Accurate and Efficient Algorithms for Boundary Element Methods in Electromagnetic Scattering - A Tribute to the Work of Professor F. Olyslager

K. Cools\textsuperscript{1}, I. Bogaert\textsuperscript{1}, J. Fostier\textsuperscript{1}, J. Peeters\textsuperscript{1}, D. Vande Ginste\textsuperscript{1}, H. Rogier\textsuperscript{1}, and D. De Zutter\textsuperscript{1}
Abstract. Boundary element methods (BEMs) are an increasingly popular approach to model electromagnetic scattering both by perfect conductors and dielectric objects. Several mathematical, numerical, and computa-

I. Bogaert, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (ignace.bogaert@intec.ugent.be)

K. Cools, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (kristof.cools@intec.ugent.be)

D. De Zutter, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (daniel.dezutter@intec.ugent.be)

J. Fostier, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (jan.fostier@ugent.be)

J. Peeters, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (joris.peeters@intec.ugent.be)

H. Rogier, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (hendrik.rogier@intec.ugent.be)

D. Vande Ginste, Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium. (dries.vandeginste@intec.ugent.be)

¹Department of Information Technology,
Ghent University, 9000 Gent, Belgium.
tional techniques pulled from the research into BEMs, enhancing its efficiency and applicability. In designing a viable implementation of the BEM, both theoretical and practical aspects need to be taken into account. Theoretical aspects include the choice of an integral equation for the sought after current densities on the geometry’s boundaries and the choice of a discretization strategy (i.e., a finite element space) for this equation. Practical aspects include efficient algorithms to execute the multiplication of the system matrix by a test vector (such as a fast multipole method) and the parallelization of this multiplication algorithm that allows the distribution of the computation and communication requirements between multiple computational nodes. In honor of our former colleague and mentor, Prof. F. Olyslager, an overview of the BEMs for large and complex EM problems developed within the Electromagnetics Group at Ghent University is presented. Recent results that ramified from Prof. Olyslager’s scientific endeavors are included in the survey.
1. Introduction

The electromagnetic (EM) scattering at complex metallic or dielectric objects often is modeled by Method of Moments (MoM) discretized boundary integral equations. From a mathematical point of view, a BEM consists from an equation for the unknown densities defined on the boundary of the geometry under consideration, and a discretization strategy. Describing a discretization strategy amounts to replacing the original equation (posed in an infinite dimensional function space) by a sequence of 'related' finite dimensional problems. Here, 'related' means that these problems are in some sense close to the original one. The most popular sequence of approximate problems yields the so-called $h$-version of the MoM. In the $h$-version, the mesh parameter $h$, i.e. the largest among the diameters of the triangles used to approximate the scatterer’s geometry, is adaptively decreased until the desired accuracy on the approximate solution is reached.

As soon as the discretization strategy is fixed, theoretically derived a priori error bounds can be derived, allowing the approximate determination of the 'coarsest' finite dimensional problem (i.e. the problem corresponding to the largest mesh parameter) that is expected to yield a solution that is within a given tolerance to the real solution.

From this discussion, the importance of choosing a good MoM scheme (equation and discretization) is evident in order to compute the approximate solution as efficiently as possible.

However, there are other, more practical aspects which determine the computational and memory resources that are required to compute the solution of the scatter-
ing problem. The discretization process described above leads to a finite dimensional system of linear equations. For large scale, high frequency scattering problems, the only reasonable method to solve these systems is an iterative algorithm.

The amount of computational resources required to solve the system iteratively depends on: (i) the number of iterations required for the algorithm to arrive at the solution (within a certain tolerance, which can be chosen, for example, to be the discretization scheme’s a priori error estimate), (ii) the computational cost required to perform one multiplication of the system matrix by a test vector.

The amount of memory resources required to solve the system iteratively is the amount of memory required to store the data that is necessary to perform the above mentioned matrix vector multiplication. Strictly speaking, each matrix element can be computed on the fly from the geometry and excitation data defining the scattering problem. In practice, this strategy gives rise to very inefficient algorithms and therefore the matrix is stored, albeit in compressed form.

Parallel execution of an iterative algorithm allows to deploy more computational and memory resources than are commonly available on one computational node. This is important, especially with regard to memory resources, since there is a practical (and budgetary) limit on the amount of memory that can be installed (and addressed efficiently) on one computational node.

Parallel algorithms for the iterative solution of linear systems arising upon discretization of boundary integral equations should satisfy two criteria; they should be: (i) **asynchronous** in that there are no fixed points of communication at which the
communication network could be overburdened, (ii) scalable in that when both the
total number of computational nodes and the problem size increase, the communication
cost remains below a fixed percentage of the total computational workload.

Under the supervision of, a.o., Prof. F. Olyslager, the Electromagnetics Group
with its longstanding history in computational electromagnetics started focusing its
research on the method of moments. All of its aspects as listed above have been sub-
ject of investigation. Improvements have been published in literature (see reference
list). More explicitly, research topics included:

- *Minimizing the number of degrees of freedom.* For most solvers, the discretiza-
tion error scales proportional to the interpolation error. Some solvers yield more
accurate solutions than others, and thus a coarser discretization and fewer degrees
of freedom (DoFs) are sufficient to reach a set accuracy. As elucidated above, the
number of degrees of freedom (and associated the required amount of computation
and memory resources) depends heavily on the choice of boundary integral equation
and discretization scheme. More particular, it is known that so-called conformal dis-
cretization schemes are more accurate than non-conformal schemes. Therefore, in
PEC scattering the electric field integral equation (EFIE) is preferred to the mag-
netic field integral equation (MFIE), since for the latter no stable conformal schemes
were known. Several solution methods have been proposed such as enriching the
finite elements space, adapting the mesh, and adjusting the magnetic field integral
recently, using both a primal and dual finite element space, a conformal well-tested 
discretization for the MFIE has been devised.

