GENERAL PURPOSE AND SCOPE: The Applied Computational Electromagnetics Society *(ACES)* Journal hereinafter known as the *ACES Journal* is devoted to the exchange of information in computational electromagnetics, to the advancement of the state-of-the art, and the promotion of related technical activities. The primary objective of the information exchange is to inform the scientific community on the developments of new computational electromagnetics tools and their use in electrical engineering, physics, or related areas. The technical activities promoted by this publication include code validation, performance analysis, and input/output standardization; code or technique optimization and error minimization; innovations in solution technique or in data input/output; identification of new applications for electromagnetics modeling codes and techniques; integration of computational electromagnetics techniques with new computer architectures; and correlation of computational parameters with physical mechanisms.

SUBMISSIONS: The *ACES Journal* welcomes original, previously unpublished papers, relating to applied computational electromagnetics. Typical papers will represent the computational electromagnetics aspects of research in electrical engineering, physics, or related disciplines. However, papers which represent research in applied computational electromagnetics itself are equally acceptable.

Manuscripts are to be submitted through the upload system of *ACES* web site http://aces.ee.olemiss.edu See "Information for Authors" on inside of back cover and at *ACES* web site. For additional information contact the Editor-in-Chief:

Dr. Atef Elsherbeni
Department of Electrical Engineering
The University of Mississippi
University, MS 386377 USA
Phone: 662-915-5382
Email: atef@olemiss.edu

SUBSCRIPTIONS: All members of the Applied Computational Electromagnetics Society are entitled to access and download the *ACES Journal* any published journal article available at http://aces.ee.olemiss.edu. Printed issues of the *ACES Journal* are delivered to institutional members. Each author of published papers receives a printed issue of the *ACES Journal* in which the paper is published.

Back issues, when available, are $50 each. Subscription to *ACES* is through the web site. Orders for back issues of the *ACES Journal* and change of address requests should be sent directly to *ACES* office at:

Department of Electrical Engineering
The University of Mississippi
University, MS 386377 USA
Phone: 662-915-7231
Email: aglisson@olemiss.edu

Allow four weeks advance notice for change of address. Claims for missing issues will not be honored because of insufficient notice, or address change, or loss in the mail unless the *ACES* office is notified within 60 days for USA and Canadian subscribers, or 90 days for subscribers in other countries, from the last day of the month of publication. For information regarding reprints of individual papers or other materials, see "Information for Authors".

LIABILITY. Neither *ACES*, nor the *ACES Journal* editors, are responsible for any consequence of misinformation or claims, express or implied, in any published material in an *ACES Journal* issue. This also applies to advertising, for which only camera-ready copies are accepted. Authors are responsible for information contained in their papers. If any material submitted for publication includes material which has already been published elsewhere, it is the author's responsibility to obtain written permission to reproduce such material.

APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY JOURNAL

**Special Issue on
Computational Electromagnetics
Workshop, CEM 11**

Guest Editor
Levent Gürel

February 2012
Vol. 27 No. 2
ISSN 1054-4887

FEBRUARY 2012 REVIEWERS

Ahmed Abdelrahman

Hakan Bagci

R. Capoglu

Ragh Vendra Chaudhary

Veysel Demir

Ozgur Ergul

Roberto Graglia

Levent Gurel

Nikolaos Kantartzis

Branislav Notaros

Abbas Omar

Andrew Peterson

C. J. Reddy

Fernando Teixeira

Ahmet Turk

Mohammed Zunoubi

THE APPLIED COMPUTATIONAL ELECTROMAGNETICS SOCIETY

JOURNAL

TABLE OF CONTENTS

Macro-Modeling of Electromagnetic Domains Exhibiting Geometric and Material Uncertainty

Juan S. Ochoa and Andreas C. Cangellaris

Department of Electrical and Computer Engineering
University of Illinois, Urbana, IL, 61801, U.S.A.
jsochoa2@illinois.edu, cangella@illinois.edu

Abstract — A methodology is presented for the development of stochastic electromagnetic macro-models for domains exhibiting geometric and material uncertainty. Focusing on the case of domains exhibiting geometric/material invariance along one of the axes of the reference coordinate system, the methodology makes use of the theory of polynomial chaos expansion and the concept of a global impedance/admittance matrix relationship defined over a circular surface enclosing the cross-sectional geometry of the domain of interest. The result is a stochastic global impedance/admittance matrix, defined on the enclosing circular surface, whose elements are truncated polynomial chaos expansions over the random space defined by the independent random variables that parameterize the geometric and material uncertainty inside the domain. Use is made of sparse Smolyak grids to reduce the computational cost of constructing the stochastic macro-model. Numerical examples are used to demonstrate some of the attributes of the proposed stochastic macro-models to the numerical solution of electromagnetic scattering problems by an ensemble of cylindrical targets exhibiting uncertainty in their shape and relative positioning.

Index Terms — Finite elements, global impedance matrix, macro-modeling, polynomial chaos expansion, random geometry, scattering/RCS.

I. INTRODUCTION

Over the years the term macro-modeling has been used in scientific and engineering modeling and simulation to describe a variety of things. In the context of electromagnetic (EM) modeling and simulation, macro-modeling is widely understood to mean the process through which a compact physical or mathematical model is defined to describe the EM attributes of a portion of the system, the detailed description of which requires a large number of degrees of freedom (state variables) for its modeling. In this context, low-order, EM macro-models have been used extensively for a variety of applications. These include, expedient calculation of the broadband response of passive EM devices; use of domain decomposition techniques for the EM field modeling of electrically-large structures of high complexity; and the abstraction of distributed portions of composite systems that include both lumped circuit components and distributed electromagnetic structures.

For the purposes of this paper, we are interested in the macro-modeling of portions of linear, passive EM structures that exhibit geometric and/or material uncertainty. In particular, we are interested in structures formed as an ensemble of multiple domains, with the aforementioned material and geometric uncertainty occurring in some or all of these domains but not in the medium in which these domains are immersed. Figure 1 depicts a representative example of such a composite structure.

The problem of electromagnetic modeling of structures that exhibit randomness is one of significant interest to the electromagnetics community because of its relevance to several application domains such as remote sensing, EMI/EMC in electronic systems, and EM wave propagation in random media. For the case of EM wave scattering by composite random structures like the one depicted in Fig. 1, the complexity of a Monte Carlo numerical solution is compounded by the need to

generate a discrete numerical model for each one of the geometries resulting from the sampling of the multi-dimensional random space defining the randomness of the structure. For example, in the context of the finite element solution of the EM boundary value problem (BVP), a new finite element grid needs to be generated for each one of the realizations of the geometry during the Monte Carlo sampling of the random space. An alternative approach, presented by Miller [1], makes use of an adaptive methodology capable of constraining an estimated radiation or scattering pattern to satisfy an uncertainty specification by employing appropriate fitting models to minimize the required number of samples.

In this paper, a macro-modeling methodology is proposed as a means of alleviate the repeated discretization of the computational domain in the numerical solution of the stochastic EM BVP. The proposed methodology makes use of the mathematical framework of polynomial chaos expansions and stochastic collocation [2– 6], which has been applied recently to the numerical solution of a variety of EM BVPs (see [7–11] for representative examples). These are combined with the concept of network matrix representation of passive EM structures to develop a compact stochastic impedance (or admittance) matrix macro-model on a fixed boundary enclosing each one of the domains that exhibits randomness. In this manner, only a single numerical grid is needed for the Monte Carlo solution of the EM scattering by the ensemble of the random domains.

The proposed approach is described in Section II. Section III presents examples from the application of the method to the two-dimensional scattering by arrays of conducting cylinders exhibiting geometric randomness. The paper concludes with some remarks on future extensions of the method.

II. STOCHASTIC MACRO-MODELING

As suggested in the introduction, the electromagnetic structures of interest to this discussion are composite structures comprising several subdomains, with a good number of them exhibiting uncertainty in their material and/or geometric composition. The EM analysis of such a structure using, for example, a Monte Carlo (MC) process, requires the development of as many FE/FD models (including the generation of an FE/FD mesh for each model) as the samples in the random space used in the MC process. One way to reduce the associated computational cost is by removing the need for the repeated mesh generation. The way stochastic macro-modeling makes this possible is demonstrated in this section.

To fix ideas, we will consider the case of electromagnetic wave scattering by a collection of targets embedded in an unbounded linear host medium. While the geometric and material attributes of each one of the targets exhibit a statistically defined randomness, the host medium does not. Figure 1 serves as a representative example of such a structure. Any randomness in geometric attributes or material properties occurs only inside the N regions $V_1, V_2, \cdots V_N$ bounded by surfaces $S_1, S_2, \cdots S_N$, respectively. The exterior medium, including the volume bounded by surface S_0, is assumed to be fixed in terms of its geometric attributes and its electromagnetic properties. In view of this, it is immediate apparent that, under the assumption that the N surfaces $S_1, S_2, \cdots S_N$ are fixed, the domain V_E bounded by these N surfaces and the surface at infinity is a fixed domain free from any geometric or material uncertainty. The way the randomness of the interiors of the domains $V_n, n = 1, 2, \cdots, N$, manifests itself in the solution of the exterior BVP in V_E is through the boundary conditions on the surfaces. This, then, suggests the idea of a stochastic macro-model for each one of the N subdomains $V_n, n = 1, 2, \cdots, N$, in terms of a global surface impedance relationship on $S_n, n = 1, 2, \cdots, N$. The way this is done is described next.
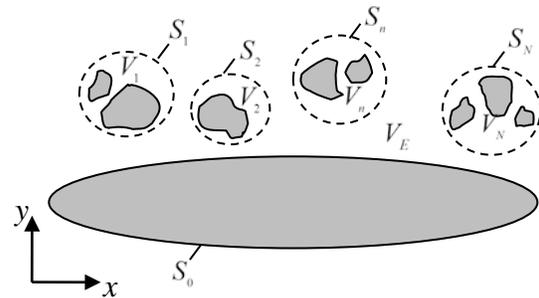
Fig. 1. Reference geometry for the discussion of the concept of stochastic macro-modeling.

A. Global impedance matrix

For the purposes of this discussion, let the geometry of Fig. 1 be the cross-sectional geometry of an infinitely long cylindrical scatterer. Under the assumption that both the cross-sectional geometry of the scatterer and the exciting electromagnetic field are z invariant, the pertinent electromagnetic boundary value problem is a two-dimensional one. Focusing on the case of TE_z polarization, where the magnetic field is linearly polarized in the z direction, we assume, without loss of generality, that S_n is a circle. A local reference coordinate system is introduced, with origin the center of the circular boundary S_n. Fourier series expansions in the polar angle ϕ are used to represent the tangential electric and tangential magnetic fields on S_n. For each Fourier mode in the expansion of the tangential magnetic field on S_n the solution of the interior BVP in V_n yields a tangential electric field on S_n. In this manner, a global impedance condition is established on S_n, defined in terms of the matrix relationship,

$$\xi_k = \sum_{m=-M}^{M} Z_{km}^{(n)} h_m, \, k = 0, \pm 1, ..., \pm M \; , \quad (1)$$

where ξ_m, h_m are, respectively, the coefficients in the Fourier series expansions of the tangential electric field and the tangential magnetic field on S_n,

$$E_\phi \approx \sum_{m=-M}^{M} \xi_m e^{jm\phi}, \, H_z \approx \sum_{m=-M}^{M} h_m e^{jm\phi}. \quad (2)$$

The truncation of the expansions in (2) is necessary for the numerical implementation of (1). The important observation here is that the global impedance matrix, $\mathbf{Z}^{(n)}$, defined through (1), (2), serves as an electromagnetic macro-model for the region V_n. Once the impedance matrices for all domains $V_n, n = 1, 2, \cdots, N$, are available, the solution to the exterior electromagnetic BVP in V_E due to an arbitrary excitation at the frequency of interest is computed in a straightforward fashion. In the presence of geometric and/or material uncertainty in V_n, the elements of the global impedance matrix $\mathbf{Z}^{(n)}$ can be used to account for the impact of the randomness of the region to the electromagnetic response of the overall structure. By abstracting the randomness of the interior region on the global impedance matrix defined on a fixed boundary, a single numerical grid is necessary for the solution of the exterior BVP. In the next subsection, a process is described for abstracting the randomness in the geometric and/or material properties of the region V_n to the elements of the global impedance matrix $\mathbf{Z}^{(n)}$ on the fixed boundary S_n.

B. Stochastic global impedance matrix

Let $\chi = (\chi_1, \chi_2, \cdots, \chi_D)$ denote the set of independent random variables necessary for describing the uncertainty in V_n. Furthermore, let $\rho(\chi)$ denote their joint probability density function. The objective is to develop a systematic and expedient process for obtaining a global impedance matrix $\mathbf{Z}^{(n)}$ that serves as an accurate macro-model of the electromagnetic attributes of V_n for any point in the D-dimensional probability space Ω defined by $\chi = (\chi_1, \chi_2, \cdots, \chi_D)$.

Toward this objective, use is made of the machinery of polynomial chaos expansion of random functions in Ω. Following the ideas in [4], a truncated polynomial chaos expansion of $Z_{km}^{(n)}$ is of the form

$$Z_{km}^{(n)} \approx \sum_{i=0}^{Q} c_i \Psi_i(\chi), \quad (3)$$

where $\Psi_i(\chi)$ are multidimensional orthogonal polynomials with regard to the inner product,

$$\langle \Psi_i, \Psi_j \rangle \equiv \int_\Omega \Psi_i(\chi) \Psi_j(\chi) \rho(\chi) d\chi = \delta_{ij} \| \Psi_i \|^2. \quad (4)$$

The type of random variables dictates the family of the polynomials to be used [6]. For example, for the case of Gaussian random variables, Hermite polynomials are used. The number of terms, Q, included in the truncated polynomial chaos expansion depends on the dimensionality D of the random space and the highest order p of the multidimensional polynomials used, and is given by

$$Q + 1 = \frac{(D+p)!}{D!\,p!}. \quad (5)$$

In view of (3), the coefficients in the polynomial chaos approximation of $Z_{km}^{(n)}$ are computed using the orthogonality relation (4),

$$c_i = \frac{1}{\| \Psi_i \|^2} \int_\Omega \rho(\chi) \Psi_i(\chi) Z_{km}^{(n)}(\chi) d\chi. \quad (6)$$

Clearly, the expedient calculation of the integral in (6) calls for an efficient multivariate quadrature rule on Ω. For example, use of the Smolyak algorithm [12] leads to the approximation of (6) through the summation,

$$c_i \approx \frac{1}{\|\Psi_i\|^2} \sum_{r=1}^{R} \rho(\boldsymbol{\chi}_r) \Psi_i(\boldsymbol{\chi}_r) Z_{km}(\boldsymbol{\chi}_r) w_r, \qquad (7)$$

where the number of nodes, R, is significantly less than the one required by a tensor product rule. The selection of the quadrature points, their weights w_r, and the level of accuracy that dictates the sparsity of the Smolyak grid are well documented in the literature and will not be repeated here (see [12–14] for details).

Equations (3) and (7) define the desired stochastic global impedance matrix macro-model of the random domain V_n. The process for its construction is summarized in terms of the following algorithm.

C. Algorithm

The stochastic global impedance matrix macro-modeling of domain V_n bounded by S_n is summarized next.

1. Choose the dimension of the global impedance matrix and, hence, the number of Fourier modes used in the expansion of the tangential electric and magnetic fields on S_n.
2. Represent geometric/material randomness in terms of D independent random variables $\boldsymbol{\chi} = (\chi_1, \chi_2, \cdots, \chi_D)$.
3. Choose polynomial family and order for truncated polynomial chaos expansion.
4. Generate Smolyak grid on probability space Ω defined by $\boldsymbol{\chi} = (\chi_1, \chi_2, \cdots, \chi_D)$.
5. For each point $\boldsymbol{\chi}_r, r = 1, 2, \cdots, R$, on the Smolyak grid, solve the deterministic interior BVP to obtain $\mathbf{Z}^{(n)}(\boldsymbol{\chi}_r)$.
6. Using the matrices obtained in Step 5, calculate the coefficients in the polynomial chaos expansion of $\mathbf{Z}^{(n)}$ using (7).

D. Solution of the exterior stochastic BVP

Once the stochastic global impedance matrices on the fixed circular boundaries $S_n, n = 1, 2, \cdots, N$, have been constructed, the numerical solution of the electromagnetic scattering problem by the union of the $N+1$ targets, $V_n, n = 0, 1, 2, \cdots, N$, amounts to solving an exterior electromagnetic BVP in V_E. As already stated, since the circular boundaries are fixed, the finite element solution of this exterior BVP requires a single numerical grid. The randomness of each one of the N regions manifests itself in terms of the polynomial chaos expansions

of the elements of its stochastic impedance matrix. With D_n denoting the number of independent random variables used to parameterize the uncertainty in V_n, the dimension of the random space Ω_E for the exterior stochastic BVP is $D_E = \sum_{n=1}^{N} D_n$. Irrespective of the process used for the solution of the exterior stochastic BVP, the global impedance matrix on each one of the circular boundaries is readily computed Ω_E from its polynomial chaos expansion for each sample in the random space.

III. NUMERICAL VALIDATION AND DEMONSTRATION STUDIES

In this section, several numerical examples involving electromagnetic wave scattering by arrays of infinitely long cylinders are used to validate the proposed methodology and demonstrate its key attributes.

We begin with the problem of TE_z time-harmonic uniform plane wave scattering by a perfect electric cylindrical conductor of circular cross section and of random radius, $a = 0.80 \, (1+\chi)$ m, where χ is a Gaussian random variable with zero mean and standard variation of 0.06 m. The cylinder is immersed in free space and its axis coincides with the z axis of the reference coordinate system. The amplitude of the incident magnetic field is 1 A/m and its angular frequency is 9×10^8 rad/s. The availability of an analytic solution for this problem makes possible the use of a standard Monte Carlo analysis to calculate the reference solution for the statistics of the scattered magnetic field on a circle of radius 1.2 m centered at the origin. Use of 10^4 sampling points in the Monte Carlo process yielded an accuracy of 10^{-5} in the calculation of the mean value of the magnitude of the scattered magnetic field.

Next, the problem was solved making use of the stochastic global impedance condition defined over the circle of radius 1.2 m. The polynomial chaos approximation of the elements of the impedance matrix is in terms of Hermite polynomials up to third order. The dimension of the impedance matrix is 11. Since the dimension of the random space is 1, the Smolyak grid reduces to a simple Gaussian quadrature rule. For accuracy level of 5, a Smolyak involving 5 grid points is required for the calculation of the integrals in (7). Since the random space for the exterior stochastic BVP is

the same with that for the interior, the same Smolyak grid used for the construction of the global impedance matrix is used for solving the exterior stochastic BVP. The mean and variance of the magnitude of the scattered magnetic field thus computed are compared with those obtained analytically in Fig. 2. The error bars represent a $\pm 3\sigma$ deviation with respect to the mean. Very good agreement is observed. More specifically, the average error in the mean value of the magnitude of the scattered magnetic field between the analytical and the numerical solution, averaged over all angles, is 1.20%.
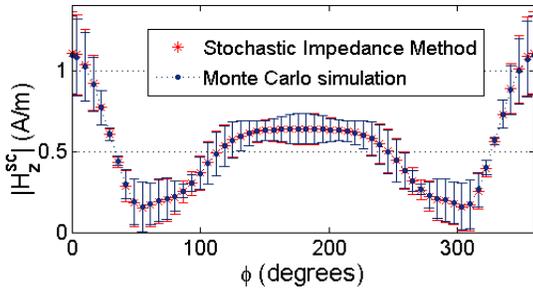
Fig. 2. Mean value of the magnitude of the scattered magnetic field. Error bars represent $\pm 3\sigma$ deviation from the mean.

Next, the case of TE_z wave scattering by an array of four elliptical cylinders is considered. All cylinders are perfect electric conductors, and the background medium is free space. The angular frequency of the excitation is 9×10^8 rad/s. In the absence of any statistical variability, the centers of the four cylinders coincide with the vertices of a square of side 2.4 m (see Fig. 3). The randomness in the cross-sectional geometry is introduced through a set of four independent random variables for each cylinder. Two of them, χ_1, χ_2, are associated with the lengths $2a$ and $2b$ of the major axis (along the x axis) and minor axis (along the y axis), respectively, of the elliptical cylinder. More specifically, with the two random variables taken to be Gaussian of mean value of 0 and standard deviation of 0.025, the lengths of the two axes (in meters) are given by

$$2a(\chi_1) = 1.4(1+\chi_1), \quad 2b(\chi_2) = 1.4(1+\chi_2). \quad (8)$$

The other two random variables, χ_x, χ_y, control the random displacement of the center of the cylinder from the vertex of the reference square. The position of the center is given by

$$\vec{r}_c = \vec{r}_o + \hat{x}\chi_x + \hat{y}\chi_y, \quad (9)$$

where $\vec{r}_o$ denotes the corresponding position of the vertex of the square. Variables χ_x and χ_y are Gaussian random variables of zero mean value and standard deviation of 0.025 m. Even though the same four variables are being used to quantify the geometric uncertainty for each one of the four cylinders, when considering the four-cylinder array, the four sets of random variables are assumed to be independent. Thus, the randomness of the cross-sectional geometry is parameterized in terms of 16 independent random variables.
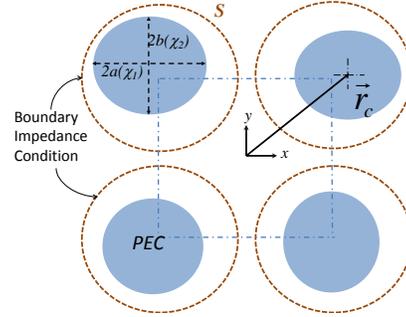
Fig. 3. Array of four PEC elliptical cylinders with random axes lengths and random positions.

Depicted in Fig. 3 are the four fixed circular boundaries on which stochastic global impedance boundary conditions will be defined, one for each one of the four cylinders. The center for each circle coincides with corresponding vertex of the reference square formed by the unperturbed centers of the four cylinders in the array. The radius of each circle is such that the cylinder associated with it remains enclosed by it for all points in the four-dimensional domain in the random space defined by the random variables χ_1, χ_2, χ_x, χ_y. For this specific example, this radius was taken to be 1 m. Given that the random variables are Gaussian distributions, orthogonal Hermite polynomials are used for the polynomial chaos expansion of the elements of the stochastic impedance matrix. The calculation of the coefficients in the polynomial chaos expansion using polynomials of up to second order is carried out efficiently through the use of a Smolyak sparse grid on the four-dimensional random space. More specifically, for

the case considered here, a Kronrod–Patterson rule [13], [14] of accuracy level 5 was used, resulting in 201 points on the four-dimensional random space. For each point on the Smolyak grid a finite element solution of the interior BVP was used to calculate a global impedance matrix of dimension 21. The computed matrices were subsequently used for the calculation of the coefficients in the polynomial chaos expansion of the stochastic impedance matrix making use of (7).

With the stochastic global impedance matrix available on each one of the four circular boundaries, a finite element model was used for the solution of the exterior stochastic BVP with excitation by a uniform, time-harmonic plane wave propagating in the $+x$ direction with magnetic field amplitude of 1 A/m. Since the circular stochastic impedance boundaries are fixed, a single finite element mesh is needed for the discretization of the geometry. The only changes to the finite element matrix are those associated with the specific values of the stochastic impedance matrices on the four boundaries for each sample realization in the 16-dimensional random space. Rather than a standard Monte Carlo process, a Smolyak sparse grid of accuracy level 3 was used to extract the statistics of the scattered fields and the radar cross section. The number of points in the sparse Smolyak grid is 513. The mean and variance of the output parameters are computed by performing the integration over the random space. These integrals are approximated by weighted summations of the scattered field, computed at each one of the 513 nodes as previously described,

$$\left\langle \left| H_z^{sc}(\rho,\varphi) \right| \right\rangle \approx \sum_{r=1}^{513} \left| H_z^{sc}(\rho,\varphi,\chi_r) \right| \rho(\chi_r) w_r \quad (10)$$

$$\text{var}\left(\left| H_z^{sc}(\rho,\varphi) \right| \right) \approx \sum_{r=1}^{513} \left| H_z^{sc}(\rho,\varphi,\chi_r) \right|^2 \rho(\chi_r) w_r$$
$$- \left\langle \left| H_z^{sc}(\rho,\varphi) \right| \right\rangle^2. \quad (11)$$

The calculated scattered magnetic field, sampled on a circle of radius 3.4 m centered at the center of the reference square defined by the unperturbed centers of the four cylinders is depicted in Fig. 4. More specifically, shown in the figure is the mean value of the magnitude of the scattered magnetic field along with error bars that indicate a $\pm 3\sigma$ deviation from the mean value. The radar cross section (RCS) is depicted in Fig. 5. Again,

the mean value is plotted, along with error bars that denote $\pm 3\sigma$ deviation from the mean.

The final numerical study considers the case where the polarization of the excitation is TM_z with the electric field linearly polarized along the z axis. For this case and in view of the fact that for the two-dimensional BVP considered the governing equation is the scalar Helmholtz equation for the z component of the magnetic field, a global admittance matrix is used instead of a global impedance matrix. The global admittance matrix relates the Fourier coefficients in the expansion of the tangential magnetic field on each one of the circular boundaries S_n to the Fourier coefficients in the expansion of the tangential electric field.
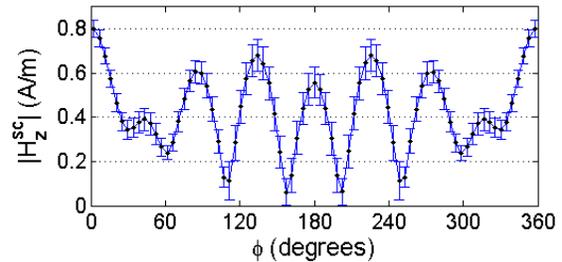
Fig. 4. Mean value of the magnitude of the scattered magnetic field on a circle of radium 3.8 m enclosing the four cylinders. Error bars represent $\pm 3\sigma$ deviation from the mean.
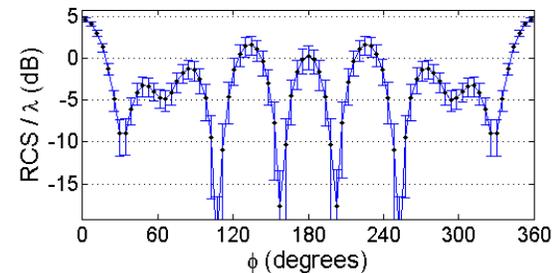
Fig. 5. Mean value of the radiation cross section (RCS) of the magnetic field. Error bars represent $\pm 3\sigma$ deviation from the mean.

For this case, the four-cylinder array depicted in Fig. 3 is illuminated by a time-harmonic, line current source of current phasor of 1 A, angular frequency 9×10^8 rad/s, and placed at position (3, 0) m with its axis parallel to the z axis. The development of the global stochastic admittance matrix, used on each one of the circular boundaries, was

carried out following the same choices for Hermite polynomial chaos expansion order and Smolyak grid accuracy level as in the computation of the impedance matrix for the TE$_z$ case.

Using expressions (10) and (11), the mean and the variance of the magnitude of the z-component of the scattered electric field were computed. Depicted in Fig. 6 is the mean value of the magnitude of the scattered electric field recorded on a circle of radius 3.4 m with its center at the center of the reference square of the four-cylinder array configuration. Also depicted in the figure are error bars indicating $\pm 3\sigma$ deviation from the mean. The large deviation obtained in the forward scattering direction (on the side of the array where the line source is placed) is attributed to the close proximity of the source to the cylinders. Also depicted in Fig. 7 is the mean value of the magnitude of the total electric field for a distance in the far-field region. Again, the error bars represent $\pm 3\sigma$ deviation from the mean.
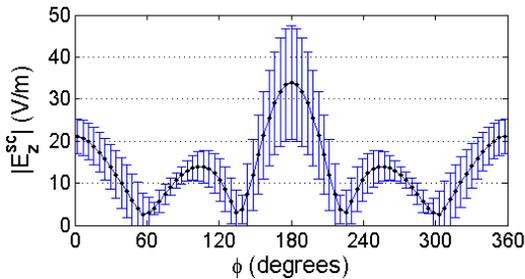
Fig. 6. Magnitude of the scattered electric field on a circle of radius 3.4 m. Error bars represent $\pm 3\sigma$ deviation from the mean.
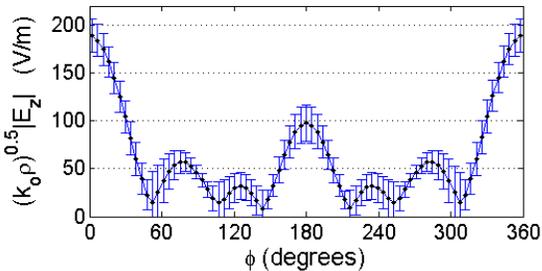
Fig. 7. Mean value of the magnitude of the far electric field at $\rho = 1000/k_o$. Error bars represent $\pm 3\sigma$ deviation from the mean.

IV. CONCLUSION

In conclusion, a methodology was proposed and numerically demonstrated for the development of stochastic macro-models of sub-domains of a complex electromagnetic structure exhibiting geometric and/or material randomness. Under the assumption that the randomness inside the sub-domain is parameterized in terms of a set of independent random variables, the proposed methodology abstracts the randomness in the sub-domain in terms of a stochastic global impedance or admittance matrix defined on a fixed surface enclosing the sub-domain. The elements of the matrix are given in terms of truncated polynomial chaos expansions on the random space defined by the independent random variables.

As demonstrated through the numerical examples presented, use of such stochastic macro-models alleviates the computational complexity of the solution of the random scattering problem by eliminating the need for the repeated numerical discretization (e.g., the repeated mesh generation) for the entire structure for each sampling point in the Monte Carlo process.

While the proposed methodology was presented in the context of two-dimensional EM scattering, its extension to three dimensions is rather straightforward. For example, for the case of an ensemble of multiple three-dimensional objects, global stochastic impedance boundary conditions can be defined on spherical surfaces enclosing each object. This extension is currently under development.

Another extension of the proposed macro-modeling involves the case where the elements of the stochastic global impedance matrix are functions of frequency. As already demonstrated in [10, 11], this extension generalizes the concept of stochastic global impedance macro-modeling – in a manner consistent with the concept of network matrix representation of passive EM multi-ports – to provide for a broadband stochastic macro-model of a portion of a composite structure exhibiting material and/or geometric randomness. Results from on-going research on these topics will appear in forthcoming papers.

ACKNOWLEDGEMENT

REFERENCES
[1] E. K. Miller, "Using Adaptive Estimation to Minimize the Number of Samples

Needed to Develop a Pattern to a Specified Uncertainty," *Applied Comput. Electromag. Soc. Journal*, vol. 17, no. 3, pp. 176-186, Nov. 2002.

[2] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, 1991.

[3] D. Xiu and J. S. Hesthaven, "Higher Order Collocation Methods for Differential Equations with Random Inputs," *SIAM Journal of Scientific Computing*, vol. 27, pp. 1118-1139, 2005.

[4] D. Xiu, "Fast Numerical Methods for Stochastic Computations: A Review," *Commun. Comput. Phys.*, vol. 5, no. 2-4, pp. 242-272, Feb. 2009.

[5] D. Xiu and J. S. Hesthaven, "Higher Order Collocation Methods for Differential Equations with Random Inputs," *SIAM Journal on Scientific Computing*, vol. 27, pp. 1118-1139, 2005.

[6] D. Xiu and G. Karniadakis, "The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal on Scientific Computing*, vol. 24, no. 2, 2002.

[7] C. Chauviere, J. S. Hesthaven, and L. C. Wilcox, "Efficient Computation of RCS from Scatterers of Uncertain Shapes," *IEEE Trans. Antennas Propagat.*, vol. 55, no. 5, pp. 1437- 1448, May 2007.

[8] H. Bagci, A. C. Yucel, J. S. Hesthaven, and E. Michielssen, "A Fast Stroud-Based Collocation Method for Statistically Characterizing EMI/EMC Phenomena on Complex Platforms," *IEEE Trans. Electromagn. Compat.*, vol. 51, pp. 301-311, 2009.

[9] A. C. Yucel, H. Bagci, and E. Michielssen, "Efficient Stochastic EMC/EMI Analysis using HDMR-Generated Surrogate Models," *Proc. XXX URSI General Assembly and Scientific Symposium of International Union of Radio Science,* Aug. 2011.

[10] P. Sumant, *Stochastic Multiphysics modeling of RF MEMS switches*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, 2010.

[11] P. Sumant, H. Wu, A. Cangellaris, and N. Aluru, "Reduced-Order Models of Finite Element Approximations of Electromagnetic Devices Exhibiting Statistical Variability," *IEEE Trans. Antennas Propagat.*, in press.

[12] V. Barthelmann, E. Novak, and K. Ritter, "High-Dimensional Integration of Smooth Functions Over Cubes," *Numerische Mathematik*, vol. 75, pp. 79-97, 1996.

[13] F. Heiss and V. Winschel, "Likelihood Approximation by Numerical Integration on Sparse Grids," *Journal of Econometrics*, vol. 144, no. 1, pp. 62-80, May 2008.

[14] K. Petras, "Smolyak Cubature of Given Polynomial Degree with Few Nodes for Increasing Dimension," *Numerische Mathematik,* vol. 93, pp. 729-753, 2003.

Juan S. Ochoa received his B.S. degrees (*summa cum laude*) in Physics and Electrical Engineering in 2008 from the University San Francisco, at Quito, Ecuador. In 2008, he joined the University of Illinois at Urbana-Champaign, where he got his M.Sc. in Electrical Engineering in 2010. He recently completed an internship at Intel as a signal integrity engineer in the summer of 2011.

Juan is currently working towards his Ph.D. in Electrical Engineering at the University of Illinois and his research interests include computational electromagnetics, stochastic MOR, sparse grid methods, metamaterials and electromagnetic scattering.

Andreas C. Cangellaris is the M. E. Van Valkenburg Professor of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign. He received his Diploma in Electrical Engineering (1981) from the Aristotle University of Thessaloniki, Greece, and the M.S. (1983), and Ph.D. (1985) degrees in Electrical Engineering from the University of California, Berkeley. Prior to joining the University of Illinois, Professor Cangellaris was on the faculty of the Department of Electrical and Computer Engineering at the University of Arizona from 1987-1997.

His research interests are in the area of applied and computational electromagnetics. He is a Fellow of IEEE, the recipient of the Alexander von Humboldt Research Award in 2005, and the recipient of the 2012 IEEE MTT-S Distinguished Educator Award.

Professor Cangellaris is the co-founder of the IEEE Topical Meeting on Electrical Performance of Electronic Packaging. He serves currently as editor of the IEEE Press Series on Electromagnetic Wave Theory.

Finite Difference Time Domain Diakoptic Strategies

Oren Harpaz and Raphael Kastner

School of Electrical Engineering
Tel Aviv University, Tel Aviv 69987, Israel
kast@eng.tau.ac.il

Abstract — In many applications, it may be advisable to "tear apart" the computational domain into several sub-domains separated by "seams," each one treated separately. The sub-domains are then sewn back together at appropriate stages of the computation. Three main diakoptic strategies have been developed in the recent past. Out of these, the diakoptics on-the-fly strategy can serve the purpose of parallelizing a FDTD process over several processors, each of which being responsible for the treatment of a certain sub-domain, as presented below.

Keywords: FDTD, diakoptics, domain decomposition, absorbing boundary conditions.

I. INTRODUCTION

Many electromagnetic scattering problems are best solved by a *diakoptic* approach; that is, by tearing them into smaller sub-problems, exciting each sub-problem with a variety of different excitations, and combining the results. The approach is facilitated by situating each sub-problem in its own sub-domain, and characterizing the "seams" between domains as boundary conditions (BCs). The diakoptic approach is most useful in two very different types of problems: problems involving tightly packed inhomogeneous sub-domains with strong interactions between them, and problems involving sub-domains separated by oceans of "white space."

Diakoptic schemes already have been successfully incorporated into frequency domain methods, e.g., [1–3] (the fast multiple method (FMM), when applied to large dense matrices [4], may also be considered as a frequency domain diakoptic method). They also hold great promise in the time domain (TD), where any sub-domain or a given group of sub-domains can be characterized by a grid-compatible Green's function over their bounding surface, as a basis for repeated computations of changes in the complementary sub-domains, e.g., for layered media [5]. In this context, a strategy presented in Section II has been recently proposed. Two more strategies, namely FDTD coupled with stabilized TD integral equations and diakoptics-on-the-fly, are presented Sections III and IV, respectively. Conclusions are drawn in Section V.

II. GREEN'S FUNCTION TYPE DIAKOPTICS

This strategy is useful in an iterative design process. The computational domain is decomposed into a fixed "basic scatterer" and a reduced computational domain to be optimized in the course of the design process (see Fig. 1). A Green's function is used to connect the sub-domains via a surface representation over the interface. A salient characteristic of this strategy is the usage of differential equation approach to generate grid-compatible time domain Green's functions that serve to represent sub-domains over the interface.

Subsets of this strategy are as follows.

A. The green's function method (GFM)

The GFM [6,7] (see Fig. 2(a)) leads to a field, or spatial domain procedure, analogous to cascaded
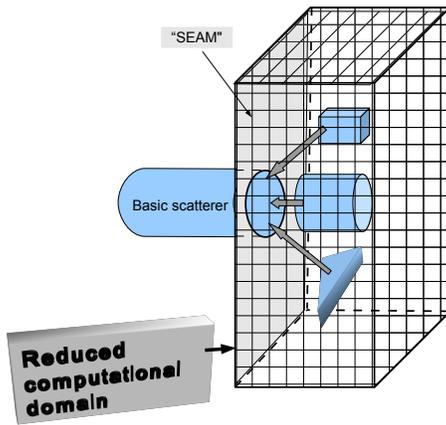
Fig. 1. A basic scatterer with different attachment, suitable for usage in a diakoptic scheme with a grid-based Green's function (see Section II-A).

impedance matrix representation in microwave circuit theory.

B. Spectral representation

This decomposition (Fig. 2(b)) is analogous to cascaded scattering matrices. Here the field is resolved into outgoing and incoming harmonics across the seam. An issue with this formulation is the restriction to convex and separable seams that limits the control of white space.

C. Source decomposition method (SDM)

The SDM [8] (Fig. 2(c)) combines advantages of the other two strategies. In one dimension, it coincides with the spectral approach, however in two or three dimensions it avoids the need for using separable geometries.

III. FDTD COUPLED WITH STABILIZED TD INTEGRAL EQUATIONS

This work is combined with the stabilizing method of D. Weile [9], who identifed the source of instabilities in the manner of the spatial integration employed, in particular, in the way numerical integration rules approximated the integrals of

functions that suddenly vanished on the boundary between illuminated and shadow regions. Stabilization is achieved through an approximated $\mathcal{Z}$-domain formulation that maps the left half of the s-plane into the unit circle in the z-plane. This recursive marching-on-in-time scheme can be merged with the FDTD, providing a stable integral equation interface at the seam.
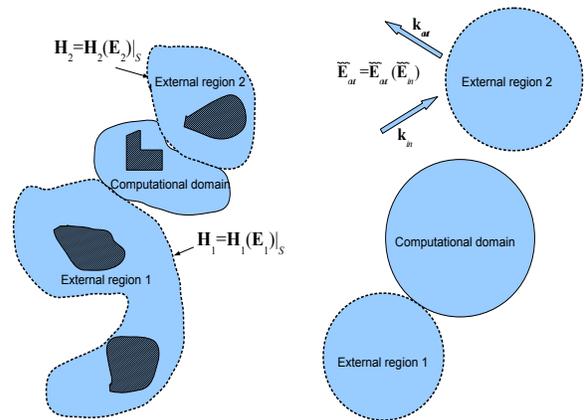
(a) GFM   (b) Spectral representation
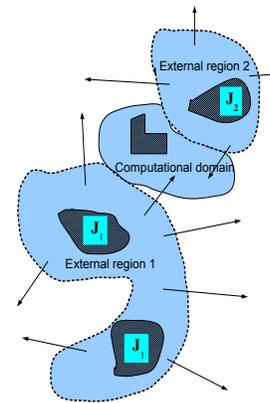
(c) Source decomposition

Fig. 2. Three grid-compatible Green's function approaches (see Introduction).

IV. DIAKOPTICS ON THE FLY

This strategy is motivated by parallel FDTD analysis. In this strategy, we tear apart and recombine
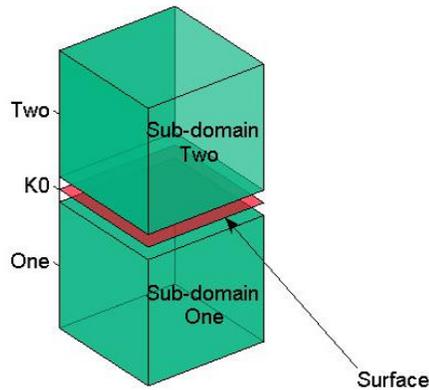
Fig. 3. Diakoptics-on-the-fly: two sub-domains with a planar seam.
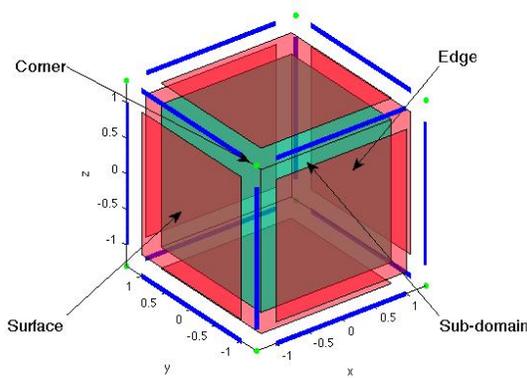
Fig. 4. The making of a seam: surfaces, edges, corners.

the sub-domains at each time step or after each small group of time steps (see Sections IV-A and IV-B, respectively). This strategy opens the door for optimization of the use of different clusters of processors, both CPUs and GPUs. The individual groups can be treated in parallel by separate processors, or sequentially by the same processor.

A. Update process: sewing at each time step

Each FDTD update sequence is divided into the following two phases: In the first phase, that can
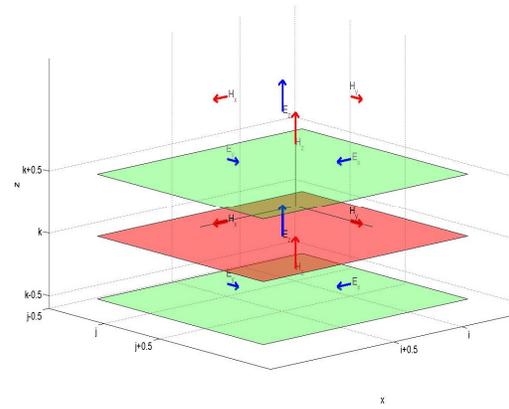
(a) Surface.
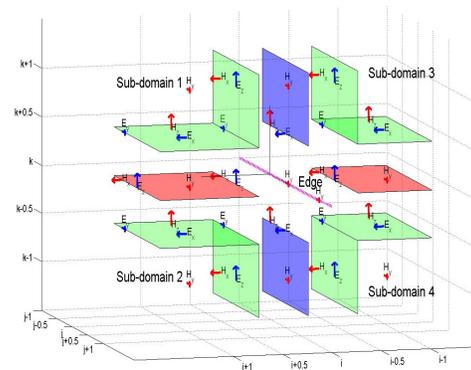
(b) Edge.
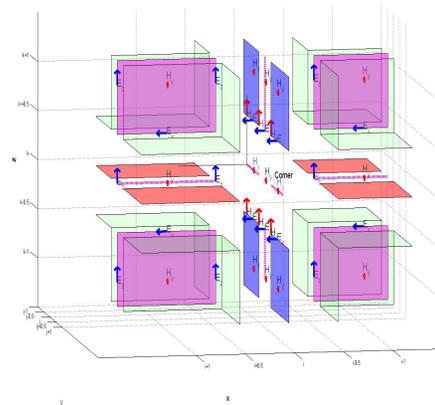
(c) Corner.

Fig. 5. The surface, edges and corners interact with two, four and eight neighboring sub-domains, respectively.

be run concurrently or sequentially for all sub-domains, the field at the the seam is considered known, acting as a boundary condition for the update of the field within each sub-domain. The second phase is the update at the seam using the known fields at the sub-domain boundaries. This second phase also includes separate updates for the edges and corners of the seam, involving four and eight neighboring sub-domains, respectively.

The seam is an appropriate Yee grid plane. If, e.g., the sub-domains "1" and "2" are aligned along the x axis (see Fig. 3), and the seam is at an E_z-field surface at $i + \frac{1}{2}$, then adjacent **H**-field surfaces are the bounding surfaces of sub-domains 1 and 2 at i and $i+1$, respectively. Divide now the FDTD update sequence into the following two phases: separate updates for sub-domains 1 and 2, followed by a "sewing" update phase.

1) The first phase: updates within the sub-domains: This phase can be run concurrently or sequentially for the two sub-domains. The field at the seam is considered a BC for the update of the **H**-field at the surfaces for i, $i + 1$:

$$\left. \begin{aligned} \frac{\gamma_x}{\eta} E^n_{z(i+\frac{1}{2},j+\frac{1}{2},k)} &\Rightarrow H^{n+\frac{1}{2}}_{y(i,j+\frac{1}{2},k)} \\ \frac{\gamma_x}{\eta} E^n_{y(i+\frac{1}{2},j,k+\frac{1}{2})} &\Rightarrow H^{n+\frac{1}{2}}_{z(i,j,k+\frac{1}{2})} \end{aligned} \right\} \text{sub-domain 1,}$$

$$\left. \begin{aligned} \frac{\gamma_x}{\eta} E^n_{z(i+\frac{1}{2},j+\frac{1}{2},k)} &\Rightarrow H^{n+\frac{1}{2}}_{y(i+1,j+\frac{1}{2},k)} \\ \frac{\gamma_x}{\eta} E^n_{y(i+\frac{1}{2},j,k+\frac{1}{2})} &\Rightarrow H^{n+\frac{1}{2}}_{z(i+1,j,k+\frac{1}{2})} \end{aligned} \right\} \text{sub-domain 2.}$$

$$(1)$$

2) The second phase: Update at the Seam: Update the field at the seam using the known **H**-fields at i and $i + 1$:

$$\eta\gamma_x \left(H^{n+\frac{1}{2}}_{z(i,j,k+\frac{1}{2})} - H^{n+\frac{1}{2}}_{z(i+1,j,k+\frac{1}{2})} \right)$$
$$\Rightarrow E^{n+1}_{y(i+\frac{1}{2},j,k+\frac{1}{2})}$$
$$\eta\gamma_x \left(H^{n+\frac{1}{2}}_{y(i+1,j+\frac{1}{2},k)} - H^{n+\frac{1}{2}}_{y(i,j+\frac{1}{2},k)} \right)$$
$$\Rightarrow E^{n+1}_{z(i+\frac{1}{2},j+\frac{1}{2},k)}.$$

$$(2)$$

This phase also includes separate updates for the edges and corners of the seam, see Fig. 4. These

are shown in Fig. 4. The update process for the surfaces, edges and corners that involve interaction with two, four and eight neighboring sub-domain, as can be seen in Figs. 5(a)–5(c), respectively.

B. Update process - sewing after several time steps

One disadvantage of the procedure in Section IV-A is the need to halt the volume calculations after each time step and allocate "overhead" time for data transfer into the separate processes of seam updates and back. This problem can be avoided, albeit at the cost of increasing the computational size at each sub-domain, by performing seam updates after each group of N time steps, as follows.

Consider the two sub-domains in Fig. 6(a). Extend each sub-domain into the region of its neighbor, as depicted by the blue boxes in Fig. 6(b). This box protrudes into the adjacent sub-domain, as seen in the red boxes in Fig. 6(c). Fill this extension with zeros, and compute the field of each sub-domain separately over N time steps, where N is chosen such that the signal would not traverse more then the length of the extension. While the main body of the sub-domain (green boxes) includes the full solution at the Nth time step, the protrusions (red and blue boxes) have a partial solution for the given sub-domain . A superposition of the partial solutions from adjacent sub-domain then produces the full solution. The entire process is then repeated for the next N time steps.

V. CONCLUSIONS

The diakoptic approach has the potential to reduce computer requirements for a large class of problems. An advantage is seen in situations including the following:

(a) Splitting computational effort between processes by using diakoptics-on-the-fly;

(b) Design processes, where the computational domain undergoes several modifications while the external boundary remains unchanged, as depicted in Fig. 3;

(a)

(b)

(c)
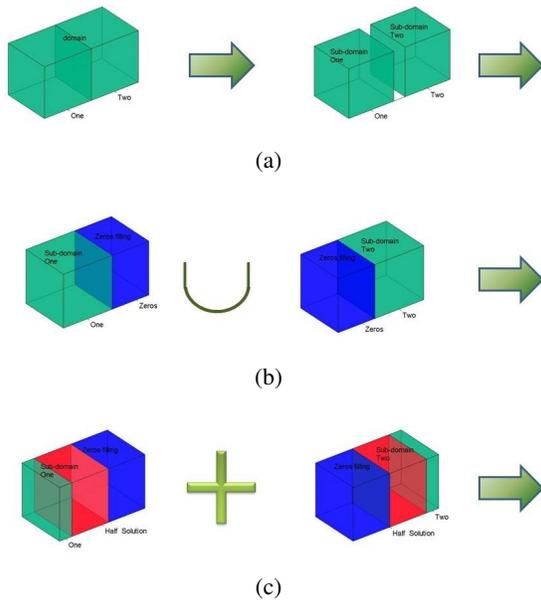
Fig. 6. Parallel FDTD computation with update after N time steps. (a) Two adjacent sub-domains, analyzed in parallel, (b) Zero padded extensions (blue boxes) of each sub-domain. The extensions protrude into the adjacent sub-domain, (c) Superposition of the overlapping solutions over each sub-domain, that includes the green and red boxes.

(c) White space elimination, in particular with non-convex scatterers. Typical absorbing boundary conditions (ABCs) require sometimes excessively large "white space" between the scatterer and the ABC boundary to ensure accuracy because standard ABCs must be applied on convex boundaries. ABCs that track the shape of the scatterer can minimize the size of the white space and allow for the inclusion of reflective external domains;

(d) Multiple-scatterer scenarios for cases where the scatterers are distinct and separated by a substantial "white space." The proposed scheme will allow the sub-problems to interact analytically over homogeneous domains;

(e) Problems best analyzed using sub-problems of different grid sizes and are solvable with methods other than FDTD, or involve moving objects.

Challenges, however, involve

(a) efficiency, since direct computations of Green's functions sometimes involve series with huge terms of alternating signs and

(b) instabilities that may occur when combining methods for boundary conditions.

REFERENCES

[1] Z. G. Qian, W. C. Chew, and R. Suaya, "Generalized Impedance Boundary Condition for Conductor Modeling in Surface Integral Equation," *IEEE Trans. Microwave Theory Tech.*, vol. 55, no. 11, pp. 2354–2364, Nov. 2007.

[2] V. K. Kollia and A. C. Cangellaris, "A Domain Decomposition Approach for Efficient Electromagnetic Analysis of the Power Distribution Network of Packaged Electronics Systems," *IEEE Trans. Electromagnetic Comp.*, vol. 52, no. 2, pp. 320—331, May 2010.

[3] A. Boag, U. Shemer, and R. Kastner, "Hybrid Absorbing Boundary Conditions Based on Fast Non-Uniform Grid Integration for Non-Convex Scatterers," *Microwave Opt. Tech. Letters*, vol. 43, 2, pp. 102–106, Oct. 20, 2004.

[4] B Carpentieri, "An Adaptive Approximate Inverse-Based Preconditioner Combined with the Fast Multipole Method for Solving Dense Linear Systems in Electromagnetic Scattering," *Applied Computational Electromagnetic Society Journal*, vol. 24, no. 5, pp. 504-510, October 2009.

[5] T. Vaupel, "A Fast Spectral Domain Solver for the Characterization of Larger Microwave Structures in Multilayered Environments," *Applied Computational Electromagnetic Society Journal*, vol. 24, no. 5, pp. 493-503, Oct. 2009.

[6] R. Holtzman and R. Kastner, "The Time–Domain Discrete Green's Function Method (GFM) Characterizing the FDTD Grid Boundary," *IEEE Trans. on Antennas and Propagation*, vol. 49, no. 7, pp. 1079–1093, July 2001.

[7] R. Holtzman, R. Kastner, E. Heyman and R. W. Ziolkowski, "Stability Analysis of the Green's Function Method (GFM) Used as an ABC for Arbitrarily-Shaped Boundaries," *IEEE Trans. on Antennas and Propagation*, vol. 50, pp. 1017–1029, July 2002.

[8] S. Malevsky, E. Heyman, and R. Kastner, "Source Decomposition as a Diakoptic Boundary Condition in FDTD with Reflecting External Regions," *IEEE Trans. Antennas Propagat.*, vol. 58, no. 11, pp. 3602–3609, Nov. 2010.

[9] X. Wang, R. A. Wildman, D. S. Weile, and P. Monk, "A Finite Difference Delay Modeling Approach to the Discretization of the Time Domain Integral Equations of Electromagnetics," *IEEE Trans. Antennas Propagat.*, Special Issue on Large and Multiscale Computational Electromagnetics, vol. 56, no. 8, Part 1, pp. 2442–2452. Aug. 2008.

Oren Harpaz received the B.Tech degree in Electrical Engineering from the Ariel University Center of Samaria in 2002 and the M.Sc degree from Tel Aviv University, Department of Physical Electronics, in 2005. His research interests include numerical techniques in electromagnetics and antennas.

