frac{\lambda}{20};\ n = 10 \ \Rightarrow\ C = 8.9 \end{array} . \tag{4.41}$$

Note that minimized memory needs about twice as many groups as minimized matrix-vector product operations. Clearly, $G \sim \sqrt{N}$ should result in near optimal results.

Implementing a choice of G within the code is not trivial, since the group dimension, D, on which the whole grouping scheme is based, is not related to G for general aperture geometries. A possible solution is to choose a starting value of D and then to calculate the number of groups, a simple estimate of the correct D can then be made, based on this result, the first choice of D and the desired value of G. Proceeding in an iterative manner one can continue until a desired error criterion is met.

4.5 Implementation issues

The FMM matrix storage and setup are based on knowledge of the groups that the cavity aperture degrees of freedom belong to, therefore a data structure containing this information is essential. All sparse FMM matrices are stored using the CRS scheme. Indices for $[V_1^m]$ and $[\mathbf{V}_2^m]$ in both normal and transpose form are established for quick multiplication in either case.

The elements of the various FMM matrices are calculated as follows:

- $[P']$ is calculated using equation (4.4) as in the standard FE-BI formulation.
- $[V_1^m]$ and $[\mathbf{V}_2^m]$ are calculated using the triangular surface quadrature rules presented in Section D.1.1.
- $[T^m]$ involves the evaluation of spherical Hankel functions and Legendre polynomials of orders $\leq L$. The spherical Hankel functions are evaluated using the routine `SPHBES` provided in [149]. If the argument becomes to large for accurate evaluation with this routine, the following asymptotic approximation for $h_l^{(2)}(x)$ is used [1]:

$$h_l^{(2)}(x) \approx \frac{1}{x} e^{-j(x-\frac{\pi}{2}l-\frac{\pi}{4})} \qquad \{x \gg l\}. \tag{4.42}$$

  The Legendre polynomials are evaluated using the following recursive formula [13]:

$$P_l(x) = \frac{(2l-1)xP_{(l-1)}(x) - (l-1)P_{(l-2)}(x)}{l} \qquad \begin{cases} P_0(x) = 1 \\ P_1(x) = x \end{cases} . \tag{4.43}$$

For the iterative solution, the FMM matrix vector product is evaluated by starting the multiplications on the side of the vector of degrees of freedom, in order to benefit from the factorization [53] (see also Section 4.4.1).

4.6 Results

The section presents results to show the computational improvements afforded by the FMM, relative to the standard BI.

Experimental evaluation of the error control scheme of Appendix F

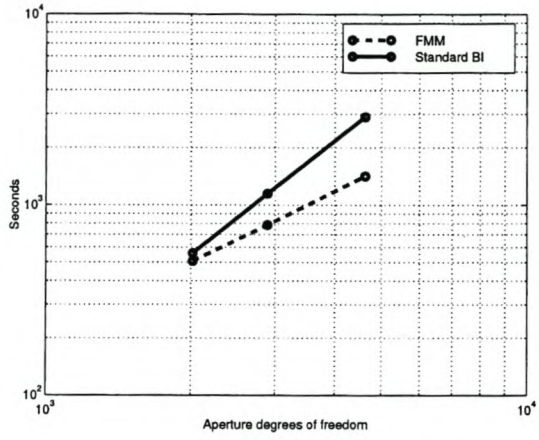
Table 4.3 shows resulting memory requirements and relative matrix element errors (measured as described in Section F.4) for a specific problem. The numbers of addition theorem terms and the numbers of spherical quadrature points were chosen as defined in equations (4.27) and (4.36). Note that the same error specification can lead to different memory requirements, depending on the other two FMM parameters ($D_{\max}, X_{\min}$). Also note that the FMM breaks down in the $\epsilon = 10^{-5}$, $X_{\min} = 2.5D_{\max}$ case as a consequence of violating the stability condition of equation (4.29). Clearly the error control techniques are very effective in this case.

Experimental complexity results

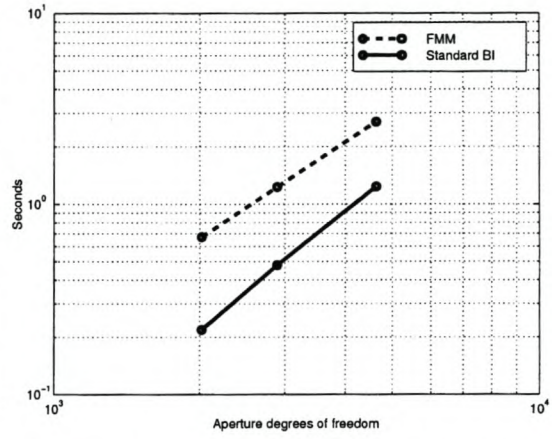
Figure 4.4 compares the FMM and the standard BI with respect to the three performance criteria discussed in Section 4.4, as functions of aperture degrees of freedom. In all three cases, the reduced order of the FMM algorithm can be observed, confirming the theoretical results of Section 4.4.3.

Number of group versus FMM performance

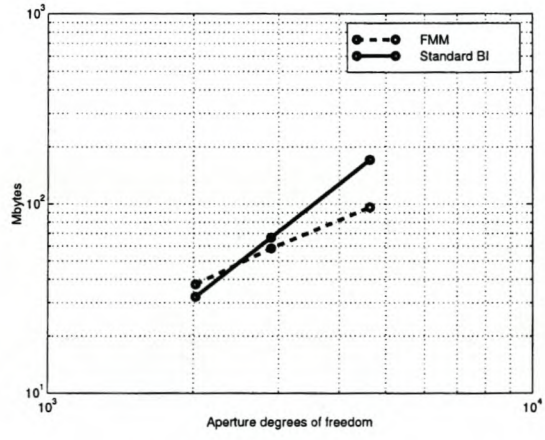
Figure 4.5 shows the effect that the number of FMM groups has on the different performance criteria discussed in Section 4.4, for a specific problem. The theoretical results of Section 4.4.3 is confirmed, where it was shown that every criterion is satisfied best at some specific number of groups for a given cavity aperture geometry and discretization and that this value is $\mathcal{O}\left(\sqrt{N_F}\right)$.



(a) System matrix setup time versus aperture degrees of freedom.



(b) System matrix-vector product time versus aperture degrees of freedom.

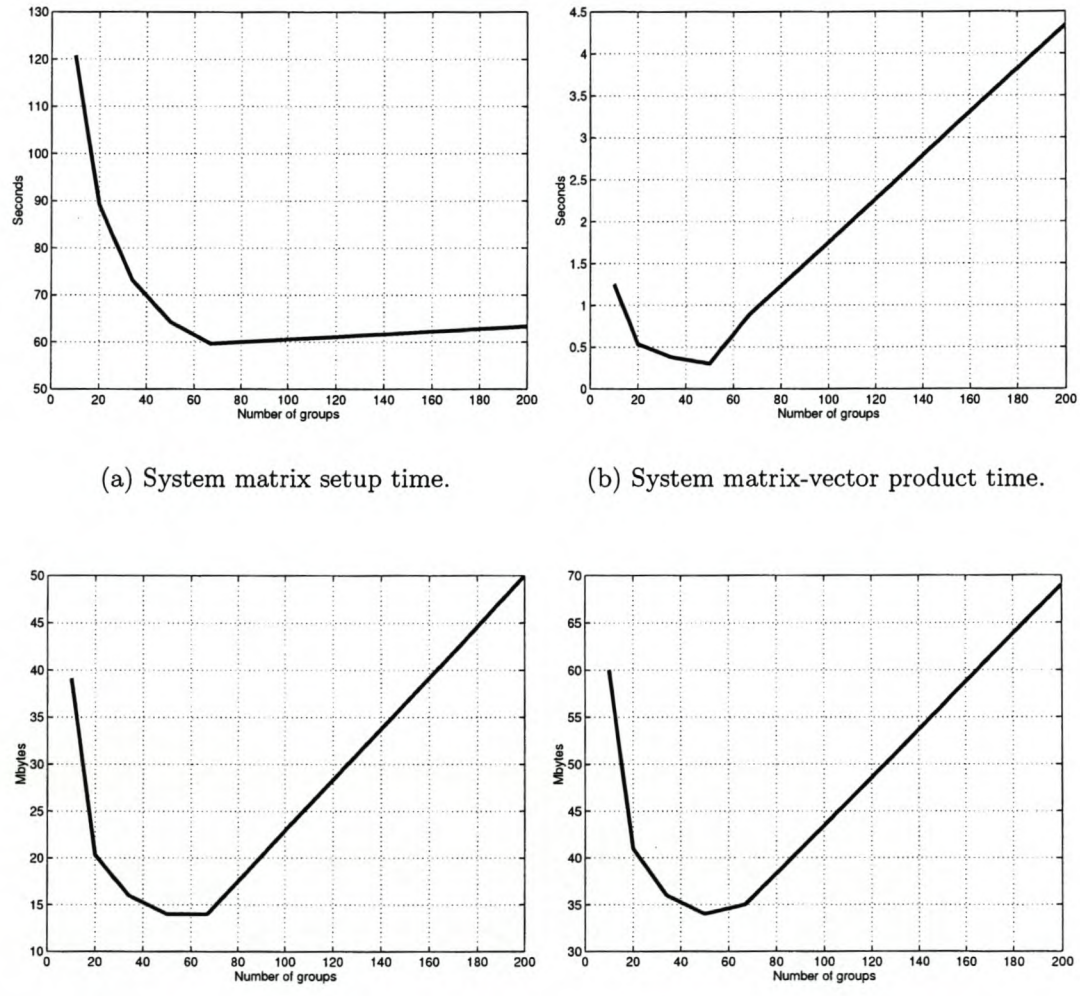


(c) BI matrix storage requirement versus aperture degrees of freedom.

Figure 4.4: Comparison between the computational performance of the FMM and the standard BI.

Table 4.3: Required FMM memory and relative matrix element error obtained with various far-interaction criteria and a constant group dimension of $D = 0.3\lambda$. This data was generated with an empty cavity with dimensions $x \times y \times z = 10\lambda \times 0.1\lambda \times 0.1\lambda$ and rotated by 45° degrees. $\frac{\lambda}{10}$ discretization with CT/LN elements. (The full BI matrix requires 1.892 MBytes.)

$X_{\min}$	$\varepsilon = 10^{-2}$		$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-5}$	
	MBytes	ε_m	MBytes	ε_m	MBytes	ε_m	MBytes	ε_m
$2.5D_{\max}$	3.725	0.580×10^{-2}	5.618	0.4673×10^{-3}	7.932	0.479×10^{-4}	10.66	67.7×10^{-5}
$3.0D_{\max}$	3.444	1.69×10^{-2}	5.231	1.19×10^{-3}	7.440	0.530×10^{-4}	10.07	3.29×10^{-5}
$5.0D_{\max}$	3.621	1.54×10^{-2}	5.409	0.927×10^{-3}	5.935	1.89×10^{-4}	7.617	3.05×10^{-5}



(a) System matrix setup time.

(b) System matrix-vector product time.

(c) BI matrix memory requirement.

(d) Total memory requirement.

Figure 4.5: Effect of the number of FMM groups on various performance criteria for an empty cavity with dimensions $x \times y \times z = 0.3\lambda \times 20\lambda \times 0.2\lambda$, $\frac{\lambda}{6.67}$ discretization and 998 aperture degrees of freedom (CT/LN elements).

4.7 Conclusion

This chapter gave a detailed presentation of the application of the FMM to the BI component of the FE-BI formulation. A thorough review of FMM error control was presented and the computational efficiency of the FMM was investigated. Results confirmed the theory and showed that the FMM has a break-even point, below which it is more profitable to use the standard BI formulation.

The following, minor contributions were made in this chapter: A thorough review of FMM error control schemes was presented (with respect to truncation and spherical quadrature). A new, empirical formula for the truncation order in the addition theorem was introduced and evaluated.

Various extensions to the FMM, as presented here, are possible. [54, 178] present a technique to reduce the FMM, as presented here, to $\mathcal{O}\left(N_F^{1.33}\right)$ by filtering the addition theorem as a function of the orders of the terms, yielding a translation operator that is more spatially directed. The same idea is applied in 2D, in [193]. [30] observes that the translation operator in its standard form becomes highly directional as the groups grow farther apart. Based on this argument, parts of the spherical integration domain are simply discarded. Finally, as already mentioned in the introduction, the FMM can be extended to a multilevel algorithm [179], based on grouping schemes of various sizes on the same mesh. The interaction between two basis function fall into the grouping scheme with the largest groups such that the interaction is still 'far', which can lead to an $\mathcal{O}\left(N_F \log(N_F)\right)$ algorithm.

The error estimation results and adaptive procedures presented in the rest of this document (Chapters 5 and 6) for the cavity FE-BI formulation, are fully compatible with the FMM formulation presented here.

Chapter 5

A posteriori error estimates: theory and implementation

5.1 Introduction

In the FEM, an infinite dimensional solution space is approximated with a finite dimensional one. There is always an error associated with such a solution. Estimation of this error is a very important aspect of the FEM as a whole, since it can increase confidence in a solution and can be used to drive adaptive refinements of the discretization, optimizing the quality of the solution relative to the number of degrees of freedom and in turn reducing computational cost.

This chapter presents two a posteriori error estimation techniques for the FEM formulation presented in Chapter 2, with the cavity aperture boundary condition presented in Chapter 3 and the general, dominant mode, coaxial port boundary condition presented in Section E.3, still using the curl-conforming, vector elements of Appendix C, as before. The techniques presented here can be used in conjunction with the FMM of Chapter 4 in the case of the cavity FE-BI formulation.

At this point, it must be reiterated that the FEM formulation at hand suffers from an instability at low frequencies (as discussed within the introduction to Chapter 2). The two error estimation techniques presented in this chapter do not estimate the effect of this instability, as a result of their direct and strict derivation from the FEM formulation itself.

In the rest of the introduction, some basic concepts are discussed and an overview of the literature is given, resulting in further motivation for the work presented in this chapter. Section 5.2 gives a detailed description of the error in the FEM formulation used here, forming the basis for the two error estimation techniques derived in Sections 5.3 and 5.4. Section 5.3 presents the first error estimation technique: an explicit, residual based error estimator. Section 5.4 presents the second error estimation technique: an implicit, element residual method, error estimator.

5.1.1 Some basic concepts

The FEM *error* is defined as the difference between the true, generally infinite dimensional, solution of the underlying BVP, and the approximate, finite dimensional, FEM solution. The error is defined over the whole problem domain.

Fundamentally, error estimation can be divided into two categories [79, 181, 156, 4]:

- *A priori* error estimation is any result involving the error, that can be established before obtaining the approximate solution, based on knowledge of the particular FE formulation and on any additional, problem specific information.
- *A posteriori* error estimation is any method of (approximately) evaluating the solution error, as measured in some related, local or global quantity, given an initial solution. This quantity can be a traditional, functional analysis norm (for example, see norm definitions in [51]) or the energy norm associated with the FEM bilinear form or an observable quantity (usually a linear functional on the solution) related to the application at hand.

A priori results generally indicate the convergence, stability and asymptotic behaviour of the FEM formulation [4]. It usually is a global quantity independent of the specific discretization and can therefore not be used to drive selective refinements of the discretization [156]. On the other hand, a posteriori results usually consist of local and global data, on which selective refinements of the discretization can be based (local data) and which can serve as a stopping criterion (global data) in an iterative simulation-refinement cycle. The local data usually sum to form the global data. The local contributions are sometimes referred to as *error indicators*, while the technique as a whole is referred to as an *error estimator*.

An a posteriori error estimator should ideally exhibit the following properties [27, 79, 138]:

- It should be based on problem related data and a given solution.
- It should result in local knowledge of the error (the relative error distribution throughout the mesh).
- It should result in global knowledge of the solution quality as a whole.
- It should be a proper upper bound on the true error (reliability).
- It should lead to (near) optimal meshes in an adaptive algorithm (efficiency).
- The estimator calculation should have a computational cost of lower asymptotic order than a global FEM solution (computationally cheap).

Given an error estimator, it can be evaluated according to these criteria. In practice, most estimators only fulfill these (very optimistic) requirements to some degree, as indeed the two estimators derived in this chapter do, but that does not mean that they are not useful.

Provided that an estimator works at all, the most important requirement is probably that it should be computationally cheap, otherwise one can just as well re-solve the problem with a uniform refinement of the discretization, making the estimator mostly obsolete with respect to selective refinement, no matter how accurate it is.

5.1.2 Overview of a posteriori error estimation techniques

This section presents a review of the different approaches to FEM error estimation available in the general FEM literature. Note that most of these approaches are FEM formulation dependent and cannot necessarily be adapted to FEM formulations in general. The categorization is made along similar lines as in [138, 4, 181].

Explicit, residual based

These estimators are usually rigorously derived in the sense that the sum of the elemental indicators present an upper-bound (with multiplicative, unknown constant) on the error as measured in some norm, typically the energy norm or a L^m norm. To obtain L^m estimates, one needs to employ a stability condition and to formulate a dual problem [79, 181, 4]. The calculation of these estimators are based on the explicit evaluation of elemental volume and face (or in 2D, face and edge-) residuals of the approximate solution in relation to the exact one.

[18] presents energy- and L^m norm estimates of 1D, scalar, deterministic- and eigenvalue problems with a second order, self-adjoint operator. Only element interior residuals are employed. [111, 81] simplify the energy norm estimator of [18], extend it to the 2D, scalar, elliptic case and investigate adaptive convergence rates experimentally. [79, 4, 181] present error estimates in both the energy- and L^2 norms for the scalar elliptic BVP case. [106, 105, 104] collectively consider hyperbolic and parabolic problems as well. [106] considers problems of linear elasticity in 3D, where the working variable is a vector field. The same goes for [190], considering the Stokes equations (but in 2D). [33, 32] present an explicit estimator based on completely diagonalizing the bilinear form in the error VBVP (as also suggested in [23, 22]). The approach is shown to be a superset of the more conventional, explicit, residual based estimator presented in [16].

Implicit, residual based

This class of error estimators is based on the solution of local VBVPs (element-wise or on groups of contiguous elements), where the data of the problems is related to the residuals of the initial solution. The solutions to the local problems are approximations of the true error and are usually solved using basis functions of higher polynomial order than the initial solution. The main issue usually is the choice of boundary conditions for the local problems. This approach offers one the flexibility to choose any appropriate norm to measure the approximate error with, since the error field itself is approximated (not in some norm).

[19] splits the global error VBVP into independent, local, Dirichlet problems, based on a partition of unity over the problem domain. [110] presents element-wise VBVPs, based on a complementary variational formulation of the original problem, guaranteeing an upper bound on the global error in the energy norm. The boundary conditions of local problems are optimized through a process of *self-equilibration*, balancing internal- and boundary data. The concept of element-wise versions of the global VBVP, but in terms of the error, is investigated in detail in [24], where various schemes are proposed to deal with the local boundary conditions (second order, elliptic problems in 2D). The approach in [24] is called the *element residual method* (ERM), characterized by the solution of elemental, Neumann

VBVPs to approximate the error. The ERM is extended in [139, 190] to more general problems in computational fluid dynamics. [23, 22] present implicit error estimates, using hierarchical, higher order polynomial spaces to model the elemental VBVPs. [3] provides a general procedure for elemental, Neumann boundary condition equilibration, minimizing the overestimation of the error (this is a comprehensive extension of [110]) and [2] presents results obtained with this procedure. Finally, [4, 138] present various of these techniques together.

Estimation through post-processing

These techniques estimate the error in a derivative of the solution field, by comparing it with an improved version. Any suitable norm can be used for making the comparison, since the improved version is available point-wise. The improved version is obtained through some form of post-processing of the solution field. These estimators are also referred to as *recovery based.*

[15] presents a general post-processing procedure for calculating improved versions of the solution field and its derivatives and establishes a bound on the accuracy of the procedure, with the means to control it. However, the procedure can potentially be computationally costly. [203, 5] present an element-wise, L^2 projection of the solution derivative onto the set of solution basis functions as a means of obtaining an improved version. This procedure is sometimes referred to as the Z^2 error estimator. In [204, 205, 206], the *superconvergent patch recovery* (SPR) error estimator is presented, which is an improved version of the Z^2 estimator. It is based on the proposed existence of points on the elements at which the derivative of the solution field exhibits convergence that is of higher order than generally expected of the polynomial order of the solution derivative itself. The procedure can be very useful, even when the superconvergence property does not apply.

Targeted quantities

Also known as *goal-oriented-* or *targeted* error estimation, these techniques bound the error in a functional output or the accuracy of a functional output, based on the solution field. See [150] and the references therein for detailed accounts of available procedures.

5.1.3 Brief survey of FEM error estimation techniques and results in engineering electromagnetics

In this section, a brief survey is made of a posteriori error estimation techniques and results, as available in the literature for the FE solution of Maxwell's equations, motivated by engineering applications.

- *Explicit, residual based:* [167] applies the energy norm estimator presented in [111], within a 2D, time-harmonic, full-wave, Galerkin FEM setting. [147] presents an estimator for 3D, electrostatic problems. [133] rigorously derives a L^2 norm estimator for the 3D, time-harmonic, full-wave, Galerkin FEM, employing spherical harmonics (similar in some respect to the unimoment method [129]) to model an open boundary.

In [96], an explicit error estimator for a transient, full-wave, 3D, FE formulation is presented, built on the explicit error estimation framework established in [79]. [28] presents an explicit error estimator for eddy current problems, using curl-conforming, vector elements. This estimator is quite similar to the one derived later in this chapter. Finally, in [185] a linear functional norm, explicit estimator for the 3D, time-harmonic, full-wave, Galerkin FEM is derived. The key step within the derivation is a local, approximate, integral representation of the error field.

- *Implicit, residual based:* [130] applies the ERM, as presented in [138], to the 2D, time-harmonic, full-wave, TM- and TE- analysis of inhomogeneous scatterers, using a hybrid, FE-BI formulation. The ERM is presented in [60] for 2D, quasi-static problems, but with local Dirichlet boundary conditions, as opposed to the more conventional Neumann case. In [144], the ERM is formulated for a 3D, time-harmonic, full-wave FEM, with a penalty term to counteract the occurrence of spurious modes that accompany the use of Lagrange-type, interpolatory elements. In [98], the third estimator presented in [24] (ERM) is applied to the Galerkin FEM for electrostatic problems in 2D and 3D. The ERM (presumably as presented in [24]), is applied to 2D eddy current problems in [7] (local, elemental problem solved with higher order basis functions) and [6] (local, elemental problem solved with first order basis functions on a subdivision of the element). [29] presents an excellent, rigourously derived, hierarchical error estimator for eddy current problems. It is based on diagonalizing the bilinear form in the global error VBVP, similar to the technique presented in [33]. The curl-conforming, element spaces of mixed orders 0.5 and 1.5 are used (see Appendix C), but it would be possible to extend the estimator to arbitrary hierarchical order. [170] applies the ERM (without reference to a full description of the method used) to closed cavity, 3D, full-wave, eigenvalue problems.

- *Heuristic- and physics based:* This is a large class with respect to the FE solution of Maxwell's equations in the engineering community. These techniques tend not to be mathematically rigorous, but rather appeal to physical insight. [127] uses an estimator based on the comparison between the coefficients of the highest order basis functions on an element with the rest of its coefficients — an approach somewhat similar to the hierarchical nature of the ERM and also an intuitive, implicit application of an interpolation estimate such as equation (C.17). The L^2 norm, over inter-element boundaries, of the discontinuity in the non-conforming part of the field modeled with conforming vector elements (or in the field derivative), is used as an error estimator in [199, 84, 73, 11]. This approach can be seen as a type of heuristic, incomplete, explicit residual-based estimator. [94] employs the derivative of a stationary, global quantity with respect to nodal positions as a local error estimator. [126] shows that this approach can lead to erroneous adaptive refinements.

- *Based on post-processing:* In [90], a very heuristic, post-processing based estimator is presented and applied to 2D, static problems. [140] applies the Z^2 estimator of [203] within a 3D, eddy current formulation.

- *Complementary VBVP:* These generally require the solution of two *complementary* VBVPs, in order to obtain an upper bound on the error energy in either solution. Strictly speaking, these do not constitute a posteriori error estimators, since they are not based on local calculations. See [46, 35] for details and more references.

- *Targeted quantities:* Adaptation based on the effect of the (estimated) error on functional outputs is presented in [82]. Various conventional error estimation schemes are compared within this adaptive framework (quasi-static problems). A very detailed, mathematically thorough scheme is presented in [134], where the full-wave scattering of PEC objects in 2D is considered with a Galerkin FE formulation. The far-field pattern value is written in terms of the bilinear- and linear forms and is then bounded explicitly.

Clearly, there is a serious lack of variety in a posteriori error estimation techniques and results for the time-harmonic, full-wave, FEM in 3D (as used in this document) and practically a complete absence of error estimators that incorporate the BI component of the standard FE-BI approach as outlined, for example, in [99]. This is echoed in [70], where the need for investigation of the established techniques within this setting is also stated. The rest of this chapter is an effort towards addressing this need to some extent.

5.2 The error

This section defines and characterizes the error in the Galerkin FEM formulation presented in Chapter 2. The error is characterized in terms of global and local BVPs and VBVPs. Starting with alternative forms of the global VBVP of the error, these are used to construct global BVPs of the error, which are then used to formulate elemental error BVPs. Finally, the elemental BVPs are converted to elemental VBVPs.

The alternative forms presented for the global VBVP of the error will be used in Section 5.3 to construct an explicit residual based error estimator. The elemental VBVPs of the error will be used in Section 5.4 to obtain an implicit residual based error estimator.

5.2.1 Definition of the error

From equation (2.10), the electric field, VBVP is

$$\begin{cases} B(\mathbf{E}, \mathbf{W}) = L(\mathbf{W}) \quad \forall\ \mathbf{W} \in W \\ \mathbf{E} \in W \end{cases} \tag{5.1}$$

and in discrete form (equation (2.23) repeated):

$$\begin{cases} B(\mathbf{E}_h, \mathbf{W}_h) = L(\mathbf{W}_h) \quad \forall\ \mathbf{W}_h \in W_h \\ \mathbf{E}_h \in W_h \end{cases} . \tag{5.2}$$

The error can be defined as

$$\mathbf{e}_h = \mathbf{E} - \mathbf{E}_h. \tag{5.3}$$

The error is orthogonal to the original testing functions (sometimes referred to as Galerkin orthogonality), since

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}_h) &= B(\mathbf{E}, \mathbf{W}_h) - B(\mathbf{E}_h, \mathbf{W}_h) \\ &= L(\mathbf{W}_h) - L(\mathbf{W}_h) \\ &= 0. \end{aligned} \tag{5.4}$$

5.2.2 The global VBVP of the error

Since

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}) &= B(\mathbf{E}, \mathbf{W}) - B(\mathbf{E}_h, \mathbf{W}) \\ &= L(\mathbf{W}) - B(\mathbf{E}_h, \mathbf{W}), \end{aligned} \tag{5.5}$$

the error VBVP is

$$\begin{cases} B(\mathbf{e}_h, \mathbf{W}) = L(\mathbf{W}) - B(\mathbf{E}_h, \mathbf{W}) \ \ \forall\ \mathbf{W} \in W \\ \mathbf{e}_h \in W \end{cases}. \tag{5.6}$$

The error VBVP can be written out on an element-by-element basis, yielding

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}) &= \sum_{i=1}^{N_\tau} \Bigg(- \int_{\partial K_i \cap \Gamma_N} \frac{1}{\mu_r} \mathbf{N} \cdot \mathbf{W} \, dS - jk_0 Z_0 \int_{K_i} \mathbf{J} \cdot \mathbf{W} \, dV \\ &\quad - \int_{K_i} \frac{1}{\mu_r} \nabla \times \mathbf{E}_h \cdot \nabla \times \mathbf{W} - k_0^2 \epsilon_r \mathbf{E}_h \cdot \mathbf{W} \, dV \Bigg) \end{aligned} \tag{5.7}$$

Now apply identities (A.10) and (A.8) in succession to the curl-dot-curl integral term in equation (5.7), obtaining

$$\begin{aligned} \int_{K_i} \frac{1}{\mu_r} \nabla \times \mathbf{E}_h \cdot \nabla \times \mathbf{W} \, dV &= \int_{K_i} \left(\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E}_h \right) \cdot \mathbf{W} \, dV \\ &\quad - \oint_{\partial K_i} \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}_h) \cdot \mathbf{W} \, dS. \end{aligned} \tag{5.8}$$

Substitute back into equation (5.7), to yield

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}) &= \sum_{i=1}^{N_\tau} \Bigg(- \int_{\partial K_i \cap \Gamma_N} \frac{1}{\mu_r} \mathbf{N} \cdot \mathbf{W} \, dS - jk_0 Z_0 \int_{K_i} \mathbf{J} \cdot \mathbf{W} \, dV \\ &\quad - \int_{K_i} \left[\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E}_h - k_0^2 \epsilon_r \mathbf{E}_h \right] \cdot \mathbf{W} \, dV \\ &\quad + \oint_{\partial K_i} \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}_h) \cdot \mathbf{W} \, dS \Bigg) \end{aligned} \tag{5.9}$$

This can be rewritten as

$$B(\mathbf{e}_h, \mathbf{W}) = \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W} \, dV + \oint_{\partial K_i} \mathbf{R}_S \cdot \mathbf{W} \, dS \right) \tag{5.10}$$

with the volume and surface residuals defined as

$$\mathbf{R}_V = -\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E}_h + k_0^2 \epsilon_r \mathbf{E}_h - jk_0 Z_0 \mathbf{J} \quad \text{in } K_i;\ i = 1, ..., N_\tau \tag{5.11}$$

$$\mathbf{R}_S = \begin{cases} \frac{1}{\mu_r} \hat{n} \times \nabla \times \mathbf{E}_h & \text{on } \partial K_i \setminus \Gamma_N;\ i = 1, ..., N_\tau \\ \frac{1}{\mu_r} \hat{n} \times \nabla \times \mathbf{E}_h - \frac{1}{\mu_r} \mathbf{N} & \text{on } \partial K_i \cap \Gamma_N;\ i = 1, ..., N_\tau \end{cases}. \tag{5.12}$$

The error VBVP can also be written in terms of element volume residuals and residuals on every face, rather than on every elemental boundary. Rewrite equation (5.10) as

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}) &= \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W}\, dV + \oint_{\partial K_i} \mathbf{R}_S \cdot \mathbf{W}\, dS \right) \\ &= \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W}\, dV \right) \\ &\quad + \sum_{m=1}^{N_f} \left(\int_{f_m \setminus \Gamma_N} \hat{n}^{(12)} \times \left[\frac{1}{\mu_r^{(1)}} \nabla \times \mathbf{E}_h^{(1)} - \frac{1}{\mu_r^{(2)}} \nabla \times \mathbf{E}_h^{(2)} \right] \cdot \mathbf{W}\, dS \right. \\ &\quad \left. + \int_{f_m \cap \Gamma_N} \frac{1}{\mu_r} [\hat{n} \times \nabla \times \mathbf{E}_h - \mathbf{N}] \cdot \mathbf{W}\, dS \right) \\ &= \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W}\, dV \right) + \sum_{m=1}^{N_f} \left(\int_{f_m} \mathbf{R}_f \cdot \mathbf{W}\, dS \right) \end{aligned} \tag{5.13}$$

with the face residual defined as

$$\mathbf{R}_f = \begin{cases} \hat{n}^{(12)} \times \left[\frac{1}{\mu_r^{(1)}} \nabla \times \mathbf{E}_h^{(1)} - \frac{1}{\mu_r^{(2)}} \nabla \times \mathbf{E}_h^{(2)} \right] & \text{on } f_m \setminus \Gamma_N;\ m = 1, ..., N_f \\ \frac{1}{\mu_r} [\hat{n} \times \nabla \times \mathbf{E}_h - \mathbf{N}] & \text{on } f_m \cap \Gamma_N;\ m = 1, ..., N_f \end{cases} . \tag{5.14}$$

f_m denotes a specific face in the mesh. The total number of faces in the mesh are N_f. The superscripts (1) and (2) designate association with the first and second elements sharing a face. The normal vector $\hat{n}^{(12)}$ points from the first element to the second element, normal to their shared face.

Note that, in equation (5.13), the normal components of $\mathbf{W}$ may experience a jump at inter-element boundaries and are therefore undefined at a boundary, since $\mathbf{W}$ is only guaranteed to have tangential continuity. This is not a problem in the case of equation (5.13), since $\mathbf{R}_f$ is purely tangential to all faces. To avoid any confusion, this fact can be used to rewrite the facial integrals such that the normal components of $\mathbf{W}$ are explicitly discarded. Observe that $\mathbf{R}_f$ is purely tangential at all faces, therefore $\hat{n} \times \hat{n} \times \mathbf{R}_f = -\mathbf{R}_f$. Substituting this into equation (5.13) and applying identity (A.8), one obtains the following alternative form:

$$B(\mathbf{e}_h, \mathbf{W}) = \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W}\, dV \right) + \sum_{m=1}^{N_f} \left(\int_{f_m} \hat{n} \times \mathbf{R}_f \cdot \hat{n} \times \mathbf{W}\, dS \right). \tag{5.15}$$

5.2.3 The global BVP of the error

Start by rewriting the double curl term in $B(\mathbf{e}_h, \mathbf{W})$ as follows:

$$\begin{aligned} \int_\Omega \frac{1}{\mu_r} \nabla \times \mathbf{e}_h \cdot \nabla \times \mathbf{W}\, dV &= \int_\Omega \left(\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h \right) \cdot \mathbf{W}\, dV \\ &\quad - \oint_\Gamma \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{e}_h) \cdot \mathbf{W}\, dS \\ &= \int_\Omega \left(\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h \right) \cdot \mathbf{W}\, dV \\ &\quad - \int_{\Gamma_N} \frac{1}{\mu_r} (\mathbf{N} - \hat{n} \times \nabla \times \mathbf{E}_h) \cdot \mathbf{W}\, dS. \end{aligned} \tag{5.16}$$

where the first step follows from identities (A.10) and (A.8). Substitute the definition of the error into the boundary integral term, then use the facts that the homogeneous Diriclet boundary condition is satisfied by $\mathbf{W}$ (see equation (2.11)) and that the inhomogeneous Neumann boundary condition is satisfied by $\mathbf{E}$ (see equation (2.3)) to obtain the second step.

Substitute the RHS of equation (5.16) into the LHS of equation (5.10). In the resulting equation, both sides are of the form $\int_\Omega(\text{vector})\cdot\mathbf{W}\,dV$. Equating the kernels, the following BVP, describing $\mathbf{e}_h$ on Ω, is obtained:

$$\begin{cases} \nabla\times\dfrac{1}{\mu_r}\nabla\times\mathbf{e}_h - k_0^2\epsilon_r\mathbf{e}_h = \displaystyle\sum_{i=1}^{N_\tau}\left(\mathbf{R}_V|_{K_i} + \delta_{\partial K_i\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_S\right) & \text{on } \Omega \\ \hat{n}\times\mathbf{e}_h = 0 & \text{on } \Gamma_D \\ \hat{n}\times\nabla\times\mathbf{e}_h = \mathbf{N} - \hat{n}\times\nabla\times\mathbf{E}_h & \text{on } \Gamma_N \end{cases} \tag{5.17}$$

where the Dirac delta function is concentrated on the surface described in its subscript (similar to a 2D equivalent used in [111]), such that

$$\int_\Omega \delta_{\partial K_i\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_S\cdot\mathbf{W}\,dV = \int_{\partial K_i\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_S\cdot\mathbf{W}\,dS. \tag{5.18}$$

An alternative form of the BVP, describing $\mathbf{e}_h$ on Ω, is obtained by substituting the RHS of equation (5.16) into the LHS of equation (5.13), yielding

$$\begin{cases} \nabla\times\dfrac{1}{\mu_r}\nabla\times\mathbf{e}_h - k_0^2\epsilon_r\mathbf{e}_h = \displaystyle\sum_{i=1}^{N_\tau}\mathbf{R}_V|_{K_i} + \sum_{m=1}^{N_f}\delta_{f_m\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_f & \text{on } \Omega \\ \hat{n}\times\mathbf{e}_h = 0 & \text{on } \Gamma_D \\ \hat{n}\times\nabla\times\mathbf{e}_h = \mathbf{N} - \hat{n}\times\nabla\times\mathbf{E}_h & \text{on } \Gamma_N \end{cases} \tag{5.19}$$

with

$$\int_\Omega \delta_{f_m\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_f\cdot\mathbf{W}\,dV = \int_{f_m\setminus(\Gamma_D\cup\Gamma_N)}\mathbf{R}_f\cdot\mathbf{W}\,dS. \tag{5.20}$$

5.2.4 Elemental VBVPs of the error

In this section, the global VBVP of the error is decomposed into elemental VBVPs. First note that a BVP on any sub-domain $\Lambda\subseteq\Omega$ can be constructed from equation (5.17) by restricting the problem domain to Λ and furnishing it with appropriate Dirichlet and/or Neumann boundary conditions, resulting in the following BVP representation of $\mathbf{e}_h$ on Λ:

$$\begin{cases} \nabla\times\dfrac{1}{\mu_r}\nabla\times\mathbf{e}_h - k_0^2\epsilon_r\mathbf{e}_h = \displaystyle\sum_{i=1}^{N_\tau}\left(\mathbf{R}_V|_{\Lambda\cap K_i} + \delta_{\Lambda\cap(\partial K_i\setminus(\Gamma_D\cup\Gamma_N))}\mathbf{R}_S\right) & \text{on } \Lambda \\ \hat{n}\times\mathbf{e}_h = \hat{n}\times\mathbf{E} - \hat{n}\times\mathbf{E}_h & \text{on } \partial\Lambda_D \\ \hat{n}\times\nabla\times\mathbf{e}_h = \hat{n}\times\nabla\times\mathbf{E} - \hat{n}\times\nabla\times\mathbf{E}_h & \text{on } \partial\Lambda_N \\ \hat{n}\times\mathbf{e}_h = 0 & \text{on } \partial\Lambda\cap\Gamma_D \\ \hat{n}\times\nabla\times\mathbf{e}_h = \mathbf{N} - \hat{n}\times\nabla\times\mathbf{E}_h & \text{on } \partial\Lambda\cap\Gamma_N \end{cases} \tag{5.21}$$

where $\partial\Lambda_D$ and $\partial\Lambda_N$ are surfaces of imposed, inhomogeneous Dirichlet- and Neumann boundary conditions on $\partial\Lambda\setminus(\Gamma_D\cup\Gamma_N)$.