- **Reducing the number of iterations required.** The number of iterations required 
to reach a solution to the linear system depends on its spectrum. In those cases 
were the spectrum is oriented mainly along one of the axes of the complex plane, the 
number of iterations can be estimated by the condition number. The spectrum of 
the MFIE and its extension to penetrable objects (the Müller equation) is situated 
around a finite non zero real number. Therefore, the number of iterations expected 
to reach a solution is small. An alternative to perform a conforming discretization of 
the MFIE is discretizing the EFIE and preconditioning the resulting system. This is 
possible because the EFIE operator exhibits a self-regularizing property.

- **Reducing the cost of a matrix-vector multiplication.** Since every iteration of the 
solution algorithm entails a matrix-vector multiplication, reducing its cost will result 
in a decreased solution time. The matrix-vector multiplication can be performed in 
nearly linear time using a fast multipole method (FMM). The electromagnetics group 
at Ghent University developed some dedicated FMMs.

Firstly, planarly layered media are of huge practical importance in microwave de-
sign. The Green's function of a layered medium can be approximated by a series 
expansion by replacing the half infinite solution domain by a layered medium that 
is closed in both directions, be it by a perfectly conducting ground plane, be it by 
a perfectly matched layer (PML) — detailed mathematically by Prof. F. Olyslager 
in *Olyslager* [2004]. Each of the terms in this series represent a two-dimensional
mode which, in its turn, can be expanded using an efficient 2D FMM \textit{Vande Ginste et al.} [2004, 2006]. These methods were extended into a multilevel context in \textit{Vande Ginste et al.} [2006, 2009a]. Matrix-vector multiplication times have been decreased significantly by means of Singular Value Decompositions \textit{Vande Ginste et al.} [2009b]. It was found that PML-based MLFMAs could also be devised for photonic crystals \textit{Pissoort et al.} [2007]. The ill-conditioning of the (iterative) solution technique for photonic crystals was mitigated by application of a rank-revealing QR based preconditioner \textit{Pissoort et al.} [2006].

Secondly, multiscale structures are common in electromagnetic design. This implies that the lowest level boxes in the MLFMA tree are electrically small and require low frequency expansions of the Green’s function, while the high level boxes in the MLFMA tree are electrically large and require high frequency expansions of the Green’s function. Recently, a novel, parametrized expansion of the Green’s function has been developed that is valid both at low and high frequencies: the nondirectional stable plane wave MLFMA (NSPWMLFMA) \textit{Bogaert et al.} [2008b]. The expansion is based on plane waves in both frequency regions. The resulting scheme is (akin to classic high frequency MLFMA) diagonal and thus efficient. The number of plane waves that need to be included in the expansion in order to reach a certain accuracy has been decreased by taking into account that the source region is a cuboid, and thus strictly smaller than its encompassing sphere \textit{Bogaert et al.} [2010].

- \textit{Asynchronous, Scalable, Adaptive Parallelization}. In the ideal case, the deployment of $P$ machines would reduce the computation time with a factor $P$. However,
the constituent processes rely on data produced by the other processes. This creates
the need for communication and creates order relations in the execution. An optimal
reduction in computation time requires the intelligent design of the parallelization
scheme used. In this paper, the asynchronous parallel MLFMA is introduced. It
can be proven to be scalable Fostier and Olyslager [2008b]. In this contribution, it is
argumented that the efficiency and load balancing can be further improved by using
an adaptive tree in the MLFMA.

2. Reducing the number of degrees of freedom

Consider a closed surface \( \Gamma \) and exterior normal \( \mathbf{n} \), in a background medium with
permittivity \( \epsilon \) and permeability \( \mu \). It is illuminated by an incident field \((\mathbf{e}^i, \mathbf{h}^i)\). The
PEC boundary conditions for the tangential traces of electric and magnetic field lead
to the following EFIE and MFIE:

\[
\mathbf{n} \times \mathbf{e}^i (\mathbf{r}) = -\eta T \mathbf{j} (\mathbf{r})
\]

\[
= -\frac{1}{j \omega \epsilon} \mathbf{n} \times \int_{\Gamma} \frac{\nabla e^{-jkR}}{4\pi R} \nabla' \cdot \mathbf{j} (\mathbf{r}') \, dS' \\
+ j \omega \mu \mathbf{n} \times \int_{\Gamma} \frac{e^{-jkR}}{4\pi R} \mathbf{j} (\mathbf{r}') \, dS',
\]

\[
\mathbf{n} \times \mathbf{h}^i (\mathbf{r}) = \left\{ \frac{1}{2} + K \right\} [\mathbf{j}] (\mathbf{r})
\]

\[
= \frac{1}{2} \mathbf{j} (\mathbf{r}) - \mathbf{n} \times \frac{1}{4\pi} \int_{\Gamma} \frac{e^{-jkR}}{R} \times \mathbf{j} (\mathbf{r}') \, dS',
\]

Devising a method of moments to discretize these boundary integral equation com-
prises
• The choice of a finite element space the approximate solution is sought in. This means the choice of a subdivision in cells (e.g. triangular), the description of the per cell behaviour (e.g. polynomial), the inter-cell continuity conditions (e.g. normal components should be continuous), and the choice of degrees of freedom (e.g. edge value of the normal component). The choice of degrees of freedom implies the choice of basis function (i.e. those functions in the finite element space that make one degree of freedom 1 and all others 0).