Raphael Kastner was born in Haifa, Israel, in 1948. He received the B.Sc. (summa cum laude) and the M.Sc. degrees in electrical engineering, engineering from the Technion, Israel Institute of Technology in 1973 and 1976, respectively, and his Ph.D. degree from the University of Illinois, Urbana, in 1982. From 1976 to 1988 he was with RAFAEL, Israel Armament Development Authority, where from 1982 to 1986 he headed the antenna section. He was a Visiting Assistant Professor at Syracuse University from 1986 to 1987, and a Visiting Scholar at the University of Illinois in 1987 and 1989. Since 1988 he has been with School of Electrical Engineering, Tel Aviv University, where is now a Professor. In 2000 he co-founded XellAnt Inc. and acted as its CEO until 2004. He is an IEEE fellow, a recipient of the IEEE Third Millenium medal and several excellence in teaching awards, and is a member of Tau Beta Pi and Eta Kappa Nu. His research interests are in computational electromagnetics and antennas.

An Electromagnetic Compatibility Problem via Unscented Transform and Stochastic Collocation Methods

S. Lalléchère [1,2], P. Bonnet [1,2], I. El Baba [1,2], and F. Paladian [1,2]

[1] Clermont University
Blaise Pascal University, LASMEA, Clermont-Ferrand, 63000, France

[2] CNRS
UMR6602, LASMEA, Aubière, 63177, France
sebastien.lallechere@lasmea.univ-bpclermont.fr, pierre.bonnet@lasmea.univ-bpclermont.fr,
ibrahim.elbaba@lasmea.univ-bpclermont.fr, francoise.paladian@lasmea.univ-bpclermont.fr

Abstract ─ This paper aims to illustrate the current interest about the use of stochastic techniques for electromagnetic compatibility (EMC) issues. This problem may be handled from various methods. First, we may focus on the Monte Carlo (MC) formalism but other techniques have been implemented more recently (the unscented transform, UT, or stochastic collocation, SC, for instance). This work deals with solving a stochastic EMC problem (transmission line) with the UT and SC techniques and to compare them with the reference MC results.

Index Terms ─ Sensibility analysis, stochastic collocation, stochastic electromagnetic compatibility, unscented transform.

I. INTRODUCTION

Nowadays, although one may note the growing interest of the electromagnetic (EM) community for measurement techniques and numerical codes with increasing accuracy, the trend is to improve their efficiency by optimizing them. Most of the computational works in electromagnetics remain deterministic (i.e., one single result per set of exact input data). Although the uncertainties are intrinsic in EM analysis (slight variations from large production, environmental factors, reproducibility drifts), very few studies are achieved to efficiently take them into account. One of the most spread techniques relies on Monte Carlo (MC) simulations [1]. The

aim of this work is to focus on two additional methods which present the simplicity of MC method with faster convergence rates. Thus, the unscented transform (UT) method [2, 3] and the stochastic collocation (SC) technique [4] will be detailed from their foundations to an electromagnetic compatibility (EMC) application.

II. THEORETICAL BASIS

Many methods are used to take uncertainties into account in EM simulations; one of the simplest ways to do so may remain as the use of MC developments. Some probabilistic techniques are also available: we can mention, without exhaustiveness, the polynomial chaos [5] or the kriging technique [6]. In this article, we focus on the interest of the UT [2] and SC [4] methods to solve EMC problems. They are two non-intrusive methods whose main advantages are simplicity and efficiency. Contrary to a classical MC simulation, well known for its slow convergence rate, the collocation methods are computationally very interesting since they need a limited number of well-chosen points related to the distribution of the random variables (RV). Thus, the UT and SC techniques share many similarities; they only differ regarding two or more random parameters. In our model, a stochastic parameter Z will be defined according to a RV $\hat{u}$ following

$$Z = Z^0 + \hat{u}, \qquad (1)$$

with Z^0 the initial value (mean) and without any loss of generality $\hat{u}$ follows a certain distribution

law (zero-mean and a given variance). The random value Z may stand for different parameters: material characteristics (dielectric), geometry (for instance size or location of a target) or the source parameters (magnitude, frequency ...).

A. UT principles

As explained in [3], the UT method is similar to the MC technique. The main difference relies on the number of realizations needed to obtain the statistical moments of a given output. Instead of several thousands of repetitions, only a few selected ones are necessary.

A single RV UT case. Some conditions are required to compute UT for a single RV: we may know both the moments of the RV $\hat{u}$ and the nonlinear mapping of the random output. Its nth order moment may be expressed as follows:

$$E\{I(\hat{u})^n\} = \int I(u)^n pdf(u)du, \qquad (2)$$

where $pdf(u)$ is the probability density function of the RV $\hat{u}$. A discrete equivalent of the relation (2) is used for the integration

$$\int I(u)^n pdf(u)du \approx \sum_i \omega_i I(S_i)^n, \qquad (3)$$

where S_i are the so-called sigma points (for the integration). If the nonlinear mapping $I(\hat{u})$ is well behaved, it could be expressed from Taylor polynomial series (g_j coefficients) as

$$\int I(u)^n pdf(u)du = \sum_{j=0}^{\infty} g_j \int u^j pdf(u)du. \quad (4)$$

From the discrete sum (3), each integration term of (4) may be expressed from $k+1$ ($k=0, 1 \dots$) equations as

$$\int u^k pdf(u)du \approx \sum_i \omega_i S_i^k = E(\hat{u}^k). \qquad (5)$$

The nonlinear system depicted in (5) allows the computation of the sigma points S_i and weights ω_i from the moments of the RV $\hat{u}$. As detailed in [2], the minimum number of S_i points for a given order of the UT technique may be derived using the Gauss quadrature schemes. Indeed, considering (5), the solution is not unique and different sets of (S_i, ω_i) may be obtained as illustrated in the following.

UT for multi-RV. Based upon the results obtained for a single RV, the Taylor polynomial representation is still suitable for two RV. In the two variables case, the system (5) may be written using statistical moments cross terms [2] following

$$\sum_i \omega_i (S_i^1)^m (S_i^2)^n = E\{\hat{u}_1^m \hat{u}_2^n\}. \qquad (6)$$

The sigma points and weights are computed for the two RV $\hat{u}_1$ and $\hat{u}_2$ and derived from the possible power combinations of m and n (natural numbers) with $0 \leq m + n \leq 4$. Once more there are several possible solutions: in the following we will give different sets of sigma points and weights solving the system (6).

B. SC foundations

This section is dedicated to the presentation of the SC technique [4].

SC basis for a single random parameter. The idea of the technique is to find a polynomial approximation of a given output I depending on a random parameter Z (1). In a first time, the function $S \rightarrow I(Z^0; S)$ is split up a Lagrangian basis with n the approximation order

$$I(Z^0; S) \approx \sum_{i=0}^{n} I_i(Z^0)L_i(S), \qquad (7)$$

with $L_i(S) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{S - S_j}{S_i - S_j}$. One of the most interesting properties of the Lagrangian basis relies on its reducing characteristic: $L_i(S_j) = \delta_{ij}$ (Kronecker δ) and we may write $I_i(Z^0) = I(Z^0; S_i)$. Then, the integration computation is based upon the Gauss quadrature with identical points S_i than the ones previously needed by the SC method

$$\int_D pdf(u) f(u)du \approx \sum_{i=0}^{n} \omega_i f(S_i). \qquad (8)$$

Similarly to the UT case, the real numbers ω_i are called integration weights. From (7), we may detail I with its polynomial approximation

$$E\{I(Z^0; S)\} = \sum_{i=0}^{n} I_i(Z^0) \int_D L_i(s)pdf(s)ds. \quad (9)$$

From (9), we may straightforward compute weights following

$$\omega_i = \int_D L_i(s)pdf(s)\, ds. \qquad (10)$$

We will detail in Section II-D the statistical moments computation enabled by the pair (S_i, ω_i).

Multi-RV SC case. The previous theoretical elements may be generalized to the multivariate case. Therefore, by considering a two-variable random problem, for instance involving two RV $\hat{u}_1$ and $\hat{u}_2$ standing for two random parameters, respectively Y and Z, we may write Y and Z from the relation (1) including two initial values (Y^0 and Z^0) and potentially two random distributions. From the same theoretical foundations, we may project the function $(s, t) \rightarrow I(Y^0, Z^0; s_i, t_j)$ on a Lagrangian basis

$$I(Y^0, Z^0; s_i, t_j) \approx \sum_{i=0}^{n} \sum_{j=0}^{n} I_{ij}(Y^0, Z^0)L_i(s)L_j(t),$$
$$(11)$$

with $I_{ij}(Y^0, Z^0) = I(Y^0, Z^0; s_i, t_j)$. It is rather simple from (11) to compute the moments of the output I through a tensor product in each direction (i.e., for each RV) based upon the generalization of (8). Comparatively to MC, the technique may appear limited when the number of RV increases. Other methods exist [7, 8] to ensure efficiency with a good level of accuracy but for few RV, UT, and SC reveal particularly precise.

C. Sigma points/weights computation

Although the UT and SC appear very similar considering the computation of their respective sigma points/weights, they differ from their basis. From the different solutions proposed in one-variable case, the minimum number of (S_i, ω_i) pairs (for a given order n) is straightforward available by the Gauss quadrature scheme (identical to the SC case [4, 9]). Therefore, the expression of the integration points/weights is similar for the UT [2] and SC [10], and the results will be identical. The Table 1 gives an overview of the points/weights in single RV case following a standard normal distribution. We may construct similarly the multi-RV set of points/weights for UT and SC. Based upon [2], it is possible to extend the previous set of points (Table 1) to the numerical examples presented in the following (one or two RV) including the distribution law variance. As depicted in [2], the UT solution is not unique when solving the system (6) and we may obtain (for a same order) different sets of sigma/weights points.

Table 1: Sigma/weights points (one RV) for a standard normal law (UT/SC)

n		Pt1	Pt2	Pt3	Pt4	Pt5
2	S_i	$-\sqrt{3}$	0	$\sqrt{3}$		
	ω_i	1/6	2/3	1/6		
4	S_i	-2.9	-1.4	0	1.4	2.9
	ω_i	0.01	0.22	0.53	0.22	0.01

D. Calculation of the statistical moments

The collocation technique gives the collocation points (S_i) and weights (ω_i) necessary to entirely compute the nth-order ($n=1, 2 ...$) Z statistical moments [10]. From previous notations ($E\{Z\}$ for instance), in order to simplify the discussion, the brackets $\langle Z \rangle$ and σ_Z symbol will

stand respectively for the mean and standard deviation of the Z output.

Table 2: Statistical moments computation with the UT/SC for one-dimensional RV case

Moment	n	Computation
Mean	1	$\langle Z \rangle = \sum_{i=0}^{n} \omega_i S_i$
Variance	2	$\sigma_Z^2 = \sum_{i=0}^{n} \omega_i S_i^2 - \langle Z \rangle^2$

The Table 2 gives the computation of the first and second statistical moments from UT/SC (one RV). The computation of the sigma points and weights (multi-RV) may be found respectively in [2, 9]. In the following, we will present the numerical differences existing between the UT and SC.

III. A STOCHASTIC EMC PROBLEM

The case of a simple transmission line of diameter d, at frequency f, length L, placed at a height h above an infinite ground plane and illuminated by a uniform linearly polarized plane wave is considered (Fig. 1). An analytical formulation can be obtained for the current $I(L)$ at load $Z_L = 1k\Omega$ [6], Z_0 is set to 50Ω.
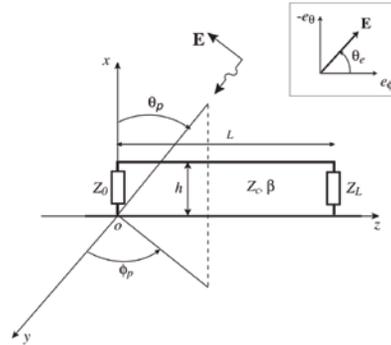
Fig. 1. Transmission line striken by a plane wave.

A. Accuracy of the UT and SC methods (single RV case)

Both for the UT and SC, the accuracy of the stochastic techniques presented in this paper depends on the order n of the approximation as presented in (8). Thus, in the following, we will talk about UT and SC "order n" which means a polynomial approximation with $n+1$ pairs (S_i, ω_i).

Using the Gauss quadrature to compute weights and sigma points leads to a similar (S_i, ω_i) set both for the UT and SC. The single variable uncertainty is first introduced considering L. The line length is associated to a RV $\hat{u}$ following a normal distribution (zero-mean, variance=0.053); different L^0 mean values are regarded in the set D_L=[1.2m; 4.5m]. In this example, the values of parameters are set following: d=1mm, f=50MHz, h=20mm, with initial electric field amplitude E_0=1kV/m, a normal striking is considered here.

Statistical moments from stochastic study. The Fig. 2 clearly shows the convergence of the stochastic techniques (UT/SC). Indeed, from the 3/5/7-points discretization (i.e., respectively n=2/4/6), the $\langle I \rangle$ curves almost overlap. For n=2, the different (S_i; ω_i) points may be summarized following (-0.4; 1/6), (0; 2/3) and (0.4; 1/6). As expected from Table 2, the 2nd order statistical moment is also computed from UT/SC. The current I standard deviation (σ_I) also appears through the numerical dispersion around the mean value $\langle I \rangle$. The 2nd order results (σ_I) agree well regarding the UT/SC 5 and 7-points accuracy and check the 2nd order convergence. Obviously, it may have been expected that the 2nd order convergence requires a higher accuracy from stochastic treatments. Moreover, one may put the focus on the interest of these stochastic formalisms. A faster approach may have been to only consider the mean value (central value $L^0 \in D_L$): it is given on Fig. 3 by the pink curve. It may be noticed that the central data gives a trend of the uncertainty impact but does not fit well to the converged stochastic (UT/SC) behavior. The differences appearing (for a relatively weak randomness uncertainty) between a single deterministic simulation (mean value) and UT/SC results ($\langle I \rangle \pm \sigma_i$ overruns) confirm the importance of the stochastic modeling.

Figure 3 gives an overview of the convergence of the UT/SC methods comparatively with a MC reference. In order to compute the relative error due to stochastic computing, we consider a converged set of MC data (here with 100,000 realizations). Then, a criterion is defined standing for this relative error

$$err_i = 100 \times \frac{\left| z_i^{UT/SC} - z_i^{MC} \right|}{\left| z_i^{MC} \right|}, \qquad (12)$$

where i corresponds to the considered random length L_i ($L_i \in D_L$) and $z^{UT/SC}$, z^{MC} stand for a given statistical moment, for instance mean or variance (standard deviation), of the random output (current I) obtained respectively from UT/SC and MC. The results from Fig. 3 show the convergence of the stochastic formulations since respectively less than 0.25% and 2% errors appear from the current mean and standard deviation computations.
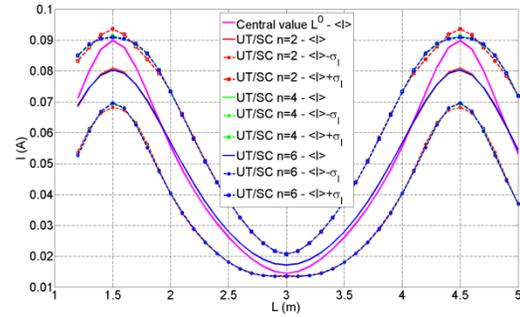
Fig. 2. $\langle I \rangle \pm \sigma_I$ from UT / SC convergence (n=2, 4, 6) for normally distributed randomness ($\sigma_{\hat{u}} = 0.231$).
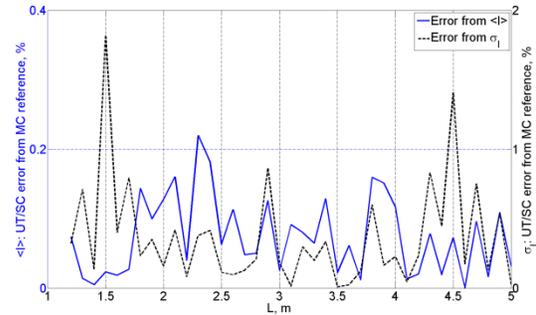
Fig. 3. Relative error between SC and 100,000 MC realizations (reference) with $\sigma_{\hat{u}} = 0.231$.

High-order statistical moments. As detailed in Table 2, both the UT and SC techniques allow computing straightforward high-order moments. The Fig. 4 shows the kurtosis convergence of the UT/SC techniques comparing to MC data. Therefore, although the results from 5-points UT/SC are in a good agreement with reference (MC), we may need an accuracy level from 7-points (n=6) UT/SC to fit with reference. The same kind of conclusion could be obtained from

the skewness (3rd order) computation. Of course, for higher statistical moments needs, the UT/SC will require higher precision.
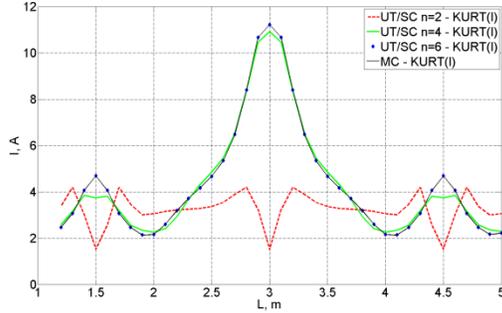
Fig. 4. Stochastic convergence from UT/SC and 100,000 MC realizations ($\sigma_{\hat{u}} = 0.231$).

B. Stochastic EMC expectations

The aim of this part is to illustrate the ability of the SC method to handle various randomness patterns (distribution, intensity ...) needed for EMC problems.

Variations around statistical distributions. In this case, we focus on two additional random distributions: uniform and log-normal ones (Figs. 5 and 6). The mean and variance of the example from Section III-A remain unchanged. For instance, comparatively to previous case, the pairs $(S_i; \omega_i)$ are (-0.31; 0.28), (0; 0.44) and (0.31; 0.28) for a uniform law and n=2. The example of the log-normal law needs $(S_i; \omega_i)$ following (-0.39; 1/6), (-0.01; 2/3) and (0.41; 1/6).
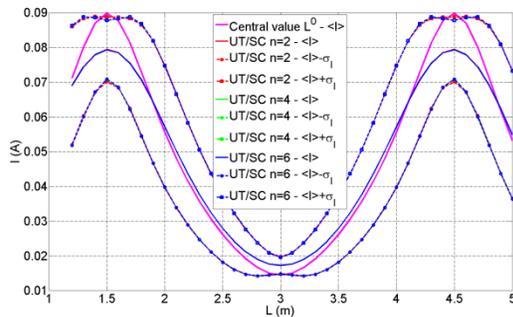
Fig. 5. Convergence (n=2, 4, 6) with a uniform random distribution, $\langle \hat{u} \rangle = 0$ and $\sigma_{\hat{u}} = 0.231$.

The SC convergence is obtained from 5-points (n=4). Even if some slight differences

appear both on the levels and the convergence rate, the SC method allows modeling different kind of uncertainties. The randomness distributed from a log-normal law seems to require a different SC order to converge (Fig. 6) comparatively to normal or uniform distributions.
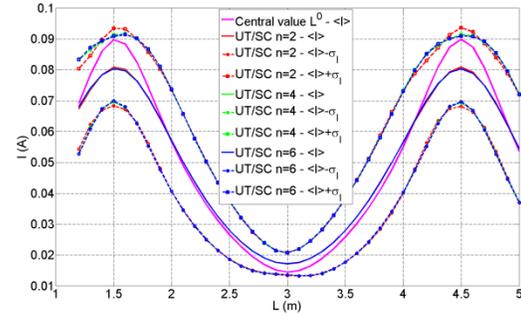
Fig. 6. Convergence (n=2, 4, 6) with a log-normal random law, $\langle \hat{u} \rangle = 0$ and $\sigma_{\hat{u}} = 0.231$.

Randomness intensity and convergence. Obviously, increasing the level of uncertainty of the studied random parameter will lead the analysis close to a "working" threshold. For a given problem, the SC method will operate well until a certain boundary. This may be easily understood since we are talking about stochastic and random parameters far from a complete parametric study. After all, we may wonder if the SC method is robust. In the following, we consider that $\hat{u}$ follows a uniform law U[-1;1] ($\langle \hat{u} \rangle = 0$ and $\sigma_{\hat{u}} = 0.577$). For n=2, the different $(S_i; \omega_i)$ points may be summarized following (-0.78; 0.28), (0; 0.44) and (0.78; 0.28). In comparison with Fig. 5, the results from Fig. 7 show the convergence of the SC technique. Obviously, it is more difficult since the magnitude of variations is huge ($\pm$ 1m over the line length for each initial value L^0). As expected, the central data (Fig. 7, pink curve) does not fit at all with the SC mean computation. Even if the variations are great due to randomness, the Fig. 8 shows the well accuracy of SC results in comparison with 100,000 MC realizations. Indeed, $\langle I \rangle$ and σ_I are close to reference data: the SC relative error from (12) is respectively lower than 0.45% and 1.5% for the current mean and standard deviation (Fig. 8). The lengths L_1=1.5m and L_2=4.5m are particular points (resonances). As depicted in Fig. 3 and Fig. 8, the highest error

levels are obtained for L_1 and L_2 (both around 1.5%) according to the current standard deviation.
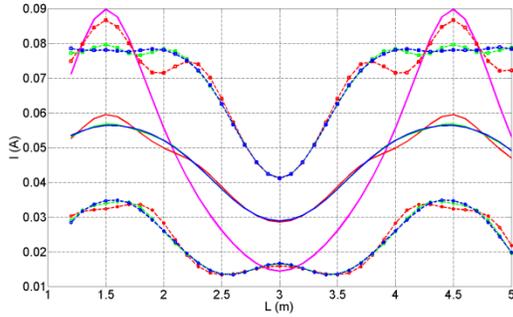
Fig. 7. Convergence (n=2, 4, 6) with a uniform random distribution ($\sigma_{\hat{u}} = 0.577$, legend Fig. 5).
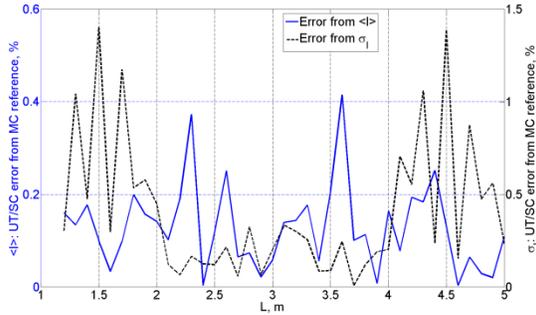
Fig. 8. Relative error between SC and 100,000 MC realizations for $\langle I \rangle$ and σ_I (randomness following a uniform distribution U[-1;1]).

C. A multivariate random problem

Since the Taylor polynomial expansion is still usable for two RV, the UT may be used to achieve multivariate stochastic problem (see Section II-A). Moreover, as illustrated in [10], the single-variable SC technique may be generalized to multi-RV problems. Based upon their distinct foundations, for the multivariate case, the different two RV (S_i, ω_i) sets jointly with the different moments computation involve variations around the numerical results. In this section, we will add the source frequency f to the line length L to achieve a stochastic treatment of the EMC problem (Fig. 1). Both L and f will be given by two independent RV ($\hat{u}_1$ and $\hat{u}_2$) following a normal distribution (zero-mean). The variances are respectively given by $\sigma_{\hat{u}_1}^2 = 2.083.10^{-2}$ and $\sigma_{\hat{u}_2}^2 = 2.083.10^{10}$. Based upon the relation (1), we may write

$$L = L^0 + \hat{u}_1$$
$$f = f^0 + \hat{u}_2, \tag{13}$$

where f^0 and L^0 stand respectively for the f and L initial values. As depicted in Fig. 9, the current variance $var(I)$ is calculated in a straightforward manner for a large set of points. Thus, each element of the set $\{L_i^0 \epsilon D_L = [1.2; 4.5]\text{m}; f_i^0 \epsilon D_f = [1; 35]MHz\}$ is subjected to the previous random variations. The SC convergence and sensibility of the model appears in Fig. 9. The results depicted show the convergence of the SC method (from the I variance). Considering the SC accuracy for 3^2 and 5^2 points (respectively SC3 and SC5), the two data sets almost overlap. Convergence is obtained considering the current mean and the SC technique approximates well the random behaviour of the system.
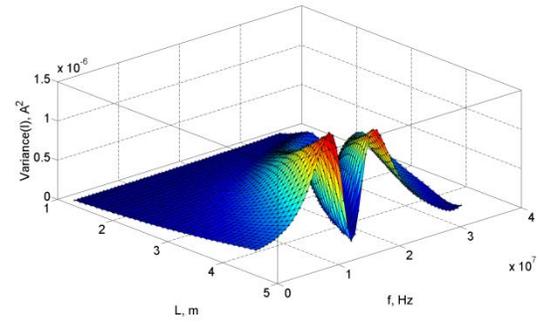
Fig. 9. Variance(I) from SC3 and SC5 (respectively colored slice and black asterisks).
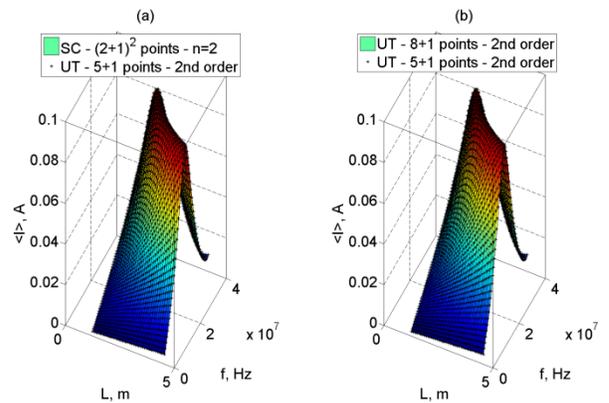
Fig. 10. Mean(I) from SC3, UT5 and UT8.

In Fig. 10, the results from UT fit very well with the "converged" data from SC. Thus, Fig. 10-a shows the agreement for $\langle I \rangle$ between SC3 and

the UT approximation (2nd order) involving 6 (=5+1) points (UT5). A great agreement appears also from Fig. 10-b considering the slight differences existing between a same UT accuracy (2nd order) involving 5+1 or 8+1 points (respectively UT5 or UT8). The UT5 and UT8 differences rely on the non-uniqueness of the solution in (6). In order to properly define the accuracy of each stochastic formalism (UT/SC), one may refer to MC simulations. First, it is necessary to determine a reference set of $\langle I \rangle$ values: empirically, 100,000 MC realizations are necessary. Relying on the relation (12) obtained for a single RV, we may define for each pair (f_k, L_k) (k=1... N_p with N_p the total number of frequency/length points) a similar parameter in bidimensional RV case.
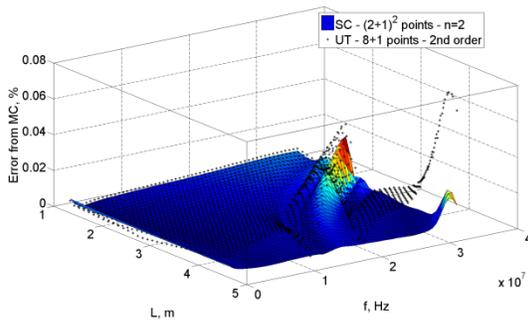
Fig. 11. Mean(I) error from MC data for SC3/UT8.

Figure 11 shows the gap between SC3 and UT8 accuracies. Even if the SC/UT numerical precision is high (less than 0.08% from MC reference), the precision remains widely better for SC (3 times) comparatively to UT. The time and memory saves appear clearly from previous examples since UT and SC need less than 10 realizations compared to the MC technique which requires about 100,000 simulations. Furthermore, it would be possible to improve the SC efficiency using techniques from [11] to reduce the number of SC realizations needed; it could be particularly interesting for multivariate stochastic problems involving many RV [12]. Another solution may be to reduce this number to a minimum regarding their relative influence. From [2], the comparison of results from 1-RV simulation with those involving a set of RV provides information on significant parameters.

D. A random sensibility analysis

This part illustrates the SC ability to achieve sensibility analysis in a random EMC problem. Among all the variables depicted on Fig. 1, we will focus on the parameters h and f. The other values are given by the previous example except the line length L which is set to 1.65m. An influence criterion is defined in [2] to characterize the sensibility of one RV. Based upon the SC results, a similar parameter is

$$In_{Z_k} = -\log\left(\left|1 - \frac{var(I(Z_k))}{var(I(Z_1, Z_2, ..., Z_n))}\right|\right), \qquad (14)$$

with $var(I(Z_k))$ and $var(I(Z_1, Z_2, ..., Z_n))$ the current I variances given respectively from one RV Z_k (k=1, ..., n) and n RV. The two random outputs are given considering two RV (RV1 $\rightarrow$ h and RV2 $\rightarrow$ f), respectively $\hat{u}_1$ and $\hat{u}_2$ both following a uniform distribution with $\sigma_{\hat{u}_1}^2 = 2.083.10^{-10}$ and $\sigma_{\hat{u}_2}^2 = 2.083.10^{10}$.
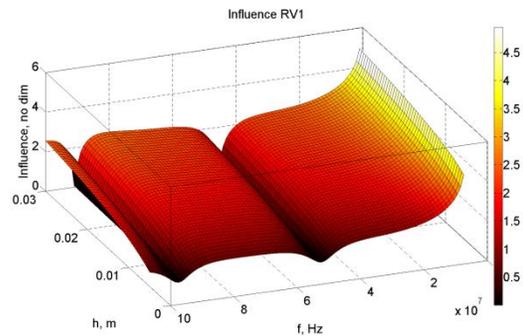
Fig. 12. Influence of RV1 (h).
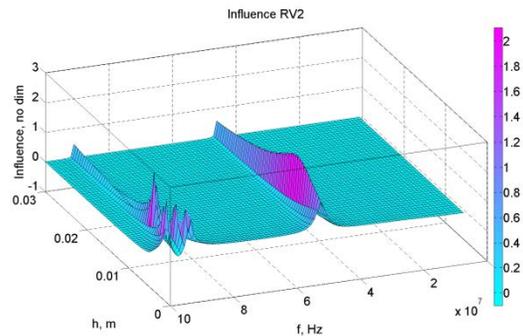
Fig. 13. Influence of RV2 (f).

The influence of each parameter is shown in Figs. 12 and 13. The computation of the influence from (14) lays emphasis on the dominant effect of height. As expected, the impact of the frequency is

relatively smooth outside resonance frequencies. The *f*-effect should not be neglected around them with influence levels *In>1dB*. Comparatively to *f*-influence, the Fig. 12 shows the global *h*-impact (1*dB<In<5dB* almost everywhere).

VI. CONCLUSION

In this paper, both UT and SC techniques to solve stochastic EMC problems are presented. Uncertainties involving source parameters (frequency) and geometry of a transmission line (length, height) have been defined considering various RV following uniform, normal and log-normal distributions. The UT and SC methods appear similar to well chosen MC simulations: their main advantages rely on their effectiveness (minimizing CPU time more than 20,000 times) and non-intrusive characteristic. Directly linked to computational electromagnetics, we may perfectly apply these methods considering other EM simulation/experimental tools.

REFERENCES

[1] N. Mishra and N. Gupta, "Quasi Monte Carlo Integration Technique for Method of Moments Solution of EFIE in Radiation Problems," *Applied Computational Electromagnetic Society (ACES) Journal*, vol. 24, no. 3, pp. 306-311, June 2009.

[2] L. De Menezes, A. Ajayi, C. Christopoulos, P. Sewell, and G. A. Borges, "Efficient Computation of Stochastic Electromagnetic Problems using Unscented Transforms," *IET Sci. Mea. Tec.*, vol. 2, no. 2, pp. 88-95, 2008.

[3] L. De Menezes, D. W. P. Thomas, C. Christopoulos, A. Ajayi, and P. Sewell, "The Use of Unscented Transforms for Statistical Analysis in EM," *Proc. EMC Europe*, 2008.

[4] C. Chauvière, J. S. Hestaven, and L. C. Wilcox, "Efficient Computation of RCS from Scatterers of Uncertain Shapes," *IEEE Trans. On Ant. And Prop.*, vol. 55, no. 5, pp. 1437-1448, May 2007.

[5] P. S. Sumant, H. Wu, A. C. Cangellaris, and N. R. Aluru, "A Sparse Grid Based Collocation Method for Model Order Reduction of Finite Element Approximations of Passive Electromagnetic Devices under Uncertainty," 2010 IEEE MTT-S International, Anaheim, CA, USA, pp. 1652-1655, May 2010.

[6] V. Rannou, F. Brouaye, M. Hélier, and W. Tabbara, "Kriging the Quantile: Application to a Simple Transmission Line Model," *Inverse Problems*, vol. 18, pp. 37-48, 2002.

[7] J. Foo and G.E. Karniadakis, "Multi-Element Probabilistic Collocation Method in High Dimensions," J. Comp. Phys., vol. 229, no. 5, pp. 1536-1577, 2010.

[8] A.C. Yucel, H. Bagci, and E. Michielsen, "An h-Adaptative Stochastic Collocation Method for Stochastic EMC/EMI Analysis," *Proc. IEEE APS*, Toronto, Canada, 2010.

[9] P. Bonnet, F. Diouf, C. Chauvière, S. Lalléchère, M. Fogli, and F. Paladian, "Numerical Simulation of a Reverberation Chamber with a Stochastic Collocation Method," *CRAS,* vol. 10, pp. 54-64, 2009.

[10] F. Paladian, P. Bonnet, and S. Lalléchère, "Modeling Complex Systems for EMC Applications by Considering Uncertainties," *XXX URSI General Assembly,* Istanbul, Turkey, August 2011.

[11] A. H. Stroud, "Remarks on the Disposition of Points in Numerical Integration Formulas," *Math. Tables Other Aids Comp.,* vol. 11, no. 3, pp. 1118-1139, October 1957.

[12] H. Bagci, A. C. Yucel, J. S. Hesthaven, and E. Michielssen, "A Fast Stroud-Based Collocation Method for Statistically Characterizing EMI/EMC Phenomena on Complex Platforms," *IEEE Trans. on EMC,* vol. 51, no. 2, pp. 301-311, 2009.

Sébastien Lalléchère received his Ph.D. degree (2006) in Electromagnetism. Since 2007, he is an assistant professor at Clermont University and is interested in numerical methods for electromagnetism.

Pierre Bonnet received his Ph.D. degree in Electronics in 1998. Since 1999, he works as an assistant professor at Clermont University on computational electromagnetics.

Ibrahim El Baba received the M.S. degree in 2008. He is currently working towards the Ph.D. degree in Electronics at Clermont University.

Françoise Paladian received her Ph.D. degree (1985) in Electronics from Clermont University and is a professor in the Physics department. Her research area concerns numerical EMC and reverberating chambers.

Combining the CORS and BiCORSTAB Iterative Methods with MLFMA and SAI Preconditioning for Solving Large Linear Systems in Electromagnetics

Bruno Carpentieri [1], Yan-Fei Jing [2], Ting-Zhu Huang [2], Wei-Chao Pi [3], and Xin-Qing Sheng [3]

[1]University of Groningen, Nijenborgh 9, PO Box 407, 9700 AK Groningen, Netherlands
b.carpentieri@rug.nl

[2]University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, P. R. China
yanfeijing@uestc.edu.cn, tzhuang@uestc.edu.cn

[3]Beijing Institute of Technology, No.5, Zhongguancun Nandajie Haidian District, Beijing 100081,
P. R. China
peao_kelvin@163.com, xsheng@bit.edu.cn

***Abstract*—** We report on experiments with a novel family of Krylov subspace methods for solving dense, complex, non-Hermitian systems of linear equations arising from the Galerkin discretization of surface integral equation models in Electromagnetics. By some experiments on realistic radar-cross-section calculation, we illustrate the numerical efficiency of the proposed class of algorithms also against other popular iterative techniques in use today.

***Index Terms*—** Krylov subspace methods, Lanczos biconjugate A-orthonormalization methods, multilevel fast multipole method, scattering problems, sparse approximate inverse preconditioning.

I. INTRODUCTION

Mathematical models based on surface integral equations are becoming very popular in computational electromagnetics. They require a simple description of the surface of the target by means of triangular facets, thus simplifying considerably the mesh generation especially in the case of moving objects. Upon discretization they typically yield smaller systems to solve in comparison with finite difference or finite element techniques applied to the same problem [1]. The potential drawback of using integral methods is that they require to solve large dense complex systems of linear equations. Therefore, robust matrix solvers are urgently needed for this problem class [2].

The Maxwell's equations can be reformulated as a set of integral equations defined in the frequency domain as the following variational problem:

Find the surface current j such that for all tangential test functions j^t, we have

$$\iint_\Gamma G(x,y)\left(j(x) \cdot j^t(y)\right)dxdy$$

$$-\frac{1}{k^2}\iint_\Gamma G(x,y)\left(div_\Gamma j(x) \cdot div_\Gamma j^t(y)\right)dxdy \quad (1)$$

$$= \frac{i}{kZ_0}\int_\Gamma E_{inc}(x)\cdot j^t(x)dx.$$

We denote by $G(x,y) = \dfrac{e^{ik|y-x|}}{4\pi|y-x|}$ the Green's function of Helmholtz equation, Γ the boundary of the object, k the wave number and $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ the characteristic impedance of vacuum (ε_0 is the electric permittivity and μ_0 the magnetic permeability), and $div j(x)$ is the divergence operator of a continuously differentiable vector field $j(x)$ defined on a 3D Euclidean space. Equation (1) expresses the electric currents in terms of the electric field and is known as electric field integral equation (EFIE). It is applied to model arbitrary geometries like objects with cavities, disconnected parts, breaks

on the surface [3,4]. For its generality, the EFIE model is very popular in industrial environment. However, it is tough to solve by iterative methods, compared to other surface integral formulations of electromagnetic scattering problems.

On discretizing Equation (1) in space by the MoM over a mesh containing n edges, the surface current j is expanded into a set of basis functions $\{\varphi_i\}_{1 \le i \le n}$ with compact support (the Rao-Wilton-Glisson basis [5] is a popular choice), then the integral equation is applied to a set of tangential test functions j^t. Selecting $j^t = \varphi_j$, we are led to compute the set of coefficients $\{\lambda_i\}_{1 \le i \le n}$ such that

$$\sum_{i=1}^{n} \lambda_i \iint_{\Gamma} G(x,y)\left(\varphi_i(x) \cdot \varphi_j(y)\right)dxdy$$
$$-\frac{1}{k^2}\sum_{i=1}^{n} \lambda_i \iint_{\Gamma} G(x,y)\left(div_{\Gamma}\varphi_i(x) \cdot div_{\Gamma}\varphi_j(y)\right)dxdy$$
$$= \frac{i}{kZ_0} \int_{\Gamma} E_{inc}(x) \cdot \varphi_j(x)dx, \quad (2)$$

for each $1 \le i \le n$. The set of equations (2) can be recast in matrix form as

$$A\lambda = b, \quad (3)$$

where $A = [A_{ij}]$ and $b = [b_i]$ have elements

$$A_{ij} = \iint_{\Gamma} G(x,y)\left(\varphi_i(x) \cdot \varphi_j(y)\right)dxdy$$
$$-\frac{1}{k^2} \iint_{\Gamma} G(x,y)\left(div_{\Gamma}\varphi_i(x) \cdot div_{\Gamma}\varphi_j(y)\right)dxdy, \quad (4)$$

$$b_j = \frac{i}{kZ_0} \int_{\Gamma} E_{inc}(x) \cdot \varphi_j(x)dx. \quad (5)$$

In Equation (3), the set of unknowns are associated with the vectorial flux across an edge in the mesh. The coefficient matrix A generated by MoM is dense complex non-Hermitian; hence the pertinent linear system cannot be solved using the conjugate gradient (CG) algorithm. The restarted generalized minimal residual (GMRES) method, its flexible variant FGMRES, and some of the short-recurrence methods such as BiCG, BiCGStab, and TFQMR are popular options, see e.g. [6,7].

In this study, we illustrate experiments with two recently developed algorithms: the conjugate A-orthogonal residual squared (CORS) and the biconjugate A-orthogonal residual stabilized (BiCORSTAB) methods for non-Hermitian linear systems, sketched in Algorithms 1-2. They compute the approximate solution x_m that belongs to the Krylov subspace $x_0 + K_m(A; v_1)$ by projecting the residual orthogonally to the constraints subspace $\mathscr{L}_m \equiv A^H K_m(A^H; w_1)$. Throughout this paper, we denote by the superscript H the Hermitian (conjugate transpose) of a vector or a matrix and the standard inner product of two complex vectors $u, v \in \mathbb{C}^n$ by

$$\langle u, v \rangle = u^H v = \sum_{i=1}^{n} \bar{u}_i v_i.$$

For the sake of conciseness, we point the reader to [8,9] for a thorough mathematical derivation of the BiCORSTAB and CORS methods.

II. NUMERICAL EXPERIMENTS

For the numerical experiments, we consider some selected scattering problems described in Table 1. We report the number of iterations required by several Krylov methods (listed in Table 2) to reduce the initial residual by five orders of magnitude, starting from the zero vector. The sequential tests are compiled with the Portland Group Fortran 90 compiler (version 9) and run on a cluster of nodes equipped with quad core Intel CPU (2.8 GHz) and 16 GB of physical RAM.

In our sequential experiments, the CORS method was the most effective non-Hermitian solver with respect to CPU time, as it is shown in Table 3. Unrestarted GMRES may outperform all other Krylov methods and should be used when memory is not a concern. We selected a value of 50 for the restart parameter in the GMRES method in our runs on small problems, reported in Table 3, and a value of 100 in the runs on large problems, reported in Tables 4-5. In Figure 1, we illustrate the convergence history of CORS and GMRES(50) on Examples 3 to show the different numerical behaviors of the two families of solvers. The residual reduction is much smoother for GMRES along the iterations. The BiCORSTAB method also

Table 1: Characteristics of the model problems

Example	Description	Size	Frequency (MHz)
1	Open cylinder	6268	362
2	Sphere	12000	535
3	Satellite	1699	57
4	Cavity	727120	300
5	Paraboloid	857862	300

Table 2: List of solvers used and relative cost. We denote by n the problem size, by i the iteration number and by m the restart value in GMRES

Solver	Products by A/A^H	Memory
CORS	2/0	matrix+$14n$
BiCORSTAB	2/0	matrix+$13n$
GMRES	1/0	matrix+$(m+3)n$
QMR	2/1	matrix+$11n$
TFQMR	4/0	matrix+$10n$
BiCGSTAB	2/0	matrix+$7n$

shows fast convergence and may be an appropriate choice. Both CORS and BiCORSTAB are based on short-term recurrences and therefore, they are very cheap in memory (see Table 2).

Finally, methods based on Lanczos biconjugation are also considered in many scattering analysis, due to their simplicity (they are parameter-free) and low memory requirements, see e.g. [10–12]. In our experiments, as shown in Table 3, BiCGSTAB and QMR-like methods are less efficient than CORS.

A. MLFMA and SAI

A straightforward implementation of Krylov methods requires $\mathcal{O}(n^2)$ memory storage, where n is the number of unknowns, to compute a solution for one excitation. The solution cost may be reduced to $\mathcal{O}(n \log n)$ algorithmic and memory complexity using the multilevel fast multipole algorithm (MLFMA) for the M-V operation. Recent progress in the developments of parallel multipole codes, provably scalable to several million discretization points, are urging the quest of robust iterative algorithms for this problem class [13,14]. In this study, we solved the two largest problems *i.e.* Examples 4 and 5, using MLFMA and a sparse approximate inverse (SAI) preconditioner. The SAI preconditioner was computed by minimizing the Frobenius-norm of

Table 3: Number of iterations and CPU time (in seconds) required by Krylov methods to reduce the initial residual to $\mathcal{O}(10^{-5})$. For each example, asterisk "*" indicates the fastest run

Solver/Example	1	2	3
CORS	601 (253*)	294 (451*)	371 (11*)
BiCORSTAB	941 (614)	423 (1099)	775 (37)
GMRES(50)	2191 (469)	1803 (1397)	871 (17)
QMR	878 (548)	430 (1045)	452 (24)
TFQMR	482 (398)	281 (863)	373 (27)
BiCGSTAB	1065 (444)	680 (1031)	566 (18)

the error matrix

$$\min_{M \in S} \left\| I - M\widehat{A} \right\|_F ,$$

where S is the set of matrices with a given sparsity pattern. We chose $\widehat{A}$ to be sparse and equal to the multipole matrix, and we selected the sparsity pattern of M equal to the nonzero structure of $\widehat{A}$. Details of the SAI preconditioner and of the highly efficient parallel implementation of MLFMA that we used in this study are found in [15] and in [16], respectively.

The first model is a cavity of size $10\lambda \times 10\lambda \times 50\lambda$, discretized with 727,120 nodes and illuminated at an incident angle $(\theta, \phi) = (45°, 0°)$. The second model is a paraboloid of radius λ and focal depth 12λ, discretized with 857,862 nodes and illuminated at an incident angle $(\theta, \phi) = (45°, 0°)$. Besides the GMRES method, in these experiments we compare CORS and BiCORSTAB also against the FGMRES method preconditioned by an inner GMRES solver. This combination of Krylov methods is reported to be amazingly effective on this problem class [6,7]. In Tables 4-5, we report the number of iterations and CPU solution time necessary to achieve convergence on 16 processors. The solution process was declared a solver failure when the initial residual was not reduced by at least four orders of magnitude after 2000 M-V products (or 50 outer iterations for FGMRES). This level of accuracy on the final residual enabled us to calculate a correct radar-cross-section, which is shown in Figure 3 for BiCORSTAB and CORS. We notice again the remarkable robustness and efficiency of the two algorithms.

(a)

Fig. 2. Geometries of the largest model problems: (left) Example 4: a cavity, and (right) Example 5: a paraboloid.

(b)

Fig. 1. Example 3: a satellite. (a) The discretized mesh. (b) The convergence history of the CORS vs. the restarted GMRES methods.

III. ENHANCING THE ROBUSTNESS OF CORS BY DEFLATION

It is known that the convergence of an iterative method is mostly dictated by the distribution of the eigenvalues of the coefficient matrix MA. For GMRES, the residual reduction after k iterations writes

$$\|r_k\|_2 / \|r_0\|_2 \leqslant \kappa(V) \min_{p_k} \max_{i=1,\dots,n} |p_k(\lambda_i)|, \quad (6)$$

where $\kappa(V)$ is the condition number of the eigenvector matrix. If $\kappa(V) \not\gg 1$, the problem of describing the convergence of GMRES reduces

Table 4: Iterations count for the experiments with MLFMA and SAI

Example 4: Cavity $10\lambda \times 10\lambda \times 50\lambda$ Size: 727,120, Setup time SAI: 57s, Nr. procs: 16				
	GMRES(100)	CORS	BiCORSTAB	FGMRES
No prec.	>2000	239	280	21
SAI	798	64	56	10
Example 5: Paraboloid of radius 25λ and focal depth of 12λ Size: 857,862, Setup time SAI: 148s, Nr. procs: 16				
	GMRES(100)	CORS	BiCORSTAB	FGMRES
No prec.	1065)	367	410	11
SAI	112	42	36	5

to a problem in approximation theory: how well can one approximate zero on the set of complex eigenvalues using a kth-degree polynomial with value 1 at the origin. From Eq. (6), we see that the presence of small eigenvalues close to the origin in the spectrum of the coefficient matrix of the preconditioned linear system may lead to highly oscillatory polynomials with high degree k, and therefore may increase the number of iterations

Table 5: CPU time for the experiments with MLFMA and SAI. For each example, asterisk "*" indicates the fastest run

Example 4: Cavity $10\lambda \times 10\lambda \times 50\lambda$ Size: 727,120, Setup time SAI: 57s, Nr. procs: 16				
	GMRES(100)	CORS	BiCORSTAB	FGMRES
No prec.	>651s	144s*	168s	2076s
SAI	325s	54s	51s*	898s
Example 5: Paraboloid of radius 25λ and focal depth of 12λ Size: 857,862, Setup time SAI: 148s, Nr. procs: 16				
	GMRES(100)	CORS	BiCORSTAB	FGMRES
No prec.	693s	449s*	495s	1071s
SAI	89s	72s	60s*	160s

(a)

(b)
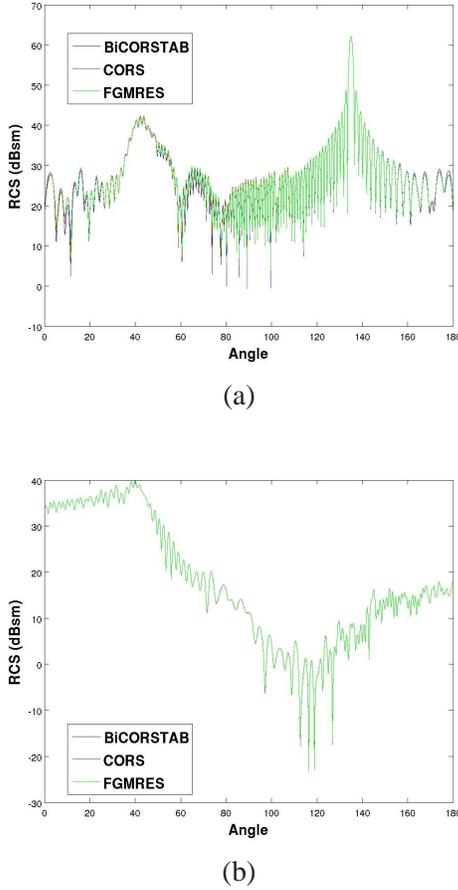
Fig. 3. Comparative curves of the radar-cross-section (RCS) calculation for (a) the cavity problem, and (b) the paraboloid problem.

of GMRES to obtain convergence. In this section, we show how to enhance the robustness of the CORS method by dumping the slowly converging components of the residuals associated to the smallest eigenvalues. This may finally result in considerably faster convergence.

Let $MA = V\Lambda V^{-1}$ be an eigendecomposition of the preconditioned matrix MA, with $\Lambda = diag(\lambda_i)$, $|\lambda_1| \leq \ldots \leq |\lambda_n|$ are the eigenvalues of MA and V is the matrix collecting the associated right eigenvectors of MA. We denote by V_ε the matrix of the right eigenvectors of MA associated to the eigenvalues λ_i such that $|\lambda_i| \leq \varepsilon$. We also denote by $A_c = V_\varepsilon^H(MA)V_\varepsilon$ the projection of MA in the eigenspace spanned by V_ε, and by $M_c = V_\varepsilon A_c^{-1} V_\varepsilon^H$ its prolongation back to the

original space.

Then, the following result holds.

Theorem 1: Let

$$\tilde{A}_c = V^H A V_\varepsilon \text{ has full rank,}$$

$$\tilde{M}_c = V_\varepsilon \tilde{A}_c^{-1} V^H,$$

and

$$\tilde{M} = M + \tilde{M}_c.$$

Then $\tilde{M}A$ is similar to a matrix whose eigenvalues are

$$\begin{cases} \eta_i = \lambda_i & \text{if} \quad |\lambda_i| > \varepsilon, \\ \eta_i = 1 + \lambda_i & \text{if} \quad |\lambda_i| \leq \varepsilon. \end{cases}$$

A_c represents the projection of the matrix MA on the coarse space defined by the approximate eigenvectors associated with its smallest eigenvalues.

Proof

We first remark that $A_c = diag(\lambda_i u_i^H v_i)$ with $|\lambda_i| \leq \varepsilon$ and so A_c is nonsingular. A_c represents the projection of the matrix MA on the space spanned by the approximate eigenvectors associated with its smallest eigenvalues.

Let $V = (V_\varepsilon, V_{\bar{\varepsilon}})$, where $V_{\bar{\varepsilon}}$ is the set of $(n - k)$ right eigenvectors associated with eigenvalues $|\lambda_i| > \varepsilon$.

Let $D_\varepsilon = diag(\lambda_i)$ with $|\lambda_i| \leq \varepsilon$ and $D_{\bar{\varepsilon}} = diag(\lambda_j)$ with $|\lambda_j| > \varepsilon$.

The following relations hold: $\tilde{M}AV_\varepsilon = V_\varepsilon(D_\varepsilon + I_k)$ and, $\tilde{M}AV_{\bar{\varepsilon}} = V_{\bar{\varepsilon}}D_{\bar{\varepsilon}} + V_\varepsilon C$ with $C = A_c^{-1} W^H A V_{\bar{\varepsilon}}$; then we have

$$\tilde{M}AV = V \begin{pmatrix} D_\varepsilon + I_k & C \\ 0 & D_{\bar{\varepsilon}} \end{pmatrix}.$$

For right preconditioning, that is $AMy = b$, similar results hold. Observe that the effect of applying the low-rank correction is to completely removed the effect of the k smallest eigenvalues from the spectrum of MA. Therefore, we may expect that an iterative method may converge faster on the transformed linear system.

The spectral corrections may be implemented in the CORS algorithm as follows.

1) Compute an approximation of the invariant subspace V_ε associated to the smallest eigenvalues of MA.

2) Construct the projected matrix $A_c = V_\varepsilon^H (MA) V_\varepsilon$ and the prolongation matrix $M_c = V_\varepsilon A_c^{-1} V_\varepsilon^H$. The matrix A_c is an $k \times k$ matrix, where k is the number of small eigenvalues that we want to remove, while M_c is $n \times n$.

3) Update the preconditioned vector quantities p in CORS as $p \leftarrow p + M_c p$

The same idea and computational scheme may be extended to the BiCORSTAB algorithm.