An alternative, equivalent BVP representation of $\mathbf{e}_h$ on Λ can be obtained from equation (5.19), yielding another BVP with the same boundary conditions as (5.21), but with the following PDE:

$$\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h - k_0^2 \epsilon_r \mathbf{e}_h = \sum_{i=1}^{N_\tau} \mathbf{R}_V|_{\Lambda \cap K_i} + \sum_{m=1}^{N_f} \delta_{\Lambda \cap (f_m \setminus (\Gamma_D \cup \Gamma_N))} \mathbf{R}_f \quad \text{on } \Lambda. \tag{5.22}$$

In both cases $\partial\Lambda_D$, $\partial\Lambda_N$ and $\partial\Lambda \cap (\Gamma_D \cup \Gamma_N)$ are disjoint and $\partial\Lambda = \partial\Lambda_D \cup \partial\Lambda_N \cup (\partial\Lambda \cap (\Gamma_D \cup \Gamma_N))$ holds.

Before choosing Λ to be an elemental volume, the following definitions are in order:

- K_i^- is the i-th elemental volume with an infinitesimally thin layer removed at its surface.
- K_i is the i-th elemental volume proper.
- K_i^+ is the i-th elemental volume with an infinitesimally thin layer added at its surface.

Choosing Λ equal to each of these volumes result in three different BVP representations of $\mathbf{e}_h$ 'on K_i'. Let the superscript $*$ indicate association with the element adjacent to K_i at the interface concerned.

- $\Lambda = K_i^-$:

$$\begin{cases} \nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h - k_0^2 \epsilon_r \mathbf{e}_h = \mathbf{R}_V|_{K_i} & \text{on } K_i \\ \hat{n} \times \mathbf{e}_h = \hat{n} \times \mathbf{E} - \hat{n} \times \mathbf{E}_h|_{K_i} & \text{on } \partial K_{iD} \\ \hat{n} \times \nabla \times \mathbf{e}_h = \hat{n} \times \nabla \times \mathbf{E} - \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i} & \text{on } \partial K_{iN} \\ \hat{n} \times \mathbf{e}_h = 0 & \text{on } \partial K_i \cap \Gamma_D \\ \hat{n} \times \nabla \times \mathbf{e}_h = \mathbf{N} - \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i} & \text{on } \partial K_i \cap \Gamma_N \end{cases} \tag{5.23}$$

- $\Lambda = K_i$:

  In this case the PDE would be

$$\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h - k_0^2 \epsilon_r \mathbf{e}_h = \mathbf{R}_V|_{K_i} + \delta_{\partial K_i \setminus (\Gamma_D \cup \Gamma_N)} \mathbf{R}_S \quad \text{on } K_i, \tag{5.24}$$

  but since

$$\frac{1}{(\mu_r)_i} \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i} \neq \frac{1}{(\mu_r)_i^*} \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i^*} \quad \text{on } \partial K_{iN}, \tag{5.25}$$

  the Neumann boundary condition required on ∂K_{iN} exactly, is undefined. The BVP cannot be posed in this case.

- $\Lambda = K_i^+$:

$$\begin{cases} \nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{e}_h - k_0^2 \epsilon_r \mathbf{e}_h = \mathbf{R}_V|_{K_i} + \sum_{m=1}^{N_f} \delta_{f_m \cap (\partial K_i \setminus (\Gamma_D \cup \Gamma_N))} \mathbf{R}_f & \text{on } K_i \\ \hat{n} \times \mathbf{e}_h = \hat{n} \times \mathbf{E} - \hat{n} \times \mathbf{E}_h|_{K_i^*} & \text{on } \partial K_{iD} \\ \hat{n} \times \nabla \times \mathbf{e}_h = \hat{n} \times \nabla \times \mathbf{E} - \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i^*} & \text{on } \partial K_{iN} \\ \hat{n} \times \mathbf{e}_h = 0 & \text{on } \partial K_i \cap \Gamma_D \\ \hat{n} \times \nabla \times \mathbf{e}_h = \mathbf{N} - \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i} & \text{on } \partial K_i \cap \Gamma_N \end{cases} \tag{5.26}$$

Again, note that ∂K_{iD} and ∂K_{iN} are surfaces of imposed, inhomogeneous Dirichlet- and Neumann boundary conditions on $\partial K_i \setminus (\Gamma_D \cup \Gamma_N)$ with ∂K_{iD}, ∂K_{iN} and $\partial K_i \cap (\Gamma_D \cup \Gamma_N)$ disjoint and that $\partial K_i = \partial K_{iD} \cup \partial K_{iN} \cup (\partial K_i \cap (\Gamma_D \cup \Gamma_N))$ holds.

These BVPs can be converted into weak forms as VBVPs, in the same way that the electric field, vector wave equation BVP was converted into a VBVP representation in Section 2.2. To this end, define the elemental function spaces

$$W_{iD} = \left\{ \mathbf{a} \in H(\text{curl}, K_i) \;\middle|\; \begin{array}{ll} \hat{n} \times \mathbf{a} = \hat{n} \times \mathbf{E} - \hat{n} \times \mathbf{E}_h|_{K_i} & \text{on } \partial K_{iD} \\ \hat{n} \times \mathbf{a} = 0 & \text{on } \partial K_i \cap \Gamma_D \end{array} \right\} \tag{5.27}$$

$$W_i = \left\{ \mathbf{a} \in H(\text{curl}, K_i) \;\middle|\; \hat{n} \times \mathbf{a} = 0 \quad \text{on } \partial K_{iD} \cup (\partial K_i \cap \Gamma_D) \right\}. \tag{5.28}$$

The VBVP related to equation (5.23) can now be stated as

$$\left\{ \begin{array}{l} B_i(\mathbf{e}_h, \mathbf{W}) = -\int_{\partial K_i \cap \Gamma_N} \frac{1}{\mu_r} \left(\mathbf{N} - \hat{n} \times \nabla \times \mathbf{E}_h|_{K_i} \right) \cdot \mathbf{W} \, dS \\ -\int_{\partial K_{iN}} \frac{1}{\mu_r} \left(\hat{n} \times \nabla \times \left[\mathbf{E} - \mathbf{E}_h|_{K_i} \right] \right) \cdot \mathbf{W} \, dS + \langle \mathbf{R}_V, \mathbf{W} \rangle_{\mathcal{C}, K_i} \\ \qquad \forall \; \mathbf{W} \in W_i \; ; \quad \mathbf{e}_h \in W_{iD} \end{array} \right. \tag{5.29}$$

Use equation (5.8) (with $\mathbf{E} \leftarrow \mathbf{E}_h$) and the definitions of $\mathbf{R}_V$ and W_i to rewrite this as

$$\left\{ \begin{array}{l} B_i(\mathbf{e}_h, \mathbf{W}) = L_i(\mathbf{W}) - B_i(\mathbf{E}_h, \mathbf{W}) - \int_{\partial K_{iN}} \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}) \cdot \mathbf{W} \, dS \\ \qquad \forall \; \mathbf{W} \in W_i \; ; \quad \mathbf{e}_h \in W_{iD} \end{array} \right. \tag{5.30}$$

with the restriction of the global bilinear- and linear forms to K_i as follows:

$$B_i(\mathbf{e}_h, \mathbf{W}) = \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{e}_h, \nabla \times \mathbf{W} \right\rangle_{\mathcal{C}, K_i} - k_0^2 \langle \epsilon_r \mathbf{e}_h, \mathbf{W} \rangle_{\mathcal{C}, K_i} \tag{5.31}$$

$$L_i(\mathbf{W}) = -\int_{\partial K_i \cap \Gamma_N} \frac{1}{\mu_r} \mathbf{N} \cdot \mathbf{W} \, dS - j k_0 Z_0 \langle \mathbf{J}, \mathbf{W} \rangle_{\mathcal{C}, K_i} . \tag{5.32}$$

After noting that $\hat{n} \times \mathbf{E}_h|_{K_i} = \hat{n} \times \mathbf{E}_h|_{K_i^*}$ on ∂K_{iD} (as a consequence of $\mathbf{E}_h \in W_h$), a careful examination of equation (5.26) reveals that it results in the same VBVP as equation (5.23).

Finally, note that equation (5.30) is in the so-called *weak residual* form (as opposed to equation (5.29)), as discussed in [120], which has the advantage that it could easily be extended to groups of elements (not to be considered here).

5.3 An explicit, residual based, error estimator

In this section, an explicit, residual based, error estimator is derived, bounding the error as measured in an approximate energy norm (to be defined), provided that the regularity requirements of the interpolation estimates employed (presented in Section C.1.3), hold. An approximate energy norm is used, since the bilinear form of the FE formulation does not naturally introduce a proper energy norm. It is interesting to note that [27] states that

residual based, a posteriori error estimation "defies a rigorous theoretical treatment" in this case, but that techniques derived with the zero order term having a positive weight, can "still be applied successfully". This is further motivation for the approximate energy norm approach followed here.

The derivation is along the same lines as in [79, 4, 181, 106, 190], but adapted to the 3D, vector wave equation, Galerkin FEM formulation with tetrahedral, curl-conforming, vector elements as presented in Appendic C. Note that here, the dual problem (in the terminology of [79, 106]) is simply the error VBVP in equation (5.6), since the 'energy norm' is being bounded.

In this section, the generic constant C will be used, resulting from the interpolation estimates. It may change values from one step to the next, but will remain independent of the approximate solution field and source terms.

5.3.1 The general VBVP

In the light of identities (A.2), (A.12) and (A.13), equation (5.10) can be changed into the following inequality:

$$|B(\mathbf{e}_h, \mathbf{W})| \leq \sum_{i=1}^{N_\tau} \left(\|\mathbf{R}_V\|_{L^2(K_i)} \|\mathbf{W}\|_{L^2(K_i)} + \|\mathbf{R}_S\|_{L^2(\partial K_i)} \|\mathbf{W}\|_{L^2(\partial K_i)} \right). \tag{5.33}$$

Now observe that the the orthogonality relation (5.4) implies that

$$B(\mathbf{e}_h, \mathbf{W}) = L(\mathbf{W} - \pi\mathbf{W}) - B(\mathbf{E}_h, \mathbf{W} - \pi\mathbf{W}), \tag{5.34}$$

with π designating the global interpolation operator $\pi : W \rightarrow W_h$, as described in Section C.1.3. Thus equation (5.33) can be rewritten as

$$|B(\mathbf{e}_h, \mathbf{W})| \leq \sum_{i=1}^{N_\tau} \left(\|\mathbf{R}_V\|_{L^2(K_i)} \|\mathbf{W} - \pi_{K_i}\mathbf{W}\|_{L^2(K_i)} + \|\mathbf{R}_S\|_{L^2(\partial K_i)} \|\mathbf{W} - \pi_{K_i}\mathbf{W}\|_{L^2(\partial K_i)} \right). \tag{5.35}$$

Note that π is now applied locally, but this is equivalent to the global version, since the interpolation operator only depends on inter-element, tangential components, which are continuous by definition ($\mathbf{W} \in W$).

Let p_i indicate the orders of the elements that were used to obtain $\mathbf{E}_h$: $p_i = m-1$ in the case of an element modeling the space $M_m(K_i)$ and $p_i = m$ in the case of an element modeling the space $(P_m(K_i))^3$. Define k as

$$0 \leq k \leq \lfloor \min \{p_1; ...; p_{N_\tau}\} \rfloor, \tag{5.36}$$

where $\lfloor \cdot \rfloor$ represents the floor function. Thus the upper bound on k represents the maximum polynomial order to which all elements are complete.

Apply equations (C.18) and (C.22) to equation (5.35) to rewrite all norms involving $\mathbf{W}$, in terms of the Sobolev semi-norm $|\cdot|_{(H^{k+1}(K))^3}$, yielding

$$|B(\mathbf{e}_h, \mathbf{W})| \leq C \sum_{i=1}^{N_\tau} \left(h_i^{k+1} \|\mathbf{R}_V\|_{L^2(K_i)} + h_i^{k+0.5} \|\mathbf{R}_S\|_{L^2(\partial K_i)} \right) |\mathbf{W}|_{(H^{k+1}(K_i))^3}$$

$$\leq C\left[\sum_{m=1}^{N_\tau}|\mathbf{W}|^2_{(H^{k+1}(K_m))^3}\right]^{\frac{1}{2}}\left[\sum_{i=1}^{N_\tau}\left(h_i^{2k+2}\|\mathbf{R}_V\|^2_{L^2(K_i)}+h_i^{2k+1}\|\mathbf{R}_S\|^2_{L^2(\partial K_i)}\right)\right]^{\frac{1}{2}} \tag{5.37}$$

with h_i defined as

$$h_i = \operatorname{diam}(K_i). \tag{5.38}$$

The last line of equation (5.37) was obtained via application of the Cauchy-Schwarz inequality (A.3). Substitute $\mathbf{W} \leftarrow \mathbf{e}_h$ and divide throughout with the semi-norm factor, to obtain

$$\|\mathbf{e}_h\|_{E^a(\Omega,\tau,k+1)} \leq C\left[\sum_{i=1}^{N_\tau}\left(h_i^{2k+2}\|\mathbf{R}_V\|^2_{L^2(K_i)}+h_i^{2k+1}\|\mathbf{R}_S\|^2_{L^2(\partial K_i)}\right)\right]^{\frac{1}{2}} \tag{5.39}$$

with the *approximate energy norm* defined as

$$\begin{aligned}\|\mathbf{v}\|_{E^a(\Omega,\tau,p)} &\equiv \frac{|B(\mathbf{v},\mathbf{v})|}{\left[\sum_{m=1}^{N_\tau}|\mathbf{v}|^2_{(H^p(K_m))^3}\right]^{\frac{1}{2}}} \\ &= \frac{\left|\int_\Omega \frac{1}{\mu_r}\nabla\times\mathbf{v}\cdot\nabla\times\mathbf{v}-k_0^2\epsilon_r\mathbf{v}\cdot\mathbf{v}\,dV\right|}{\left[\sum_{m=1}^{N_\tau}|\mathbf{v}|^2_{(H^p(K_m))^3}\right]^{\frac{1}{2}}}.\end{aligned} \tag{5.40}$$

Note that $\|\mathbf{e}_h\|_{E^a(\Omega,\tau,p)}$ is not a proper norm of the error field, because it does not conform to the well known specifications of a proper norm [158, 115], since $\|\mathbf{e}_h\|_{E^a(\Omega,\tau,p)} = 0 \not\Rightarrow \mathbf{e}_h = 0$. However, $\mathbf{e}_h = 0 \Rightarrow \|\mathbf{e}_h\|_{E^a(\Omega,\tau,p)} = 0$. For these reasons, the 'norm' is termed *approximate*. Regarding the naming *energy norm*: it is commonly used in the literature when referring to a norm definition following directly from the bilinear form of the VBVP (see for example [4, 79, 24]).

One would like the RHS of equation (5.39) also to be equal to zero when there is no error. Consider first what happens to $\|\mathbf{R}_V\|_{L^2(K_i)}$: since the vector wave equation is exactly satisfied by $\mathbf{E}_h$, it is clear from the definition of $\mathbf{R}_V$ that this term will be zero. Next, consider the $\|\mathbf{R}_S\|_{L^2(\partial K_i)}$ term: $\mathbf{R}_S$ will not be zero, but from the Maxwell continuity conditions it is clear that $\mathbf{R}_S$ will cancel itself out at inter element boundaries, since every shared boundary is integrated twice, with reversed sign, in equation (5.10). Unfortunately, by taking the norm of $\mathbf{R}_S$ before this cancellation takes place, the RHS of equation (5.39) does not incorporate this behaviour and will therefore not be zero when the error is zero.

To rectify this problem, start by applying the same steps as between equations (5.10) and (5.35), to the alternative form of the error VBVP given by equation (5.15). This yields

$$\begin{aligned}|B(\mathbf{e}_h,\mathbf{W})| \leq\ & \sum_{i=1}^{N_\tau}\|\mathbf{R}_V\|_{L^2(K_i)}\|\mathbf{W}-\pi_{K_i}\mathbf{W}\|_{L^2(K_i)} \\ & +\sum_{m=1}^{N_f}\|\hat{n}\times\mathbf{R}_f\|_{L^2(f_m)}\|\hat{n}\times(\mathbf{W}-\pi\mathbf{W})\|_{L^2(f_m)}.\end{aligned} \tag{5.41}$$

Since the $\hat{n}\times$ operation on $\mathbf{W} - \pi\mathbf{W}$ just discards one of its components and rotates the other two (which are equal in both elements sharing the face), it clearly follows that

$$\|\hat{n} \times (\mathbf{W} - \pi\mathbf{W})\|_{L^2(f_m)} \leq \|\mathbf{W} - \pi\mathbf{W}\|_{L^2(\partial K_m)} \tag{5.42}$$

where K_m is the element with the largest diameter of the two elements sharing f_m (the most conservative choice in light of the next step). After making this substitution in equation (5.41), apply the steps leading from equation (5.35) to equation (5.39) to it, yielding

$$\begin{aligned} \|\mathbf{e}_h\|_{E^a(\Omega,\tau,k+1)} &\leq C\left[\sum_{i=1}^{N_\tau} h_i^{2k+2}\|\mathbf{R}_V\|^2_{L^2(K_i)} + \sum_{m=1}^{N_f} h_m^{2k+1}\|\hat{n}\times\mathbf{R}_f\|^2_{L^2(f_m)}\right]^{\frac{1}{2}} \\ &= C\left[\sum_{i=1}^{N_\tau} h_i^{2k+2}\|\mathbf{R}_V\|^2_{L^2(K_i)} + \sum_{m=1}^{N_f} h_m^{2k+1}\|\mathbf{R}_f\|^2_{L^2(f_m)}\right]^{\frac{1}{2}}, \end{aligned} \tag{5.43}$$

with h_i and h_m defined as

$$h_i = \operatorname{diam}(K_i) \tag{5.44}$$

$$h_m = \begin{cases} \max\left\{\operatorname{diam}(K^{(1)}), \operatorname{diam}(K^{(2)})\right\} & \text{internal face} \\ \operatorname{diam}(K^{(1)}) & \text{boundary face} \end{cases}, \tag{5.45}$$

where the superscripts (1) and (2) again designate association with the first and second elements sharing f_m. The last step in equation (5.43) follows from the fact that $\mathbf{R}_f$ is tangential to all faces, therefore $\|\mathbf{R}_f\|_{L^2(f_m)} = \|\hat{n}\times\mathbf{R}_f\|_{L^2(f_m)}$.

Concerning face residuals with $f \subseteq \Gamma_D$: they must be discarded, since $\mathbf{W} \in W$ in equation (5.15) has no tangential components to such faces.

Equation (5.43) is in the desired format. The approximate energy norm of the error is bounded by a sum of elemental volume- and face residual contributions, all of which will go to zero when $\mathbf{E}_h$ satisfies the vector wave equation and the Maxwell continuity conditions (i.e. $\mathbf{e}_h = 0$). This behaviour is consistent with the behaviour of the approximate energy norm itself. Therefore: the RHS of equation (5.43) can reliably indicate the presence of an error, but not the absence thereof. It can be used to construct elemental, relative error indicators, but does not give a computable global error bound on the solution, due to the presence of the unknown constant C. The construction of the elemental indicators will be discussed in Chapter 6.

Finally, note that deriving a lower bound on the approximate energy norm (to show efficiency, as in [24] for example) would not make sense, since it is not a proper norm.

5.3.2 The cavity backed aperture VBVP

In this section, the explicit, residual based bound on the approximate energy norm, as established for the general VBVP (see equation (5.43)), will be established for the VBVP describing the cavity backed aperture problem, as formulated in Chapter 3.

Start by using equation (3.35) to define the electric field, boundary integral operator:

$$F_E^{\mathrm{BI}}(S_a, \mathbf{v}) = 2\int_{S_a} \{\hat{z}\times\mathbf{v}(\mathbf{r}')|_{z'=0}\}\cdot\left\{\nabla'\times\bar{\bar{\mathrm{I}}}G_0(\mathbf{r}',\mathbf{r})\right\} dS'. \tag{5.46}$$

Note that

$$\mathbf{E}|_{z>0} = \mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a, \mathbf{E}). \tag{5.47}$$

Now rewrite the bilinear and linear forms of the cavity backed aperture VBVP in terms of the electric field, boundary integral operator, yielding

$$B(\mathbf{E}, \mathbf{W}) = \left\langle \frac{1}{\mu_r}\nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{C,\Omega} - k_0^2 \langle \epsilon_r \mathbf{E}, \mathbf{W} \rangle_{C,\Omega} + \int_{S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(S_a, \mathbf{E}) \right\} \cdot \mathbf{W}\, dS \tag{5.48}$$

$$L(\mathbf{W}) = -jk_0 Z_0 \langle \mathbf{J}, \mathbf{W} \rangle_{C,\Omega} + 2jk_0 Z_0 \int_{S_a} \left[\hat{z} \times \mathbf{H}^{\text{inc}} \right] \cdot \mathbf{W}\, dS. \tag{5.49}$$

Construct the cavity backed aperture error VBVP from equation (5.6) and write it out in an element-by-element fashion, yielding

$$\begin{aligned} B(\mathbf{e}_h, \mathbf{W}) \;=\; & \sum_{i=1}^{N_\tau} \Bigg(- \int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times \left[\mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a, \mathbf{E}_h) \right] \right\} \cdot \mathbf{W}\, dS \\ & - \int_{K_i} \frac{1}{\mu_r} \nabla \times \mathbf{E}_h \cdot \nabla \times \mathbf{W} - k_0^2 \epsilon_r \mathbf{E}_h \cdot \mathbf{W} + jk_0 Z_0 \mathbf{J} \cdot \mathbf{W}\, dV \Bigg). \end{aligned} \tag{5.50}$$

Apply equation (5.8) to equation (5.50) and rewrite in terms of elemental and facial contributions, to obtain

$$B(\mathbf{e}_h, \mathbf{W}) = \sum_{i=1}^{N_\tau} \left(\int_{K_i} \mathbf{R}_V \cdot \mathbf{W}\, dV \right) + \sum_{m=1}^{N_f} \left(\int_{f_m} \mathbf{R}_f \cdot \mathbf{W}\, dS \right) \tag{5.51}$$

with

$$\mathbf{R}_V = -\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E}_h + k_0^2 \epsilon_r \mathbf{E}_h - jk_0 Z_0 \mathbf{J} \quad \text{in } K_i;\ i = 1, ..., N_\tau \tag{5.52}$$

$$\mathbf{R}_f = \begin{cases} \hat{n}^{(12)} \times \left[\dfrac{1}{\mu_r^{(1)}} \nabla \times \mathbf{E}_h^{(1)} - \dfrac{1}{\mu_r^{(2)}} \nabla \times \mathbf{E}_h^{(2)} \right] & \text{on } f_m \setminus S_a;\ m = 1, ..., N_f \\ \dfrac{1}{\mu_r} \hat{n} \times \nabla \times \mathbf{E}_h - \hat{n} \times \nabla \times \left[\mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a, \mathbf{E}_h) \right] & \text{on } f_m \cap S_a;\ m = 1, ..., N_f \end{cases} \tag{5.53}$$

Proceed as in Section 5.3.1 to obtain an upper bound on the approximate energy norm, exactly like equation (5.43), but with $\mathbf{R}_V$ and $\mathbf{R}_f$ as defined in this section and with $\|\mathbf{e}_h\|_{E^a(\Omega,\tau,k+1)}$ as defined in equation (5.40), but using the cavity backed aperture bilinear form of this section in the numerator.

Practically, the L^2 norms of the residuals can be calculated with the tetrahedral and triangular quadrature rules presented in Section D.1. The point-wise evaluation of $F_E^{\text{BI}}(S_a, \mathbf{E}_h)$ in the aperture is discussed in Section 3.7.3.

5.3.3 The general, dominant mode, coaxial port VBVP

In this section, the explicit, residual based bound on the approximate energy norm, as established for the general VBVP (see equation (5.43)), will be established for the VBVP that includes the general, dominant mode, coaxial port model, as presented in Section E.3.

From equations (2.12), (2.13) and (E.32), the bilinear and linear forms of the coaxial port, VBVP follow as

$$B(\mathbf{E},\mathbf{W}) = \left\langle \frac{1}{\mu_r}\nabla\times\mathbf{E}, \nabla\times\mathbf{W} \right\rangle_{\mathcal{C},\Omega} - k_0^2 \langle \epsilon_r\mathbf{E},\mathbf{W}\rangle_{\mathcal{C},\Omega} - \frac{jk_c}{\mu_{rc}}\int_{S_c}\{\hat{n}\times(\hat{n}\times\mathbf{E})\}\cdot\mathbf{W}\,dS \quad (5.54)$$

$$L(\mathbf{W}) = -jk_0Z_0\langle\mathbf{J},\mathbf{W}\rangle_{\mathcal{C},\Omega} - \frac{2jk_c}{\mu_{rc}}\int_{S_c}\left\{\hat{n}\times(\hat{n}\times\mathbf{E}_c^{\text{inc}})\right\}\cdot\mathbf{W}\,dS. \quad (5.55)$$

Construct the coaxial port, error VBVP from equation (5.6) and write it out in an element-by-element fashion, yielding

$$\begin{aligned} B(\mathbf{e}_h,\mathbf{W}) \;=\; \sum_{i=1}^{N_\tau}\Bigg(& -\frac{jk_c}{\mu_{rc}}\int_{\partial K_i\cap S_c}\left\{\hat{n}\times(\hat{n}\times 2\mathbf{E}_c^{\text{inc}}) - \hat{n}\times(\hat{n}\times\mathbf{E}_h)\right\}\cdot\mathbf{W}\,dS \\ & -\int_{K_i}\frac{1}{\mu_r}\nabla\times\mathbf{E}_h\cdot\nabla\times\mathbf{W} - k_0^2\epsilon_r\mathbf{E}_h\cdot\mathbf{W} + jk_0Z_0\mathbf{J}\cdot\mathbf{W}\,dV\Bigg). \end{aligned} \quad (5.56)$$

Apply equation (5.8) to equation (5.56) and rewrite in terms of elemental and facial contributions, to obtain

$$B(\mathbf{e}_h,\mathbf{W}) = \sum_{i=1}^{N_\tau}\left(\int_{K_i}\mathbf{R}_V\cdot\mathbf{W}\,dV\right) + \sum_{m=1}^{N_f}\left(\int_{f_m}\mathbf{R}_f\cdot\mathbf{W}\,dS\right) \quad (5.57)$$

with

$$\mathbf{R}_V = -\nabla\times\frac{1}{\mu_r}\nabla\times\mathbf{E}_h + k_0^2\epsilon_r\mathbf{E}_h - jk_0Z_0\mathbf{J} \quad \text{in } K_i;\ i=1,\ldots,N_\tau \quad (5.58)$$

$$\mathbf{R}_f = \begin{cases} \hat{n}^{(12)}\times\left[\dfrac{1}{\mu_r^{(1)}}\nabla\times\mathbf{E}_h^{(1)} - \dfrac{1}{\mu_r^{(2)}}\nabla\times\mathbf{E}_h^{(2)}\right] & \text{on } f_m\setminus S_c;\ m=1,\ldots,N_f \\ \dfrac{1}{\mu_r}\hat{n}\times\nabla\times\mathbf{E}_h - \dfrac{jk_c}{\mu_{rc}}\hat{n}\times\left[\hat{n}\times\left(2\mathbf{E}_c^{\text{inc}} - \mathbf{E}_h\right)\right] & \text{on } f_m\cap S_c;\ m=1,\ldots,N_f \end{cases} \quad (5.59)$$

Proceed as in Section 5.3.1 to obtain an upper bound on the approximate energy norm, exactly like equation (5.43), but with $\mathbf{R}_V$ and $\mathbf{R}_f$ as defined in this section and with $\|\mathbf{e}_h\|_{E^a(\Omega,\tau,k+1)}$ as defined in equation (5.40), but using the coaxial port, bilinear form of this section in the numerator.

Practically, the L^2 norms of the residuals can be calculated with the tetrahedral and triangular quadrature rules presented in Section D.1.

5.4 The implicit, element residual method

This section presents the element residual method (ERM) [4, 138, 24, 144], as applied to the 3D, vector wave equation, Galerkin FEM (Chapter 2) with tetrahedral, curl-conforming, vector elements (as defined in Appendix C) and including the cavity BI boundary condition (Chapter 3) and the dominant mode, coaxial aperture boundary condition (Section E.3). The ERM is based on solving elemental VBVPs of the error, using an approximation of the true elemental boundary conditions.

The exact version of the elemental error VBVP employed is presented first, followed by the approximation made in order to calculate the elemental boundary conditions. The technique is then extended to handle the cavity backed aperture- and general, coaxial, dominant mode port boundary conditions.

5.4.1 The exact, elemental, general VBVP of the error

The elemental VBVP of the error, as presented in equation (5.29), is employed, but with $\partial K_{iD} = \emptyset$. The local Dirichlet boundary condition is discarded in favour of the local Neumann boundary condition for various reasons. An elemental problem with Neumann boundary conditions results in more local degrees of freedom, which will (hopefully) lead to greater solution accuracy (as also noted in [170]). In the Dirichlet case, the only obvious, intuitive approximation of $\hat{n} \times \mathbf{E}$ on ∂K_{iD} is 0, else one would need to resort to $\hat{n}\times$ some smoothed version of $\mathbf{E}_h$, based on its value in the vicinity of ∂K_{iD} (see [15] for such techniques). On the other hand, approximation of the Neumann boundary condition required on ∂K_{iN}, does leave some opportunity for simple, intuitive choices that contribute more information to the local BVP – which is the essence of the element residual method. (Obviously $\frac{1}{\mu_r}\hat{n} \times \nabla \times \mathbf{E}$ on ∂K_{iN} could also be approximated by $\hat{n}\times$ some smoothed version of $\frac{1}{\mu_r}\nabla \times \mathbf{E}_h$ in the vicinity of ∂K_{iN}.) Finally, note that the local, Neumann BVP is well defined, since static field solutions are inadmissible (see [24] for comments on this issue).

The exact, elemental, general VBVP of the error, which will be solved approximately (defining the element residual method), can now be stated as

$$\begin{cases} B_i(\mathbf{e}_h, \mathbf{W}) = L_i(\mathbf{W}) - B_i(\mathbf{E}_h, \mathbf{W}) - \displaystyle\int_{\partial K_{iN}} \frac{1}{\mu_r} (\hat{n} \times \nabla \times \mathbf{E}) \cdot \mathbf{W}\, dS \\ \qquad\qquad \forall\ \mathbf{W} \in W_i\ ; \quad \mathbf{e}_h \in W_i \end{cases} \tag{5.60}$$

with ∂K_{iN} and $\partial K_i \cap (\Gamma_D \cup \Gamma_N)$ disjoint and $\partial K_i = \partial K_{iN} \cup (\partial K_i \cap (\Gamma_D \cup \Gamma_N))$.

5.4.2 The approximate, elemental, general VBVP of the error

The essence of the element residual method, as applied to the FE formulation considered here, is the intuitive approximation of the true inter-element, tangential magnetic field intensity ($\hat{n} \times \frac{1}{\mu_r}\nabla \times \mathbf{E}$). This approximation takes the form of a weighted sum of the approximate magnetic field intensities obtained within the two elements sharing the inter-element boundary, evaluated at the boundary. To this end, let $\xi^{(1)}$ and $\xi^{(2)}$ be the weighting factors associated with face f, with

$$\xi^{(1)} + \xi^{(2)} = 1, \tag{5.61}$$

then the element residual method proposes the following approximation:

$$\hat{n} \times \frac{1}{\mu_r^{(1)}}\nabla \times \mathbf{E}^{(1)} = \hat{n} \times \frac{1}{\mu_r^{(2)}}\nabla \times \mathbf{E}^{(2)} \approx \hat{n} \times \left(\frac{\xi_m^{(1)}}{\mu_r^{(1)}}\nabla \times \mathbf{E}_h^{(1)} + \frac{\xi_m^{(2)}}{\mu_r^{(2)}}\nabla \times \mathbf{E}_h^{(2)} \right)$$
$$\text{on } f_m \setminus (\Gamma_D \cup \Gamma_N);\ m = 1, ..., N_f \tag{5.62}$$

where the superscripts (1) and (2) denote association with the two elements sharing f_m, as in equation (5.14). Note that the approximation is not valid on homogeneous Dirichlet boundaries, where a surface current density $\mathbf{K}$ is generally present.

Substitute this approximation into equation (5.60) to yield the desired, approximate, solvable, elemental VBVP of the error:

$$\begin{cases} B_i(\mathbf{e}_h, \mathbf{W}) = L_i(\mathbf{W}) - B_i(\mathbf{E}_h, \mathbf{W}) \\ -\sum_{f \subseteq K_{iN}} \int_f \left(\hat{n} \times \left[\frac{\xi^{(1)}}{\mu_r^{(1)}} \nabla \times \mathbf{E}_h^{(1)} + \frac{\xi^{(2)}}{\mu_r^{(2)}} \nabla \times \mathbf{E}_h^{(2)} \right] \right) \cdot \mathbf{W} \, dS \\ \forall \ \mathbf{W} \in W_i \ ; \quad \mathbf{e}_h \in W_i \end{cases} \tag{5.63}$$

The above procedure has reduced the element residual method to a choice of the splitting factors $\xi_m^{(1)}$ and $\xi_m^{(2)}$ with $m = 1, ..., N_f$. To solve equation (5.63) in practice, the task of approximating the infinite dimensional space W_i still remains.

Supposing that $\mathbf{E}_h = \pi\mathbf{E}$, then $\pi\mathbf{e}_h = 0$. If the interpolate spans the space $U_{p_i}(K_i)$, then the approximation of W_i should be an approximation of the bubble space $V_{p_i,\infty}(K_i)$ (see Section C.1.2 for definitions of these spaces).

With the hierarchical basis functions described in Appendix C, the approximation of $V_{p_i,\infty}(K_i)$ is a straight forward procedure. Simply choose a set of basis functions that span the space $V_{p_i,q}(K_i)$, with $p_i < q$. This choice of basis functions to approximate the error is also advocated in [138, 4, 24, 111, 144].

The above discussion represents the ideal case. In general, $\mathbf{E}_h \neq \pi\mathbf{E}$, especially due to dispersion [20], therefore $\pi\mathbf{e}_h$ will not be exactly zero. In practice it has been reported that $V_{p_i,q}(K_i)$ remains a good choice for approximating W_i [144, 130], but it is important to note that $U_{p_i}(K_i)$ represents the most comprehensive choice.

Finally, one should be aware that some of the higher order basis functions may be near-orthogonal to the error, as mentioned in [4, 81], therefore choosing $q - p_i \leq 1$ may result in all additional basis function modeling the error quite badly. In general though, a *saturation assumption* is implicitly made (see [29, 24, 33] for example), i.e. it is assumed that any increase in the polynomial order of the basis functions will lead to an increase in solution accuracy.

5.4.3 The approximate, elemental, cavity VBVP of the error

Start by noting the cavity Neumann boundary condition, with the electric field, boundary integral operator as defined in equation (5.46):

$$\frac{1}{\mu_r}\mathbf{N} = \hat{n} \times \nabla \times \left[\mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a, \mathbf{E}) \right] \quad \text{on } S_a. \tag{5.64}$$

The exact, elemental VBVP is obtained by substituting this equation into equation (5.60). However, since $\mathbf{E}$ is not known on S_a, the non-local boundary condition requires further approximation. The first possible version of the element residual, approximate VBVP is obtained by simply assuming

$$F_E^{\text{BI}}(S_a, \mathbf{E}) \approx F_E^{\text{BI}}(S_a, \mathbf{E}_h) \tag{5.65}$$

(as was done in [130] in the case of a scalar, two-dimensional FE-BI formulation), yielding equation (5.63) again

$$\begin{cases} B_i(\mathbf{e}_h, \mathbf{W}) = L_i(\mathbf{W}) - B_i(\mathbf{E}_h, \mathbf{W}) \\ -\sum_{f \subseteq K_{iN}} \int_f \left(\hat{n} \times \left[\frac{\xi^{(1)}}{\mu_r^{(1)}} \nabla \times \mathbf{E}_h^{(1)} + \frac{\xi^{(2)}}{\mu_r^{(2)}} \nabla \times \mathbf{E}_h^{(2)} \right] \right) \cdot \mathbf{W} \, dS \\ \forall \, \mathbf{W} \in W_i \; ; \quad \mathbf{e}_h \in W_i \end{cases} \tag{5.66}$$

with

$$B_i(\mathbf{E}, \mathbf{W}) = \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{\mathcal{C}, K_i} - k_0^2 \left\langle \epsilon_r \mathbf{E}, \mathbf{W} \right\rangle_{\mathcal{C}, K_i} \tag{5.67}$$

$$L_i(\mathbf{W}) = -\int_{\partial K_i \cap S_a} \hat{n} \times \nabla \times \left[\mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a, \mathbf{E}_h) \right] \cdot \mathbf{W} \, dS - jk_0 Z_0 \left\langle \mathbf{J}, \mathbf{W} \right\rangle_{\mathcal{C}, K_i}. \tag{5.68}$$

A more sophisticated approximation of the double integral term in the Neumann boundary condition is possible, resulting in an improved version of the approximate problem. The approximation is as follows:

$$\begin{aligned} & -\int_{\partial K_i \cap S_a} \hat{n} \times \nabla \times F_E^{\text{BI}}(S_a, \mathbf{E}) \cdot \mathbf{W} \, dS \\ &= -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(\partial K_i \cap S_a, \mathbf{e}_h) \right\} \cdot \mathbf{W} \, dS \\ &\quad -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(\partial K_i \cap S_a, \mathbf{E}_h) \right\} \cdot \mathbf{W} \, dS \\ &\quad -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(S_a \setminus (\partial K_i \cap S_a), \mathbf{E}) \right\} \cdot \mathbf{W} \, dS \\ &\approx -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(\partial K_i \cap S_a, \mathbf{e}_h) \right\} \cdot \mathbf{W} \, dS \\ &\quad -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(\partial K_i \cap S_a, \mathbf{E}_h) \right\} \cdot \mathbf{W} \, dS \\ &\quad -\int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(S_a \setminus (\partial K_i \cap S_a), \mathbf{E}_h) \right\} \cdot \mathbf{W} \, dS. \end{aligned} \tag{5.69}$$

Thus the first assumption (5.65) can be reduced to the following assumption:

$$F_E^{\text{BI}}(S_a \setminus (\partial K_i \cap S_a), \mathbf{E})\big|_{\partial K_i \cap S_a} \approx F_E^{\text{BI}}(S_a \setminus (\partial K_i \cap S_a), \mathbf{E}_h)\big|_{\partial K_i \cap S_a}. \tag{5.70}$$

This reduced assumption results in a BI component entering into the approximate local problem. Equation (5.66) still represents the local problem, but the elemental bilinear- and linear forms must be redefined as

$$\begin{aligned} B_i(\mathbf{E}, \mathbf{W}) &= \left\langle \frac{1}{\mu_r} \nabla \times \mathbf{E}, \nabla \times \mathbf{W} \right\rangle_{\mathcal{C}, K_i} - k_0^2 \left\langle \epsilon_r \mathbf{E}, \mathbf{W} \right\rangle_{\mathcal{C}, K_i} \\ &\quad + \int_{\partial K_i \cap S_a} \left\{ \hat{z} \times \nabla \times F_E^{\text{BI}}(\partial K_i \cap S_a, \mathbf{E}) \right\} \cdot \mathbf{W} \, dS \end{aligned} \tag{5.71}$$

$$\begin{aligned} L_i(\mathbf{W}) &= -\int_{\partial K_i \cap S_a} \hat{n} \times \nabla \times \left[\mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} + F_E^{\text{BI}}(S_a \setminus (\partial K_i \cap S_a), \mathbf{E}_h) \right] \cdot \mathbf{W} \, dS \\ &\quad -jk_0 Z_0 \left\langle \mathbf{J}, \mathbf{W} \right\rangle_{\mathcal{C}, K_i}. \end{aligned} \tag{5.72}$$

The double surface integral term in the above bilinear form can be rewritten in the same way as equation (3.23), so that the same integration scheme (presented in Section D.3) can be used to calculate the BI contribution the the system matrix of the local problem.