• The choice of a finite element space used for testing. This is done similarly to the construction of the finite element space the approximate solution is sought in. More in particular this construction yields a set of testing functions.

If both these spaces are subspaces of the functional spaces the continuous equation is posed in, the method is called conforming. In classical implementations, the solution of the MFIE is sought after in the space spanned by RWG functions \( f_m \) subordinate to a triangular approximation of the scatterer’s surface. This is a subspace of the space \( H^{-1/2}(\text{div}; \Gamma) \) of currents radiating finite energy fields. The testing is done using these same divergence conforming RWG functions \( f_m \). However, since the tangential traces \( n \times h \) of the magnetic fields belong to the function space \( H^{-1/2}(\text{div}; \Gamma) \) of divergence conforming functions whose \( L^2(\Gamma) \)-dual consists of curl-conforming functions in \( H^{-1/2}(\text{curl}; \Gamma) \), the testing space is not conforming.

The rationale behind discretizing the MFIE in this manner is that the discretized identity operator is well-conditioned, resulting in fast convergence of iterative solution algorithms. To achieve a conforming and well-conditioned discretization of the MFIE,
it suffices to find a set of basis functions that is curl conforming and that results in a well-conditioned discretization of the identity operator Cools et al. [2009]. The set of ‘rotated’ Buffa-Christiansen (BC) functions $n \times g_n$ Buffa and Christiansen [2007] constitutes such a set. These functions are defined as linear combination of RWG functions subordinate to the barycentric refinement of the original mesh (Fig. 1).

This scheme yields the following system: find the current density $\sum_{i=n}^{N} J_n f_n (r)$ such that $J = (J_n)_{n=1}^{N}$ fulfills

$$\left( \frac{1}{2} G + K \right) \cdot J = H. \quad (3)$$

Here, $(G)_{m,n} = (n \times g_m, f_n)$,

$$(K)_{m,n} = -\frac{1}{4\pi} \int_{\Gamma \times \Gamma} g_m (r) \cdot \left( \nabla \frac{e^{-jkR}}{R} \times f_n (r') \right) dS' dS, \quad (4)$$

and $(H)_{m} = (g_m, h^i)$.

The surface current distribution for scattering by a PEC cube at 150 MHz has been computed for ever decreasing values of the mesh parameter $h$ using the classical discretization of the MFIE and the mixed discretization of the MFIE (Fig. 2). The difference w.r.t. the EFIE solution subordinate to the same mesh has been computed in the $L^2(\Gamma)$, $H^{-1/2}(\text{div}; \Gamma)$, and $H(\text{div}; \Gamma)$ norms. It is clear that the mixed scheme outperforms the classic scheme. Note moreover, that in the $H(\text{div}; \Gamma)$ norm, the classic solution does not converge!

Finally, in the past, other strategies to improve the accuracy of the MFIE have been investigated. In particular, the use of curl conforming rotated RWG functions $n \times f_n$ has been studied Ubeda and Rius [2005]; Ergül and Gürel [2006]; Peterson [2006]; Ubeda and Rius [2008]; Peterson [2008]. They have been shown to improve accuracy
at least in a class of simulations. The improvement reported in these references can not be explained, however, by the above framework, since these methods are non-conforming.

3. Reducing the number of iterations required

The EFIE suffers from dense grid breakdown, i.e. the condition number of the system increases if the mesh parameter decreases. Calderón preconditioning Andriulli et al. [2008], has proven to be an efficient strategy to remove this problem. Perhaps less known is that the Poggio-Miller-Chang-Harrington-Wu-Tsai (PMCHWT) equation as described in Poggio and Miller [1973] used to model scattering by penetrable bodies also suffers from dense grid breakdown. It is not immediately clear in this case whether Calderón preconditioning is applicable and what form it should take. Inspired by the CP-EFIE, the following Calderón preconditioned PMCHWT (CP-PMCHWT) equation is proposed:

\[
\left( \begin{array}{cc}
K + K' & -\eta T - \eta' T' \\
T/\eta + T'/\eta' & K + K'
\end{array} \right) \cdot \left( \begin{array}{c}
e \times n \\
n \times h
\end{array} \right) = - \left( \begin{array}{cc}
K + K' & -\eta T - \eta' T' \\
T/\eta + T'/\eta' & K + K'
\end{array} \right) \cdot \left( \begin{array}{c}
e^{inc} \times n \\
n \times h^{inc}
\end{array} \right).
\]

(5)

Here, \((\eta, k)\) and \((\eta', k')\) are the characteristic impedance of the exterior and exterior medium, respectively. Leveraging the Calderón identities, the diagonal blocks of the operator involved can be seen to be

\[
\frac{1}{2} - \frac{\eta'}{\eta} TT' - \frac{\eta}{\eta'} T'T + KK' + K'K,
\]