We applied deflated CORS (or, shortly DCORS) to a complete RCS calculation, which requires to solve linear systems with multiple right-hand sides. Take as incident field a plane wave of general form in spherical coordinates

$$\vec{E}_{inc}(x, \varphi, p_\theta, p_\varphi) = p_\theta \hat{u}_\theta e^{ikx \cdot \hat{u}_r \varphi} + p_\varphi \hat{u}_\varphi e^{ikx \cdot \hat{u}_r \varphi},$$

where p_θ, p_φ are two complex numbers and $\hat{u}_r, \hat{u}_\theta, \hat{u}_\varphi$ are the unitary vectors:

$$\hat{u}_r = \begin{pmatrix} \cos\varphi\cos\theta \\ \sin\varphi\cos\theta \\ \sin\theta \end{pmatrix}, \hat{u}_\theta = \begin{pmatrix} -\cos\varphi\sin\theta \\ -\sin\varphi\sin\theta \\ \cos\theta \end{pmatrix},$$

$$\hat{u}_\varphi = \begin{pmatrix} -\sin\varphi\cos\theta \\ -\cos\varphi\cos\theta \\ \sin\theta \end{pmatrix}.$$

Choose $\theta = 0$ and increase φ of one degree each time from 0 to π. Then, we obtain a sequence of 180 linear systems, each of them having the same coefficient matrix and a different right-hand side, associated to the following expression for the incident field

$$\vec{E}_{inc}(x) = \hat{z} e^{ikx \cdot \hat{u}_r(\varphi)} = \hat{z} e^{ik(x_1 \cos\varphi + x_2 \sin\varphi)}.$$

In the experiments reported in Table 6, the preconditioner M is the SAI method computed using sixty nonzeros per column, and we deflate an approximate invariance eigenspace V_ε of dimension 10. The invariant subspace V_ε is computed using the ARPACK library [17]. In our runs, the extra cost to setup DCORS is quickly amortized using only three right-hand sides for the satellite problem and four for the cylinder problem.

IV. CONCLUSIONS

We have analyzed the performance of two novel Krylov projection methods computed from

Table 6:   Experiments with deflated CORS

On the satellite: SAI = 60 - AV = 3		
	CORS	DCORS(10)
Avg Its on 180 RHS	62.8	38.2
Total solution time (sec)	275.8	198.9
On the cylinder - SAI = 60		
	CORS	DCORS(10)
Avg Its on 180 RHS	193.4	123.7
Total solution time (sec)	365.5	249.4

the Lanczos biconjugate A-orthonormalization method for solving dense complex non-Hermitian linear systems in radar-cross-section calculation. This family of solvers shows good convergence properties, is cheap in memory as it is derived from short-term vector recurrences, is parameter-free and does not suffer from the restriction to require a symmetric preconditioner. Additionally, it does not necessitate of matrix multiplication by A^H that might be tricky to implement in some integral application codes combined with MLFMA. Finally, we have illustrated how to possibly enhance the efficiency of the methods for solving linear systems with multiple right-hand sides, a typical scenario arising in realistic RCS calculation in industry.

The numerical results indicate that the proposed solvers may be an efficient alternative to other popular methods especially when robustness and memory are concerns.

Algorithm 1 *Left preconditioned BiCORSTAB method.*

1: Compute $r_0 = b - Ax_0$ for some initial guess x_0.
2: Choose $r_0^* = P(A)r_0$ such that $\langle r_0^*, Ar_0 \rangle \neq 0$, where $P(t)$ is a polynomial in t. (For example, $r_0^* = Ar_0$).
3: **for** $j = 1, 2, \ldots$ **do**
4:    **solve** $Mz_{j-1} = r_{j-1}$
5:    $\hat{z} = Az_{j-1}$
6:    $\rho_{j-1} = \langle r_0^*, \hat{z} \rangle$
7:    **if** $\rho_{j-1} = 0$, **method fails**
8:    **if** $j = 1$ **then**
9:       $p_0 = r_0$
10:      **solve** $Mzp_0 = p_0$
11:      $q_0 = \hat{z}$
12:    **else**
13:      $\beta_{j-2} = (\rho_{j-1}/\rho_{j-2}) \times (\alpha_{j-2}/\omega_{j-2})$
14:      $p_{j-1} = r_{j-1} + \beta_{j-2}(p_{j-2} - \omega_{j-2}q_{j-2})$
15:      $zp_{j-1} = z_{j-1} + \beta_{j-2}(zp_{j-2} - \omega_{j-2}zq_{j-2})$
16:      $q_{j-1} = \hat{z} + \beta_{j-2}(q_{j-2} - \omega_{j-2}\hat{z}q_{j-2})$
17:    **end if**
18:    **solve** $Mzq_{j-1} = q_{j-1}$
19:    $\hat{z}q_{j-1} = Azq_{j-1}$
20:    $\alpha_{j-1} = \rho_{j-1}/\langle r_0^*, \hat{z}q_{j-1} \rangle$
21:    $s = r_{j-1} - \alpha_{j-1}q_{j-1}$
22:    if $\|s\|_2$ is small, set $x_j = x_{j-1} + \alpha_{j-1}zp_{j-1}$ and stop
23:    $zs = z_{j-1} - \alpha_{j-1}zq_{j-1}$
24:    $t = \hat{z} - \alpha_{j-1}\hat{z}q_{j-1}$
25:    $\omega_{j-1} = \langle t, s \rangle / \langle t, t \rangle$
26:    $x_j = x_{j-1} + \alpha_{j-1}zp_{j-1} + \omega_{j-1}zs$
27:    $r_j = s - \omega_{j-1}t$
28:    check convergence; continue if necessary and $\omega_{j-1} \neq 0$
29: **end for**

Algorithm 2 *Left preconditioned CORS method.*

1: Compute $r_0 = b - Ax_0$ for some initial guess x_0.
2: Choose $r_0^* = P(A)r_0$ such that $\langle r_0^*, Ar_0 \rangle \neq 0$, where $P(t)$ is a polynomial in t. (For example, $r_0^* = Ar_0$).
3: **for** $j = 1, 2, \ldots$ **do**
4:    **solve** $Mz_{j-1} = r_{j-1}$
5:    $\hat{r} = Az_{j-1}$
6:    $\rho_{j-1} = \langle r_0^*, \hat{r} \rangle$
7:    **if** $\rho_{j-1} = 0$, **method fails**
8:    **if** $j = 1$ **then**
9:       $e_0 = r_0$
10:      solve $Mze_0 = e_0$
11:      $d_0 = \hat{r}$
12:      $q_0 = \hat{r}$
13:    **else**
14:      $\beta_{j-2} = \rho_{j-1}/\rho_{j-2}$
15:      $e_{j-1} = r_{j-1} + \beta_{j-2}h_{j-2}$
16:      $ze_{j-1} = z_{j-1} + \beta_{j-2}zh_{j-2}$
17:      $d_{j-1} = \hat{r} + \beta_{j-2}f_{j-2}$
18:      $q_{j-1} = d_{j-1} + \beta_{j-2}(f_{j-2} + \beta_{j-2}q_{j-2})$
19:    **end if**
20:    **solve** $Mq = q_{j-1}$
21:    $\hat{q} = Aq$
22:    $\alpha_{j-1} = \rho_{j-1}/\langle r_0^*, \hat{q} \rangle$
23:    $h_{j-1} = e_{j-1} - \alpha_{j-1}q_{j-1}$
24:    $zh_{j-1} = ze_{j-1} - \alpha_{j-1}q$
25:    $f_{j-1} = d_{j-1} - \alpha_{j-1}\hat{q}$
26:    $x_j = x_{j-1} + \alpha_{j-1}(2ze_{j-1} - \alpha_{j-1}q)$
27:    $r_j = r_{j-1} - \alpha_{j-1}(2d_{j-1} - \alpha_{j-1}\hat{q})$
28:    check convergence; continue if necessary
29: **end for**

REFERENCES

[1] W. C. Chew, J. M. Jin, C. C. Lu, E. Michielssen, and J. M. Song, "Fast Solution Methods in Electromagnetics," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, pp. 533–543, 1997.

[2] B. Carpentieri, "Fast Iterative Solution Methods in Electromagnetic Scattering," *Progress in Electromagnetic Research*, vol. 79, pp. 151–178, 2007.

[3] F. Bilotti and C. Vegni, "MoM Entire Domain Basis Functions for Convex Polygonal Patches," *Journal of Electromagnetic Waves and Applications*, vol. 17, no. 11, pp. 1519–1538, 2003.

[4] J. Li, L. Li, and Y. Gan, "Method of Moments Analysis of Waveguide Slot Antennas using the EFIE," *Journal of Electromagnetic Waves and Applications*, vol. 19, no. 13, pp. 1729–1748, 2005.

[5] S. Rao, D. Wilton, and A. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shape," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 409–418, 1982.

[6] B. Carpentieri, I. Duff, L. Giraud, and G. Sylvand, "Combining Fast Multipole Techniques and an Approximate Inverse Preconditioner for Large Electromagnetism Calculations," *SIAM J. Scientific Computing*, vol. 27, no. 3, pp. 774–792, 2005.

[7] T. Malas, Ö. Ergül, and L. Gürel, "Sequential and Parallel Preconditioners for Large-Scale Integral-Equation Problems," in *Computational Electromagnetics Workshop, August 30-31, 2007, Izmir, Turkey*, pp. 35–43, 2007.

[8] Y.-F. Jing, T.-Z. Huang, Y. Zhang, L. Li, G.-H. Cheng, Z.-G. Ren, Y. Duan, T. Sogabe, and B. Carpentieri, "Lanczos-Type Variants of the COCR Method for Complex Nonsymmetric Linear Systems," *Journal of Computational Physics*, vol. 228, no. 17, pp. 6376–6394, 2009.

[9] B. Carpentieri, Y.-F. Jing, and T.-Z. Huang, "The BiCOR and CORS Algorithms for Solving Nonsymmetric Linear Systems," *SIAM J. Scientific Computing*, vol. 33, no. 5, pp. 3020–3036, 2011.

[10] J. Song, C.-C. Lu, and W. Chew, "Multilevel Fast Multipole Algorithm for Electromagnetic Scattering by Large Complex Objects," *IEEE Transactions on Antennas and Propagation*, vol. 45(10), pp. 1488–1493, 1997.

[11] J. Rahola and S. Tissari, "Iterative Solution of Dense Linear Systems Arising from the Electrostatic Integral Equation in MEG," *Physics in Medicine and Biology*, vol. 47(6), pp. 961–975, 2002.

[12] J. Lee, C.-C. Lu, and J. Zhang, "Sparse Inverse Preconditioning of Multilevel Fast Multipole Algorithm for Hybrid Integral Equations in Electromagnetics," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 9, pp. 2277–2287, 2004.

[13] Ö. Ergül and L. Gürel, "Efficient Parallelization of the Multilevel Fast Multipole Algorithm for the Solution of Large-Scale Scattering Problems," *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 8, pp. 2335–2345, 2008.

[14] X. Pan and X. Sheng, "A Highly Efficient Parallel Approach of Multi-Level Fast Multipole Algorithm," *Journal of Electromagnetic Waves and Applications*, vol. 20, no. 8, pp. 1081–1092, 2006.

[15] X.-M. Pan and X.-Q. Sheng, "An Efficient Parallel SAI Preconditioner for Multilevel Fast Multipole Algorithm for Scattering by Extremely Large Complex Targets," in *Microwave Technology and Computational Electromagnetics, 2009. ICMTCE. International Conference on*, pp. 407 –410, Nov. 2009.

[16] X.-M. Pan and X.-Q. Sh, "A Sophisticated Parallel MLFMA for Scattering by Extremely Large Targets [EM Programmer's Notebook]," *Antennas and Propagation Magazine, IEEE*, vol. 50, no. 3, pp. 129 –138, June 2008.

[17] R. Lehoucq, D. Sorensen, and C. Yang, *ARPACK User's Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[18] N. Nakashima, S. Fujino, and M. Tateiba, "Performance Evaluation of State-of-the-Art Linear Iterative Solvers Based on IDR Theorem for Large Scale Electromagnetic Multiple Scattering Simulations," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 1, pp. 37–44, 2011.

[19] Y. Zhang and Q. Sun, "Complex Incomplete Cholesky Factorization Preconditioned Bi-conjugate Gradient Method," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 9, pp. 750–754, 2010.

[20] D. Ding, J. Ge, and R. Chen, "Well-Conditioned CFIE for Scattering from Dielectric Coated Conducting Bodies above a Half-Space," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 11, pp. 936–946, 2010.

Bruno Carpentieri received a Laurea degree in Applied Mathematics from Bari University, Italy, in 1998, and a Ph.D. degree in Computer Science from the Institut National Polytechnique of Toulouse, France, in 2002. After completing his Ph.D., he was as post-doctoral fellows at the CERFACS Institute in Toulouse, France (2003-2004), and at Karl-Franzens University in Graz, Austria (2005-2008). He also served as a consultant on European projects at CRS4 in Sardinia, Italy (2008-2009). Since January 2010, he is an University Assistant at the Institute of Mathematics and Computing Science of the University of Groningen, The Netherlands. His research interests include numerical linear algebra, parallel computing, Computational Fluid Dynamics, Computational Electromagnetics, Cardiac Modelling.

Yan-Fei Jing received the B.S. and Ph.D. degrees in Applied Mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2005 and 2011, respectively. He is currently an Associated Professor in the School of Mathematical Sciences, UESTC. He is the author or coauthor of more than 18 research papers and is currently a member of International Linear Algebra Society (ILAS). His research interests include iterative and preconditioning methods for linear systems with applications in computational electromagnetics.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in Computational Mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992 and 2000, respectively. During 2005, he was a visiting scholar in Dept. of Computer Science, Loughborough University of UK.

He is currently a professor in the School of Mathematical Sciences, UESTC. He is currently an editor of several journals such as Advances in Numer. Anal., J. Pure and Appl. Math.: Advances and Appl, Computer Science and Appl., J. Electro. Sci. and Tech. of China, Advances in Pure Math., etc. He is the author or coauthor of more than 100 research papers. His current research interests include numerical linear algebra with applications in computational electromagnetics and image science, and matrix analysis.

Wei-Chao Pi received the B.S. degree in Electronic Engineering from the school of Information and Electronics, University of Beijing Institute of Technology (BIT), Beijing, China, in 2009. Then he joined the Computational Electromagnetics Center of BIT. His research interest mainly focus on the development of highly efficient methods in multilevel fast multipole algorithm.

Xin-Qing Sheng received the B.S., M.S., and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 1991, 1994, and 1996, respectively. From 1996 to 1998, Dr. Sheng was a Postdoctoral Research Fellow with the Center for Computational Electromagnetics, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. From 1998 to 2002, Dr. Sheng was with the Department of Electronic Engineering, City University of Hong Kong, first as a Research Fellow and then as a Senior Research Fellow. From 2002-2005, Dr. Sheng

was a Professor with the Institute of Electronics, Chinese Academy of Sciences. Since 2005, Dr. Sheng has become a Chang-Jiang Professor of the Department of Electronic Engineering and Director of Center for Electromagnetic Simulation at the Beijing Institute of Technology. He has authored and coauthored over 80 papers in refereed journals. He has authored A Brief Treatise on Computational Electromagnetics (Beijing: Science Press, 2004) and A Treatise on Electromagnetic Wave (Beijing: Science Press, 2007). He has coauthored SINOCOM, a simulation software for scattering by complex targets. His research interests include computational electromagnetics, scattering and antenna analysis, electromagnetic compatibility, microwave imaging. Dr. Sheng is a recipient of the 1995 President Awards of the Chinese Academy of Sciences, the 2001 One Hundred Talents Program awarded by the Chinese Academy of Sciences, the 2004 Cheung Kong Scholar Program awarded by the Ministry of Education, China. Dr. Sheng received the first prize of Beijng Science and Technology Awards in 2009.

Investigating the Composite Step Biconjugate A-Orthogonal Residual Method for Non-Hermitian Dense Linear Systems in Electromagnetics

Yan-Fei Jing [1], Ting-Zhu Huang [1], Bruno Carpentieri [2], and Yong Duan [1]

[1]University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, P. R. China
yanfeijing@uestc.edu.cn, tzhuang@uestc.edu.cn, duanyong72@yahoo.com.cn

[2]University of Groningen, Nijenborgh 9, PO Box 407, 9700 AK Groningen, Netherlands
b.carpentieri@rug.nl

Abstract— An interesting stabilizing variant of the biconjugate A-orthogonal residual (BiCOR) method is investigated for solving dense complex non-Hermitian systems of linear equations arising from the Galerkin discretization of surface integral equations in electromagnetics. The novel variant is naturally based on and inspired by the composite step strategy employed for the composite step biconjugate gradient method from the point of view of pivot-breakdown treatment when the BiCOR method has erratic convergence behaviors. Besides reducing the number of spikes in the convergence history of the norm of the residuals to the greatest extent, the present composite step BiCOR method can provide some further practically desired smoothing behavior towards stabilizing the numerical performance of the BiCOR method in the case of irregular convergence.

Index Terms— Krylov subspace methods, Lanczos biconjugate A-orthonormalization methods, scattering problems, sparse approximate inverse preconditioning.

I. INTRODUCTION

Solution of large linear systems is crucial to many numerical simulations in science and engineering [1,2]. Many real-world applications demand an accurate numerical solution of physical problems arising from fields such as fluid mechanics, structural engineering, computational electromagnetics, design and computer analysis of circuits, power system networks, and economics models [3]. Take scattering problems of determining the diffraction pattern irradiated by an illuminated object for instance. They include medical imaging, electromagnetic compatibility, radar cross section (RCS) calculation of large objects. Krylov subspace methods, which are considered as one of the "Top Ten Algorithms of the 20th Century" [4], are one of the most widespread and extensively accepted techniques for iterative solution of today's large-scale linear systems [5]. The starting point for this work was the investigation of the applicability of a new interesting stabilizing variant of the biconjugate A-orthogonal residual (BiCOR) method [7] to iterative solution of non-Hermitian systems of linear equations in electromagnetism applications.

In recent years, there have been many advances in Krylov subspace methods for solution of large linear systems [5]. Different variants of restarted, augmented, deflated, flexible, nested, and inexact methods are involved in these new developments. Various methods differ in the way they extract information from Krylov spaces [8–10]. Observing from earlier studies on surface scattering problems, different Krylov subspace methods have both advantages and disadvantages [11]. For instance, the GMRES method is robust but expensive due to long recurrence in the underlying Arnoldi procedure. Restarting the GMRES deteriorates convergence significantly. The BiCGSTAB method typically requires many more iterations than the GMRES method, especially on complex

geometries. QMR-like methods are only slightly more competitive than the BiCGSTAB method, but less robust than the GMRES method.

A novel Lanczos-type biconjugate A-orthonormalization procedure has recently been established to give birth to a new family of efficient short-recurrence methods for large real nonsymmetric and complex non-Hermitian systems of linear equations, named as the Lanczos biconjugate A-orthonormalization methods [7]. As observed from numerous numerical experiments carried out with the Lanczos biconjugate A-orthonormalization methods, it has been numerically demonstrated that this family of solvers shows competitive convergence properties, is cheap in memory as it is derived from short-term vector recurrences, is parameter-free and does not suffer from the restriction to require a symmetric preconditioner like other methods [12–15]. However, the family of Lanczos biconjugate A-orthonormalization methods is often faced with apparently irregular convergence behaviors appearing as "spikes" in the convergence history of the norm of the residuals, possibly leading to substantial build-up of rounding errors and worse approximate solutions, or possibly even overflow. Therefore, it is quite necessary to tackle their irregular convergence properties to obtain more stabilized variants so as to improve the accuracy of the desired physical numerical solutions.

Our main attention in this paper is focused on the demonstration of the straightforward natural enhancement of the BiCOR method, which is the basic underlying variant of the above-mentioned family of Lanczos biconjugate A-orthonormalization methods. The content of this paper can be considered as the natural follow-up to the paper [16]. In particular, we exploit the composite step strategy taken for the composite step biconjugate gradient (CSBCG) method [17,18] from the point of view of pivot-breakdown treatment when the BiCOR method has erratic convergence behaviors. The outline of the paper is organized as follows. In the coming section, the good performance of the BiCOR algorithm in electromagnetics will be illustrated numerically by recalling some introductory comparative

experiments. In Section III, we present the interesting stabilizing variant–the composite step BiCOR (CSBiCOR) method with applications on a set of model problems representative of realistic RCS calculation to show the improved numerical performance with respect to the stabilizing effect of the composite step strategy on the BiCOR method. Conclusions and perspectives are finally made with some future research issues.

Throughout the paper, denote the overbar ("−") the conjugate complex of a scalar, vector or matrix and the superscript "T" the transpose of a vector or matrix. For a non-Hermitian matrix $A = (a_{ij})_{N \times N} \in \mathbb{C}^{N \times N}$, the Hermitian conjugate of A is denoted as

$$A^H \equiv \overline{A}^T = (\overline{a}_{ji})_{N \times N}.$$

The standard Hermitian inner product of two complex vectors $u, v \in \mathbb{C}^N$ is defined as

$$\langle u, v \rangle = u^H v = \sum_{i=1}^{N} \overline{u}_i v_i.$$

The nested Krylov subspace of dimension t generated by A from v is of the form

$$\mathcal{K}_t(A, v) = \operatorname{span}\{v, Av, A^2 v, \ldots, A^{t-1} v\}.$$

In addition, e_i denotes the ith column of the appropriate identity matrix.

II. PRELIMINARY REVIEW FOR THE BICOR METHOD

First, we briefly recall a version of the Lanczos biconjugate A-orthonormalization procedure [7] as in Algorithm 1, which can ideally build up a pair of biconjugate A-orthonormal bases for the dual Krylov subspaces $\mathcal{K}_m(A, v_1)$ and $\mathcal{K}_m(A^H, w_1)$, where v_1 and w_1 are chosen initially to satisfy certain conditions.

Observe that the above algorithm is possible to have Lanczos-type breakdown whenever δ_{j+1} vanishes while $\hat{w}_{j+1}$ and $A\hat{v}_{j+1}$ are not equal to $\mathbf{0} \in \mathbb{C}^N$ appearing in line 8. In the interest of counteraction against such breakdowns, we refer the reader to possible remedies proposed in earlier studies, such as so-called look-ahead strategies [19–27] which can enhance stability

Algorithm 1 *Biconjugate A-Orthonormalization Procedure*

1: *Choose v_1, ω_1, such that $\langle \omega_1, Av_1 \rangle = 1$*
2: Set $\beta_1 = \delta_1 \equiv 0, \omega_0 = v_0 \equiv \mathbf{0} \in \mathbb{C}^N$
3: **for** $j = 1, 2, \dots$ **do**
4: $\quad \alpha_j = \langle \omega_j, A(Av_j) \rangle$
5: $\quad \hat{v}_{j+1} = Av_j - \alpha_j v_j - \beta_j v_{j-1}$
6: $\quad \hat{\omega}_{j+1} = A^H \omega_j - \bar{\alpha}_j \omega_j - \delta_j \omega_{j-1}$
7: $\quad \delta_{j+1} = |\langle \hat{\omega}_{j+1}, A\hat{v}_{j+1} \rangle|^{\frac{1}{2}}$
8: $\quad \beta_{j+1} = \dfrac{\langle \hat{\omega}_{j+1}, A\hat{v}_{j+1} \rangle}{\delta_{j+1}}$
9: $\quad v_{j+1} = \dfrac{\hat{v}_{j+1}}{\delta_{j+1}}$
10: $\quad \omega_{j+1} = \dfrac{\hat{\omega}_{j+1}}{\bar{\beta}_{j+1}}$
11: **end for**

while increasing cost modestly. However, the analysis of these strategies is outside the scope of this paper and we shall not pursue that here. For more details, please refer to [5,10] and the references therein. In the present paper, we suppose there is no such Lanczos-type breakdown encountered during algorithm implementations because most of our considerations concern the exploration of the composite step strategy [17,18] to handle the pivot breakdown occurring in the BiCOR method for solving non-Hermitian linear systems in electromagnetics.

Next, some properties of the vectors produced by Algorithm 1 are reviewed [7] in the following proposition for the preparation of the theoretical basis of the composite step method.

Proposition 1: If Algorithm 1 proceeds m steps, then the right and left Lanczos-type vectors $v_j, j = 1, 2, \dots, m$ and $w_i, i = 1, 2, \dots, m$ form a biconjugate A-orthonormal system in exact arithmetic, i.e.,

$$\langle \omega_i, Av_j \rangle = \delta_{i,j}, 1 \le i, j \le m.$$

Furthermore, denote by $V_m = [v_1, v_2, \dots, v_m]$ and $W_m = [w_1, w_2, \dots, w_m]$ the $N \times m$ matrices and by $\underline{T_m}$ the extended tridiagonal matrix of the form

$$\underline{T_m} = \begin{bmatrix} T_m \\ \delta_{m+1} e_m^T \end{bmatrix},$$

where

$$T_m = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \end{bmatrix},$$

whose entries are the coefficients generated during the algorithm implementation, and in which $\alpha_1, \dots, \alpha_m, \beta_2, \dots, \beta_m$ are complex while $\delta_2, \dots, \delta_m$ are positive. Then with the biconjugate A-orthonormalization procedure, the following four relations hold

$$AV_m = V_m T_m + \delta_{m+1} v_{m+1} e_m^T, \tag{1}$$
$$A^H W_m = W_m T_m^H + \bar{\beta}_{m+1} \omega_{m+1} e_m^T, \tag{2}$$
$$W_m^H AV_m = I_m, \tag{3}$$
$$W_m^H A^2 V_m = T_m. \tag{4}$$

Given an initial guess x_0 to the non-Hermitian linear system $Ax = b$ associated with the initial residual $r_0 = b - Ax_0$, define a Krylov subspace $\mathcal{L}_m \equiv A^H \text{span}(W_m) = A^H \mathcal{K}_m(A^H, w_1)$, where W_m is defined in Proposition 1, $v_1 = \frac{r_0}{\|r_0\|_2}$ and w_1 is chosen arbitrarily such that $\langle w_1, Av_1 \rangle \neq 0$. But w_1 is often chosen to be equal to $\frac{Av_1}{\|Av_1\|_2^2}$ subjecting to $\langle w_1, Av_1 \rangle = 1$. It is worthy noting that this choice for w_1 plays a significant role in establishing the empirically observed superiority of the BiCOR method to the BiCR [6] method as well as to the BCG method [7]. Thus running Algorithm 1 m steps, we can seek an mth approximate solution x_m from the affine subspace $x_0 + \mathcal{K}_m(A, v_1)$ of dimension m, by imposing the Petrov-Galerkin condition

$$b - Ax_m \perp \mathcal{L}_m,$$

which can be mathematically written in matrix formulation as

$$(A^H W_m)^H (b - Ax_m) = 0. \tag{5}$$

Analogously, an mth dual approximation x_m^* of the corresponding dual system $A^H x^* = b^*$ is

sought from the affine subspace $x_0^* + \mathcal{K}_m(A^H, w_1)$ of dimension m by satisfying

$$b^* - A^H x_m^* \perp A\mathcal{K}_m(A, v_1),$$

which can be mathematically written in matrix formulation as

$$(AV_m)^H(b^* - A^H x_m) = 0, \qquad (6)$$

where, x_0^* is an initial dual approximate solution and V_m is defined in Proposition 1 with $v_1 = \frac{r_0}{\|r_0\|_2}$.

Consequently, the BiCOR iterates $x_j's$ can be computed by the coming Algorithm 2, which is just the unpreconditioned BiCOR method with the preconditioner M there taken as the identity matrix [7] and has been rewritten with the algorithmic scheme of the unpreconditioned BCG method as presented in [10,18].

Algorithm 2 *Algorithm BiCOR*

1: Compute $r_0 = b - Ax_0$ for some initial guess x_0.
2: Choose $r_0^* = P(A)r_0$ such that $\langle r_0^*, Ar_0 \rangle \neq 0$, where $P(t)$ is a polynomial in t. (For example, $r_0^* = Ar_0$).
3: Set $p_0 = r_0, p_0^* = r_0^*, q_0 = Ap_0, q_0^* = A^H p_0^*, \hat{r}_0 = Ar_0, \rho_0 = \langle r_0^*, \hat{r}_0 \rangle$.
4: **for** $n = 0, 1, \ldots$ **do**
5: $\quad \sigma_n = \langle q_n^*, q_n \rangle$
6: $\quad \alpha_n = \rho_n / \sigma_n$
7: $\quad x_{n+1} = x_n + \alpha_n p_n$
8: $\quad r_{n+1} = r_n - \alpha_n q_n$
9: $\quad x_{n+1}^* = x_n^* + \bar{\alpha}_n p_n^*$
10: $\quad r_{n+1}^* = r_n^* - \bar{\alpha}_n q_n^*$
11: $\quad \hat{r}_{n+1} = Ar_{n+1}$
12: $\quad \rho_{n+1} = \langle r_{n+1}^*, \hat{r}_{n+1} \rangle$
13: $\quad$ **if** $\rho_{n+1} = 0$, **method fails**
14: $\quad \beta_{n+1} = \rho_{n+1} / \rho_n$
15: $\quad p_{n+1} = r_{n+1} + \beta_{n+1} p_n$
16: $\quad p_{n+1}^* = r_{n+1}^* + \bar{\beta}_{n+1} p_n^*$
17: $\quad q_{n+1} = \hat{r}_{n+1} + \beta_{n+1} q_n$
18: $\quad q_{n+1}^* = A^H p_n^*$
19: $\quad$ check convergence; continue if necessary
20: **end for**

Before ending this section, we review some introductory comparative experiments to see

Table 1: Characteristics of the model problems

Example	Description	Size	Frequency (MHz)
1	Open cylinder	6,268	362
2	Sphere	12,000	535
3	Satellite	1,699	57

the good numerical performance of the BiCOR algorithm [12]. The set of linear systems selected for the numerical experiments arise from RCS calculations of realistic targets. They are dense complex non-Hermitian. We report the characteristics of the model problems in Table 1. Although not very large, the selected problems are representative of realistic RCS calculation. Their solution demands considerable computer resources as it can be seen in the table. Larger problems require using the multilevel fast multipole algorithm (MLFMA) [28–31] for the M-V products to reduce the memory requirement and effective preconditioners to accelerate the convergence, and they are out of the scope of this study. We carried out the M-V product at each iteration using dense linear algebra packages, i.e. the ZGEMV routine available in the LAPACK library and we did not use preconditioning. In addition to the BiCOR method, we considered the other two evolving variants known as the conjugate A-orthogonal residual squared (CORS) method and the biconjugate A-orthogonal residual stabilized (BiCORSTAB) method, complex versions of iterative algorithms based on Lanczos biorthogonalization, such as BiCGSTAB and QMR, and on Arnoldi orthogonalization, such as GMRES. In Table 2, we list the complete set of solvers used in our experiments and their algorithmic and memory complexity. All the runs were done on one node of the *Entu* cluster facility located at CRS4. Each node features a quad core Intel CPU at 2.8GHz and 16 GB of physical RAM. The codes were compiled in Fortran with the Portland Group Fortran 90 compiler version 9.

In Table 3, we show the number of iterations and CPU time (in seconds) required by Krylov methods to reduce the initial residual to $\mathcal{O}(10^{-5})$ starting from the zero vector. The right-hand side of the linear system is set up so that the exact solution is

Table 2: Algorithmic cost and memory expenses of the implementation of Krylov algorithms that are used for the experiments. We denote by n the problem size, by i the iteration number and by m the restart value in GMRES

Solver	Products by A/A^H	Memory
BiCOR	1/1	matrix+$10n$
CORS	2/0	matrix+$14n$
BiCORSTAB	2/0	matrix+$13n$
GMRES	1/0	matrix+$(m+3)n$
QMR	2/1	matrix+$11n$
TFQMR	4/0	matrix+$10n$
BiCGSTAB	2/0	matrix+$7n$

Table 3: Number of iterations and CPU time (in seconds) required by Krylov methods to reduce the initial residual to $\mathcal{O}(10^{-5})$; for each example, an asterisk "*" indicates the fastest run

Solver/Example	1	2	3
CORS	601 (253*)	294 (451*)	371 (11*)
BiCOR	785 (334)	338 (525)	431 (15)
BiCORSTAB	941 (614)	423 (1099)	775 (37)
GMRES(50)	2191 (469)	1803 (1397)	871 (17)
QMR	878 (548)	430 (1045)	452 (24)
TFQMR	482 (398)	281 (863)	373 (27)
BiCGSTAB	1065 (444)	680 (1031)	566 (18)

the vector of all ones. We observe the effectiveness of the BiCOR method, that is the the second fastest non-Hermitian solver with respect to CPU time on most selected examples. Its performance is very close to that of the CORS method and may be an appropriate choice.

III. INVESTIGATION OF THE CSBICOR METHOD IN ELECTROMAGNETICS

Suppose Algorithm 2 runs successfully to step n, that is $\sigma_i \neq 0$, $\rho_i \neq 0$, $i = 0, 1, \ldots, n-1$. The BiCOR iterates satisfy the following properties [7].

Proposition 2: Let $R_{n+1} = [r_0, r_1, \ldots, r_n]$, $R^*_{n+1} = [r^*_0, r^*_1, \ldots, r^*_n]$ and $P_{n+1} = [p_0, p_1, \ldots, p_n]$, $P^*_{n+1} = [p^*_0, p^*_1, \ldots, p^*_n]$. We have

(1) $Range(R_{n+1}) = Range(P_{n+1}) = \mathcal{K}_{n+1}(A, r_0)$,
    $Range(R^*_{n+1}) = Range(P^*_{n+1}) = \mathcal{K}_{n+1}(A^H, r^*_0)$.

(2) $R^{*H}_{n+1} A R_{n+1}$ is diagonal.

(3) $P^{*H}_{n+1} A^2 P_{n+1}$ is diagonal.

Similarly to the breakdowns of the BCG method [18], it is observed from Algorithm 2 that there also exist two possible kinds of breakdowns for the BiCOR method:

(1) $\rho_n \equiv \langle r^*_n, \hat{r}_n \rangle \equiv \langle r^*_n, Ar_n \rangle = 0$ but r^*_n and Ar_n are not equal to $\mathbf{0} \in \mathbb{C}^N$ appearing in line 14;

(2) $\sigma_n \equiv \langle q^*_n, q_n \rangle \equiv \langle A^H p^*_n, Ap_n \rangle = 0$ appearing in line 6.

Although the computational formulae for the quantities where the breakdowns reside are different between the BiCOR method and the BCG method, we do not have a better name for them. Therefore, we still call the two cases of breakdowns described above as Lanczos breakdown and pivot breakdown, respectively.

The Lanczos breakdown can be cured using look-ahead techniques [19–27] as mentioned in the previous section, but such techniques require a careful and sophisticated way so as to make them become necessarily quite complicated to apply. This aspect of applying look-ahead techniques to the BiCOR method demands further research.

In this paper, we attempt to resort to the composite step idea employed for the CSBCG method [17,18] to handle the pivot breakdown of the BiCOR method with the assumption that the underlying biconjugate A-orthonormalization procedure depicted as in Algorithm 1 does not breakdown; that is the situation where $\sigma_n = 0$ while $\rho_n \neq 0$.

Suppose Algorithm 2 comes across a situation where $\sigma_n = 0$ after successful algorithm implementation up to step n with the assumption that $\rho_n \neq 0$, which indicates that the updates of x_{n+1}, r_{n+1}, x^*_{n+1}, r^*_{n+1} are not well defined. Taking the composite step idea, we will avoid division by $\sigma_n = 0$ via skipping this $(n+1)$th update and exploiting a composite step update to directly obtain the quantities in step $(n+2)$ with scaled versions of r_{n+1} and r^*_{n+1} as well as with the previous primary search direction vector p_n and shadow search direction vector p^*_n. The following process for deriving the CSBiCOR method is the same as that of the derivation of the CSBCG method [18] except for the different underlying procedures involved to correspondingly generate different Krylov subspace bases.

Analogously, define auxiliary vectors $z_{n+1} \in \mathcal{K}_{n+2}(A, r_0)$ and $z_{n+1}^* \in \mathcal{K}_{n+2}(A^H, r_0^*)$ as follows

$$
\begin{aligned}
z_{n+1} &= \sigma_n r_{n+1} \\
&= \sigma_n r_n - \rho_n A p_n, \quad (7)
\end{aligned}
$$

$$
\begin{aligned}
z_{n+1}^* &= \bar{\sigma}_n r_{n+1}^* \\
&= \bar{\sigma}_n r_n^* - \bar{\rho}_n A^H p_n^*, \quad (8)
\end{aligned}
$$

which are then used to look for the iterates $x_{n+2} \in x_0 + \mathcal{K}_{n+2}(A, r_0)$ and $x_{n+2}^* \in x_0^* + \mathcal{K}_{n+2}(A^H, r_0^*)$ in step $(n+2)$ as follows

$$
\begin{aligned}
x_{n+2} &= x_n + [p_n, z_{n+1}] f_n, \\
x_{n+2}^* &= x_n^* + [p_n^*, z_{n+1}^*] f_n^*,
\end{aligned}
$$

where, f_n, $f_n^* \in \mathbb{C}^2$. Correspondingly, the $(n+2)$th primary residual $r_{n+2} \in \mathcal{K}_{n+3}(A, r_0)$ and shadow residual $r_{n+2}^* \in \mathcal{K}_{n+3}(A^H, r_0^*)$ are respectively computed as

$$
r_{n+2} = r_n - A[p_n, z_{n+1}] f_n, \quad (9)
$$

$$
r_{n+2}^* = r_n^* - A^H[p_n^*, z_{n+1}^*] f_n^*. \quad (10)
$$

The biconjugate A-orthogonality condition between the BiCOR primary residuals and shadow residuals shown as Property (2) in Proposition 2 requires

$$
\begin{aligned}
\langle [p_n^*, z_{n+1}^*], A r_{n+2} \rangle &= 0, \\
\langle [p_n, z_{n+1}], A^H r_{n+2}^* \rangle &= 0,
\end{aligned}
$$

combining with Eqns. (9) and (10) gives rise to the following two 2×2 systems of linear equations for respectively solving f_n and f_n^*

$$
\begin{bmatrix} \langle A^H p_n^*, A p_n \rangle & \langle A^H p_n^*, A z_{n+1} \rangle \\ \langle A^H z_{n+1}^*, A p_n \rangle & \langle A^H z_{n+1}^*, A z_{n+1} \rangle \end{bmatrix} \begin{bmatrix} f_n^{(1)} \\ f_n^{(2)} \end{bmatrix} =
$$
$$
\begin{bmatrix} \langle p_n^*, A r_n \rangle \\ \langle z_{n+1}^*, A r_n \rangle \end{bmatrix}, \quad (11)
$$

$$
\begin{bmatrix} \langle A p_n, A^H p_n^* \rangle & \langle A p_n, A^H z_{n+1}^* \rangle \\ \langle A z_{n+1}, A^H p_n^* \rangle & \langle A z_{n+1}, A^H z_{n+1}^* \rangle \end{bmatrix} \begin{bmatrix} f_n^{*(1)} \\ f_n^{*(2)} \end{bmatrix} =
$$
$$
\begin{bmatrix} \langle A p_n, r_n^* \rangle \\ \langle A z_{n+1}, r_n^* \rangle \end{bmatrix}. \quad (12)
$$

Similarly, the $(n+2)$th primary search direction vector $p_{n+2} \in \mathcal{K}_{n+3}(A, r_0)$ and shadow search direction vector $p_{n+2}^* \in \mathcal{K}_{n+3}(A^H, r_0^*)$ in a composite step are computed with the following form

$$
p_{n+2} = r_{n+2} + [p_n, z_{n+1}] g_n, \quad (13)
$$

$$
p_{n+2}^* = r_{n+2}^* + [p_n^*, z_{n+1}^*] g_n^*, \quad (14)
$$

where, g_n, $g_n^* \in \mathbb{C}^2$.

The biconjugate A^2-orthogonality condition between the BiCOR primary search direction vectors and shadow search direction vectors shown as Property (3) in Proposition 2 requires

$$
\begin{aligned}
\langle [p_n^*, z_{n+1}^*], A^2 p_{n+2} \rangle &= 0, \\
\langle [p_n, z_{n+1}], (A^H)^2 p_{n+2}^* \rangle &= 0,
\end{aligned}
$$

combining with Eqns. (13) and (14) results in the following two 2×2 systems of linear equations for respectively solving g_n and g_n^*

$$
\begin{bmatrix} \langle A^H p_n^*, A p_n \rangle & \langle A^H p_n^*, A z_{n+1} \rangle \\ \langle A^H z_{n+1}^*, A p_n \rangle & \langle A^H z_{n+1}^*, A z_{n+1} \rangle \end{bmatrix} \begin{bmatrix} g_n^{(1)} \\ g_n^{(2)} \end{bmatrix} =
$$
$$
- \begin{bmatrix} \langle A^H p_n^*, A r_{n+2} \rangle \\ \langle A^H z_{n+1}^*, A r_{n+2} \rangle \end{bmatrix}, \quad (15)
$$

$$
\begin{bmatrix} \langle A p_n, A^H p_n^* \rangle & \langle A p_n, A^H z_{n+1}^* \rangle \\ \langle A z_{n+1}, A^H p_n^* \rangle & \langle A z_{n+1}, A^H z_{n+1}^* \rangle \end{bmatrix} \begin{bmatrix} g_n^{*(1)} \\ g_n^{*(2)} \end{bmatrix} =
$$
$$
- \begin{bmatrix} \langle A p_n, A^H r_{n+2}^* \rangle \\ \langle A z_{n+1}, A^H r_{n+2}^* \rangle \end{bmatrix}. \quad (16)
$$

Therefore, it could be able to advance from step n to step $(n + 2)$ to provide x_{n+2}, r_{n+2}, x_{n+2}^*, r_{n+2}^*, p_{n+2}, p_{n+2}^* by solving the above four 2×2 linear systems represented as in Eqns. (11), (12), (15), and (16). With an appropriate combination of 1×1 and 2×2 steps, the CSBiCOR method can be simply obtained with only a minor modification to the usual implementation of the BiCOR method. The pseudocode for the preconditioned CSBiCOR with a left preconditioner B can be represented by Algorithm 3. For full details on the derivation and analysis of the CSBiCOR method, please refer to our recent work [32].

Algorithm 3 *Left preconditioned CSBiCOR method*

1: Compute $r_0 = b - Ax_0$ for some initial guess x_0.
2: Choose $r_0^* = P(A)r_0$ such that $\langle r_0^*, Ar_0 \rangle \neq 0$, where $P(t)$ is a polynomial in t. (For example, $r_0^* = Ar_0$). Set $p_0 = r_0, \tilde{p}_0 = \tilde{r}_0, q_0 = Ap_0, \tilde{q}_0 = A^H \tilde{p}_0$.

3: Compute $\rho_0 = \langle \tilde{r}_0, Ar_0 \rangle$.
4: Begin LOOP $(n = 0, 1, 2, \ldots)$
5:     $\sigma_n = \langle \tilde{q}_n, q_n \rangle$
6:     $s_{n+1} = \sigma_n r_n - \rho_n q_n$
7:     $\tilde{s}_{n+1} = \bar{\sigma}_n \tilde{r}_n - \bar{\rho}_n \tilde{q}_n$
8:     $y_{n+1} = As_{n+1}$
9:     $\tilde{y}_{n+1} = A^H \tilde{s}_{n+1}$
10:    $\theta_{n+1} = \langle \tilde{s}_{n+1}, y_{n+1} \rangle$
11:    $\zeta_{n+1} = \langle \tilde{y}_{n+1}, y_{n+1} \rangle$
12: **if** 1×1 step **then**
13:        $\alpha_n = \rho_n / \sigma_n$
14:        $\rho_{n+1} = \theta_{n+1} / \sigma_n^2$
15:        $\beta_{n+1} = \rho_{n+1} / \rho_n$
16:        $x_{n+1} = x_n + \alpha_n p_n$
17:        $r_{n+1} = r_n - \alpha_n q_n$
18:        $\tilde{r}_{n+1} = \tilde{r}_n - \bar{\alpha}_n \tilde{q}_n$
19:        $p_{n+1} = s_{n+1}/\sigma_n + \beta_{n+1} p_n$
20:        $\tilde{p}_{n+1} = \tilde{s}_{n+1}/\bar{\sigma}_n + \bar{\beta}_{n+1} \tilde{p}_n$
21:        $q_{n+1} = y_{n+1}/\sigma_n + \beta_{n+1} q_n$
22:        $\tilde{q}_{n+1} = \tilde{y}_{n+1}/\bar{\sigma}_n + \bar{\beta}_{n+1} \tilde{q}_n$
23:        $n \leftarrow n + 1$
24: **else**
25:        $\delta_n = \sigma_n \zeta_{n+1} \rho_n^2 - \theta_{n+1}^2$
26:        $\alpha_n = \zeta_{n+1} \rho_n^3 / \delta_n$
27:        $\alpha_{n+1} = \theta_{n+1} \rho_n^2 / \delta_n$
28:        $x_{n+2} = x_n + \alpha_n p_n + \alpha_{n+1} s_{n+1}$
29:        $r_{n+2} = r_n - \alpha_n q_n - \alpha_{n+1} y_{n+1}$
30:        $\tilde{r}_{n+2} = \tilde{r}_n - \bar{\alpha}_n \tilde{q}_n - \bar{\alpha}_{n+1} \tilde{y}_{n+1}$
31:        **solve** $Bz_{n+2} = r_{n+2}$
32:        **solve** $B^H \tilde{z}_{n+2} = \tilde{r}_{n+2}$
33:        $\hat{z}_{n+2} = Az_{n+2}$
34:        $\hat{\tilde{z}}_{n+2} = A^H \tilde{z}_{n+2}$
35:        $\rho_{n+2} = \left\langle \hat{\tilde{z}}_{n+2}, r_{n+2} \right\rangle$
36:        $\beta_{n+1} = \rho_{n+2} / \rho_n$
37:        $\beta_{n+2} = \rho_{n+2} \sigma_n / \theta_{n+1}$
38:        $p_{n+2} = z_{n+2} + \beta_{n+1} p_n + \beta_{n+2} s_{n+1}$
39:        $\tilde{p}_{n+2} = \tilde{z}_{n+2} + \bar{\beta}_{n+1} \tilde{p}_n + \bar{\beta}_{n+2} \tilde{s}_{n+1}$
40:        $q_{n+2} = \hat{z}_{n+2} + \beta_{n+1} q_n + \beta_{n+2} y_{n+1}$
41:        $\tilde{q}_{n+2} = \hat{\tilde{z}}_{n+2} + \bar{\beta}_{n+1} \tilde{q}_n + \bar{\beta}_{n+2} \tilde{y}_{n+1}$
42:        $n \leftarrow n + 2$
43: **end if**
44: Check convergence; continue if necessary
45: End LOOP

This study illustrates the applicability of the CSBiCOR method in electromagnetics to show its improved numerical behaviors in comparison with the BiCOR method. The dense linear systems considered in these experiments arise from RCS calculation of perfectly conducting objects. They are generated by applying the method of moments discretization to the electric field integral equation for surface scattering problems (see e.g. [33]). In all the experiments, we use ten discretization points per wavelength and a physical right-hand for the linear system. We precondition the linear systems using a sparse approximate inverse method based on the minimization of the Frobenius norm. The preconditioner is computed by minimizing the Frobenius-norm of the error matrix

$$\min_{M \in S} \left\| I - M\widehat{A} \right\|_F,$$

where S is the set of matrices with a given sparsity pattern. We construct the approximate inverse M from a sparse approximation $\widehat{A}$ of the dense coefficient matrix A. The sparsity patterns of $\widehat{A}$ and M are computed in advance by selecting a fixed number of the largest entries in each column of A. Details of the preconditioner are found in [11]. The stopping criterion for solving the linear system consists in reducing the initial residual by six orders of magnitude, starting from the zero vector. In the experiments reported in [11], it was shown that this value of the tolerance is sufficient to enable a correct reconstruction of the RCS signal for engineering purposes. These experiments are run in double precision complex arithmetic in Fortran on a PC equipped with an Intel(R) Core(TM)2 Duo CPU P8700 running at 2.53GHz, and with 4 GB of RAM.

The stabilizing and robust effect of the composite step strategy on the BiCOR method can be observed according to the comparative figures presented in Table 4 and the convergence histories depicted in Fig. 1.

IV. CONCLUSIONS

We have presented an investigation of a new interesting variant of the BiCOR method for solving dense complex non-Hermitian systems of

Table 4: Number of iterations and CPU time required by BiCOR and CSBiCOR to reduce the initial residual by six orders of magnitude on some dense linear systems from electromagnetics

Problem	size	CSBiCOR		BiCOR	
		Iter	CPU time	Iter	CPU time
guide	1080	29	0.63	37	0.71
sphere	2430	36	2.97	44	3.46
parallelepipede	2016	32	1.95	41	2.25
cube	1800	53	2.54	67	2.95
paraboloid	1980	39	2.21	49	2.59
satellite	1701	98	4.48	126	5.10

linear equations in electromagnetics. Our approach is naturally based on and inspired by the composite step strategy taken for the CSBCG method [17, 18]. The present CSBiCOR method can be both theoretically and numerically demonstrated to avoid near pivot breakdowns and compute all the well-defined BiCOR iterates stably with only minor modifications with the assumption that the underlying biconjugate A-orthonormalizaion procedure does not break down [32]. Besides reducing the number of spikes in the convergence history of the norm of the residuals to the greatest extent, the CSBiCOR method could provide some further practically desired smoothing behavior towards stabilizing the behavior of the BiCOR method when it has erratic convergence behaviors. Additionally, the CSBiCOR method seems to be superior to the CSBCG method to some extent because of the inherited promising advantages of the empirically observed stability and fast convergence rate of the BiCOR method over the BCG method.

Since the BiCOR method is the most basic variant of the family of Lanczos biconjugate A-orthonormalization methods, its improvement will analogously lead to similar improvements for the CORS and BiCORSTAB methods, which is under investigation.

ACKNOWLEDGMENT

The authors would like to thank Professor Randolph E. Bank for showing us the "PLTMG"software package and for his insightful and beneficial discussions and suggestions.

We gratefully thank the EMC Team at CERFACS in Toulouse for providing us with some test examples used in the numerical experiments.

REFERENCES

[1] W. M. Coughran, Jr. and R. W. Freund, "Recent Advances in Krylov Subspace Solvers for Linear Systems and Applications in Device Simulation," in *Proc. IEEE International Conference on Simulation of Semiconductor Processes and Devices, SISPAD*, 1997, pp. 9–16.

[2] G. Meurant, "Computer Solution of Large Linear Systems," in *Studies in Mathematics and its Applications*, North-Holland: Amsterdam, 1999, vol. 28.

[3] Y. Saad and H. A. van der Vorst, "Iterative Solution of Linear Systems in the 20th Century," *J. Comput. Appl. Math.*, vol. 43, pp. 1155–1174, 2005.

[4] J. Dongarra and F. Sullivan, "Guest Editors' Introduction to the Top 10 Algorithms," *Comput. Sci. Eng.*, vol. 2, no. 1, pp. 22–23, 2000.

[5] V. Simoncini and D. B. Szyld, "Recent Computational Developments in Krylov Subspace Methods for Linear Systems," *Numer. Lin. Alg. Appl.*, vol. 14, pp. 1–59, 2007.

[6] T. Sogabe, M. Sugihara, and S.-L. Zhang, "An Extension of the Conjugate Residual Method to Nonsymmetric Linear Systems," *J. Comput. Appl. Math.*, 226 (2009), pp. 103–113.

[7] Y.-F. Jing, T.-Z. Huang, Y. Zhang, L. Li, G.-H Cheng, Z.-G. Ren, Y. Duan, T. Sogabe, and B. Carpentieri, "Lanczos-Type Variants of the COCR Method for Complex Nonsymmetric Linear Systems," *J. Comput. Phys.*, vol. 228, pp. 6376–6394, 2009.

[8] B. Philippe and L. Reichel, "On the Generation of Krylov Subspace Bases," *Appl. Numer. Math.*, In Press, Corrected Proof, Available online 7 January 2011.

[9] A. Greenbaum, *Iterative Methods for Solving Linear Systems*. Philadelphia: SIAM, 1997.

[10] Y. Saad, *Iterative Methods for Sparse Linear Systems* (2nd edn). Philadelphia: SIAM, 2003.

[11] B. Carpentieri, I. Duff, L. Giraud, and G. Sylvand, "Combining Fast Multipole Techniques and an Approximate Inverse Preconditioner for Large Electromagnetism Calculations," *SIAM J. Scientific Computing*, vol. 27, no. 3, pp. 774–792, 2005.