The BI component of the above local problem is quite similar to one presented in [108], except that the integral over $S_a \setminus (\partial K_i \cap S_a)$ in the local linear form, is dropped in [108].

5.4.4 The approximate, elemental, dominant mode, general coaxial port VBVP of the error

Start with the general, dominant mode, coaxial port, Neumann boundary condition, as presented in Section E.3:

$$\frac{1}{\mu_r}\mathbf{N} = \frac{jk_c}{\mu_{rc}}\left[2\hat{n}\times(\hat{n}\times\mathbf{E}_c^{\text{inc}}) - \hat{n}\times(\hat{n}\times\mathbf{E})\right] \quad \text{on } S_c. \tag{5.73}$$

The approximate, element residual method, VBVP is obtained by substituting this equation into equation (5.63), yielding

$$\left\{\begin{array}{c} B_i(\mathbf{e}_h,\mathbf{W}) = L_i(\mathbf{W}) - B_i(\mathbf{E}_h,\mathbf{W}) \\ -\sum_{f\subseteq K_{iN}}\int_f\left(\hat{n}\times\left[\frac{\xi^{(1)}}{\mu_r^{(1)}}\nabla\times\mathbf{E}_h^{(1)} + \frac{\xi^{(2)}}{\mu_r^{(2)}}\nabla\times\mathbf{E}_h^{(2)}\right]\right)\cdot\mathbf{W}\,dS \\ \forall\ \mathbf{W}\in W_i\ ;\quad \mathbf{e}_h\in W_i \end{array}\right. \tag{5.74}$$

with

$$\begin{aligned} B_i(\mathbf{E},\mathbf{W}) &= \left\langle\frac{1}{\mu_r}\nabla\times\mathbf{E},\nabla\times\mathbf{W}\right\rangle_{\mathcal{C},K_i} - k_0^2\left\langle\epsilon_r\mathbf{E},\mathbf{W}\right\rangle_{\mathcal{C},K_i} \\ &\quad + \frac{jk_c}{\mu_{rc}}\int_{\partial K_i\cap S_c}(\hat{n}\times\mathbf{E})\cdot(\hat{n}\times\mathbf{W})\,dS \end{aligned} \tag{5.75}$$

$$L_i(\mathbf{W}) = \frac{2jk_c}{\mu_{rc}}\int_{\partial K_i\cap S_c}\left(\hat{n}\times\mathbf{E}_c^{\text{inc}}\right)\cdot(\hat{n}\times\mathbf{W})\,dS - jk_0Z_0\left\langle\mathbf{J},\mathbf{W}\right\rangle_{\mathcal{C},K_i}. \tag{5.76}$$

5.4.5 Inter-element boundary condition approximation: uniform weighting

This is the most obvious approximation of the inter-element tangential magnetic field intensity. The discontinuity in the approximate, finite element solution is simply averaged. This technique of obtaining a more accurate inter-element, tangential magnetic field intensity value is also discussed in [24, 4, 111], in terms of scalar, one- and two dimensional problems. The resulting weighting factors to be used in equation (5.63) are as follows:

$$\left.\begin{array}{l} \xi_m^{(1)} = \frac{1}{2} \\ \xi_m^{(2)} = \frac{1}{2} \end{array}\right\} \quad m = 1,\ldots,N_f. \tag{5.77}$$

5.4.6 Other inter-element boundary condition approximations

Intuitive, possibly improved choices can be made, other than the simple averaging approach, e.g. to make the relationship between the face weights correspond to the relationship between the two element volumes [144, 17], or even to weigh the real and imaginary parts separately [7].

The above 'improved' choices were not implemented, because they are fundamentally still heuristic in nature. A clear improvement would be to extend the ERM, as presented here, to incorporate local Neumann boundary condition equilibration, as presented in [4, 3, 110]. This possibility is not explored further.

5.5 Conclusion

This chapter has presented two error estimation techniques for the FE formulation considered in this document, incorporating the cavity aperture boundary condition of Chapter 3 and the dominant mode, coaxial aperture boundary condition presented in Section E.3. The two techniques have been put in perspective, within the vast literature available on a posteriori, FEM error estimation.

The first error estimation technique, an explicit residual based error estimator, was shown to be reliable in the sense that it bounds the error, measured in the approximate energy norm, in terms of a sum of local, residual contributions. In its present form, it can only be used to estimate the relative error distribution. It cannot be used to estimate the global solution quality, due to the unspecified, multiplicative constant in the global bound. Obviously, the estimator can also be formulated when both coaxial ports and the cavity aperture are present, as indeed it will be used in Chapter 6. Note that, strictly speaking, the upper bound only applies when the regularity requirements of the interpolation error estimates are met — which excludes the presence of electric field singularities. In Chapter 6 it will be demonstrated that this fact does not impact visibly on the identification of singular field regions, even thought the error is underestimated in such regions.

The second error estimation technique, an implicit residual based estimator of ERM type, calculates an explicit approximation of the error over the entire mesh. This can obviously be used on a local indication level as well as to calculate some global norm of this approximate error as an indication of the global solution quality. However, no guarantee has been provided that this approach is reliable or effective in some global sense. For example, this means that the effect of dispersion error (pollution error) is not necessarily taken into account by the estimator in its present form. Pollution error is a common occurrence with FE solutions of the Helmholz equation, as discussed in [20].

Thus, the principal difference between the two error estimation schemes presented is that the explicit scheme bounds the error as measured in the approximate energy norm, while the error itself is directly approximated by the implicit scheme. The first has the advantage that it represents a true bound on the approximate energy norm of the error (thought the approximate energy norm is not a proper norm), while the second has the advantage that the approximated error can be measured in any preferred way, but without any guarantee on its accuracy.

These two error estimators clearly do not satisfy all the ideal error estimator requirements, as listed in Section 5.1.1, but they are rigorously derived in the sense that all approximations and assumptions made, are clearly defined. This is a somewhat unique property among the current a posteriori error estimators available for the 3D, full-wave, Maxwell Galerkin FEM, especially in the engineering literature.

The main contributions of this chapter are the following: An overview and categorization of the extensive FEM, a posteriori error estimation literature, in general as well as specifically concerning Maxwell's equations was given. A thorough characterization of the error in the FE formulation considered in this document was presented. The derivation of an explicit residual based error estimator — novel to the FE formulation considered here, was derived. Its novel extensions to the cavity backed aperture- and dominant mode, coaxial port boundary conditions were made. The extension of the well-known ERM to the FE formulation considered here was presented. The novel extensions of the presented ERM to the cavity backed aperture- and dominant mode, coaxial port boundary conditions were made. The cavity backed aperture boundary condition required special attention, since another fundamental approximation had to be made (on top of the ERM approximation). Finally, a well-defined framework for further error estimation work on this particular FE formulation was established.

Possible extensions to the work presented in this chapter, are as follows: The explicit scheme could be improved by estimating the constants occurring in the upper bound, as discussed in [79, 106]. Some form of boundary condition equilibration for the ERM could possibly be devised. Galerkin orthogonality could possibly be used to optimize the ERM. The estimators could be extended to other boundary conditions, such as the free space BI boundary condition and absorbing boundary conditions. Ways could be sought to adapt the estimators, in order that the low frequency instability of the FEM formulation is also assessed.

In the next chapter, it will be shown how to use the error estimators constructed here, in order to drive adaptive refinements to the FE discretization. This will give a concrete, practical perspective on the somewhat abstract concepts of the current chapter.

Chapter 6

p Adaptation: results

6.1 Introduction

This chapter presents adaptive results based on the a posteriori error estimators presented in Chapter 5. The adaptive procedures themselves are straightforward, with the aim being to show the potential of the error estimators and upgrading schemes, rather that to present a complete, robust adaptive algorithm.

Both p adaptation (upgrading elemental polynomial orders) and/or h adaptation (upgrading the mesh size) could in principle be used to demonstrate the application of the error estimators within an adaptive setting. However, only p adaptation is considered here for a number of reasons. Firstly, using only one type of adaptation at a time, demonstrates more clearly the relative performance of the respective error estimators. Secondly, the hierarchical elements available within the code are ideally suited to p adaptation. The third reason is a practical issue, which is a consequence of the commercial meshing program that is used here (FEMAP, see [183]): an external program cannot set the meshing parameters and retrieve a new mesh, making h adaptation very difficult within the current version of this FEM implementation.

Ideally, one would like to experimentally evaluate the performance of an error estimation procedure by considering problems with analytical solutions, i.e. through calculation of the *effectivity index*, as in [24, 2] for example. Unfortunately, very few such problems are available for the type of structures considered here, except for the trivial coaxial through, which has a highly regular solution in any case. Therefore, error estimation results will be evaluated by comparison with higher order solutions, comparison with randomly upgraded discretizations, by intuitive reasoning or by comparing adaptive results from different error estimators.

Three different p upgrading schemes are proposed. These are evaluated by comparing adaptive results based on the same error estimation procedure. Apart from this, the adaptive algorithms are not central to this chapter.

In the literature, adaptive results nearly always accompany the presentation of an error estimation technique (see the references in Section 5.1), but it is not always emphasized how crucial the adaptive algorithm itself is to the effective application of an a posteriori error estimator [106]. See [153, 207, 80, 141, 81] for some ideas concerning effective adaptive algorithms. Note that an optimal adaptive rate of convergence can in general only be achieved

via *hp* adaptation [207, 155], [167, Sec. 2.5.3], especially when the solution field is irregular (i.e. singularities in the field or its derivatives).

An outline of this chapter is as follows: Section 6.2 discusses the adaptive procedure employed, presenting the error indicators and the refinement strategies. Section 6.3 presents adaptive results for the stepped cavity patch antenna of Figure 3.7. The performances of the explicit- and implicit estimators are compared as well as two different upgrading schemes. Results of the same form are presented in Section 6.4, but for a waveguide filter. The waveguide port formulation is briefly discussed. Section 6.5 considers the effect of relative volume and face weighting on the explicit residual estimator of Section 5.3. A representative set of three simple waveguide problems are used for experimental purposes.

6.2 Adaptive procedure

Single-level p adaptation is employed. An initial solution is used to calculate elemental error indicators (positive real values). Based on these error indicators, the polynomial orders of certain elements are increased, called the *refinement strategy*. (A multilevel adaptive scheme would involve more than one simulation and refinement cycles.) In this section, the error indicators are first discussed, followed by the refinement strategy.

6.2.1 Error indicators

Three different error indicators are considered. Two are based on the explicit error estimation scheme presented in Section 5.3 and one on the implicit error estimation scheme presented in Section 5.4. The indicators η_i, associated with elemental volumes K_i, $i = 1, .., N_\tau$, are as follows:

- This indicator is based on the explicit scheme. Rewrite equation (5.43), with $k = 0$, as

$$\|\mathbf{e}_h\|^2_{E^a(\Omega,\tau,1)} \leq C \sum_{i=1}^{N_\tau} \left(h_i^2 \|\mathbf{R}_V\|^2_{L^2(K_i)} + 0.5 \sum_{f \subset \partial K_i} h_f \|\mathbf{R}_f\|^2_{L^2(f)} \right), \tag{6.1}$$

  with the face and volume residuals defined in Section 5.3. It is assumed that the facial contributions are shared equally between elements. The boundary face contributions are also scaled by 0.5 even though they are not shared, since they represent the same Maxwell continuity condition as the internal face residuals and should therefore be treated in the same way. This equation implies an elemental error indicator:

$$\eta_i = h_i^2 \|\mathbf{R}_V\|^2_{L^2(K_i)} + 0.5 \sum_{f \subset \partial K_i} h_f \|\mathbf{R}_f\|^2_{L^2(f)}. \tag{6.2}$$

- This indicator is a generalization of the first. A more detailed version of equation (6.1) is

$$\|\mathbf{e}_h\|^2_{E^a(\Omega,\tau,1)} \leq \sum_{i=1}^{N_\tau} \left(C_V h_i^2 \|\mathbf{R}_V\|^2_{L^2(K_i)} + 0.5 C_f \sum_{f \subset \partial K_i} h_f \|\mathbf{R}_f\|^2_{L^2(f)} \right), \tag{6.3}$$

where the constant $C = \max(C_V, C_f)$ was split. Since C_V and C_f emanate from different interpolation error estimates, the above equation is indeed a more detailed version of equation (6.1). Now re-introduce a generic constant C, as well as the constant α, in order to rewrite the above equation as

$$\|\mathbf{e}_h\|^2_{E^a(\Omega,\tau,1)} \leq C \sum_{i=1}^{N_\tau} \left(\alpha h_i^2 \|\mathbf{R}_V\|^2_{L^2(K_i)} + 0.5(1-\alpha) \sum_{f \subset \partial K_i} h_f \|\mathbf{R}_f\|^2_{L^2(f)} \right), \tag{6.4}$$

with

$$0 \leq \alpha \leq 1. \tag{6.5}$$

Equation (6.4) implies an elemental error indicator:

$$\eta_i = \alpha h_i^2 \|\mathbf{R}_V\|^2_{L^2(K_i)} + 0.5(1-\alpha) \sum_{f \subset \partial K_i} h_f \|\mathbf{R}_f\|^2_{L^2(f)}. \tag{6.6}$$

- This indicator is based on the implicit scheme, as presented in Section 5.4. Suppose $\mathbf{e}_h^a$ is the approximate solution to the approximate elemental error VBVPs. Measuring $\mathbf{e}_h^a$ on an element-wise basis in the L^2 norm yields an elemental error indicator:

$$\eta_i = \|\mathbf{e}_h^a\|^2_{L^2(K_i)}. \tag{6.7}$$

6.2.2 Refinement strategy

The refinement strategy consists of two steps: first a group of elements is identified for p upgrading (the *marking* stage), second, new polynomial orders must be assigned to these elements.

The first step is straight forward. The elements are sorted according to their error indicator values. A percentage of elements with the largest error indicators are marked for p upgrading.

The second step presents some options. The following p upgrading schemes are considered:

- *Equal* upgrading: All the marked elements are upgraded to the same higher polynomial order. In the case of an initial CT/LN solution, this implies upgrading to either LT/LN, LT/QN or QT/QN.
- *Graded* upgrading: Suppose there are available n higher orders of elements than that of the initial solution (assumed uniform with respect to element order). Split the marked elements into n equal sized groups according to the value of their error indicators. The group with the largest error indicators are upgraded to the highest element order, the group with the second largest error indicators are upgraded to the second highest element order and so forth. For example, the elements marked for refinement after an initial CT/LN solution will be split into 3 groups, those with the largest error indicators will be upgraded to QT/QN, the middle group will be upgraded to LT/QN and the group with the smallest error indicators will be upgraded to LT/LN.
- *Model based* upgrading: In regions where the electric field can be considered quasi-static, it is important to use elements that model the gradient space $\nabla\psi$ as well as

possible, since the approximate representation $\mathbf{E} \approx \nabla\psi$ holds in such regions [93]. Considering the basis functions available here, this implies using full order elements. It is further assumed that full order elements do not represent an advantage in regions where $\mathbf{E}$ and $\nabla \times \mathbf{E}$ are of equal importance, i.e. non-static fields (see Appendix C).

Therefore, to intelligently upgrade to mixed- or full order elements, calculate the ratio between the $\nabla \times \mathbf{E}$ and $\mathbf{E}$ terms in the stationary functional (see Section 2.3), for every marked element:

$$\beta_i = \frac{1}{|\mu_r \epsilon_r| k_0^2} \cdot \frac{\|\nabla \times \mathbf{E}_h\|_{L^2(K_i)}}{\|\mathbf{E}_h\|_{L^2(K_i)}}. \tag{6.8}$$

If β_i is below a certain threshold, the element is upgraded to a full order, otherwise to a mixed order. For example, based on an initial CT/LN solution, the above criterion can be used to decide between upgrading to QT/QN or LT/QN elements.

In [197] an indicator is proposed, where the errors in the gradient- and rotational function components of $\mathbf{E}_h$ are estimated separately in an intuitive manner. Here, the estimation of the quasi-static nature of the field is directly based on the same argument used to motivate the use of mixed order elements.

Finally, note that when an element is upgraded, the adjacent elements obviously need to share some of the additional degrees of freedom in order to maintain the curl-conformity of the approximation.

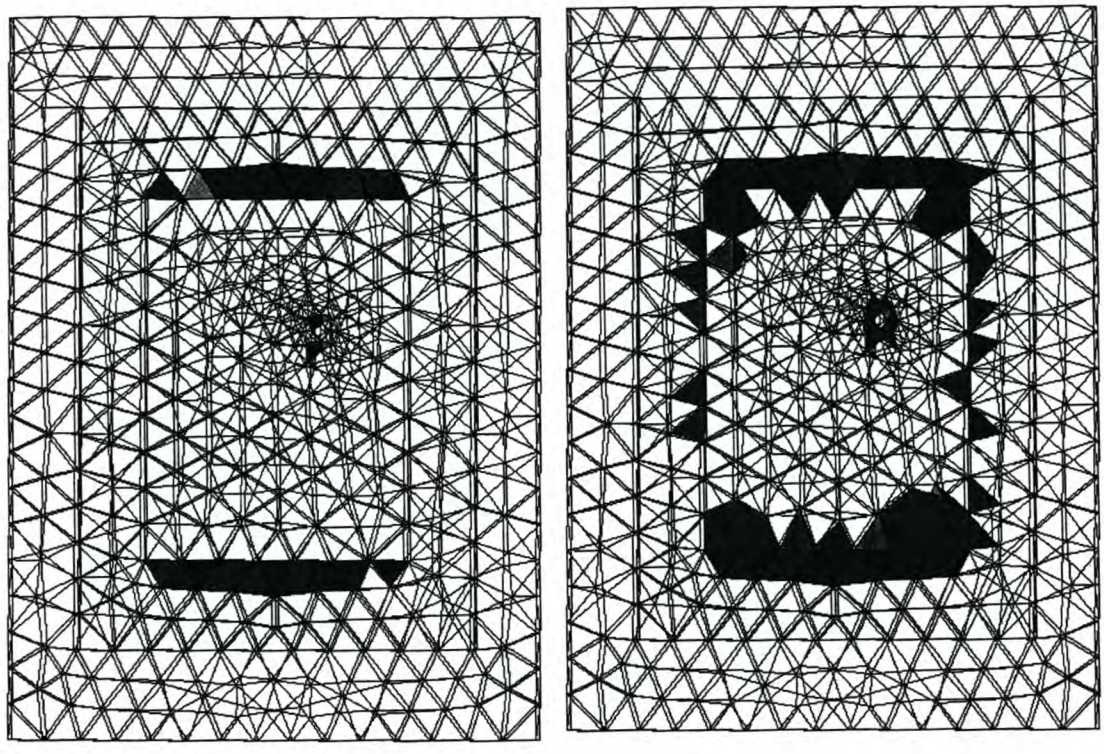
6.3 Patch antenna in a stepped cavity

This section presents estimated error distributions and adaptive results for the problem of a patch antenna in a stepped cavity, as already considered with uniform element orders in Section 3.8. The geometry and mesh are shown in Figure 3.7. The uniform element order results are shown in Figure 3.9. The indicators of equations (6.2) and (6.7) are used with equal- and model based upgrades.

Figures 6.1 and 6.2 present the error distributions as estimated by the explicit- and implicit error indicators of equations (6.2) and (6.7). In both cases the estimation is based on an initial CT/LN solution at the uniform CT/LN resonant frequency of 2.78 GHz. Both techniques target the patch edges, where the electric field is known to be singular and to change direction rapidly — in other words: where one would expect large errors to occur. The explicit scheme targets the two strongly radiating edges more intensely than the implicit scheme. The implicit scheme indicates the error to be more spread out than the explicit scheme. Concerning the implicit scheme, the elemental VBVPs were solved on the spaces $U_2(K_i)$, $i = 1, .., N_\tau$.

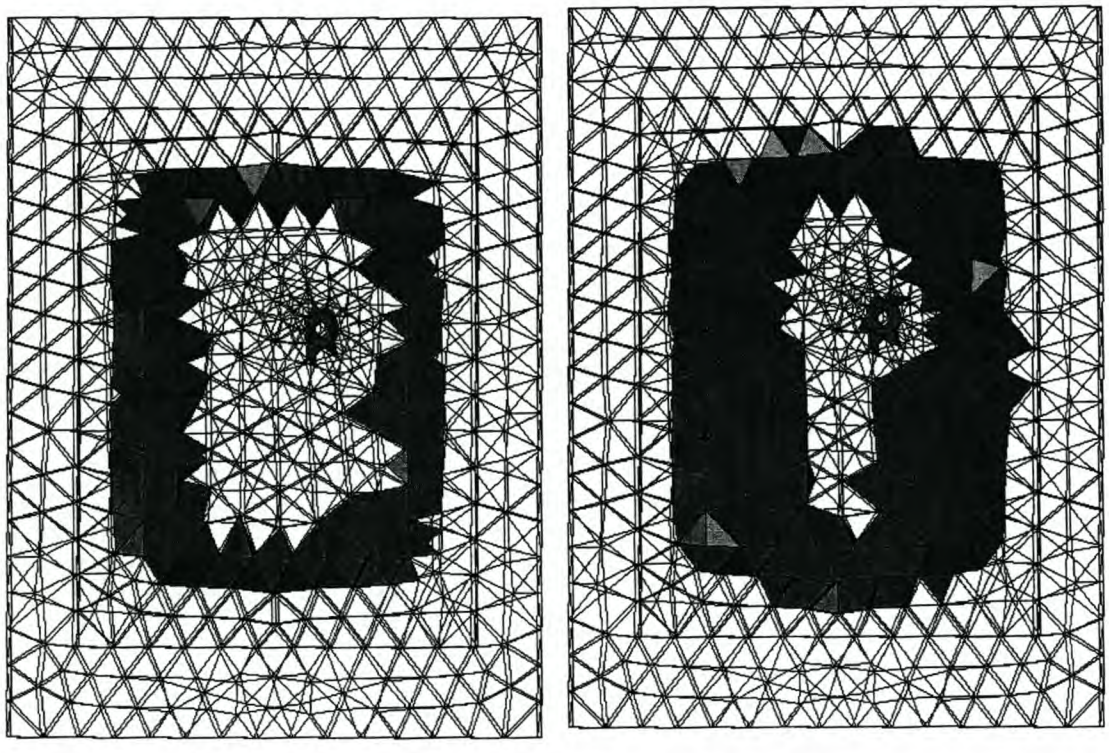
Figure 6.3 compares the performance of the two indicators in a single-level adaptive scheme with equal upgrading to QT/QN elements. Only upgrading to QT/QN elements is considered, because these elements lead to the most accurate result in the uniform element order case. The error in resonant frequency is used to measure the performance of an upgraded discretization. A curve results from using different percentages of marked elements. Clearly, the explicit indicator is superior in this case.

The model based upgrading strategy is tested by comparing the explicit indicator's curve in Figure 6.3 (equal upgrading to QT/QN) with another curve obtained with the same indicator, but using model based upgrading to LT/QN- or QT/QN elements. Figure 6.4 shows the result. The threshold for full order upgrading was set to $\beta_i < 0.01$ in the model based case. This value resulted in a nearly equal splitting of full- and mixed order upgrades with 20% marked elements. In this case, the performance of the model based upgrading scheme is superior to the equal upgrading scheme.



(a) 2.5% largest error.

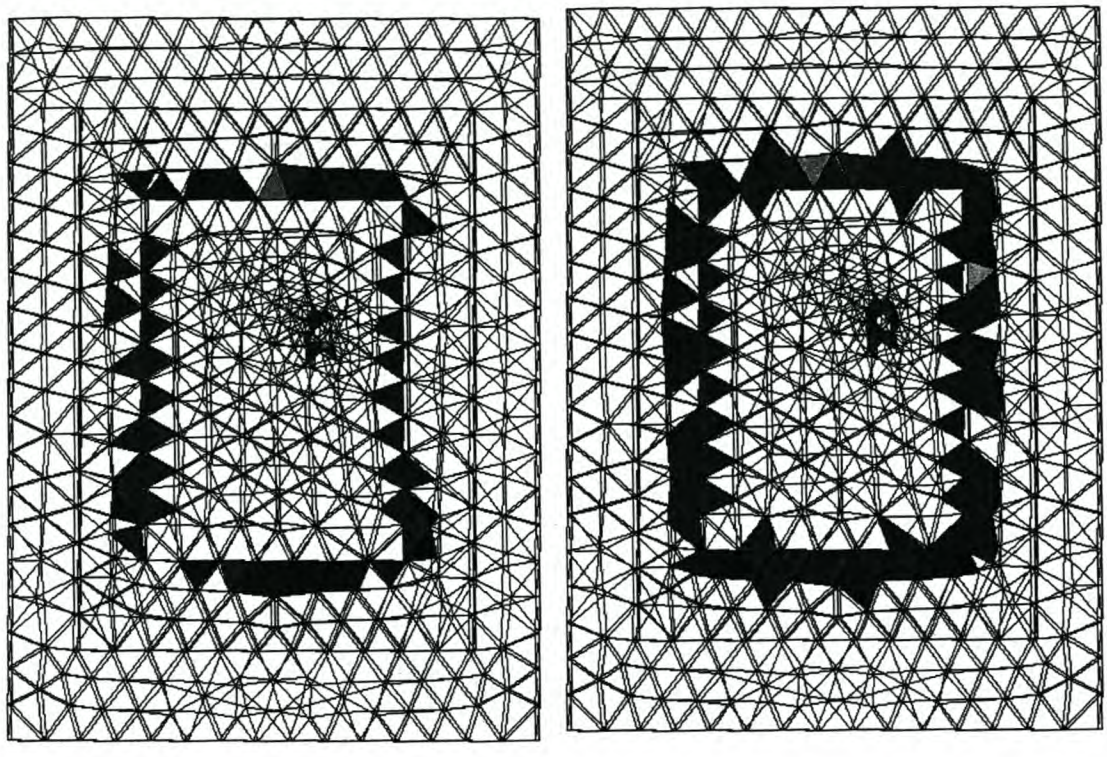
(b) 5.0% largest error.



(c) 10.0% largest error.

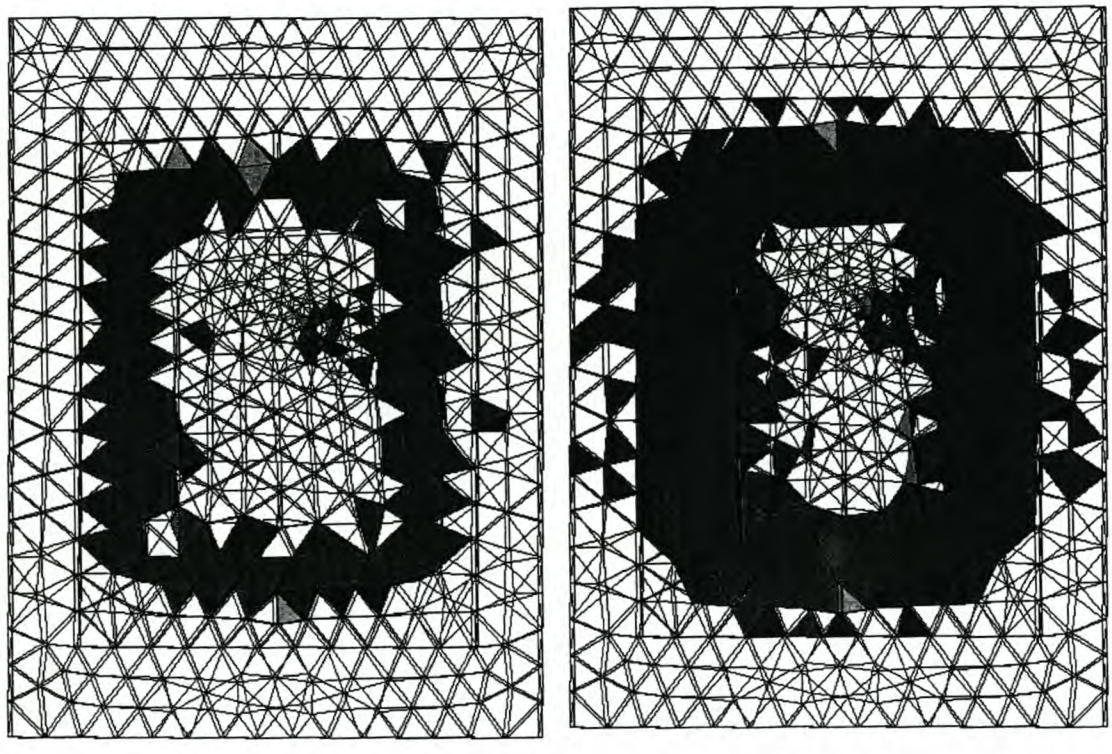
(d) 20.0% largest error.

Figure 6.1: Error distribution obtained with the explicit error indicator of equation (6.2), based on an initial CT/LN solution of the stepped cavity problem in Figure 3.7.



(a) 2.5% largest error.

(b) 5.0% largest error.



(c) 10.0% largest error.

(d) 20.0% largest error.

Figure 6.2: Error distribution obtained with the implicit error indicator of equation (6.7), based on an initial CT/LN solution of the stepped cavity problem in Figure 3.7.

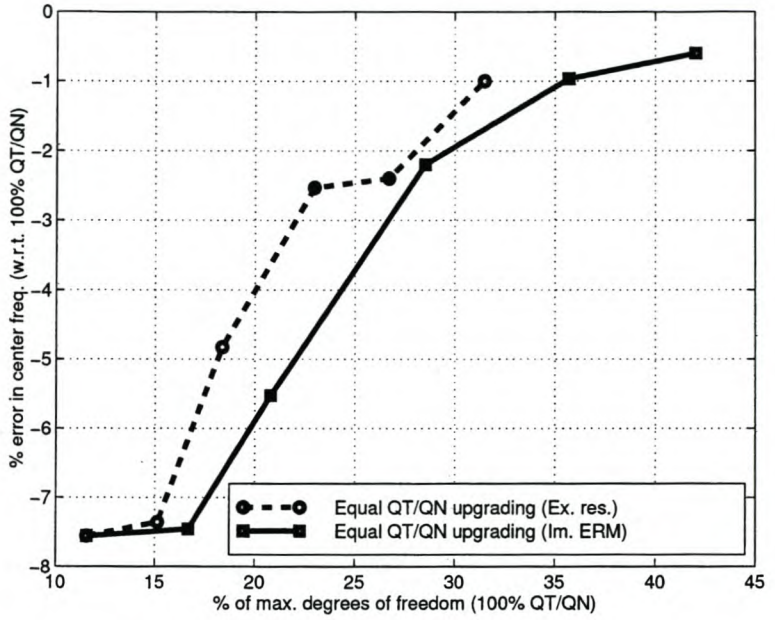


Figure 6.3: Center frequency error versus degrees of freedom. Equal upgrading to QT/QN elements. Comparison between the explicit- and implicit indicators of equations (6.2) and (6.7). Based on an initial CT/LN solution of the stepped cavity problem in Figure 3.7.

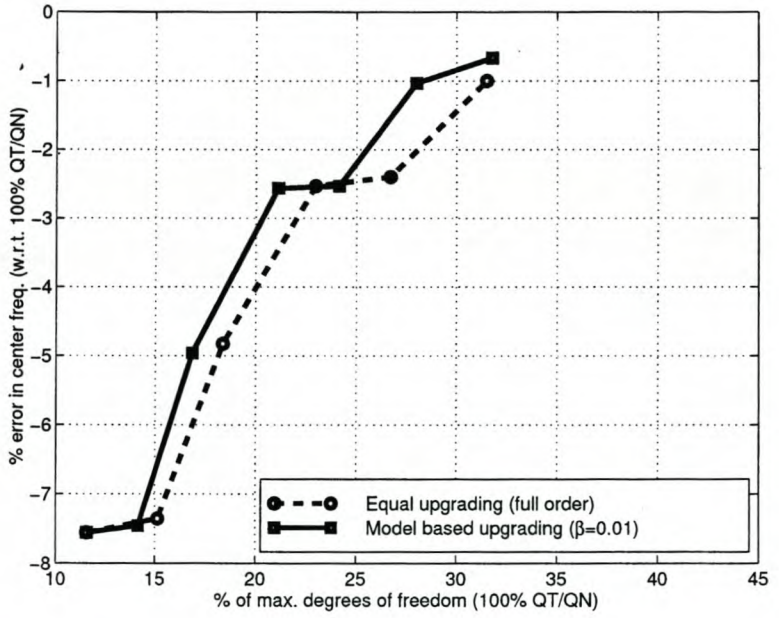


Figure 6.4: Center frequency error versus degrees of freedom. Comparison between equal- and model based upgrading to QT/QN elements, using the explicit indicator of equation (6.2). Based on an initial CT/LN solution of the stepped cavity problem in Figure 3.7.

6.4 Waveguide filter

This section presents estimated error distributions and adaptive results for the problem of a square waveguide filter and is partly based on [42]. The indicators of equations (6.2) and (6.7) are used with equal- and graded upgrades. A dominant mode (TE_{10}), rectangular waveguide port, Neumann boundary condition [99, 159, 64] is employed to incorporate the ports within the general VBVP described in Chapter 2. The port formulation is an exact analogy of the coaxial port formulation presented in Section E.3 and will not be discussed further. The implementation of the rectangular waveguide port formulation (without error estimation), within the computer code used here, is discussed in [64].

The residuals necessary for calculating the explicit error indicator of equation (6.2) are exactly the same as in Section 5.3.3, but with $\mathbf{E}_c^{\text{inc}}$, μ_{rc}, k_c and S_c replaced with $\mathbf{E}_w^{\text{inc}}$, μ_{rw}, k_w and S_w. The ERM VBVP necessary for calculating the implicit indicator of equation (6.7) is exactly the same as in Section 5.4.4, but also with these replacements.

S_w indicates the waveguide port aperture. $\mathbf{E}_w^{\text{inc}}$ and k_w represent the incident, TE_{10} wave at the port and the feeding waveguide, TE_{10} mode propagation constant, respectively. They are defined in terms of the local port coordinate system shown in Figure 6.5, as follows [154]:

$$\mathbf{E}_w^{\text{inc}} = E_w^{\text{inc}} \sin\left(\frac{\pi x}{a}\right) \hat{y} \tag{6.9}$$

$$k_w = \sqrt{\epsilon_{rw}\mu_{rw}k_0^2 - \left(\frac{\pi}{a}\right)^2}. \tag{6.10}$$

Figure 6.6 shows the waveguide filter geometry [131]. The filter consists of an X-band waveguide (aperture dimensions 22.86mm $\times$ 10.16mm) with three metallic septa along its center and normal to the broad walls. The septa are 0.2mm thick and spaced 12.148mm apart. Their respective lengths are 6.556mm, 16.788mm and 6.556mm. Figure 6.6 also shows the tetrahedral finite element mesh. Within the simulated geometry, lengths (22.902mm each) of empty waveguide are added on either side of the filter to ensure that only the propagating, TE_{10} mode is present at the ports, as required by the formulation. (At such a distance, any higher order, non-propagating modes will have attenuated sufficiently within the frequency range of interest [154].) The septa are modeled as infinitely thin (without loss of accuracy – this is true from a design viewpoint and has been verified experimentally).

Figure 6.7 compares the uniform element order results with a mesurement from [131]. Since the electric field does not dominate the problem (it is non-quasi-static and no electric field singularities are present), the mixed- and full order elements perform similarly, except that the LT/LN result is worse than the CT/LN result, which can again be attributed to the strong imbalance between the modeling of $\mathbf{E}$ and its curl in the LT/LN case (see Appendix C). The LT/QN and QT/QN solutions are practically identical.

Figures 6.8 and 6.9 present the error distributions as estimated by the explicit- and implicit error indicators of equations (6.2) and (6.7). Concerning the implicit scheme, the elemental VBVPs were solved on the spaces $U_2(K_i)$, $i = 1, .., N_\tau$. The error indication results were computed at $f = 10.867$ GHz, the first null of the CT/LN solution. Using the center of the band of the CT/LN solution produced inferior refined results. Note how the elements start by clustering around the spaces between the septa, where one would expect the greatest variation in the true field to take place and therefore the worst FE approximation. As in

Section 6.3, the implicit scheme indicates the error to be more spread out than the explicit scheme.

Figure 6.10 compares the performance of the two indicators in a single-level adaptive scheme with equal upgrading to LT/QN elements. Only upgrading to LT/QN elements is considered, because these elements, together with the QT/QN elements, lead to the most accurate results in the uniform element order case, but with less degrees of freedom in the LT/QN case. The error in resonant frequency is again used to measure the performance of an upgraded discretization. The explicit indicator is also superior in this case.

Figure 6.11 compares the performance of the graded upgrading strategy with the equal QT/QN upgrading strategy, driven by the explicit indicator. The graded scheme fails completely, which can be attributed to the presence of the LT/LN elements (see the uniform element order results in Figure 6.7).