(6)
which is the addition of the a constant operator and a compact perturbation. The upper-right off-diagonal block reads

\[-(K + K')(\eta T + \eta' T') - (\eta T + \eta' T')(K + K')\]

\[= -\eta\{K + K', T\} - \eta'\{K + K', T'\}. \quad (7)\]

Here, curly braces denote the anti-commutator of their arguments, \(\{A, B\} = AB + BA\). Since equal material operators \(K\) and \(T\) anti-commute, (7) reduces to

\[-\eta\{K', T\} - \eta'\{K, T'\}\]

\[= -\eta k\{K', kT\} - \eta' k\{K, k'T'\}\]

\[= -\eta k\{K', k'T' + (kT - k'T')\} - \eta' k\{K, kT + (k'T' - kT)\}\]

\[= -\eta k\{K', (kT - k'T')\} - \eta' k\{K, (k'T' - kT)\}. \quad (8)\]

This expression, although not zero, only contains compact contributions since the hypersingular parts of \(kT\) and \(k'T'\) cancel each other (cfr. the construction of the Müller equation).

Similarly, the lower-left off-diagonal block can be shown to be compact:

\[\left(\frac{T}{\eta} + \frac{T'}{\eta'}\right)(K + K') + (K + K')\left(\frac{T}{\eta} + \frac{T'}{\eta'}\right)\]

\[= \frac{1}{\eta}\{T, K'\} + \frac{1}{\eta'}\{T', K\}\]

\[= \frac{1}{k\eta}\{kT - k'T' + k'T', K'\} + \frac{1}{k'\eta'}\{k'T' - kT + kT, K\}\]

\[= \frac{1}{k\eta}\{kT - k'T', K'\} + \frac{1}{k'\eta'}\{k'T' - kT, K\}. \quad (9)\]
The complete block operator thus is the addition of a constant operator and a compact perturbation, which is expected to yield a well conditioned system matrix, regardless the mesh parameter.

The discretization of this CP-PMCHWT equation requires, as is the case for the CP-EFIE, both RWG functions $f_n$ and BC functions $g_n$. More in particular, the electric and magnetic currents are approximated by expansions in RWG basis functions:

$$e \times n \approx \sum_{n=1}^{N} M_n f_n$$  \hspace{1cm} (10)

$$n \times h \approx \sum_{n=1}^{N} J_n f_n.$$ \hspace{1cm} (11)

The first PMCHWT operator is tested using curl-conforming 'rotated' RWG functions $n \times f_n$. These RWG testing coefficients are mapped onto BC expansion coefficients by multiplication with the inverse of the mixed Gram matrix (3). Finally, the second PMCHWT operator is testing using curl-conforming 'rotated' BC functions $n \times g_n$.

This discretization scheme yields

$$
\begin{pmatrix}
K_{gg} + K'_{gg} & -\eta T_{gg} - \eta' T'_{gg} \\
T_{gg}/\eta + T_{gg}/\eta' & K_{gg} + K'_{gg}
\end{pmatrix}
G^{-1}
\begin{pmatrix}
K_{ff} + K'_{ff} & -\eta T_{ff} - \eta' T'_{ff} \\
T_{ff}/\eta + T_{ff}/\eta' & K_{ff} + K'_{ff}
\end{pmatrix}
\begin{pmatrix}
M_f \\
J_f
\end{pmatrix}
= -
\begin{pmatrix}
K_{gg} + K'_{gg} & -\eta T_{gg} - \eta' T'_{gg} \\
T_{gg}/\eta + T_{gg}/\eta' & K_{gg} + K'_{gg}
\end{pmatrix}
G^{-1}
\begin{pmatrix}
M^{inc}_f \\
J^{inc}_f
\end{pmatrix}.$$

(12)

Here the subscripts carried by the matrices indicate the expansion/testing scheme. For example

$$(K_{ff})_{mn} = (n \times f_m, K f_n).$$ \hspace{1cm} (13)$$
Similar expressions exist for other combinations of operators, expansion functions, and testing functions. Moreover,

\[(\mathbf{M}^{inc}_f)_m = (n \times f_m, e^{inc} \times n), \]  

\[(\mathbf{J}^{inc}_f)_m = (n \times f_m, n \times h^{inc}). \]  

The vectors \(\mathbf{M}_f\) and \(\mathbf{J}_f\), finally, contain the expansion coefficients defined by (10) and (11), respectively.

As an example, the system matrix was constructed for scattering by a 1 meter radius sphere of relative permittivity and permeability 1.5. The sphere is illuminated by an incident wave

\[e^{inc}(r) = xe^{-j\frac{2\pi}{\lambda}z}, \]  

with \(\lambda = 2\) meter. The sphere was modeled by different meshes. The condition number remains constant, whereas the condition number of the classic PMCHWT equation tends to infinity quadratically (Fig. 3). For a mesh parameter of 0.4 meter, the number of iterations plummeted from 256 to 9, corresponding to a speedup of 14. Calderón Preconditioner for the PMCHWT has been extended to PMCHWT based modelling of scattering by Chiral media Beghein et al. [2011], and a Calderón regularized electric current equation for scattering by a configuration of penetrable objects containing junctions has been described in Ylä-Oijala et al. [2011].