(a)

(b)

(c)

(d)
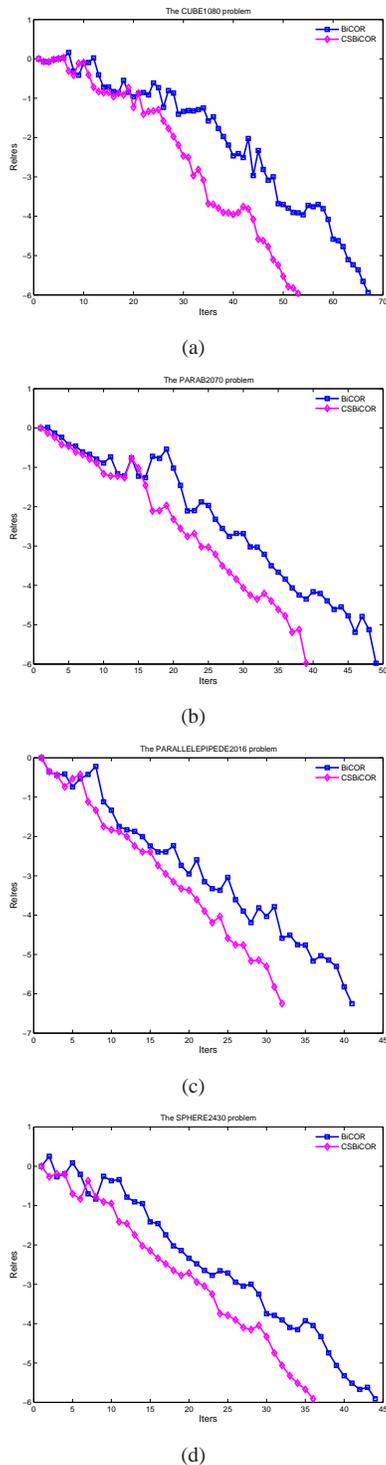
Fig. 1. Comparative experiments between the BiCOR and CSBiCOR methods on dense linear systems from electromagnetics

[12] Y.-F. Jing, B. Carpentieri, and T.-Z. Huang, "Experiments with Lanczos Biconjugate A-Orthonormalization Methods for MoM Discretizations of Maxwell's Equations," *Prog. Electromagn. Res.*, vol. 99, pp. 427–451, 2009.

[13] B. Carpentieri, Y.-F. Jing, and T.-Z. Huang, "Lanczos Biconjugate A-Orthonormalization Methods for Surface Integral Equations in Electromagnetism," in *Proc. the Progress In Electromagnetics Research Symposium*, 2010, pp. 678–682.

[14] Y.-F. Jing, T.-Z. Huang, Y. Duan, and B. Carpentieri. "A Comparative Study of Iterative Solutions to Linear Systems Arising in Quantum Mechanics," *J. Comput. Phys.*, vol. 229, pp. 8511–8520, 2010.

[15] B. Carpentieri, Y.-F. Jing, and T.-Z. Huang, "The BiCOR and CORS Algorithms for Solving Nonsymmetric Linear Systems," *SIAM J. Scientific Computing*, vol. 33, no. 5, pp. 3020–3036, 2011.

[16] B. Carpentieri, Y.-F. Jing, T.-Z. Huang, W.-C. Pi, and X.-Q. Sheng. "A Novel Family of Iterative Solvers for Method of Moments Discretizations of Maxwell's Equations," *the CEM'11 International Workshop on Computational Electromagnetics*, Izmir, Turkey, August 2011.

[17] R. E. Bank and T. F. Chan, "An Analysis of the Composite Step Biconjugate Gradient Method," *Numer. Math.*, vol. 66, pp. 295–319, 1993.

[18] R. E. Bank and T. F. Chan, "A Composite Step Biconjugate Gradient Algorithm for Nonsymmetric Linear Systems," *Numer. Algo.*, vol. 7, pp. 1–16, 1994.

[19] B. Parlett, D. Taylor, and Z.-S. Liu, "A Look-Ahead Lanczos Algorithm for Unsymmetric Matrices," *Math. Comp.*, vol. 44, pp. 105–124, 1985.

[20] B. Parlett, "Reduction to Tridiagonal Form and Minimal Realizations," *SIAM J. Matrix Anal. Appl.*, vol. 13, pp. 567–593, 1992.

[21] C. Brezinski, M. R. Zaglia, and H. Sadok, "Avoiding Breakdown and Near-Breakdown in Lanczos Type Algorithms," *Numer. Algo.*, vol. 1, pp. 261–284, 1991.

[22] C. Brezinski, M. R. Zaglia, and H. Sadok, "A Breakdown-Free Lanczos Type Algorithm for Solving Linear Systems," *Numer. Math.*, vol. 63, pp. 29–38, 1992.

[23] C. Brezinski, M. R. Zaglia, and H. Sadok, "Breakdowns in the Implementation of the Lanczos Method for Solving Linear Systems," *Comput. Math. Appl.*, vol. 33, pp. 31–44, 1997.

[24] C. Brezinski, M. R. Zaglia, and H. Sadok, "New

Look-Ahead Lanczos-Type Algorithms for Linear Systems," *Numer. Math.*, vol. 83, pp. 53–85, 1999.

[25] N. M. Nachtigal, *A Look-Ahead Variant of the Lanczos Algorithm and its Application to the Quasi-Minimal Residual Method for Non-Hermitian Linear Systems*. Ph.D. Thesis, Massachussets Institute of Technology, Cambridge, MA, 1991.

[26] R. Freund, M. H. Gutknecht, and N. Nachtigal, "An Implementation of the Look-Ahead Lanczos Algorithm for Non-Hermitian Matrices," *SIAM J. Sci. Stat. Comput.*, vol. 14, pp. 137–158, 1993.

[27] M. H. Gutknecht, "Lanczos-Type Solvers for Nonsymmetric Linear Systems of Equations," *Acta Numer.*, vol. 6, pp. 271–397, 1997.

[28] E. Darve, "The Fast Multipole Method (i) : Error Analysis and Asymptotic Complexity, " *SIAM J. Numerical Analysis*, vol. 38,no. 1, pp. 98–128, 2000.

[29] B. Dembart and M. A. Epton, "A 3D Fast Multipole Method for Electromagnetics with Multiple Levels," *Tech. Rep. ISSTECH-97-004*, The Boeing Company, Seattle, WA, 1994.

[30] L. Greengard and V. Rokhlin, "A Fast Algorithm for Particle Simulations," *Journal of Computational Physics*, vol. 73, pp. 325–348, 1987.

[31] J. M. Song, C.-C. Lu, and W. C. Chew, "Multilevel Fast Multipole Algorithm for Electromagnetic Scattering by Large Complex Objects," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 10, pp. 1488–1493, 1997.

[32] Y.-F. Jing, T.-Z. Huang, B. Carpentieri, and Y. Duan, "Exploiting the Composite Step Strategy to the Biconjugate A-orthogonal Residual Method for Non-Hermitian Linear Systems," *SIAM J. Numerical Analysis*, 2011, submitted.

[33] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill Book Company, 1961.

[34] T. F. Chan and T. Szeto, "A Composite Step Biconjugate Gradients Squared Algorithm for Solving Nonsymmetric Linear Systems," *Numer. Algo.*, vol. 7, pp. 17–32, 1994.

[35] T. F. Chan and T. Szeto, "Composite Step Product Methods for Solving Nonsymmetric Linear Systems," *SIAM J. Sci. Comput.*, vol. 17, no. 6, pp. 1491–1508, 1996.

[36] S. He, C. Li, F. Zhang, G. Zhu, W. Hu, and W. Yu, "An Improved MM-PO Method with UV Technique for Scattering from an Electrically Large Ship on a Rough Sea Surface at Low Grazing Angle," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 2, pp. 87–95, 2011.

[37] Z. Jiang, Z. Fan, D. Ding, R. Chen, and K. Leung, "Preconditioned MDA-SVD-MLFMA for Analysis of Multi-Scale Problems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 11, pp. 914–925, 2010.

[38] M. Li, M. Chen, W. Zhuang, Z. Fan, and R. Chen, "Parallel SAI Preconditioned Adaptive Integral Method For Analysis of Large Planar Microstrip Antennas," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 11, pp. 926–935, 2010.

[39] Y. Zhang and Q. Sun, "Complex Incomplete Cholesky Factorization Preconditioned Bi-conjugate Gradient Method," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 9, pp. 750–754, 2010.

[40] D. Ding, J. Ge, and R. Chen, "Well-Conditioned CFIE for Scattering from Dielectric Coated Conducting Bodies above a Half-Space," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 11, pp. 936–946, 2010.

[41] B. Carpentieri, "An Adaptive Approximate Inverse-Based Preconditioner Combined with the Fast Multipole Method for Solving Dense Linear Systems in Electromagnetic Scattering," *Applied Computational Electromagnetics Society Journal (ACES)*, vol. 24, no. 5, pp. 504–510, 2009.

Yan-Fei Jing received the B.S. and Ph.D. degrees in Applied Mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2005 and 2011, respectively. He is currently an Associated Professor in the School of Mathematical Sciences, UESTC. He is the author or coauthor of more than 18 research papers and is currently a member of International Linear Algebra Society (ILAS). His research interests include iterative methods of linear systems and preconditioning techniques with applications in computational electromagnetics.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in Computational Mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992 and 2000, respectively. During 2005, he was a visiting scholar in Dept. of Computer Science, Loughborough University of UK.

He is currently a professor in the School of Mathematical Sciences, UESTC. He is currently an editor of several journals such as Advances in Numer. Anal., J. Pure and Appl. Math.: Advances and Appl, Computer Science and Appl., J. Electro. Sci. and Tech. of China, Advances in Pure Math., etc. He is the author or coauthor of more than 100 research papers. His current research interests include numerical linear algebra with applications in computational electromagnetics and image science, and matrix analysis.

Yong Duan received the D.Sc. degree in applied mathematics from Fudan University, Shanghai, China, in 2005. His research interests include applied partial differential equations and numerical methods for partial differential equations as well as computational electromagnetics.

Bruno Carpentieri received a Laurea degree in Applied Mathematics from Bari University, Italy, in 1998, and a Ph.D. degree in Computer Science from the Institut National Polytechnique of Toulouse, France, in 2002. After completing his Ph.D., he was as post-doctoral fellows at the CERFACS Institute in Toulouse, France (2003-2004), and at Karl-Franzens University in Graz, Austria (2005-2008). He also served as a consultant on European projects at CRS4 in Sardinia, Italy (2008-2009). Since January 2010, he is an University Assistant at the Institute of Mathematics and Computing Science of the University of Groningen, The Netherlands. His research interests include numerical linear algebra, parallel computing, Computational Fluid Dynamics, Computational Electromagnetics, Cardiac Modelling.

Advanced Techniques for Efficient Modeling of Electrically Large Structures on Desktop PCs

Branko M. Kolundzija [1], Miodrag S. Tasic [1], Dragan I. Olcan [1]
Dusan P. Zoric [2], and Srdjan M. Stevanetic [2]

[1] School of Electrical Engineering
University of Belgrade, King Alexander's Blvd 73, 11120 Belgrade, Serbia
kol@etf.rs, tasic@etf.rs, olcan@etf.rs

[2] WIPL-D d.o.o.
Gandijeva 7/32, 11073 Belgrade, Serbia
dusan.zoric@wipl-d.com, srdjan.stevanetic@wipl-d.com

Abstract — The method of moments (MoM), applied to surface integral equations (SIEs) in the frequency domain, enables very accurate analysis of composite metallic and dielectric structures. Particularly, on a desktop PC, the electrical size of solvable problems is very limited by memory and time resources. This limit can be significantly extended by using advanced techniques, which are shortly revealed. The focus of the paper is to present results that illustrate current possibilities of MoM/SIEs solution on a desktop PC: (1) monostatic RCS of a cube of side 80 λ, (2) beam steering of an array of 30 by 30 microstrip patch antennas at 9.2 GHz, and (3) beam steering of 4 by 4 patch antennas at 5 GHz, placed on a 19 m long helicopter.

Index Terms — Higher order basis functions, method of moments, surface integral equations.

I. INTRODUCTION

Electromagnetic modelling of composite metallic and dielectric structures in the frequency domain can be performed very accurately by solving surface integral equations (SIEs) using the method of moments (MoM). Using the MoM theory [1, 2], induced currents over metallic surfaces and equivalent currents over material boundary surfaces are approximated by a series of known basis functions multiplied by unknown coefficients. The SIE is transformed into a system of linear equations, which is solved for the unknown coefficients. However, the size of the solvable problem in terms of the number of unknowns, N, is limited by memory and time resources of the computer used for the simulation. By increasing N, the memory occupation and matrix-fill time increase as N^2 and the matrix-solution time (in case of direct methods as Gaussian elimination or LU decomposition) increases as N^3. Typical size of operative memory (RAM) of modern PCs used for number crunching is 8 GB. In that case, systems of linear equations in the complex domain of up to about $N_{max} = 30,000$ unknowns can be solved "incore", which means keeping the whole matrix of the system in RAM during solving. On the other hand, the electrical size of the solvable problem (in λ^2 of surface area) is dependent on the choice of basis functions. For the Rao-Wilton-Glisson (RWG) basis functions defined over triangles, typical edge length of triangles is $\lambda/10$, resulting in about 300 unknowns per λ^2 for metallic surfaces [3]. Particularly, for $N_{max} = 30,000$, the electrical size of the surface area of a structure is limited to about $100\ \lambda^2$.

Many techniques have been developed to increase the electrical size of the solvable structure within the limits of PC computer resources. In what follows, we shall focus on techniques implemented in the commercial software package WIPL-D Pro v9.0 [4]. Generally, these techniques can be grouped into three classes.

In the first class, there are techniques that decrease the number of unknowns:

    (a) application of higher order basis functions [5, 6],

    (b) exploiting symmetry of the problem [4],

    (c) "smart reduction" of expansion order [7],

    (d) construction of macro-basis functions by physical optics (PO) driven MoM [8], and

    (e) hybrid MoM-PO methods [9].

In the second class, there are techniques that decrease the memory resources and matrix-fill/solution time for given number of unknowns:

    (a) iterative techniques [10-13],

    (b) fast multipole method (FMM), and

    (c) multilevel fast multipole algorithm (MLFMA) [14-18].

Finally, in the third class, there are techniques that enable efficient usage of modern hardware resources:

    (a) out-of-core solution of matrix equation [19],

    (b) parallelization on CPU based on OpenMP [20], and

    (c) parallelization on GPU based on CUDA [21-23].

The goals of the paper are: (1) to reveal various techniques for increasing the electrical size of the solvable structure within the limits of PC computer resources, (2) to compare these techniques and discuss optimal usage of these techniques, and (3) to show numerical results for some typical electrically large structures.

II. TECHNIQUES THAT REDUCE NUMBER OF UNKNOWNS

A. Higher-order basis functions

The basic way to decrease the number of unknowns is to apply higher-order basis functions (HOBFs). For interpolatory HOBFs defined over triangles [6], the maximum edge length of triangles can be extended to 0.5~1λ, resulting in 40–70 unknowns per λ^2 [15]. In case of polynomial HOBFs defined over quadrilaterals [2], [5], the maximum edge can be extended to 1~2λ, resulting in 20–35 unknowns per λ^2 [5]. In both cases, the expansion orders are chosen according to the electrical size of patches. Thus, the electrical size of a solvable structure within the limit of computer resources is increased by an order of magnitude when compared with RWG basis functions.

B. Exploitation of geometrical symmetry of the problem

If the geometry of a structure is symmetrical with respect to one plane, and the excitation is either symmetrical or anti-symmetrical with respect to this plane, the unknown coefficients on one side of the plane are equal to the coefficients from the other side of the plane multiplied by ±1, so the original number of unknowns is halved. The memory requirements are decreased four times and the matrix solution time is decreased eight times. In some cases, there are two or even three mutually orthogonal planes of symmetry, so that the number of unknowns can be decreased four or eight times, respectively.

The symmetry of geometry can be exploited to decrease the number of unknowns even if the excitation is not (anti) symmetrical [4]. In that case, the excitation can be decomposed into a set of symmetrical and anti-symmetrical excitations. Consequently, the problem is decomposed into set of sub-problems, for which not only the geometry is symmetrical, but also the excitations are symmetrical and/or anti-symmetrical. In that case, the number of unknowns is halved and the number of sub-problems to be solved is doubled for each symmetry plane. Once all sub-problems are solved, the final results are obtained by superposition. For each symmetry plane, the memory requirements are reduced four times, the matrix fill time is unchanged, while the matrix solution time is shortened by a factor of four.

C. Smart reduction of expansion order

In the case of an antenna placed at some platform (e.g., airplane fuselage), the currents, which are induced over the fuselage, are of the largest magnitude in the vicinity of the antenna, and decrease going away from the antenna. Since the currents of the lower magnitude have a smaller impact on global quantities, such as antenna input impedance or gain, their distribution can be determined with lower accuracy. With this in mind, the expansion orders of currents can be linearly reduced going from the antenna to the most distant part of the fuselage, for which the maximum reduction is specified (e.g., in %). For a maximum reduction of 100%, the expansion order is reduced to order one along each side of the patch, i.e., approximation of currents over relatively large

patches (up to 2 by 2 λ) is performed using rooftop basis functions.

In the case of two antennas placed at some platform, for mutual coupling between them, the most important currents are in the 1st Fresnel region. In that sense, the space around two antennas can be subdivided into the 1st and higher-order Fresnel regions. Then expansion orders for currents can be reduced gradually, from the 1st Fresnel region to the highest Fresnel region that still contains parts of the fuselage.

Finally, we know that, in the "shadow region" of a fuselage (i.e., in the region without optical visibility from the antennas), the induced currents are also much smaller than those in the lit region. In particular, if these currents are farther away from the border of the lit region, we consider them to be deeper in the shadow. In that sense, different levels of reduction in expansion orders can be specified for a given depth of the shadow.

Generally, by increasing the level of reduction of expansion orders, the number of unknowns is decreased and the solution error is decreased. It is shown that, by proper choice of reduction techniques and levels, the number of unknowns can be significantly reduced with negligible loss of accuracy.

As an example in [7], a half-wavelength dipole at 2 GHz is placed above a payload fairing of length 8.9 m and largest diameter 2.9 m. Basically, with one symmetry plane applied, the model requires 44,614 unknowns. By combining the reduction techniques explained above the number of unknowns can be reduced almost ten times, down to 4,983 unknowns.

D. PO driven MoM

The main goal of the PO driven MoM method is to solve electrically large problems using a small numbers of unknowns, and thus, reduce memory and time resources [8]. In case of scatterers, the method starts from the PO solution and improves it iteratively. In each iteration, the structure is excited by the solution from the previous iteration and correctional PO currents are determined and grouped into a small number of macro-basis functions (MBFs). Unknown coefficients multiplying all macro basis functions are obtained by minimizing the residuum of the original MoM solution. Thus, the solution after the 1st iteration is

better than the PO solution, and in each next iteration, it approaches the original MoM solution. The number of MBFs added per iteration is adopted to be comparable with the square root of the number of unknowns. An acceptable solution is obtained even after the 1st iteration with a reduction in the number of unknowns of two orders of magnitude.

In the case of an antenna placement problem, the only difference is that the starting solution is obtained as the MoM solution of the antenna isolated from the fuselage.

III. TECHNIQUES THAT REDUCE MEMORY REQUIREMENTS AND NUMBER OF OPERATIONS

A. Iterative methods

The number of operations needed to solve a matrix equation can significantly be decreased by using iterative methods, generally speaking from $N^3/3$ to MN^2, where M is number of iterations. The number of iterations generally depends on: (a) the maximum allowed mean square value of residuum of matrix equation (e.g., $R = 0.001$), (b) the type of problem to be solved, (c) the type of MoM/SIE method used to obtain the solution, and (d) the type of iterative procedure itself [10]. For example, scattering problems (distributed excitation) require fewer iterations than antenna problems (localised excitation), which require fewer iterations than closed problems (e.g., resonant cavities). The convergence for different types of problems can be improved by using various preconditioners, or other techniques that decrease the matrix condition number [11].

In particular, HOBFs cannot be efficiently used for iterative solutions if they are not orthogonalized in some way [12, 13]. The best results are obtained by using maximally orthogonalized HOBFs [13], which enable almost the same convergence of the matrix solution for higher-order basis functions as for the rooftop basis functions.

Fully developed iterative techniques can solve the matrix equation in a relatively small number of iterations, which is much smaller than the number of unknowns, i.e., $M \ll N$, and thus significantly reduce the time needed for matrix solution. However, in many cases, the iterative solution cannot outperform the direct solver (e.g., LU

decomposition), not only because of slow convergence. Namely, in case of a set of independent excitations (e.g., multiple excitation), the iterative procedure is performed from the beginning for each excitation. On the other hand, in case of a direct solver, once the LU decomposition is performed, the solution for each excitation is obtained by the so-called forward and backward substitution in N^2 operations, which corresponds to a single iteration in an iterative procedure.

B. Multilevel fast multipole algorithm

Both direct solvers and iterative procedures for the solution of matrix equations are limited by the memory required to store the matrix. In case of iterative solvers, the memory requirements, as well as solution time, can significantly be reduced by using FMM and MLFMA [14-16]. In both cases, the method is based on the acceleration of matrix-vector multiply operations performed in each iteration of an iterative solver, by taking into account interactions between widely separated groups of basis functions instead of interactions between individual basis functions belonging to these groups. (Particularly, in case of MLFMA the grouping is performed in few levels.) Reduction of interaction between individual basis functions to interaction between their widely separated groups is enabled by the multipole expansion of the free-space Greens function. The most efficient reduction is obtained for groups, which are far enough away so that the multipole expansion can be represented by a single term, which corresponds to a far-field approximation (FFA).

Generally, the application of FMM and MLFMA introduces an additional error in the MoM solution. This error can be decreased (e.g., by increasing the number of terms in the multipole expansion, or by increasing the relative distance at which FFA is applied), which results in an increase of memory resources and simulation time. Particularly, the error issues are critical in the case of application of HOBFs [15, 17]. In order to enable the efficient application of HOBFs, an improved far-field approximation is proposed in [18]. Even with this improvement, efficient MLFMA solutions are only possible for maximally 2nd and 3rd orders of HOBFs.

As an example in [24], the fighter scatterer, 12 m long, is analyzed at 4 GHz (160 λ). The problem requires 307,170 unknowns and 754 GB of RAM for storage of the full matrix. The MLFMA solution is obtained in 3.2 hours using a standard quad core PC with 8 GB of RAM. In this case, the application of MLFMA reduces the memory needs to 7.2 GB.

IV. TECHNIQUES THAT ENABLE EFFICIENT USAGE OF HARDWARE RESOURCES

A. Out-of-core solver

In the case when the MoM matrix is too large to be stored in RAM, it must be split into blocks, which are stored on a hard disk. In particular, if there are many excitation columns (multiple excitation) they can be also stored in the disk. The solution of such stored matrix equation is obtained using an appropriate out-of-core solver.

In the case of out-of-core direct solution, Gaussian elimination or LU decomposition is performed using two by two blocks from the disk. The size of the blocks is limited so that two of them can be stored in RAM. Since the total number of readings of the full matrix is equal to the half of the total number of blocks, it is optimal that these two blocks have size slightly smaller than the size of RAM. As an example, consider a problem of 120,000 unknowns that should be solved using 8 GB of RAM. In that case, the matrix size is about 107.3 GB and each block can have a size of 3.8 GB, so that for the full solution the matrix is read 15 times.

In the case of an out-of-core iterative solution, the matrix must be read at least once during each iteration. The number of iterations needed to obtain a sufficiently accurate solution is usually much larger than the number of blocks into which the matrix is split. Thus, the total time needed for an out-of-core iterative solution is much larger than for a direct solution. Hence, the out-of-core solver is used only for direct solution.

B. Parallelization on CPU based on OpenMP

Modern CPUs contain more than one core, usually 4 cores. Using OpenMP, the calculation can be performed in parallel at an arbitrary number of threads, but an efficient parallelization uses a number of threads approximately equal to the number of cores. Since the memory used for calculation is shared by all threads, for an efficient parallelization, it is very important not to access the same variable with more than one thread. Generally,

the code that was optimal for serial performance should be reorganized to support efficient parallel performance. For example, the efficient matrix fill in serial mode is organized so that mutual coupling between two patches is calculated one at a time. However, the basis functions that provide continuity between patches (doublets in general case, roof top basis functions in special case) belong to both patches. So, it can happen that if such couplings are calculated in parallel two or more threads approach to the same element in the matrix. This can be avoided if the couplings are calculated for groups of test and basis function patches, such that all patches in a group have no common basis functions. Using such grouping, the efficiency of parallelization of 90% (70%) is achieved for matrix fill at 4 (8) cores.

In the case of the matrix solution, the MKL [25] library enables almost 100% efficiency for 4 (8) cores.

C. Parallelization on GPU based on CUDA

In cases when the calculation consists of many repetitions of the same operation, graphical processing units (GPU) can perform the calculation up to 10 times faster than a quad-core CPU. However, the code that is executed in parallel mode using OpenMP on CPU cannot be simply transferred into CUDA on the GPU. Again, the algorithms should be reorganized to support parallel execution on thousands of threads. In particular, it is desirable that tasks performed in parallel at all threads use the same sequence of operations the same number of times. Since not all parts of the MoM code are suitable for GPU parallelization, the optimal code is one that combines parallelization on the CPU and the GPU.

For example, in the case of an out-of-core solver on a CPU, there are two critical operations: (1) storing/reading the matrix to/from disk, and (2) calculations performed in LU decomposition. So, for significant acceleration, it is not enough to parallelize the calculations on the GPU. It is also necessary to perform storing and reading of the matrix blocks in parallel on the CPU and in parallel with the calculations on the GPU [22].

V. OPTIMAL COMBINATION OF TECHNIQUES FOR EFFICIENT AND ACCURATE EM MODELING

The basic combination for efficient and accurate modelling on PCs is that based on higher-order basis functions, incore/out-of-core direct solvers and combined CPU/GPU parallelization. In particular, the simulation can be accelerated by exploiting geometrical symmetry and by using the "smart reduction" for antenna placement problems. In the case of electrically large structures with a single excitation (e.g., bistatic RCS for scatterers or radiation pattern for antenna placement), the simulation can be further accelerated by using iterative methods, e.g., MLFMA or PO driven MoM.

VI. NUMERICAL RESULTS

Figure 1 shows the monostatic RCS of a cube, of side 80λ, illuminated by a vertically polarized plane wave in the *xOy*-plane from 3,552 directions. Using expansions of 5th order and one symmetry plane, as shown in the left inset, the problem is reduced to 655,380 unknowns. The results are obtained by using all three symmetry planes, as shown in right inset. The problem is decomposed into four sub-problems, each containing 163,845 unknowns. Total simulation time for all 4 sub-problems and post-processing is 33 hours.

The second problem considered is a phased array consisting of *n* by *n* probe-fed microstrip patch antennas, where $n = 10, 20, 30$, and 40. (The array for $n = 10$ is shown in Fig. 2.) The task is to determine the gain for a set of directions of the main beam, with angle θ going from 20 to 90 degrees and angle φ going from 0 to 90 degrees, both with a step of 5 degrees, so that the total number of excitations is 285.

It is also required that the finite size and finite thickness of the substrate, as well as the finite thickness of the metallization, are taken into account, as shown in Fig. 3. The size of a square metallic patch is 10 mm by 10 mm, while the distance between antenna centers is 15 mm. The thickness and relative permittivity of the teflon substrate are 0.5 mm and 2.1, respectively. The thickness of the metallization is 34 μm. The operating frequency is 9.2 GHz.

The original problem for $n = 30$ has about 450,000 unknowns. Two symmetry planes are used

to facilitate the analysis, so that the original problem is decomposed into four sub-problems, each requiring about 112,500 unknowns. The simulation is performed using one generator at time, so that for each generator turned on, all others are turned off. For each such excitation, the radiation pattern of the array is determined. In addition, all these excitations are used to determine the matrices of mutual admittances and impedances.
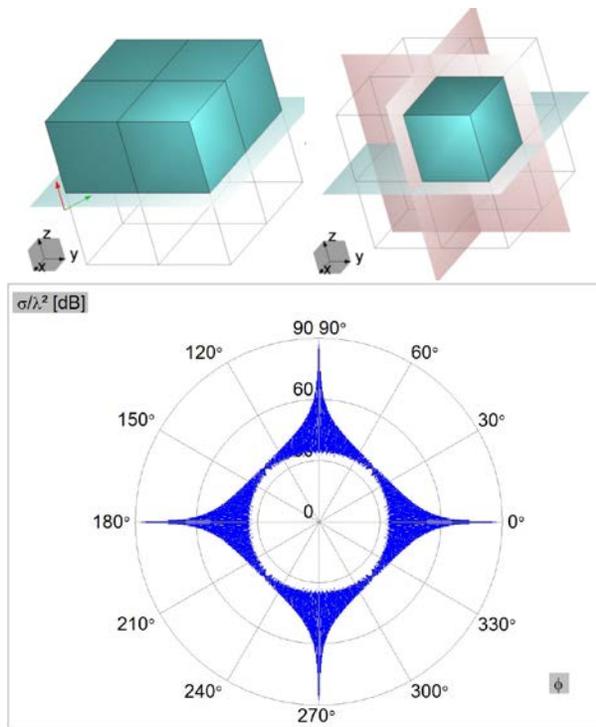
Fig. 1. Monostatic RCS of a cube of side $80\,\lambda$ in *xOy*-plane versus angle φ.

Once the 3D EM simulation is performed, data are imported as an *n* by *n*-port device into the schematic feature of WIPL-D [26]. The current sources are attached to these ports. Once this simple circuit is solved for voltages at the ports, the total radiation pattern is easily obtained by superimposing radiation patterns due to each of these voltages. By properly adjusting the current sources, the main beam of the array can be positioned in a desired direction.
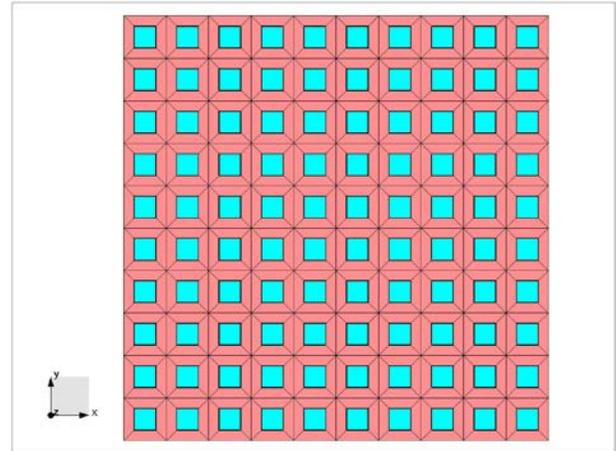
Fig. 2. Geometrical model of an array of 10 by 10 microstrip patch antennas.
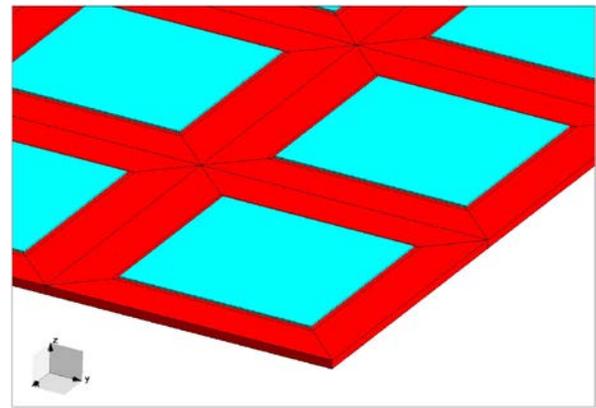
Fig. 3. Geometrical model of an array of 10 by 10 microstrip patch antennas.

Figure 4 shows the antenna gain for various beam directions in the plane φ = 45°, which is 16 beam directions in total. Total simulation time including post-processing for 285 beam directions is 8 hours and 20 minutes.

The last problem considered is the radiation of a 4 by 4 microstrip patch antenna array placed at the bottom of a helicopter fuselage, as shown in Fig. 5. The array is adjusted to operate at a frequency of 5 GHz. Total length of helicopter is 19 m, so that its electrical length at this frequency is 316.7 λ. The original problem having 1,254,034 unknowns is solved using a GPU accelerated PO driven MoM method in 28.25 hours. The radiation pattern is shown in Fig. 6.
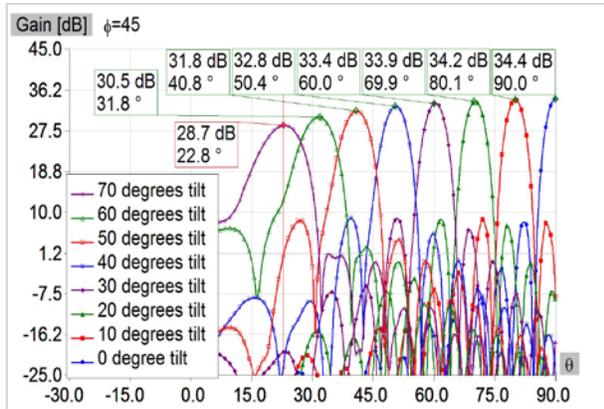
Fig. 4. Gain of array of 30 by 30 microstrip patch antennas for various directions of main beam in plane φ = 45°.
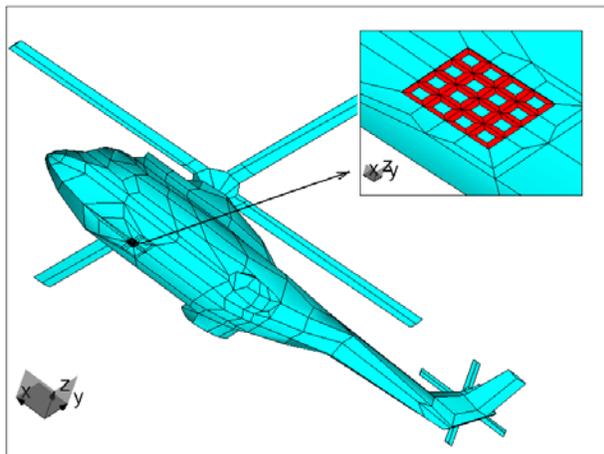
Fig. 5. Geometrical model of an array of 4 by 4 microstrip patch antennas placed on helicopter.
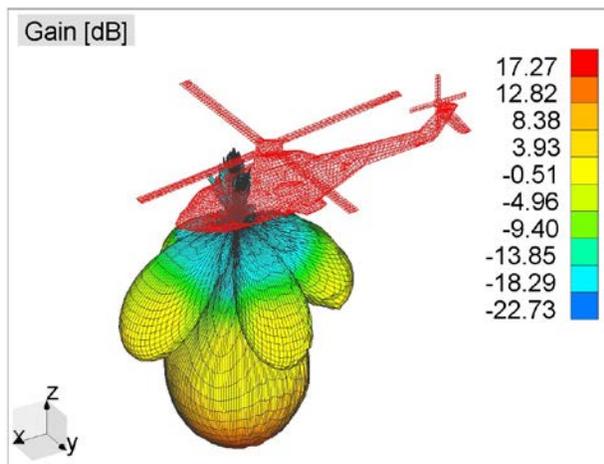
Fig. 6. Radiation pattern at 5 GHz of array of 4 by 4 microstrip patch antennas placed at helicopter.

VII. CONCLUSION

The basic combination for efficient and accurate EM modelling in the frequency domain on PCs is based on higher-order basis functions, incore/out-of-core direct solvers, and combined CPU/GPU parallelization. Where possible, the simulation can be accelerated by exploiting geometrical symmetry and the "smart reduction" of expansion orders. In particular, for single excitation problems (e.g., bistatic RCS and antenna placement), the analysis can be further accelerated using MLFMA and PO driven MoM.

REFERENCES

[1] R. F. Harrington, *Field Computations by Moment Methods.* New York: MacMillan, 1968.

[2] B. M. Kolundzija and A. R. Djordjevic, *Electromagnetic Modelling of Composite Metallic and Dielectric Structures*, Norwood, USA: Artech House, 2002.

[3] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shape," *IEEE Trans. Antennas Propagat.,* vol. AP-30, pp. 409-418, May 1982.

[4] WIPL-D Pro v9.0, 3D EM solver, http://www.wipl-d.com, 2011.

[5] B. M. Kolundžija and B. D. Popović, "Entire-Domain Galerkin Method for Analysis of Metallic Antennas and Scatterers," *IEE Proceedings-H*, vol. 140, no. 1, pp. 1-10, Feb. 1993.

[6] R. D. Graglia, D. R. Wilton, and A. F. Peterson, "Higher Order Interpolatory Vector Bases for Computational Electromagnetics," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 329-342, Mar. 1997.

[7] B. M. Kolundzija and D. S. Sumic, "Adaptive Higher Order Modeling of Antennas Placed on Large Platforms," *Proc of EuCAP,* ESA SP-626, Nice, France, Nov. 2006.

[8] M. S. Tasić and B. M. Kolundžija, "Efficient Analysis of Large Scatterers by Physical Optics Driven Method of Moments," *IEEE Trans. Antennas and Propagat.,* vol. AP-59, no. 8, Aug. 2011 (to be published).

[9] S. Y. He, C. Li, F. Zhang, G. Q. Zhu, W. D. Hu, and W. X. Yu, "An Improved MM-PO Method with UV Technique for Scattering from an Electrically Large Ship on a Rough Sea Surface at Low Grazing Angle," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 2, pp. 87 – 95, February 2011.

[10] B. M. Kolundzija and T.K. Sarkar, "Iterative Solvers in Frequency Analysis of Complex Structures Based on MoM Solution of Surface

Integral Equations," *Proc. IEEE AP-S Int. Symp.,* vol. 2, pp. 588-591, July 2001.

[11] B. M. Kolundzija and D. S. Sumic, "Optimal Weights of Basis Functions for Efficient Iterative Solution of surface Integral Equations," *Proc. IEEE AP-S Int. Symp., CD ROM Edition: s116p04a.pdf*, Washington, July 2005.

[12] E. Jorgensen, J. L. Volakis, P. Meincke, and O. Breinbjerg, "Higher-Order Hierarchical Basis Functions for Electromagnetic Modeling," *IEEE Trans. Antennas and Propagat.,* vol. 52, no. 11, pp. 2985-2995, Nov. 2004.

[13] D. S. Sumic and B. M. Kolundzija, "Efficient Iterative Solution of Surface Integral Equations Based on Maximally Orthogonalized Higher Order Basis Functions," *Proc. IEEE AP-S Int. Symp., CD ROM Edition: s116p05a.pdf*, Washington, July 2005.

[14] J. M. Song, C. C. Lu, and W. C. Chew, "Multilevel Fast Multipole Algorithm for Electromagnetic Scattering by Large Complex Objects," *IEEE Trans. Antennas Propagat.,* vol. 45, pp. 1488–1493, Oct. 1997.

[15] K. Donnepudi, J. M. Song, J. M. Jin, G. Kang, and W. C. Chew: "A Novel Implementation of Multilevel Fast Multipole Algorithm for Higher Order Galerkin's Method," *IEEE Trans. Antennas and Propagat.,* vol. AP-48, pp. 1192-1197, Aug. 2000.

[16] H. Fangjing, N. Zaiping, and H. Jun, "An Efficient Parallel Multilevel Fast Multipole Algorithm for Large-scale Scattering Problems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 4, pp. 381 – 387, April 2010.

[17] B. M. Kolundzija and D. S. Sumic, "Multilevel Fast Multipole Method for Higher Order Basis Functions Implemented in WIPL-D Pro, "*Proc of EuCAP,* pp. 2136-2140, Berlin, Geramny, Nov. 2009.

[18] B. M. Kolundzija and D. S. Sumic, "Improvement of Far-Field Approximation in EM Modeling of Electrically Large Structures," *26th Annual Review of Progress in Applied Electromagnetics (ACES),* Tampere, Finland, pp. 178-182, April 2010.

[19] M. Yuan, T. K. Sarkar, and B. M. Kolundzija, "Solution of Large Complex Problems in Computational Electromagnetic using Higher-Order Basis in MoM with Out-of-Core Solvers," *IEEE Antennas and Propagat. Magazine*, vol. 48, no. 2, pp. 55-62, 2006.

[20] http://openmp.org

[21] http://www.nvidia.com

[22] D. P. Zoric, D. I. Olcan, and B. M. Kolundzija, "Solving Electrically Large EM Problems by using Out-of-Core Solver Accelerated with Multiple Graphic Processing Units," *Proc. IEEE AP-S Int. Symp., CD ROM Edition: 1751.pdf*, Spokane, WA, July 2011.

[23] O. Kilic, M. Huang, C. Conner, and M. S. Mirotznik, "Hardware Accelerated Design of Millimeter Wave Antireflective Surfaces: A Comparison of Field-Programmable Gate Array (FPGA) and Graphics Processing Unit (GPU) Implementations," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 3, pp. 188 – 198, March 2011.

[24] D. S. Sumic and B. M. Kolundzija, "Extending the Reach of WIPL-D Pro on PCs – Revisited 5 years Later," *26th Annual Review of Progress in Applied Electromagnetics (ACES)*, Tampere, Finland, pp. 667-672, April 2010.

[25] http://software.intel.com

[26] WIPL-D Microwave, v2.3, Circuit solver, http://www.wipl-d.com, 2011.

Branko Kolundzija is a full professor at the Dept. of EE at the University of Belgrade, where he received his B.Sc. in 1981, M.Sc. in 1987, and D.Sc. in 1990. He is the head developer of WIPL-D software and founder of WIPL-D d.o.o. He is author of several books and many scientific papers and articles in his main areas of interest: numerical electromagnetics, antennas, microwaves, and EMC.

Miodrag Tasic received his B. Sc. at the Dept. of EE, University of Belgrade in 1998, and M. Sc. at the same department in 2004. His main interests are: optimal and automatic meshing of polygonal geometrical models, and EM modeling of large structures using MoM and macro-basis functions.

Dragan Olcan is an Associate Professor at the Dept. of EE, University of Belgrade, Serbia where he received his B.Sc. M.Sc. and Ph.D. in 2001, 2004, and 2008, respectively. His main interests are: diakoptic approach to numerical electromagnetic analysis and optimization algorithms.

Dusan Zoric received his B. Sc. at the Dept. of EE, University of Belgrade in 2009, and M. Sc. at the same department in 2010. His main interest is GPU accelerated computing.

Srdjan Stevanetic received his B. Sc. at the Dept. of EE, University of Belgrade in 2009, and M. Sc. at the same department in 2010. His main interests are: efficient EM modeling of symmetrical structures, meshing algorithms, and interpolation techniques for efficient graphic presentation.

Distributed-Memory Parallelization of an Explicit Time-Domain Volume Integral Equation Solver on Blue Gene/P

Ahmed Al-Jarro[1], Mark Cheeseman[2], and Hakan Bağcı[1]

[1] Division of Physical Sciences and Engineering
King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia
ahmed.aljarro@kaust.edu.sa, hakan.bagci@kaust.edu.sa

[2] KAUST Supercomputing Laboratory
King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia
mark.cheeseman@kaust.edu.sa

Abstract ─ Two distributed-memory schemes for efficiently parallelizing the explicit marching-on-in-time based solution of the time domain volume integral equation on the IBM Blue Gene/P platform are presented. In the first scheme, each processor stores the time history of all source fields and only the computationally dominant step of the tested field computations is distributed among processors. This scheme requires all-to-all global communications to update the time history of the source fields from the tested fields. In the second scheme, the source fields as well as all steps of the tested field computations are distributed among processors. This scheme requires sequential global communications to update the time history of the distributed source fields from the tested fields. Numerical results demonstrate that both schemes scale well on the IBM Blue Gene/P platform and the memory-efficient second scheme allows for the characterization of transient wave interactions on composite structures discretized using three million spatial elements without an acceleration algorithm.

Index Terms ─ Distributed-memory parallelization, explicit solvers, IBM Blue Gene/P, marching-on-in-time, time domain volume integral equation.

I. INTRODUCTION

Time domain volume integral equation (TDVIE) solvers [1- 6] are becoming an attractive alternative to time domain finite element [7, 8] and finite-difference time domain [8, 9] methods for analyzing transient electromagnetic wave interactions on inhomogeneous dielectric structures. TDVIEs, which relate total electric fields to equivalent polarization source fields/currents induced in the dielectric volume, are constructed using the volume equivalence principle. Often times, marching-on-in-time (MOT) schemes are the method of choice for discretizing the TDVIEs and solving the resulting system of equations. The MOT scheme requires, at each time step, tested (retarded scattered) fields to be computed from the discretized spatial and temporal convolutions of the source fields/currents' time history with the free-space time-domain Green function. Unlike the differential equation based finite element and finite difference time domain methods, the Green function approach requires discretization of only the dielectric volume, avoids the need for absorbing boundary conditions, and virtually eliminates numerical phase dispersion. On the other hand, it renders MOT-TDVIE solvers susceptible to late-time instabilities and significantly increases their computational complexity. The computational cost of the MOT-TDIE solvers has been reduced with the development of the plane wave time domain
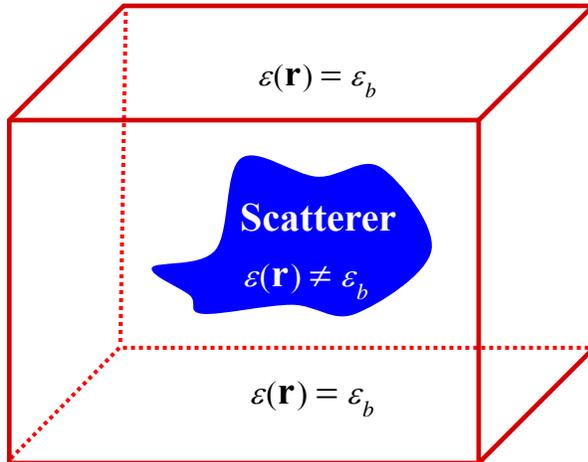
Fig. 1. Pictorial description of the volumetric scatterer in a background medium with relative permittivity ε_b.

(PWTD) algorithm [2, 3] and the (blocked) fast Fourier transform (FFT) based schemes [10-13] aimed at accelerating the computation of the discretized spatial and temporal convolutions. The problem of late-time instability has been practically alleviated with the development of implicit MOT schemes [1-3], which make use of accurate temporal interpolation rules [14, 15] and highly accurate (semi-) analytic integration techniques [16, 17].

When compared to the implicit techniques, explicit MOT schemes are less stable but more efficient: they do not require a matrix inversion at every time step and also, as a consequence, they do not suffer from possible ill-conditioning problems [18-20]. Until recently, instability of the explicit MOT-TDVIE solvers has been alleviated using effective but computationally expensive temporal filtering techniques [4]. To eliminate the use of these filtering techniques while maintaining the stability and explicitness of the solver, a predictor-corrector scheme is proposed in [5, 6].

In this work, to allow for the application of this predictor-corrector based TDVIE solver to the analysis of transient electromagnetic wave interactions with electrically large dielectric structures, two distributed-memory schemes are proposed. It should be noted here that the research efforts on the development of parallelization strategies for MOT-TDIE solvers are as recent as the work on the development of PWTD and FFT-based acceleration engines [11, 12, 21].

Especially, efficient distributed-memory parallelization schemes have been as indispensable as the acceleration engines in enabling the use of MOT-TDIE solvers in the analysis of transient electromagnetic wave interactions on electrically large structures.

Before reading the description of the parallelization schemes, one should be reminded here that the predictor-corrector based TDVIE solver parallelized in this work does not pre-compute or store any interaction matrices, which represent discretized retarded field interactions [5, 6]. Since it utilizes a nodal discretization scheme, the computation of the interactions is rather fast and it can be repeated without increasing the MOT time drastically. This approach makes the predictor-corrector based TDVIE solver memory efficient since only the time history of the source fields are stored. It should also be added here that the predictor-corrector based TDVIE solver approximates space and time derivatives, which operate on the scattered field by finite differences. It is well known that in parallel implementations of finite-difference time-domain method, computation of spatial finite differences require "halo" type localized communications between processors [22, 23]. Similar types of communications are needed in parallelized computation of the spatial finite differences in implementations of the TDVIE solver. The cost of these localized communications is much smaller than the global communications needed for parallel computation of the discretized integral present in the retarded field interactions. Therefore, the focus of this work is on comparing the performance of two different distributed-memory parallelization schemes for computing this integral discretized in space and time.

The first parallelization scheme implemented in this work is rather straightforward; each processor stores the time history of all source fields and only the computationally dominant step of the test field computations is distributed among processors. In this scheme, all processors compute the part of the tested fields that is assigned to them from the time history of all the source fields they store. "All-to-all" global communications are needed to update, from the tested fields, the time history of the source fields, which will be used in the computation of the next time step's tested fields. The second parallelization scheme is

slightly more difficult to implement; the source fields as well as all steps of the tested field computations are distributed among the processors. In this scheme, all processors compute in parallel the tested fields of a given processor only from the history of the source fields they store. After this step, tested fields are communicated to the given processor via an 'all reduce sum' operation to update the history of the source fields stored on that processor. These steps are repeated in a sequential manner for all processors storing different parts of the tested and source fields.

The two distributed-memory parallelization schemes described above run efficiently on Shaheen, an IBM Blue Gene/P platform, located at the Supercomputing Laboratory of the King Abdullah University of Science and Technology. The IBM Blue Gene/P platform possesses software and hardware optimizations that significantly improve MPI global communication operations, which form the essential core of both parallelization schemes. The hardware support comes in the form of a low latency tree network specifically dedicated to MPI global communication operations. This tree network is a unique defining feature of the IBM Blue Gene/P platform. Additionally, IBM has included its own MPI implementation, which is called DCMF (deep computing message framework) and provides optimized global MPI operations such as "MPI All-to-All" and "MPI Reduce". All of these optimizations result in substantial performance benefits for the proposed parallelization schemes as shown in Section IV.

Numerical results demonstrate that both schemes scale well on the IBM Blue Gene/P platform and the memory-efficient second scheme allows for characterization of transient electromagnetic wave interactions on dielectric structures discretized using three million spatial elements without any acceleration engine. Additionally, the effectiveness of the parallelized predictor-corrector based TDVIE solver is demonstrated via its application to the characterization of scattering of light from a red blood cell [24-26].

II. FORMULATION

In this section, the formulation and the space-time discretization scheme underlying the predictor-corrector based TDVIE solver is reviewed. For more details on the formulation and the discretization scheme, the reader is referred to [6].

Consider a scatterer compromising potentially inhomogeneous dielectric volumes represented by V with relative permittivity and permeability, $\varepsilon(\mathbf{r})$ and μ (Fig. 1). The scatterer resides in an infinite homogeneous (background) medium with relative permittivity and permeability, ε_b and μ_b. It is assumed that $\varepsilon(\mathbf{r})$, ε_b, μ, and μ_b, are frequency independent and $\mu = \mu_b = 1$. The wave speed in the background medium is given by $c_b = c_0 / \sqrt{\varepsilon_b \mu_b}$, where c_0 is the wave speed in free space. Let $\mathbf{E}_0(\mathbf{r},t)$ represent an incident electric field that is vanishingly small for $\mathbf{r} \in V$ and $t \leq 0$. $\mathbf{E}_0(\mathbf{r},t)$ excites the scatterer; in return the equivalent currents induced in V generate the scattered electric field $\mathbf{E}^{\text{sca}}(\mathbf{r},t)$. Expressing $\mathbf{E}^{\text{sca}}(\mathbf{r},t)$ in terms of equivalent currents, currents in terms of the total electric field $\mathbf{E}(\mathbf{r},t)$, and enforcing the fundamental field electric relation $\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r},t) + \mathbf{E}^{\text{sca}}(\mathbf{r},t)$ for $\mathbf{r} \in V$ yields the TDVIE in the unknown $\mathbf{E}(\mathbf{r},t)$, $\mathbf{r} \in V$ [6]:

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}_0(\mathbf{r},t) + \left[\nabla\nabla \cdot - \frac{\partial_t^2}{c_b^2} \right]$$
$$\int_V d\mathbf{r}' \mathbf{E}(\mathbf{r}',t') \frac{\varepsilon(\mathbf{r}') - \varepsilon_b}{4\pi\varepsilon_b R}, \ \mathbf{r} \in V. \tag{1}$$

Here, $R = |\mathbf{r} - \mathbf{r}'|$ is the distance between the observation and source points, $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z$ and $\mathbf{r} = \hat{\mathbf{x}}x' + \hat{\mathbf{y}}y' + \hat{\mathbf{z}}z'$ located in V, $t' = t - R/c_b$ is the retarded time, and ∂_t^2 represents the second order partial derivative with respect to time.