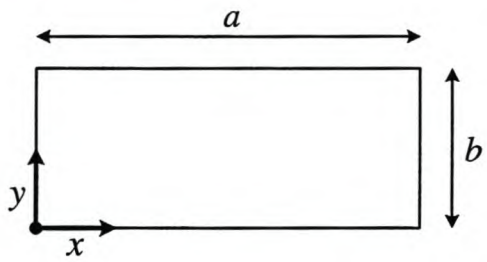
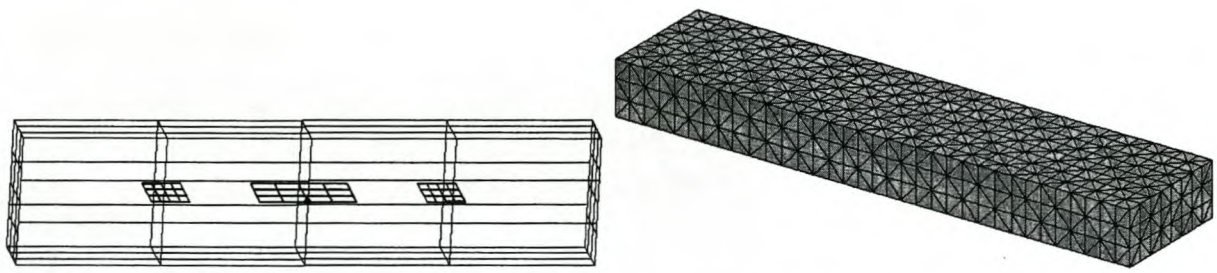


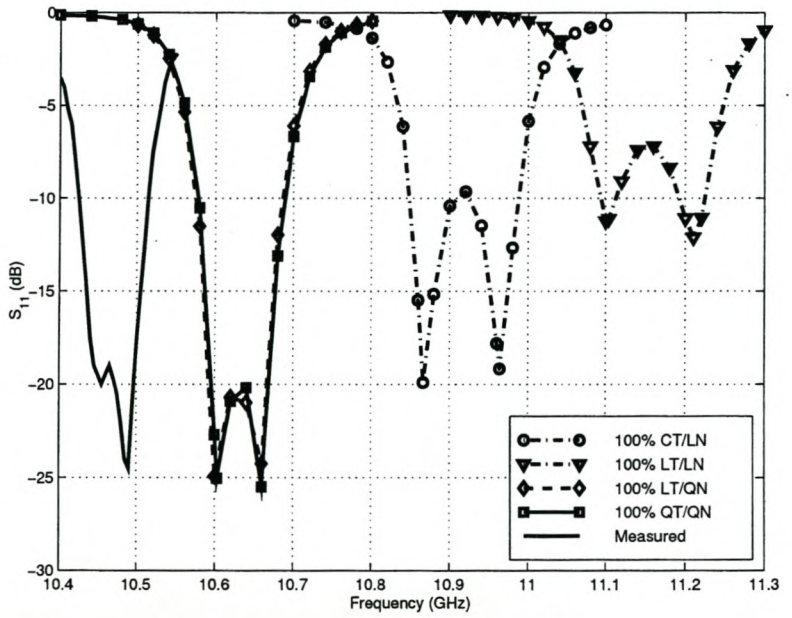
Figure 6.5: Rectangular waveguide aperture. Definitions of the local coordinate system and dimensions. $a \leq b$.



(a) Geometry.

(b) 4968 elements. 3.74mm average edge length.

Figure 6.6: The waveguide filter geometry and mesh.

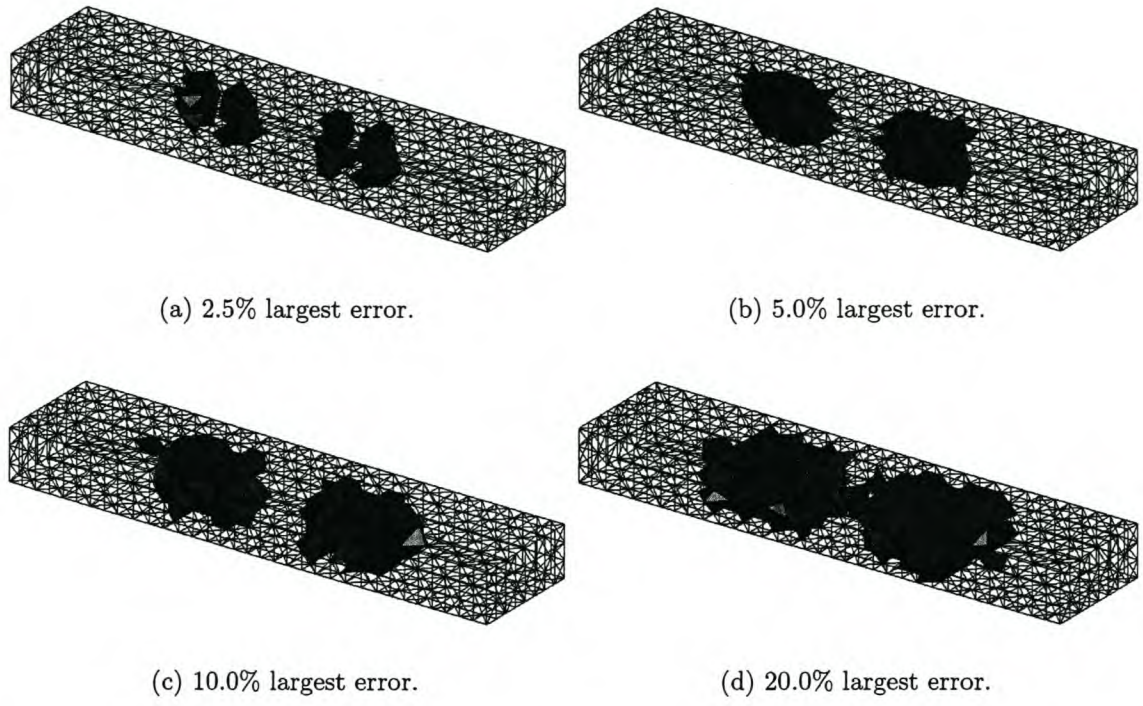


(a) s_{11} versus frequency.

Order	D.o.f.
CT/LN	4666
LT/LN	9332
LT/QN	27684
QT/QN	41526

(b) Degrees of freedom.

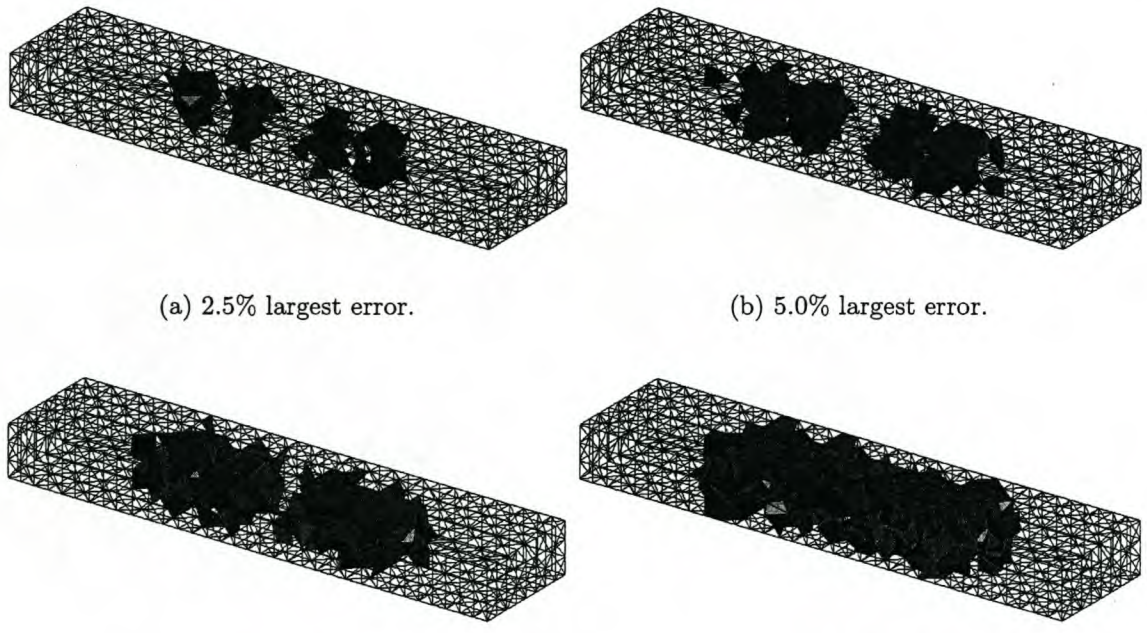
Figure 6.7: s_{11} versus frequency. Uniform element order results for the waveguide filter shown in Figure 6.6.



(a) 2.5% largest error. (b) 5.0% largest error.

(c) 10.0% largest error. (d) 20.0% largest error.

Figure 6.8: Error distribution obtained with the explicit error indicator of equation (6.2), based on an initial CT/LN solution of the waveguide filter problem in Figure 6.6.



(a) 2.5% largest error. (b) 5.0% largest error.

(c) 10.0% largest error. (d) 20.0% largest error.

Figure 6.9: Error distribution obtained with the implicit error indicator of equation (6.7), based on an initial CT/LN solution of the waveguide filter problem in Figure 6.6.

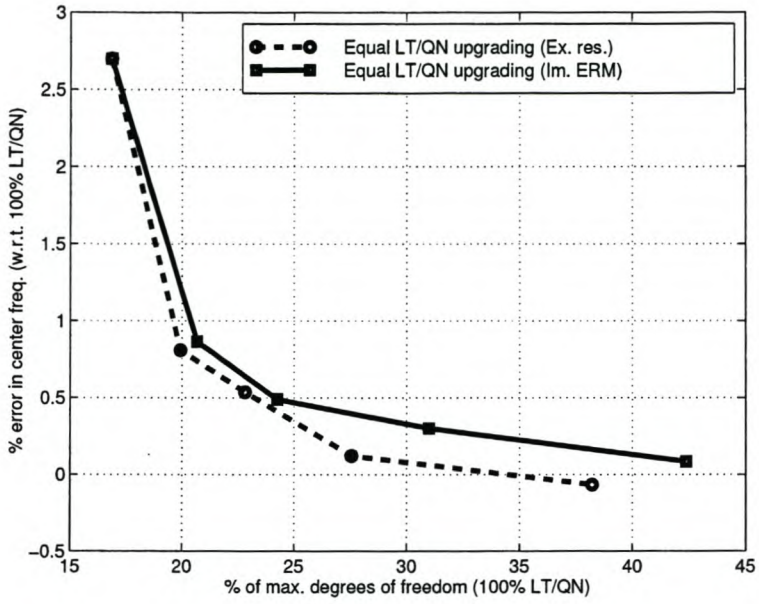


Figure 6.10: Center frequency error versus degrees of freedom. Equal upgrading to LT/QN elements. Comparison between the explicit- and implicit indicators of equations (6.2) and (6.7). Based on an initial CT/LN solution of the waveguide filter problem in Figure 6.6.

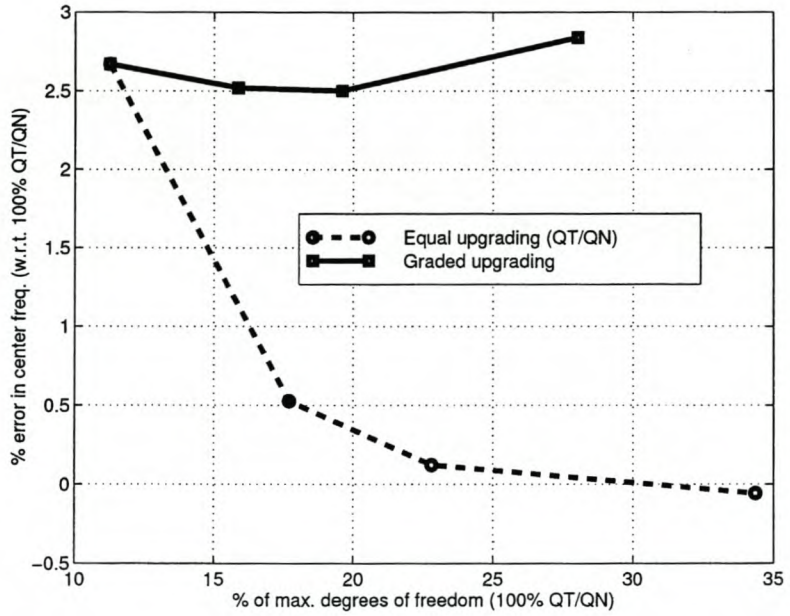


Figure 6.11: Center frequency error versus degrees of freedom. Comparison between equal- and graded upgrading to QT/QN elements, using the explicit indicator of equation (6.2). Based on an initial CT/LN solution of the waveguide filter problem in Figure 6.6.

6.5 Simple waveguide problems

This section is based on [41]. This section considers the explicit error indicator of equation (6.6). The relative importance of the volume and face residuals to the indicator is evaluated experimentally. Three simple square waveguide geometries are considered: a through problem (regular solution), a waveguide iris problem (strong electric field singularity) and an E-plane bend (weaker electric field singularity). The investigative procedure is discussed first, followed by the presentation and discussion of the results for the three geometries.

6.5.1 Investigative procedure

In equation (6.6), α clearly represents the relative contributions of the volume- and facial residuals to the elemental indicators. This section describes a procedure for evaluating the effect of α on the performance of the error indicator for a specific problem, at a specific frequency.

After a uniformly CT/LN solution, the elemental indicators are evaluated with fixed α. (Section 6.4 discusses the evaluation of the residuals for the waveguide VBVP.) Based on the indicator values, elements are marked, equally upgraded to LT/QN and the problem resolved with the upgraded discretization. Since the quality of the upgraded solution must lie between that the uniform CT/LN- and LT/QN solutions, the *relative solution quality error* ϵ_Q, measured in terms of the calculated reflection coefficient S_{11}, is defined as follows:

$$\epsilon_Q = \left| \frac{S_{11} - S_{11}^{\mathrm{LT/QN}}}{S_{11}^{\mathrm{LT/QN}}} \right|. \tag{6.11}$$

Various ϵ_Q values are obtained for the current value of α, by changing the marking percentage. The following set of marking percentages were used to generate all graphs: 0.0%, 2.5%, 5.0%, 7.5%, 10.0%, 12.5% and 100.0%. This defines a curve of ϵ_Q as a function of the number of degrees of freedom. A set of such curves is generated at a given frequency point by considering a range of α values and will henceforth be referred to as a *performance graph*. On every performance graph a curve denoted "Random" is included for reference purposes. These curves were generated by random marking. Considering a specific problem, a distinct performance graph can be generated at any frequency by the above procedure. The performance graphs are functions of the number of degrees of freedom (rather than the upgrade percentage), since it is a better measure of relative computational effort.

For example, consider the performance graph of a waveguide through problem at $f = 8.5$GHz in Figure 6.14. The first cluster of data points, around 1500 degrees of freedom, represents a marking percentage of 2.5%. Following clusters represent the other upgrade percentages used. These clusters can be quite spread out, since the upgrading of two neighbouring elements results in fewer additional degrees of freedom than the upgrading of two free-standing elements (see Section 6.2), therefore the resulting number of degrees of freedom is indicator dependent.

For every problem, performance graphs at $f = 8.5$ GHz, $f = 9.5$ GHz and $f = 10.5$ GHz are presented.

6.5.2 Waveguide through problem

This section considers a waveguide through problem. The geometry of the problem is a straight, empty length of standard X-band waveguide. Figure 6.12 shows the finite element mesh. Figure 6.13 compares the reflection coefficient values obtained with uniform CT/LN- and uniform LT/QN elements, with the analytical solution, showing that the LT/QN result is indeed an improvement upon the CT/LN result.

Concerning the performance graphs, the solution quality error was not divided by $\left|S_{11}^{\mathrm{LT/QN}}\right|$, because the true reflection coefficient is zero. Figure 6.14 shows the performance graphs.

There seems to be no consistent tendency in the performance graphs. The error indicator performance is generally poor. The following reason for this behaviour is proposed: The actual field possesses no variation in amplitude along the guide length, only a sinusoidal variation in phase. In the transverse plane there is only a sinusoidal amplitude variation in the local x-direction (see Figure 6.5). Since the actual field variations are clearly very slow and uniform throughout the whole structure, the actual error distribution is relatively flat, compared to the other problems considered in this paper. Thus, one actually needs to upgrade the mesh uniformly, rather than selectively, for optimal error reduction.

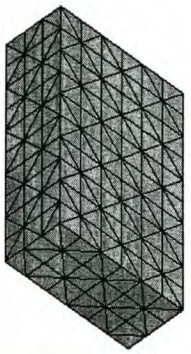


Figure 6.12: Waveguide through mesh. 1194 elements. 4.5mm average edge length. The ports are transverse to the longest dimension of the structure. The transverse waveguide geometry is as shown in Figure 6.5, with $a = 22.86$mm and $b = 10.16$mm.

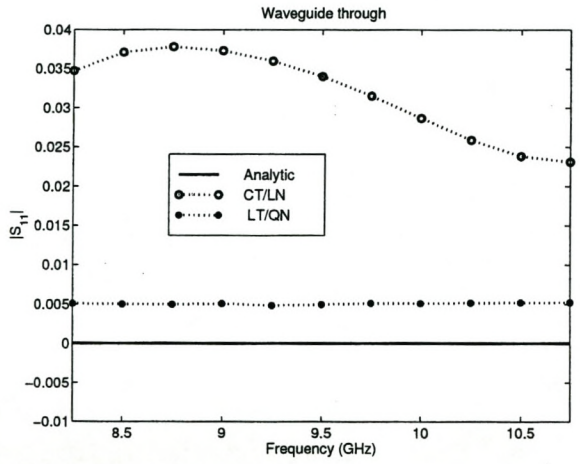
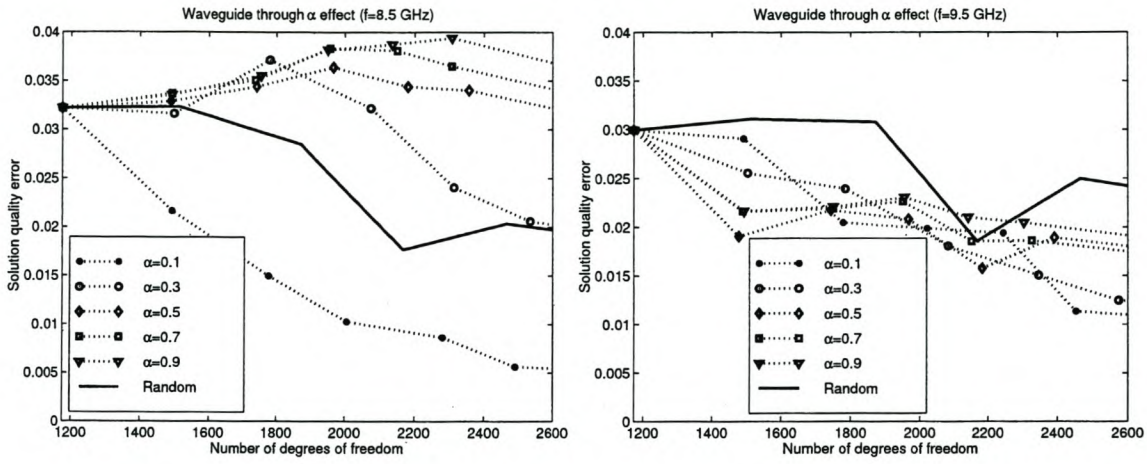
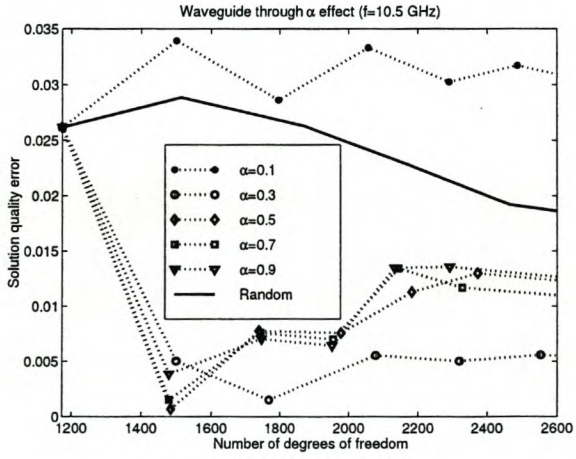


Figure 6.13: S_{11} vs. frequency for the waveguide through problem of Figure 6.12.



(a) $f = 8.5$ GHz.

(b) $f = 9.5$ GHz.



(c) $f = 10.5$ GHz.

Figure 6.14: Performance graphs for the waveguide through problem of Figure 6.12. 6836 uniform LT/QN degrees of freedom.

6.5.3 Waveguide iris problem

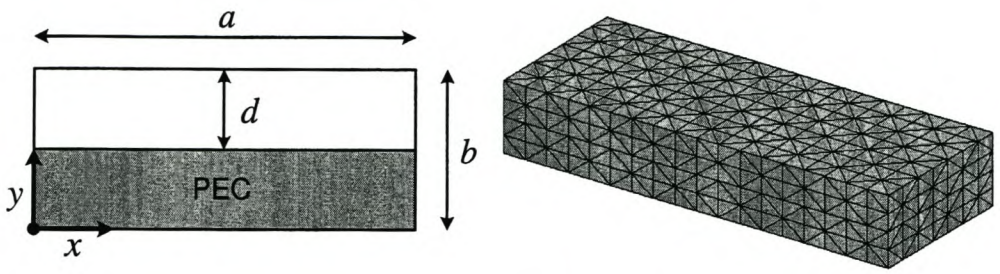
This section considers a waveguide iris problem. The geometry of the problem is a straight, empty length of X-band waveguide, except for an infinitely thin PEC iris located at its center. Figure 6.15 shows the iris geometry and the FE mesh. Figure 6.16 compares the reflection coefficient values obtained with uniform CT/LN- and LT/QN elements, with an approximate, analytical result by Marcuvitz [125], showing that the LT/QN result is indeed an improvement upon the CT/LN result. Marcuvitz's results are lumped-element circuit models; in [64] the procedure required to obtain s-parameters from these is outlined.

Figure 6.17 shows the performance graphs. Observe the following tendency in the performance graphs: when considering only a small increase in the number of degrees of freedom ($\leq 2.5\%$ upgraded elements), a dominant surface contribution leads to superior results ($\alpha < 0.5$), but if one intends to upgrade $\geq 5\%$ of the elements, a value of $\alpha \geq 0.5$ seems to be required.

A possible explanation for this tendency, which is also confirmed by inspection of the geometric distribution of the volume and face residual values, is as follows: When a small enough number of elements are to be upgraded, exclusive use of the face residuals leads to the best results, because they are most effective in identifying the elements along the iris edge, where one would expect the greatest error in the approximate field representation to occur. It is well known that the electric field strength at such a re-entrant corner is singular and changes direction extremely rapidly in its vicinity [195]. The elements are of finite size and the polynomial orders of the basis functions are also finite, thus large inter-element discontinuities will be present as a matter of course. Away from the singularity, the variation in the true field is less intense and the volume residuals overshadow the face residuals in importance.

Figure 6.18 shows the 2.5% of elements with the largest error indicator values at $f = 9.5\text{GHz}$, as identified by the $\alpha = 0.1$ and $\alpha = 0.9$ indicators, respectively. Comparison of these two distributions clearly shows the initial, superior capability of the $\alpha = 0.1$ indicator in identifying the elements along the iris edge in the middle of the waveguide.

From the performance graphs it can be seen via inspection that $\alpha = 0.5$ leads to the best all-round results for the waveguide iris problem. $\alpha = 0.5$ causes the relative solution quality error to decrease at a near optimal initial gradient in two out of three cases and leads to optimal relative solution quality error values at the highest upgrade percentage (12.5%) in all three cases.



(a) $a = 22.86$mm, $b = 10.16$mm and $d = 5.08$mm.

(b) 1889 elements. 4.3mm average edge length.

Figure 6.15: Waveguide iris problem. Iris geometry and FE mesh.

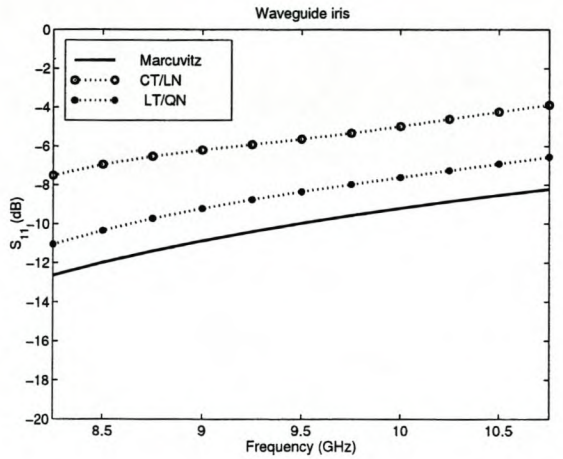
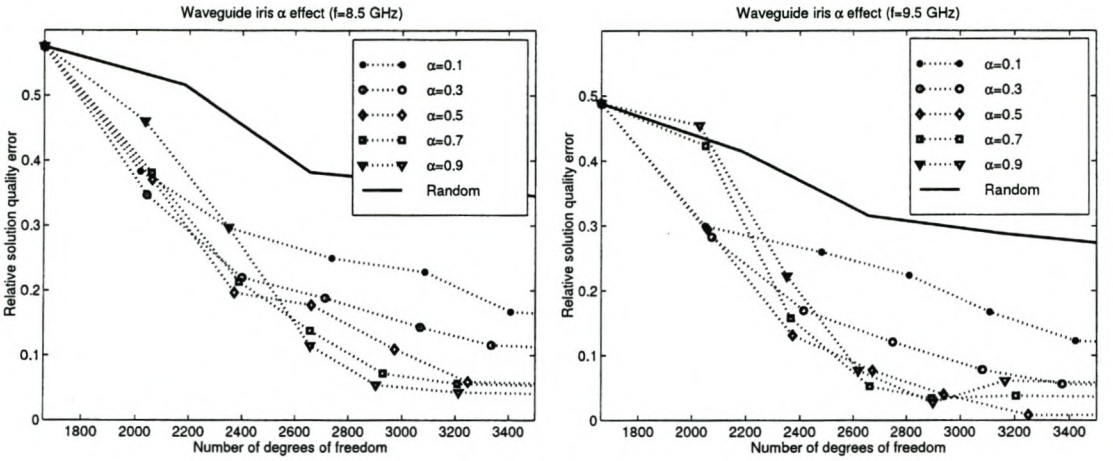
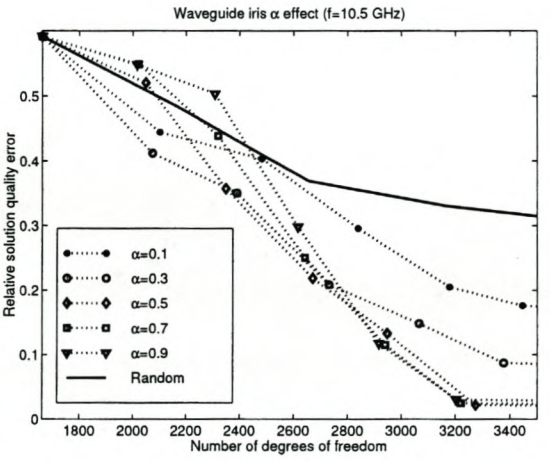


Figure 6.16: S_{11} vs. frequency for the waveguide iris problem of Figure 6.15.



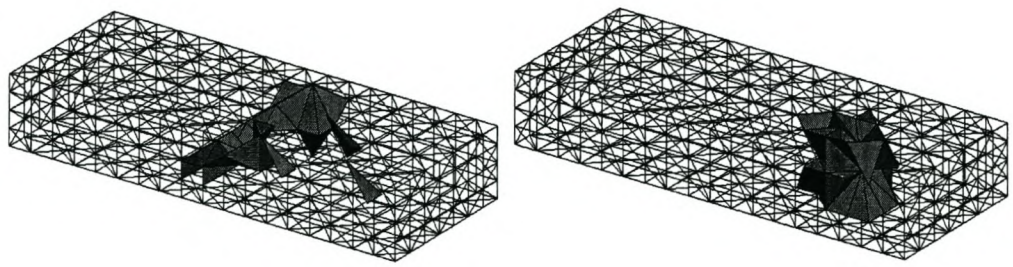
(a) $f = 8.5$ GHz.

(b) $f = 9.5$ GHz.



(c) $f = 10.5$ GHz.

Figure 6.17: Performance graphs for the waveguide iris problem of Figure 6.15. 10144 uniform LT/QN degrees of freedom.



(a) $\alpha = 0.1$. (b) $\alpha = 0.9$.

Figure 6.18: The 2.5% elements with the largest error indicator values for the waveguide iris problem of Figure 6.15 at $f = 9.5$ GHz, according to equation (6.6).

6.5.4 Waveguide bend problem

This section considers a waveguide bend problem. The problem geometry is an E-plane, 90°, standard X-band, waveguide bend. Figure 6.19 shows the finite element mesh. Figure 6.20 compares the S_{11} values obtained with uniform CT/LN- and LT/QN elements, with an approximate, analytical result by Marcuvitz [125], showing that the LT/QN result is indeed an improvement upon the CT/LN result. Again, [64] discusses the relevant manipulations of Marcuvitz's lumped-element model.

Figure 6.21 shows the performance graphs. Observe the following tendency in the performance graphs: throughout the range of degrees of freedom (upgrade percentages) considered, the $\alpha \geq 0.5$ indicators resulted in superior, near-identical performances in every graph.

The observed tendency is close to that of the waveguide iris problem in Section 6.5.3, except that at small upgrade percentages ($\leq 5\%$), the $\alpha \geq 0.5$ indicators remain superior to the $\alpha < 0.5$ indicators. In the light of this similarity, the proposes reason for the behaviour exhibited by the waveguide bend performance graphs is the same as that proposed for the waveguide iris problem's performance graphs. The difference in behaviour in the case of small upgrade percentages can be accounted for by noting that the field singularity at the re-entrant corner of the waveguide bend is of a lower order than that of the iris problem ($r^{-\frac{1}{3}}$ vs. $r^{-\frac{1}{2}}$, where r is a radial coordinate perpendicular to the re-entrant corner – see [93, p.178] for details). This means that the upgrade percentage below which the exclusive use of face residuals leads to superior results, is smaller than in the waveguide iris case. In fact, this percentage is below 2.5% and thus, it is not shown in Figure 6.21.

Figure 6.22 shows the 2.5% elements with the highest error indicator values in the case of $\alpha = 0.5$ and $f = 9.5$ GHz. Note how the re-entrant corner of the bend is covered, as one would expect (as motivated in Section 6.5.3 for the iris edge).

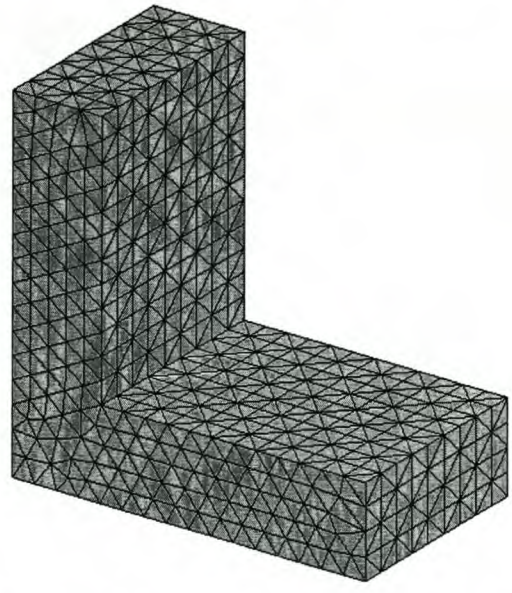


Figure 6.19: Waveguide bend problem mesh. 3331 elements. 3.5mm average edge length. The port geometries are as shown in Figure 6.5, with $a = 22.86$mm and $b = 10.16$mm.

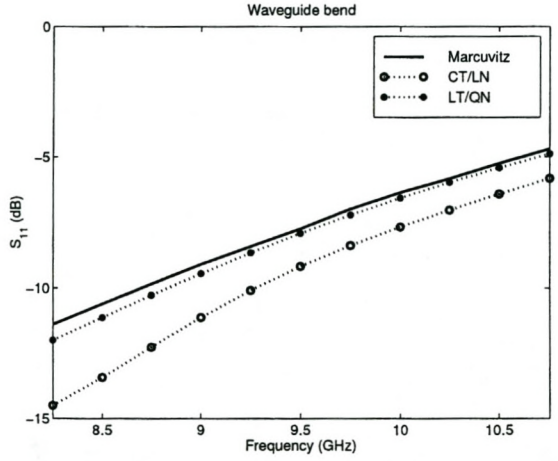
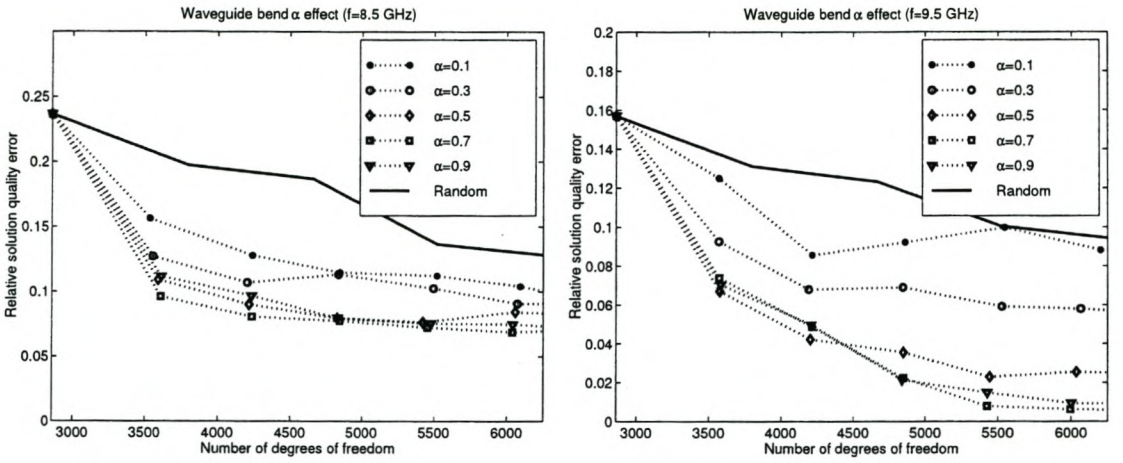
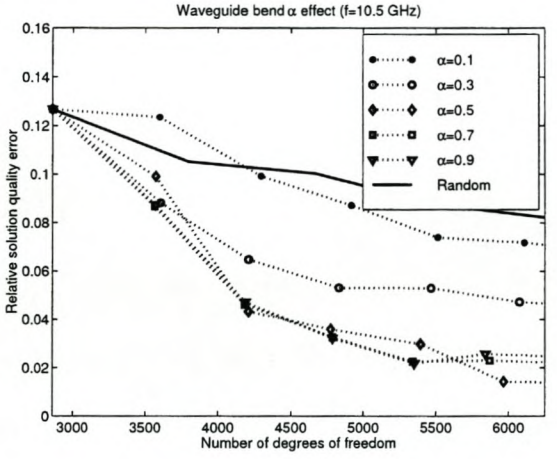


Figure 6.20: S_{11} vs. frequency for the waveguide bend problem of Figure 6.19.



(a) $f = 8.5$ GHz.

(b) $f = 9.5$ GHz.



(c) $f = 10.5$ GHz.

Figure 6.21: Performance graphs for the waveguide bend problem of Figure 6.19. 17628 uniform LT/QN degrees of freedom.

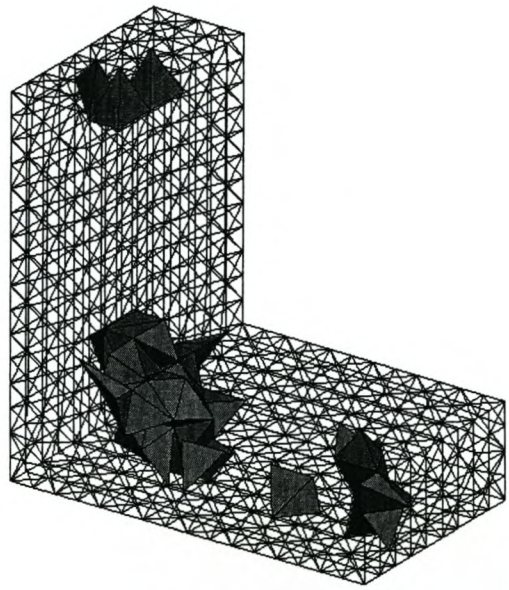


Figure 6.22: The 2.5% elements with the largest error indicator values for the waveguide bend problem of Figure 6.19 at $f = 9.5$ GHz, with $\alpha = 0.5$, according to equation (6.6).

6.6 Conclusion

Generally, this chapter demonstrated the excellent potential that the error estimation techniques developed in Chapter 5 have to successfully drive adaptive FE schemes for the FE formulations considered in this document. In two different problem settings (stepped cavity and waveguide filter), it was found that the explicit residual indicator is superior to the implicit ERM indicator, but this does not represent a comprehensive result. Note that the explicit scheme is substantially easier to implement. It was found that the performance of an indicator can be enhanced by using a quasi-static condition (equation (6.8)) to direct full- or mixed order upgrading. Graded upgrades were found to be inefficient. Concerning the relative weighting of the volume and face residual contributions to the explicit residual indicator, it was found that in some cases, one or the other is more important, but generally both should be present and $\alpha = 0.5$ is a good choice — keeping in mind that both residuals together are needed to form an upper bound on the approximate energy norm of the error field, this is not surprising.

Since this chapter is based on two error estimators which are both novel in some respects, all the adaptive results presented are novel. The results as a whole are the main contribution, extending the application of explicit- and implicit residual based error indicators to the FE formulations considered in this document, which are highly relevant in the field of microwave engineering. The quasi-static condition, for enhancing adaptive FE analysis for the full-wave Maxwell equations, using curl-conforming elements, represents a separate, important contribution.

The work presented here can be extended by considering these indicators within different adaptive schemes, for example: *h*- and *hp* upgrading and multiple adaptive cycles. The comparative indication results should be extended to include other, established indicators such as the Z^2 indicator [203], the SPR indicator [204] and the intuitive indicators found in the microwave engineering literature (see Section 5.1.3). Further experimentation with the quasi-static condition is warranted — it could possibly be used to distinguish between *p*- and *hp*- upgrades, needed in the case of irregular solutions [155]. Finally, a standard procedure to evaluate error indicators for microwave FE calculations is needed and stopping criteria need to be investigated.