4. Reducing the matrix-vector multiplication cost

In the iterative solution of integral equations for electromagnetic scattering the computationally most intensive step is the multiplication of the system matrix with
a vector. When there are $N$ basis and test functions, $O(N^2)$ operations are needed. In the past, many fast matrix multiplication methods have been developed to perform this task more efficiently. An important class of fast matrix multiplication methods are the so-called fast multipole methods (FMMs), which can reduce the computational and memory complexity to $O(N)$ or $O(N \log N)$. In FMMs, the geometry of the problem is subdivided into a hierarchy (a tree) of boxes. Then a decomposition of the Green’s function is invoked to evaluate interactions between entire boxes. Because the boxes interact as a whole, there is no need to keep track of all the individual interactions between basis and test functions in these boxes. Therefore massive efficiency gains are realized.

The MLFMA Coifman et al. [1993], Chew et al. [1997], Song et al. [1998] is arguably the most successful and widely used FMM. It is based on a propagating plane wave expansion of the Green’s function. However, this expansion suffers from a low-frequency breakdown due to numerical roundoff errors. As a consequence it is impossible to evaluate the interactions between groups that are significantly less than one wavelength apart. These interactions must therefore be performed classically, destroying the linear complexity for problems with a lot of subwavelength geometrical detail. In modern FMMs, this problem is usually circumvented by switching to an alternative, low-frequency stable, expansion of the Green’s function when the group size becomes too small.

Most of these alternative expansions are either based on spherical modes Jiang and Chew [2005], Bogaert et al. [2008b], Bogaert et al. [2008a], Bogaert and Olyslager...
[2009b], Bogaert and Olyslager [2008] or on the spectral representation of the Green’s function Darve and Havé [2004], Cheng et al. [2006], Wallen and Sarvas [2005]. Deep down, all spherical mode-based expansions are based on Gegenbauer’s addition theorem Abramowitz and Stegun [1965]. As a consequence these expansions only converge if the spheres circumscribing two interacting boxes are not touching. Although the spheres circumscribing interacting FMM boxes in three dimensional FMM are indeed not touching, they can be very close if the two interacting boxes are separated by only one buffer box. For low frequencies, this leads to a very slow convergence of the expansion (see also Bogaert et al. [2011]). This problem does not occur when using the MLFMA at high frequencies, because the number of plane waves is nearly constant as a function of the distance between the boxes. It is only when the two boxes enter each other’s near field that this convergence problem manifests itself. The only expansions that successfully avoid this problem are based on the spectral representation. The reason for this is the half-space-shaped convergence region. Unfortunately, this also leads to the need for six radiation patterns, which increases the computational cost. Apparently, all these expansions are sub-optimal in at least some circumstances. In the following, a partial removal of this sub-optimality for the two dimensional case will be presented.

The starting point of our derivation is a novel expansion of the Green’s function that was introduced in Bogaert and Olyslager [2009a]. In the following, the wavenumber
will be denoted by \( k \), and the aggregation and translation vector respectively by

\[
\rho_A = \rho_A \left[ \cos \phi_A x + \sin \phi_A y \right],
\]

\[
\rho_T = \rho_T \left[ \cos \phi_T x + \sin \phi_T y \right].
\]

(17) (18)

The sum of the aggregation and translation vector will be denoted by \( \rho = \rho_A + \rho_T \).

In the following, we will assume that \( \rho_A \) is always confined to a square region with \( x \) and \( y \) ranges given by \([-b, b] \times [-b, b]\). This square region will be conveniently called 'FMM box' in the following, although it actually is the Minkowski sum of the source and observer box (see also Bogaert et al. [2008b]). It will also be assumed that \( \rho_T > b \sqrt{2} \). The expansion introduced in Bogaert and Olyslager [2009a] is defined in terms of the following inner product

\[
\langle f(\phi) | g(\phi) \rangle_{k\rho_T} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}_{k\rho_T}(\phi_i) \int_0^{2\pi} f(\phi_r + j\phi_i) \overline{g(\phi_r + j\phi_i)} d\phi_r d\phi_i.
\]

(19)

where \( \overline{z} \) denotes the complex conjugation of \( z \). Also the function \( \mathcal{F}_{k\rho_T}(\phi_i) \) is defined by the property

\[
\int_{-\infty}^{\infty} \mathcal{F}_{k\rho_T}(\phi_i) e^{-2n\phi_i} d\phi_i = \left| H_n^{(2)}(k\rho_T) \right|^2.
\]

(20)

Although this property does not uniquely determine \( \mathcal{F}_{k\rho_T}(\phi_i) \), the ambiguity can be easily resolved by enforcing (20) not only for integer \( n \), but for all real values of \( n \). When this is done, it is easily seen that the left hand side of (20) represents the two-sided Laplace transform of \( \mathcal{F}_{k\rho_T}(\phi_i) \). This Laplace transform can be inverted,
yielding an integral representation for $F_{k\rho T}(\phi_i)$

$$F_{k\rho T}(\phi_i) = \frac{1}{\pi} \int_{-\infty}^{\infty} H_{js}^{(2)}(k\rho_T) H_{js}^{(1)}(T\rho_T) e^{2js\phi_i} ds. \quad (21)$$

Note that the Hankel functions occurring in this expression are holomorphic functions of their index as long as $k\rho_T \neq 0$ Abramowitz and Stegun [1965]. The novel expansion of the Green’s function is now given by

$$H_{0}^{(2)}(k\rho) = \langle e^{-j\mathbf{k} \cdot \mathbf{\rho}_A} | \mathcal{T}(k\rho_T, \phi) \rangle_{k\rho_T}, \quad (22)$$

in which $\mathbf{k}(\phi)$ equals $k [\cos \phi x + \sin \phi y]$ and the translation operator $\mathcal{T}(k\rho_T, \phi)$ is given by

$$\mathcal{T}(k\rho_T, \phi) = \sum_{m=-\infty}^{\infty} j^{-m} e^{jm(\phi - \phi_T)} \frac{H_{m}^{(2)}(k\rho_T)}{H_{m}^{(2)}(k\rho_T)}. \quad (23)$$

The fact that the translation operator features a division by a Hankel function instead of a multiplication makes expansion (22) numerically stable for all frequencies.