Equation (1) is discretized using a nodal discretization scheme to approximate the volume integral and finite differences to approximate the second order derivatives in space and time. Consider a spatial discretization, where cubic elements of dimension Δd are used to divide V into N_e number of elements and a uniform time discretization, where Δt and N_t represent the time step size and the number of total time steps. Electric field is sampled at the centers of the cubic elements at $\mathbf{r} = \mathbf{r}_i$, $i = 1,..,N_e$ and at times $t = t_n = n\Delta t$, $n = 1,..,N_t$. Following the detailed derivation in [6], the final form of the predictor-corrector algorithm is provided below in pseudo-code format:

Assume zero initial conditions for all variables

for $n = 1 : N_t$

$$\mathbf{F}(\mathbf{r}_i, t_n) =$$
$$\sum_{i \neq j}^{N_e} \Delta d^3 \frac{\varepsilon(\mathbf{r}_j) - \varepsilon_b}{4\pi\varepsilon_b R_{ij}} \mathbf{E}(\mathbf{r}_j, t_n - R_{ij}/c_b), \; i = 1, ..., N_e, \quad (2)$$

$$\mathbf{F}^s(\mathbf{r}_i, t_n) =$$
$$\mathbf{F}(\mathbf{r}_i, t_n) + \frac{\varepsilon(\mathbf{r}_i) - \varepsilon_b}{4\pi\varepsilon_b} \mathbf{E}(\mathbf{r}_i, t_n) S(\mathbf{r}_i), \; i = 1, ..., N_e, \quad (3)$$

$$\mathbf{f}(\mathbf{r}_i, t_n) = \tilde{\nabla}\tilde{\nabla} \cdot \mathbf{F}^s(\mathbf{r}_i, t_n) - \frac{\tilde{\partial}_t^2}{c_b^2} \mathbf{F}(\mathbf{r}_i, t_n), i = 1, ..., N_e, \quad (4)$$

$$\mathbf{E}^p(\mathbf{r}_i, t_n)$$
$$= \frac{1}{(1 + M\Delta t^2)} \Big[2\mathbf{E}(\mathbf{r}_i, t_{n-1}) - \mathbf{E}(\mathbf{r}_i, t_{n-2}) \quad (5)$$
$$+ M\Delta t^2 \{ \mathbf{E}_0(\mathbf{r}_i, t_n) + \mathbf{f}(\mathbf{r}_i, t_n) \} \Big], \; i = 1, ..., N_e,$$

$$\mathbf{E}(\mathbf{r}_i, t_n) = \mathbf{E}^p(\mathbf{r}_i, t_n), \; i = 1, ..., N_e, \quad (6)$$

Update $\mathbf{f}(\mathbf{r}_i, t_n)$ *for* $\forall (\mathbf{r}_i, \mathbf{r}_j) \, | \, |\mathbf{r}_i - \mathbf{r}_j| < 2c_b\Delta t$, (7)

$$\mathbf{E}^c(\mathbf{r}_i, t_n)$$
$$= \frac{1}{(1 + M\Delta t^2)} \Big[2\mathbf{E}(\mathbf{r}_i, t_{n-1}) - \mathbf{E}(\mathbf{r}_i, t_{n-2})$$
$$+ M\Delta t^2 \{ \mathbf{E}_0(\mathbf{r}_i, t_n) + 0.5\mathbf{f}(\mathbf{r}_i, t_n) + \quad (8)$$
$$0.5\mathbf{f}(\mathbf{r}_i, t_{n-1}) \} \Big], \; i = 1, ..., N_e,$$

$$\mathbf{E}(\mathbf{r}_i, t_n) = \mathbf{E}^c(\mathbf{r}_i, t_n), \; i = 1, ..., N_e, \quad (9)$$

Update $\mathbf{f}(\mathbf{r}_i, t_n)$ *for* $\forall (\mathbf{r}_i, \mathbf{r}_j) \, | \, |\mathbf{r}_i - \mathbf{r}_j| < 2c_b\Delta t$, (10)

end for

In the algorithm provided above

$$M = \frac{c_b^2 4\pi\varepsilon_b}{[\varepsilon(\mathbf{r}_i) - \varepsilon_b] S(\mathbf{r}_i)}, \quad (11)$$

$R_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between the test and the source points, $\mathbf{r}_i$ and $\mathbf{r}_j$, the singular integral

$$S(\mathbf{r}_i) = \int_{V_i} \frac{d\mathbf{r}'}{|\mathbf{r}_i - \mathbf{r}'|}, \quad (12)$$

where V_i is the support of the ith cubic element, is evaluated analytically as described in [27], and the operators "$\tilde{\nabla}\tilde{\nabla}\cdot$" and "$\tilde{\partial}_t^2$" are finite difference approximations of the continuous operators "$\nabla\nabla\cdot$" and "∂_t^2". Several comments about the above algorithm are in order: (i) Steps (5) and (8) are the predictor and the corrector steps, respectively, and the samples $\mathbf{E}^p(\mathbf{r}_i, t_n)$ and $\mathbf{E}^c(\mathbf{r}_i, t_n)$ are collectively termed "tested fields" while the samples $\mathbf{E}(\mathbf{r}_i, t_n)$ are termed "source fields". (ii) At step (4), $\tilde{\partial}_t^2$ is evaluated using a backward difference formula for test-source point pairs $(\mathbf{r}_i, \mathbf{r}_j)$, which satisfy the condition $|\mathbf{r}_i - \mathbf{r}_j| < 2c_b\Delta t$ and using a central difference formula for all other pairs. At step (7), $\tilde{\partial}_t^2$ is re-evaluated using a central difference formula for test-source point pairs $(\mathbf{r}_i, \mathbf{r}_j)$, which satisfy the condition $|\mathbf{r}_i - \mathbf{r}_j| < 2c_b\Delta t$. The use of central difference is now possible since $\mathbf{E}(\mathbf{r}_i, t_n)$ that were not known at step (4) (due to causality) are replaced by $\mathbf{E}^p(\mathbf{r}_i, t_n)$ obtained at the predictor step. Also note that, at step (10), $\mathbf{f}(\mathbf{r}_i, t_n)$ is "corrected" only for the same test-source point pairs since now $\mathbf{E}(\mathbf{r}_i, t_n)$ are replaced by $\mathbf{E}^c(\mathbf{r}_i, t_n)$ computed at the corrector step. The corrected $\mathbf{f}(\mathbf{r}_i, t_n)$ is used at step (8) of the next time step. This approach increases the accuracy of the finite difference approximations while maintaining the explicitness of the MOT scheme. (iii) When $t_n - R_{ij}/c_b$ is not an integer multiple of Δt, $\mathbf{E}(\mathbf{r}_j, t_n - R_{ij}/c_b)$ is approximated using a linear interpolation between $\mathbf{E}(\mathbf{r}_j, t_n)$ and $\mathbf{E}(\mathbf{r}_j, t_{n-1})$. (v) Note that, in (2), $\mathbf{E}(\mathbf{r}_j, t_n - R_{ij}/c_b)$, which satisfy the condition $t_n - R_{ij}/c_b < 0$ do not contribute to $\mathbf{F}(\mathbf{r}_i, t_n)$ since the fields radiated from the source point $\mathbf{r}_j$ have not yet reached the test point $\mathbf{r}_j$ at time $t = t_n$. (iv) The length of the temporal history of the source fields stored, in terms of time steps, is $\min(n-1, N_g)$; i.e., only $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $m = 1, ..., \min(n-1, N_g)$ are stored. Here, $N_g = \lfloor D_{\max}/c_b\Delta t \rfloor + 2$, where $D_{\max}$ is the maximum distance between any test-source point pair on V. (vi) Unlike the classical MOT schemes, the scheme described above does not pre-compute or store any interaction matrices; the full computation of $\mathbf{f}(\mathbf{r}_i, t_n)$ at step (4) and its correction at steps (7) and (10), in a sense, replaces the multiplication of these interaction matrices with the samples of the source fields' temporal history. (vii) Numerical results presented
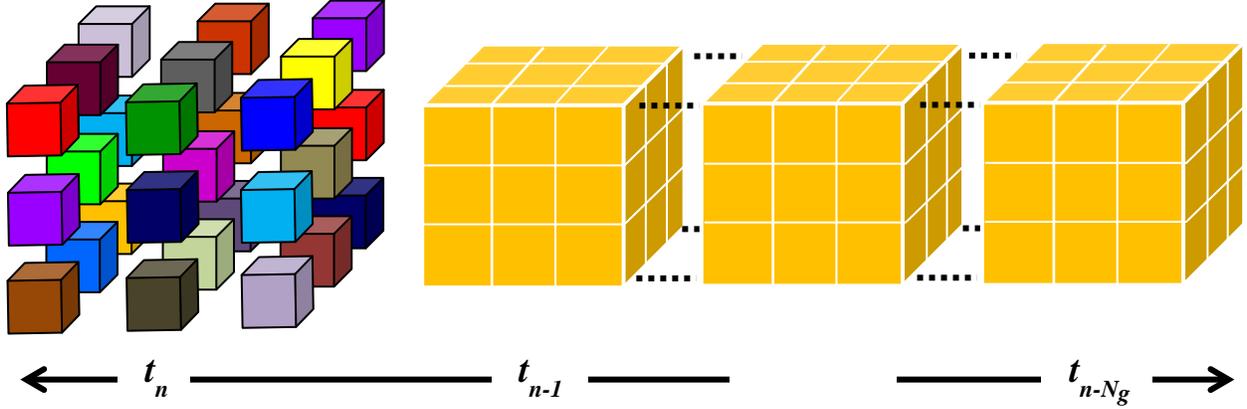
Fig. 2. Pictorial description of scheme 1. Each box of sampling points (represented with a different color) is assigned to a processor. Only computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$, at time step t_n is parallelized; every processor computes $\mathbf{F}(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box using the time history of the all source fields, $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1,...,N_e$, $m = 1,...,\min(n-1, N_g)$, which all processors store (represented with solid yellow blocks).

here and in [5, 6] show that the MOT scheme described above is stable when $\varepsilon(\mathbf{r}) - \varepsilon_b$ is not large.

III. PARALLELIZATION

The per-time-step computational cost of the predictor-corrector based MOT scheme described in Section II is dominated by the computational cost of evaluating the discrete convolutions $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$ at step (2). At the N_g^{th} time step, the fields radiated from all source points $\mathbf{r}_j$, $j = 1,...,N_e$ reach all test points $\mathbf{r}_i$, $i = 1,...,N_e$. From this time step onwards, the cost of evaluating $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$ at time step t_n for $n > N_g$ is $O(N_e^2)$, since all source points interact with all test points. All other operations that are carried out at steps (2)-(10) of the MOT scheme are localized in space and time. Their contribution to the computational cost is very limited especially for large N_e. As a result, the total computational cost of the predictor-corrector based MOT scheme scales as $O(N_t N_e^2)$ under the assumption that $N_t \gg N_g$. This high computational cost could be reduced by integrating PWTD- [2, 3] or blocked FFT-based [10-13] schemes into the predictor-corrector based MOT scheme. Another way of rendering the MOT scheme applicable to the analysis of transient electromagnetic wave interactions on electrically large dielectric structures is through parallelization; which allows for executing the scheme on distributed memory clusters with thousands of cores.

In this section, two distributed-memory parallelization schemes are proposed for accelerating the predictor-corrector based MOT-TDVIE solver. Scheme 1 is a straightforward MPI/OpenMP hybrid parallelization scheme that involves global all-to-all operations; and scheme 2 is a slightly more complex algorithm that involves sequential global reductions. Both schemes are fundamentally aimed at accelerating the computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$ at step (2). Detailed descriptions of the schemes are presented next.

A. Scheme 1

The partitioning of the geometry has no effect on the parallelization efficiency of scheme 1. The space sampling points, $\mathbf{r}_i$, $i = 1,...,N_e$ can be randomly assigned to processors; as long as they are equally distributed, the scheme will provide the highest efficiency. For the sake of simplicity in the visualization, one can assume that the space sampling points on a rectangular scatterer are assigned to processors as shown in Fig. 2. In this figure, each box of sampling points (shown with a different color on the left) is assigned to a processor.

In this scheme, each processor computes and stores all tested fields, $\mathbf{E}^p(\mathbf{r}_i, t_n)$ and $\mathbf{E}^c(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$ at a given time step t_n, and updates and stores the time history of all source fields, $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1,...,N_e$, $m = 1,...,\min(n-1, N_g)$. Only the computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1,...,N_e$ at step (2) is parallelized. At time step t_n, each
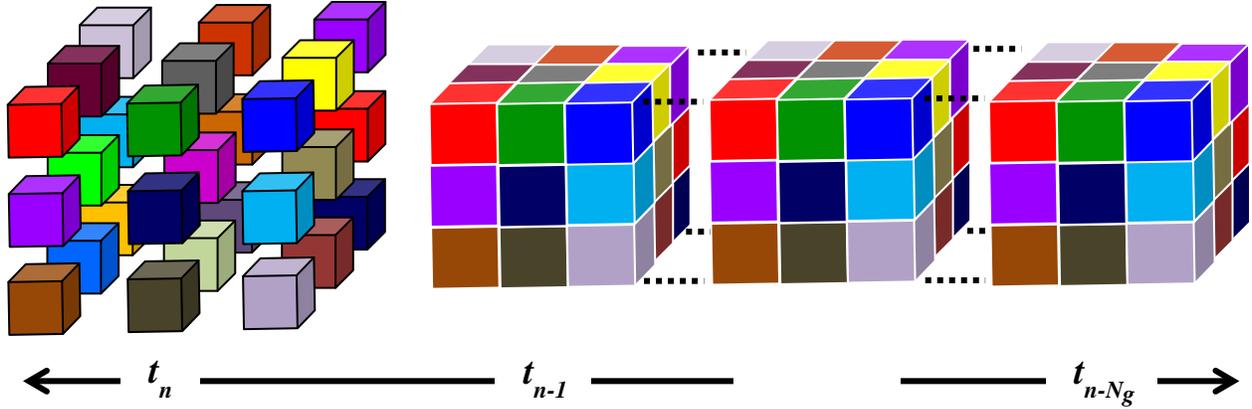
Fig. 3. Pictorial description of scheme 2. Each box of sampling points (represented with a different color) is assigned to a processor. All steps in the computation of tested fields $\mathbf{E}^p(\mathbf{r}_i, t_n)$ and $\mathbf{E}^c(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$, at time step t_n is parallelized; every processor stores the tested fields $\mathbf{E}^p(\mathbf{r}_i, t_n)$, $\mathbf{E}^c(\mathbf{r}_i, t_n)$, and the time history of the source fields, $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1, ..., N_e$, $m = 1, ..., \min(n-1, N_g)$, for all $\mathbf{r}_i$ that reside in its box.

processor independently computes $\mathbf{F}(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box. The parts of $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$, which are computed and stored on different processors are communicated to every processor via a global all-to-all operation. At this point, each processor stores $\mathbf{F}(\mathbf{r}_i, t_n)$, for all $i = 1, ..., N_e$. Thereafter, computations at steps (3), (4), and (5) are replicated on all processors. $\mathbf{E}^p(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ is computed and stored on all processors. At step (6), $\mathbf{E}(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ is updated and stored on all processors. At step (7), each processor updates $\mathbf{f}(\mathbf{r}_i, t_n)$ for every test-source pair $(\mathbf{r}_i, \mathbf{r}_j)$ that satisfies $|\mathbf{r}_i - \mathbf{r}_j| < 2c_b \Delta t$. At step (8), $\mathbf{E}^p(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ is computed and stored on all processors. At step (9), $\mathbf{E}(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ is updated and stored on all processors. Step (10) is the same as step (7).

Several remarks about scheme 1 are in order: (i) The scheme is straightforward to implement. The parallelization can be carried out on the serial version of the predictor-corrector based MOT scheme by simply incorporating calls to "MPI_Allgatherv" subroutine after step (2). (ii) Since $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1, ..., N_e$, $m = 1, ..., \min(n-1, N_g)$ is stored on every processor, the largest spatial discretization that can be handled by this approach is limited by the memory available on a given processor. (iii) Storing $\mathbf{E}^p(\mathbf{r}_i, t_n)$, $\mathbf{E}^c(\mathbf{r}_i, t_n)$, and $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ as well as $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1, ..., N_e$, $m = 1, ..., \min(n-1, N_g)$ on all processors avoids the local communications needed to compute the finite differences in steps

(4), (7), and (10). This simplifies the implementation considerably and provides the flexibility in geometry partitioning mentioned at the beginning of Section III-A. (iv) The extension of the scheme to include shared memory hybridization is straightforward: OpenMP is used throughout, that is in steps (2)-(10) to parallelize the computations when they are distributed over multi-core processors.

B. Scheme 2

Unlike scheme 1, the partitioning of the geometry has an effect on the parallelization efficiency, even though it is small, as explained in the text below. Similar to the description of scheme 1, for the sake of simplicity in the visualization, one can assume that the space sampling points on a rectangular scatterer are assigned to processors as shown in Fig. 3. In this figure, each box of sampling points is assigned to a processor.

In this scheme, each processor stores, for all $\mathbf{r}_i$ that reside in its box, the tested fields, $\mathbf{E}^p(\mathbf{r}_i, t_n)$ and $\mathbf{E}^c(\mathbf{r}_i, t_n)$ at a given time step t_n and the corresponding time history of source fields, $\mathbf{E}(\mathbf{r}_i, t_{n-m})$, $m = 1, ..., \min(n-1, N_g)$. Note that unlike scheme 1, all steps of the MOT scheme are parallelized. The computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, $i = 1, ..., N_e$ at step (2) is parallelized in a sequential manner (over processors) that makes use of global reductions. Assume that, at time step t_n, $\mathbf{F}(\mathbf{r}_i, t_n)$ is being computed for all $\mathbf{r}_i \subset RB$, i.e., all

sampling points reside in red box (Fig. 4(a) left side); and let P_R represent the processor assigned to RB. Note that each $\mathbf{F}(\mathbf{r}_i, t_n)$ with $\mathbf{r}_i \subset RB$ is a summation of sampled field contributions radiated from all source points, $\mathbf{r}_j$, $j = 1, ..., N_e$, which are distributed among the processors. Each processor independently computes its contributions to $\mathbf{F}(\mathbf{r}_i, t_n)$, for all $\mathbf{r}_i \subset RB$, from only $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $m = 1, ..., \min(n-1, N_g)$ that it stores. Then, these contributions are communicated to P_R via a global reduction call, which is executed with the "sum" flag. At this point, $\mathbf{F}(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i \subset RB$ is fully computed and stored on P_R. Then, the scheme moves, for example, to the purple box, represented with PB (Fig. 4(b), left side), and repeats the parallel computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, for all $\mathbf{r}_i \subset PB$. This step is repeated processor by processor until $\mathbf{F}(\mathbf{r}_i, t_n)$, for all $\mathbf{r}_i = 1, ..., N_e$ is computed and stored part-by-part on all processors. Step (3) is executed independently on every processor without any communications. At step (4), if the computation of $\tilde{\nabla}\tilde{\nabla} \cdot \mathbf{F}(\mathbf{r}_i, t_n)$ requires a (local) finite difference grid that strides across the boundary of two partitioning boxes, then a local communication must occur between the two processors that are assigned to those boxes. At step (5), each processor computes and stores $\mathbf{E}^p(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box. At step (6), each processor updates $\mathbf{E}(\mathbf{r}_i, t_n)$ from $\mathbf{E}^p(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box. At step (7), each processor updates $\mathbf{f}(\mathbf{r}_i, t_n)$ for every test-source pair $(\mathbf{r}_i, \mathbf{r}_j)$ that satisfies $|\mathbf{r}_i - \mathbf{r}_j| < 2c_b \Delta t$. Similar to step (4), local communications are required. At step (8), each processor computes and stores $\mathbf{E}^c(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box. At step (9), each processor updates $\mathbf{E}(\mathbf{r}_i, t_n)$ from $\mathbf{E}^c(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in its box. Step (10) is the same as step (7).

Several remarks about scheme 2 are in order: (i) Scheme 2 is slightly more complicated to implement than scheme 1. The global reduction call at step (2) is implemented by incorporating calls to "MPI_Reduce" subroutine with a "sum" flag. (ii) Since, $\mathbf{E}(\mathbf{r}_j, t_{n-m})$, $j = 1, ..., N_e$, $m = 1, ..., \min(n-1, N_g)$ are distributed among processors in scheme 2, it is more memory efficient when compared to scheme 1. (iii) The halo type local communications needed at steps (4), (7), and (10) which are also used in many other parallelization schemes [22, 23] have little effect on the scheme's overall parallelization performance. Note that to

(a)

(b)
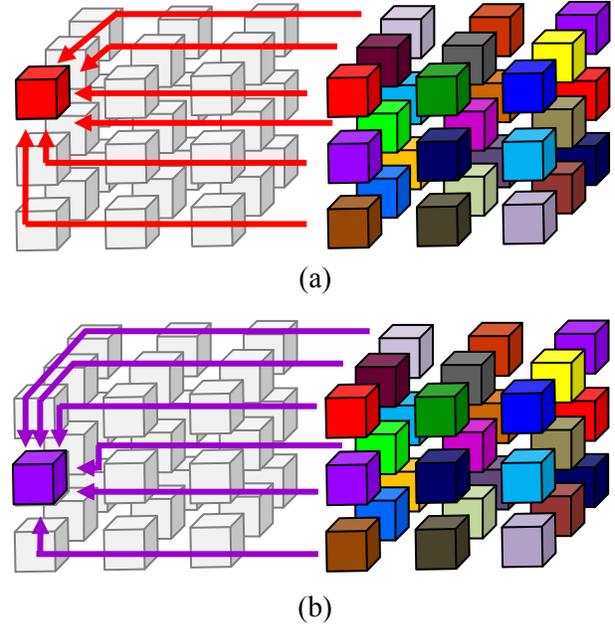
Fig. 4. Pictorial description of the parallelization of the computation of $\mathbf{F}(\mathbf{r}_i, t_n)$, at time step t_n in scheme 2. (a) $\mathbf{F}(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in the red box (on the left) is computed in parallel. (b) $\mathbf{F}(\mathbf{r}_i, t_n)$ for all $\mathbf{r}_i$ that reside in the purple box is computed in parallel.

minimize halo type communications; one may need to find an optimal strategy to partition the geometry.

IV. NUMERICAL EXPERIMENTS

Scalability tests of the two parallelization schemes proposed in this paper for accelerating the predictor-corrector based MOT TDVIE solver are performed on the IBM Blue Gene/P platform located at Supercomputing Laboratory of the King Abdullah University of Science and Technology. The IBM Blue Gene/P platform, named Shaheen, possesses an IBM design, which was awarded the National Medal of Technology and Innovation in U.S. in 2009. Shaheen has 16384 compute nodes, each of which contains four processing cores and 4 GB shared physical memory. Each processing core runs at a modest clock rate of 850 MHz. However, the addition of a double floating-point unit, an 8MB high speed cache memory, and a fast main memory bandwidth of 13.6 GB/sec raises the peak processing limit of each core to a respectable 13.6 GF/sec. Superior connectivity between the IBM Blue Gene/P platform's compute nodes is

Table 1: Tabulated scaling results for the problem with $N_e = 531441$

$4N_p$	$S_{N_p}^1$	$S_{N_p}^2$
512	0	0
1024	0.9781	0.9865
2048	1.9561	1.9931
4096	2.8087	2.8004
8192	3.6827	3.6994
16384	4.2971	4.8783
32768	4.7659	5.1977

Table 2: Tabulated scaling results for the problem with $N_e = 3048625$

$4N_p$	$S_{N_p}^1$	$S_{N_p}^2$
512	0	0
1024	0.9884	0.9257
2048	1.9783	1.7916
4096	2.9376	2.5422
8192	3.8745	3.4167
16384	4.7459	4.4652
32768	5.5259	5.2762
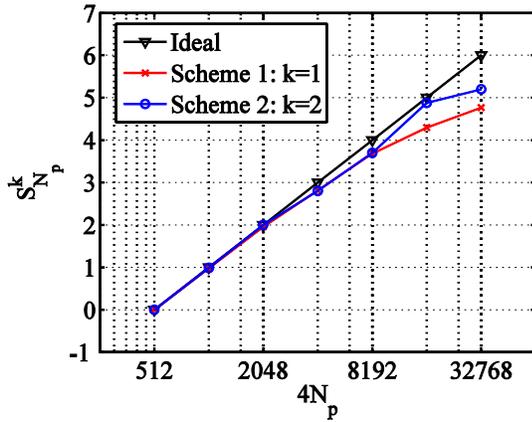
Fig. 5. Scalability of schemes 1 and 2 for a moderate-size problem with $N_e = 531441$.
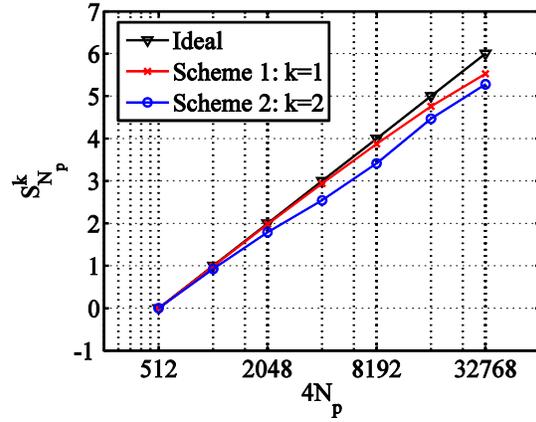
Fig. 6. Scalability of schemes 1 and 2 for a large problem with $N_e = 3048625$.

provided via two dedicated communication networks. A 3D torus network is available for fast point-to-point MPI communication between compute nodes. This network has a peak bandwidth of 5.1 GB/sec and a low latency of only 3.5 microseconds. The second network is dedicated to MPI global communication operations. Every compute node possesses three connections to this tree-based network providing a low latency of 2.5 microseconds per MPI message. Lastly, installed on the IBM Blue Gene/P is a software stack that includes the MPI library called DCMF (deep computing message framework). DCMF library provides optimized versions of normal MPI operations including 'MPI All-to-All' and 'MPI Reduce', which are heavily optimized to run efficiently on the Blue Gene/P platform including the two dedicated networks interconnecting the compute nodes. It should be emphasized here that the parallelization schemes proposed in this work benefit from the superiority

of the second network dedicated to global communications and the optimized DCMF library, since the schemes extensively utilize global MPI communications.

In the remainder of this section, first the weak scalability of the two parallelization schemes are compared, then the effectiveness of the parallelized predictor-corrector based MOT TDVIE solver is demonstrated via its application to the characterization of scattering of light from a red blood cell [24-26].

A. Scalability

Weak scalability for the proposed schemes is investigated here for two scenarios: a moderately sized problem with $N_e = 531441$ and a much larger problem with $N_e = 3048625$. For both examples, scalability results are presented in Tables 1 and 2 and plotted in Figs. 5 and 6. Here, N_p represents the number of compute nodes of

the IBM Blue Gene/P platform used in the simulations. This means that for scheme 1, which is hybridized with OpenMP that uses four threads per node, the number of parallel tasks is $4N_p$. Similarly, scheme 2, which is a pure distributed memory implementation, is executed on $4N_p$ cores since each compute node of the IBM Blue Gene/P platform has four processing cores. The weak scaling is defined as $S_{N_p}^k = \log_2(T_{N_p}^k / T_{ref}^k)$. Here, the subscript "$N_p$" refers to the simulation carried on N_p nodes, and the superscript "k" refers to the parallelization scheme used. The recorded total times, $T_{N_p}^k$, include both communication and computation times. T_{ref}^k is the reference total time recorded for the simulation that is executed with the lowest N_p. It is clear from Fig. 5 that, for the smaller size problem, the scalability of scheme 1 is hindered by the communication costs when $4N_p$ is larger than 8192. On the other hand, Fig. 5 shows that, for the larger problem, the scalability of scheme 1 carries over even for large values of $4N_p$ around 32768. Figs. 5 and 6 also demonstrate that there is no distinct difference in scheme 2's scalability behaviour for the two problems up to $4N_p = 32768$. Even if the scalability behaviour of the two schemes is different, as clearly demonstrated by the results presented here, they both scale very well on the IBM Blue Gene/P platform.

It should be noted here that, to be able to run the large problem with parallelization scheme 1, N_g was artificially set to a small number. If the actual N_g was used in the simulation, 4GB memory of a single node of the IBM Blue Gene/P platform would not be large enough to store the time history of the source fields. Note that this problem does not exist for the parallelization scheme 2; its memory efficient implementation allows for storing the time history of the source fields using the actual value of N_g.

B. Light scattering from red blood cells

Over the last two decades many biomedical devices utilizing lasers for disease diagnosis have been developed. Consequently, there is an increasing interest in understanding how electromagnetic waves interact with biological cells and tissue. In particular, the analysis of light-scattering from red blood cells (RBCs) have attracted the interest of many researchers [24-26]
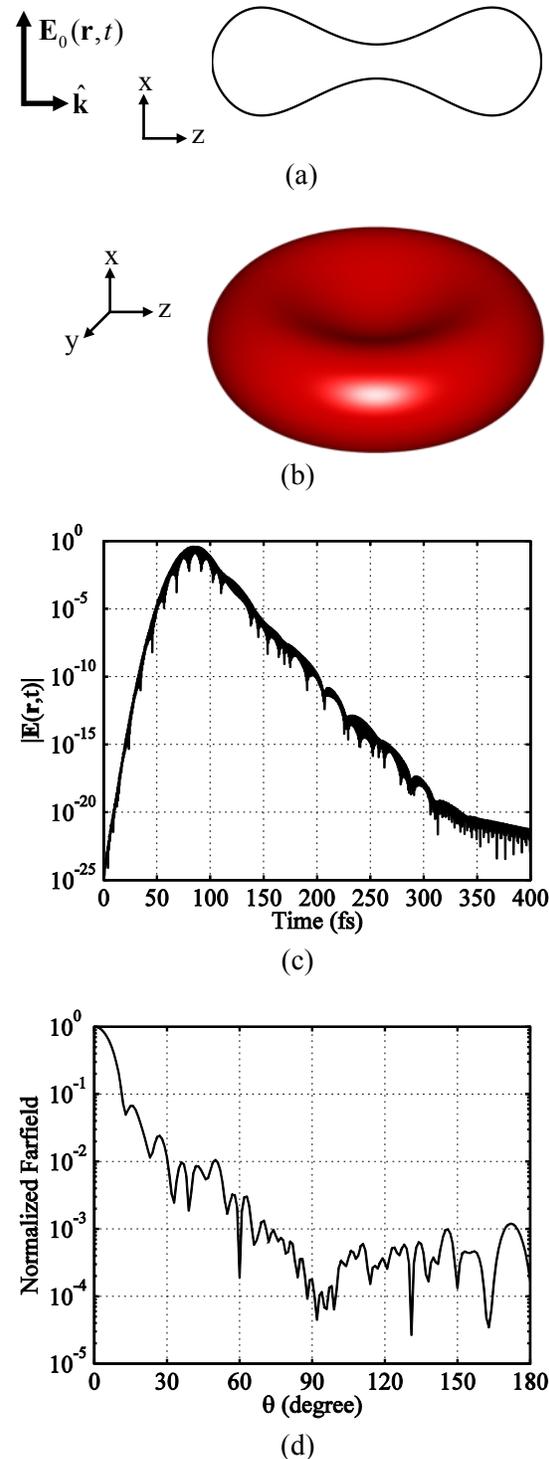
Fig. 7. Analysis of light scattering from an RBC. (a) Cross-section of the RBC model along the xz - plane and the plane wave excitation. (b) Three dimension view of the model. (c) The amplitude of the transient electric field induced at the center of the RBC. (d) Normalized amplitude of the electric far-field on the xz - plane.

since this type of analysis may provide essential information for the diagnosis of blood related diseases [26].

For this example, the parallelized predictor-corrector MOT-TDVIE solver is used to analyze light scattering from an RBC residing in plasma. The membrane of RBCs has a negligible effect on the scattered field [24], and hence the RBC model does not include the membrane or any other internal structure and is formulated as a biconcave volume as described in [24]. Figures 7 (a) and (b) depict the xz-cross section and three-dimensional view of this model, respectively. The diameter of the cell is $7.82\,\mu m$, its largest and smallest thickness values on xz-plane are $2.565\,\mu m$ and $0.81\,\mu m$, respectively, producing a volume of $94\,\mu m^3$. The relative permittivities of the RBC and the background plasma are $\varepsilon(\mathbf{r}) = 1.9768$ and $\varepsilon_b = 1.8089$, respectively [24]. Note that while absorption can also be handled by the implemented MOT-TDVIE scheme, for the excitation frequencies considered, it is very small and thus is neglected [24, 25]. The excitation is an $\hat{\mathbf{x}}$ polarized plane wave with a modulated Gaussian time signature, which propagates in the $\hat{\mathbf{z}}$ direction; the incident electric field is then expressed as

$$\mathbf{E}_0(\mathbf{r},t) = \hat{\mathbf{x}}E_0 G(t - z/c_b), \qquad (13)$$

$$G(t) = \cos\left[2\pi f_o(t - t_p)\right]\exp\left[-(t - t_p)^2 / \varsigma^2\right], \quad (14)$$

where $E_0 = 1\,\text{V/m}$ is the electric field amplitude and $G(t)$ represents a Gaussian pulse, $f_o = 637.2\,\text{THz}$, $\varsigma = 3/(2\pi f_{bw})$, $f_{bw} = f_0/15$, and $t_p = 6\varsigma$ are its modulation frequency, duration, bandwidth, and delay, respectively. The grid space is chosen as $\Delta d = 0.045\,\mu m$, which produces $N_e = 1031550$ discretization elements within the volume of the RBC. The time step $\Delta t = 0.15\,\text{fs}$ and the simulation is carried out for $N_t = 2666$ time steps.

The amplitude of the electric field at the center of the RBC, $\mathbf{E}(\mathbf{r}_o,t)$, $\mathbf{r}_o = (0,0,0)$, which is recorded during the simulation, is presented in Fig. 7(c); the figure clearly demonstrates the late-time stability of the scheme. Additionally, frequency-domain scattered farfields on the xz-plane are computed. For this purpose, at $f = 473.8\,\text{THz}$ ($\lambda = 0.6328\mu m$, in the plasma), the Fourier transform of the currents induced in the RBC volume is computed during time marching using a running discrete Fourier transform (DFT) summation. Note that the DFT of the currents is normalized by the Fourier transform of $G(t)$ to produce the time-harmonic currents. Then, as a post-processing step, the farfields are easily computed from the frequency-domain currents. Figure 7(d) plots the normalized amplitude of the electric farfield on the xz-plane. The results agree well with those generated in [24] using the finite difference time domain method.

V. CONCLUSIONS

Two distributed-memory schemes are proposed to efficiently parallelize the predictor-corrector based MOT-TDVIE solver on the IBM Blue Gene/P platform. The first scheme distributes the computationally dominant step of the tested field computations among the processors using the MPI standard. To achieve an easy-to-implement and highly-scalable parallelization scheme, the time history of the source fields are stored simultaneously on all processors. Within each multi-core processor, OpenMP standard is used to further accelerate the computation of the tested fields. Obviously, the fundamental limitation of this scheme is its high memory requirement due to the storage of the time history of all the source fields on each processor.

The second scheme alleviates this limitation by distributing the time history of the source fields as well as all steps of the tested field computations among the processors. Even though the implementation of scheme 2 is slightly more complicated than scheme 1, numerical results demonstrate that scheme 2 scales as well as scheme 1 on the IBM Blue Gene/P platform. The memory efficient scheme 2 allows the predictor-corrector based MOT-TDVIE solver to simulate transient electromagnetic wave interactions on electrically large structures discretized using more than three million spatial elements.

It should be emphasized here again that both methods benefit from the IBM Blue Gene/P platform's superior tree network dedicated to global communications and optimized MPI library since they heavily utilize "MPI_Allgatherv" and "MPI_Reduce" subroutines. Other computing platforms do not have this level of hardware and software support for MPI global communications. Development of the extensions to the method

proposed here, which would scale well on other platforms, is underway.

ACKNOWLEDGEMENT

REFERENCES

[1] N. T. Gres, A. A. Ergin, E. Michielssen, and B. Shanker "Volume-Integral-Equation-Based Analysis of Transient Electromagnetic Scattering from Three-Dimensional Inhomogeneous Dielectric Objects," *Radio Sci.*, vol. 36, no. 3, pp. 379-386, May 2001.

[2] B. Shanker, K. Aygun, and E. Michielssen, "Fast Analysis of Transient Scattering from Lossy Inhomogeneous Dielectric Bodies," *Radio Sci.*, vol. 39, pp. 1-14, Mar. 2004.

[3] G. Kobidze, J. Gao, B. Shanker, and E. Michielssen, "A Fast Time Domain Integral Equation Based Scheme for Analyzing Scattering from Dispersive Objects," *IEEE Trans. Antennas Propag.*, vol. 53, no. 3, pp. 1215-1226, Mar. 2005.

[4] Al-Jarro, P. Sewell, T. M. Benson, A. Vukovic, and J Paul, "Transient Time-Dependent Electric Field of Dielectric Bodies using the Volterra Integral Equation in Tthree Dimensions," *Progress Electromag. Res.*, vol. 110, pp. 179-197, 2010.

[5] A. Al-Jarro and H. Bagci, "A Predictor-Corrector Scheme for Solving the Volterra Integral Equation," *The XXX General Assembly and Scientific Symposium of URSI*, Istanbul, Aug. 2011.

[6] A. Al-Jarro, M. A. Salem, H. Bagci, T. M. Benson, P. Sewell, and A. Vukovic, "Explicit Solution of the Time Domain Volume Integral Equation using a Stable Predictor-Corrector Scheme," submitted for publication, 2011.

[7] P. P. Silvester and R. L. *Ferrari, Finite Elements for Electrical Engineers*. Cambridge, U.K, Cambridge University Press, 1990.

[8] F. L. Teixeira, "A Summary Review on 25 Years of Progress and Future Challenges in FDTD and FETD Techniques," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 1, pp. 1-14, Jan. 2010.

[9] A. Taflove and Susan C. Hagness, *Computational Electrodynamics: The Finite Difference Time Domain Method*, Artech House, 2005.

[10] A. E. Yilmaz, J. M. Jin, and E. Michielssen, "Time Domain Adaptive Integral Method for Surface Integral Equations," *IEEE Trans. Antennas Propag.*, vol. 52, no. 10, pp. 2692-2708, Oct. 2004.

[11] A. E. Yilmaz, J. M. Jin, and E. Michielssen, "A Parallel FFT Accelerated Transient Field-Circuit Simulator," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2851-2865, Sep. 2005.

[12] H. Bagci, A. E. Yilmaz, J.-M. Jin, and E. Michielssen, "Fast and Rigorous Analysis of EMC/EMI Phenomena on Electrically Large and Complex Cable-Loaded Structures," *IEEE Trans. Electromagn. Comp.*, vol. 49, no. 9, pp. 361-381, May 2007.

[13] H. Bagci, A. E. Yilmaz, and E. Michielssen, "An FFT-Accelerated Time-Domain Multiconductor Transmission Line Simulator," *IEEE Trans. Electromagn. Comp.*, vol. 52, no. 1, pp. 199-214, Feb. 2010.

[14] G. Manara, A. Monorchio, and R. Reggiannini, "A Space-Time Discretization Criterion for a Stable Time-Marching Solution of the Electric Field Integral Equation," *IEEE Trans. Antennas Propag.*, vol. 45, no. 3, pp. 527-532, Mar. 1997.

[15] D. S. Weile, G. Pisharody, N.-W. Chen, B. Shanker, and E. Michielssen, "A Novel Scheme for the Solution of the Time-Domain Integral Equations of Electromagnetics," *IEEE Trans. Antennas Propag.*, vol. 52, no.1, pp. 283-295, Jan. 2004.

[16] Y. Shi, M. Xia, R. Chen, E. Michielssen, and M. Lu, "Stable Electric Field TDIE Solvers via Quasi-Exact Evaluation of MOT Matrix Elements," *IEEE Trans. Antennas Propag.*, vol. 59, no. 2, pp. 574-585, Feb. 2011.

[17] H. A. Ulku and A. A. Ergin, "Application of Analytical Retarded-Time Potential Expressions to the Solution of Time Domain Integral Equations," *IEEE Trans. Antennas Propag.*, vol. 59, no. 11, pp. 4123- 4131, Nov. 2011.

[18] F. P. Andriulli, H. Bagci, F. Vipiana, G. Vecchi, and E. Michielssen, "A Marching-on-in-Time Hierarchical Scheme for the Time Domain Electric Field Integral Equation," *IEEE Trans. Antennas Propag.*, vol. 55, no. 12, pp. 3734-3738, Dec. 2007.

[19] F. P. Andriulli, H. Bagci, F. Vipiana, G. Vecchi, and E. Michielssen, "Analysis and Regularization of the TD-EFIE Low-Frequency Breakdown,"

IEEE Trans. Antennas Propag., vol. 57, no. 7, pp. 2034-2046, July 2009.

[20] H. Bagci, F. P. Andriulli, F. Vipiana, G. Vecchi, and E. Michielssen, "A Well-Conditioned Integral-Equation Formulation for Efficient Transient Analysis of Electrically Small Microelectronic Devices," *IEEE Trans. Adv. Packag.*, vol. 33, no. 2, pp. 468-480, May 2010.

[21] F. Wei and A. E. Yilmaz, "A Hybrid Message Passing/Shared Memory Parallelization of the Adaptive Integral Method for Multi-Core Clusters," *Parallel Comp.*, vol. 37, no. 6-7, pp. 279-301, June-July 2011.

[22] X. Duan, X. Chen, K. Huang, and H. Zhou, "A High Performance Parallel FDTD Based on Winsock and Multi-Threading on a PC-Cluster," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 3, pp. 241-249, March 2011.

[23] J. E. Lump, S. K. Mazumder, and S. D. Gedney, "Performance Modeling of the Finite-Difference Time-Domain Method on Parallel Systems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 19, no. 2, pp. 147-159, July 1998.

[24] J. He, A. Karlsson, J. Swartling, and S. Andersson-Engels, "Light Scattering by Multiple Red Blood Cells," *J. Opt. Soc. Am. A*, vol. 21, no. 10, pp. 1953-1961, Oct. 2004.

[25] J. Q. Lu, P. Yang, and X.H. Hu, "Simulations of Light Scattering from a Biconcave Red Blood Cell using the Finite-Difference Time-Domain Method," *J. Biomed Opt.*, vol. 10, no. 2, pp. 024022-10, Mar./Apr. 2005.

[26] O. Ergul, A. Arslan-Ergul, and L. Gurel, "Computational Study of Scattering from Healthy and Diseased Red Blood Cells using Surface Integral Equations and the Multilevel Fast Multipole Algorithm," *J. Biomed. Opt.*, vol. 15, no. 4, pp. 045004-8, July/Aug. 2010.

[27] J. Waldvogel, "The Newtonian Potential of Homogeneous Cube," *J. Applied Math. Phys.*, vol. 27, no. 6, pp. 867-871, 1979.

Ahmed Al-Jarro received the B.Eng. degree in Electronic Engineering with Spanish and the Ph.D. degree in Electrical and Electronic Engineering from the University of Nottingham, UK, in 2001 and 2004, respectively. From 2004 to 2009, he was a Research Assistant and Research Fellow at the George Green Institute for Electromagnetics Research (GGIEMR), University of Nottingham, UK. In 2010, he was awarded with the Knowledge Transfer Secondment from the University of Nottingham, UK, in collaboration with Photon Design Ltd., Oxford, UK. In 2011, he joined the Division of Physical Sciences and Engineering at King Abdullah University of Science and Technology (KAUST), Saudi Arabia, as a Postdoctoral Research Fellow. Dr. Al-Jarro was the recipient of the Japan Society for the Promotion of Science, JSPS, Fellowship Award in 2010.

Dr. Al-Jarro's research interests are in the field of computational electromagnetics for the analysis of photonic and optical devices. He is currently working on the development of explicit and stable marching-on-in-time based schemes for solving the time domain volume integral equation, and their efficient parallelization.

Mark Cheeseman received the B.Sc. degree in Physics from the Memorial University of Newfoundland, Canada, in 1999 and the M.Sc. degree in Earth and Atmospheric Science from the University of Alberta, Canada, in 2003. He has been a computational scientist at the Supercomputing Laboratory of the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, since 2009. He possesses over eight years of experience at various HPC centers; including the Swiss National Supercomputing Center, Switzerland, and the National Oceanography Center in Southampton, UK.

The majority of his work concerns the optimization and re-engineering of existing scientific applications for efficient use on massively parallel platforms. These applications range from combustion models, to acoustic wave simulation, and distributed 3D Fast Fourier Transform libraries. His technical interests are particularly focused on IO optimization and the scalability of present-day mathematical libraries.

Hakan Bağcı received the B.Sc. degree in Electrical and Electronics Engineering from the Bilkent University, Ankara, Turkey, in 2001 and the M.Sc. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2003 and 2007, respectively. From June 1999 to July 2001, he worked as an Undergraduate Researcher at the Computational Electromagnetics Group, Bilkent University. From 2001 to 2007, he was a Research Assistant at the Center for Computational Electromagnetics and Electromagnetics Laboratory, UIUC. From 2007 to 2009, he worked as a Research Fellow at the Radiation Laboratory, University of

Michigan. In 2009, he joined the Division of Physical Sciences and Engineering at the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, as an Assistant Professor.

His research interests include various aspects of computational electromagnetics with emphasis on time-domain integral equations and their fast marching-on-in-time based solutions, well-conditioned integral-equation formulations, and development of fast hybrid methods for analyzing statistical EMC/EMI phenomena on complex and fully loaded platforms.

Dr. Bağcı was the recipient of the 2008 International Union of Radio Scientists (URSI) Young Scientist Award and the 2004–2005 Interdisciplinary Graduate Fellowship from the Computational Science and Engineering Department, UIUC. His paper titled "Fast and rigorous analysis of EMC/EMI phenomena on electrically large and complex structures loaded with coaxial cables" was one of the three finalists (with honorable mention) for the 2008 Richard B. Schulz Best Transactions Paper Award given by the IEEE Electromagnetic Compatibility Society. He authored and co-authored three finalist papers and another paper, which is awarded honorable mention, in the student paper competitions at the 2005, 2008, and 2010, IEEE Antennas and Propagation Society International Symposiums.

On-The-Fly Mesh Generation for a High Performance Physical Optics Radar Backscattering Simulator

Sebastian Hegler, Ronny Hahnel, and Dirk Plettemeier

Communications Laboratory
Dresden University of Technology, Dresden, 01127, Germany
sebastian.hegler@tu-dresden.de, ronny.hahnel@tu-dresden.de, dirk.plettemeier@tu-dresden.de

Abstract — In this paper, we present a radar backscattering simulator based on the method of physical optics (PO). Our simulation tool closely intertwines the tessellation of the simulation geometry with the physical optics method kernel, which enables on-the-fly refinement of input model data while still yielding high precision and computational performance. The algorithms for the physical optics method as well as the parallelization scheme will be presented. Also, performance comparisons will be shown and explained, both in regard to accuracy of the results and computation time.

Index Terms — Remote sensing, vector and parallel computation.

I. INTRODUCTION

Radar clutter is loosely defined as "the part of the received signal that is undesired." For ground penetrating radar systems (aboard satellites or other space craft, for example), this definition applies to the part of the signal that is backscattered from the surface of the sounded object. Separating surface clutter from the received signal is a tremendous aid for the correct interpretation of radar images.

In order to calculate the backscattered signal of known terrains, we developed a simulation tool based on the method of physical optics. The development of this tool was driven by the need to simulate huge objects, i.e., entire moons or planets, and the possibility to choose arbitrary radiation patterns for the sender and receiver. It is implemented in C++, using OpenMP[1] for parallelization.

The following section is dedicated to introducing the method of physical optics. In Section III, our simulation tool is presented. Section IV gives a description of the generation of on-the-fly meshes. In Section V, results of different simulation scenarios will be compared. The final section is dedicated to drawing conclusions and outlining further work.

II. METHOD OF PHYSICAL OPTICS

A. Method overview

The method of physical optics is a method to calculate the electromagnetic field backscattered from the surface of an object. As this method neglects electromagnetic coupling, its demands for computation time and main memory are modest compared to other methods, for example the method of moments. Assuming far field conditions further simplifies the equations; this assumption is justified for our use cases.

The formulae given in this section describe the method of physical optics for dielectric bodies, assuming far field conditions, a homogeneous dielectric permittivity ε_r and permeability μ_r.

B. Surface current densities

Given a far field radiation pattern (as calculated by antenna simulation tools, for example, or as an analytical formulation where possible), the incident electric and magnetic far field for a given point in free space can be calculated. These are denoted by $\mathbf{E}_i$ and $\mathbf{H}_i$. For a dielectric surface with a known homogeneous dielectric permittivity ε_r and permeability μ_r, the equivalent magnetic and

electric surface current densities can be calculated as follows[2–4]:

$$\mathbf{J}_s^{PO} = 2\hat{\mathbf{n}} \times \left[\left(\frac{1}{1 + \zeta cos\vartheta_i} \, \mathbf{e}_{i\|} \otimes \mathbf{e}_{i\|} \right) \cdot \mathbf{H}_i \right]$$
$$+ 2\hat{\mathbf{n}} \times \left[\left(\frac{1}{1 + \frac{\zeta}{cos\vartheta_i}} \, \mathbf{e}_{i\perp} \otimes \mathbf{e}_{i\perp} \right) \cdot \mathbf{H}_i \right],$$

and

$$\mathbf{M}_s^{PO} = - 2\hat{\mathbf{n}} \times \left[\left(\frac{\frac{\zeta}{cos\vartheta_i}}{1 + \frac{\zeta}{cos\vartheta_i}} \, \mathbf{e}_{i\|} \otimes \mathbf{e}_{i\|} \right) \cdot \mathbf{E}_i \right]$$
$$- 2\hat{\mathbf{n}} \times \left[\left(\frac{\zeta cos\vartheta_i}{1 + \zeta cos\vartheta_i} \, \mathbf{e}_{i\perp} \otimes \mathbf{e}_{i\perp} \right) \cdot \mathbf{E}_i \right],$$

where

$$\zeta = \frac{Z_F}{Z_0} = \frac{\sqrt{\frac{\varepsilon_r}{\mu_r}}}{\sqrt{\frac{\varepsilon_0}{\mu_0}}},$$

$\hat{\mathbf{n}}$ is the surface unit normal, and ϑ_i is the angle of incidence. $\mathbf{e}_{i\perp}$ and $\mathbf{e}_{i\|}$ are the perpendicular and parallel base unit vectors of the plane of incidence, separating the incident field into perpendicular and parallel components. Thereby, for both $\mathbf{J}_s^{PO}$ and $\mathbf{M}_s^{PO}$ polarization is accounted for.

C. Backscattered field

Using dyadic Green's function in vectorial form, the electric field based on the equivalent surface current densities can be calculated[5, 6]. Assuming far field conditions again, the equations can be simplified. The electric field caused by the equivalent magnetic surface current, denoted $\mathbf{E}_m$, can be written as follows:

$$\mathbf{E}_m(\mathbf{r}', \mathbf{M}_s^{PO}) = jk\hat{\boldsymbol{\beta}}_r \times \int_S \frac{e^{-jkr}}{4\pi r} \mathbf{M}_s^{PO}(\mathbf{r}) \, ds,$$

whereas the magnetic field caused by the equivalent electric surface current densities, denoted $\mathbf{H}_e$, is calculated as follows:

$$\mathbf{H}_e(\mathbf{r}', \mathbf{J}_s^{PO}) = -jk\hat{\boldsymbol{\beta}}_r \times \int_S \frac{e^{-jkr}}{4\pi r} \mathbf{J}_s^{PO}(\mathbf{r}) \, ds,$$

where $\mathbf{r}$ is the source point and $\mathbf{r}'$ is the target point. $r = ||\mathbf{r}' - \mathbf{r}||$ is the distance between these

points, and $\hat{\beta}_r = r^{-1}(\mathbf{r}' - \mathbf{r})$ is the unit vector of the direction. The total field can then be written as:

$$\mathbf{E}_{total} = \mathbf{E}_m - Z_0\hat{\boldsymbol{\beta}}_r \times \mathbf{H}_e.$$

For each surface element, $\mathbf{E}_{total}$ is weighted with the correspondent antenna gain before numerical integration is performed. By calculating a frequency series and transforming it to time domain, the radar echo backscattered by the given surface is calculated.

III. PHYSICAL OPTICS SIMULATOR

A. Overview

In its current state, our physical optics simulator is capable of calculating the backscattered radar echo of a given input topography in monostatic as well as in bistatic mode. Antenna radiation patterns for a Hertzian dipole or a finite length dipole are provided as analytical formulations. Arbitrary antenna radiation patterns can be used, by importing far field ϑ-φ maps for $\mathbf{E}$ and $\mathbf{H}$. Post processing is performed to generate radargram images from the frequency domain simulation output using the Python programming (scripting) language[7].

B. Software evolution

Starting as a set of MATLAB scripts, the software was first ported to Fortan90 in order to build a standalone binary application[8]. This allowed a much higher degree of parallelism, as the bottlenecks of MATLAB and its MEX interface, namely the limitation of parallel constructs to compiled MEX files and the inability to re-use memory, could be overcome this way.

For a final re-implementation, C++ was chosen as programming language, because of its advantages over the Fortran family of programming languages. These include, amongst others, a concise integration of the paradigm of object oriented programming (while refraining from making its use obligatory), stricter type-safety enforcement (while offering well-defined ways to circumvent it where necessary), support for generic programming (using C++'s keyword `template`), and complete configurability at a low level (by operator overloading), making the resulting code easier to understand, maintain and evolve.
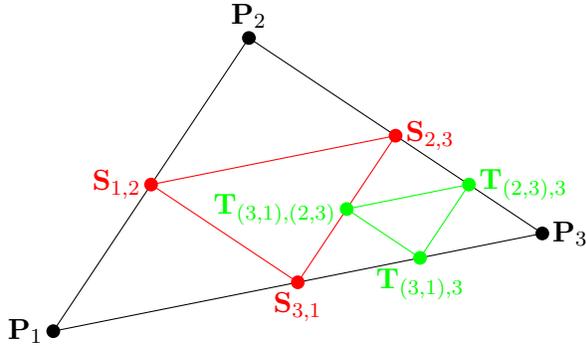
Fig. 1.   Sierpinski tessellation of a triangle. Points S are the edge bisection points of the original triangle. Points T, also edge bisection points, illustrate one further step in the refinement process.

Thanks to a change in the internal data structures and data flow layout, the allocation of memory for intermediate results could be removed, which was a limiting factor of the degree of parallelism. This resulted in a speedup of factor 2.5 (for serial execution): a test model with 235.623 triangles completes in $45s$, compared to $120s$ for the Fortran90 version on an Intel Xeon X5365.

Parallel processing is implemented using OpenMP instructions. This is sufficient, as the problem resident set can be kept small enough to fit into the main memory of small compute nodes. The following section presents how this is achieved.