Chapter 7

General conclusions

In this document, a FEM formulation for microwave engineering problems was introduced, as commonly used within the microwave engineering community. The FEM was extended to incorporate a cavity backed aperture, BI boundary condition as well as a dominant mode, coaxial port, boundary condition, both of which are quite standard. After establishing this formulation, two techniques for enhancing its computational efficiency were investigated. The first was the application of the FMM to the BI component of the formulation, resulting in minor research contributions with respect to the control of the error introduced by the FMM. The second technique, which is more directly related to the FEM and larger in scope, is the use of adaptive mesh refinement in order to optimize the FE discretization. This second investigation resulted in the main research contributions of this document.

The research contributions made in this document are discussed in more detail in the conclusions to every chapter. Here follows a list of the most important results:

- The cavity backed aperture FE-BI formulation with the FMM applied to the BI component, a dominant mode coaxial port formulation and a p adaptive analysis cycle have all been successfully integrated within a single FE computer code implementation, based on the use of hierarchical, curl-conforming, mixed- and complete order, tetrahedral vector elements. All these components are highly relevant within the field of microwave engineering, FE analysis.
- Novel results on the coupling between microstrip patch antennas on a perforated substrate were obtained with the cavity FE-BI formulation.
- Error control schemes for the FMM were thoroughly reviewed and an additional, novel scheme was devised.
- A posteriori error estimation techniques were reviewed and two known approaches within the applied mathematics literature were adapted to the FE formulation at hand, resulting in two novel error estimation procedures for this FE formulation.
- p adaptive results have been presented, demonstrating the power of the two error estimators within an adaptive algorithm. A novel, quasi-static condition was introduced and successfully used to enhance the adaptive algorithm's effectiveness, independently of the error estimation procedure employed.

Further extensions to the work presented here are clearly possible, as discussed within the conclusions to every chapter. Regarding the FMM, it would appear that most research contributions currently still being made are of an incremental nature, which can be attributed to the high level of maturity to which the FMM algorithm has already been developed. In contrast, the field of research on adaptive FE analysis for the formulation considered here still presents many unanswered questions and challenges, especially regarding the following issues:

- Extensive critical assessment of the performance of a posteriori error estimates is needed. This is important to increase confidence in the techniques. Comparative studies over a wide range of problems would enhance the usability of available techniques.
- More techniques for estimating the global solution quality is needed. This concept usually goes hand in hand with an associated a posteriori error estimator, but every error estimator does not necessarily imply a global measure of the solution quality.
- The available set of error estimators should be further expanded and improved. As pointed out earlier in this document, there is a lack of variety in error estimators for the general FEM formulation considered in this document, which implies potentially inefficient adaptive procedures.

Appendix A

Mathematical identities

A.1 Scalar identities

General inequalities [1]:

$$|z_1| - |z_2| \leq \Big||z_1| - |z_2|\Big| \leq |z_1 \pm z_2| \leq |z_1| + |z_2| \tag{A.1}$$

$$\left|\sum_{i=1}^{N} z_i\right| \leq \sum_{i=1}^{N} |z_i| \tag{A.2}$$

The Cauchy-Schwarz inequality [1]:

$$\sum_{i=1}^{N} |\alpha_i \beta_i| \leq \sqrt{\sum_{k=1}^{N} |\alpha_k|^2} \sqrt{\sum_{m=1}^{N} |\beta_m|^2} \tag{A.3}$$

Addition theorems: Suppose that $\mathbf{X}$ and $\mathbf{d}$ are two position vectors, then it follows from [1, eqs (10.1.45) and (10.1.46)] that

$$\frac{\sin k_0|\mathbf{X}+\mathbf{d}|}{k_0|\mathbf{X}+\mathbf{d}|} = \sum_{l=0}^{\infty} (2l+1)\, j_l(k_0 d)\, j_l(k_0 X)\, P_l(-\hat{d}\cdot\hat{X}) \tag{A.4}$$

$$-\frac{\cos k_0|\mathbf{X}+\mathbf{d}|}{k_0|\mathbf{X}+\mathbf{d}|} = \sum_{l=0}^{\infty} (2l+1)\, j_l(k_0 d)\, y_l(k_0 X)\, P_l(-\hat{d}\cdot\hat{X}) \quad \{d < X\} \tag{A.5}$$

where $j_l(z)$ $y_l(z)$ are the spherical Bessel functions of the first and second kinds and $P_l(z)$ is the Legendre polynomial – all of the lth order, as defined in [1]. These addition theorems are also provided in [164].

Legendre polynomial symmetry property [13, 1]:

$$P_l(-z) = (-1)^l P_l(z) \tag{A.6}$$

where $P_l(z)$ is the Legendre polynomial of the lth order.

A.2 Vector identities

General identities [187, 180]:

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} \tag{A.7}$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) \tag{A.8}$$

$$\nabla \times (\nabla a) = 0 \tag{A.9}$$

First vector Green's theorem [187, 99]:

$$\begin{aligned} &\int_V [u(\nabla \times \mathbf{a}) \cdot (\nabla \times \mathbf{b}) - \mathbf{a} \cdot (\nabla \times u\nabla \times \mathbf{b})]\, dV \\ &\quad = \oint_S u(\mathbf{a} \times \nabla \times \mathbf{b}) \cdot \hat{n}\, dS \end{aligned} \tag{A.10}$$

Second vector Green's theorem [187, 99]:

$$\begin{aligned} &\int_V [\mathbf{b} \cdot (\nabla \times u\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times u\nabla \times \mathbf{b})]\, dV \\ &\quad = \oint_S u(\mathbf{a} \times \nabla \times \mathbf{b} - \mathbf{b} \times \nabla \mathbf{a}) \cdot \hat{n}\, dS \end{aligned} \tag{A.11}$$

From the Schwarz inequality [115], it follows for the vector space $(L^2(V))^3$, endowed with the inner product

$$\langle \mathbf{A}, \mathbf{B} \rangle = \int_V \mathbf{A} \cdot \mathbf{B}^* dV,$$

that

$$\left| \int_V \mathbf{A} \cdot \mathbf{B}\, dV \right| \leq \left[\int_V \mathbf{A} \cdot \mathbf{A}^* dV \right]^{\frac{1}{2}} \left[\int_V \mathbf{B} \cdot \mathbf{B}^* dV \right]^{\frac{1}{2}}. \tag{A.12}$$

Similarly, for the vector space $(L^2(S))^3$:

$$\left| \int_S \mathbf{A} \cdot \mathbf{B}\, dS \right| \leq \left[\int_S \mathbf{A} \cdot \mathbf{A}^* dS \right]^{\frac{1}{2}} \left[\int_S \mathbf{B} \cdot \mathbf{B}^* dS \right]^{\frac{1}{2}}. \tag{A.13}$$

A.3 Dyadic identities

General identities [187, 194]:

$$\mathbf{a} \cdot \overline{\overline{C}} = \left[\overline{\overline{C}}\right]^T \cdot \mathbf{a} \tag{A.14}$$

$$\mathbf{a} \cdot (\mathbf{b} \times \overline{\overline{C}}) = -\mathbf{b} \cdot (\mathbf{a} \times \overline{\overline{C}}) = (\mathbf{a} \times \mathbf{b}) \cdot \overline{\overline{C}} \tag{A.15}$$

$$\nabla \times \left(\nabla \times \overline{\overline{C}}\right) = \nabla \left(\nabla \cdot \overline{\overline{C}}\right) - \nabla^2 \overline{\overline{C}} \tag{A.16}$$

$$\nabla \times \nabla \mathbf{a} = 0 \tag{A.17}$$

If $\mathbf{a}$ is a constant vector, then in tensor notation [14]

$$\begin{aligned}\nabla \times \left\{\overline{\overline{C}} \cdot \mathbf{a}\right\} &= \nabla \times C_{ij} a^i \mathbf{g}^j \\ &= C_{ij} a^i|_k \epsilon^{kjl} \mathbf{g}_l \\ &= C_{ij}|_k a^i \epsilon^{kjl} \mathbf{g}_l \quad \{\mathbf{a} \text{ is constant}\}\end{aligned}$$

and

$$\begin{aligned}\left\{\nabla \times \overline{\overline{C}}\right\} \cdot \mathbf{a} &= C_{ij}|_k \epsilon^{jlk} \mathbf{g}_l \mathbf{g}^i \cdot \mathbf{a} \\ &= C_{ij}|_k a^i \epsilon^{jlk} \mathbf{g}_l \\ &= C_{ij}|_k a^i \epsilon^{kjl} \mathbf{g}_l \quad \{\epsilon^{jlk} = \epsilon^{kjl}\}\end{aligned}$$

thus

$$\nabla \times \left\{\overline{\overline{C}} \cdot \mathbf{a}\right\} = \left\{\nabla \times \overline{\overline{C}}\right\} \cdot \mathbf{a} \quad \{\text{if } \mathbf{a} \text{ is constant}\}. \tag{A.18}$$

Second vector-dyadic Green's theorem [187]:

$$\begin{aligned}&\int_V \left[(\nabla \times \nabla \times \mathbf{a}) \cdot \overline{\overline{B}} - \mathbf{a} \cdot (\nabla \times \nabla \times \overline{\overline{B}})\right] dV \\ &= \oint_S \hat{n} \cdot \left[\mathbf{a} \times \nabla \times \overline{\overline{B}} + (\nabla \times \mathbf{a}) \times \overline{\overline{B}}\right] dS\end{aligned} \tag{A.19}$$

A.4 Analytic integrals

$$\int k \, dx = kx + C \tag{A.20}$$

$$\int x \, dx = \frac{1}{2}x^2 + C \tag{A.21}$$

Appendix B

Electromagnetic theory

B.1 Maxwell's equations

The foundation of classical electromagnetics is Maxwell's equations [97, 93]. These equations are written in terms of the quantities listed in Table B.1.

Table B.1: Quantities of classical electromagnetic theory.

Quantity	Order	Symbol	Extended SI units
Material permittivity	scalar	ϵ	$C \cdot V^{-1} \cdot m^{-1}$
Material permeability	scalar	μ	$V \cdot s^2 \cdot C^{-1} \cdot m^{-1}$
Electric charge density	scalar	ρ	$C \cdot m^{-3}$
Electric field intensity	vector	$\mathbf{E}$	$V \cdot m^{-1}$
Electric flux density	vector	$\mathbf{D}$	$C \cdot m^{-2}$
Magnetic field intensity	vector	$\mathbf{H}$	$A \cdot m^{-1}$
Magnetic flux density	vector	$\mathbf{B}$	$V \cdot s \cdot m^{-2}$
Electric current density	vector	$\mathbf{J}$	$A \cdot m^{-2}$

The following constitutive relations hold:

$$\mathbf{B} = \mu\mathbf{H} \tag{B.1}$$

$$\mathbf{D} = \epsilon\mathbf{E}. \tag{B.2}$$

Maxwell's equations can be stated in differential form:

Gauss' law:

$$\nabla \cdot \mathbf{D} = \rho \tag{B.3}$$

Ampère's law:

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\mathbf{D} \tag{B.4}$$

Faraday's law:

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} \tag{B.5}$$

Magnetic flux continuity law:

$$\nabla \cdot \mathbf{B} = 0 \tag{B.6}$$

Charge conservation law (derived from Gauss' law and Ampère's law):

$$\nabla \cdot \mathbf{J} + j\omega\rho = 0. \tag{B.7}$$

B.2 Field continuity conditions

These can be derived from the Maxwell equations, as described in [93, 97]. Suppose $\hat{n}_{12}$ is a unit vector normal to an interface between regions 1 and 2, pointing from 1 to 2, then the following continuity conditions hold at the interface:

$$\begin{aligned}
\hat{n}_{12} \cdot (\mathbf{D}_2 - \mathbf{D}_1) &= \sigma && \text{(B.8)}\\
\hat{n}_{12} \cdot (\mathbf{B}_2 - \mathbf{B}_1) &= 0 && \text{(B.9)}\\
\hat{n}_{12} \times (\mathbf{E}_2 - \mathbf{E}_1) &= 0 && \text{(B.10)}\\
\hat{n}_{12} \times (\mathbf{H}_2 - \mathbf{H}_1) &= \mathbf{K} && \text{(B.11)}
\end{aligned}$$

where σ and $\mathbf{K}$ represent surface electric charge- and surface electric current densities, respectively.

When there is no net current flowing into- or out of the interface region, then it follows from the charge conservation law (B.7) that

$$\hat{n}_{12} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = 0, \tag{B.12}$$

which is the usual case.

B.3 The electric field, vector wave equation

Also called the electric field Helmholtz equation, it is obtained by eliminating $\mathbf{H}$ from Maxwell's equation.

Divide Faraday's law by μ (in general, μ and ϵ are functions of position), then take the curl on both sides to obtain

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{E} = -j\omega \{\nabla \times \mathbf{H}\}. \tag{B.13}$$

Differentiate Ampère's law once with respect to time, to obtain

$$j\omega \{\nabla \times \mathbf{H}\} = j\omega\mathbf{J} - \epsilon\omega^2\mathbf{E}. \tag{B.14}$$

Now substitute equation (B.14) into equation (B.13), yielding the electric field, vector wave equation:

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{E} - \epsilon\omega^2\mathbf{E} = -j\omega\mathbf{J}. \tag{B.15}$$

This equation can be rewritten in terms of the free-space wavenumber and wave impedance, as follows:

$$\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E} - k_0^2\epsilon_r\mathbf{E} = -jk_0Z_0\mathbf{J}. \tag{B.16}$$

B.4 The Sommerfeld radiation condition

The Sommerfeld radiation condition is a formal statement of the assumption that no wave, propagating towards the origin, originating at infinity, can be part of the vector wave equation solution when all sources are located within a sphere of finite radius. The Sommerfeld radiation condition, stated in terms of the electric field, is as follows [58, 99]:

$$\lim_{r\to\infty} r\left[\nabla\times\mathbf{E}(\mathbf{r}) + jk_0\hat{r}\times\mathbf{E}(\mathbf{r})\right] = 0. \tag{B.17}$$

This condition can be derived by demanding that the surface integral component of the integral equation representation of the electric field in free space, vanish at infinity [187], or from power conservation principles [107].

B.5 Green functions

B.5.1 Free space, dyadic Green function of electric type

Definition of $\overline{\overline{\mathrm{G}}}_0(\mathbf{r},\mathbf{r}')$

The free space, dyadic Green function of electric type is defined as follows [187]:

$$\overline{\overline{\mathrm{G}}}_0(\mathbf{r},\mathbf{r}') = \left(\overline{\overline{\mathbf{I}}} - \frac{1}{k_0^2}\nabla\nabla'\right) G_0(\mathbf{r},\mathbf{r}') \tag{B.18}$$

with

$$G_0(\mathbf{r},\mathbf{r}') = \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{B.19}$$

and

$$\overline{\overline{\mathbf{I}}} = \hat{x}\hat{x} + \hat{y}\hat{y} + \hat{z}\hat{z}. \tag{B.20}$$

It represents the three electric field distributions corresponding to the three possible components ($\hat{x}$, $\hat{y}$, $\hat{z}$) of a spatially impulsive driving function to the electric field, vector wave equation (B.16) with the Sommerfeld radiation condition (described in Section B.4) imposed at the boundary of an infinite free space region. Thus:

$$\nabla\times\nabla\times\overline{\overline{\mathrm{G}}}_0(\mathbf{r},\mathbf{r}') - k_0^2\overline{\overline{\mathrm{G}}}_0(\mathbf{r},\mathbf{r}') = \overline{\overline{\mathbf{I}}}\delta(\mathbf{r}-\mathbf{r}'). \tag{B.21}$$

The free space, dyadic Green function of electric type can be derived by first solving the alternative representation of the Maxwell equations in terms of potentials. Note that $G_0(\mathbf{r},\mathbf{r}')$ is the solution of the scalar wave equation $\nabla^2 G_0(\mathbf{r},\mathbf{r}') + k_0^2 G_0(\mathbf{r},\mathbf{r}') = -\delta(\mathbf{r}-\mathbf{r}')$ with the scalar, Sommerfeld radiation condition $\lim_{r\to\infty} r\left[\frac{\partial G_0}{\partial r} - jk_0 G_0\right] = 0$ at infinity [187].

Alternative forms of the free space, dyadic Green function of electric type, based on the fact that $\nabla G_0(\mathbf{r},\mathbf{r}') = -\nabla' G_0(\mathbf{r},\mathbf{r}')$, are

$$\left(\overline{\overline{\mathbf{I}}} + \frac{1}{k_0^2}\nabla\nabla\right) G_0(\mathbf{r},\mathbf{r}')$$

$$\left(\bar{\bar{\mathbf{I}}} + \frac{1}{k_0^2}\nabla'\nabla'\right) G_0(\mathbf{r}, \mathbf{r}')$$
$$\left(\bar{\bar{\mathbf{I}}} - \frac{1}{k_0^2}\nabla'\nabla\right) G_0(\mathbf{r}, \mathbf{r}')$$

however these forms are not suitable for defining the half space, dyadic Green function of electric type as in Section B.5.2. The last form in particular may seem like a viable alternative, but it will lead to the transpose of the true half space, dyadic Green function (see equation (B.36)). According to the definition of dyadic Green functions of electric type, $-jk_0Z_0\overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}')\, dV$ must be a valid, incremental contribution to the electric field and *not* $-jk_0Z_0\mathbf{J}(\mathbf{r}') \cdot \overline{\overline{\mathrm{G}}}_e(\mathbf{r}, \mathbf{r}')\, dV$ [194].

Symmetry properties of $\overline{\overline{\mathrm{G}}}_0(\mathbf{r}, \mathbf{r}')$

$\overline{\overline{\mathrm{G}}}_0(\mathbf{r}, \mathbf{r}')$ has the following symmetry properties [187]:

$$\left[\overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r})\right]^T = \overline{\overline{\mathrm{G}}}_0(\mathbf{r}, \mathbf{r}') \tag{B.22}$$

$$\overline{\overline{\mathrm{G}}}_0(\mathbf{r}', \mathbf{r}) = \overline{\overline{\mathrm{G}}}_0(\mathbf{r}, \mathbf{r}'). \tag{B.23}$$

Derivative properties of $G_0(\mathbf{r}, \mathbf{r}')$ and $G_0(\mathbf{r}', \mathbf{r})$

In this section, let

$$R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} \tag{B.24}$$

$$T_1 = \frac{jk_0}{R} + \frac{1}{R^2} \tag{B.25}$$

$$T_2 = \frac{jk_0}{R^3} + \frac{2}{R^4} \tag{B.26}$$

$$\alpha, \beta \in \{x, y, z\} \mid \alpha \neq \beta, \tag{B.27}$$

$$G_0(\mathbf{r}, \mathbf{r}') = G_0(\mathbf{r}', \mathbf{r}) \equiv G_0 \tag{B.28}$$

then

$$\frac{\partial G_0}{\partial \alpha} = -(\alpha - \alpha')T_1 G_0 \tag{B.29}$$

$$\frac{\partial G_0}{\partial \alpha} = -\frac{\partial G_0}{\partial \alpha'} \tag{B.30}$$

$$\frac{\partial^2 G_0}{\partial \alpha^2} = -\frac{\partial^2 G_0}{\partial \alpha \partial \alpha'} = \frac{\partial^2 G_0}{\partial \alpha'^2} = \left[(\alpha - \alpha')^2(T_1^2 + T_2) - T_1\right] G_0 \tag{B.31}$$

$$\frac{\partial^2 G_0}{\partial \alpha \partial \beta} = -\frac{\partial^2 G_0}{\partial \alpha \partial \beta'} = \frac{\partial^2 G_0}{\partial \alpha' \partial \beta'} = (\alpha - \alpha')(\beta - \beta')(T_1^2 + T_2) G_0. \tag{B.32}$$

Calculation of $\nabla' \times \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r})$

$$\nabla' \times \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) = \begin{pmatrix} 0 & -\dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial z'} \hat{x}\hat{y} & \dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial y'} \hat{x}\hat{z} \\ \dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial z'} \hat{y}\hat{x} & 0 & -\dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial x'} \hat{y}\hat{z} \\ -\dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial y'} \hat{z}\hat{x} & \dfrac{\partial G_0(\mathbf{r}', \mathbf{r})}{\partial x'} \hat{z}\hat{y} & 0 \end{pmatrix}. \tag{B.33}$$

Calculation of $\nabla \times \nabla \times \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r})$

$$\begin{aligned} \nabla \times \nabla \times \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) &= -\nabla^2 \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) + \nabla \left(\nabla \cdot \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) \right) \quad \{\text{from equation (A.16)}\} \\ &= -\bar{\bar{\mathbf{I}}} \nabla^2 G_0(\mathbf{r}', \mathbf{r}) - \nabla' \left(\nabla \cdot \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) \right) \quad \{\text{from equation (B.30)}\} \\ &= -\bar{\bar{\mathbf{I}}} \nabla^2 G_0(\mathbf{r}', \mathbf{r}) - \nabla' \nabla G_0(\mathbf{r}', \mathbf{r}) \\ &= -\bar{\bar{\mathbf{I}}} R^2 (T_1^2 + T_2) G_0(\mathbf{r}', \mathbf{r}) + 3\bar{\bar{\mathbf{I}}} T_1 G_0(\mathbf{r}', \mathbf{r}) \\ &\quad -\nabla' \nabla G_0(\mathbf{r}', \mathbf{r}) \quad \{\text{from equation (B.31)}\} \\ &= k_0^2 \bar{\bar{\mathbf{I}}} G_0(\mathbf{r}', \mathbf{r}) - \nabla' \nabla G_0(\mathbf{r}', \mathbf{r}) \\ &= k_0^2 \bar{\bar{\mathbf{G}}}_0(\mathbf{r}', \mathbf{r}). \end{aligned} \tag{B.34}$$

B.5.2 Half space, dyadic Green function of electric type

Definition of $\bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}')$

The half space, dyadic Green function of electric type, $\bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}')$, is the solution of the vector wave equation in the presence of a PEC half space at $z \leq 0$:

$$\nabla \times \nabla \times \bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}') - k_0^2 \bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}') = \bar{\bar{\mathbf{I}}} \delta(\mathbf{r} - \mathbf{r}'). \tag{B.35}$$

$\bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}')$ is defined in terms of $\bar{\bar{\mathbf{G}}}_0(\mathbf{r}, \mathbf{r}')$ (defined in equation (B.18)) through the use of image theory [21, 92] for electric currents to satisfy the boundary condition at $z = 0$ [187], yielding

$$\bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}') = \bar{\bar{\mathbf{G}}}_0(\mathbf{r}, \mathbf{r}') - \bar{\bar{\mathbf{G}}}_0(\mathbf{r}, \mathbf{r}_i') + 2\hat{z}\hat{z} G_0(\mathbf{r}, \mathbf{r}_i'), \tag{B.36}$$

with the image position

$$\mathbf{r}_i' = x'\hat{x} + y'\hat{y} - z'\hat{z}. \tag{B.37}$$

$\bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}')$ satisfies $\hat{z} \times \bar{\bar{\mathbf{G}}}_e(\mathbf{r}, \mathbf{r}') = 0$ at $z = 0$ and the Sommerfeld radiation condition at $r \to \infty$ (with $z > 0$).

In equation (B.36) some ambiguity is present, since it is not clear that the relevant derivatives within $\bar{\bar{\mathbf{G}}}_0(\mathbf{r}, \mathbf{r}_i')$ are with respect to z' and *not* with respect to $-z'$. To avoid this ambiguity, the function $G_{0i}(\mathbf{r}, \mathbf{r}')$ is introduced, defined as

$$G_{0i}(\mathbf{r}, \mathbf{r}') = G_0(\mathbf{r}, \mathbf{r}_i'). \tag{B.38}$$

The half space, dyadic, Green function of electric type can now be redefined in terms of $G_{0i}(\mathbf{r},\mathbf{r}')$, as

$$\overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}') = \overline{\overline{\mathrm{G}}}_0(\mathbf{r},\mathbf{r}') - \left(\overline{\overline{\mathrm{I}}} - \frac{1}{k_0^2}\nabla\nabla'\right) G_{0i}(\mathbf{r},\mathbf{r}') + 2\hat{z}\hat{z}G_{0i}(\mathbf{r},\mathbf{r}'). \tag{B.39}$$

Symmetry properties of $\overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}')$

$\overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}')$ has the following symmetry properties:

$$\left[\overline{\overline{\mathrm{G}}}_e(\mathbf{r}',\mathbf{r})\right]^T = \overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}') \quad \{\text{See [187]}\} \tag{B.40}$$

$$\overline{\overline{\mathrm{G}}}_e(\mathbf{r}',\mathbf{r}) \neq \overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}') \tag{B.41}$$

$$\overline{\overline{\mathrm{G}}}_e(\mathbf{r}',\mathbf{r})_{ij} \neq -\overline{\overline{\mathrm{G}}}_e(\mathbf{r},\mathbf{r}')_{ji} \quad \text{with } i \neq j \text{ and } i,j \in \{1,2,3\}. \tag{B.42}$$

Derivative properties of $G_{0i}(\mathbf{r},\mathbf{r}')$ and $G_{0i}(\mathbf{r}',\mathbf{r})$

In this section, let

$$R = \sqrt{(x-x')^2 + (y-y')^2 + (z+z')^2} \tag{B.43}$$

$$T_1 = \frac{jk_0}{R} + \frac{1}{R^2} \tag{B.44}$$

$$T_2 = \frac{jk_0}{R^3} + \frac{2}{R^4} \tag{B.45}$$

$$\alpha,\beta \in \{x,y\} \mid \alpha \neq \beta, \tag{B.46}$$

$$G_{0i}(\mathbf{r},\mathbf{r}') = G_{0i}(\mathbf{r}',\mathbf{r}) \equiv G_{0i} \tag{B.47}$$

then

$$\frac{\partial G_{0i}}{\partial \alpha} = -\frac{\partial G_{0i}}{\partial \alpha'} = -(\alpha-\alpha')T_1 G_{0i} \tag{B.48}$$

$$\frac{\partial G_{0i}}{\partial z} = \frac{\partial G_{0i}}{\partial z'} = -(z+z')T_1 G_{0i} \tag{B.49}$$

$$\frac{\partial^2 G_{0i}}{\partial \alpha^2} = -\frac{\partial^2 G_{0i}}{\partial \alpha \partial \alpha'} = \frac{\partial^2 G_{0i}}{\partial \alpha'^2} = \left[(\alpha-\alpha')^2(T_1^2+T_2) - T_1\right] G_{0i} \tag{B.50}$$

$$\frac{\partial^2 G_{0i}}{\partial z^2} = \frac{\partial^2 G_{0i}}{\partial z \partial z'} = \frac{\partial^2 G_{0i}}{\partial z'^2} = \left[(z+z')^2(T_1^2+T_2) - T_1\right] G_{0i} \tag{B.51}$$

$$\frac{\partial^2 G_{0i}}{\partial \alpha \partial \beta} = -\frac{\partial^2 G_{0i}}{\partial \alpha \partial \beta'} = \frac{\partial^2 G_{0i}}{\partial \alpha' \partial \beta'} = (\alpha-\alpha')(\beta-\beta')(T_1^2+T_2)G_{0i} \tag{B.52}$$

$$\frac{\partial^2 G_{0i}}{\partial \alpha \partial z} = \frac{\partial^2 G_{0i}}{\partial \alpha \partial z'} = -\frac{\partial^2 G_{0i}}{\partial \alpha' \partial z} = -\frac{\partial^2 G_{0i}}{\partial \alpha' \partial z'} = (\alpha-\alpha')(z+z')(T_1^2+T_2)G_{0i}. \tag{B.53}$$

Calculation of $\nabla' \times (\hat{x}\hat{x}G_{0i}(\mathbf{r}',\mathbf{r}) + \hat{y}\hat{y}G_{0i}(\mathbf{r}',\mathbf{r}) - \hat{z}\hat{z}G_{0i}(\mathbf{r}',\mathbf{r}))$

$$\nabla' \times (\hat{x}\hat{x}G_{0i}(\mathbf{r}',\mathbf{r}) + \hat{y}\hat{y}G_{0i}(\mathbf{r}',\mathbf{r}) - \hat{z}\hat{z}G_{0i}(\mathbf{r}',\mathbf{r}))$$

$$= \begin{pmatrix} 0 & -\dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial z'}\hat{x}\hat{y} & -\dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial y'}\hat{x}\hat{z} \\ \dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial z'}\hat{y}\hat{x} & 0 & \dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial x'}\hat{y}\hat{z} \\ -\dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial y'}\hat{z}\hat{x} & \dfrac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial x'}\hat{z}\hat{y} & 0 \end{pmatrix} \tag{B.54}$$

From equations (B.29), (B.48) and (B.49) it follows that

$$\left.\frac{\partial G_0(\mathbf{r}',\mathbf{r})}{\partial \alpha'}\right|_{z'=0} = \left.\frac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial \alpha'}\right|_{z'=0} \qquad \alpha \in \{x, y\} \tag{B.55}$$

$$\left.\frac{\partial G_0(\mathbf{r}',\mathbf{r})}{\partial z'}\right|_{z'=0} = -\left.\frac{\partial G_{0i}(\mathbf{r}',\mathbf{r})}{\partial z'}\right|_{z'=0} \tag{B.56}$$

Comparing equations (B.33) and (B.54) in light of the previous two equations, it follows that

$$\hat{\alpha} \cdot \nabla' \times (\hat{x}\hat{x}G_{0i}(\mathbf{r}',\mathbf{r}) + \hat{y}\hat{y}G_{0i}(\mathbf{r}',\mathbf{r}) - \hat{z}\hat{z}G_{0i}(\mathbf{r}',\mathbf{r}))|_{z'=0} = -\hat{\alpha} \cdot \nabla' \times \bar{\bar{\mathrm{I}}}G_0(\mathbf{r}',\mathbf{r})\Big|_{z'=0} \qquad \hat{\alpha} \in \{\hat{x}, \hat{y}\}. \tag{B.57}$$

B.6 Calculation of $\mathbf{H}^{\text{inc}}$: the TEM plane wave case

See [97]. A TEM plane wave, traveling in the positive γ direction can be defined in phasor form as

$$\mathbf{H}^{\text{inc}} = \left\{H_0 e^{-jk_0\gamma}\right\}\hat{\eta}. \tag{B.58}$$

γ is the coordinate measured along the vector of propagation, calculated as follows:

$$\gamma = \hat{v} \cdot \mathbf{r} \tag{B.59}$$

with

$$\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z} \tag{B.60}$$

$$\hat{v} = -\sin\theta\cos\phi\hat{x} - \sin\theta\sin\phi\hat{y} - \cos\theta\hat{z} \tag{B.61}$$

where $\hat{v}$ is the unit vector defining the positive, wave traveling direction in terms of the angles θ and ϕ, which in turn define, in spherical coordinates, the direction from which the wave is incident upon the coordinate system origin.

$\hat{\eta}$ is the magnetic field polarization unit vector, which is normal to the vector of propagation and is defined in terms of the polarization angle η, as follows:

$$\hat{\eta} = -\cos\eta\hat{\theta} + \sin\eta\hat{\phi}. \tag{B.62}$$

Appendix C

The FEM elements

C.1 Elemental properties

C.1.1 Basic properties

Tetrahedral

The elements are tetrahedral and rectilinear. The tetrahedron is the simplest rectilinear shape in three dimensions. Any polyhedral shape can be represented as the union of a number of tetrahedrons [175]. Curvilinear elements [158, 174, 145] are not considered here.

Vector elements

Given an unknown vector field $\mathbf{v}$, it is represented on the elemental domain K, as a sum of *elemental basis functions*, each multiplied by an unknown coefficient. These unknown coefficients are referred to as the *elemental degrees of freedom*.

Vector elements refer to basis functions that are vector-valued as opposed to scalar-valued ones. Scalar elements are usually interpolatory and node-based [195, 135], i.e. the basis functions interpolate to the unknown coefficient values at a specified set of elemental nodes, usually including the element vertices. To model $\mathbf{v}$ with scalar elements, separate basis functions must be used for the three vector components. Examples of scalar elements are numerous: [43, 167] (Lagrange elements in two- and three dimensions), [174] (simplex elements in three dimensions) and [71] (two dimensional, rectangular, h-p elements on 1-irregular meshes).

Vector elements are used here, approximating an elemental field as follows:

$$\mathbf{v}_h^K = \sum_{i=1}^{N_F^K} v_i^K \mathbf{N}_i^K, \tag{C.1}$$

where $\mathbf{N}_i^K$ and v_i^K, with $i = 1, .., N_F^K$, represent the elemental sets of vector basis functions and degrees of freedom, respectively.

Curl-conforming

Most vector elements in electromagnetics are *curl-conforming* [145, 168], defined in [136, 145] as follows: The approximation is curl-conforming if the tangential components of the approximation in any two elements that share parts of their boundaries (edges and/or faces) are equal at the shared boundary. Curl-conforming elements are also referred to as *tangentially-continuous* elements [198].

This implies the construction of basis functions that allow enforcing only tangential continuity, providing a natural way of enforcing boundary conditions in terms of tangential components only. Such bases have the capability of correctly modeling the Maxwell continuity conditions (see Section B.2) on inter-element boundaries (possibly between different media, i.e. discontinuous normal electric field), in a distributional sense, as shown in [26, 167]. The non-enforcement of normal continuity also presents an advantage over total continuity when modeling the (infinite) electric field strength near a re-entrant PEC boundary [195].

[135] notes a slight drawback to field modeling with curl-conforming elements as opposed to scalar, interpolatory elements: on the same mesh, the latter leads to less degrees of freedom than the former, since enforcing full continuity as opposed to continuity of only the two tangential components results in more shared degrees of freedom. On the other hand, [167] remarks that the sparsity of the system matrix is higher in the case of curl-conforming elements, because of the lower connectivity.

Curl-conforming elements play an important role in the spurious modes issue (see [167] for a survey on this matter). According to [167, 195, 145, 45, 184], spurious modes are severely inaccurate representations of eigenmodes that are supposed to correspond to zero eigenvalues of the eigenvalue problem $\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E} = k_0^2 \epsilon_r \mathbf{E}$ (see equation (B.16)). This results in contamination of the eigenvalue spectrum of interest by these erroneous, non-zero eigenvalues. Clearly these eigenmodes must be elements of the curl operator's null space, that is $\nabla\phi$. [167, 195, 184] shows how the non-enforcement of normal continuity (using curl-conforming elements) allows for much more comprehensive modeling of the curl operator's null space, resulting in numerically zero eigenvalues. The spurious modes issue is relevant to the driven problem as well, since the solution can be seen as a superposition of eigenmodes [167, 195].

[167, Section 7.2.2.2] gives a thorough review of curl-conforming element properties. Finally, it is interesting to note that [157, 136, 137] define div-conforming degrees of freedom on various 2D and 3D elements.

Hierarchical

This property describes the way in which different sets of elemental basis functions relate to each other. Suppose $\{N_p^K\}$ and $\{N_q^K\}$ are two sets of basis functions on elemental domain K, such that

$$\text{span}\{N_p^K\} \subset \text{span}\{N_q^K\}. \tag{C.2}$$

If the bases are *hierarchical*, then

$$\{N_p^K\} \subset \{N_q^K\}. \tag{C.3}$$

Interpolatory bases, of which all the members are of the same polynomial order, form a large subset of the non-hierarchical bases.

Complete- and mixed order

These concepts involve the polynomial spaces spanned by the vector basis functions within an element. Define the gradient spaces G:

$$G_S = \text{span}\{\nabla\psi\} \quad \text{with} \quad \psi \in S. \tag{C.4}$$

From equation (A.9) it can be observed that

$$\text{Ker}\,(\nabla \times \mathbf{W}) = G_{C^0(\mathcal{R}^3)}. \tag{C.5}$$

Suppose that $\mathbf{W}_p \in (P_p(K))^3$ (where K is an arbitrary, elemental volume), then

$$\begin{aligned} \text{Ker}\,(\nabla \times \mathbf{W}_p) &= G_{P_{p+1}(K)} \\ &= G_{\overline{P}_1(K)} \oplus G_{\overline{P}_2(K)} \oplus \ldots \oplus G_{\overline{P}_p(K)} \oplus G_{\overline{P}_{p+1}(K)} \end{aligned} \tag{C.6}$$

and observe that

$$G_{\overline{P}_{p+1}(K)} \subset (\overline{P}_p(K))^3. \tag{C.7}$$

If basis functions spanning at least $(P_p(K))^3$ are employed, then they are said to be *complete* up to polynomial order p (at least). Suppose that a basis complete to order p is used, then $\mathbf{E}$ will be approximated to order p complete, but $\nabla \times \mathbf{E}$ will only be approximated to order $p-1$ complete; thus the largest complete order up to which both are modeled is $p-1$.

Define the mixed order spaces M as defined in [137] and discussed in [145, 137, 136]:

$$(P_{p-1}(K))^3 \subset M_p(K) \equiv (P_p(K))^3 \setminus G_{\overline{P}_{p+1}(K)}. \tag{C.8}$$

Now it follows that if $\mathbf{E}$ is represented by a basis spanning $M_p(K)$, then the maximum order to which both $\mathbf{E}$ and $\nabla \times \mathbf{E}$ are represented completely is still $p-1$, since $G_{\overline{P}_{p+1}(K)} \subset \text{Ker}\,(\nabla \times \mathbf{W}_p)$.

The dimensions of the spaces $(P_p(K))^3$, $G_{\overline{P}_{p+1}(K)}$ and $M_p(K)$ are as follows [137, 136, 196]:

$$\dim\,(P_p(K))^3 = \frac{(p+1)(p+2)(p+3)}{2} \tag{C.9}$$

$$\begin{aligned} \dim\, G_{\overline{P}_{p+1}(K)} &= \dim\, \overline{P}_{p+1}(K) \\ &= \frac{(p+2)(p+3)}{2} \end{aligned} \tag{C.10}$$

$$\begin{aligned} \dim\, M_p(K) &= \dim\,(P_p(K))^3 - \dim\, G_{\overline{P}_{p+1}(K)} \\ &= \frac{p(p+2)(p+3)}{2}. \end{aligned} \tag{C.11}$$

Assume that the accuracy of the discretized VBVP depends on the highest complete order to which both $\mathbf{E}$ and $\nabla \times \mathbf{E}$ are modeled, since their roles are of equal importance in the

bilinear form (equation (2.12), i.e. $\int_\Omega [\frac{1}{\mu_r} \nabla \times \mathbf{E} \cdot \nabla \times \mathbf{W} - k_0^2 \epsilon_r \mathbf{E} \cdot \mathbf{W}] \, dV$). Then it follows that bases spanning $M_p(K)$ rather than $(P_p(K))^3$, allows for an optimal representation of both $\mathbf{E}$ and $\nabla \times \mathbf{E}$, complete to the same order.