The integral over the real and imaginary part of $\phi$ in (19) can be numerically computed by means of a suitable quadrature rule $\phi_p \forall p \in [1, P]$. However, in general this leads to a very large $P$, making expansion (22) uncompetitive with any of the expansions mentioned before. To reduce the number of integration points a.k.a. plane waves, a QR-based technique can be applied (as was done in Bogaert et al. [2010]). This technique takes advantage of the fact that $P$ is much larger than the number of degrees of freedom in the radiation pattern of any box. Indeed, the number of linearly independent radiation patterns that a box can radiate is limited. Therefore it is possible to sample the radiation pattern in a subset of the $\phi_p$ points, denoted by $\phi_{il} \forall l \in [1, L]$, and reconstruct the radiation pattern in all $\phi_p$ by means of interpolat-
The QR decomposition (with pivoting) is ideally suited to obtain this subset. First, the following matrix is constructed

\[
[M]_{iqp} = \sqrt{\mathcal{F}(\Im \phi_p)} e^{-jk(\phi_p) \cdot \rho_q}
\]  

(24)

with \(\rho_q \forall q \in [1, Q]\) a collection of points, which will be discussed further on. Up to the scaling factor \(\sqrt{\mathcal{F}(\Im \phi_p)}\), the rows of \(M\) are the radiation patterns of \(Q\) sources, located at \(\rho_q\). The rank of \(M\) is mainly determined by the number of linearly independent radiation patterns, which is influenced by the points \(\rho_q\). The size \(L\) of the selected subset will then in turn be determined by this rank. The selection procedure starts by taking the QR decomposition of \(M\). The QR algorithm returns as its output a unitary matrix \(Q\), an upper triangular matrix \(R\) and a permutation matrix \(P\), such that \(M \cdot P = Q \cdot R\). These matrices can be partitioned as follows

\[
[M] = \begin{bmatrix} Q_{11} & Q_{12} \end{bmatrix} \cdot \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \cdot \begin{bmatrix} P_{11}^T \\ P_{12}^T \end{bmatrix},
\]  

(25)

The size \(L\) of the square matrix \(R_{11}\) is chosen such that \(\frac{\|R_{22}\|}{\|R_{11}\|}\) is below the required precision. This means that \(L\) is the numerical rank of \(M\). The other matrices are then partitioned such that \(P_{11}\) and \(Q_{11}\) have size \(P\) by \(L\) and \(Q\) by \(L\) respectively. Now consider the matrix \(M \cdot P_{11}\). This matrix contains only the columns that were selected by the QR algorithm. Since, the columns from \(M\) selected by \(P_{11}\) span in very good approximation the range of \(M\), the other columns can be written approximately as linear combinations of the former. This yields the following low rank representation.
of $M$:

$$M = [M \cdot P_{11}] \cdot \left[ [1 \cdot R_{11}^{-1} \cdot R_{12}] \cdot P^T + E \right],$$

(26)

where $\|E\| = O(\|R_{22}\|)$, as can be seen by substituting (25) in (26). The matrix $D$ can be interpreted as an interpolation matrix that maps the selected points onto all the $\phi_p$ points. Therefore it becomes possible to use only the selected points which are given by

$$\phi_i^s = \sum_{p=1}^{P} \phi_p [P_{11}]_{p,i}, \forall l \in [1, L].$$

(27)

We now turn to the choice of the points $\rho_q$. These points determine the space of radiation patterns that can be radiated and guide the QR in its choice of columns. In Bogaert et al. [2010] the points were chosen on a circle around the FMM box. This point configuration is shown in Figure 4. Figure 5 shows the size $L$ of the selected subset as a function of the attained accuracy for an example configuration with $b = 1m$, $\rho_T = 2m$, $\phi_T = 0$ and $k = 0.01m^{-1}$. As can be seen, the number of plane waves selected by the QR is only slightly higher than the number of plane waves that would be required by the MLFMA (obtained using Gegenbauer’s addition theorem) if it were stable. This is due to the fact that the MLFMA is based on spherical modes and the fact that we chose a circular configuration of points.

In this contribution, however, we make the observation that points $\rho_q$ on a circle around the FMM box allow for too much freedom in the currents. Indeed, because the currents cannot be outside of the square-shaped FMM box, we could just as well place the points $\rho_q$ on the boundary of the FMM box, as depicted in Figure 4. The
net result is that the QR selects fewer plane waves. Figure 5 shows that the number of plane waves is drastically reduced when compared to the case where the points $\rho_q$ are placed on a circle. A significant gain in comparison to the MLFMA is also observed, showing that taking into account the square shape of FMM boxes can significantly reduce the number of plane waves. To the best of the author’s knowledge, this is also the first plane wave expansion that has its plane waves tailored to the square shape of the FMM boxes.