IV. ON-THE-FLY MESH GENERATION

A. Overview and requirements

A crucial factor for electrically large problem sets is the generation of a tessellation of the computational domain. Unfortunately, currently no tessellation software packages are available which feature a multi-threaded implementation and/or the ability to deal with huge datasets. For our typical use cases, the area to be tessellated is huge, ranging from 6100 km^2 in a basic case, to 1.4 million km^2 and beyond, usually consisting of some 10000^2 points. Pre-calculating a tessellation for these areas would consume enormous amounts of time, and possibly exhaust main memory even on HPC platforms. For example, for the first given case, a tessellation occupies 12 GB of hard

disk space, with 5 GB being the data relevant to the simulation, and the rest being adjacency information for the tessellation graph; however, this pre-generated mesh still has a resolution that is greater than the wavelength, and is therefore not fine enough to yield precise results.

The original terrain data is provided as a *Digital Elevation Model* (DEM), i.e., as a matrix of values denoting a difference in radius compared to a reference sphere or ellipsoid. Adjacency relationship is implicitly available as the point's neighborhoods. The only preprocessing stage that is necessary is to convert the implicit coordinates of the DEM to explicit Cartesian coordinates; the matrix structure, however, is retained, and therefore also the neighborhood relationship is. This data structure can then be used in the simulator to generate an on-the-fly tessellation of the topography of interest.

B. Triangle selection strategies

For the on-the-fly meshing procedure, a triangle selection strategy has to be selected first. As the selection strategy plays is important to the performance of the code, it is implemented as a C++ template argument. The neighborhood relationship of the DEM allows four possible combinations
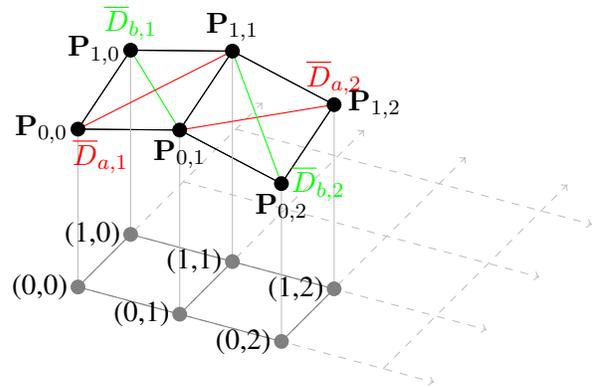
Fig. 2.   Triangle construction from elevation map point set. Points $\mathbf{P}$ illustrate the heights at a given matrix coordinate. The diagonals $\overline{D}$ illustrate possible diagonals for the triangular tessellation. Four possible selection strategies can be applied: left-upper-right-lower (red), left-lower-right-upper (green), or choosing the diagonal by length (shorter or longer).

with two general cases, as illustrated in Figure 2. The first two cases are static: the same diagonal is chosen for all points, by either splitting the area into a triangle residing up left and one below right (illustrated by the red diagonals $\overline{D}_a$), or the other way round (illustrated by the green diagonals $\overline{D}_b$). The remaining cases are dynamic, by choosing the triangles based on the length of the diagonal.

After having chosen a selection strategy $\mathcal{S}$, a refinement criterion $\mathcal{C}$ needs to be defined. The default setting for $\mathcal{C} = \lambda_{min} = c_0/f_{max}$, i.e. the wavelength of the highest frequency of the sweep that is to be simulated.

C. Parallel iteration

Given the input dataset, the selection strategy $\mathcal{S}$, and a refinement criterion $\mathcal{C}$, an iterator object, denoted $\mathcal{I}_\mathcal{S}(DEM, \mathcal{C})$, is generated.

Generating and using an iterator object is a natural choice when C++ is chosen as the implementation language. The class describing the iterator is implemented as a random access iterator, allowing consistent use within the idiomatics of C++. Since the iteration over the triangles provided by the input data set is the outermost loop, this is where the OpenMP instructions to parallelize the computation are placed.

This approach has several advantages. Firstly, it allows a single thread to work on adjacent memory items, allowing hardware-level pre-fetching to come into effect, thereby reducing memory latency. Using OpenMP's dynamic scheduling with a moderate chunk size, load imbalance caused by triangles that are discarded from the calculation can easily be accounted for. Experience shows that setting the chunk size to 5000 to 10000 works well, balancing scheduling overhead versus parallel computation time. Secondly, as an added benefit, this implementation is very easy to use as a (header-only) library function, since the user only has to provide an iterator class to feed the geometry to the simulator core. Parallelization and infrastructure is already taken care of.

D. Mesh refinement

Using the algorithm to calculate the back-scattered echo for a triangle T from Section II, and

inspired by [9] a tessellation using Sierpinski triangles was implemented in order to increase precision where the original tessellation does not meet the necessary numerical criterion $\mathcal{C}$ for precise simulation results. The tessellation itself is implemented as a recursive function. The following piece of pseudo-code describes the function intertwined with the PO method. An illustration of the tessellation is shown in Figure 1.

Given a triangle $T : (\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$ generated by $\mathcal{I}_\mathcal{S}(DEM, \mathcal{C})$:

Call decomposition function $\mathcal{D}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$:

1) Are all edge lengths (l_1, l_2, l_3) of $T < \mathcal{C}$?
   yes: Calculate back-scattered signal of T, return
2) Calculate edge bisection points $(\mathbf{S}_{1,2}, \mathbf{S}_{2,3}, \mathbf{S}_{3,1})$ from $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$
3) Calculate $\mathcal{D}(\mathbf{P}_1, \mathbf{S}_{1,2}, \mathbf{S}_{3,1})$
4) Calculate $\mathcal{D}(\mathbf{S}_{1,2}, \mathbf{P}_2, \mathbf{S}_{2,3})$
5) Calculate $\mathcal{D}(\mathbf{S}_{3,1}, \mathbf{S}_{2,3}, \mathbf{P}_3)$
6) Calculate $\mathcal{D}(\mathbf{S}_{1,2}, \mathbf{S}_{2,3}, \mathbf{S}_{3,1})$

E. Evaluation

This implementation has several advantages. Firstly, generation and storage of a surface tessellation can be completely omitted, removing the primary computational bottleneck of the simulator tool chain. The resident memory set size of the simulator is reduced to only the pre-processed input elevation map plus the memory to store the results, which is another advantage. Therefore, thirdly, this allows for a much higher degree of parallelism, as the simulator now iterates over triangles which are generated on-the-fly, using only the neighboring points of the current point. Although this incurs a small penalty in total computation time, this allows simulation of very large radar targets.

V. COMPARISON OF RESULTS

This section is dedicated to a comparison of the simulator in regard to computational performance and precision. To this end, two very different scenarios are used. For both cases, the input data set is provided as a DEM.

Test case A represents a very large area with generally smooth slopes, where the distance between points is large. The total number of triangles

Fig. 3. Frequency domain signal for test case A, with varying segmentation.
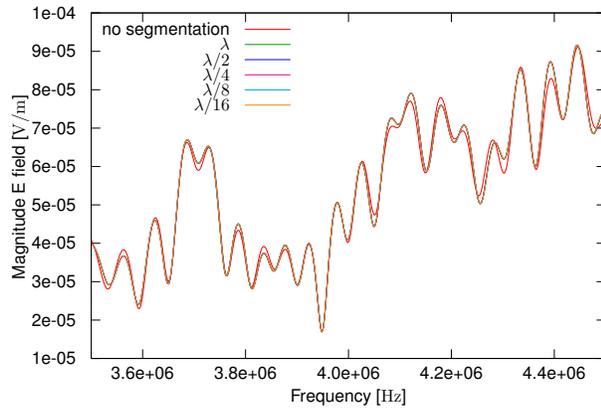
Fig. 5. Frequency domain signal for test case B, with varying segmentation.

in this dataset can be calculated from the number of points of the DEM: 12000^2 points yield 287952002 initial triangles.

Test case B represents a closed body, which is smaller than the radar foot print of the simulated antenna system for the given simulation distance. The Cartesian distance between two points at the object's equator is smaller than the average distance for test case A, while it is in the range of a few meters at the poles. Given a DEM of 14000 by 7000 points, this results in 195958002 initial triangles.

All computation time measurements were done with the time(1) utility. The simulations were run on a 4x4 core Xeon E7430 system using the Linux operating system. The compiler used was gcc-4.5[10], compilation flags where chosen to yield best performance, even when code size would grow, also using the compiler's "unsafe"
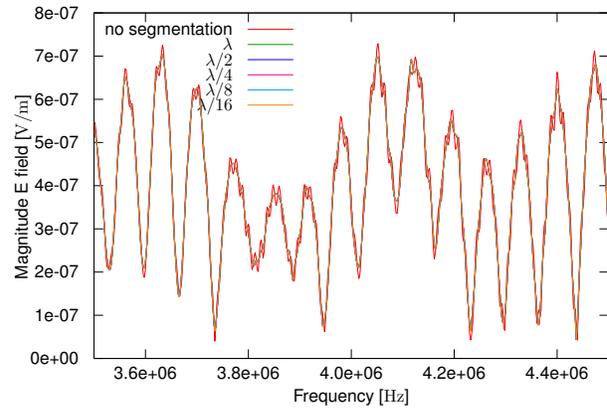
math optimizations (these optimizations may violate the IEEE 754[11] standard by changing rounding behavior and floating point unit error reporting; also see [12]).

A. Precision of the simulation

As can be seen in Figures 3 and 5, for both test case A and test case B the curves of the simulations with segmentation disabled diverge notably. However, for segmentations using values smaller or equal to the wavelength, no increase in precision is gained.

When comparing the time domain signal for test case A shown in Figure 4, it can be seen that the segmentation plays only a little role in the simulation signal results, as the raw data already yields a good result.

Fig. 4. Time domain signal for test case A, with varying segmentation.
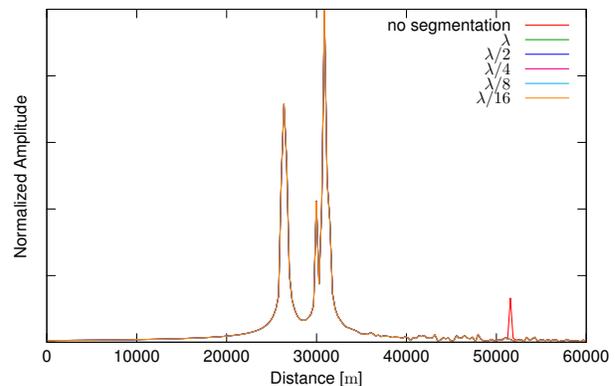
Fig. 6. Time domain signal for test case B, with varying segmentation.

Table 1: Simulation time for test case A, with varying segmentation

Segmentation	λ_{rel}	t_{user} [s]	Increase of t_{user}
none	n/a	178.22	n/a
λ	66.6	2389.40	13.407
$\lambda/2$	33.3	9545.24	3.994
$\lambda/4$	16.6	37964.96	3.977
$\lambda/8$	8.3	151567.09	3.992
$\lambda/16$	4.1	605593.99	3.995

Table 2: Simulation time for test case B, with varying segmentation

Segmentation	λ_{rel}	t_{user} [s]	Increase of t_{user}
none	n/a	21215.08	n/a
λ	66.6	21212.09	0.999
$\lambda/2$	33.3	21222.50	1.000
$\lambda/4$	16.6	21249.54	1.001
$\lambda/8$	8.3	22705.27	1.068
$\lambda/16$	4.1	75746.32	3.336

However, for test case B, the distortion introduced by the insufficiently fine discretization causes phantom echoes, as shown in Figure 6. This can be explained by the small size of the simulated object: for a closed surface which fits into the radar foot print and many triangles with edges that are bigger than the wavelength, side lobes are introduced with increasing angle of incidence. This can be seen in Figure 5. With mesh refinement enabled, the side lobes disappear, as do, in consequence, the phantom echoes.

B. Computational performance

The Sierpinski refinement approach described in Section IV has a clear disadvantage: the increase of computation time by factor four for each level of refinement. This can clearly be seen in Table 1. The increase in computation time for test case A by a factor of 13.407 between disabled segmentation and λ refinement indicates that the original triangles are indeed very large; implying an average of 1.8725 necessary refinement stages.

Test case B shows that for all refinement levels, except for $\lambda/16$, the simulations run in nearly the same computation time. For $\lambda/16$, an increase of factor 3.3 was measured, indicating that only for this case, a further segmentation was necessary for a majority of the first level triangles. The minuscule decrease of computation time between the first two lines was probably caused by different loads on the compute host.

The differences between the initial and refined computation times for cases A and B can be explained by the simulated geometries. For case A, about 80% of the first level triangles are discarded, since they are not within the radar footprint and therefore do not contribute to the result. For case B however, about 50% of the first level triangles

need to be considered for the simulation, resulting in a higher total computation time.

VI. CONCLUSIONS AND FURTHER WORK

A simulation tool to calculate the backscattered radar echo of large dielectric surfaces was successfully implemented. Switching to C++ as the implementation language resulted in an immense gain of computational performance, and also in degrees of freedom in regard to programming and further development of the code base. The code base is now easier to maintain, and extensions are easier to implement. Core code features are available as template classes, which allow easy component reuse and compile-time configuration, without sacrificing computational performance. For example, the iterator class was adapted for a different geometry format. The only required changes were the loading routines for the different format, and the iterator itself, boiling down to 20 lines of code.

As a task for further investigation, integrating the method described in [13] as another simulation kernel is considered, as it may yield considerable savings in computation time.

REFERENCES

[1] OpenMP.org – The OpenMP API specification for parallel programming. [Online]. Available: http://openmp.org/wp/

[2] C. Balanis, *Advanced Engineering Electromagnetics*. Wiley, 1989.

[3] E. Yamashita, *Analysis Methods for Electromagnetic Wave Problems*, ser. The Artech House antenna library. Artech House, no. 2, 1996.

[4] U. Jakobus, *Intelligente Kombination verschiedener numerischer Berech-*

nungsverfahren zur effizienten Analyse elektromagnetischer Streuprobleme unter besonderer Berücksichtigung der Parallelverarbeitung.   Shaker, 1999.

[5] L. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*, ser. IEEE Press series on electromagnetic waves.   IEEE Press, 1994.

[6] L. Díaz and T. Milligan, *Antenna Engineering using Physical Optics: Practical CAD Techniques and Software*, ser. Artech House antenna library.   Artech House, no. 1, 1996.

[7] Python Programming Language – Official Website.   [Online].   Available: http://www.python.org

[8] S. Hegler, R. Hahnel, and D. Plettemeier, "Implementation of a High Performance Numerical Simulator for Radar Surface Echoes," in *Proceedings of The 2nd International Multi-Conference on Engineering and Technological Innovation: IMETI 2009*.   International Institute of Informatics and Systemics, 2009.

[9] M. Bader and C. Zenger, "Efficient Storage and Processing of Adaptive Triangular Grids Using Sierpinski Curves," in *Computational Science – ICCS 2006*, ser. Lecture Notes in Computer Science, V. Alexandrov, G. van Albada, P. Sloot, and J. Dongarra, Eds.   Springer Berlin / Heidelberg, vol. 3991, pp. 673–680, 2006.

[10] GCC, the GNU Compiler Collection. [Online]. Available: http://gcc.gnu.org/

[11] IEEE Task P754, *IEEE 754-2008, Standard for Floating-Point Arithmetic*, Aug. 2008.

[12] GCC 4.5.3 Manual. [Online]. Available: http://gcc.gnu.org/onlinedocs/gcc-4.5.3/gcc/

[13] J.-F. Nouvel, A. Herique, W. Kofman, and A. Safaeinili, "Radar Signal Simulation: Surface Modeling with the Facet Method," in *Radio Science*, vol. 39, 2004.

Exploiting FPGAs and GPUs for Electromagnetics Applications: Interferometric Imaging in Random Media Case Study

E. El-Araby, O. Kilic, and V. Dang

Department of Electrical Engineering and Computer Sciences
The Catholic University of America, Washington, DC 20064, U.S.A.
aly@cua.edu, kilic@cua.edu, and 13dang@cardinalmail.cua.edu

Abstract ─ There is a growing need for reliable and efficient numerical methods for electromagnetic applications. This is important for addressing the complex designs with fine features on electrically large platforms. As designs become more complex, a good prediction of overall system performance becomes essential for cost reduction especially in the conceptualization stage. Researchers have attempted to address this issue by developing hybrid methods based on asymptotic techniques that can avoid the numerical inefficiency while maintaining high degrees of accuracy. Another approach is to implement fast computational methods that utilize parallel computing platforms. This paper focuses on the latter; i.e. by investigating the use of field programmable gate arrays (FPGA) and general purpose graphics processing units (GPGPU) as coprocessors to parallelize numerically challenging problems. The weaknesses and strengths of both platforms will be investigated in the context of their ease of use, efficiency, and potential for accelerated computations.

Index Terms ─ FPGA, GPU, hardware accelerated computing, high performance computing, imaging, interferometry, numerical methods, random media.

I. INTRODUCTION

In the field of electromagnetic modeling, the complex designs for engineered materials coupled with performance analysis of radio frequency components integrated within their natural environment drive the need for highly efficient numerical techniques. This cannot be achieved by conventional computer systems, but rather through using the so-called high performance computing (HPC) systems. HPC systems often utilize an integrated package of multiprocessors, multicore processors, and specialized hardware enabling rapid computations by exploiting the parallelism of these hardware platforms. Various configurations and platforms exist including clusters that utilize multiple CPUs (e.g., Cray), PCs supported by GPGPU (e.g., NVIDIA, AMD) and FPGA (e.g., Xilinx, Altera) based systems. In the current state of the art, both GPGPU and FPGA systems use these hardware as coprocessors. Such configuration and usage is typically referred to as hardware acceleration.

GPGPUs have been increasingly utilized for accelerated computing in numerous fields as they are readily integrated in PCs. Their main advantages are the high memory bandwidth as well as the availability of multiple vendors developing commercial tools that enable high level language support. FPGAs, on the other hand, are highly customizable and reconfigurable chips which can be optimally configured for a specific application.

This paper investigates these hardware acceleration platforms for computational electromagnetics applications, and is an extension of the work reported earlier in [1]. The application used for this investigation is the interferometric imaging of objects behind random media. A detailed discussion on the different architectures and programming environments for FPGA and GPGPU based systems is presented in Section II. Interferometric imaging of targets behind random media is chosen as the application. The details of the interferometric imaging are introduced in Section III. The implementation approaches on both platforms are given in Section IV. Performance analysis and metrics for evaluation are presented in Section V. The experimental

results are provided in Section VI, followed by the conclusions in Section VII.

II. HARDWARE ACCELERATION

Hardware acceleration is the use of hardware to enable parallel processing for higher computation speed than is possible in software running on the general purpose CPU. Examples of hardware acceleration are systems that include field programmable gate arrays (FPGA) and/or general purpose graphics processing units (GPGPU).

A. FPGA-based systems

The evolution of hardware acceleration based on reconfigurable computers (RC), such as FPGAs, has been progressing along two orthogonal technology-characterizing paths, namely performance and flexibility. RCs evolved from originally being discrete components used mainly as glue-logic devices in larger systems to accelerator boards and recently to parallel reconfigurable supercomputers often referred to as high-performance reconfigurable computers (HPRCs). Examples of such supercomputers are the SRC-7 and SRC-6 [2], the SGI Altix/RASC [3] and the Cray $XT5_h$ and Cray XD1 [4].

Reconfigurable computing (RC) architectures

A reconfigurable computer system typically consists of microprocessor and reconfigurable processor sub-systems closely coupled with each other through a common interface. The microprocessor sub-system includes all major components of a traditional computer system such as general purpose microprocessors, microprocessor memories, and I/O interfaces. On the other hand, a reconfigurable processor sub-system consists of one or more FPGAs, FPGA memories, and an I/O interface. A generalized architecture of a reconfigurable computer is shown in Fig. 1. The functionality of each component within the system is influenced by its interconnection topology. Within a reconfigurable processor sub-system, every FPGA can be connected to every other FPGA, or FPGAs can be grouped into clusters. FPGA memories can be shared by a group of FPGAs or each can be dedicated to a single FPGA. Additional hardware, such as a cross-bar switch, might be necessary to make connections as flexible as possible.
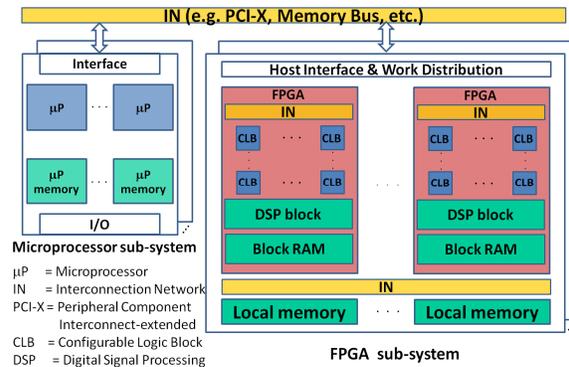
Fig. 1. General architecture of RC systems.

RC programming models

Application development on RC systems typically requires software and hardware programming expertise for which design paradigms and tools have been traditionally separate [1]. These products aim to abstract underlying hardware design details and streamline the disparate design flows [5]. They often tradeoff performance for programmability [6]. Dataflow design tools, based on the graphical user interface, e.g., DSPLogic, seem to offer an interesting compromise between high-level languages (HLLs) and hardware description languages (HDLs), e.g. VHDL and/or Verilog. They allow a tradeoff between a shorter development time and a performance overhead imposed by HLLs [7]. Streamlining hardware description using HLLs typically used in software programming, or at least using dataflow languages, is a major and distinctive feature of high performance RCs that potentially allows domain scientists to develop entire applications without relying on hardware designers. However, an HLL compiler for RCs must combine the capabilities of tools for traditional microprocessor compilation and tools for computer-aided design with FPGAs. It must also extend these two separate set of tools with capabilities for mutual synchronization and data transfer between microprocessors and reconfigurable processor sub-systems [8].

B. GPGPU-based systems

Due to their powerful floating-point computational capabilities and massively parallel processor architecture, GPUs are increasingly being used as application accelerators in the high-performance computing arena. Thus, a wide range

of HPC systems now incorporate GPUs as hardware accelerators including systems ranging from clusters of compute nodes to parallel supercomputer systems. Several examples of GPU-based computer clusters may be found in academia, such as [9]. Latest offerings from supercomputer vendors have begun to include GPUs in the compute blades of their parallel machines; examples include the SGI Altix UV and the recently announced plans from Cray to include NVIDIA GPUs into the Cray XE6 supercomputer.

GPGPU architecture

Similar to an RC system, a GPGPU computer system typically consists of a traditional microprocessor sub-system and a GPU sub-system closely coupled with each other through a common interface. A GPU sub-system consists of a multitude (texture/cluster) of processors often referred to as streaming multiprocessor (SM) and GPU local memory. Each SM contains a number of processor cores usually called streaming processors (SP). A generalized architecture of a GPGPU computer is shown in Fig. 2. For example, the Fermi architecture, which is one of the most recent GPU products by NVIDIA, is composed of 16 streaming multiprocessors (SMs) each of which consists of 32 streaming processor (SP) cores.
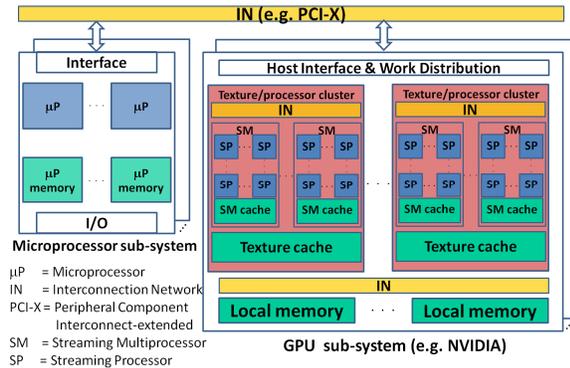
Fig. 2. General architecture of GPGPU-based systems.

GPGPU programming models

Among the different parallel programming approaches for GPGPU platforms, the most commonly followed programming style is the single program multiple data (SPMD) model [10]. Under the SPMD scenario, multiple processes execute the same program on different CPU cores, simultaneously operating on different data sets in parallel. By allowing autonomous execution of processes at independent points of the same program, SPMD serves as a convenient yet powerful approach for efficiently making use of the available hardware parallelism.

Two widely used GPU programming models, i.e., CUDA (by NVIDIA) and OpenCL (by Khronos Working Group) are similar and follow SPMD flow by executing data-parallel kernel functions within the GPU. Both models also provide abstractions of thread (basic unit for parallel execution) group hierarchy and shared memory hierarchy. In terms of thread group hierarchy, both provide three hierarchy levels: grid, block, and thread. GPU kernels are launched per grid and a grid is composed of a number of blocks, which have the access to the global device memory. Each block consists of a group of threads, which are executed concurrently and share the accesses to the on-chip shared memory. Each thread is a very lightweight execution of the kernel function. From programming perspective, the programmer needs to write the kernel program for one thread and decides the total number of threads to be executed on the GPU device while dividing the threads into blocks based on the data-sharing pattern, memory sharing, and architectural considerations.

III. INTERFEROMETRIC IMAGING IN RANDOM MEDIA

The earliest applications of interferometric measurements have been reported in the field of radio astronomy [11,12]. The main advantage of interferometry is the higher resolution achieved by the use of multiple antennas instead of using an equivalent larger antenna.

An interferometric image is created using the complex correlations of intensities obtained from all possible pair combinations in a detector array [13,14]. This is expressed as follows:

$$\sigma_E(\zeta,\eta) = \sum_{i=1}^{N(N-1)/2} A_i \cos(\Delta\phi_i + k(\mu_i\zeta + \nu_i\eta)), \quad (1)$$

where σ_E denotes the time average intensity of the source, ζ and η correspond to the coordinates of each pixel in the image, $\mu_i = x_n - x_m$ and $\nu_i = y_n - y_m$ are the x and y baselines provided by detector pairs (x_m, y_m) and (x_n, y_n), k is the wave number, and N is the number of detectors in the array. The summation is evaluated over all possible

combinations of detector pairs. A_i and $\Delta\phi_i$ correspond to the amplitude and phase terms of the correlated fields for the mn^{th} pair such that, $A_i = |E_m E_n^*|$ and $\Delta\phi_i = \phi_m - \phi_n$.

A. Target behind random media

Interferometric imaging has recently been applied to targets behind random media [15], and implemented on GPGPU platform [16]. A summary of these research outcomes are presented in this section. Furthermore, the same algorithm is implemented on an FPGA platform to gain insight into the differences in implementation on GPGPU versus FPGA platforms. We report on how the different computer architectures are utilized in the implementation, and their performance levels in terms of speed and efficiency.

Our model for interferometric imaging of targets behind random media consists of three major building blocks: (i) field calculation at the detector array (includes scattered field from the target and random medium), (ii) image construction, and (iii) image plotting. The field calculations include a direct scattering term from the target, a direct scattering term from the media as well as an indirect term scattered from the target to the medium before arriving at the detector array, see Fig. 3. The scattering field calculations from the random media uses the distorted Born approximation following the work reported earlier, [17]. Attenuation effects are accounted for as the fields travel through the random medium.
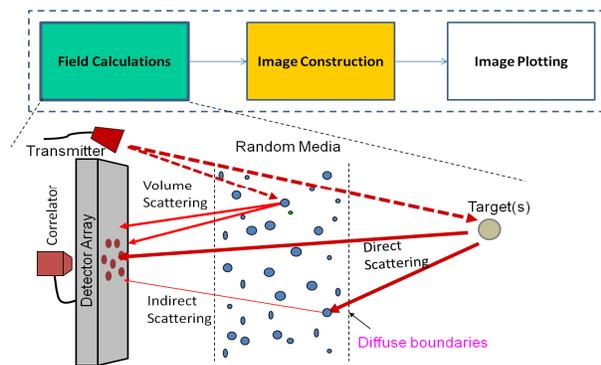
Fig. 3. Computational steps of the model.

B. Parallelization

The most time consuming computation in the model is the image construction, which consists of the summation of correlated fields for each pair of detectors as given in (1). Since the intensity for each pixel can be calculated independently, this part of the algorithm can benefit from parallel implementation. Therefore, the image construction was selected as the candidate for hardware implementation, while the rest of the algorithm is carried out on the CPU.

IV. HARDWARE IMPLEMENTATION

For comparison purposes, we start with a highly productive programming environment for both GPGPU and FPGA implementations to benefit from a fast learning curve. For the FPGA implementation, we employed the SRC-6 system and its proprietary Carte programming environment. For the GPGPU implementation, we utilized Jacket version 1.5.0 by AccelerEyes on an NVIDIA C1060 based system. All field calculations and image plotting are performed on the CPU. The image construction stage, which is implemented on hardware, is divided into four stages: (i) setup of the hardware, (ii) transfer of input parameters to the hardware, (iii) parallelized calculations, and (iv) transfer of results back to the CPU. The specifics of the implementation on these two systems are discussed in the following sub-sections.

A. Implementation on FPGA

The FPGA platform employed for the implementation is the SRC-6 scalable system which is a cluster-based system. The SRC-6 system includes two MAP (multi-adaptive processor) reconfigurable boards and each consists of two FPGAs, which can all be programmed simultaneously in one application. Each MAP has 24 MB of memory and the processing speed of FPGAs is 100 MHz. SRC's proprietary Carte-C programming environment is used in the code development [2].

In the FPGA implementation, the image is divided into two regions and run in parallel on a single FPGA. The computations and the summation given in (1) are pipelined such that the accumulation pipeline can start before all terms are computed. SRC's Carte development environment provides a relatively easy interface to implement code on the FPGA, while allowing the programmer some control over the data streaming and parallelization.

B. Implementation on GPGPU

The GPGPU platform used in the implementation is a single workstation which consists of a 16-core 2.67 GHz Intel Xeon CPU with four NVIDIA Tesla C1060 graphics processors with a total 16 GB of memory and runs at 1.3 GHz. The workstation operates on Microsoft Windows 7 Professional, and MATLAB version 7.7.0.471 (R2008b) is utilized along with Jacket version 1.5.0 by AccelerEyes. This version of Jacket uses NVIDIA's CUDA version 3.1. The architecture of the system is similar to the general architecture shown earlier in Fig. 2.

Jacket programming environment from Accelereyes is used in the implementation. This package can be easily integrated with MATLAB and provides a MATLAB-like environment to enable a fairly easy learning curve. However, this comes at the expense of limited control over how the data streaming and parallelization are implemented, resulting in reduced efficiency as compared to that of FPGAs.

V. PERFORMANCE ANALYSIS AND METRICS FOR EVALUATION

In our analysis, we assumed the fixed-time model (Gustafson's Law) [18] and used larger computational workloads (larger data/image size) while maintaining the same performance on larger configurations. Larger configurations utilize more processing units (FPGAs or GPUs) which results in increasing the overall system utilization.

The image construction is composed of 5 stages as depicted in Fig. 4. The algorithm first sets up the parallel environment, then transfers data from CPU memory space to accelerator memory space, processes data using accelerator(s), transfers the image data back to the CPU and finally releases the allocated hardware resources. The associated execution times for each of these stages are defined as shown in Fig. 4.
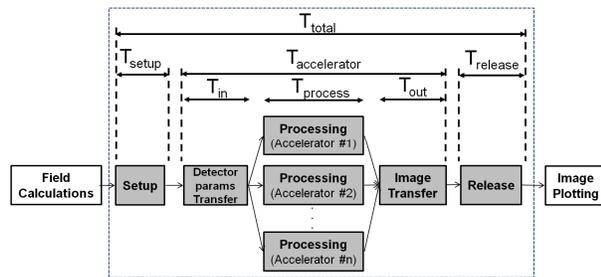
Fig. 4. Execution model and performance metrics.

Execution scenarios

Two scenarios were considered for hardware implementation, namely no-streaming and streaming scenarios. In the no-streaming scenario, operations are performed in a non-overlapped execution while in the streaming scenario operations are allowed to overlap, as shown in Figs. 5 and 6, respectively. These scenarios affect the total execution time as given in (2) and (3).

$$T_{total}^{no-streaming} = T_{setup} + T_{in} + T_{process} + T_{out} + T_{release}, \quad (2)$$

$$T_{total}^{streaming} = T_{setup} + T_{in} + \max(T_{process}, T_{out}) + T_{release}. \quad (3)$$
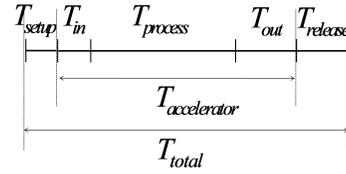
Fig. 5. No-streaming (non-overlapped execution).
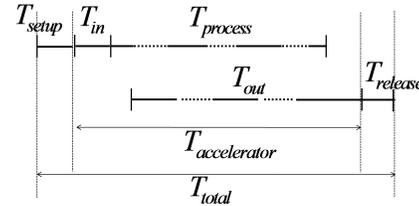
Fig. 6. Streaming (overlapped execution).

Metrics

Three metrics were investigated for performance comparison between the two platforms: (i) speedup, (ii) efficiency, and (iii) scalability.

The speedup is defined as the performance gain by using multiple hardware processing units (i.e. GPU, FPGA) in reference to a single CPU. This can be expressed as follows:

$$S(N_{PU}) = \frac{T_{total}^{CPU}(1, N_{PU}D_0)}{T_{total}^{PU}(N_{PU}, N_{PU}D_0)}, \quad (4)$$

where D_0 is the image size for a single processing unit, and N_{PU} is the number of processing units. $T_{total}^{CPU}(1, N_{PU}D_0)$ is the total execution time of a single CPU with image size of $N_{PU}D_0$, and $T_{total}^{PU}(N_{PU}, N_{PU}D_0)$ is the total execution time of multiple processing units with image size of $N_{PU}D_0$.

The hardware efficiency compares the execution time measured through experiments to

the expected/theoretical performance. This is expressed as:

$$E(N_{PU}, N_{PU}D_0) = \frac{T_{total}^{th}(1, D_0)}{T_{total}^{meas}(N_{PU}, N_{PU}D_0)} (\%), \quad (5)$$

where $T_{total}^{th}(1, D_0)$ is the theoretical total execution time of a single processing unit with image size D_0, and $T_{total}^{meas}(N_{PU}, N_{PU}D_0)$ is the measured total execution time of multiple processing units with image size of $N_{PU}D_0$. The theoretical execution time, $T_{total}^{th}(1, D_0)$ is calculated using (2) or (3) depending on the implemented execution scenario. Each term in (2) and (3) is calculated as discussed below.

The setup and release times represent the system overhead associated with setting up and releasing multiple processing units and therefore they are both equal to zero for a single GPU. However, the setup time for a single FPGA is non-zero, i.e. 65 ms, due to configuring the FPGA.

The input transfer time is the ratio of the total amount of input transfer data, D_{in}, to the bandwidth, B_{in}, for the input data transfer from the CPU's memory to accelerator's memory. The input data size, D_{in}, is calculated as:

$$D_{in} = N_{det-pairs}.N_{param}^{detector}.D_{bytes}, \quad (6)$$

where $N_{det-pairs}$ is the number of detector pairs, D_{bytes} is the number of precision bytes, and $N_{param}^{detector}$ is the number of detector parameters. The input bandwidth values are as shown in Table 1.

Table 1: Bandwidth parameters

(MB/s)	B_{in}	$B_{process}$	B_{out}
GPU	2,169	61,360	1,151
FPGA	1,415	4,800	1,260

The processing time is the ratio of the total amount of data for processing, $D_{process}$ to the processing throughput, $B_{process}$. $B_{process}$ is given in Table 1 and $D_{process}$ is calculated by equation (7):

$$D_{process} = N_{pixels}.N_{det-pairs}.(N_{param}^{detector} + N_{param}^{pixel}).D_{bytes}. \quad (7)$$

The output transfer time is the ratio of the total amount of output transfer data, D_{out} to the bandwidth for the output data transfer, B_{out}, from the accelerator's memory to the CPU's memory. The output transfer data size is calculated by

$$D_{out} = N_{pixels}.D_{bytes}, \quad (8)$$

where N_{pixels} is the number of pixels in the image, D_{bytes} is the number of precision bytes as before, and the bandwidth, B_{out} is given in Table 1.

Finally, the scalability factor, Ω, is defined as the normalized speedup. In other words, it is the performance achieved by multiple processing units as compared to the performance achieved by a single processing unit as given by equation (9).

$$\Omega(N_{PU}) = \frac{S(N_{PU})}{S(1)}. \quad (9)$$

VI. EXPERIMENTAL RESULTS

The image size for a single processing unit (i.e. GPU or FPGA) is chosen as $D_0 = 250,000$ pixels for comparison purposes. The implementation for both GPU and FPGA were parallelized using up to four processing units in each system. Therefore, as the number of processors was increased, proportionally more pixels were created in the image (following Gustafson's model). A comparison is provided in Fig. 7 in terms of the end-to-end computation time for each system, as well as the hardware-only time. Figure 7 also shows the speedup of total execution time of GPUs/FPGAs over the total execution time of a single CPU which are put in parentheses next to their corresponding execution times. It is observed for the unit data size implemented on a single processing unit, GPU outperforms the FPGA (speedup of 71 versus 12). As the image size increases, increasing the number of FPGAs compensates for added computational workload, and the total time remains constant. However, the GPU system shows signs of decreased efficiency with increased workload (image size). This can be better observed if one considers the ratio of the time it took to construct the image on the FPGA system versus the GPU system. It is observed that for the single accelerator case, the GPU system is approximately six times faster than the FPGA. As the number of accelerators is increased (proportionally with the pixels generated), the ratio drops to roughly two. This is due to the higher efficiency achieved by the FPGA implementation as well as the faster decreasing efficiency of the GPU implementation. The decreasing efficiency of the GPU is due to the overhead associated with the setup and merging of the results before transferring the output back to the CPU . The GPU requires more resources to be
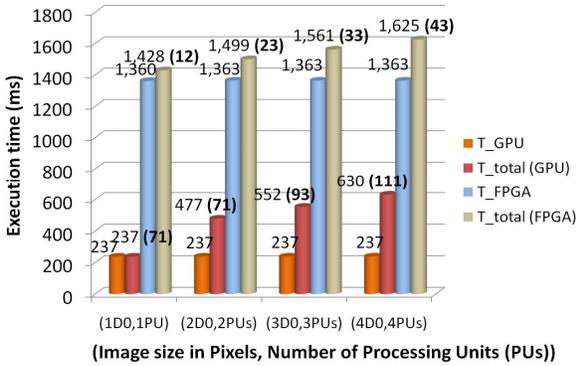
Fig. 7. Total execution time.

allocated and becomes less efficient.

The system efficiency was calculated by taking the ratio of the measured time to the theoretically expected time for each system to complete the calculations. The theoretical time calculations are based on the system parameters such as the bandwidth, processing speed, etc. The efficiencies for both implementations are depicted in Fig. 8, demonstrating that despite being slower than the GPUs, FPGA implementations show higher efficiency.
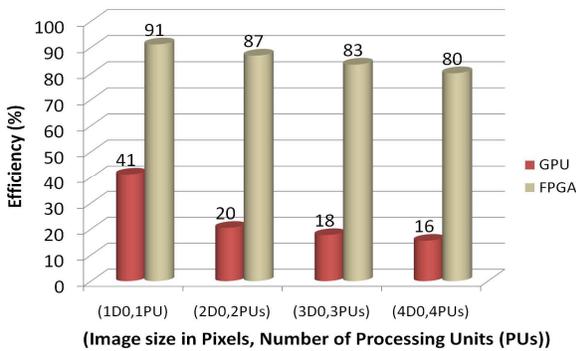
Fig. 8. Efficiency of FPGA vs. GPU.
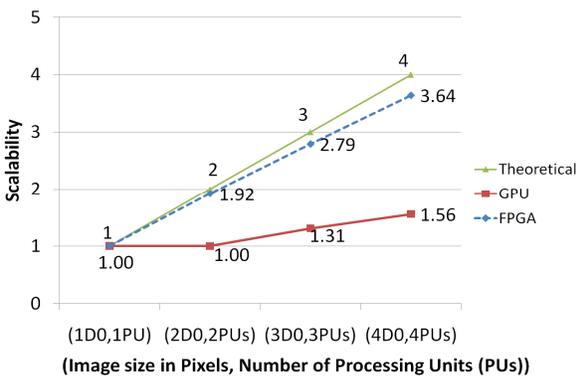
Fig. 9. Scalability of algorithm on FPGA vs. GPU.

In order to investigate the scalability, we compare the speedup scale with increasing image size as depicted in Fig. 9. It is observed that the FPGA computations when compared to GPU computations scaled more closely to the theoretical linear expectations.

VII. CONCLUSIONS

While the comparison of performance in terms of execution time shows that the GPU based implementation performs better than the FPGA, it should be noted that the FPGA system has significantly more limited resources in terms of clock speed (100 MHz vs. 1.3 GHz) and on board memory (48 MB total vs. 16 GB total). From that perspective, the FPGA implementation proves to be significantly more efficient than the GPU implementation. This is shown in the comparison of the system efficiencies, where the FPGA implementation achieved about 90% efficiency as opposed to the 40% observed for the GPU implementation. It is worthwhile noting that the SRC Carte programming environment allows the user flexibility in controlling the data utilization in terms of pipelining and parallelization, whereas with Jacket environment the user is oblivious to how the GPU is controlled. This comes at the expense of ease of use, as Jacket is very much like MATLAB and very easy to use, while the learning curve in Carte is steeper. The tradeoff between productivity versus efficiency is one of the key features for hardware accelerated computing on these platforms. While FPGAs can be programmed to perform a specific task in a highly efficient way, they currently lack the wide commercial support enjoyed by GPGPUs. Consequently, they require a good knowledge of hardware and ability to program in hardware languages such as VHDL and/or Verilog to program them most efficiently.

REFERENCES

[1] O. Kilic, E. El-Araby, and V. Dang, "Hardware Accelerated Computing for Electromagnetics Applications," *International Workshop on Computational Electromagnetics (CEM'11)*, Izmir, Turkey, August 2011.

[2] SRC Computers, Inc., "SRC Carte^TM C Programming Environment v2.2," *(SRC-007-18)*, Aug. 2006.

[3] Silicon Graphics Inc., "Reconfigurable Application-Specific Computing User's Guide," *(007-4718-005)*, Jan. 2007.

[4] Cray Inc., "Cray XD1^TM FPGA Development," *(S-6400-14)*, 2006.

[5] E. El-Araby, M. Taher, M. Abouellail, T. El-Ghazawi, and G. B. Newby, "Comparative Analysis of High Level Programming for Reconfigurable Computers: Methodology and Empirical Study," *III Southern Conference on Programmable Logic (SPL2007)*, Mar del Plata, Argentina, February, 2007.

[6] O. Kilic, "FPGA Accelerated Phased Array Design using the Ant Colony Optimization," *Applied Comp. Electromag. Soc. Journal*, Vol. 20, No. 1, pp. 23-30, February 2010.

[7] E. El-Araby, P. Nosum, and T. El-Ghazawi, "Productivity of High-Level Languages on Reconfigurable Computers: An HPC Perspective," *IEEE International Conference on Field-Programmable Technology (FPT 2007)*, Japan, December, 2007.

[8] El-Araby, S. G. Merchant, and T. El-Ghazawi, "A Framework for Evaluating High-Level Design Methodologies for High-Performance Reconfigurable Computers," *IEEE Trans. Parallel Distrib. Syst.,* vol. 22, pp. 33-45, Jan. 2011.

[9] V. V. Kindratenko, J. J. Enos, G. Shi, M. T. Showerman, G. W. Arnold, J. E. Stone, J. C. Phillips, and W. Hwu, "GPU Clusters for High-Performance Computing," *Proc. Workshop on Parallel Programming on Accelerator Clusters* , 2009.

[10] F. Darema, "The SPMD Model: Past, Present and Future," *Proc. 8th European PVM/MPI Users' Group Meeting on Recent Advances in Parallel Virtual Machine and Message Passing interface,* Lecture Notes In Computer Science, vol. 2131/2001, pp. 1, Sept. 2001.

[11] M. Ryle, "A New Radio Interferometer and Its Applications to the Observation of Weak Radio Stars," *Proc. Roy. Soc. London Ser. A,* vol. 211, pp. 351-375, 1952.

[12] R. N. Bracewell, "Radio Interferometry of Discrete Sources," *Proc. IRE,* vol. 46, pp. 97-105, 1958.

[13] J. F. Frederici, D. Gray, B. Schulkin, F. Huang, H. Altan, R. Barat, and D. Zimdars, "Terahertz Imaging Using an Interferometric Array," *Appl. Phys. Lett.,* vol. 83, p. 2477, 2004.

[14] K. P. Walsh, B. Schulkin, D. Gary, J. F. Federici, R. Barat, and D. Zimdars, "Terahertz Near-Field Interferometric and Synthetic Aperture Imaging," *Proc. SPIE,* vol. 5411, p. 9, 2003.

[15] O. Kilic and A. Smith, "Imaging Through Random Media," *European Conference on Antennas and Propagation* (*EuCAP),* Rome, Italy, April 2011.

[16] O. Kilic, A. Smith, E. El-Araby, and V. Dang, "Interferometric Imaging Through Random Media using GPU," *The Applied Computational Electromagnetics Society (ACES),* Williamsburg, VA, USA, April 2011.

[17] O. Kilic and R. H. Lang, "Scattering of a Pulsed Beam by a Random Medium Over Ground," *J. of Electromagnetic Waves and Appl.,* vol. 15, no. 4, pp. 481-516, 2001.

[18] K. Hwang, and Z. Xu, "Scalable Parallel Computing: Technology, Architecture, Programming," McGraw-Hill, 1998.

Dr. Esam El-Araby received his B.Sc. degree in Electronics and Telecommunications Engineering and his M.Sc. degree in Automatic Control and Computer Engineering from Assiut University, Egypt, in 1991 and 1997, respectively. He received his M.Sc. and Ph.D. degrees in Computer Engineering from the George Washington University (GWU), USA, in 2005 and 2010, respectively. Dr. El-Araby joined the Catholic University of America (CUA) as an Assistant Professor in the Department of Electrical Engineering and Computer Science in 2010. He is the founder and director of the Heterogeneous and Biologically-inspired Architectures (HEBA) laboratory at CUA. Prior to that, he worked at the High Performance Computing Laboratory (HPCL) at GWU as well as the NSF Center for High-Performance Reconfigurable Computing (NSF-CHREC). His research interests include computer architecture, hybrid/heterogeneous architectures, hardware acceleration, reconfigurable computing, embedded systems, evolvable hardware, performance evaluation, and digital signal/image processing and remote sensing.

Dr. Ozlem Kilic joined the Catholic University of America as an Assistant Professor in the Department of Electrical Engineering and Computer Science in 2005. Prior to that, she was an Electronics Engineer at the U.S. Army Research Laboratory, Adelphi MD where she managed Small Business Innovative Research (SBIR) Programs for the development of hybrid numerical electromagnetic tools to analyze and design electrically large structures, such as the Rotman lens. She has also designed, fabricated and tested various prototypes. Dr. Kilic has over five years of industry experience at COMSAT Laboratories as a Senior Engineer and Program Manager with specialization in satellite communications, link modeling and analysis, and modeling, design and test of phased arrays and reflector antennas for satellite communications system. Her research interests include numerical electromagnetics, antennas, wave propagation, satellite communications systems, and microwave remote sensing.

Vinh Dang received his B.Sc. degree in Electronics and Telecommunications Engineering from the Posts and Telecommunications Institute of Technology, Vietnam, in 2003. He received his M.Eng. degree in Electrical Engineering from the University of Technology, Vietnam, in 2006. He is currently a Ph.D. candidate at the Department of Electrical Engineering and Computer Science, the Catholic University of America (CUA), USA. During the period from 2008 to 2010, he was a Lecturer at the School of Electrical Engineering, International University, Vietnam. Since 2010, he has been a Research Assistant in the Department of Electrical Engineering and Computer Science at CUA. His current research interests include high performance computing, embedded systems, biomedical image processing, and remote sensing.

Rigorous Analysis of Double-Negative Materials with the Multilevel Fast Multipole Algorithm

Özgür Ergül[1] and Levent Gürel[2,3]

[1]Department of Mathematics and Statistics
University of Strathclyde, G11XH, Glasgow, UK
ozgur.ergul@strath.ac.uk

[2]Department of Electrical and Electronics Engineering
[3]Computational Electromagnetics Research Center (BiLCEM)
Bilkent University, TR-06800, Bilkent, Ankara, Turkey
lgurel@bilkent.edu.tr

Abstract — We present rigorous analysis of double-negative materials (DNMs) with surface integral equations and the multilevel fast multipole algorithm (MLFMA). Accuracy and efficiency of numerical solutions are investigated when DNMs are formulated with two recently developed formulations, i.e., the combined tangential formulation (CTF) and the electric and magnetic current combined-field integral equation (JMCFIE). Simulation results on canonical objects are consistent with previous results in the literature on ordinary objects. MLFMA is also parallelized to solve extremely large electromagnetics problems involving DNMs.

Index Terms — Double-negative materials, metamaterials, multilevel fast multipole algorithm, surface integral equations.

I. INTRODUCTION

Double-negative materials (DNMs) are commonly used as simplified models of metamaterials at resonance frequencies [1]. Specifically, a metamaterial structure at a resonance frequency can be modeled (homogenized [2]) as a homogeneous object with negative permittivity and permeability. Using the equivalence principle, a DNM can be formulated with surface integral equations, which can be discretized and solved numerically. Recently, various surface formulations, such as the Poggio-Miller-Chang-Harrington-Wu-Tsai (PMCHWT) formulation [3], the Müller formulation [4], and the electric and magnetic current combined-field integral equation (JMCFIE) [5], have been used to analyze DNMs [6],[7]. It has been shown that homogenization can provide fast analysis of metamaterial structures before their detailed analysis via full-wave solvers [8].

Although electromagnetics problems obtained via homogenization are relatively easy compared to the original problems, their efficient solutions may not be trivial. Surface integral equations require only the discretization of boundaries, but the resulting matrix equations can be very large because realistic metamaterials are usually large with respect to wavelength. Hence, fast and efficient methods, such as the multilevel fast multipole algorithm (MLFMA) [9]–[13], are required for the solution of large metamaterial problems, even when they are homogenized. Applying MLFMA to homogeneous materials, including DNMs, is straightforward, but the number of iterations must be small for efficient solutions, and thus the choice of the surface formulation is critical for efficient solutions.

In this paper, we present iterative solutions of DNMs using MLFMA. Problems are formulated with two recently developed formulations, namely, the combined tangential formulation (CTF) [14] and JMCFIE [5], and discretized with the Rao-Wilton-Glisson (RWG) functions [15]. Accuracy and efficiency of numerical solutions are investi-

gated on canonical problems involving the sphere geometry. We show that the conventional JMCFIE (with $\alpha = 0.5$ combination parameter) provides efficient solutions but relatively inaccurate results. In addition, accuracy of simulations can significantly be improved using CTF, instead of JMCFIE. These observations are consistent with earlier results obtained for ordinary materials [16]. We also show that the combination parameter of JMCFIE can be increased towards unity to improve the accuracy of JMCFIE, without sacrificing the efficiency. Finally, MLFMA is parallelized using the hierarchical partitioning strategy [17] to solve very large problems involving DNMs. The resulting implementation based on JMCFIE (with high combination parameter) and parallel MLFMA seems to be a suitable solver for the fast and accurate analysis of DNMs.

II. NUMERICAL SOLUTIONS OF SURFACE FORMULATIONS

For homogeneous penetrable objects, discretizations of surface formulations lead to $2N \times 2N$ dense matrix equations in the form of

$$\begin{bmatrix} \bar{\mathbf{Z}}^{(11)} & \bar{\mathbf{Z}}^{(12)} \\ \bar{\mathbf{Z}}^{(21)} & \bar{\mathbf{Z}}^{(22)} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{v}^{(1)} \\ \mathbf{v}^{(2)} \end{bmatrix}. \quad (1)$$

Using JMCFIE and a Galerkin discretization,

$$\bar{\mathbf{Z}}^{(ab)}[m,n] = \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \mathcal{Z}^{(ab)}\{\mathbf{b}_n\}(\mathbf{r}) \quad (2)$$

for $a = 1, 2$ and $b = 1, 2$, where $\mathbf{t}_m$ and $\mathbf{b}_n$ represent the testing and basis functions with spatial supports of S_m and S_n, respectively, for $m, n = 1, 2, \ldots, N$. The combined operators are defined as

$$\mathcal{Z}^{(11)} = \mathcal{Z}^{(22)} = -\alpha \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times (\mathcal{T}_o + \mathcal{T}_i)$$
$$+ (1-\alpha)\hat{\mathbf{n}} \times (\mathcal{K}_o - \mathcal{K}_i) - (1-\alpha)\mathcal{I} \quad (3)$$
$$\mathcal{Z}^{(12)} = (1-\alpha)\hat{\mathbf{n}} \times (\eta_o^{-1}\mathcal{T}_o - \eta_i^{-1}\mathcal{T}_i)$$
$$+ \alpha \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times (\eta_o^{-1}\mathcal{K}_o + \eta_i^{-1}\mathcal{K}_i)$$
$$- \frac{1}{2}\alpha(\eta_o^{-1} - \eta_i^{-1})\hat{\mathbf{n}} \times \mathcal{I} \quad (4)$$
$$\mathcal{Z}^{(21)} = -(1-\alpha)\hat{\mathbf{n}} \times (\eta_o\mathcal{T}_o - \eta_i\mathcal{T}_i)$$
$$- \alpha \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times (\eta_o\mathcal{K}_o + \eta_i\mathcal{K}_i)$$
$$+ \frac{1}{2}\alpha(\eta_o - \eta_i)\hat{\mathbf{n}} \times \mathcal{I}, \quad (5)$$

where $\alpha \in [0, 1]$ is the combination parameter, $\hat{\mathbf{n}}$ is the unit normal vector at the observation point $\mathbf{r}$, and $\eta_u = \sqrt{\mu_u}/\sqrt{\epsilon_u}$ is the wave impedance in the outer ($u = o$) and inner ($u = i$) media. The integro-differential operators are defined as

$$\mathcal{T}_u\{\mathbf{b}_n\}(\mathbf{r}) = \frac{i}{k_u} \int_{S_n} d\mathbf{r}' \nabla' \cdot \mathbf{b}_n(\mathbf{r}') \nabla g_u(\mathbf{r}, \mathbf{r}')$$
$$+ ik_u \int_{S_n} d\mathbf{r}' \mathbf{b}_n(\mathbf{r}') g_u(\mathbf{r}, \mathbf{r}') \quad (6)$$
$$\mathcal{K}_u\{\mathbf{b}_n\}(\mathbf{r}) = \int_{PV,S_n} d\mathbf{r}' \mathbf{b}_n(\mathbf{r}') \times \nabla' g_u(\mathbf{r}, \mathbf{r}'), \quad (7)$$

where PV indicates the principal value of the integral, $k_u = \omega\sqrt{\epsilon_u}\sqrt{\mu_u}$ is the wavenumber, and

$$g_u(\mathbf{r}, \mathbf{r}') = \frac{\exp(ik_u|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|} \quad (8)$$

denotes the homogeneous-space Green's function in the phasor domain using the $e^{-i\omega t}$ time dependence. The elements of the right-hand-side vectors in (1) are derived similarly as

$$v_m^{(a)} = \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \boldsymbol{\psi}^{(a)}(\mathbf{r}) \quad (9)$$

for $a = 1, 2$, where

$$\boldsymbol{\psi}^{(1)}(\mathbf{r}) = -(1-\alpha)\hat{\mathbf{n}} \times \mathbf{H}^{inc}(\mathbf{r})$$
$$+ \alpha\eta_o^{-1}\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{E}^{inc}(\mathbf{r}) \quad (10)$$
$$\boldsymbol{\psi}^{(2)}(\mathbf{r}) = (1-\alpha)\hat{\mathbf{n}} \times \mathbf{E}^{inc}(\mathbf{r})$$
$$+ \alpha\eta_o\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{H}^{inc}(\mathbf{r}). \quad (11)$$

In (10) and (11), $\mathbf{E}^{inc}$ and $\mathbf{H}^{inc}$ represent the incident electric and magnetic fields created by external sources located in the outer medium.