If a vector field is $\in M_p(K)$, then the variation in the tangential component along any straight line will always be of polynomial order $\leq p-1$ [136] [167, p. 106], as a consequence of the optimal representation. However, it may not be possible to model $M_p(K)$ (or $(P_p(K))^3$) exactly on an element, since some additional dimensions might be needed to satisfy the conformity condition, depending on the element shape. On tetrahedrons, $M_p(K)$ and $(P_p(K))^3$ can be exactly modeled in terms of curl-conforming basis functions [136, 137].

Examples of mixed- and complete order, curl-conforming bases are numerous. In 2D, [117] presents interpolatory, triangular elements, spanning a 2D version of $M_2(K)$. [10] presents mixed order, hierarchical bases on triangles, up to $M_3(K)$. [72, 189] present interpolatory bases of arbitrary, mixed- and complete orders, on triangles. In 3D, [198] presents tetrahedral, hierarchical bases spanning the full- and mixed spaces up to $(P_2(K))^3$. [169, 145] present non-hierarchical bases on tetrahedrons, spanning the mixed spaces up to $M_3(K)$. [87] presents mixed order, interpolatory bases of arbitrary order on triangles, quadrilaterals, tetrahedrons and bricks. [168, 9] present hierarchical bases on tetrahedrons, spanning the mixed spaces up to $M_4(K)$ and $M_3(K)$, respectively. Finally, [196] presents a general procedure for obtaining hierarchical, curl-conforming bases on tetrahedrons, spanning the mixed and complete spaces to arbitrary order. The basis functions are separated into two groups: those spanning $\left((P_p(K))^3 \setminus G_{P_{p+1}(K)}\right) \oplus (P_0(K))^3$ and those spanning $G_{P_{p+1}(K)} \setminus G_{\overline{P}_1(K)}$, which together form the complete space $(P_p(K))^3$.

C.1.2 Definitions of elemental approximation spaces

This section defines the spaces spanned by the elemental basis functions presented in this appendix. The elemental spaces used for approximating the solution of the electric field VBVP (see Section 2.4) is defined first, followed by the definition of bubble spaces, which are relevant to the approximation of the finite element solution error (see Section 5.4).

Solution approximation spaces

The solution approximation spaces $U_p(K)$, of polynomial order p on elemental volume K, are defined as follows:

$$U_p(K) = \begin{cases} M_m(K) & \Rightarrow \quad p = m - 0.5 \\ (P_m(K))^3 & \Rightarrow \quad p = m \end{cases} . \tag{C.12}$$

This half-order terminology is also used in [198, 10], for example.

Bubble spaces

The bubble spaces $V_{p,q}(K)$, of 'polynomial orders' p to q on elemental volume K, are defined in terms of the solution approximation spaces, as defined in equation (C.12), as follows:

$$V_{p,q}(K) = U_q(K) \setminus U_p(K) \quad \{p < q\} \tag{C.13}$$

or equivalently, as

$$U_q(K) = U_p(K) \oplus V_{p,q}(K) \quad \{p < q\}. \tag{C.14}$$

C.1.3 Interpolation error estimates

The interpolation operator

The elemental interpolation operator $\pi_K : H(\text{curl}, K) \rightarrow U_p(K)$ is a way of approximately representing an arbitrary, known field $\mathbf{v} \in H(\text{curl}, K)$ in terms of a sum of degrees of freedom multiplying basis functions spanning the space $U_p(K)$, in the form of equation (C.1). In other words, the elemental interpolation operator defines the calculation of the elemental degrees of freedom $\{v_i^K \mid i = 1, .., N_F^K\}$ associated with the set of elemental basis functions $\{N_i^K \mid i = 1, .., N_F^K\}$, given an arbitrary, known field $\mathbf{v}$ on elemental volume K. In [136, 137], curl-conforming degrees of freedom are defined for tetrahedral, vector elements, spanning the spaces $M_p(K)$ and $(P_p(K))^3$. The degrees of freedom are linear functionals on the edges, faces and volume of the tetrahedron and are *unisolvent*, which means that

$$v_i^K = 0 \ \ \forall\, i \in [1, N_F^K] \Rightarrow \mathbf{v} = 0 \quad \{\mathbf{v} \in U_p(K)\}. \tag{C.15}$$

The elemental interpolation operator has a global counterpart, such that

$$\pi\mathbf{v}|_K = \pi_K \mathbf{v} \quad \{\mathbf{v} \in H(\text{curl}, \Omega)\} \tag{C.16}$$

which is an obvious consequence of the curl-conforming nature of the degrees of freedom.

The basis functions can be derived from the definitions of the degrees of freedom [167], but since a set of previously published basis functions will be used throughout, this issue will not be discussed further.

A scalar, Sobolev semi-norm, interpolation error estimate

This interpolation estimate is presented and proved in [51, Thm 3.1.5]. Here it is presented in a form restricted to the current needs. A similar result is presented in [43, pp.104-105].

Let tetrahedral finite element K with associated scalar, interpolation operator π_K, model the space S, with $P_{m-1}(K) \subset S \subset H^m(K)$, $m \geq 1$, $0 \leq i \leq m$ and $v \in H^m(K)$, then it follows that

$$|v - \pi_K v|_{H^i(K)} \leq C h_K^{m-i} |v|_{H^m(K)}, \tag{C.17}$$

with $h_K = \text{diam}(K)$ and the assumption that the diameter of the largest inscribed sphere within K is > 0. C is dependent on the element geometry, π_K and m.

A vector, Sobolev semi-norm, interpolation error estimate

Substituting the components of the vector field $\mathbf{v} \in (H^m(K))^3$ individually into equation (C.17), adding the squares and taking the square root of the sum, one obtains the following vector, Sobolev, semi-norm, interpolation error estimate:

$$|\mathbf{v} - \pi_K \mathbf{v}|_{(H^i(K))^3} \leq C h_K^{m-i} |\mathbf{v}|_{(H^m(K))^3}, \tag{C.18}$$

with π_K the curl-conforming, interpolation operator that maps to the space $U_p(K)$, with $1 \leq m \leq \lceil p \rceil$.

This estimate corresponds to the interpolation estimates provided in [136, 137] for the tetrahedral elements spanning the spaces $M_m(K)$ and $(P_m(K))^3$, with $i = 0$.

A scalar, trace estimate

The trace of a function on element K is its limiting value on ∂K as the boundary is approached from inside the element [158]. [191], [133, eq.(4.14)] and [151] provide the following scalar trace estimate on element K:

$$\|v\|^2_{L^2(\partial K)} \leq C \left(\|v\|_{L^2(K)} \|\nabla v\|_{L^2(K)} + \frac{1}{h_K} \|v\|^2_{L^2(K)} \right), \tag{C.19}$$

with $v \in H^1(K)$ and $h_K = \text{diam}(K)$.

A vector, L^2 norm, trace interpolation error estimate

Apply equation (C.19) separately to the components of a vector and sum the resulting inequalities to yield

$$\|\mathbf{v}\|^2_{L^2(\partial K)} \leq C \|\mathbf{v}\|_{L^2(K)} \left(|\mathbf{v}|_{(H^1(K))^3} + \frac{1}{h_K} \|\mathbf{v}\|_{L^2(K)} \right). \tag{C.20}$$

Replace $\mathbf{v}$ with $\mathbf{v} - \pi_K \mathbf{v}$ to obtain

$$\begin{aligned} \|\mathbf{v} - \pi_K \mathbf{v}\|^2_{L^2(\partial K)} &\leq C \|\mathbf{v} - \pi_K \mathbf{v}\|_{L^2(K)} \left(|\mathbf{v} - \pi_K \mathbf{v}|_{(H^1(K))^3} + \frac{1}{h_K} \|\mathbf{v} - \pi_K \mathbf{v}\|_{L^2(K)} \right) \\ &= C |\mathbf{v} - \pi_K \mathbf{v}|_{(H^0(K))^3} \left(|\mathbf{v} - \pi_K \mathbf{v}|_{(H^1(K))^3} + \frac{1}{h_K} |\mathbf{v} - \pi_K \mathbf{v}|_{(H^0(K))^3} \right) \end{aligned} \tag{C.21}$$

where the fact that $\|\mathbf{v} - \pi_K \mathbf{v}\|_{L^2(K)} = |\mathbf{v} - \pi_K \mathbf{v}|_{(H^0(K))^3}$ was used.

Now construct a trace interpolation estimate in the L^2 norm by substituting equation (C.18) (with the enclosed requirements $\mathbf{v} \in (H^m(K))^3$ and $1 \leq m \leq \lceil p \rceil$) into the above inequality, resulting in

$$\|\mathbf{v} - \pi_K \mathbf{v}\|_{L^2(\partial K)} \leq C h_K^{m-0.5} |\mathbf{v}|_{(H^m(K))^3}. \tag{C.22}$$

C.2 Elemental basis functions

C.2.1 Local numbering conventions

The edge and face definitions in terms of local node numbers are adopted from [169]. The local node numbers relate to the global node numbers in ascending order (i.e. local node 1 associates with the local node of lowest global number). Tables C.1, C.2 and C.3 list the edge and face definitions.

Table C.1: Element edge definitions in terms of local node numbers.

Edge No.	Node 1	Node 2
1	1	2
2	1	3
3	1	4
4	2	3
5	2	4
6	3	4

Table C.2: Element face definitions in terms of local node numbers.

Face No.	Node 1	Node 2	Node 3
1	1	2	3
2	1	2	4
3	1	3	4
4	2	3	4

Table C.3: Element face definitions in terms of local edge numbers.

Face No.	Edge 1	Edge 2	Edge 3
1	1	2	4
2	1	3	5
3	2	3	6
4	4	5	6

C.2.2 Simplex coordinates on tetrahedral elements

The elemental basis functions are defined in terms of *simplex coordinates* (also known as barycentric- or volume coordinates) [174]. This is a coordinate system used to reference points within a tetrahedron, independent of its orientation and specific shape, therefore only relative to the ordered set of four nodes. Consider the following equation:

$$[S_c] = \begin{bmatrix} b_1 & c_1 & d_1 & a_1 \\ b_2 & c_2 & d_2 & a_2 \\ b_3 & c_3 & d_3 & a_3 \\ b_4 & c_4 & d_4 & a_4 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \tag{C.23}$$

where (x_i, y_i, z_i) with $i = 1, .., 4$, are the four nodes of the tetrahedron. The simplex coordinates, $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, of a point, (x, y, z), are calculated as follows:

$$\begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{Bmatrix} = [S_c] \begin{Bmatrix} x \\ y \\ z \\ 1 \end{Bmatrix} \tag{C.24}$$

where λ_i is said to be the simplex coordinate associated with node i and where $\{\lambda\}$ exhibits the following properties:

- λ_i varies linearly from 1 at node i to 0 at the other three nodes.
- $\sum_{i=1}^{4} \lambda_i = 1$, which follows directly from the above definition of simplex co-ordinates.
- $0 \leq \lambda_i \leq 1$, $i = 1, .., 4$, for all points within the tetrahedron concerned.

Simplex coordinates can also be used with triangular elements — then all the above still applies except that $[S_c]$ is a 3×3 matrix and there are only 3 simplex coordinates.

C.2.3 Basis functions

These basis functions are presented in [196] and are defined as polynomial in λ_i, multiplied by $\nabla\lambda_i$. In the case of rectilinear elements, the λ_i are linear functions of x, y and z (and the $\nabla\lambda_i$ are constant vectors), thus the simplex coordinate polynomial order of a basis function is equal to its Cartesian coordinate polynomial order. As noted earlier, the degrees of freedom are associated with edges, faces and the element volume [136, 137] — the same associations hold for the basis functions. These associations, the definition in terms of simplex coordinates and the consistent relation between local and global node numbers, together ensure that the basis functions are curl-conforming. Table C.4 presents the tetrahedral, hierarchical, curl-conforming, vector basis functions, with their elemental geometric associations and the elemental spaces that they span. The designations are commonly used in the literature, see [145, 169, 9]. The lowest order, CT/LN element is also referred to as the Whitney element [34].

Table C.4: Definitions of the basis functions in terms of simplex coordinates [196].

Cumulative designation	Cumulative space spanned	Actual space spanned	Geometric association	Number	Definition
CT/LN	$U_{0.5}(K)$	$U_{0.5}(K)$	edge	6	$\lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i \quad \{i < j\}$
LT/LN	$U_1(K)$	$V_{0.5,1}(K)$	edge	6	$\nabla (\lambda_i \lambda_j)$
LT/QN	$U_{1.5}(K)$	$V_{1,1.5}(K)$	face	$4+4$	$\lambda_j \lambda_k \nabla \lambda_i + \lambda_i \lambda_k \nabla \lambda_j - 2\lambda_i \lambda_j \nabla \lambda_k$ $\{i,j,k\} = \text{Face nodes}\{1,2,3\}\ \forall$ faces $\{i,j,k\} = \text{Face nodes}\{2,3,1\}\ \forall$ faces
QT/QN	$U_2(K)$	$V_{1.5,2}(K)$	edge face	6 4	$\nabla (\lambda_i \lambda_j [\lambda_i - \lambda_j]) \quad \{i < j\}$ $\nabla (\lambda_i \lambda_j \lambda_k) \quad \{i,j,k\} = \text{Face nodes}\{1,2,3\}\ \forall$ faces

C = Constant; L = Linear; Q = Quadratic; T = Tangential; N = Normal.

Appendix D

Integration schemes

D.1 Tables of elemental quadrature rules

This section lists the symmetric quadrature rules used on elemental faces and volumes. The rules are specified in terms of simplex coordinates. If they were not symmetric, one would not be able to express them in terms of simplex coordinates.

D.1.1 Triangular surfaces

Gaussian quadrature rules for triangular surfaces presented in [74]. Kernel evaluation points are specified in terms of simplex coordinates and every point is assigned a weight. A rule is applied as follows:

$$\int_S f(\lambda_1, \lambda_2, \lambda_3) dS = \text{Area}(S) \sum_{i=1}^{N_Q} w^i f(\lambda_1^i, \lambda_2^i, \lambda_1^i), \tag{D.1}$$

where the ith integration point has simplex coordinates $(\lambda_1^i, \lambda_2^i, \lambda_3^i)$ and is weighted with w^i.

All rules used are stated in Table D.1. These rules are symmetric, which is the reason for the multiplicity column, indicating the number of permutations of the coordinate values, all with the same weight.

D.1.2 Tetrahedral volumes

Kernel evaluation points are specified in terms of simplex coordinates and every point is assigned a weight. A rule is applied as follows:

$$\int_V f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, dV = \text{Volume}(V) \sum_{i=1}^{N_Q} w^i f(\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i), \tag{D.2}$$

where $(\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i)$ and w^i represent the ith quadrature point and weight.

Table D.2 presents a 4-point, second order complete rule from [103] and a 11-point, fourth order complete rule from [109]. These rules are symmetric, which is the reason for the multiplicity column, indicating the number of permutations of the coordinate values, all with the same weight.

Table D.1: Symmetric quadrature rules for integration on a triangle. N indicates the polynomial order to which the rule is complete. M indicates the quadrature point multiplicity.

$N = 3$:

M	λ_1	λ_2	λ_3	w
1	0.3333333333	0.3333333333	0.3333333333	−0.5625
3	0.6	0.2	0.2	0.5208333333

$N = 5$:

M	λ_1	λ_2	λ_3	w
1	0.3333333	0.3333333	0.3333333	0.225
3	0.059715872	0.470142064	0.470142064	0.132394153
3	0.797426985	0.10128651	0.10128651	0.1259391805

$N = 7$:

M	λ_1	λ_2	λ_3	w
1	0.3333333	0.3333333	0.3333333	−0.1495700444677
3	0.4793080678	0.26034596608	0.26034596608	0.175615257433
3	0.869739794196	0.065130102902	0.065130102902	0.053347235609
6	0.048690315425	0.312865496005	0.6384441885698	0.0771137608903

D.2 Calculation of $[K^e]$

This section discusses the analytical- and numerical evaluation of the elemental FE integral (3.25).

Considering the analytical case first. Note that a vector function in terms of simplex coordinates, representing a generic term of the basis functions defined in Appendix C, along with its curl, can be expressed as

$$\mathbf{F} = F(\lambda)\nabla\lambda_\alpha \tag{D.3}$$

$$\begin{aligned}\nabla\times\mathbf{F} &= \nabla\times F(\lambda)\nabla\lambda_\alpha \\ &= \nabla F(\lambda)\times\nabla\lambda_\alpha \\ &= T_1(\lambda)(\nabla\lambda_1\times\nabla\lambda_\alpha)+T_2(\lambda)(\nabla\lambda_2\times\nabla\lambda_\alpha) \\ &\quad +T_3(\lambda)(\nabla\lambda_3\times\nabla\lambda_\alpha)+T_4(\lambda)(\nabla\lambda_4\times\nabla\lambda_\alpha),\end{aligned} \tag{D.4}$$

where F, T_1, T_2, T_3 and T_4 are polynomials in λ (that is $\lambda_1,...,\lambda_4$) and $\alpha\in\{1,2,3,4\}$. Observed that both $\mathbf{F}$ and $\nabla\times\mathbf{F}$ consist of terms that are products of a polynomial in λ and a constant vector. The products $\mathbf{F}\cdot\mathbf{F}$ and $\nabla\times\mathbf{F}\cdot\nabla\times\mathbf{F}$ will thus be polynomials in λ.

The following general integration formula for simplex coordinates on tetrahedrons [103] can be used to evaluate these integrals of polynomials in λ over the elemental volume:

$$\int_{V^e}\lambda_1^i\lambda_2^j\lambda_3^k\lambda_4^l\,dV=\frac{3!\;i!\;j!\;k!\;l!}{(3+i+j+k+l)!}V. \tag{D.5}$$

In [169] this analytical procedure has been applied to vector elements very similar to those used here. In [66] the procedure was employed for calculation of the contributions to the system matrix of some of the basis functions presented in Appendix C — those results, in computer code form, were directly used here.

To evaluate the integral numerically, the quadrature rules for tetrahedral volumes are employed, as defined in Section D.1.2. For elements of polynomial order 1, the 4-point, second order complete rule is used. For elements of polynomial order 2, the 11-point, fourth order complete rule is used.

D.3 Calculation of $[P^{st}]$

This section refers to equation (3.26). The four dimensional integrals must be calculated for every possible combination of surface triangles, st. Triangular surface quadrature is used for both the inner and outer surface integrals when the non-self-terms are considered (no quadrature point lies on the triangular boundary, therefore adjacent faces are included here). However the self-terms require special attention, because then the inner kernels will become singular.

For the self-terms, one possibility is to choose two sets of non-coinciding quadrature points, but then great uncertainty exists about the accuracy around the singularity. A techniques of which the accuracy is stable is needed. Another option is to subtract the singularity [99, 200], leaving two kernel terms. A well behaved part $G_0(\mathbf{r},\mathbf{r}') - \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}$ that can be handled with non-overlapping quadrature rules for the inner and outer integrals, and a singular part $\frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}$ that can possibly be integrated analytically with techniques similar to those presented in [200, 95], provided that the basis functions used here can successfully be taken into account. Still another technique is used in [119], where the inner integral's domain is divided into three triangles and these domains are in turn transformed to squares.

The technique proposed here for the self-terms is different from the above, since it removes the singularity completely, before integration. This is achieved by transforming to a cylindrical coordinate system. This procedure has the distinct advantage that an extension to basis functions of arbitrary polynomial order is trivial. The technique presented here is based on a simpler version presented in [124].

D.3.1 Evaluating the inner surface integrals in the self-term case

The inner integral is converted to a polar coordinate system. This results in a well behaved kernel for the inner integral. The procedure is as follows:

$$I_{\text{inner}} = 4\pi \int_{S^t} H(x',y')\, G_0 \, dS' \tag{D.6}$$

where the function $H(x',y')$ represents either $\frac{1}{4\pi}(\nabla' \cdot \hat{z} \times \mathbf{N}^t(\mathbf{r}'))$ or a component of $\frac{1}{4\pi}(\hat{z} \times \mathbf{N}^t(\mathbf{r}'))$ (see equation (3.26)). $\mathbf{N}^t(\mathbf{r}')$ represents any basis function with a tangential component in face t.

Now make the following substitutions:

$$\begin{aligned} X &= x - x' \\ Y &= y - y' \end{aligned} \quad . \tag{D.7}$$

The integral can be rewritten as

$$I_{\text{inner}} = \int\int_{S_T^t} H(X,Y) \frac{e^{-jk_0\sqrt{X^2+Y^2}}}{\sqrt{X^2+Y^2}} \, dXdY \tag{D.8}$$

where S_T^t represents the translated area of face t. Now convert to polar coordinates ($X \leftarrow r\cos\theta$, $Y \leftarrow r\sin\theta$, $r \leftarrow \sqrt{X^2+Y^2}$ and $dXdY \leftarrow rdrd\theta$):

$$I_{\text{inner}} = \int\int_{S_T^t} H(r\cos\theta, r\sin\theta)\, e^{-jk_0 r} \, drd\theta. \tag{D.9}$$

When $s = t$, the center of the (r,θ) coordinate system lies within S_T^t, the integration in r is carried out from this center to the edge of S_T^t and repeated at incremented values of θ until the whole S_T^t is covered once. Should one wish to use this technique in the $s \neq t$ case, then the r integration is carried out between the two crossing points of the line of integration (at angle θ) and the edge of S_T^t. In this case the range of θ is obviously $\theta_{\text{range}} \leq \pi$, where in the previous case it is $\theta_{\text{range}} = 2\pi$.

Inspection of the two possible functions H in equation (3.26) reveals that in both cases H itself or its components can be represented by the following general form (due to the elemental basis functions employed):

$$H(X,Y) = c_5X^2 + c_4Y^2 + c_3XY + c_2X + c_1Y + c_0, \tag{D.10}$$

where $c_0, ..., c_5$ are constants. Rewriting this in the (r,θ) coordinate system and setting $\theta = \theta_i$ (the current value of θ at which the radial integral is being evaluated), results in

$$H(r,\theta_i) = Ar^2 + Br + C, \tag{D.11}$$

where A, B and C are constants. Substituting this form for H into equation (D.9), the r integration can be carried out analytically, yielding

$$\begin{aligned} \int_{r_1}^{r_2} (Ar^2 + Br + C)e^{-jk_0 r} dr \;=\; & A\left[\frac{r^2e^{-jk_0r}}{(-jk_0)} - \frac{2re^{-jk_0r}}{(-jk_0)^2} + \frac{2e^{-jk_0r}}{(-jk_0)^3}\right]_{r_1}^{r_2} \\ & + B\left[\frac{re^{-jk_0r}}{(-jk_0)} - \frac{e^{-jk_0r}}{(-jk_0)^2}\right]_{r_1}^{r_2} \\ & + C\left[\frac{e^{-jk_0r}}{(-jk_0)}\right]_{r_1}^{r_2}. \end{aligned} \tag{D.12}$$

The constant coefficients can be calculated from the evaluation of H at three arbitrary, distinct locations (chosen as r_1, $\frac{r_1+r_2}{2}$ and r_2) on the $\theta = \theta_i$ line, as follows:

$$\begin{Bmatrix} A \\ B \\ C \end{Bmatrix} = \begin{bmatrix} r_1^2 & r_1 & 1 \\ \left(\frac{r_1+r_2}{2}\right)^2 & \left(\frac{r_1+r_2}{2}\right) & 1 \\ r_2^2 & r_2 & 1 \end{bmatrix}^{-1} \begin{Bmatrix} H(r_1,\theta_i) \\ H\left(\frac{r_1+r_2}{2},\theta_i\right) \\ H(r_2,\theta_i) \end{Bmatrix} \tag{D.13}$$

Since the r integration is carried out analytically, I_{inner} can be evaluated as a one dimensional integral over θ, using a trapezoidal rule. A higher order rule is not used for the θ integration, because the kernel's first derivative is discontinuous at the corners of S_T^t. Since this elaborate procedure is only employed in the self-term case, computational efficiency is not that important, therefore a large number of θ integration points is employed (100), to ensure convergence for any of the basis functions considered.

D.3.2 General case

The outer integrals and the inner integrals in the non-self-term case are evaluated using the triangular surface quadrature rules defined in Section D.1.1. It was found that the observable quantities converge completely when using a $N = 3$ order complete rule with basis functions of polynomial order 1 and when using a $N = 5$ order complete rule with basis functions of polynomial order 2.

Table D.2: Symmetric quadrature rules for integration on a tetrahedron. N indicates the polynomial order to which the rule is complete. M indicates the quadrature point multiplicity.

$N = 2$:

M	λ_1	λ_2	λ_3	λ_4	w
4	0.585410196624968	0.138196601125015	0.138196601125015	0.138196601125015	0.25

$N = 4$:

M	λ_1	λ_2	λ_3	λ_4	w
1	0.25	0.25	0.25	0.25	$-6 \times 0.0131555555555555$
4	0.0714285714285714	0.0714285714285714	0.0714285714285714	0.785714285714285	$6 \times 0.00762222222222222$
6	0.399403576166799	0.399403576166799	0.100596423833200	0.100596423833200	$6 \times 0.0248888888888888$

Appendix E

FEM modeling of a coaxial port

E.1 Introduction

This appendix describes the modeling of a coaxial port located within a PEC boundary of the FEM domain, as shown in Figure E.1. The model is based on the assumption that only the dominant, TEM coax mode with zero cutoff frequency is present at the port. Therefore, the coaxial aperture must be small enough that it is operating below the cutoff frequency of the first higher order mode. The cutoff wavelength of the first higher order mode is of the order of the arithmetic mean of the inner conductor and the outer conductor circumferences [88].

Similar procedures can be found in the literature. [159] presents the same formulation with provisions for higher order modes, but little details are provided. Only CT/LN elements are considered. [85, 192] consider only the dominant mode. A stationary functional approach is followed. The coaxial aperture triangulation is restricted to a hexagon comprising of six equilateral faces (thus the inner conductor is always modeled as infinitely thin) and the formulation is restricted to CT/LN elements. [99, 159] consider similar formulations for rectangular and circular waveguide ports, again only implemented with CT/LN elements.

With a FE-BI formulation, the coaxial aperture could be incorporated exactly (similar to the handling of the cavity backed aperture in Chapter 3): apply an appropriate form of the equivalence principle and employ the coaxial Green functions provided in [187]. This is excessive except if one is interested in the analysis of electrically large coaxial apertures.

Section E.2 describes s- and z parameters for general, multi-, TEM port devices. Procedures for arbitrary transformations are derived. Section E.3 presents the dominant mode coaxial port model, which is an application of the dominant mode, waveguide port procedure of [99], to the coaxial geometry. Section E.4 describes the formulation of [85, 192] and shows that it is a special case of the procedure in Section E.3.

In Sections E.3 and E.4, let the subscript c indicate association with the coaxial feed line and aperture.

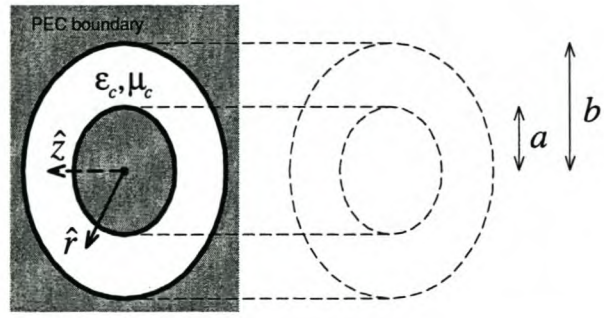


Figure E.1: General coaxial port geometry.

E.2 A note on s- and z parameters for multi-, TEM port devices

This section is an extension of material found in [148, 57, 201].

Along any TEM, two conductor transmission line, the voltage and current can be unambiguously established [154]. The voltage and current can be expressed in terms of forward- and backward propagating waves, as follows:

$$V = V^+ e^{-jkz} + V^- e^{+jkz} \tag{E.1}$$
$$I = I^+ e^{-jkz} + I^- e^{+jkz} \tag{E.2}$$

with

$$Z^c I^+ = V^+ \tag{E.3}$$
$$-Z^c I^- = V^- \tag{E.4}$$

where Z^c is the characteristic impedance [93, 154].

Suppose one has a N port device. Port n is connected to a TEM transmission line with characteristic impedance Z_n^c, with scaled input and output wave amplitudes defined as

$$a_n = \frac{V_n^+}{\sqrt{Z_n^c}} \tag{E.5}$$
$$b_n = \frac{V_n^-}{\sqrt{Z_n^c}}. \tag{E.6}$$

The scaling is included so that the squares of the scaled amplitude at all the ports are directly proportional to power (with proportionality constant $\frac{1}{2}$ in this case).

The s parameters of the device are defined as follows:

$$\{b\} = [S]\{a\} \tag{E.7}$$

with

$$\{a\} = \begin{Bmatrix} a_1 \\ \vdots \\ a_N \end{Bmatrix}; \quad \{b\} = \begin{Bmatrix} b_1 \\ \vdots \\ b_N \end{Bmatrix} \tag{E.8}$$

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0\ \{k\neq j\}}. \tag{E.9}$$

The z parameters of the device are defined as follows:

$$\{V\} = [Z]\{I\} \tag{E.10}$$

with

$$\{I\} = \begin{Bmatrix} I_1 \\ \vdots \\ I_N \end{Bmatrix}; \quad \{V\} = \begin{Bmatrix} V_1 \\ \vdots \\ V_N \end{Bmatrix} \tag{E.11}$$

$$Z_{ij} = \left.\frac{V_i}{I_j}\right|_{I_k=0\ \{k\neq j\}}. \tag{E.12}$$

Note the following relations:

$$\{V\} = \left[\sqrt{Z^c}\right](\{a\}+\{b\}) \tag{E.13}$$

$$\{I\} = \left[\sqrt{Z^c}\right]^{-1}(\{a\}-\{b\}) \tag{E.14}$$

$$\{a\} = \frac{1}{2}\left(\left[\sqrt{Z^c}\right]^{-1}\{V\} + \left[\sqrt{Z^c}\right]\{I\}\right) \tag{E.15}$$

$$\{b\} = \frac{1}{2}\left(\left[\sqrt{Z^c}\right]^{-1}\{V\} - \left[\sqrt{Z^c}\right]\{I\}\right), \tag{E.16}$$

with

$$\left[\sqrt{Z^c}\right] = \begin{bmatrix} \sqrt{Z_1^c} & & \\ & \ddots & \\ & & \sqrt{Z_N^c} \end{bmatrix}. \tag{E.17}$$

Transform from s parameters to z parameters as follows:

$$\begin{aligned} \{b\} &= [S]\{a\} \\ \left[\sqrt{Z^c}\right]^{-1}\{V\} - \left[\sqrt{Z^c}\right]\{I\} &= [S]\left[\sqrt{Z^c}\right]^{-1}\{V\} + [S]\left[\sqrt{Z^c}\right]\{I\} \\ [U-S]\left[\sqrt{Z^c}\right]^{-1}\{V\} &= [U+S]\left[\sqrt{Z^c}\right]\{I\}, \end{aligned} \tag{E.18}$$

thus

$$[Z] = \left[\sqrt{Z^c}\right][U-S]^{-1}[U+S]\left[\sqrt{Z^c}\right]. \tag{E.19}$$

Transform from z parameters to s parameters as follows:

$$\begin{aligned} \{V\} &= [Z]\{I\} \\ \left[\sqrt{Z^c}\right](\{a\}+\{b\}) &= [Z]\left[\sqrt{Z^c}\right]^{-1}(\{a\}-\{b\}) \\ \left[[Z]\left[\sqrt{Z^c}\right]^{-1} + \left[\sqrt{Z^c}\right]\right]\{b\} &= \left[[Z]\left[\sqrt{Z^c}\right]^{-1} - \left[\sqrt{Z^c}\right]\right]\{a\}, \end{aligned} \tag{E.20}$$

thus

$$[S] = \left[[Z]\left[\sqrt{Z^c}\right]^{-1} + \left[\sqrt{Z^c}\right]\right]^{-1}\left[[Z]\left[\sqrt{Z^c}\right]^{-1} - \left[\sqrt{Z^c}\right]\right]. \tag{E.21}$$

Clearly the s parameters are dependent on the characteristic impedances of the systems connected to the ports, while the z parameters are fixed. By using the equations presented in this section, one can transform a given set of s parameters to another set of s parameters, with arbitrarily specified port, characteristic impedances. This is accomplished by first transforming the given set of s parameters to z parameters, employing $[\sqrt{Z^c}]$ associated with the given s parameters. Then transform back to s parameters, using a $[\sqrt{Z^c}]$, constructed from the desired, new set of port, characteristic impedances.

E.3 Dominant mode, coaxial port model

E.3.1 Formulation

The coaxial port is incorporated into the variational boundary value problem via an inhomogeneous Neumann boundary condition.

According to Maxwell's equations, $\hat{n} \times \frac{1}{\mu_r} \nabla \times \mathbf{E}$ is continuous across any interface where a surface current is not present, such as the coax aperture. Calculate $\mathbf{N}_c$ at the coax aperture, on the coax side, as follows:

$$\mathbf{N}_c = \hat{n} \times \nabla \times \mathbf{E}_c, \tag{E.22}$$

where $\hat{n}$ is the outward pointing normal to the FEM region, as in Chapter 2. Then the continuity condition at the aperture implies that

$$\frac{1}{\mu_r}\mathbf{N} = \frac{1}{\mu_{rc}}\mathbf{N}_c, \tag{E.23}$$

where $\mathbf{N}$ is as defined in equation (2.3). Thus the following term must be added to the linear form of equation (2.13) in order to model the coax aperture:

$$-\int_{S_c} \frac{1}{\mu_{rc}}\mathbf{N}_c \cdot \mathbf{W}\, dS. \tag{E.24}$$

Next, assume a local, cylindrical coordinate system centered at the center of the coax aperture. $\hat{z}$ points in the propagating direction of the incident wave, i.e. $\hat{z} = -\hat{n}$.

Assuming that the electromagnetic fields in the aperture is the sum of purely TEM incident and reflected parts, one starts with the specification that the phasor value of the coaxial current wave traveling into the cavity, at the aperture, is I_0. The magnetic and electric fields within the coax follows as (see [93]):

$$\begin{aligned} \mathbf{H}_c &= \frac{I_0}{2\pi}(e^{-jk_c z} - \Gamma e^{jk_c z})\frac{1}{r}\hat{\phi} \\ &= \mathbf{H}_c^{\text{inc}} + \mathbf{H}_c^{\text{ref}} \qquad \text{(E.25)} \\ \mathbf{E}_c &= \frac{I_0\sqrt{\mu_c}}{2\pi\sqrt{\epsilon_c}}(e^{-jk_c z} + \Gamma e^{jk_c z})\frac{1}{r}\hat{r} \\ &= \mathbf{E}_c^{\text{inc}} + \mathbf{E}_c^{\text{ref}} \qquad \text{(E.26)} \end{aligned}$$

where Γ represents the reflection coefficient seen by the cable at the aperture, and

$$k_c = \frac{2\pi\sqrt{\epsilon_{rc}\mu_{rc}}}{\lambda_0} \tag{E.27}$$

represents the propagation constant inside the cable.

Note that due to the transverse nature of the electric field, the following properties hold:

$$-\sqrt{\frac{\epsilon_c}{\mu_c}}\mathbf{E}_c^{\text{inc}} = \hat{z} \times \mathbf{H}_c^{\text{inc}} \tag{E.28}$$

$$\sqrt{\frac{\epsilon_c}{\mu_c}}\mathbf{E}_c^{\text{ref}} = \hat{z} \times \mathbf{H}_c^{\text{ref}}. \tag{E.29}$$

Next, calculate $\mathbf{N}_c$:

$$\begin{aligned}
\mathbf{N}_c &= \hat{n} \times (\nabla \times \mathbf{E}_c)|_{z=0} \\
&= -\hat{z} \times (\nabla \times \mathbf{E}_c)|_{z=0} \\
&= j\omega\mu_c \left(\hat{z} \times \mathbf{H}_c^{\text{inc}} + \hat{z} \times \mathbf{H}_c^{\text{ref}}\right)\Big|_{z=0} \\
&= jk_c \left(\mathbf{E}_c^{\text{ref}} - \mathbf{E}_c^{\text{inc}}\right)\Big|_{z=0} \\
&= jk_c \left(\mathbf{E}_c - 2\mathbf{E}_c^{\text{inc}}\right)\Big|_{z=0}.
\end{aligned} \tag{E.30}$$

In equation (E.30), $\mathbf{E}_c^{\text{inc}}$ is known, but $\mathbf{E}_c$ is unknown. Rewrite $\mathbf{E}_c$ in terms of $\mathbf{E}$ by applying the continuity condition $\hat{n} \times \mathbf{E}_c = \hat{n} \times \mathbf{E}$ (assuming that only the dominant mode is present at the aperture):

$$\begin{aligned}
\mathbf{N}_c &= jk_c \left(\mathbf{E}_c - 2\mathbf{E}_c^{\text{inc}}\right)\Big|_{z=0} \\
&= jk_c \left(2\hat{n} \times (\hat{n} \times \mathbf{E}_c^{\text{inc}}) - \hat{n} \times (\hat{n} \times \mathbf{E}_c)\right)\Big|_{z=0} \\
&= jk_c \left(2\hat{n} \times (\hat{n} \times \mathbf{E}_c^{\text{inc}}) - \hat{n} \times (\hat{n} \times \mathbf{E})\right)\Big|_{z=0}
\end{aligned} \tag{E.31}$$

where the tangential nature of $\mathbf{E}_c^{\text{inc}}$ and $\mathbf{E}_c$ was employed.

Substitute equation (E.31) into equation (E.24), yielding

$$\begin{aligned}
-\int_{S_c} \frac{1}{\mu_{rc}} \mathbf{N}_c \cdot \mathbf{W}\, dS &= -\frac{jk_c}{\mu_{rc}} \int_{S_c} \left(2\hat{n} \times (\hat{n} \times \mathbf{E}_c^{\text{inc}}) - \hat{n} \times (\hat{n} \times \mathbf{E})\right) \cdot \mathbf{W}\, dS \\
&= \frac{jk_c}{\mu_{rc}} \int_{S_c} \left(2\hat{n} \times \mathbf{E}_c^{\text{inc}} - \hat{n} \times \mathbf{E}\right) \cdot (\hat{n} \times \mathbf{W})\, dS
\end{aligned} \tag{E.32}$$

where identity (A.8) was employed.