5. Distributed memory parallelization

Large scale simulations are often impractical using a sequential code, either because the memory requirements exceed the capacity of a single workstation, or because the runtime on a single CPU is simply too high. The distribution of both the computations and memory over a large number of machines (i.e. parallelization) allows for larger simulations and a significant reduction in execution time. For techniques such as Finite Difference Time Domain, Finite Elements and Finite Volumes, this parallelization is readily achieved. Indeed, these techniques make use of a very sparse and highly structured system matrix, which can be evenly distributed among a number of machines while requiring very little communication between these machines. However, in the case of the Method of Moments, the impedance matrix is dense and the algorithm is further complicated by the use of Fast Multipole Methods. Consequently, the distribution of computational effort and memory usage requires a significant effort and has lately been the topic of intense research. The original focus
was on high frequency problems of ever increasing size, which advanced from a purely spatial partitioning (i.e. the distributing of the boxes in the tree) to a hybrid scheme that includes k-space partitioning (i.e. the distribution of the sampling points within a box) and finally, to a fully scalable hierarchical partitioning, as will be elaborated upon further. A second trend focuses on the distribution problems that arise when modeling complicated structures, featuring many different objects of various size. Advances include an asynchronous approach and the usage of a so-called adaptive tree, both of which will be discussed below.

First, however, let us take a closer look at what has historically been the prime interest of parallelization research, namely the attempt to model electrically huge objects (e.g. airplanes at radar frequencies). The most recent (and possibly definitive) breakthrough is the hierarchical partitioning approach Ergül and Gürel [2008], which can be proven to lead to a scalable parallel implementation Fostier and Olyslager [2008b]. The term scalable denotes that larger simulations can gracefully be handled using a proportionally higher number of CPUs, without loss of efficiency. Spatial and hybrid partitioning strategies Velamparambil and Chew [2005] do not lead to scalable algorithms. The problem with spatial partitioning is that, as we progress upwards in the tree, the top levels will have fewer boxes than there are machines available. Obviously, this means that the load can no longer be balanced properly in an algorithm that assigns the entire radiation pattern for one box to a single machine. Clearly, by increasing the number of machines, this problems occurs even earlier in the tree, hence limiting the number of machines that can effectively be used.
One solution is to introduce a so-called ‘split level’ or ‘level of distribution’, above which the radiation patterns are distributed over all machines. While a significant improvement, this does not lead to a scalable algorithm either, as can be shown by the required amount of communication between the nodes. The hierarchical approach refines this concept by introducing a gradual transition from spatial partitioning to purely k-space partitioning, by distributing radiation patterns over an increasing amount of nodes while progressing upwards in the tree. In two dimensions, this concept has been proven both theoretically and experimentally Fostier and Olyslager [2008b]. At the moment of writing, an implementation in three dimensions is still a work in progress and has, to the best of the authors’ knowledge, not yet been achieved anywhere. This is due to the fact that, at least from an implementation point of view, the algorithm is very complicated.

A second advance is the development of asynchronous parallel schemes Fostier and Olyslager [2008a]. Essentially, it allows for the different machines to perform different tasks at any given moment, while still satisfying the necessary order relations imposed by FMM. This is in contrast with the first (and still the majority of) parallelization schemes, that were purely synchronous, i.e. where all nodes are performing the same kind of computations (e.g. an aggregation) before simultaneously moving on to the next phase (e.g. communication of the radiation patterns to those nodes that need it). This leads to a number of issues, one of them being congestion of the network due to these communication bursts. An asynchronous technique works by means of a priority queue, to which work packages are added as soon as they are allowed to be
calculated (i.e. a disaggregation at certain box can not happen before all incoming radiation patterns have been added to it). The ordering of the work within the priority queue is largely heuristic and favors computations that could be needed by other nodes. When the structure contains many objects, the asynchronous approach becomes even more powerful. The boxes are distributed among the nodes through a global space filling curve, such that large objects are distributed over many nodes whereas very small objects may be contained by as little as a single node. Also, as the FMM computations in different dielectric domains are fully decoupled from each other, the asynchronous algorithm gains a lot by its additional degree of flexibility.

A final technique that was developed lately is the adaptive tree. When a traditional FMM tree is used for a strongly non-uniform mesh, this will invariably lead to problems. The lowest level minimal box size (MBS) is determined by the radius of the largest expansion function, in order to maintain accuracy. However, if regions of both high and low mesh density occur, the efficiency of the FMM algorithm is destroyed by this lower limit on the MBS. One way to effectively counter this issue is by allowing certain branches of the tree to extend as deep as needed, depending on the local density of the mesh. This results in an algorithm where the lowest level aggregation (from basis functions to radiation pattern) differs throughout the geometry. A first step, that allowed only one leaf, was made in Peeters [2010]. A further generalization, that also allows aggregation of basis functions at different levels within the same branch, has recently been achieved. Aside from reducing the number of calculations (single processor complexity), this algorithm also lends itself well to
parallelization, because the adaptive tree is more suitable for efficient load balancing. However, it is precisely with regard to this load balancing that a lot of work has yet to be covered in the future. Dynamic load balancing, which allows reorganization of the work packages during the iterative solution, is probably not within immediate reach. Hence, we are limited to static load balancing, which attempts to approach ideal load balancing before the iterative process, in a semi-heuristic manner. Due to the complicated nature of the FMM algorithm, this can currently only be done in a rather approximative manner. One can easily imagine how a perfect optimization would require detailed information about the speed of the network, the computational cost of various tasks (like the matrix-vector product with BLAS), etc. On a fairly slow network, one may even notice that iteration times vary throughout the solution process. Our current approach is to disregard individual hardware specifics and balance the load with regard to the memory required for near interactions and aggregations from basis functions to radiation patterns. As it happens, this memory-oriented approach is also very successful in balancing the calculation time for all practical examples that we tried to date.