JMCFIE is a mixed formulation involving directly and rotationally tested electromagnetic fields. Using a Galerkin discretization (using the same set of the RWG functions as the basis and testing functions), JMCFIE involves a well-tested identity operator, i.e.,

$$\bar{\mathbf{I}}_{mn} = \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{b}_n(\mathbf{r})$$
$$= \int_{S_m} d\mathbf{r}\, \mathbf{t}_m(\mathbf{r}) \cdot \int_{S_n} \delta(\mathbf{r}, \mathbf{r}')\mathbf{b}_n(\mathbf{r}'), \quad (12)$$

which is a major error source for low-order discretizations [18]. CTF can be seen as a special case of JMCFIE and it is obtained by setting

$\alpha = 1$ in (3)–(5), (10), and (11). Note that well-tested identity operators disappear in CTF; this explains why it is more accurate than JMCFIE.

Employing the conventional formulations, such as PMCHWT and JMCFIE, for DNMs is extensively discussed in [6],[7]. Using $k_u = \omega\sqrt{\epsilon_u}\sqrt{\mu_u}$ and $\eta_u = \sqrt{\mu_u}/\sqrt{\epsilon_u}$ leads to negative wavenumber and positive wave impedance when the permittivity (ϵ_u) and permeability (μ_u) are negative. In order to construct a tree structure for a DNM, we use the absolute value of k_u in the excess bandwidth formula, i.e.,

$$\tau_{l,u} \approx 1.73|k_u|a_l + 2.16(d_0)^{2/3}(|k_u|a_l)^{1/3}, \quad (13)$$

to determine truncation numbers $\tau_{l,u}$ and samples on the unit sphere. In (13), a_l is the box size at level l and d_0 is the number of accurate digits for the far-field interactions.

III. NUMERICAL RESULTS

In order to test the accuracy and efficiency of solutions of DNMs with MLFMA, we consider increasingly large scattering problems involving a sphere of radius 0.3 m. The object is located in free space and illuminated by plane waves at various frequencies. Problems are formulated with CTF and JMCFIE and discretized with the RWG functions on $\lambda_o/10$ triangles, where λ_o is the wavelength in the host medium (free space). Both near-field and far-field interactions are calculated with maximum 1% error. Solutions are performed using the biconjugate-gradient-stabilized (BiCGStab) algorithm [19] accelerated with MLFMA. Iterative convergences are also accelerated with the four-partition block-diagonal preconditioner (4PBDP) [16] for the conventional JMCFIE ($\alpha = 0.5$).

Fig. 1 presents the solution of a small scattering problem involving a sphere of radius 0.3 m at 500 MHz. Both the relative permittivity and permeability of the sphere are selected as -2.0. The sphere is illuminated by a plane wave from the top and the problem is formulated with CTF. For numerical solutions, the problem is discretized with 1860 unknowns. Fig. 1 depicts the total electric field in the vicinity of the sphere on the E-plane for the inner and outer problems. The maximum electric field value is normalized to 0 dB. For the inner/outer problem, the equivalent currents
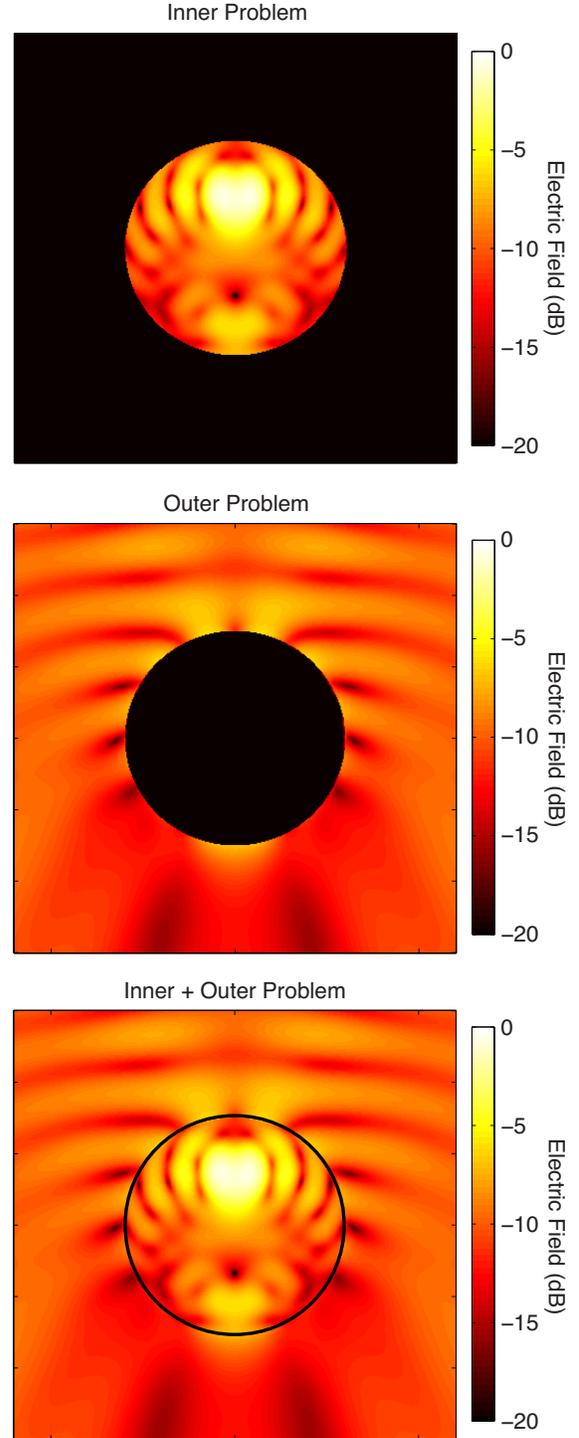
Fig. 1. Solution of a scattering problem involving a sphere of radius 0.3 m at 500 MHz. Both the relative permittivity and permeability of the sphere are -2.0.
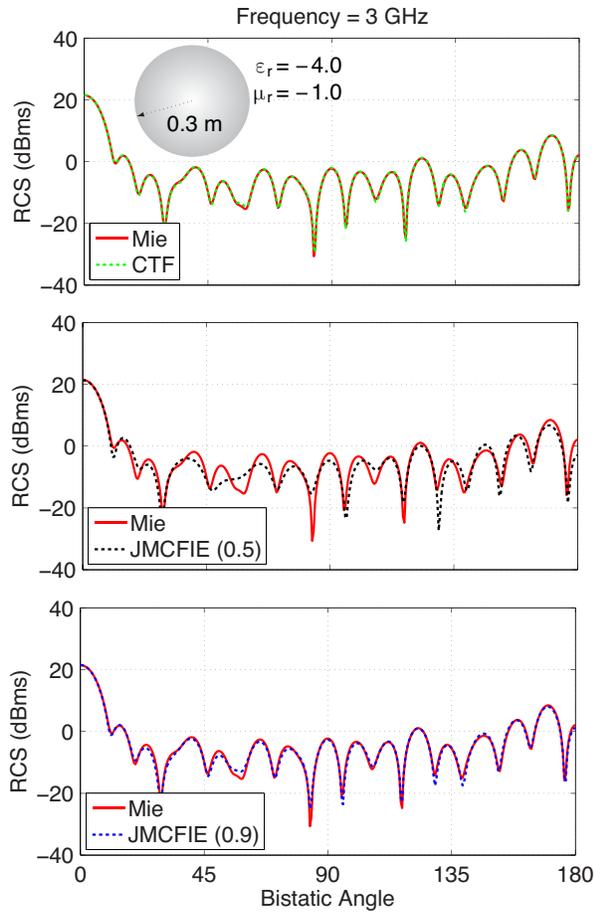
Fig. 2.   Solutions of a scattering problem involving a sphere of radius 0.3 m at 3 GHz. The relative permittivity and permeability of the sphere are -4.0 and -1.0, respectively.

Fig. 3.   The relative error and the number of BiCGStab iterations (for 10^{-3} residual error) required in numerical solutions of scattering problems involving a sphere of radius 0.3 m at 3 GHz.

provided by MLFMA are allowed to radiate into a homogeneous space with the electrical parameters of the inner/outer medium assumed everywhere. Hence, for the inner/outer problem, any radiation outside/inside the sphere can be interpreted as numerical error. It can be observed that these unwanted radiations are below -20 dB, verifying the high accuracy of the solution. As also depicted in Fig. 1, the complete plot can be obtained by superimposing the plots for the inner and outer problems. It is remarkable that field values become maximum in the upper part of the sphere as a result of the negative refractive index of the object.

Fig. 2 presents the solution of a scattering problem involving a sphere of radius 0.3 m at 3 GHz. In this problem, the relative permittivity and permeability of the sphere are selected as -4.0 and -1.0, respectively. For numerical solutions, the problem is discretized with 65,724 unknowns. Fig. 2 depicts the radar cross section (RCS) on the E-plane as a function of the bistatic observation angle from $0°$ to $180°$, where $0°$ and $180°$ correspond to the forward-scattering and backscattering directions, respectively. Computational values obtained with CTF, the conventional JMCFIE ($\alpha = 0.5$), and JMCFIE with a high combination parameter ($\alpha = 0.9$) are compared to the analytical Mie-series results. It can be observed that CTF results agree very well with the analytical results. However, the same level of accuracy is not obtained with the conventional JMCFIE. In addition, as also depicted in Fig. 2, increasing the combination parameter to 0.9 significantly improves the accuracy.

For a more quantitative comparison of the formulations, Fig. 3 presents the results of scattering problems involving a sphere with different material properties. A sphere of radius 0.3 m is again investigated at 3 GHz and discretized with 65,724
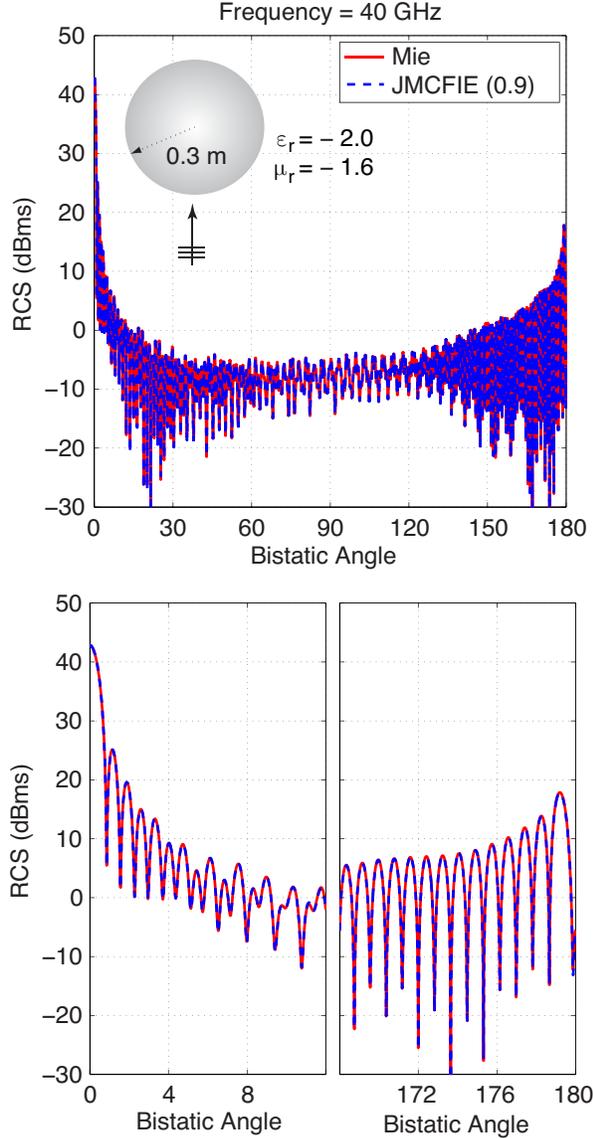
Fig. 4. Solution of a scattering problem involving a sphere of 0.3 m at 40 GHz. The relative permittivity and permeability of the sphere are -2.0 and -1.6, respectively.

unknowns. The first plot of Fig. 3 depicts the relative error in the far-zone electric field obtained with different formulations. To find the relative error, the co-polar electric field in the far-zone on the E-plane is sampled at $\pi/360$ intervals, i.e., we compute

$$f[n] = \lim_{r \to \infty} \{ r E_\varphi(r, \varphi[n]) \}, \qquad (14)$$

where $\varphi[n] = (n-1)\pi/360$ for $n = 1, 2, \ldots, 361$. Then, the error is defined as

$$\Delta = \frac{\|\boldsymbol{f}_C - \boldsymbol{f}_A\|}{\|\boldsymbol{f}_A\|}, \qquad (15)$$

where $\|\cdot\|$ represents the 2-norm and $\boldsymbol{f}_C$ and $\boldsymbol{f}_A$ are vectors of 361 elements containing the computational and analytical values, respectively. It can be observed that the relative error for the conventional JMCFIE ($\alpha = 0.5$) is generally higher than 1%, which may not be acceptable. As also shown in the same plot, the accuracy is significantly improved by using CTF or increasing the amount of CTF in JMCFIE, i.e., using JMCFIE with $\alpha = 0.9$, instead of the conventional JMCFIE. As complementary data, the second plot of Fig. 3 depicts the number of BiCGStab iterations for 0.001 residual error. Iterative solutions of the conventional JMCFIE ($\alpha = 0.5$) are accelerated with 4PBDP; but this preconditioner is not useful for CTF and JMCFIE with $\alpha = 0.9$. It can be observed that JMCFIE (both with $\alpha = 0.5$ and with $\alpha = 0.9$) provides very efficient solutions, compared to CTF. Considering the results in both plots of Fig. 3, JMCFIE with $\alpha = 0.9$ seems to be a good choice for efficient and accurate solutions.

Finally, Fig. 4 presents the solution of a large scattering problem involving a sphere of radius 0.3 m at 40 GHz. The relative permittivity and permeability of the sphere are -2.0 and -1.6, respectively. The problem is formulated with JMCFIE using $\alpha = 0.9$ and discretized with 11,702,832 unknowns. MLFMA is parallelized into 64 processes on a cluster of Intel Xeon Nehalem quad-core processors with 2.80 GHz clock rate. The total time including the setup and 176 iterations (for 0.005 residual error) is approximately 10 hours. Fig. 4 depicts the bistatic RCS values on the z-x plane as a function of the bistatic angle θ from $0°$ to $180°$. RCS values around the forward-scattering ($0°$) and backscattering ($180°$) directions are also focused in separate plots. It can be observed that the computational values obtained by using JMCFIE and parallel MLFMA agree very well with the analytical results. For this large-scale problem, the relative error in (15) is found to be 0.21%.

IV. CONCLUSIONS

This paper presents the analysis of DNMs with surface integral equations and MLFMA. Numerical results obtained with conventional formulations are in agreement with previous results obtained for ordinary materials. Numerical experiments on canonical objects show that JMCFIE with $\alpha = 0.9$ is a good choice for efficient and accurate solutions.

REFERENCES

[1] G. Monti and L. Tarricone, "Dispersion analysis of a negative group velocity medium with MATLAB," *ACES J.*, vol. 24, no. 5, pp. 478–486, Oct. 2009.

[2] H. Kettunen, J. Qi, H. Wallén, and A. Sihvola, "Homogenization of thin dielectric composite slabs: techniques and limitations," *ACES J.*, vol. 26, no. 3, pp. 179–187, Mar. 2011.

[3] A. J. Poggio and E. K. Miller, "Integral equation solutions of three-dimensional scattering problems," in *Computer Techniques for Electromagnetics*, R. Mittra, Ed. Oxford: Pergamon Press, 1973, Chap. 4.

[4] C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves*. New York: Springer, 1969.

[5] P. Ylä-Oijala and M. Taskinen, "Application of combined field integral equation for electromagnetic scattering by dielectric and composite objects," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 3, pp. 1168–1173, Mar. 2005.

[6] Y. A. Liu and W. C. Chew, "Stability of surface integral equation for left-handed materials," *IET Microw. Antennas Propag.*, vol. 1, no. 1, pp. 84–89, Mar. 2007.

[7] J. Rivero, J. M. Taboada, L. Landesa, F. Obelleiro, and I. Garcia-Tunon, "Surface integral equation formulation for the analysis of left-handed metamaterials," *Opt. Express*, vol. 18, no. 15, pp. 15876–15886, Jul. 2010.

[8] L. Gürel, Ö. Ergül, A. Ünal, and T. Malas, "Fast and accurate analysis of large metamaterial structures using the multilevel fast multipole algorithm," *Prog. Electromagn. Res.*, vol. 95, pp. 179–198, 2009.

[9] J. Song, C.-C. Lu, and W. C. Chew, "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects," *IEEE Trans. Antennas Propag.*, vol. 45, no. 10, pp. 1488–1493, Oct. 1997.

[10] J. Lee, J. Zhang, and C.-C. Lu, "Performance of preconditioned Krylov iterative methods for solving hybrid integral equations in electromagnetics," *ACES J.*, vol. 18, no. 4, pp. 54–61, Nov. 2003.

[11] R. J. Burkholder, P. H. Pathak, K. Sertel, R. J. Marhefka, and J. L. Volakis, "A hybrid framework for antenna/platform analysis," *ACES J.*, vol. 21, no. 3, pp. 177–195, Nov. 2006.

[12] M. Vikram and B. Shanker, "An incomplete review of fast multipole methods–from static to wideband–as applied to problems in computational electromagnetics," *ACES J.*, vol. 24, no. 2, pp. 79–108, Apr. 2009.

[13] H. Fangjing, N. Zaiping, and H. Jun, "An efficient parallel multilevel fast multipole algorithm for large-scale scattering problems," *ACES J.*, vol. 25, no. 4, pp. 381–387, Apr. 2010.

[14] P. Ylä-Oijala, M. Taskinen, and S. Järvenpää, "Surface integral equation formulations for solving electromagnetic scattering problems with iterative methods," *Radio Sci.*, vol. 40, RS6002, doi:10.1029/2004RS003169, Nov. 2005.

[15] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 409–418, May 1982.

[16] Ö. Ergül and L. Gürel, "Comparison of integral-equation formulations for the fast

and accurate solution of scattering problems involving dielectric objects with the multi-level fast multipole algorithm," *IEEE Trans. Antennas Propag.*, vol. 57, no. 1, pp. 176–187, Jan. 2009.

[17] Ö. Ergül and L. Gürel, "A hierarchical partitioning strategy for an efficient parallelization of the multilevel fast multipole algorithm," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1740–1750, Jun. 2009.

[18] Ö. Ergül and L. Gürel, "Discretization error due to the identity operator in surface integral equations," *Comput. Phys. Comm.*, vol. 180, no. 10, pp. 1746–1752, Oct. 2009.

[19] H. van der Vorst, "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.*, vol. 13, no. 2, pp. 631–644, Mar. 1992.

Özgür Ergül received B.Sc., M.S., and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 2001, 2003, and 2009, respectively, all in electrical and electronics engineering. He is currently a lecturer in the Department of Mathematics and Statistics at the University of Strathclyde, Glasgow, UK. He is also a lecturer of the Centre for Numerical Algorithms and Intelligent Software (NAIS), an EPSRC/SFC funded centre.

Dr. Ergül served as a Teaching and Research Assistant in the Department of Electrical and Electronics Engineering at Bilkent University from 2001 to 2009. He was also affiliated with the Computational Electromagnetics Group at Bilkent University from 2000 to 2005 and with the Computational Electromagnetics Research Center (BiLCEM) from 2005 to 2009. His research interests include fast and accurate algorithms for the solution of electromagnetics problems involving large and complicated structures, integral equations, parallel programming, iterative methods, and high-performance computing.

Dr. Ergül is a recipient of the 2007 IEEE Antennas and Propagation Society Graduate Fellowship, the 2007 Leopold B. Felsen Award for Excellence in Electrodynamics, the 2010 Serhat Ozyar Young Scientist of the Year Award, and the 2011 URSI Young Scientists Award. He is the coauthor of more than 140 journal and conference papers.

Levent Gürel (Fellow of ACES, IEEE, and EMA) is the Director of the Computational Electromagnetics Research Center (BiLCEM) at Bilkent University, Ankara, Turkey. He received the B.Sc. degree from the Middle East Technical University (METU), Ankara, Turkey, in 1986, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign (UIUC) in 1988 and 1991, respectively, in electrical and computer engineering. He joined the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, in 1991, where he worked as a Research Staff Member. Since 1994, he has been a faculty member in the Department of Electrical and Electronics Engineering of the Bilkent University, Ankara, where he is currently a Professor. He was a Visiting Associate Professor at the Center for Computational Electromagnetics (CCEM) of the UIUC for one semester in 1997. He returned to the UIUC as a Visiting Professor in 2003–2005, and as an Adjunct Professor for several years after 2005. He founded the Computational Electromagnetics Research Center (BiLCEM) at Bilkent University in 2005. Prof. Gürel's research interests include the development of fast algorithms for computational electromagnetics (CEM) and the application thereof to scattering and radiation problems involving large and complicated scatterers, antennas and radars, optical and imaging systems, nanostructures, and metamaterials. Since 2006, his research group has been breaking several world records by solving extremely large integral-equation problems.

Among the recognitions of Prof. Gürel's accomplishments, the two prestigious awards from the

Turkish Academy of Sciences (TUBA) in 2002 and the Scientific and Technical Research Council of Turkey (TUBITAK) in 2003 are the most notable. He is currently serving as an associate editor for Radio Science, IEEE Antennas and Wireless Propagation Letters, Journal of Electromagnetic Waves and Applications (JEMWA), and Progress in Electromagnetics Research (PIER). He is a member of the USNC of the International Union of Radio Science (URSI) and the Chairman of Commission E (Electromagnetic Noise and Interference) of URSI Turkey National Committee. He served as a member of the General Assembly of the European Microwave Association (EuMA) during 2006–2008. Prof. Gürel served as the Chairman of the AP/MTT/ED/EMC Chapter of the IEEE Turkey Section in 2000–2003. He founded the IEEE EMC Chapter in Turkey in 2000. He served as the Cochairman of the 2003 IEEE International Symposium on Electromagnetic Compatibility. Prof. Gürel is a member of the ACES Board of Directors and the Guest Editor for a special issue of the ACES Journal. He is the organizer and General Chair of the CEM'07, CEM'09, and CEM'11 Computational Electromagnetics International Workshops held in 2007–2011 in Izmir, Turkey. He is named an IEEE Distinguished Lecturer for 2011–2013 and invited to address the 2011 ACES Conference as a Plenary Speaker.

Analysis of Electromagnetic Wave Propagation
in Frequency Bands of Nonlinear Metamaterials

Nikolaos L. Tsitsas

Department of Informatics,
Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
ntsitsas@csd.auth.gr

Abstract— Electromagnetic wave propagation phenomena in nonlinear metamaterials are investigated for waves propagating either in the left-handed frequency band or in the frequency band gaps. In the left-handed band, we implement directly the reductive perturbation method to Faraday's and Ampére's laws and derive a second-and a third-order nonlinear Schrödinger (NLS) equation, describing solitons of moderate and ultra-short pulse widths, respectively. Then, we find necessary conditions and derive exact bright and dark soliton solutions of these equations. On the other hand, in the frequency band gaps with negative linear effective permittivity and positive permeability (where linear electromagnetic waves are evanescent), we derive two short-pulse equations (SPEs) for the high- and low-frequency band gaps. The structure of the SPEs solutions is discussed, and connections with the NLS soliton solutions are presented. Numerical simulations of the SPEs solutions are included and compared with those of the reduced wave equations. Directions towards the modelling of wave propagation in nonlinear chiral metamaterials are pointed out.

Index Terms— Frequency band gaps, left-handed media, negative refractive index, nonlinear metamaterials, nonlinear Schrödinger equations, ultra-short pulses.

I. INTRODUCTION

The remarkable electromagnetic (EM) properties and numerous potential applications of metamaterials (MMs) have been meticulously documented in the literature (see e.g. the classic books [1,2]. By considering a Drude model for the effective permittivity ϵ and a Lorentz model for the magnetic permeability μ of the MM, it is seen that there exist three different types of frequency bands in the MM: (i) those displaying a right-handed (RH) behavior characterized by $\epsilon > 0$, $\mu > 0$, or (ii) a left-handed (LH) behavior characterized by $\epsilon < 0$, $\mu < 0$, and thus exhibiting negative refractive index, or (iii) frequency band gaps, namely domains where linear EM waves are evanescent, e.g., for $\epsilon < 0$ and $\mu > 0$. Each of the above bands exhibits significantly different properties referring to the associated propagation and localization of EM waves. The RH band has been extensively analyzed; here, we focus on waves propagating either in the LH regime or in the band gaps.

Concerning first LH MMs, they exhibit negative refractive index at the microwave [3-5] as well as at the optical frequencies [6], and hence they are characterized by unique properties, including: reversal of Snell's law, backward wave propagation, reversal of the Doppler shift and Cherenkov effect, collecting lens behavior forming 3-D images, perfect lens performance, and so on [7,8]. To this end, they have become a subject of intense research activity; see the reviews [9-12]. Such MMs are experimentally realized mainly by periodic arrays of thin conducting wires, exhibiting plasma behavior, and split-ring resonators (SRR's), resembling parallel plate capacitors, generating negative ϵ and μ, respectively [3,4]. For other realizations of LH MMs and related discussions see [13-15].

So far, MMs have been mainly investigated in the linear regime, where ϵ and μ are independent of the EM fields intensities. Nevertheless, *nonlinear* MMs, which may be created by embedding an array of wires and SRR's into a nonlinear dielectric [16,17], are useful in various applications. In particular, it has been demonstrated that the field intensity acts as a control mechanism, altering the MM properties from LH to RH and back. Hence, the study of MMs nonlinear properties may assist in the implementation of tunable structures with intensity controlled transmission and in studying nonlinear effects in negative refraction photonic

crystals. Furthermore, it was shown in [18] that LH weakly nonlinear MMs support propagation of vector solitons.

On the other hand, concerning the frequency band gaps, they are characterized by $\epsilon < 0$ and $\mu > 0$ and hence linear EM waves are evanescent. In particular, there exist two band gaps which will be named as the "low-frequency" (LF) and the "high-frequency" (HF) gap. However, when a nonlinearity occurs, e.g. in the dielectric response of the MM [16-22], then nonlinearity-induced localization of EM waves is possible. Such localization is indicated by the formation of *gap solitons*, occurring mainly in nonlinear optics [23]. Gap solitons were also predicted in nonlinear MMs as solutions of a nonlinear Klein-Gordon equation [24].

In this work, first we analyze EM wave propagation in nonlinear lossy LH MMs. The methodology consists in starting from the Maxwell's equations, and using the reductive perturbation method to derive systematically a nonlinear Schrödinger (NLS) equation and a higher-order NLS (HNLS) equation for the EM fields envelopes, governing the propagation of ultra-short solitons. Analyzing the NLS and HNLS equations, we find necessary conditions for the formation of bright or dark solitons in the LH regime, and derive exact ultra-short solitons propagating in nonlinear LH MMs. The developed direct analysis of the Maxwell's equations, rather than the coupled wave equations for the EM fields envelopes, shows that the electric field envelope is proportional to the magnetic field one (their ratio being the linear wave-impedance). Thus, for each of the EM wave components, we derive a *single* NLS (for moderate pulse widths) or a *single* HNLS equation (for ultra-short pulse widths), rather than a system of *coupled* NLS equations (as in existing literature, see e.g. [18]). The HNLS equation, which incorporates higher-order dispersive and nonlinear terms, generalizes the one describing short pulse propagation in nonlinear optical fibers.

Then, we proceed with the analysis of nonlinear localized EM waves propagating in the frequency band gaps of nonlinear MMs. More precisely, we derive nonlinear evolution equations describing *ultra-short* pulses (possessing pulse widths of the order of a few cycles of the carrier frequency) that can be formed in these band gaps. In doing so, we consider a MM characterized by the permittivity ϵ and permeability μ of [3], as well as a weak Kerr-type nonlinearity in its dielectric response [20,21,24]. By modifying the techniques of [25,26], we derive appropriate expressions for μ in the HF

and LF band gaps. Then, we use a multiscale perturbation method, with different small parameters for the HF and LF gaps depending on the MM characteristics, to derive from Maxwell's equations two *short-pulse equations* SPEs. Each of these equations describes the evolution of ultra-short pulses either in the HF or the LF gap. The SPE has been shown to be the proper model for describing the evolution of ultra-short pulses in nonlinear fiber optics [25]; moreover in [25] the solutions of Maxwell's equations were compared numerically to the ones of the SPE and NLS models and it was shown that the accuracy of the SPE (NLS) increases (decreases) as the pulse width shortens. We also discuss the structure of the resulting peakon-like solutions of the SPEs derived in our context of nonlinear MMs, and draw parallels to NLS-like soliton solutions (which can be regarded as ultra-short gap solitons in nonlinear MMs). Numerical simulations of the peakon-like and breather-like solutions are also included both in the context of the SPEs as well as in that of the reduced wave equations originating from Maxwell's equations.

Investigations of the above described phenomena were initiated in [27] and [28]. This paper contains a unified treatment of these investigations, provides appropriate extensions and points out generalizations concerning the modelling of wave propagation in more complicated types of nonlinear MMs (e.g. chiral MMs).

II. DESCRIPTION OF THE NONLINEAR METAMATERIAL

We consider a composite lossy nonlinear MM, consisting of an array of conducting wires and SRRs, with the slits of the SRRs filled with a weakly nonlinear dielectric [16,18-20,22]. The MM is characterized by a frequency dependent effective complex permittivity $\hat{\epsilon}(\omega)$ and magnetic permeability $\hat{\mu}(\omega)$ (below f and $\hat{f}$ denote any function f in the time- and frequency-domain, respectively) given by [3,16]:

$$\hat{\epsilon}(\omega; |\mathbf{E}|^2) = \epsilon_0 \left(\epsilon_D(|\mathbf{E}|^2) - \frac{\omega_p^2}{\omega^2 + i\omega\gamma_\epsilon} \right),$$

$$\hat{\mu}(\omega; |\mathbf{H}|^2) = \mu_0 \left(1 - \frac{F\omega^2}{\omega^2 + i\omega\gamma_\mu - \omega_{0NL}^2(|\mathbf{H}|^2)} \right),$$

where ϵ_0 and μ_0 are the vacuum permittivity and permeability, ω is the EM wave's frequency, ω_p is the plasma frequency, F is the filling factor, γ_ϵ and γ_μ are the linear loss frequencies, ω_{0NL} is the nonlinear resonant SRR frequency [16], $\mathbf{E}$ and $\mathbf{H}$

are the electric and magnetic field intensities. In the linear limit, $\epsilon_D \to 1$ and $\omega_{0NL} \to \omega_{res}$ (ω_{res} is the linear resonant SRR frequency).

In particular, we assume the decompositions [18-22,29]:

$$\hat{\epsilon}(\omega; |\mathbf{E}|^2) = \hat{\epsilon}_L(\omega) + \epsilon_{NL}(|\mathbf{E}|^2), \quad (1)$$

$$\hat{\mu}(\omega; |\mathbf{H}|^2) = \hat{\mu}_L(\omega) + \mu_{NL}(|\mathbf{H}|^2). \quad (2)$$

For the linear parts we assume a Drude (Lorentz) model for the permittivity (permeability) [27]:

$$\hat{\epsilon}_L(\omega) = \epsilon_0 \left(1 - \frac{\omega_p^2}{\omega^2 + i\omega\gamma_\epsilon} \right), \quad (3)$$

$$\hat{\mu}_L(\omega) = \mu_0 \left(1 - \frac{F\omega^2}{\omega^2 + i\omega\gamma_\mu - \omega_{res}^2} \right), \quad (4)$$

while the weakly nonlinear parts of the permittivity and permeability exhibit a Kerr-type behavior and are, respectively, given by [18-22]:

$$\epsilon_{NL}(|\mathbf{E}|^2) = \epsilon_0 \alpha |\mathbf{E}|^2, \quad (5)$$

$$\mu_{NL}(|\mathbf{H}|^2) = \mu_0 \beta |\mathbf{H}|^2, \quad (6)$$

where α and β are the Kerr coefficients; $\alpha > 0$ and $\alpha < 0$ corresponds to focusing and defocusing dielectrics.

Assuming that $\omega_p > \omega_{res}$ and that the linear losses γ_ϵ and γ_μ are small compared to the operating frequency, the dispersion relation $k_0^2 = \omega_0^2 \mathrm{Re}[\hat{\epsilon}_L(\omega_0)] \mathrm{Re}[\hat{\mu}_L(\omega_0)]$ (k_0 the wavenumber corresponding to the carrier frequency ω_0) shows that linear waves exist ($k_0^2 > 0$) in the bands $\omega > \omega_p$ and $\omega_{res} < \omega < \omega_M \equiv \omega_{res}/\sqrt{1-F}$, where the medium is right-handed (RH) and left-handed (LH), respectively. These bands will be called the RH and LH bands. On the other hand, there exist two bands, with $\mathrm{Re}[\hat{\epsilon}_L] < 0$ and $\mathrm{Re}[\hat{\mu}_L] > 0$, where linear waves are evanescent ($k_0^2 < 0$). These "band-gaps" are $0 < \omega < \omega_{res}$ ["low-frequency" (LF) gap], and $\omega_M < \omega < \omega_p$ ["high-frequency" (HF) gap]. Typical dispersion curves of $\mathrm{Re}[\hat{\epsilon}_L](\omega)$ and $\mathrm{Re}[\hat{\mu}_L](\omega)$ are shown in Fig. 1.

In the sequel, we consider the physically relevant [18] values of parameters: $F = 0.4$, $\omega_{res} = 2\pi \times 1.45$ GHz, $\omega_M = 2\pi \times 1.87$ GHz, and $\omega_p = 2\pi \times 10$ GHz, for which α may be positive or negative, while β is positive.

III. EM WAVE PROPAGATION IN THE LEFT-HANDED BAND

First, we analyze the frequency band $2\pi \times 1.45$ GHz $< \omega < 2\pi \times 1.87$ GHz, inside which the SRRs are negative-index LH media (with $\mathrm{Re}[\hat{\epsilon}_L] < 0$ and $\mathrm{Re}[\hat{\mu}_L] < 0$), as was discussed above.
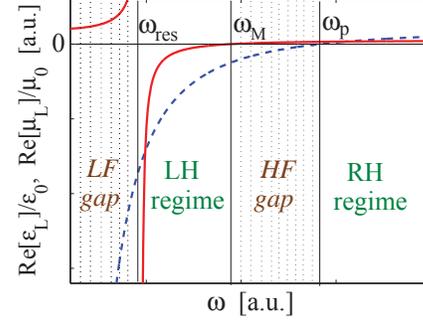
Fig. 1. $\mathrm{Re}[\hat{\mu}_L]/\mu_0$ [solid (red) line], and $\mathrm{Re}[\hat{\epsilon}_L]/\epsilon_0$ [dashed (blue) line] versus ω in arbitrary units (a.u.). Linear waves propagate ($\mathrm{Re}[\hat{\epsilon}_L]\mathrm{Re}[\hat{\mu}_L] > 0$) for $\omega > \omega_p$ (RH regime) and $\omega_{res} < \omega < \omega_M$ (LH regime). Dotted regions with $\mathrm{Re}[\hat{\epsilon}_L] < 0$ and $\mathrm{Re}[\hat{\mu}_L] > 0$ depict the LF gap ($0 < \omega < \omega_{res}$) and the HF gap ($\omega_M < \omega < \omega_p$).

A. Propagation along the $+\hat{z}$ axis

We consider the propagation along the $+\hat{z}$ direction of a x- (y-) polarized electric (magnetic) field, namely,

$$\mathbf{E}(z,t) = \hat{\mathbf{x}}E(z,t) \quad , \quad \mathbf{H}(z,t) = \hat{\mathbf{y}}H(z,t). \quad (7)$$

Then, using the constitutive relations (in frequency domain) $\hat{\mathbf{D}} = \hat{\epsilon}\hat{\mathbf{E}}$ and $\hat{\mathbf{B}} = \hat{\mu}\hat{\mathbf{H}}$ ($\hat{\mathbf{D}}$ and $\hat{\mathbf{B}}$ are the electric flux density and the magnetic induction), Faraday's and Ampére's laws respectively read (in the time domain):

$$\partial_z E = -\partial_t(\mu * H), \quad \partial_z H = -\partial_t(\epsilon * E), \quad (8)$$

where $*$ denotes the convolution integral, i.e., $f(t) * g(t) = \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau$, for functions $f(t)$ and $g(t)$.

Note that Eqs. (8) may be used in either the RH or the LH regime of a MM: once the dispersion relation $k_0 = k_0(\omega_0)$ and the evolution equations for the fields E and H are found, then $k_0 > 0$ ($k_0 < 0$) corresponds to the RH (LH) regime. Alternatively, for fixed $k_0 > 0$, one should shift the fields as $[E, H]^T \to [\pm E, \mp H]^T$ (either up or down sign combinations), thus inverting the orientation of the magnetic field and associated Poynting vector. Below, in our consideration we will assume that the wavenumber $k_0 = k_0(\omega_0)$ obtained from the linear dispersion relation [see Eq. (20) below] will be $k_0 < 0$ for the LH regime.

Next, we consider that:

$$[E(z,t), H(z,t)]^T = [q(z,t), p(z,t)]^T e^{i(k_0 z - \omega_0 t)}, \quad (9)$$

where q and p are the unknown EM fields envelopes.

B. Nonlinear evolution equations by the reductive perturbation method

Nonlinear evolution equations for the field envelopes can be found by the *reductive perturbation method* [30] as follows. First, we assume that the temporal spectral width of the nonlinear term with respect to the spectral width of the quasi-plane-wave dispersion relation is characterized by the formal small parameter ε [31,32].

Then, we introduce the slow variables:

$$Z = \varepsilon^2 z, \qquad T = \varepsilon(t - k_0' z), \qquad (10)$$

where $k_0' \equiv v_g^{-1}$ is the inverse of the group velocity (hereafter, primes will denote derivatives with respect to ω_0). Additionally, we express q and p as asymptotic expansions in terms of ε,

$$q(Z,T) = q_0(Z,T) + \varepsilon q_1(Z,T) + \varepsilon^2 q_2(Z,T) + \cdots$$
$$(11)$$

$$p(Z,T) = p_0(Z,T) + \varepsilon p_1(Z,T) + \varepsilon^2 p_2(Z,T) + \cdots$$

and assume that the Kerr coefficients α and β are of order $\mathcal{O}(\varepsilon^2)$ (see, e.g., [18,31] as well as [33,34]). Moreover, the linear components of $\hat{\epsilon}_L$ and $\hat{\mu}_L$ are decomposed into real and imaginary parts, as follows:

$$\hat{\epsilon}_L(\omega) = \hat{\epsilon}_R(\omega) - i\hat{\epsilon}_I(\omega), \qquad (12)$$

$$\hat{\mu}_L(\omega) = \hat{\mu}_R(\omega) - i\hat{\mu}_I(\omega), \qquad (13)$$

where the imaginary parts are assumed to be $\mathcal{O}(\varepsilon^2)$. Note that the effect of losses in $\hat{\epsilon}_L$ and $\hat{\mu}_L$ is taken here for the first time into account since the analysis of [27] concerns the lossless case.

Then, substituting Eqs. (11) into Eqs. (8), using Eqs. (1), (2), and (10), and Taylor expanding $\hat{\epsilon}_L$, and $\hat{\mu}_L$, we arrive at the following hierarchy of equations:

$$\mathcal{O}(\varepsilon^0): \quad \mathbf{W}\mathbf{x}_0 = \mathbf{0}, \qquad (14)$$

$$\mathcal{O}(\varepsilon^1): \quad \mathbf{W}\mathbf{x}_1 = -i\mathbf{W}'\partial_T \mathbf{x}_0, \qquad (15)$$

$$\mathcal{O}(\varepsilon^2): \quad \mathbf{W}\mathbf{x}_2 = -i\mathbf{W}'\partial_T \mathbf{x}_1 + \frac{1}{2}\mathbf{W}''\partial_T^2 \mathbf{x}_0$$
$$+ \frac{k_0''}{2}\partial_T^2 \mathbf{x}_0 - i\partial_Z \mathbf{x}_0 - (\mathbf{A} - i\mathbf{B})\,\mathbf{x}_0, \qquad (16)$$

$$\mathcal{O}(\varepsilon^3): \quad \mathbf{W}\mathbf{x}_3 = -i\mathbf{W}'\partial_T \mathbf{x}_2 + \frac{1}{2}\mathbf{W}''\partial_T^2 \mathbf{x}_1$$
$$+ \frac{i}{6}\mathbf{W}'''\partial_T^3 \mathbf{x}_0 + \frac{ik_0'''}{6}\partial_T^3 \mathbf{x}_0 + \frac{k_0''}{2}\partial_T^2 \mathbf{x}_1$$
$$- i\partial_Z \mathbf{x}_1 - (\mathbf{A} - i\mathbf{B})\,\mathbf{x}_1 - (\mathbf{B}'\partial_T - i\mathbf{C})\mathbf{x}_0, \qquad (17)$$

with $\mathbf{x}_i = [q_i, p_i]^{\mathrm{T}}$ $(i = 0, 1, 2, 3)$ unknown vectors, and

$$\mathbf{W} = \begin{bmatrix} -k_0 & \omega_0\hat{\mu}_R \\ \omega_0\hat{\epsilon}_R & -k_0 \end{bmatrix}, \; \mathbf{A}\mathbf{x}_i = \omega_0 \begin{bmatrix} \beta|p_0|^2 p_i \\ \alpha|q_0|^2 q_i \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & \omega_0\hat{\mu}_I \\ \omega_0\hat{\epsilon}_I & 0 \end{bmatrix}, \qquad (18)$$

$$\mathbf{C}\mathbf{x}_0 = \begin{bmatrix} -\beta\partial_T(|p_0|^2 p_0) + i\omega_0\beta(p_0 p_1^\star + p_0^\star p_1)p_0 \\ -\alpha\partial_T(|q_0|^2 q_0) + i\omega_0\alpha(q_0 q_1^\star + q_0^\star q_1)q_0 \end{bmatrix}$$

with $\star$ denoting complex conjugate.

The compatibility conditions required for the solvability of Eqs. (14)–(17) (known also as Fredholm alternatives [30,31]) are: $\mathbf{L}\mathbf{W}\mathbf{x}_i = 0$, where $\mathbf{L} = [1, \hat{Z}_L]$ is a left eigenvector of $\mathbf{W}$, such that $\mathbf{L}\mathbf{W} = \mathbf{0}$, and $\hat{Z}_L = \sqrt{\hat{\mu}_R/\hat{\epsilon}_R}$ is the linear wave-impedance, when dissipation is small enough to be ignored.

The leading-order Eq. (14) provides the following results. First, the solution $\mathbf{x}_0$ of Eq. (14) has the form:

$$\mathbf{x}_0 = \mathbf{R}\phi(Z,T), \qquad (19)$$

where $\phi(Z,T)$ an under determination scalar field and $\mathbf{R} = [1, \hat{Z}_L^{-1}]^{\mathrm{T}}$ a right eigenvector of $\mathbf{W}$, such that $\mathbf{W}\mathbf{R} = \mathbf{0}$. Second, by using the compatibility condition $\mathbf{L}\mathbf{W}\mathbf{x}_0 = 0$ and Eq. (19), we obtain that $\mathbf{L}\mathbf{W}\mathbf{R} = 0$, which is actually the linear dispersion relation:

$$k_0^2 = \omega_0^2 \hat{\epsilon}_R \hat{\mu}_R, \qquad (20)$$

with all functions of frequency evaluated at ω_0. Eq. (20) is also obtained by the nontrivial solution condition $\det\mathbf{W} = 0$. Third, the electric and magnetic field envelopes are proportional to each other, i.e. $q_0 = p_0 \hat{Z}_L$.

Next, at $\mathcal{O}(\varepsilon^1)$, the compatibility condition for Eq. (15) results in $\mathbf{L}\mathbf{W}'\mathbf{R} = 0$, written as:

$$2k_0 k_0' = \omega_0^2(\hat{\epsilon}_R \hat{\mu}_R' + \hat{\epsilon}_R' \hat{\mu}_R) + 2\omega_0 \hat{\epsilon}_R \hat{\mu}_R. \quad (21)$$

This is actually the definition of the group velocity $v_g = 1/k_0'$, as can also be found by differentiating the dispersion relation Eq. (20) with respect to ω. Furthermore, by combining Eqs. (15) and (19), we get that:

$$\mathbf{x}_1 = i\mathbf{R}'\partial_T \phi(Z,T) + \mathbf{R}\psi(Z,T), \qquad (22)$$

where $\psi(Z,T)$ is an unknown scalar field. Next, at order $\mathcal{O}(\varepsilon^2)$, the compatibility condition for Eq. (16), combined with Eqs. (19) and (22), yields the following nonlinear Schrödinger (NLS) equation:

$$i\partial_Z \phi - \frac{1}{2}k_0''\partial_T^2 \phi + \vartheta|\phi|^2 \phi = i\tilde{g}\phi, \qquad (23)$$

where k_0'' is the group-velocity dispersion coefficient, while $\tilde{g}$ and ϑ are, respectively, the linear loss and nonlinear coefficients which are given by:

$$\tilde{g} = \frac{k_0}{2\hat{\epsilon}_R\hat{\mu}_R}\left(\hat{\epsilon}_R\hat{\mu}_I + \hat{\mu}_R\hat{\epsilon}_I\right), \qquad (24)$$

$$\vartheta = \frac{\omega_0^2}{2k_0}\left(\epsilon_0\alpha\hat{\mu}_R + \mu_0\beta\hat{\epsilon}_R\hat{Z}_L^{-2}\right). \qquad (25)$$

Importantly, once ϕ is obtained from the NLS Eq. (23), the electric and magnetic field envelopes are, respectively, determined by means of Eq. (19) as $q_0 = \phi$ and $p_0 = \hat{Z}_L^{-1}\phi$, similarly to the case of a linear medium.

Finally, to order $\mathcal{O}(\varepsilon^3)$, we use the compatibility condition for Eq. (17), as well as Eqs. (16), (19), and (22), and obtain a NLS equation, incorporating higher-order dispersive and nonlinear terms. This equation describes the evolution of ψ, and yet contains ϕ, which obeys Eq. (23). By following [31] and [32], we introduce a new combined function $\Phi = \phi + \varepsilon\psi$, and then combine the NLS equations obtained at $\mathcal{O}(\varepsilon^2)$ and $\mathcal{O}(\varepsilon^3)$ to find that Φ obeys the higher-order NLS (HNLS) equation:

$$i\partial_Z\Phi - \frac{1}{2}k_0''\partial_T^2\Phi + \vartheta|\Phi|^2\Phi$$
$$= i\varepsilon\left[\frac{1}{6}k_0'''\partial_T^3\Phi - \frac{\vartheta}{\omega_0}\partial_T(|\Phi|^2\Phi) + i\tilde{g}'\partial_T\Phi\right].$$
$$\qquad (26)$$

For $\varepsilon = 0$, the HNLS Eq. (26) is reduced to the NLS Eq. (23), while for $\varepsilon \neq 0$ generalizes the HNLS equation describing ultra-short pulse propagation in optical fibers [31,32] (where $\mu = \mu_0$, while dispersion and nonlinearity appear solely in the fiber dielectric properties). As in the case of the NLS Eq. (23), Eq. (26) provides the field Φ which determines the electric and magnetic fields at $\mathcal{O}(\varepsilon^3)$, respectively, as $q_0 + \varepsilon q_1 = \Phi$ and $p_0 + \varepsilon p_1 = \hat{Z}_L^{-1}\Phi$ [see Eqs. (19) and (22)]. Note that Eqs. (23) or (26) can be used in the LH (RH) regime, taking k_0, $\hat{\epsilon}_R$, and $\hat{\mu}_R$ negative (positive), as per the discussion above.

In the sequel, we derive analytically exact soliton solutions of Eqs. (23) and (26).

C. Solitons solutions of the NLS equation

First, we analyze in detail the NLS Eq. (23): we start by measuring length, time, and the field intensity $|\phi|^2$ in units of the dispersion length $L_D = t_0^2/|k_0''|$, initial pulse width t_0, and $L_D/|\vartheta|$, respectively, and reduce Eq. (23) to the following dimensionless form:

$$i\partial_Z\phi - \frac{s}{2}\partial_T^2\phi + \sigma|\phi|^2\phi = ig\phi, \qquad (27)$$

TABLE I: Conditions for the formation of bright or dark solitons (BS or DS) for the NLS Eq. (27)

		$s = +1$	$s = -1$
$\sigma = +1$	$\alpha > 0$	DS	BS
$\sigma = -1$	$\alpha < 0$, $\left|\frac{\alpha}{\beta}\right| > \frac{Z_0^2}{\hat{Z}_L^4}$	BS	DS
$\sigma = +1$	$\alpha < 0$, $\left|\frac{\alpha}{\beta}\right| < \frac{Z_0^2}{\hat{Z}_L^4}$	DS	BS

where $s = \text{sign}(k_0'')$, $\sigma = \text{sign}(\vartheta)$, and $g = L_D\tilde{g}$. The NLS Eq. (27) admits bright (dark) soliton solutions for $s\sigma = -1$ ($s\sigma = +1$).

For our choice of parameters, numerical simulations indicate (see Fig. 2 of [27]) that inside the LH regime $s = +1$ ($k_0'' > 0$) for $2\pi \times 1.76 < \omega < 2\pi \times 1.87$ GHz, while $s = -1$ ($k_0'' < 0$) for $2\pi \times 1.45 < \omega < 2\pi \times 1.76$ GHz. On the other hand, the linear loss coefficient g attains very small values and can to a certain approximation be ignored. Moreover, regarding σ, it can take either the value $\sigma = +1$ or $\sigma = -1$, depending on the magnitudes and signs of the Kerr coefficients α and β. Note that we have assumed that $\beta > 0$, and hence $\sigma = +1$ either for a focusing dielectric, with $\alpha > 0$, or for a defocusing dielectric, $\alpha < 0$, with $|\alpha/\beta| < Z_0^2/\hat{Z}_L^4$ ($Z_0 = \sqrt{\mu_0/\epsilon_0}$ is the vacuum wave-impedance).

Thus, for $\sigma = +1$, bright (dark) solitons occur in the anomalous (normal) dispersion regimes, namely, for $k_0'' < 0$ ($k_0'' > 0$), respectively. On the other hand, $\sigma = -1$ for a defocusing dielectric ($\alpha < 0$), with $|\alpha/\beta| > Z_0^2/\hat{Z}_L^4$ and, bright (dark) solitons occur in the normal (anomalous) dispersion regimes. The above conclusions are summarized in Table I.

The "flexibility" due to the extra "degree of freedom" provided by the dispersion and nonlinearity properties of the magnetic response of the LH MM (missing in fiber optics), allows for the formation of bright (dark) solitons in the anomalous (normal) dispersion regimes for defocusing dielectrics (see third line of Table I).