Clearly, there are in fact contributions to both the linear and the bilinear forms of the variational boundary value problem (2.10). After discretization (described in Section 2.4), the following elemental contributions to the system matrix equation results ($[A_C]$ must be added to $[A]$ and $\{b_C\}$ to $\{b\}$):

$$[A_C^s] = \frac{jk_c}{\mu_{rc}} \int_{S^s} \{\hat{n} \times \mathbf{N}^s\} \cdot \{\hat{n} \times \mathbf{N}^s\}^T\, dS, \tag{E.33}$$

$$\{b_C^s\} = \frac{2jk_c}{\mu_{rc}} \int_{S^s} \left(\hat{n} \times \mathbf{E}_c^{\text{inc}}\right) \cdot \{\hat{n} \times \mathbf{N}^s\}\, dS \tag{E.34}$$

where s refers to a triangular face in the coax aperture. $\{\mathbf{N}^s\}$ represents the column vector of the vector basis functions associated with the element to which the face s belongs. Only those basis functions with components tangential to face s need to be included.

E.3.2 Calculation of port parameters

Define the dominant, TEM, coaxial, electric field, normalized mode as follows:

$$\mathbf{e}^E = \frac{a}{r}\hat{r}. \tag{E.35}$$

with the property

$$\int_{S_c} \mathbf{e}^E \cdot \mathbf{e}^E \, dS = 2\pi a^2 \ln(b/a) \tag{E.36}$$

and a and b as defined in Figure E.1.

The dominant mode, incident and reflected fields can be written in terms of $\mathbf{e}^E$, resulting in

$$\mathbf{E}_c^{\text{inc}} = E^+ e^{-jk_c z}\mathbf{e}^E \tag{E.37}$$

$$\mathbf{E}_c^{\text{ref}} = E^- e^{+jk_c z}\mathbf{e}^E. \tag{E.38}$$

By the orthogonality of the coaxial transmission line modes with respect to an inner product over the coaxial aperture [92], one can extract the amplidude of the dominant mode amplitude at the port. The incident and reflected wave amplitudes are thus calculated as follows:

$$E^+ = \frac{I_0\sqrt{\mu_c}}{2\pi a\sqrt{\epsilon_c}} \tag{E.39}$$

$$\begin{aligned} E^- &= \frac{1}{2\pi a^2 \ln(b/a)} \int_{S_c} \left(\mathbf{E}_h - \mathbf{E}_c^{\text{inc}}\right) \cdot \mathbf{e}^E \, dS \\ &= \frac{1}{2\pi a^2 \ln(b/a)} \int_{S_c} \mathbf{E}_h \cdot \mathbf{e}^E \, dS - E^+ \\ &= \frac{1}{2\pi a^2 \ln(b/a)} \int_{S_c} (\hat{n} \times \mathbf{E}_h) \cdot (\hat{n} \times \mathbf{e}^E) \, dS - E^+ \end{aligned} \tag{E.40}$$

where $\mathbf{E}_h$ is the approximate, FEM solution. The last step employs the transverse property of $\mathbf{e}^E$ and equation (A.8). The last step was performed in order to write the integral in terms of the same basis function operations as in equations (E.33) and (E.34), for convenience of implementation.

The transmission line voltage wave amplitudes can be obtained from the electric field, using the following relation (derived from Faraday's law and the definition of potential, see [97, 154]):

$$\begin{aligned} V = V_a - V_b &= -\int_b^a \mathbf{E} \cdot \hat{l} \, dl \\ &= -E \int_b^a \mathbf{e}^E \cdot \hat{r} \, dr \\ &= a \ln(b/a) E. \end{aligned} \tag{E.41}$$

Therefore

$$V^+ = a \ln(b/a) E^+ \tag{E.42}$$

$$V^- = a \ln(b/a) E^- \tag{E.43}$$

$$Z^c = \frac{\ln(b/a)}{2\pi}\sqrt{\frac{\mu_c}{\epsilon_c}} \tag{E.44}$$

are the desired values needed for every port, in order to obtain the s- and/or z parameters of the simulated device, using the procedures described in Section E.2. The characteristic impedance equation can be found in [154] (note that the superscript c does not indicate the coaxial context).

E.4 Dominant mode, coaxial port model restricted to Whitney elements

This section relates the port model presented in [85, 192], which requires the use of Whitney (CT/LN) elements and a hexagonal descretization in the port aperture, consisting of 6 equilateral faces (as mentioned earlier). It is shown that this approach is a special case of the approach presented in Section E.3. The stationary functional approach is used, but it is also briefly discussed, how to obtain the same results with the VBVP formulation.

E.4.1 Formulation

Substitute equation (E.30) into equation (E.24), yielding

$$-\int_{S_c} \frac{1}{\mu_{rc}} \mathbf{N}_c \cdot \mathbf{W}\, dS = \frac{jk_c}{\mu_{rc}} \int_{S_c} \left(2\mathbf{E}_c^{\text{inc}} - \mathbf{E}_c\right) \cdot \mathbf{W}\, dS. \tag{E.45}$$

Thus the following term must be added to the functional to be rendered stationary (see Section 2.3):

$$F_c = \frac{jk_c}{\mu_{rc}} \int_{S_c} \left(\frac{1}{2}\mathbf{E}_c - 2\mathbf{E}_c^{\text{inc}}\right) \cdot \mathbf{E}_c\, dS. \tag{E.46}$$

At this point it must be noted that the coax aperture is discretized with 6 equilateral, triangular faces forming a hexagon. Whitney elements are used.

The Maxwell continuity condition at the port states that

$$\hat{n} \times \mathbf{E}|_{S_c} = \hat{n} \times \mathbf{E}_c|_{S_c}. \tag{E.47}$$

In order to relate $\mathbf{E}_c$ to the FE electric field representation, equation (E.47) obviously cannot be used directly; it is used indirectly by equating the potential differences on the two sides of the aperture (which implicitly selects only the tangential components, as was done explicitly in equation (E.31)), as follows:

$$\begin{aligned}
\int_a^b \mathbf{E}_h|_{S_c} \cdot \hat{r}\, dr &= \int_a^b \mathbf{E}_c|_{S_c} \cdot \hat{r}\, dr \\
\int_a^b \frac{E_i}{b}\, dr &= \int_a^b (\mathbf{E}_c|_{S_c})_r\, dr \\
\frac{(b-a)}{b} E_i &= r(\mathbf{E}_c|_{S_c})_r \ln(b/a) \\
\Rightarrow \quad (\mathbf{E}_c|_{S_c})_r &= \frac{(b-a)E_i}{rb\ln(b/a)} \\
&= \frac{(b-a)\left(\sum_{l=1}^{6} E_l\right)}{6b\ln(b/a)} \frac{1}{r}.
\end{aligned} \tag{E.48}$$

The second line is obtained by noting that the integral is carried out along one of the 6 aperture edges and then employing the constant tangential property of the Whitney elements [99, eq. (8.55)]. The final result is obtained by assuming that all 6 degrees of freedom in the aperture are equal and thus

$$E_i = \frac{1}{6}\sum_{l=1}^{6} E_l. \tag{E.49}$$

Noting that

$$\mathbf{E}_c^{\text{inc}}|_{S_c} = \frac{I_0\sqrt{\mu_c}}{2\pi\sqrt{\epsilon_c}}\frac{1}{r}\hat{r}, \tag{E.50}$$

equation (E.46) is rewritten with substitution of equations (E.48) and (E.50), yielding

$$\begin{aligned} F_c &= \frac{jk_c}{\mu_{rc}}\int_0^{2\pi}\int_a^b \left[\left(\frac{(b-a)\left(\sum_{l=1}^6 E_l\right)}{12b\ln(b/a)}\frac{1}{r}\hat{r} - \frac{I_0\sqrt{\mu_c}}{\pi\sqrt{\epsilon_c}}\frac{1}{r}\hat{r}\right)\cdot\frac{(b-a)\left(\sum_{l=1}^6 E_l\right)}{6b\ln(b/a)}\frac{1}{r}\hat{r}\right] r dr d\phi \\ &= \frac{j\pi k_c(b-a)}{3\mu_{rc}b\ln(b/a)}\int_a^b \left[\frac{(b-a)\left(\sum_{l=1}^6 E_l\right)^2}{12b\ln(b/a)}\frac{1}{r} - \frac{I_0\sqrt{\mu_c}}{\pi\sqrt{\epsilon_c}}\frac{\left(\sum_{l=1}^6 E_l\right)}{r}\right] dr \\ &= \frac{j\pi k_c(b-a)^2}{36\mu_{rc}b^2\ln(b/a)}\left(\sum_{l=1}^6 E_l\right)^2 - \frac{jk_cI_0(b-a)\sqrt{\mu_c}}{3\mu_{rc}b\sqrt{\epsilon_c}}\left(\sum_{l=1}^6 E_l\right). \end{aligned} \tag{E.51}$$

Note that the integration is carried out over a circular aperture, rather than the actual, hexagonal, aperture mesh.

Rendering F_c stationary, results in

$$\begin{aligned} \frac{\partial F_c}{\partial E_i} &= \frac{j\pi k_c(b-a)^2}{18\mu_{rc}b^2\ln(b/a)}\left(\sum_{l=1}^6 E_l\right) - \frac{jk_cI_0(b-a)\sqrt{\mu_c}}{3\mu_{rc}b\sqrt{\epsilon_c}} \\ &= \frac{j\pi k_c(b-a)^2}{3\mu_{rc}b^2\ln(b/a)}E_i - \frac{jk_cI_0(b-a)\sqrt{\mu_c}}{3\mu_{rc}b\sqrt{\epsilon_c}}, \end{aligned} \tag{E.52}$$

where equation (E.49) was used to obtain the last equation.

Thus the following additions must be made to the system matrix equation:

$$A(g_i, g_i) = A(g_i, g_i) + \frac{j\pi k_c(b-a)^2}{3\mu_{rc}b^2\ln(b/a)} \quad i = 1, .., 6 \tag{E.53}$$

$$b(g_i) = b(g_i) + \frac{jk_cI_0(b-a)\sqrt{\mu_c}}{3\mu_{rc}b\sqrt{\epsilon_c}} \quad i = 1, .., 6 \tag{E.54}$$

where g_i is the global degree of freedom number associated with the local coax aperture edge i.

If Whitney elements scaled by their associated edge lengths are used (as defined in [99]), then the b^2, b dividing factors must be replaced by unity in the above two equations, which then correspond to the results obtained in [85, 192].

The same results can be obtained with the VBVP approach, as follows: Start with equation (E.45). By the assumption of equal aperture degrees of freedom and the CT/LN basis function properties, use $\mathbf{W}_h^i \approx \frac{1}{br}\hat{r}$ with $\phi[\frac{\pi}{3}(i-1), \frac{\pi}{3}i]$ and $i = 1, .., 6$. Use $\mathbf{E}_c = (\mathbf{E}_c|_{S_c})_r\hat{r}$ from equation (E.48). Carry out the integration in (E.45) over a circular aperture. Equations (E.53) and (E.54) result again.

E.4.2 Calculation of port parameters

One needs V^+, V^- and Z^c for every port, in order to calculate the s and/or z parameters of the (multi) coaxial port device, as described in Section E.2. As stated earlier,

$$Z^c = \frac{\ln(b/a)}{2\pi}\sqrt{\frac{\mu_c}{\epsilon_c}}. \tag{E.55}$$

Since $I_0 = I^+$, equation (E.3) is used to obtain

$$V^+ = I_0 Z^c. \tag{E.56}$$

Since the assumption was made that the degrees of freedom of the 6 coaxial aperture edges are equal, the total coaxial aperture voltage is taken to be the average of the voltages implied by every edge (see equations (E.41) and (E.48)), that is

$$V = \frac{(b-a)}{6b}\sum_{l=1}^{6} E_l \tag{E.57}$$

where the dividing factor of b must be set equal to unity if Whitney elements scaled by edge lengths are used.

With V known, V^- be calculated as

$$V^- = V - V^+. \tag{E.58}$$

As a consequence of the restriction on aperture discretization, the coax centre conductor is modeled as infinitely thin. It was found that a ratio of at least $a : b \equiv 1 : 100$ should be used for the analytical and numerical models to converge. In the case of Section E.3, the center conductor cross section must be modeled as a circular area (radius a) in the boundary of the computational domain, on which a homogeneous Dirichlet boundary condition is applied.

Appendix F

Empirical FMM investigation

This chapter is based on the scalar, 3D, free space Green function FMM formulation, as set out in Chapter 4.

F.1 General definitions

Define the true scalar Green function as

$$G_0 = \frac{e^{-jk_0|\mathbf{X}+\mathbf{d}|}}{4\pi|\mathbf{X}+\mathbf{d}|}. \tag{F.1}$$

Define the addition theorem truncated at $L_{\text{tot}} \equiv L+1$ terms, as

$$G_0^L = -\frac{jk_0}{4\pi}\sum_{l=0}^{L}(-1)^l(2l+1)\, j_l(k_0 d)\, h_l^{(2)}(k_0 X)\, P_l(\hat{d}\cdot\hat{X}) \quad \{d<X\}. \tag{F.2}$$

Define the addition theorem truncated at $L_{\text{tot}} \equiv L+1$ terms, with the spectral domain representation of $j_l(k_0 d)\, P_l(\hat{d}\cdot\hat{X})$ and with a M point quadrature rule for the spherical surface (i.e. a truncated version of equation (4.3) with the spherical surface integration carried out numerically), as

$$G_0^{L,M} = -\frac{jk_0}{(4\pi)^2}\sum_{l=0}^{L}(-j)^l(2l+1)\, h_l^{(2)}(k_0 X)\sum_{m=1}^{M} w_{k_0}^m e^{-j\mathbf{k}_0^m\cdot\mathbf{d}}\, P_l(\hat{k}_0^m\cdot\hat{X}) \quad \{d<X\}. \tag{F.3}$$

In this chapter, adopt the following abbreviated notation for a BI matrix entry calculation (see equation (4.4)):

$$P(a,b) = \int_{S_a}\int_{S_b} f(\mathbf{N}_a(\mathbf{r}),\mathbf{N}_b(\mathbf{r}'))\, G_0\, dS'dS \tag{F.4}$$

with

$$f(\mathbf{N}_a(\mathbf{r}),\mathbf{N}_b(\mathbf{r}')) = 2\{\nabla\cdot\hat{z}\times\mathbf{N}_a(\mathbf{r})\}\{\nabla'\cdot\hat{z}\times\mathbf{N}_b(\mathbf{r}')\} - 2k_0^2\{\hat{z}\times\mathbf{N}_a(\mathbf{r})\}\cdot\{\hat{z}\times\mathbf{N}_b(\mathbf{r}')\}. \tag{F.5}$$

Based on these definitions, the following relative Green function- and relative matrix element errors are defined:

- Relative Green function error (with truncation):

$$\varepsilon^L = \left| \frac{G_0^L - G_0}{G_0} \right|. \tag{F.6}$$

- Relative Green function error (with truncation and spherical surface quadrature):

$$\varepsilon^{L,M} = \left| \frac{G_0^{L,M} - G_0}{G_0} \right|. \tag{F.7}$$

- Relative BI matrix element error (with truncation):

$$\varepsilon_P^L = \left| \frac{\int_{S_a} \int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}')) \left[G_0^L - G_0\right] dS' dS}{\int_{S_a} \int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\, G_0\, dS' dS} \right|. \tag{F.8}$$

- Relative BI matrix element error (with truncation and spherical surface quadrature):

$$\varepsilon_P^{L,M} = \left| \frac{\int_{S_a} \int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}')) \left[G_0^{L,M} - G_0\right] dS' dS}{\int_{S_a} \int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\, G_0\, dS' dS} \right|. \tag{F.9}$$

F.2 Explicit formula for the truncation order of the addition theorem

The aim of this section is to obtain an explicit, approximate formula for the minimum number of terms in the 3D FMM addition theorem in terms of all the FMM parameters ($D_{\max}$, $X_{\min}$ and ε^L), by empirical means. The effect of spherical quadrature is not considered.

As noted in Chapter 4, for fixed $d = D_{\max}$, $X = X_{\min}$ and L_{tot}, the largest ε^L results when $\hat{d} \cdot \hat{X} = 1$. Figure F.1 shows plots of ε^L versus L_{tot} for various values $-1 \leq \hat{d} \cdot \hat{X} \leq 1$ with fixed X and d, confirming this result. In the rest of this section, $\hat{d} \cdot \hat{X} = 1$ is always assumed.

The effect of the variables $D_{\max}$, $X_{\min}$ and ε^L on L_{tot} is investigated by starting with various $X_{\min}$ values and noting the number of terms needed for $\varepsilon^L = 10^{-3}$, $\varepsilon^L = 10^{-4}$ and $\varepsilon^L = 10^{-5}$. Table F.1 lists the criteria considered. By generating figures such as Figure F.1, with X set equal to the values in Table F.1 and varying $D_{\max}$ over the range $0 < D_{\max} \leq 4\lambda$, the amount of terms needed for the error levels considered can be obtained graphically. The results are shown in Figure F.2. (L was recorded as a fractional value to facilitate in the analytical approximation of the curves, although the number of orders added is obviously an integer greater than zero.)

From the results presented in Figure F.2, the conclusion is drawn that L_{tot} could be approximated by the sum of a linear part to describe the asymptotic behaviour and an exponential part to approximate the behaviour of L_{tot} as $D_{\max}$ nears zero, thus:

$$L_{\text{tot}}(D_{\max}) = ke^{-\alpha D_{\max}} + mD_{\max} + c, \tag{F.10}$$

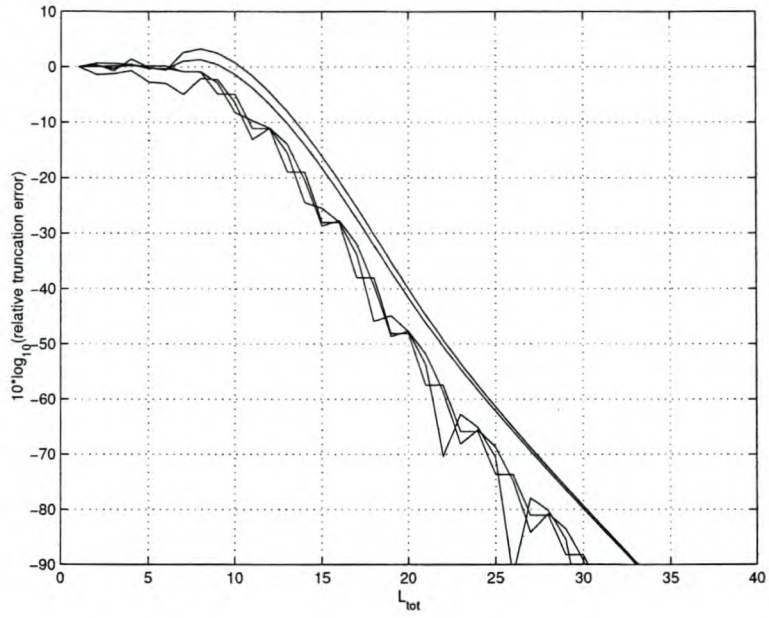


Figure F.1: Example of the relative Green function truncation error versus the amount of terms retained. The various plots correspond to different relative directions of $\mathbf{d}$ and $\mathbf{X}$ while keeping their magnitudes constant. $d = 1.5\lambda$ and $X = 3\lambda$. The worst convergence result corresponds to $\hat{d} \cdot \hat{X} = 1$.

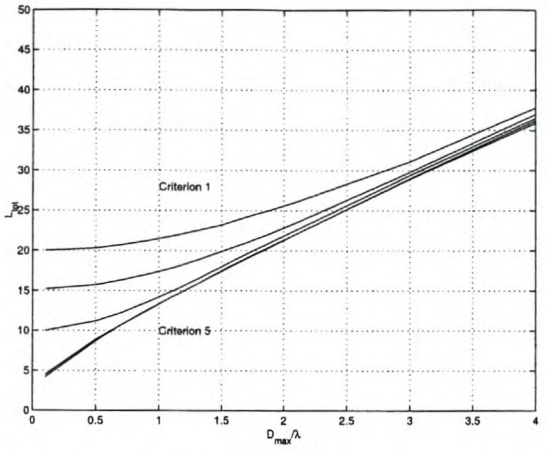
Table F.1: Far interaction criteria considered for the numerical investigation of the FMM truncation error.

Criterion 1	$X_{\min} = \sqrt{2}D_{\max}$
Criterion 2	$X_{\min} = \sqrt{2.5}D_{\max}$
Criterion 3	$X_{\min} = 2D_{\max}$
Criterion 4	$X_{\min} = 6D_{\max}$
Criterion 5	$X_{\min} = 10D_{\max}$

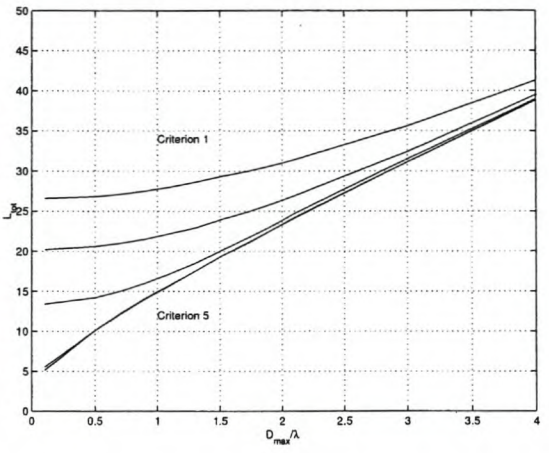
where the coefficients k, α, m and c should be functions of ε^L and $X_{\min}$.

Least squares, first order fits were performed on the asymptotic data to obtain m and c. These least squares fits were subtracted from the original data. The difference is the part to be approximated by an exponential function, therefore another least squares, first order approximation was performed on the natural logarithm of the difference, because $\ln(ke^{-\alpha D_{\max}}) = \ln(k) - \alpha D_{\max}$. The second least squares fit was performed on the range where the exponential part manifests itself the strongest. As an example, Table F.2 shows the results for the first three criteria.

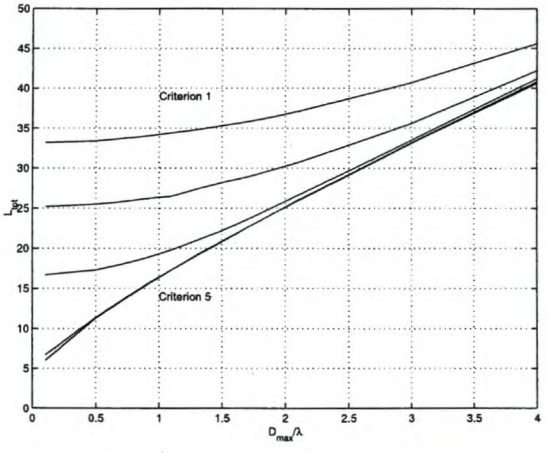
m stays roughly constant and c is approximated as linearly varying with $\log_{10}\varepsilon^L$. It can be observed that the values of α for a certain $X_{\min}$ differs on average by a factor of 0.8. It can also be observed that the values of k for a certain $X_{\min}$ differs on average by a factor of 1.3.



(a) $\varepsilon^L = 10^{-3}$



(b) $\varepsilon^L = 10^{-4}$



(c) $\varepsilon^L = 10^{-5}$

Figure F.2: Results of the numerical investigation into the FMM truncation error. L_{tot} vs. $D_{\max}$ for various error levels.

Table F.2: Least squares fit to the values of L_{tot} for the range $0 \leq \frac{D_{\max}}{\lambda} \leq 4$.

ε	m	c	Criterion 1		Criterion 2		Criterion 3	
			k	α	k	α	k	α
10^{-3}	7.8183	5.5941	14.4335	0.5790	9.6286	0.8763	4.2626	1.8174
10^{-4}	7.8379	7.5717	19.0714	0.4368	12.8211	0.6944	5.2203	1.3875
10^{-5}	7.8231	9.4293	23.9499	0.3514	15.9354	0.5536	7.4369	1.2855

Based on these observation, α and k are expressed as

$$\alpha = K_1 0.8^{-\log_{10}\varepsilon^L} \tag{F.11}$$

$$k = K_2 1.3^{-\log_{10}\varepsilon^L}. \tag{F.12}$$

K_1 and K_2 had to be functions of only

$$x' \equiv \frac{X_{\min}}{D_{\max}}, \tag{F.13}$$

which is the only variable not yet used.

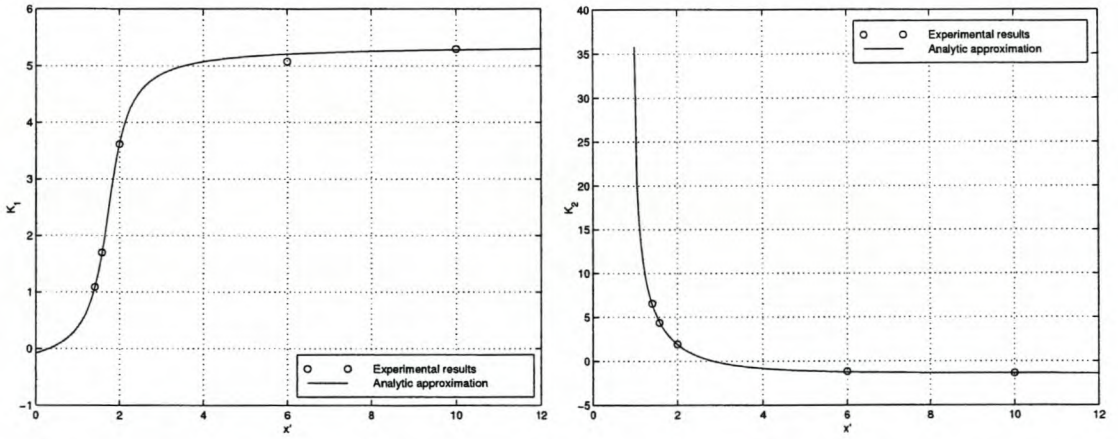
Various functions were experimented with and efforts were guided by the observations that

$$\lim_{x'\to\infty} \frac{dK_1}{dx'} = 0 \tag{F.14}$$

$$\lim_{x'\to\infty} \frac{dK_2}{dx'} = 0 \tag{F.15}$$

$$\lim_{x'\to 1} K_2 = \infty. \tag{F.16}$$

The first two limits results from all curves in Figure F.2 tending towards the same bottom curve as $X_{\min}$ is increased. (In fact the $X_{\min} = 10D_{\max}$ case practically represents the asymptotic behaviour as $\frac{X_{\min}}{D_{\max}} \to \infty$.) The third limit states the fact that K_2 tends towards infinity as $X_{\min} \to D_{\max}$. The end results are shown in Figure F.3.



(a) K_1 versus x' (b) K_2 versus x'

Figure F.3: Analytic approximations of experimental K_1 and K_2 data.

Thus the coefficients of equation (F.10) are now defined as follows:

$$\varepsilon_{10}^L = \log_{10}\varepsilon^L \tag{F.17}$$

$$x' = X_{\min}/D_{\max} \tag{F.18}$$

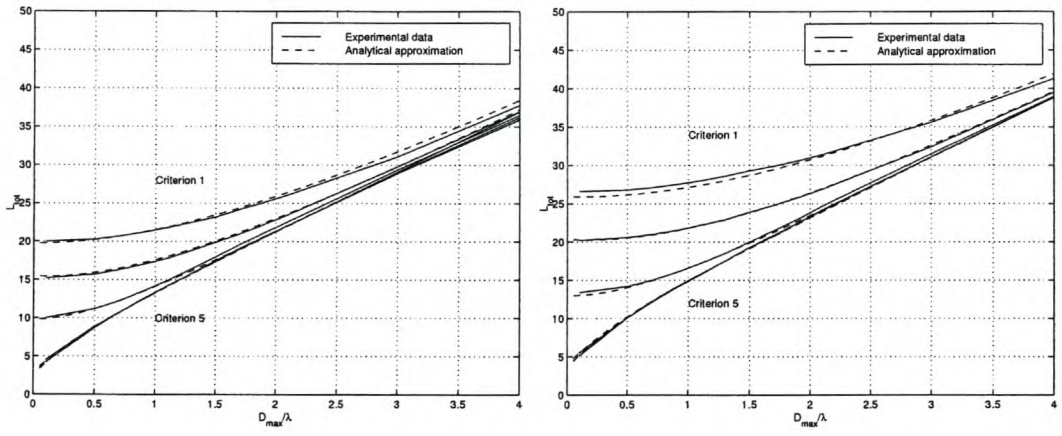
$$m = 7.826 \tag{F.19}$$

$$c = -1.92\varepsilon_{10}^{L} - 0.138 \tag{F.20}$$

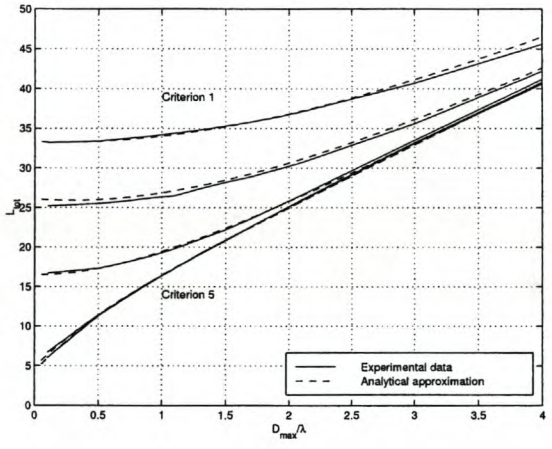
$$\alpha = [1.85 \arctan(2.8(x' - 1.74)) + 2.455] \cdot 0.8^{-\varepsilon_{10}^{L}} \tag{F.21}$$

$$k = \left[37.2\, e^{-2.42\sqrt{x'-1}} - 1.4\right] \cdot 1.3^{-\varepsilon_{10}^{L}}. \tag{F.22}$$

Figure F.4 shows the experimental data in comparison with the analytical approximation presented. Further verification of equation (F.10) is presented in Section 4.3.1.



(a) $\varepsilon = 10^{-3}$ (b) $\varepsilon = 10^{-4}$



(c) $\varepsilon = 10^{-5}$

Figure F.4: Comparison between the experimental L_{tot} vs. D_{max} data of Figure F.2 and the analytical approximation of equation (F.10).

F.3 Investigation of the spherical quadrature rule order

As set out in Section 4.3.2, a quadrature rule of product type is used for the FMM spherical quadrature. A cosinusoidally transformed, N_θ point Gauss-Legendre rule is used for θ $[0, \pi]$ and a N_ϕ point, trapezoidal rule for ϕ $[0, 2\pi]$.

This section explicitly establishes the linear relations of N_θ and N_ϕ with respect to L_{tot}, such that the influence of the relative Green function truncation error dominates the relative matrix element error. The relations are established in a semi-empirical manner, since their linear form has a theoretical backing (as set out in Section 4.3.2, also see [62]), but the constants are found experimentally.

In general, note that

$$\varepsilon^{L,M} \geq \varepsilon^{L} \tag{F.23}$$
$$\varepsilon_P^{L,M} \geq \varepsilon_P^{L}, \tag{F.24}$$

because the spherical quadrature is an additional source of error on top of the truncation error (see Section F.1 for definitions). The second inequality follows directly from the first. Clearly, as M is increased, the spherical quadrature error should decrease, therefore

$$\lim_{M\to\infty} \varepsilon^{L,M} = \varepsilon^{L} \tag{F.25}$$
$$\lim_{M\to\infty} \varepsilon_P^{L,M} = \varepsilon_P^{L}. \tag{F.26}$$

A further relationship can be established between ε_P^L and ε^L. The FMM only applies to basis function pairs belonging to far interacting groups. The domains of the basis functions are generally an order of magnitude smaller than the wavelength. Based on these two observations, it follows that the exponential part as well as the $\frac{1}{|\mathbf{X}+\mathbf{d}|}$ part of the Green function will remain fairly constant on the two basis function domains. Based on the last observation, it follows from the definition of ε_P^L that

$$\begin{aligned}
\varepsilon_P^L &= \left| \frac{\int_{S_a}\int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\,[G_0^L - G_0]\,dS'dS}{\int_{S_a}\int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\,G_0\,dS'dS} \right| \\
&\approx \left| \frac{G_0^L - G_0}{G_0} \right| \cdot \left| \frac{\int_{S_a}\int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\,dS'dS}{\int_{S_a}\int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}'))\,dS'dS} \right| \\
&= \varepsilon^L.
\end{aligned} \tag{F.27}$$

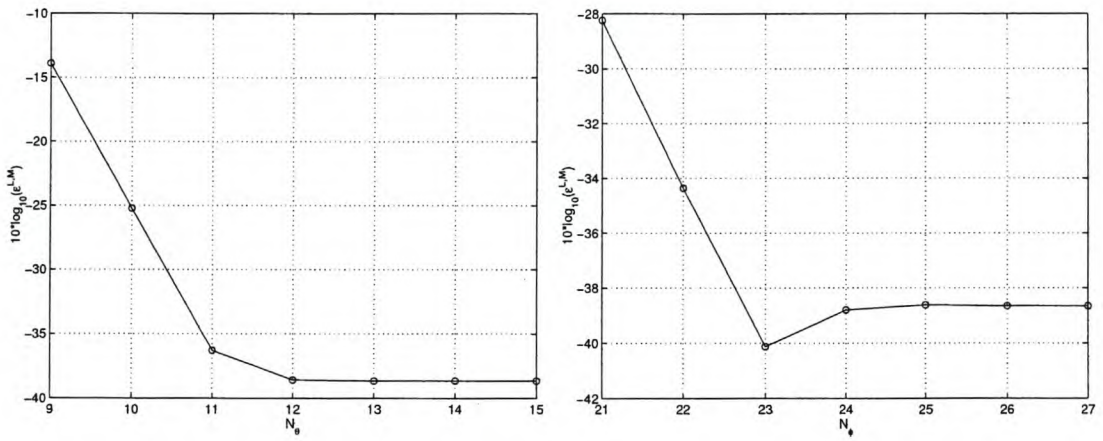
A similar relationship can be established between $\varepsilon_P^{L,M}$ and $\varepsilon^{L,M}$. Therefore

$$\varepsilon_P^{L} \approx \varepsilon^{L} \tag{F.28}$$
$$\varepsilon_P^{L,M} \approx \varepsilon^{L,M}. \tag{F.29}$$

Equations (F.29) and (F.25) together show that by choosing L_{tot} for a specified ε^L (as discussed in Section 4.3.1), one also fixes $\varepsilon_P^{L,M}$ to the same value — provided enough quadrature

points are used. Therefore, to experimentally study the convergence of $\varepsilon_P^{L,M}$ with respect to M, one only needs to consider the convergence of $\varepsilon^{L,M}$ to ε^L.

To measure the minimum M, such that $\varepsilon^{L,M} \approx \varepsilon^L$, set N_θ and N_ϕ each in turn to a very large value (assume exact integration) and increase the other until $\varepsilon^{L,M} \approx \varepsilon^L$ is obtained. Figure F.5 shows an example of the way in which the experimental N_θ- and N_ϕ versus L_{tot} data was obtained. Table F.3 presents representative minimum (N_θ, N_ϕ) versus L_{tot} data obtained in this manner. The fact that this data is only dependent on L_{tot} (linearly) and *not* on $D_{\max}$ may seem strange, since $D_{\max}$ occurs in the exponent of e in equation (F.3), but consider increasing $D_{\max}$ while keeping L_{tot} and its related M constant: the quadrature error will increase, but so will the truncation error.



(a) $\varepsilon^{L,M}$ vs. N_θ. $N_\phi = 27$.

(b) $\varepsilon^{L,M}$ vs. N_ϕ. $N_\theta = 15$.

Figure F.5: Example of the convergence of $\varepsilon^{L,M}$ with respect to the number of spherical quadrature points. $L_{\text{tot}} = 12$, $\frac{D_{\max}}{\lambda} = 0.7$, $X_{\min} = 5D_{\max}$.

Table F.3: Experimental data. Minimum (N_θ, N_ϕ) versus L_{tot}, such that $\varepsilon^{L,M} \approx \varepsilon^L$ is measured.

L_{tot}	N_θ	N_ϕ
9	9	18
12	12	24
16	16	32
25	24	50

From Table F.3 the desired linear relations now follow:

$$N_\theta = L_{\text{tot}} \tag{F.30}$$

$$N_\phi = 2L_{\text{tot}}. \tag{F.31}$$

It is gratifying to note that these results were re-enforced by evaluations of $\varepsilon_P^{L,M}$ rather than $\varepsilon^{L,M}$. See Section F.4 for a note on the evaluation of $\varepsilon_P^{L,M}$.

F.4 Evaluating $\varepsilon_P^{L,M}$ for the BI matrix as a whole

The FMM approximation to the standard BI matrix is obtained by multiplying the FMM factorization terms with each other, yielding the approximate BI matrix.

Evaluating $\varepsilon_P^{L,M}$ for the BI matrix as a whole is not entirely trivial, since every matrix element has a different $\varepsilon_P^{L,M}$. A naive approach would be to average $\varepsilon_P^{L,M}$ over all FMM-calculated matrix elements, but such a value will have little relation with the specified Green function truncation error. Most basis functions are closer to their group centers than $\frac{D_{\max}}{2}$, and will therefore result in smaller relative errors (see the experimental data in Figure F.2). The directions of most basis functions relative to their group centers are not exactly collinear with $\hat{X}$ either, which will also cause smaller relative errors (see the experimental data in Figure F.1). Finally, most inter-group distances will be larger than $X_{\min}$, therefore the relative errors of basis function combinations from such groups will always be less than or equal to the specified value (see the experimental data in Figure F.2 again). Another practical issue is the finite computing precision. The absolute Green function error $|G_0^{L,M} - G_0|$ decreases as the inter-group distance increases, therefore the finite precision calculation of $\varepsilon_P^{L,M}$ grows exceedingly inaccurate as the inter-group distance increases. In light of these observations, define $\varepsilon_P^{L,M}$ for the BI matrix as a whole, as follows:

1. Find the matrix element with the largest absolute error

$$\left| \int_{S_a} \int_{S_b} f(\mathbf{N}_a(\mathbf{r}), \mathbf{N}_b(\mathbf{r}')) \, [G_0^{L,M} - G_0] \, dS' dS \right|. \tag{F.32}$$

2. Calculate $\varepsilon_P^{L,M}$ for this element, which then represents the relative matrix element error of the matrix as a whole.