As an example, we will look at the simulation of a solar cell that has small silver nanoparticles in the active layer to benefit from the plasmonic resonance. The field is localized around these particles, which could potentially lead to higher efficiencies. The solar cell itself is modeled as a (finite) layered structure of $3\mu m \times 3\mu m \times 0.9\mu m$, as also shown in Figure 6. The wavelength is 470nm. The characteristic size of the basis functions for the solar cell itself is about 40nm. The nanoparticles, which have
a radius of 5nm are meshed in triangles that are roughly 1nm in size. Without the adaptive tree algorithm, the lowest level MBS would have to be larger than 100nm. Each nanoparticle requires about 3500 degrees of freedom and in this example we consider a closely packed 9 × 9 grid. This would result in roughly 275,000 unknowns in one box, which requires 1 TByte of memory for the near interactions alone. In the adaptive tree, however, the MBS near the nanoparticles is 4nm and most expansion functions on the solar cell boundary are located 5 levels higher (near corners and junctions the mesh density is slightly higher and those expansion functions are automatically stocked one level below the majority). The total number of unknowns is 475,000. One iteration on 64 nodes requires roughly 15 seconds. Previous implementations, that completely disregarded adaptivity or did so only partially, would not fit in the memory or take multiple days to solve. The global solution time for the fully adaptive case is under one hour. Figure 7 shows the absolute value of the x-component of the total electric field. As can be observed, there is a sharp peak at z = 550nm, where the output line goes straight through a nanoparticle.

**Figure Captions**

1. Definition of a Buffa-Christiansen function as a linear combination of RWGs on the barycentric refinement. All edges are oriented away from the vertices of the original mesh. Weights are multiplied by the number indicated on the vertex of the original mesh.
2. Relative $L^2(\Gamma)$, $H^{-1/2}(\text{div}; \Gamma)$, and $H(\text{div}; \Gamma)$ errors on the current resulting from a classic and a mixed discretization of the MFI applied to a unit side cube.

3. Condition number of the CP-PMCHWT system matrix.

4. The used circular and square point configurations. The solid line denotes the area where currents may occur.

5. The required number of plane waves as a function of the target accuracy for the circular and square point configurations. The solid line shows the required number of plane waves for the MLFMA if it were stable.

6. The geometry of the solar cell, including the nanoparticles (not to scale). The incident plane wave ($\lambda = 470 \text{nm}$) is z-directed and has a linearly polarized E-field along the x-direction. The nanoparticles have a radius of $5 \text{nm}$ and their centers are located on a grid with $15 \text{nm}$ spacing. The relative permittivity for the $\text{SiO}_2$ is $2.13$, that for the active layer is $4 - i0.4$, that for the electrode is $-2 - i$ and finally that for the nanoparticles is $-8 - i$.

7. The absolute value of $E_x$ along the central axis of the solar cell. The sharp peak is due to the plasmonic resonance of the nanoparticles. The interface between air and solar cell is located at $z = 0$. 


**Figure 1.** Definition of a Buffa-Christiansen function as a linear combination of RWGs on the barycentric refinement. All edges are oriented away from the vertices of the original mesh. Weights are multiplied by the number indicated on the vertex of the original mesh.
Figure 2. Relative $L^2(\Gamma)$, $H^{-1/2}(\text{div}; \Gamma)$, and $H(\text{div}; \Gamma)$ errors on the current resulting from a classic and a mixed discretization of the MFIE applied to a unit side cube.

Figure 3. Condition number of the CP-PMCHWT system matrix.
Figure 4. The used circular and square point configurations. The solid line denotes the area where currents may occur.

Figure 5. The required number of plane waves as a function of the target accuracy for the circular and square point configurations. The solid line shows the required number of plane waves for the MLFMA if it were stable.
Figure 6. The geometry of the solar cell, including the nanoparticles (not to scale).

The incident plane wave ($\lambda = 470\,nm$) is z-directed and has a linearly polarized E-field along the x-direction. The nanoparticles have a radius of 5$nm$ and their centers are located on a grid with 15$nm$ spacing. The relative permittivity for the $SiO_2$ is 2.13, that for the active layer is $4 - i0.4$, that for the electrode is $-2 - i$ and finally that for the nanoparticles is $-8 - i$.

Figure 7. The absolute value of $E_x$ along the central axis of the solar cell. The sharp peak is due to the plasmonic resonance of the nanoparticles. The interface between air and solar cell is located at $z = 0$. 

References


Ubeda, E., and J. Rius, Advantages of rectangular and uniform-triangluar discretization on the scattering analysis of very small sharp-edged objects with the magnetic field integral equation, in *USNC/URSI National Radio Science Meeting*, 2006a.


Vande Ginste, D., E. Michielssen, F. Olyslager, and D. D. Zutter, A high-performance upgrade of the perfectly matched layer multilevel fast multipole algorithm for large planar