With few exceptions (due to specific mesh geometries), this element will always result from the interaction between two basis functions located at the edges of the two far interaction groups that are closest to each other — as one would expect. This definition results in a true representation of the relative matrix element error, in the sense that is should be approximately equal to the specified, relative Green function truncation error.

Finally, when measuring $\varepsilon_P^{L,M}$, it is very important to use a triangular surface integration scheme that is sufficiently accurate that it does not influence the measurement.

Bibliography

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions.* Dover, New York, 1972.

[2] M. Ainsworth and J. T. Oden. A procedure for a posteriori error estimation for *h-p* finite element methods. *Computer Meth. in Appl. Mech. and Eng*, 101:73–96, 1992.

[3] M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numerische Mathematik*, 65:23–50, 1993.

[4] M. Ainsworth and J. T. Oden. A posteriori error estimation in finite element analysis. *Computer Meth. in Appl. Mech. and Eng*, 142:1–88, 1997.

[5] M. Ainsworth, J. Z. Zhu, A. W. Craig, and O. C. Zienkiewicz. Analysis of the Zienkiewicz-Zhu *a-posteriori* error estimator in the finite element method. *Int. J. Num. Meth. Engng.*, 28:2161–2174, 1989.

[6] P. Alotto, P. Girdinio, P. Molfino, and M. Nervi. Time-harmonic mesh adaption with error estimate based on the "Local Field Error" approach. *IEEE Trans. Mag.*, 33(2):1744–1747, March 1997.

[7] P. Alotto, P. Girdinio, M. Nervi, and P. Fernandes. Mesh adaptation in finite element analysis of 2D steady state time harmonic eddy current problems. *IEEE Trans. Mag.*, 32(3):1361–1364, May 1996.

[8] S. Amini and A. Profit. Analysis of the truncation errors in the fast multipole method for scattering problems. *Jnl of Comp. and Appl. Math.*, 115:23–33, 2000.

[9] L. S. Andersen and J. L. Volakis. Hierarchal tangential vector finite elements for tetrahedra. *IEEE Microwave Guided Wave Lett.*, 8(3):127–129, March 1998.

[10] L. S. Andersen and J. L. Volakis. Development and application of a novel class of hierarchical tangential vector finite elements for electromagnetics. *IEEE Trans. Antennas Propagat.*, 47(1):112–120, January 1999.

[11] L. S. Andersen and J. L. Volakis. Adaptive multiresolution antenna modeling using hierarchical mixed-order tangential vector finite elements. *IEEE Trans. Antennas Propagat.*, 49(2):211–222, February 2001.

[12] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK user's guide.* SIAM, Philadelphia, 1992.

[13] G. Arfken. *Mathematical Methods for Physicists.* Academic Press, New York, 1970.

[14] R. Aris. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics.* Dover, New York, 1989.

[15] I. Babuška and A. Miller. The post-processing approach in the finite element method—Part 1: Calculation of displacements, stresses and other higher derivatives of the displacements. *Int. J. Num. Meth. Engng.*, 20:1085–1109, 1984.

[16] I. Babuška and A. Miller. A feedback finite element method with a posteriori error estimation: Part I: The finite element method and some basic properties of the a posteriori error estimator. *Computer Meth. in Appl. Mech. and Eng*, 61:1–40, 1987.

[17] I. Babuška, L. Planck, and R. Rodriquez. Basic problems of a posteriori error estimation. *Computer Meth. in Appl. Mech. and Eng*, 101:97–112, 1992.

[18] I. Babuška and W. C. Rheinboldt. *A-posteriori* error estimates for the finite element method. *Int. J. Num. Meth. Engng.*, 12:1597–1615, 1978.

[19] I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM Journal on Numerical Analysis*, 15(4):736–754, August 1978.

[20] I. Babuška and S. A. Sauter. Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers? *SIAM Review*, 42(3):451–484, 2000.

[21] C. A. Balanis. *Advanced Engineering Electromagnetics.* John Wiley and Sons, New York, 1989.

[22] R. E. Bank. Hierarchical bases and the finite element method. *Acta Numerica*, pages 1–43, 1996.

[23] R. E. Bank and R. K. Smith. A posteriori error estimates based on hierarchichal bases. *SIAM Journal on Numerical Analysis*, 30:921–935, 1993.

[24] R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations. *Mathematics of Computation*, 44:283–301, 1985.

[25] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods.* SIAM, Philadelphia, PA, 1994.

[26] M. L. Barton and Z. J. Cendes. New vector finite elements for three-dimensional magnetic field computation. *Journal of Applied Physics*, 61(8):3919–3921, April 1987.

[27] R. Beck, P. Deuflhard, R. Hiptmair, R. H. W. Hoppe, and B. Wohlmuth. Adaptive multilevel methods for edge element discretizations of Maxwell's equations. *Surveys on Mathematics for Industry*, 8:271–312, 1999.

[28] R. Beck, R. Hiptmair, R. H. W. Hoppe, and B. Wohlmuth. Residual based a posteriori error estimators for eddy current computation. *Mathematical Modelling and Numerical Analysis*, 34(1):159–182, 2000.

[29] R. Beck, R. Hiptmair, and B. Wohlmuth. Hierarchical error estimator for eddy current computation. In *Proceedings of the 3rd European Conference on Numerical Mathematics and Advanced Applications*, July 1999. Jyvskyl, Finland.

[30] S. S. Bindiganavale and J. L. Volakis. Comparison of three FMM techniques for solving hybrid FE-BI systems. *IEEE Antennas and Propagation Magazine*, 39(4):47–60, August 1997.

[31] S. S. Bindiganavale and J. L. Volakis. A hybrid FE-FMM technique for electromagnetic scattering. *IEEE Trans. Antennas Propagat.*, 45(1):180–181, January 1997.

[32] F. A. Bornemann, B. Erdmann, and R. Kornhuber. Adaptive multilevel methods in three space dimensions. *Int. J. Num. Meth. Engng.*, 36:3187–3203, 1993.

[33] F. A. Bornemann, B. Erdmann, and R. Kornhuber. A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM Journal on Numerical Analysis*, 33(3):1188–1204, June 1996.

[34] A. Bossavit. *Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements.* Academic Press, San Diego, 1998.

[35] A. Bossavit. A posteriori error bounds by "local corrections" using the dual mesh. *IEEE Trans. Magn.*, 35(3):1350–1353, May 1999.

[36] M. M. Botha and D. B. Davidson. Analyzing cavity-backed antennas using the hybrid FEM/MoM technique. In *Proceedings of SATCAM 2000*, September 2000. Somerset West, South Africa.

[37] M. M. Botha and D. B. Davidson. Modelling cavity-backed patch antennas with the FEM. Presentation at the 5th International Workshop on Finite Elements for Microwave Engineering, June 2000. Boston, Massachusetts.

[38] M. M. Botha and D. B. Davidson. Analyzing cavity backed, perforated substrate, microstrip patch antennas with a FMM, FE-BI hybrid formulation. In *Proceedings of the 2001 URSI International Symposium on Electromagnetic Theory*, pages 627–629, May 2001. Victoria,

Canada.

[39] M. M. Botha and D. B. Davidson. Application of the fast multipole method to the FE-BI analysis of cavity backed structures with comprehensive FMM error control. *Electromagnetics*, 22(5):393–404, July 2002.

[40] M. M. Botha and D. B. Davidson. Comparison of a posteriori error estimation results for FE analysis of waveguides and FE-BI analysis of cavity backed apertures. Presentation at the 6th International Workshop on Finite Elements for Microwave Engineering, May 2002. Chios, Greece.

[41] M. M. Botha and D. B. Davidson. Investigation of an explicit, residual-based, a posteriori error indicator for the adaptive FEM analysis of multi-port waveguide structures. *Applied Computational Electromagnetics Society Journal*, 2002. Submitted for publication.

[42] M. M. Botha and D. B. Davidson. A posteriori error estimation results for the FEM analysis of a waveguide filter. In *Proceedings of IEEE AFRICON'02*, volume 2, pages 541–544, October 2002. George, South Africa.

[43] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods.* Texts in applied mathematics 15. Springer-Verlag, New York, 1994.

[44] F. X. Canning. Improved impedance matrix localization method. *IEEE Trans. Antennas Propagat.*, 41(5):659–667, May 1993.

[45] Z. J. Cendes. Vector finite elements for electromagnetic field computation. *IEEE Trans. Magn.*, 27(5):3958–3966, September 1991.

[46] Z. J. Cendes and D. N. Shenton. Adaptive mesh refinement in the finite element computation of magnetic fields. *IEEE Trans. Magn.*, 21(5):1811–1816, September 1985.

[47] C. H. Chen and C.-D. Lien. The variational principle for non-self-adjoint electromagnetic problems. *IEEE Trans. Microwave Theory Tech.*, 28(8):878–886, August 1980.

[48] W. C. Chew, J.-M. Jin, C.-C. Lu, E. Michielssen, and J. M. Song. Fast solution methods in electromagnetics. *IEEE Trans. Antennas Propagat.*, 45(3):533–543, March 1997.

[49] W. C. Chew, J.-M. Jin, E. Michielssen, and J. Song, editors. *Fast and efficient algorithms in computational electromagnetics.* Artech House, Boston, 2001.

[50] W. C. Chew, S. Koc, J. M. Song, C. C. Lu, and E. Michielssen. A succinct way to diagonalize the translation matrix in three dimensions. *Microwave and Optical Technology Letters*, 15(3):144–147, June 1997.

[51] P. G. Ciarlet. *The finite element method for elliptic problems.* Studies in mathematics and its applications 4. North-Holland, Amsterdam, 1978.

[52] R. Coccioli, F.-R. Yang, K.-P. Ma, and T. Itoh. Aperture-coupled patch antenna on UC-PBG substrate. *IEEE Trans. Microwave Theory Tech.*, 47(11):2123–2130, November 1999.

[53] R. Coifman, V. Rohklin, and S. Wandzura. The fast multipole method for the wave equation: A pedestrian prescription. *IEEE Antennas and Propagation Magazine*, 35(3):7–12, June 1993.

[54] R. Coifman, V. Rohklin, and S. Wandzura. Faster single-stage multipole method for the wave equation. In *Proceedings of the 10th Annual Review of Progress in Applied Computational Electromagnetics*, pages 19–24, March 1994. Monterey, CA.

[55] J. S. Colburn and Y. Rahmat-Samii. Printed antenna pattern improvement through substrate perforation of high dielectric constant material: an FDTD evaluation. *Microwave and Optical Technology Letters*, 18(1):27–32, May 1998.

[56] J. S. Colburn and Y. Rahmat-Samii. Patch antennas on externally perforated high dielectric constant substrates. *IEEE Trans. Antennas Propagat.*, 47(12):1785–1794, December 1999.

[57] R. E. Collin. *Foundations for microwave engineering.* McGraw-Hill, New York, 2nd edition, 1992.

[58] R. E. Collin and F. J. Zucker, editors. *Antenna Theory.* McGraw-Hill, New York, 1969.

[59] J. D. Collins, J. L. Volakis, and J.-M. Jin. A combined finite element–boundary integral formulation for two-dimensional scattering problems via CGFFT. *IEEE Trans. Antennas Propagat.*, 38(11):1852–1858, November 1990.

[60] W. Daigang and J. Kexun. P-version adaptive computation of FEM. *IEEE Trans. Mag.*, 30(5):3515–3518, September 1994.

[61] E. Darve. The fast multipole method: a mathematical study. *C. R. Acad. Sci. Paris, Série I*, 325:1037–1042, 1997.

[62] E. Darve. The fast multipole method I: Error analysis and asymptotic complexity. *SIAM J. Numer. Anal.*, 38(1):98–128, 2000.

[63] E. Darve. The fast multipole method: Numerical implementation. *Journal of Computational Physics*, 160(1):195–240, May 2000.

[64] D. B. Davidson. Higher-order (LT/QN) vector finite elements for waveguide analysis. *Applied Computational Electromagnetics Society Journal*, 17(1):1–10, March 2002. Special Issue on Approaches to Better Accuracy/Resolution in Computational Electromagnetics.

[65] D. B. Davidson and M. M. Botha. User and code manual for program FEMFEKO version 0.97 (release 10). Technical Report US-EMSS-00/01-1-DBD, University of Stellenbosch / Electromagnetic Software and Systems, Stellenbosch, South Africa, June 2001.

[66] D. B. Davidson and R. H. Hansmann. Hierarchal 2D and 3D vector finite elements for electromagnetic wave eigenvalue problems. In *Proceedings of the 15th Annual Review of Progress in Applied Computational Electromagnetics*, pages 518–521, March 1999. Monterey, CA.

[67] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration.* Academic Press, Orlando, Florida, 2nd edition, 1984.

[68] L. M. Delves and J. L. Mohamed. *Computational Methods for Integral Equations.* Cambridge University Press, Cambridge, UK, 1985.

[69] B. Dembart and E. Yip. The accuracy of fast multipole methods for Maxwell's equations. *IEEE Computational Science and Engineering*, 5(3):48–56, July-September 1998.

[70] L. Demkowicz. An *hp*-adaptive finite element method for Maxwell's equations. In *Proceedings of the 16th Annual Review of Progress in Applied Computational Electromagnetics*, pages 18–27, 2000. Monterey, CA.

[71] L. Demkowicz, J. T. Oden, W. Rachowicz, and O. Hardy. Toward a universal *h-p* adaptive finite element strategy, part 1. constrained approximation and data structure. *Computer Meth. in Appl. Mech. and Eng*, 77:79–112, 1989.

[72] L. Demkowicz and L. Vardapetyan. Modeling of electromagnetic absorbsion/scattering problems using *hp*-adaptive finite elements. *Computer Meth. in Appl. Mech. and Eng*, 152:103–124, 1998.

[73] A. Díaz-Morcillo, L. Nuño, and J. V. Balbastre. A new error indicator for the analysis of waveguiding structures by the adaptive finite-element method. *Microwave and Optical Technology Letters*, 27(5):361–366, December 2000.

[74] D. A. Dunavant. High degree efficient symmetrical gaussian quadrature rules for the triangle. *International Journal for Numerical Methods in Engineering*, 21:1129–1148, 1985.

[75] R. Dyczij-Edlinger, G. Peng, and J.-F. Lee. Efficient finite element solvers for the Maxwell equations in the frequency domain. *Computer Meth. in Appl. Mech. and Eng*, 169:297–309, 1999.

[76] EM Software & Systems, Technopark, Stellenbosch 7600, South Africa. *FEKO User's Manual, Suite 3.0*, 2000. www.feko.co.za.

[77] N. Engheta, W. D. Murphy, V. Rokhlin, and M. S. Vassiliou. The fast multipole method (FMM) for electromagnetic scattering problems. *IEEE Trans. Antennas Propagat.*, 40(6):634–641, June 1992.

[78] M. A. Epton and B. Dembart. Multipole translation theory for the three-dimensional Laplace and Helmholtz equations. *SIAM J. Sci. Comput.*, 16(4):865–897, July 1995.

[79] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. *Acta Numerica*, pages 105–158, 1995.

[80] K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems.

Mathematics of Computation, 50(182):361–383, April 1988.

[81] J. P. de S. R. Gago, D. W. Kelly, O. C. Zienkiewicz, and I. Babuska. *A posteriori* error analysis and adaptive processes in the finite element method: Part II—adaptive mesh refinement. *Int. J. Num. Meth. Engng.*, 19:1621–1656, 1983.

[82] M. M. Gavrilovic and J. P. Webb. Targeted error indicators for use in finite-element *p*-adaption. *IEEE Trans. Mag.*, 34(5):3280–3283, September 1998.

[83] N. Geng, A. Sullivan, and L. Carin. Fast multipole method for scattering from an arbitrary PEC target above or buried in a lossy half space. *IEEE Trans. Antennas Propagat.*, 49(5):740–748, May 2001.

[84] N. A. Golias and T. D. Tsiboukis. 3-D eddy-current computation with a self-adaptive refinement technique. *IEEE Trans. Magn.*, 31(3):2261–2268, May 1995.

[85] J. Gong and J. L. Volakis. An efficient and accurate model of the coax cable feeding structure for FEM simulations. *IEEE Trans. Antennas Propagat.*, 43(12):1474–1478, December 1995.

[86] R. Gonzalo, P. de Maagt, and M. Sorolla. Enhanced patch-antenna performance by suppressing surface waves using photonic-bandgap substrates. *IEEE Trans. Microwave Theory Tech.*, 47(11):2131–2138, November 1999.

[87] R. D. Graglia, D. R. Wilton, and A. F. Peterson. Higher order interpolatory vector bases for computational electromagnetics. *IEEE Trans. Antennas Propagat.*, 45(3):329–342, March 1997.

[88] H. E. Green. Determination of the cutoff of the first higher order mode in a coaxial line by the transverse resonance technique. *IEEE Trans. Microwave Theory Tech.*, 37(10):1652–1653, October 1989.

[89] M. F. Gyure and M. A. Stalzer. A prescription for the multilevel Helmholtz FMM. *IEEE Computational Science and Engineering*, 5(3):39–47, July-September 1998.

[90] S. Hahn, C. Calmels, G. Meunier, and J. L. Coulomb. A posteriori error estimate for adaptive finite element mesh generation. *IEEE Trans. Magn.*, 24(1):315–317, January 1988.

[91] P. S. Hall and C. M. Hall. Coplanar corporate feed effects in microstrip patch array design. *IEE Proceedings H*, 135(12):1785–1794, June 1988.

[92] R. F. Harrington. *Time-Harmonic Electromagnetic Fields.* McGraw-Hill, New York, 1961.

[93] H. A. Haus and J. R. Melcher. *Electromagnetic Fields and Energy.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[94] G. Henneberger, G. Meunier, J. C. Sabonnadière, P. K. Sattler, and D. Shen. Sensitivity analysis of the nodal position in the adaptive refinement of finite element meshes. *IEEE Trans. Magn.*, 26(2):787–790, March 1990.

[95] R. E. Hodges and Y. Rahmat-Samii. The evaluation of MFIE integrals with the use of vector triangle basis functions. *Microwave and Optical Technology Letters*, 14(1):9–14, January 1997.

[96] J. Hoffman. Adaptive finite element methods for the unsteady Maxwell's equations. Preprint 2000-001, Chalmers Finite Element Center, Chalmers University of Technology, Göteborg, Sweden, 2000.

[97] J. D. Jackson. *Classical Electrodynamics.* John Wiley and Sons, New York, 3rd edition, 1999.

[98] L. Jänicke and A. Kost. Error estimation and adaptive mesh generation in the 2D and 3D finite element method. *IEEE Trans. Mag.*, 32(3):1334–1337, May 1996.

[99] J.-M. Jin. *The Finite Element Method in Electromagnetics.* John Wiley and Sons, New York, 1993.

[100] J.-M. Jin and J. L. Volakis. A finite element-boundary integral formulation for scattering by three-dimensional cavity-backed apertures. *IEEE Trans. Antennas Propagat.*, 39(1):97–104, January 1991.

[101] J.-M. Jin and J. L. Volakis. A hybrid Finite Element method for scattering and radiation by microstrip patch antennas and arrays residing in a cavity. *IEEE Trans. Antennas Propagat.*, 39(11):1598–1604, November 1991.

[102] J.-M. Jin, J. L. Volakis, and J. D. Collins. A finite-element—boundary-integral method for scattering and radiation by two- and three-dimensional structures. *IEEE Antennas and Propagation Magazine*, 33(3):22–32, June 1991.

[103] Y. Jinyun. Symmetric gaussian quadrature formulae for tetrahedronal regions. *Computer Meth. in Appl. Mech. and Eng*, 43:349–353, 1984.

[104] C. Johnson. Adaptive finite element methods for diffusion and convection problems. *Computer Meth. in Appl. Mech. and Eng*, 82:301–322, 1990.

[105] C. Johnson. Discontinuous Galerkin finite element methods for second order hyperbolic problems. *Computer Meth. in Appl. Mech. and Eng*, 107:117–129, 1993.

[106] C. Johnson and P. Hansbo. Adaptive finite element methods in computational mechanics. *Computer Meth. in Appl. Mech. and Eng*, 101:143–181, 1992.

[107] D. S. Jones. *The Theory of Electromagnetism.* Pergamon Press, Oxford, 1964.

[108] J. Jou and J.-L. Liu. A posteriori boundary element error estimation. *Jnl of Comp. and Appl. Math.*, 106:1–19, 1999.

[109] P. Keast. Moderate-degree tetrahedral quadrature formulas. *Computer Meth. in Appl. Mech. and Eng*, 55:339–348, 1986.

[110] D. W. Kelly. The self-equilibration of residuals and complementary *a posteriori* error estimates in the finite element method. *Int. J. Num. Meth. Engng.*, 20:1491–1506, 1984.

[111] D. W. Kelly, J. P. de S. R. Gago, O. C. Zienkiewicz, and I. Babuska. *A posteriori* error analysis and adaptive processes in the finite element method: Part I—error analysis. *Int. J. Num. Meth. Engng.*, 19:1593–1619, 1983.

[112] A. Kirsch and P. Monk. A finite element/spectral method for approximating the time-harmonic Maxwell system in $\Re^3$. *SIAM Journal on Applied Mathematics*, 55(5):1324–1344, October 1995.

[113] S. Koc, J. Song, and W. C. Chew. Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem. *SIAM J. Numer. Anal.*, 36(3):906–921, 1999.

[114] H. Kopka and P. W. Daly. *A Guide to LaTeX: Document Preparation for Beginners and Advanced Users.* Addison-Wesley, Reading, MA, 3rd edition, 1999.

[115] E. Kreyszig. *Introductory Functional Analysis with Applications.* John Wiley and Sons, New York, 1978.

[116] L. Lamport. *LaTeX: A document preparation system.* Addison-Wesley, Massachusetts, 1986.

[117] J.-F. Lee, D.-K. Sun, and Z. J. Cendes. Full-wave analysis of dielectric waveguides using tangential vector finite elements. *IEEE Trans. Microwave Theory Tech.*, 39(8):1262–1271, August 1991.

[118] J.-Y. Li, L.-W. Li, B.-L. Ooi, P.-S. Kooi, and M.-S. Leong. On the accuracy of the addition theorem for a scalar Green's function used in the FMM. *Microwave and Optical Technology Letters*, 31(6):439–442, December 2001.

[119] J. Liu and J.-M. Jin. A special higher order finite-element method for scattering by deep cavities. *IEEE Trans. Antennas Propagat.*, 48(5):794–703, May 2000.

[120] J.-L. Liu. On weak residual error estimation. *SIAM J. Sci. Comput.*, 17(5):1249–1268, September 1996.

[121] C. C. Lu and W. C. Chew. Fast algorithm for solving hybrid integral equations. *IEE Proceedings H*, 140(6):455–460, December 1993.

[122] C. C. Lu and W. C. Chew. A multilevel algorithm for solving a boundary integral equation of wave scattering. *Microwave and Optical Technology Letters*, 7(10):466–470, July 1994.

[123] N. Lu and J.-M. Jin. Application of fast multipole method to finite-element boundary-integral solution of scattering problems. *IEEE Trans. Antennas Propagat.*, 44(6):781–786, June 1996.

[124] D. H. Malan. Hybrid finite element and moment method modelling of microstrip patches. Master's thesis, University of Stellenbosch, Stellenbosch, South Africa, November 1995.

[125] N. Marcuvitz. *Waveguide Handbook.* Peter Peregrinus, on behalf of IEE, 1986. Originally published 1951.

[126] S. McFee and D. Giannacopoulos. The implications of second-order functional derivatives on error estimation in adaptive finite element analysis for electromagnetics. *IEEE Trans. Mag.*, 35(3):1330–1333, May 1999.

[127] S. McFee and J. P. Webb. Adaptive finite element analysis of microwave and optical devices using hierarchal triangles. *IEEE Trans. Mag.*, 28(2):1708–1711, March 1992.

[128] A. D. McLaren. Optimal numerical integration on a sphere. *Mathematics of Computation*, 17:361–383, 1963.

[129] K. K. Mei. Unimoment Method for Electromagnetic Wave Scattering. *Journal of Electromagnetic Waves and Applications*, 1(3):201–222, 1987.

[130] F. J. C. Meyer. *The Two-Dimensional Finite Element/Boundary Element Method in Electromagnetics: Formulation, Applications, Error Estimates and Mesh Adaptive Procedures.* PhD thesis, University of Stellenbosch, Stellenbosch, South Africa, 1994.

[131] P. Meyer. Die ontwerp van 'n nouband E-vlak golfleierfilter. Master's thesis, University of Stellenbosch, Stellenbosch, South Africa, 1988.

[132] P. Monk. A finite element method for approximating the time-harmonic Maxwell equations. *Numerische Mathematik*, 63:243–261, 1992.

[133] P. Monk. A posteriori error indicators for Maxwell's equations. *Jnl of Comp. and Appl. Math.*, 100:173–190, 1998.

[134] P. Monk and E. Süli. The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals. *SIAM Journal on Numerical Analysis*, 36(1):251–274, 1998.

[135] G. Mur. Edge elements, their advantages and their disadvantages. *IEEE Trans. Magn.*, 30(5):3552–3557, March 1994.

[136] J. C. Nédélec. Mixed finite elements in $\Re^3$. *Numerische Mathematik*, 35:315–341, 1980.

[137] J. C. Nédélec. A new family of mixed finite elements in $\Re^3$. *Numerische Mathematik*, 50:57–81, 1986.

[138] J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. Toward a universal *h-p* adaptive finite element strategy, part 2. a posteriori error estimation. *Computer Meth. in Appl. Mech. and Eng*, 77:113–180, 1989.

[139] J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. A posteriori error analysis in finite elements: the element residual method for symmetrizable problems with applications to compressible Euler and Navier-Stokes equations. *Computer Meth. in Appl. Mech. and Eng*, 82:183–203, 1990.

[140] P. Olszewski, T. Nakata, N. Takahashi, and K. Fujiwara. A simple algorithm for adaptive refinement of tetrahedral meshes combined with edge elements. *IEEE Trans. Magn.*, 29(2):1898–1901, March 1993.

[141] A. Papastavrou and R. Verfürth. A posteriori error estimators for stationary convection-diffusion problems: a computational comparison. *Computer Meth. in Appl. Mech. and Eng*, 189:449–462, 2000.

[142] K. D. Paulsen and D. R. Lynch. Elimination of vector parasites in finite element Maxwell solutions. *IEEE Trans. Microwave Theory Tech.*, 39(3):395–404, March 1991.

[143] L. W. Pearson, A. F. Peterson, L. J. Bahrmasel, and R. A. Whitaker. Inward-looking and outward-looking formulations for scattering from penetrable objects. *IEEE Trans. Antennas Propagat.*, 40(6):714–720, June 1992.

[144] Ü. Pekel and R. Lee. An *a posteriori* error reduction scheme for the three-dimensional finite element solution of Maxwell's equations. *IEEE Trans. Microwave Theory Tech.*, 43(2):421–427, February 1995.

[145] A. F. Peterson and D. R. Wilton. Curl-conforming mixed-order edge elements for discretizing the 2D and 3D vector Helmholtz equation. In T. Itoh, G. Pelosi, and P. P. Silvester, editors, *Finite Element Software for Microwave Engineering*, pages 101–125. John Wiley and Sons, New York, 1996.

[146] A. J. Poggio and E. K. Miller. Integral equation solutions of three dimensional scatter-

ing problems. In R. Mittra, editor, *Computer Techniques for Electromagnetics*. Pergamon, Oxford, 1973.

[147] S. Polstyanko and J.-F. Lee. Adaptive finite element electrostatic solver. *IEEE Trans. Mag.*, 37(5):3120–3124, September 2001.

[148] D. M. Pozar. *Microwave engineering*. John Wiley and Sons, New York, 2nd edition, 1998.

[149] W. H. Press, S. A. Teukolsky, W.T. Vettering, and B. R. Flannery. *Numerical Recipes in Fortran: the Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 2nd edition, 1992.

[150] S. Prudhomme and J. T. Oden. On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Computer Meth. in Appl. Mech. and Eng*, 176:313–331, 1999.

[151] S. Prudhomme, F. Pascal, J. T. Oden, and A. Romkes. Review of a priori error estimation for discontinuous Galerkin methods. Technical Report TICAM REPORT 00-27, Texas Institute for Computational and Applied Mathematics, the University of Texas at Austin, October 2000.

[152] W. Rachowicz and L. Demkowicz. An *hp*-adaptive finite element method for electromagnetics Part 1: Data structure and constrained approximation. *Computer Meth. in Appl. Mech. and Eng*, 187:307–335, 2000.

[153] W. Rachowicz, J. T. Oden, and L. Demkowicz. Toward a universal *h-p* adaptive finite element strategy, part 3. design of *h-p* meshes. *Computer Meth. in Appl. Mech. and Eng*, 77:181–212, 1989.

[154] S. Ramo, J. R. Whinnery, and T. van Duzer. *Fields and Waves in Communication Electronics*. John Wiley and Sons, 3rd edition, 1994.

[155] E. Rank. Adaptive *h*-, *p*- and *hp*-versions for boundary integral element methods. *Int. J. Num. Meth. Engng.*, 28:1335–1349, 1989.

[156] R. Rannacher. Error control in finite element computations. Preprint 1998-54, Universität Heidelberg, 1998.

[157] P. A. Raviart and J. M. Thomas. A mixed finite element method for 2nd order elliptic problems. In A. Dold and B. Eckmann, editors, *Mathematical Aspects of Finite Element Methods*. Springer-Verlag, New York, 1977.

[158] B. D. Reddy. *Introductory functional analysis: with applications to boundary value problems and finite elements*. Texts in applied mathematics 27. Springer, New York, 1998.

[159] C. J. Reddy, M. D. Deshpande, C. R. Cockrell, and F. B. Beck. Analysis of three-dimensional-cavity-backed aperture antennas using a combined finite element method/method of moments/geometrical theory of diffraction technique. Technical Report 3548, NASA, Langley Research Center, November 1995.

[160] C. J. Reddy, M. D. Deshpande, C. R. Cockrell, and F. B. Beck. Electromagnetic scattering analysis of a three-dimensional-cavity-backed aperture in an infinite ground plane using a combined finite element method/ method of moments approach. Technical Report 3544, NASA, Langley Research Center, October 1995.

[161] J. N. Reddy. *Applied functional analysis and variational methods in engineering*. McGraw-Hill, New York, 1986.

[162] S. R. Rengarajan and Y. Rahmat-Samii. The field equivalence principle: Illustration of the establishment of the non-intuitive null fields. *IEEE Antennas Propagat. Mag.*, 42(4):122–128, August 2000.

[163] V. Rokhlin. Rapid solution of integral equations of scattering theory in two dimensions. *Journal of Computational Physics*, 86:414–439, 1990.

[164] V. Rokhlin. Diagonal forms of translation operators for the Helmholtz equation in three dimensions. Technical Report YALEU/DCS/RR-894, Department of Computer Science, Yale University, March 1992.

[165] Y. Saad. *Iterative methods for sparse linear systems*. PWS Publishing Company, Boston,

MA, 1996.

[166] H. Sagan. *Introduction to the Calculus of Variations.* Dover, New York, 1992.

[167] M. Salazar-Palma, T. K. Sarkar, L.-E. García-Castillo, T. Roy, and A. Djordjević. *Iterative and self-adaptive finite-elements in electromagnetic modeling.* Artech House, Boston, 1998.

[168] J. S. Savage. Comparing high order vector basis functions. In *Proceedings of the 14th Annual Review of Progress in Applied Computational Electromagnetics*, pages 742–749, March 1998. Monterey, CA.

[169] J. S. Savage and A. F. Peterson. Higher-order vector finite elements for tetrahedral cells. *IEEE Trans. Microwave Theory Tech.*, 44(6):874–879, June 1996.

[170] S. Savage and J. Manges. Local error estimation for high-frequency problems using hierarchal tangential vector finite elements. In *Proceedings of the 15th Annual Review of Progress in Applied Computational Electromagnetics*, pages 524–529, March 1999. Monterey, CA.

[171] R. A. Serway. *Physics for scientists and engineers with modern physics.* Saunders Colledge Publishing, Philadelphia, 3rd edition, 1992.

[172] X.-Q. Sheng, J.-M. Jin, J. Song, C.-C. Lu, and W. C. Chew. On the formulation of hybrid finite-element and boundary-integral methods for 3-D scattering. *IEEE Trans. Antennas Propagat.*, 46(3):303–311, March 1998.

[173] X.-Q. Sheng, E. K.-N. Yung, C. H. Chan, J.-M. Jin, and W. C. Chew. Scattering from a large body with cracks and cavities by the fast and accurate finite-element boundary-integral method. *IEEE Trans. Antennas Propagat.*, 48(8):1153–1160, August 2000.

[174] P. P. Silvester and R. L. Ferrari. *Finite Elements for Electrical Engineers.* Cambridge University Press, Cambridge, 3rd edition, 1996.

[175] P. P. Silvester and G. Pelosi, editors. *Finite Elements for Wave Electromagnetics.* IEEE Press, New York, 1994.

[176] J. Song and W. C. Chew. Error analysis for the truncation of multipole expansions of vector Green's functions. *IEEE Microwave and Wireless Components Letters*, 11(7):311–313, July 2001.

[177] J. Song, C.-C. Lu, and W. C. Chew. Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects. *IEEE Trans. Antennas Propagat.*, 45(10):1488–1493, October 1997.

[178] J. M. Song and W. C. Chew. Fast multipole method solution using parametric geometry. *Microwave and Optical Technology Letters*, 7(16):760–765, November 1994.

[179] J. M. Song and W. C. Chew. Multilevel fast-multipole algorithm for solving combined field integral equations of electromagnetic scattering. *Microwave and Optical Technology Letters*, 10(1):14–19, September 1995.

[180] M. R. Spiegel and J. Liu. *Mathematical handbook of formulas and tables.* McGraw-Hill, New York, 2nd edition, 1999.

[181] J. R. Stewart and T. J. R. Hughes. A tutorial in elementary finite element error analysis: A systematic presentation of a priori and a posteriori error estimates. *Computer Meth. in Appl. Mech. and Eng*, 158:1–22, 1998.

[182] A. H. Stroud and D. Secrest. *Gaussian Quadrature Formulas.* Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

[183] Structural Dynamics Research Corp., P.O. Box 1172, Exton, PA 19341, USA. *FEMAP Version 8.0 User Guide*, 2000. www.femap.com.

[184] D. Sun, J. Manges, X. Yuan, and Z. Cendes. Spurious modes in finite-element methods. *IEEE Antennas and Propagation Magazine*, 37(5):12–24, October 1995.

[185] D. K. Sun, Z. Cendes, and J.-F. Lee. Adapive mesh refinement, h-version, for solving multiport microwave devices in three dimensions. *IEEE Trans. Mag.*, 36(4):1596–1599, July 2000.

[186] D.-K. Sun, J.-F. Lee, and Z. Cendes. Construction of nearly orthogonal Nedelec bases for rapid convergence with multilevel preconditioned solvers. *SIAM J. Sci. Comput.*, 23(4):1053–

1076, 2001.

[187] C. T. Tai. *Dyadic Green Functions in Electromagnetic Theory.* IEEE Press, New York, 2nd edition, 1994.

[188] L. Vardapetyan. *hp-Adaptive Finite Element Method for Electromagnetics with Applications to Waveguiding Structures.* PhD thesis, The University of Texas at Austin, Austin, Texas, 1999.

[189] L. Vardapetyan and L. Demkowicz. *hp*-adaptive finite elements in electromagnetics. *Computer Meth. in Appl. Mech. and Eng*, 169:331–344, 1999.

[190] R. Verfürth. A posteriori error estimators for the Stokes equations. *Numerische Mathematik*, 55:309–325, 1989.

[191] R. Verfürth. A posteriori error estimators for convection-diffusion equations. *Numerische Mathematik*, 80:641–663, 1998.

[192] J.L. Volakis, A. Chatterjee, and L.C. Kempel. *Finite Element Method for electromagnetics : Antennas, Microwave Cicuits and Scattering Applications.* Oxford University Press and IEEE Press, Oxford & New York, 1998.

[193] R. L. Wagner and W. C. Chew. A ray-propagation fast multipole algorithm. *Microwave and Optical Technology Letters*, 7(10):435–438, July 1994.

[194] J. J. H. Wang. *Generalized Moment Methods in Electromagnetics.* John Wiley and Sons, New York, 1991.

[195] J. P. Webb. Edge elements and what they can do for you. *IEEE Trans. Magn.*, 29(2):1460–1465, March 1993.

[196] J. P. Webb. Hierarchal vector basis functions of arbitrary order for triangular and tetrahedral finite elements. *IEEE Trans. Antennas Propagat.*, 47(8):1244–1253, August 1999.

[197] J. P. Webb. *P*-adaptive methods for electromagnetic wave problems using hierarchal tetrahedral edge elements. *Electromagnetics*, 22(5):443–451, July 2002.

[198] J. P. Webb and B. Forghani. Hierarchal scalar and vector tetrahedra. *IEEE Trans. Magn.*, 29(2):1495–1498, March 1993.

[199] J. P. Webb and B. Forghani. Adaptive improvement of magnetic fields using hierarchal tetrahedral finite elements. *IEEE Trans. Magn.*, 30(5):3511–3514, September 1994.

[200] D. R. Wilton, S. M. Rao, A. W. Glisson, D. H. Schaubert, O. M. Al-Bundak, and C. M. Butler. Potential integrals for uniform and linear source distributions on polygonal and polyhedral domains. *IEEE Trans. Antennas Propagat.*, 32(3):276–281, March 1984.

[201] P. C. L. Yip. *High-Frequency Circuit Design and Measurements.* Chapman & Hall, London, 1990.

[202] J.-S. Zhao, W. C. Chew, C.-C. Lu, E. Michielssen, and J. Song. Thin-stratified medium fast-multipole algorithm for solving microstrip structures. *IEEE Trans. Microwave Theory Tech.*, 46(4):395–403, April 1998.

[203] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Num. Meth. Engng.*, 24:337–357, 1987.

[204] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and *a posteriori* error estimates. Part 1: The recovery technique. *Int. J. Num. Meth. Engng.*, 33:1331–1364, 1992.

[205] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and *a posteriori* error estimates. Part 2: Error estimates and adaptivity. *Int. J. Num. Meth. Engng.*, 33:1365–1382, 1992.

[206] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery (SPR) and adaptive finite element refinement. *Computer Meth. in Appl. Mech. and Eng*, 101:207–224, 1992.

[207] O. C. Zienkiewicz, J. Z. Zhu, and N. G. Gong. Effective and practical *h-p*-version adaptive analysis procedures for the finite element method. *Int. J. Num. Meth. Engng.*, 28:879–891, 1989.