D. Ultra-short solitons solutions of the HNLS equation

Now, we analyze the HNLS Eq. (26) which, by using the same dimensionless units as before, is expressed as:

$$i\partial_Z\Phi - \frac{s}{2}\partial_T^2\Phi + \sigma|\Phi|^2\Phi =$$
$$i\delta_1\partial_T^3\Phi - i\sigma\delta_2\partial_T(|\Phi|^2\Phi) - \delta_3\partial_T\Phi, \qquad (28)$$

where the coefficients δ_1, δ_2, and δ_3 are given by:

$$\delta_1 = \varepsilon \frac{k_0'''}{6t_0|k_0''|}, \quad \delta_2 = \varepsilon \frac{1}{\omega_0 t_0}, \quad \delta_3 = \varepsilon \frac{L_D \tilde{g}'}{t_0}. \tag{29}$$

Ultra-short solitons in nonlinear LH MMs can be predicted by means of Eq. (28). More precisely, following [35], and assuming negligible contribution from the linear losses (i.e. $\delta_3 \simeq 0$), we seek travelling-wave solutions of Eq. (28) of the form:

$$\Phi(Z, T) = U(\eta) \exp[i(KZ - \Omega T)], \tag{30}$$

where $U(\eta)$ is the unknown real envelope function, $\eta = T - \Lambda Z$, and the real parameters Λ, K, and Ω denote the wave's inverse velocity, wavenumber, and frequency. Substituting Eq. (30) into Eq. (28), the real and imaginary parts of the resulting equation read:

$$\ddot{U} + \frac{K - \frac{s}{2}\Omega^2 - \delta_1\Omega^3}{\frac{s}{2} + 3\delta_1\Omega}U - \frac{\sigma(1 + \delta_2\Omega)}{\frac{s}{2} + 3\delta_1\Omega}U^3 = 0, \tag{31}$$

$$\delta_1\dddot{U} + (\Lambda - s\Omega - 3\delta_1\Omega^2)\dot{U} - 3\sigma\delta_2 U^2\dot{U} = 0, \tag{32}$$

where overdots denote differentiations with respect to η.

Notice that when $\delta_1 = \delta_2 = 0$, Eq. (32) is automatically satisfied if $\Lambda = s\Omega$ and the profile of "long" soliton pulses [governed by Eq. (27)] is determined by Eq. (31). On the other hand, for ultra-short solitons (corresponding to $\delta_1 \neq 0$, $\delta_2 \neq 0$), Eqs. (31) and (32) are consistent if:

$$\frac{K - \frac{s}{2}\Omega^2 - \delta_1\Omega^3}{\frac{s}{2} + 3\delta_1\Omega} = \frac{\Lambda - s\Omega - 3\delta_1\Omega^2}{\delta_1} \equiv \kappa, \tag{33}$$

$$-\frac{\sigma\delta_2}{\delta_1} = -\frac{\sigma(1 + \delta_2\Omega)}{\frac{s}{2} + 3\delta_1\Omega} \equiv \nu, \tag{34}$$

where κ and ν are nonzero constants. In such a case, Eqs. (31) and (32) are equivalent to the equation of motion of the unforced and undamped Duffing oscillator:

$$\ddot{U} + \kappa U + \nu U^3 = 0. \tag{35}$$

For $\kappa\nu < 0$, Eq. (35) possesses two exponentially localized solutions [35], having the form of a hyperbolic secant (tangent) for $\kappa < 0$ and $\nu > 0$ ($\kappa > 0$ and $\nu < 0$), and thus corresponding to the bright, U_{BS} (dark, U_{DS}) solitons of Eq. (28):

$$U_{\text{BS}}(\eta) = (2|\kappa|/\nu)^{1/2} \operatorname{sech}(\sqrt{|\kappa|}\eta), \tag{36}$$

$$U_{\text{DS}}(\eta) = (2\kappa/|\nu|)^{1/2} \tanh(\sqrt{\kappa/2}\eta). \tag{37}$$

These are *ultra-short* solitons of the HNLS Eq. (28), valid even for $\varepsilon = \mathcal{O}(1)$: since both δ_1, δ_2 of Eq. (28) scale as $\varepsilon(\omega_0 t_0)^{-1}$, it is clear that for $\omega_0 t_0 = \mathcal{O}(1)$, or for soliton widths $t_0 \sim \omega_0^{-1}$, the higher-order terms can be neglected and propagation is governed by Eq. (27). On the other hand, if $\omega_0 t_0 = \mathcal{O}(\varepsilon)$, the higher-order terms become important and solitons governed by the HNLS Eq. (28) are *ultra-short*, of width $t_0 \sim \varepsilon\omega_0^{-1}$, satisfying Faraday's and Ampére's laws in Eqs. (8) up to $\mathcal{O}(\varepsilon^3)$.

Concerning the condition for bright or dark soliton formation, namely $\kappa\nu < 0$, we note that κ depends on the free parameters K and Ω (and, thus, can be tuned on demand), while the parameter ν has the opposite sign from σ. This means that bright solitons are formed for $\kappa < 0$ and $\sigma = -1$ (i.e., $\alpha < 0$ with $|\alpha/\beta| > Z_0^2/\hat{Z}_L^4$), while dark ones are formed for $\kappa > 0$ and $\sigma = +1$ (i.e., $\alpha > 0$, or $\alpha < 0$ with $|\alpha/\beta| < Z_0^2/\hat{Z}_L^4$).

IV. EM WAVE PROPAGATION IN THE FREQUENCY BAND GAPS

Now, we consider the EM wave propagation in the two frequency band-gaps of the nonlinear MM (characterized by $\operatorname{Re}[\hat{\epsilon}_L] < 0$ and $\operatorname{Re}[\hat{\mu}_L] > 0$), where, as mentioned above, linear EM waves are evanescent ($k_0^2 < 0$). To this end, we assume that the MM exhibits a Kerr-type nonlinearity exclusively in its dielectric response, while the magnetic permeability is considered a linear function; this assumption is justified in [20,21,24].

As in (7), also here we assume propagation along the $+z$ direction of a x- (y-) polarized electric (magnetic) field. Then, Maxwell's equations lead to the following time-domain nonlinear wave equation for $E(z, t)$:

$$\partial_z^2 E - \partial_t^2(\epsilon_L * \mu_L * E) - \epsilon_0\alpha\partial_t^2(\mu_L * E^3) = 0. \tag{38}$$

Note that once the electric field $E(z, t)$ is obtained from Eq. (38), the magnetic field $H(z, t)$ is derived from Faraday's law.

In the sequel, we analyze the EM wave propagation separately in the "high-frequency" (HF) and "low-frequency" (LF) gaps. We assume that the MM is lossless; extensions to the lossy case are discussed in Sec. V below.

A. The high-frequency band gap

The HF band gap is defined by $\omega_M < \omega < \omega_p$ (see Fig. 1). In this regime, by assuming $\omega \gg \omega_{\text{res}}$,
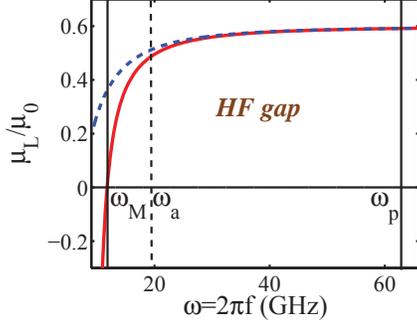
Fig. 2. Permeability $\hat{\mu}_L/\mu_0$ in the HF gap. Solid (red) and dashed (blue) lines correspond, respectively, to the exact [Eq. (4)] and approximate [Eq. (39)] expressions of $\hat{\mu}_L(\omega)/\mu_0$ in this band. The approximation produces a relative error less than 5% for $\omega_a \equiv 2\pi \times 3.1$ GHz $< \omega < \omega_p = 2\pi \times 10$ GHz.

we find that $\hat{\mu}_L(\omega)$ in Eq. (4) is approximated by:

$$\hat{\mu}_L(\omega) \approx \mu_0(1 - F) - \mu_0 F \frac{\omega_{\text{res}}^2}{\omega^2}. \qquad (39)$$

Using the parameter values discussed above (namely $F = 0.4$, $\omega_{\text{res}} = 2\pi \times 1.45$ GHz, $\omega_M = 2\pi \times 1.87$ GHz, and $\omega_p = 2\pi \times 10$ GHz), in Fig. 2 we show the exact [Eq. (4)] and approximate [Eq. (39)] expressions for the effective permeability in this band. As seen, in Fig. 2, the above approximation produces a relative error from the exact form of $\hat{\mu}_L(\omega)/\mu_0$ less than 5% in a wide sub-interval of frequencies in this band, namely for $\omega_a \equiv 2\pi \times 3.1$ GHz $< \omega < \omega_p = 2\pi \times 10$ GHz.

Next, by using Eq. (39) and employing the techniques of [28], we reduce Eq. (38) to:

$$\partial_z^2 E \;-\; \frac{1 - F}{c^2}\partial_t^2 E - \frac{1}{c^2}[F\omega_{\text{res}}^2 + (1 - F)\omega_p^2]E$$
$$-\; \frac{\alpha}{c^2}\left[F\omega_{\text{res}}^2 E^3 + (1 - F)\partial_t^2 E^3\right] = 0, (40)$$

where c is the velocity of light in vacuum. Next, we measure time, space, and the field intensity E^2 in units of ω_{res}^{-1}, v/ω_{res} [where $v = c(1 - F)^{-1/2}$], and $|\alpha|^{-1}$, respectively, and express Eq. (40) in dimensionless form:

$$\left(\partial_z^2 - \partial_t^2 - \tilde{\rho}\right)E = s\left(\frac{F}{1 - F} + \partial_t^2\right)E^3, \quad (41)$$

where $s = \text{sgn}(\alpha) = \pm 1$ for focusing or defocusing nonlinearity, respectively, and $\tilde{\rho} = F/(1 - F) + (\omega_p/\omega_{\text{res}})^2$. We have assumed above that $(\omega_p/\omega_{\text{res}})^2 \gg 1$ in order to ensure the validity of the approximation (39) in a wide sub-interval of the HF band gap. Then, $\tilde{\rho}$ is a large parameter, which

suggests that $\tilde{\rho} = \rho/\varepsilon$, where ε is a formal small parameter, and $\rho = \mathcal{O}(1)$. Furthermore, considering propagation of small-amplitude short pulses, we introduce a multiple scale ansatz of the form:

$$E = \varepsilon^{3/2}E_1(T_{\text{HF}}, Z_1, \cdots)$$
$$+ \varepsilon^{5/2}E_2(T_{\text{HF}}, Z_1, \cdots) + \cdots, \qquad (42)$$

where $T_{\text{HF}} = \varepsilon^{-2}(t - z)$ and $Z_n = \varepsilon^n z$ ($n = 1, 2, \cdots$). Substituting Eq. (42) into Eq. (41), we obtain various equations at different orders of ε. In particular, terms at $\mathcal{O}(\varepsilon^{-5/2})$ cancel, there are no terms at $\mathcal{O}(\varepsilon^{-3/2})$, while terms at $\mathcal{O}(\varepsilon^{-1/2})$, cancel provided that:

$$2\partial_\zeta \partial_{T_{\text{HF}}} E_1 + \rho E_1 + s\partial_{T_{\text{HF}}}^2 E_1^3 = 0, \qquad (43)$$

where $\zeta \equiv Z_1$. Eq. (43) is the short pulse equation (SPE), which was derived in Ref. [25] in the context of ultra-short pulses propagation in nonlinear silica optical fibers.

B. The low-frequency band gap

The LF gap is defined by $0 < \omega < \omega_{\text{res}}$ (see Fig. 1). For $\omega \ll \omega_{\text{res}}$ we approximate $\hat{\mu}_L(\omega)$ in Eq. (4) by:

$$\hat{\mu}_L(\omega) \approx \mu_0\left(1 + F\frac{\omega^2}{\omega_{\text{res}}^2}\right). \qquad (44)$$

Figure 3 depicts the exact [Eq. (4)] and approximate [Eq. (44)] expressions for $\hat{\mu}_L(\omega)$ in the LF band gap for $F = 0.02$, $\omega_{\text{res}} = 2\pi \times 1.45$ GHz. The relative error is less than 5% for $0 < \omega < \omega_b \equiv 2\pi \times 1.28$ GHz.

Fig. 3. Permeability $\hat{\mu}_L/\mu_0$ in the LF gap. Solid (red) and dashed (blue) lines correspond, respectively, to the exact [Eq. (4)], and approximate [Eq. (44)] expressions of $\hat{\mu}_L(\omega)/\mu_0$; a relative error of less than 5% is achieved for $\omega_b \equiv 2\pi \times 1.28$ GHz $< \omega < \omega_{\text{res}} = 2\pi \times 1.45$ GHz.

Now, we use Eq. (44) and follow a similar procedure with that of the HF gap, to reduce Eq. (38) to:

$$\partial_z^2 E - \frac{1}{c^2}\left(1 - F\frac{\omega_p^2}{\omega_{\mathrm{res}}^2}\right)\partial_t^2 E - \frac{\omega_p^2}{c^2}E =$$
$$-\frac{F}{\omega_{\mathrm{res}}^2 c^2}\partial_t^4 E + \frac{\alpha}{c^2}\partial_t^2 E^3 - \frac{\alpha F}{\omega_{\mathrm{res}}^2 c^2}\partial_t^4 E^3. \quad (45)$$

Notice that in Eq. (45), the ratio $(\omega_p/\omega_{\mathrm{res}})^2$ is considered to be a $\mathcal{O}(1)$ parameter (as, e.g., in Ref. [24]) since ω_p is not involved in the band width of the LF band gap. In this band, it is convenient to use the filling factor F as the small parameter; this is a physically relevant choice for SRRs [36], as well as for other types of MMs [37].

Next, we measure time, space, and the field intensity E^2 in units of $\omega_{\mathrm{res}}^{-1}$, c/ω_{res} and $|\alpha|^{-1}$, respectively, and thus reduce Eq. (45) to the dimensionless form:

$$\left(\partial_z^2 - \partial_t^2 - \frac{\omega_p^2}{\omega_{\mathrm{res}}^2}\right)E = s\partial_t^2 E^3. \quad (46)$$

From Eq. (46), we derive an SPE for the LF band by using the asymptotic expansion:

$$E = \varepsilon E_1(T_{\mathrm{LF}}, Z_1, \cdots) + \varepsilon^2 E_2(T_{\mathrm{LF}}, Z_1, \cdots) + \cdots, \quad (47)$$

where $T_{\mathrm{LF}} = \varepsilon^{-1}(t-z)$ and $Z_n = \varepsilon^n z$ ($n = 1, 2, \cdots$). Substituting Eq. (47) into Eq. (46), the terms at $\mathcal{O}(\varepsilon^{-1})$ cancel, there are no terms at $\mathcal{O}(\varepsilon^0)$, while terms at $\mathcal{O}(\varepsilon)$, cancel provided that E_1 satisfies the SPE:

$$2\partial_\zeta \partial_{T_{\mathrm{LF}}} E_1 + \frac{\omega_p^2}{\omega_{\mathrm{res}}^2}E_1 + s\partial_{T_{\mathrm{LF}}}^2 E_1^3 = 0. \quad (48)$$

C. Solutions of the short pulse equations

We discuss solutions of the SPEs by unifying Eqs. (43) and (48) in the single equation with respect to $u \equiv E_1$:

$$\partial_\zeta \partial_\tau u + \Gamma u + \frac{1}{2}s\partial_\tau^2 u^3 = 0, \quad (49)$$

where $\Gamma = \rho/2$ and $\tau = T_{\mathrm{HF}}$ for the HF gap, and $\Gamma = \omega_p^2/(2\omega_{\mathrm{res}}^2)$ and $\tau = T_{\mathrm{LF}}$ for the LF gap. Seeking travelling wave solutions $u = u(\xi)$, where $\xi = \zeta - C\tau$, Eq. (49) is reduced to the ordinary differential equation:

$$-Cu_{\xi\xi} + \Gamma u + \frac{1}{2}sC^2(6uu_\xi^2 + 3u^2 u_{\xi\xi}) = 0, \quad (50)$$

which by means of the transformation $u_\xi^2 = w(u)$ yields:

$$u_\xi^2 = -\frac{\Gamma u^2}{C}\frac{3sCu^2 - 4}{(3sCu^2 - 2)^2}, \quad (51)$$

subject to the initial condition $u_\xi(\pm\infty) = u(\pm\infty) = 0$ or $w(u(\pm\infty)) = w(0) = 0$. For $sC < 0$ the solutions (51) are always either ascending or descending, while for $sC > 0$ bounded solutions are allowed. In the case $s = +1$ (i.e., focusing dielectrics with $\alpha > 0$), which implies that $C > 0$, the maximum of the travelling wave occurs when $u = \sqrt{4/3C}$. To avoid the singularity at $u = \sqrt{2/3C}$, we consider small amplitude pulses and Eq. (51) is reduced to $u_\xi^2 = (\Gamma/C)u^2$, which possesses a *peakon*-like solution (see, e.g. [38]) of the form:

$$u(\xi) = A\exp(-\sqrt{\Gamma/C}|\xi|), \quad (52)$$

whose derivative has a discontinuity at $\xi = 0$, with amplitude $A < \sqrt{2/3C}$. Also, when the above approximation is not used, the right-hand side of Eq. (51) needs to be positive giving $|u| < \sqrt{4/3|C|}$. These characteristics are consistent with the fact that the SPE exhibits loop-solitons (see [39]).

Moreover, based on the formal connection between the SPE and the sine-Gordon equation (SGE), a smooth, sech-shaped, envelope soliton solution of the SPE, based on the breather solution of the SGE, was derived in [39]. In the framework of Eq. (49), this solution has the form:

$$u \approx \frac{4m}{\sqrt{3\Gamma}}\cos\left(\zeta + \Gamma\tau\right)\mathrm{sech}\left[m\left(\zeta - \Gamma\tau\right)\right], \quad (53)$$

where $0 < m < 1$. The shape of the SPE pulse in Eq. (53) bears resemblance to the NLS soliton, as it consists of a sech-shaped pulse modulating a cos-shaped function. In particular, a connection is established in [28] between the SPE and NLS equations by showing that the envelope function A satisfies the NLS equation:

$$i\left(\partial_{\zeta_2}A + k'\partial_{\tau_2}A\right) - \frac{k''}{2}\partial_{\bar{\tau}_1}^2 A + \frac{3}{2}\omega|A|^2 A = 0, \quad (54)$$

where $\zeta_n = \varepsilon^n\zeta$, $\tau_n = \varepsilon^n\tau$ ($\varepsilon \ll 1$ and $n = 0, 1, \cdots$), $\bar{\tau}_1 = \tau_1 - k'\zeta_1$, $k' = -k/\omega$, and $k'' = 2k/\omega^2$. Clearly the well-known sech-shaped envelope soliton solution of the NLS Eq. (54) resembles the soliton of Eq. (53), and scales in space and time similarly. Such smooth SPEs solutions are weak gap solitons that can be formed in the HF and LF band gaps of the nonlinear MM.

D. Numerical simulations

We end up by giving some indicative numerical simulations concerning Eqs. (49) and (41); for further details, analysis and discussions on these

simulations see [28]. The utilized methodology for these simulations relies on Fourier transforming Eq. (49) with respect to τ, then solving the ensuing first order ODE in ζ (for each frequency), via a fourth-order Runge-Kutta scheme, and then Fourier transforming back to obtain $u(\zeta, \tau)$. We use the constant set of parameters $s = 1$, $F = 0.4$, and $\omega_p/\omega_{\mathrm{res}} = 10/1.45$ and $\varepsilon = 0.1$. The initial condition is depicted in Fig. 4a and is obtained as the exact breather solution of the SPE equation (Eq. (22) in [39], with $m = 0.32$). The evolution of the breather is shown both in Fig. 4b depicting the contour plot of the space-time evolution as well as in Fig. 4d, depicting the center of mass of the solution vs. time. This is a robust localized structure propagating through the domain with constant speed in time. We have also integrated Eq. (41) with the same type of breather initial profile. In the latter case, however, from the multiple scale ansatz, the initial condition of Eq. (41) was $E(0, \tau) = \varepsilon u(0, \tau)$, $E_z(0, \tau) = -u_\tau(0, \tau)$. The result of the time integration is shown in Fig. 4c. Evidently, the breather is robust in this setting as well, although its propagation speed is slightly smaller than that of the SPE breather. This result confirms our prediction that such "gap breathers" should be observable in the considered type of nonlinear MMs.

V. CONCLUSIONS AND FUTURE WORK

In conclusion, we analyzed electromagnetic (EM) pulse propagation in the left-handed (LH) regime and in the frequency band gaps of nonlinear metamaterials (MMs). In the LH regime, we used the reductive perturbation method to derive from Maxwell's equations a nonlinear Schrödinger (NLS) and a higher-order NLS (HNLS) equation. We found necessary conditions for the formation of bright or dark ultra-short solitons, as well as analytical expressions for these solitons. In the frequency band gaps [with negative (positive) linear permittivity (permeability) hence not allowing propagation of linear EM waves] we derived short-pulse equations (SPEs) describing the propagation of ultra-short pulses. We discussed the structure of the SPEs solutions and presented peakon-like and NLS-like solitons, which can be regarded as weak ultra-short gap solitons. The existence of such structures, indicates the possibility of nonlinear localization of EM waves in the gaps of nonlinear MMs.

Interesting subjects for future research may include a systematic study of the stability and dy-
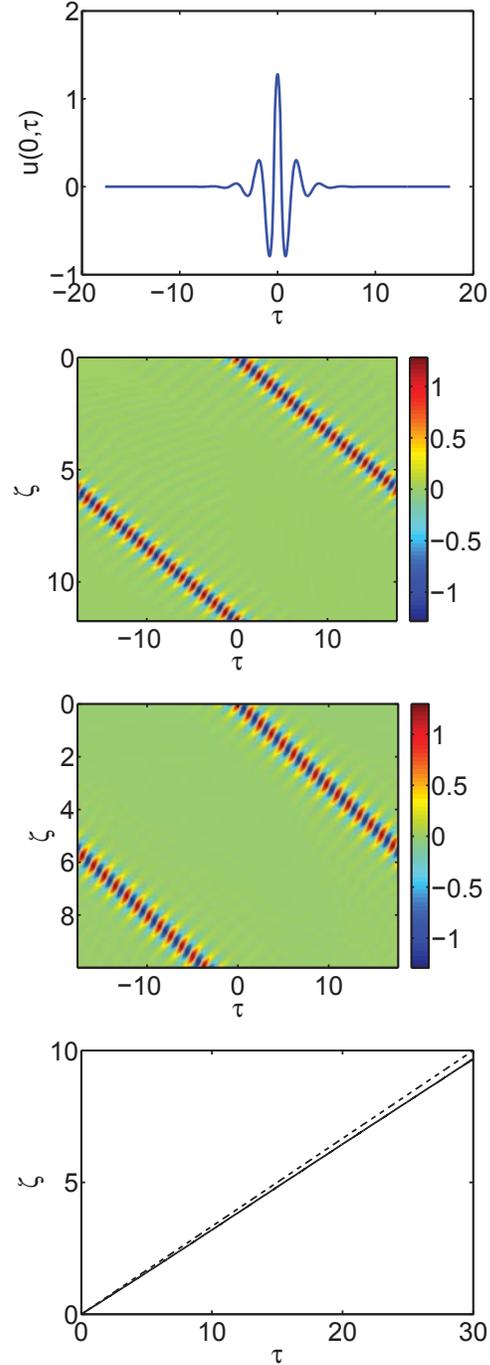
Fig. 4.   (a) Breather initial condition used in Eqs. (49) and (41), (b) space-time contour plot of the field evolution in Eq. (49), (c) the same as in (b) but for Eq. (41), (d) evolution of the breathers' center of mass in the two respective models by dotted and solid lines.

namics of the ultra-short or gap solitons, both in the framework of the HNLS or SPE as well as in that of Maxwell's equations. Moreover, it is also worthwhile to analyze the MM in its initial form of conducting wires and SRR's by a suitable CEM methodology (e.g. FDTD or MoM) in order to derive the propagating pulses and then compare them with (i) the solutions of the Maxwell's system (8) after assuming an effective medium approach for the material parameters (see Eqs. (3) and (4)), and (ii) the ultra-short solitons solutions derived in Section III-D. On the other hand, the consideration in the SPE models of the linear losses related to the permittivity and permeability of the MM would result to more general and interesting to investigate SPEs, which would also incorporate first temporal derivative terms of the electric field constituting control mechanisms of the evolution of the gap soliton.

Furthermore, it is highly worthwhile to carry out the corresponding investigations, concerning solitons formation and propagation, in other more complicated types of nonlinear MMs, exhibiting negative refractive index in a certain frequency band. For example, a representative material of this class is an isotropic chiral MM for which one of the two refractive indices can have a negative real part [40,41]; for the EM modelling of linear chiral media see [42], while for applications of negative index chiral MMs see [43]. For a nonlinear chiral MM, application of the reductive perturbation method leads to a system of two coupled NLS equations for the LH and RH Beltrami components of the EM field. For a sufficiently large chirality parameter there exists a certain spectral regime where the refractive index for the LH/RH Beltrami component is real and negative but that for the RH/LH Beltrami component is real and positive [40]. To this direction, it is interesting to approximate, inside the above regime, the coupled system of the NLS equations by a completely integrable system, and hence predict various classes of exact vector soliton solutions (bright-bright, dark-dark, as well as dark-bright solitons) that can be supported in the nonlinear chiral MM. Certain relevant investigations are reported in [44].

ACKNOWLEDGMENT

REFERENCES

[1] G. V. Eleftheriades and K. G. Balmain, *Negative-Refraction Metamaterials. Fundamental Principles and Applications*, John Wiley, New Jersey, 2005.

[2] C. Caloz and T. Itoh, *Electromagnetic Metamaterials. Transmission Line Theory and Microwave Applications*, John Wiley, New Jersey, 2006.

[3] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, "Composite Medium with Simultaneously Negative Permeability and Permittivity," *Phys. Rev. Lett.*, vol. 84, pp. 4184-4187, May 2000.

[4] A. Shelby, D. R. Smith, and S. Schultz, "Experimental Verification of a Negative Index of Refraction," *Science*, vol. 292, pp. 77-79, Apr. 2001.

[5] M. C. Tang, S. Q. Xiao, D. Wang, J. Xiong, K. Chen, and B. Z. Wang, "Negative Index of Reflection in Planar Metamaterial Composed of Single Split-Ring Resonators," *ACES J.*, vol. 26, pp. 250-258, Mar. 2011.

[6] V. Shalaev, "Optical Negative-Index Metamaterials," *Nature Photonics*, vol. 1, pp. 41-48, Jan. 2007.

[7] V. G. Veselago, "Electrodynamics of Substances with Simultaneously Negative Values of Sigma and Mu," *Sov. Phys. Usp.*, vol. 10, pp. 509, 1968.

[8] J. B. Pendry, "Negative Refraction Makes a Perfect Lens," *Phys. Rev. Lett.*, vol. 85, pp. 3966-3969, Oct. 2000.

[9] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire, "Metamaterials and Negative Refractive Index," *Science*, vol. 305, pp. 788-792, Aug. 2004.

[10] J. B. Pendry and D. R. Smith, "Reversing Light with Negative Refraction," *Phys. Today*, vol. 57, pp. 37-43, Jun. 2004.

[11] S. A. Ramakrishna, "Physics of Negative Refractive Index Materials," *Rep. Prog. Phys.*, vol. 68, pp. 449-521, Feb. 2005.

[12] C. M. Soukoulis, M. Kafesaki, and E. N. Economou, "Negative-Index Materials: New Frontiers in Optics," *Adv. Materials*, vol. 18, pp. 1941-1952, Aug. 2006.

[13] J. C. Liu, W. Shao, and B. Z. Wang, "A Dual-Band Metamaterial Design using Double SRR Structures," *ACES J.*, vol. 26, pp. 459-463, Jun. 2011.

[14] Y. Huang, G. Wen, T. Li, and K. Xie, "Positive-Negative-Positive Metamaterial Consisting of Ferrimagnetic Host and Wire Array," *ACES J.*, vol. 25, pp. 696-702, Aug. 2010.

[15] G. Antonini, "Reduced Order Models for Metamaterial Transmission Lines," *ACES Newsletter*, vol. 21, pp. 78-103, 2006.

[16] A. A. Zharov, I. V. Shadrivov, and Yu. S. Kivshar, "Nonlinear Properties of Left-Handed Metamaterials," *Phys. Rev. Lett.*, vol. 91, 037401, Jul. 2003.

[17] V. M. Agranovich, Y. R. Shen, R. H. Baughman, and A. A. Zakhidov, "Linear and Nonlinear Wave Propagation in Negative Refraction Metamaterials," *Phys. Rev. B*, vol. 69, 165112, Apr. 2004.

[18] N. Lazarides and G. P. Tsironis, "Coupled Nonlinear Schrödinger Field Equations for Electromagnetic Wave Propagation in Nonlinear Left-Handed Materials," *Phys. Rev. E*, vol. 71, 036614, Mar. 2005.

[19] I. V. Shadrivov, A. A. Zharov, N. A. Zharova, and Yu. S. Kivshar, "Nonlinear Left-Handed Metamaterials," *Radio Science*, vol. 40, RS3S90, May 2005.

[20] M. Scalora, M. S. Syrchin, N. Akozbek, E. Y. Poliakov, G. D'Aguanno, N. Mattiucci, M. J. Bloemer, and A. M. Zheltikov, "Generalized Nonlinear Schrodinger Equation for Dispersive Susceptibility and Permeability: Application to Negative Index Materials," *Phys. Rev. Lett.*, vol. 95, 013902, Jul. 2005.

[21] S. C. Wen, Y. J. Xiang, X. Y. Dai, Z. X. Tang, W. H. Su, and D. Y. Fan, "Theoretical Models for Ultrashort Electromagnetic Pulse Propagation in Nonlinear Metamaterials," *Phys. Rev. A*, vol. 75, 033815, Mar. 2007.

[22] I. V. Shadrivov and Yu. S. Kivshar, "Spatial Solitons in Nonlinear Left-Handed Metamaterials," *J. Opt. A: Pure Appl. Opt.*, vol. 7, pp. S68-S72, Feb. 2005.

[23] A. Aceves, "Optical Gap Solitons: Past, Present, and Future; Theory and Experiments," *Chaos*, vol. 10, pp. 584-589, Sep. 2000.

[24] S. Longhi, "Gap Solitons in Metamaterials," *Waves in Random and Complex Media*, vol. 15, pp. 119-126, Feb. 2005.

[25] T. Schäfer and C. E. Wayne, "Propagation of Ultra-Short Optical Pulses in Cubic Nonlinear Media," *Physica D*, vol. 196, pp. 90-105, Sep. 2004.

[26] N. Costanzino, V. Manukian, and C. K. R. T. Jones, "Solitary Waves of the Regularized Short Pulse and Ostrovsky Equations," *SIAM J. Math. Anal.*, vol. 41, pp. 2088-2106, 2009.

[27] N. L. Tsitsas, N. Rompotis, I. Kourakis, P. G. Kevrekidis, and D. J. Frantzeskakis, "Higher-Order Effects and Ultrashort Solitons in Left-Handed Metamaterials," *Phys. Rev. E*, vol. 79, 037601, Mar. 2009.

[28] N. L. Tsitsas, T. P. Horikis, Y. Shen, P. G. Kevrekidis, N. Whitaker, and D. J. Frantzeskakis, "Short Pulse Equations and Localized Structures in Frequency Band Gaps of Nonlinear Metamaterials," *Phys. Let. A*, vol. 374, pp. 1384-1388, Mar. 2010.

[29] I. Kourakis and P. K. Shukla, "Nonlinear Propagation of Electromagnetic Waves in Negative-Refraction-Index Composite Materials," *Phys. Rev. E*, vol. 72, 016626, Jul. 2005.

[30] T. Taniuti, "Reductive Perturbation Method and Far Fields of Wave-Equations," *Prog. Theor. Phys. Suppl.*, vol. 55, pp. 1-35, 1974.

[31] Y. Kodama and A. Hasegawa, "Nonlinear Pulse-Propagation in a Monomode Dielectric Guide," *IEEE J. Quantum Electron.*, vol. 23, pp. 510-524, May 1987.

[32] A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Clarendon Press, Oxford, 1995.

[33] J. V. Moloney and A. C. Newell, *Nonlinear Optics*, Westview Press, Oxford, 2004.

[34] G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, London, 2007.

[35] K. Hizanidis, D. J. Frantzeskakis, and C. Polymilis, "Exact Travelling Wave Solutions for a Generalized Nonlinear Schrodinger Equation," *J. Phys. A: Math. Gen.*, vol. 29, pp. 7687-7703, Dec. 1996.

[36] L. Kang, Q. Zhao, H. Zhao, and J. Zhou, "Magnetically Tunable Negative Permeability Metamaterial Composed by Split Ring Resonators and Ferrite Rods," *Opt. Express*, vol. 16, pp. 8825-8834, Jun. 2008.

[37] S. Zhang, W. Fan, K. J. Malloy, S. R. J. Brueck, N. C. Panoiu, and R. M. Osgood, "Near-Infrared Double Negative Metamaterials," *Opt. Express*, vol. 13, pp. 4922-4930, Jun. 2005.

[38] R. Camassa and D. D. Holm, "An Integrable Shallow-Water Equation with Peaked Solitons," *Phys. Rev. Lett.*, vol. 71, pp. 1661-1664, Sep. 1993.

[39] A. Sakovich and S. Sakovich, "Solitary Wave Solutions of the Short Pulse Equation," *J. Phys. A: Math. Gen.*, vol. 39, pp. L361-L367, Jun. 2006.

[40] T. G. Mackay and A. Lakhtakia, "Plane Waves with Negative Phase Velocity in Faraday Chiral Mediums," *Phys. Rev. E*, vol. 69, 026602, Feb. 2004.

[41] J. B. Pendry, "A Chiral Route to Negative Refraction," *Science*, vol. 306, pp. 1353-1355, Nov. 2004.

[42] V. Demir, A. Z. Elsherbeni, and E. Arvas, "FDTD Formulation for Dispersive Chiral Media using the Z Transform Method," *IEEE Trans. Antennas Propagat.*, vol. 53, pp. 3374-3384, Oct. 2005.

[43] N. Wongkasem, A. Akyurtlu, K. A. Marx, Q. Dong, J. Li, and W .D. Goodhue, "Development of Chiral Negative Refractive Index Metamaterials for the Terahertz Frequency Regime," *IEEE Trans. Antennas Propagat.*, vol. 55, pp. 3052-3062, Nov. 2007.

[44] N. L. Tsitsas, A. Lakhtakia, and D. J. Frantzeskakis, "Vector Solitons in Nonlinear Isotropic Chiral Metamaterials," *J. Phys. A: Math. Theor.*, vol. 44, 435203, Oct. 2011.

Nikolaos L. Tsitsas was born in Athens, Greece, in 1979. He received the Diploma and Ph.D. Degree in Electrical Engineering from the National Technical University of Athens (NTUA) in 2002 and 2006 respectively, and the M.Sc. Degree in Applied Mathematics from the National and Kapodistrian University of Athens in 2005.

From 2008 to 2011, he was an Adjunct Lecturer at the School of Applied Mathematical and Physical Sciences of the NTUA. From 2009 to 2011, he was an Adjunct Lecturer at the Hellenic Army Academy. Since 2012, he has been an Assistant Professor at the Department of Informatics of the Aristotle University of Thessaloniki. He is the author or coauthor of 30 papers in scientific journals and 36 papers in conference proceedings. His research interests include analytical and numerical methods in direct and inverse wave scattering and propagation theory as well as numerical solution techniques of integral and partial differential equations.

Dr. Tsitsas is a member of the IEEE, the Optical Society of America, the American Mathematical Society, and the Technical Chamber of Greece.

Optical Aspects of the Interaction of Focused Beams with Plasmonic Nanoparticles

Kürşat Şendur

Faculty of Engineering and Natural Sciences
Sabancı University, Istanbul, 34956, Turkey
sendur@sabanciuniv.edu

Abstract — In this study, the interaction of nanoparticles with focused beams of various angular spectra is investigated. This study demonstrates that the focused light can be used to manipulate the near-field radiation around nanoparticles. It is shown that suppressing strong lobes and enhancing weaker lobes is possible for spherical particles by altering the angular spectrum. This can have an impact on plasmonic applications where strong and weak fields are desired at specific locations.

Index Terms — Nanoparticles, nanoscale, plasmonics, radiative energy transfer, scattering, surface plasmons.

I. INTRODUCTION

Surface plasmons [1-2], optical nanoantennas [3-6], and radiative energy transfer at the nanoscale [7-9] have led to significant advances in nanotechnology. Plasmonics and interaction of light with nanoparticles have fascinated scientists because of their ability to manipulate light beyond the diffraction limit. Such an achievement at the nanoscale has enabled scientists to overcome technological barriers and expand the frontiers for scientific breakthroughs in near-field imaging [10], solar cells [11], optical data storage [12], heat assisted magnetic recording [13], light emitting devices [14], spectroscopy [15], medical applications [16], and bio-chemical sensors [17].

The interest in the interaction of photons with metallic nanoparticles for emerging nanotechnology applications is driven mainly due to the large enhancement and tight localization of electromagnetic fields in the vicinity of nanoparticles. Although there has been much effort to understand the effects of various parameters on the plasmon resonances of nanoparticles, the interaction of nanoparticles with a focused beam of light has not been significantly investigated in the context of particle plasmons. Similarly, the interaction of a tightly focused beam of incident light with metallic prolate spheroids has also been largely overlooked in the literature.

In this study, we investigate the effect of the angular spectrum of a focused beam of light on the near-field radiation from spherical nanoparticles and prolate spheroids. This paper is an extended version of previous conference papers [18, 19]. Focused beams with various angular spectra are used in this study to understand field distributions and their formations over the nanoparticles.

II. RICHARDS-WOLF VECTOR FIELD FORMALISM

To obtain an accurate representation of a tightly focused beam of light, the theory established by Richards and Wolf [20, 21] is used. In this approach, rays that are incident onto a lens are collected and focused based on the rules of geometric optics. After each ray is diffracted by the lens system, the overall contribution is calculated by summing up the individual rays. Formulas based on the Richards and Wolf vector field theory have been previously used in the literature [22-25] for focused beams of various polarizations.

A linearly polarized beam in the *x*-direction, which is focused onto the *x*-*y* plane and propagating in the *z*-direction is used. The beam is cylindrically symmetric around the *z*-axis with the focal point at (0, 0, 0). The incident electric field near the focus is given as [24]

$$\vec{E}(\vec{r}) = -\frac{i}{\lambda} \int_0^\alpha \int_0^{2\pi} \vec{a}(\theta,\phi)\, e^{i\vec{k}\cdot\vec{r}} \sin\theta\, d\phi\, d\theta \,, \quad (1)$$

for a linearly polarized focused beam, where α is the half angle of the beam, r is the observation point, and

$$\vec{a}(\theta,\phi) = \begin{bmatrix} \cos\theta\cos^2\phi + \sin^2\phi \\ \cos\theta\cos\phi\sin\phi - \cos\phi\sin\phi \\ \sin\theta\cos\phi \end{bmatrix} \sqrt{\cos\theta} \cdot \quad (2)$$

The half-beam angle α, which defines the upper limit of the integral in Eq. (1), represents the span of angles at which the corresponding plane wave has a non-zero contribution to the focused field. The half-beam angle is illustrated in Fig. 1. Any lens has a finite size, which results in the integral in Eq. (1) having an upper limit α. The half-beam angle is determined by the physical configuration and the size of the lens system. The largest α used in this study is 60°, which yields a full-width half-maximum beam waist of 405 nm at a 700 nm wavelength. The amplitude of the focused beam is normalized so that the value at the focus is 1 V/m.
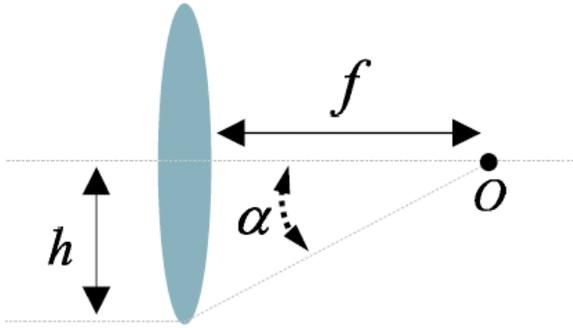
Fig. 1. A graphical illustration of the half-beam angle, α. The focus of the lens is illustrated with the point O.

The spatial distributions of incident focused beams are plotted in Fig. 2 for various α. The results in Fig. 2 illustrate the field distributions in the absence of a metallic particle. The results in subsequent figures illustrate the impact of placing a spherical nanoparticle on the electric field distributions. In Figs. 2(a), (c), and (e), the x-component of the electric field $E_x(x, y, 0)$ is plotted on the x-y plane for $\alpha=0°$, $\alpha=30°$, and $\alpha=60°$, respectively. The beam with $\alpha=0°$ corresponds to a plane wave. As α is increased, the beam becomes more tightly focused. Similarly

in Figs. 2(b), (d), and (f), the x-component of the electric field $E_x(x, 0, z)$ is plotted on the x-z plane for $\alpha=0°$, $\alpha=30°$, and $\alpha=60°$, respectively. The field distribution is elongated in the z-direction due to the poor axial resolution of lenses [22].
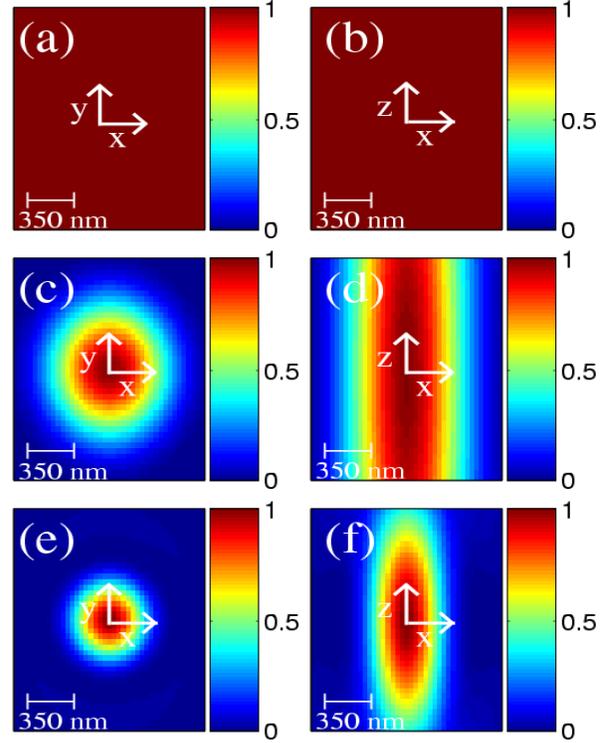
Fig. 2. (a) $E_x(x, 0, z)$ on the x-y plane for $\alpha = 0°$, (b) $E_x(x, 0, z)$ on the x-z plane for $\alpha = 0°$, (c) $E_x(x, 0, z)$ on the x-y plane for $\alpha = 30°$, (d) $E_x(x, 0, z)$ on the x-z plane for $\alpha =30°$, (e) $E_x(x, 0, z)$ on the x-y plane for $\alpha = 60°$, and (f) $E_x(x, 0, z)$ on the x-z plane for $\alpha =60°$.

In the representation in Eq. (1), each ray is identified by angles θ and ϕ depending on their incidence angle. Only the rays incident on the lens can be collected. The rays beyond the size of the lens cannot be focused by the lens system. The half-beam angle imposes a cut-off for the upper limit of the integral in Eq. (1). Therefore, the half-beam angle incorporates the physics of the lens system into Eq. (1). The finite size of the lens results in the integral in Eq. (1) having an upper limit $\theta = \alpha$ where $\alpha \le 90°$. If the size of the lens approaches to infinity, the upper limit of the integral in Eq. (1) will approach to 90°. To study the interaction of nanoparticles with incident beams with different angular spectra, the half

beam angle will be varied.

III. RESULTS

The plasmon resonances of nanoparticles have been investigated in the literature to understand the effects of various parameters [26] and to engineer the spectral response [27-29]. Until recently, the interaction of nanoparticles with a focused beam of light has not been significantly investigated in the context of particle plasmons. With emerging potential applications, there is an increased interest in the interaction of focused beams and nanoparticles [25, 30-38] as well as developing computational solutions to numerical aspects of scattering problems [39, 40]. To analyze the effect of the angular spectrum, a silver nanoparticle is illuminated using a focused beam of light with small and large α. The focused beam propagates in the z-direction, and is polarized in the x-direction.

comparison of Figs. 3(a) and (b) suggests that the field distribution at $\lambda = 400$ nm for $\alpha=5°$ shows a significant difference compared to the results of $\alpha=60°$. In Figs. 3(c)-(f), deviations are observed at other wavelengths as well. The $E_y(x, 0, z)$ component is negligible for the solutions. The impact of altering the angular spectrum is more drastic for the $E_z(x, 0, z)$ component, as shown in Fig. 4. For example, when the angular spectrum is narrowly distributed along the direction of propagation, as shown in Fig. 4(e), two stronger lobes are observed at the back of the spherical particle. As α is increased, and therefore the angular spectrum is widened, the stronger lobes are moved from the back of the particle to the front of the particle, as shown in Fig. 4(f). This was achieved without changing the frequency, geometry, or composition of the particle.

Fig. 3. $E_x(x, 0, z)$ on the x-z plane for various [α (°), λ (nm)]: (a) $E_x(x, 0, z)$ for [5, 400], (b) $E_x(x, 0, z)$ for [60, 400], (c) $E_x(x, 0, z)$ for [5, 500], (d) $E_x(x, 0, z)$ for [60, 500], (e) $E_x(x, 0, z)$ for [5, 600], and (f) $E_x(x, 0, z)$ for [60, 600].
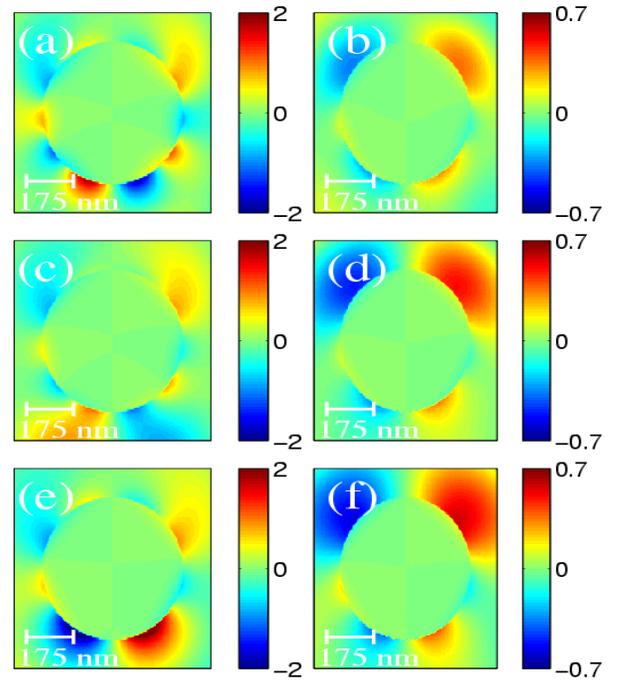
Fig. 4. $E_z(x, 0, z)$ on the x-z plane for various [α(°), λ (nm)]: (a) $E_z(x, 0, z)$ for [5, 400], (b) $E_z(x, 0, z)$ for [60, 400], (c) $E_z(x, 0, z)$ for [5, 500], (d) $E_z(x, 0, z)$ for [60, 500], (e) $E_z(x, 0, z)$ for [5, 600], and (f) $E_z(x, 0, z)$ for [60, 600].

In Fig. 3, the electric field is computed at various wavelengths on the x-z plane for a silver sphere with a 250 nm radius. The field distribution $E_x(x, 0, z)$ is plotted for $\alpha=5°$ and $\alpha=60°$. A

The results in Figs. 3 and 4 demonstrate that suppressing strong lobes and enhancing weaker lobes is possible by altering the angular spectrum. These results have important implications for

plasmonic applications. Some plasmonic applications require strong fields at specific locations. For example, in heat assisted magnetic recording (HAMR) [13] strong fields are necessary at the air-bearing surface [41] to change the magnetic properties of the recording medium. The results in this study suggest that the location of strong fields, such as the air-bearing fields used in HAMR, can be adjusted using the angular spectrum of the incident field. Additional potential applications include single molecule spectroscopy [42, 43], single molecule fluorescence enhancement [44], plasmonic waveguides [45-47], directional plasmonic emitters [48-50], plasmonic routers and switches, which can benefit from manipulating the location of strong optical spots using the angular spectrum of the incident beam.

nanoparticle is not impacted by the wide range of spectral components, even if the incident light has a wide angular spectrum. A very small particle does not feel the variation in the field, since it is much smaller than the variations shown in Fig. 2. However, as the sphere gets larger it will start to interact with various components of the angular spectrum.
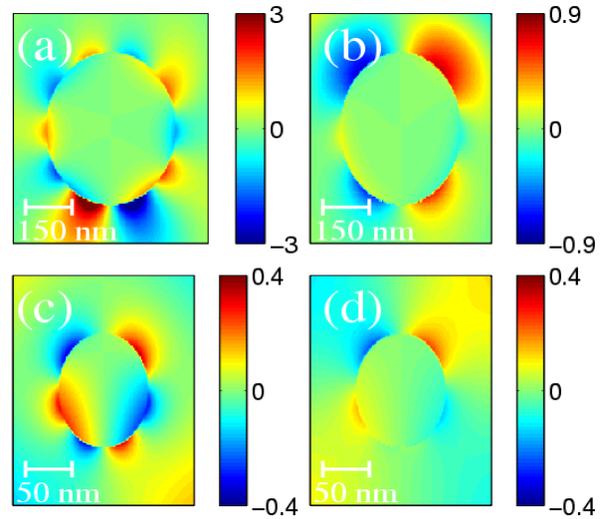
Fig. 6. $E_z(x, 0, z)$ on the x-z plane for various [α (°), r (nm)]: (a) [5, 200], (b) [60, 200], (c) [5, 50], and (d) [60, 50].

Fig. 5. $E_x(x, 0, z)$ on the x-z plane for various [α (°), r (nm)]: (a) [5, 200], (b) [60, 200], (c) [5, 50], and (d) [60, 50].

$E_x(x, 0, z)$ and $E_z(x, 0, z)$ are plotted for various particle sizes at $\lambda = 400$ nm in Figs. 5 and 6, respectively. The results suggest that the $E_x(x, 0, z)$ and $E_z(x, 0, z)$ distributions vary for small and large α, even when the particle size is small. The difference becomes smaller as the particle size decreases. If a particle smaller than $\lambda/10$ is placed at the focus of a tightly focused beam of light, the interaction becomes quasistatic. Therefore, for small particles the results of the small and large α begin to resemble each other. For very small particles smaller than $\lambda/10$, the

Figure 7(a) shows the results for a linearly polarized plane wave at the incidence angle 0°. Figures 7(b)-(d) illustrate the results for a linearly polarized focused wave with half beam angles of 5°, 45°, and 60°. A comparison of Figs. 7(a) and (b) suggests that the result of the small α is similar to that of a plane wave at $\theta = 0°$. Figures 7(c) and (d) show that the plasmon distribution for a large α deviates from that of a plane wave. This deviation can be interpreted using Figs. 7(a), (e), and (f). As shown in Figs. 7(a), (e), and (f), the field distribution changes for plane waves of different incidence angles. Plane waves with different angles are scaled and summed up based on the angular spectrum of the focused beam. For small half-beam angle α, the plane wave at an incidence angle 0° is the dominant contributor, as shown in Fig. 7(b). Contributions from larger incidence angles start to dominate as the angular spectrum gets wider, which impacts the distributions in Figs. 7(c) and (d).
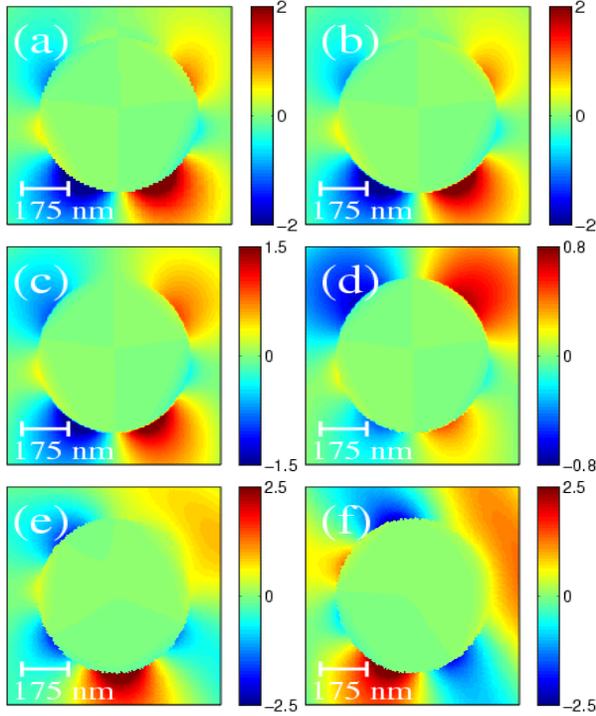
Fig. 7. $E_z(x, 0, z)$ on the x-z plane: (a) plane wave at 0°, (b) focused beam with $\alpha = 0°$, (c) focused beam with $\alpha = 45°$, (d) focused beam with $\alpha = 60°$, (e) plane wave at 30°, and (f) plane wave at 60°.

Although the field distributions over the spheres resulting from off-angle plane waves in Figs. 7(e) and (f) are asymmetric, the field distributions over spheres excited with focused beams in Fig. 7 are symmetric. The primary reason why the asymmetry in the results in Figs. 7(e) and (f) does not result in asymmetry in the focused beam results can be explained as follows: In Figs. 7(e) and (f), the results for the plane waves with non-zero incidence angles are presented in the global coordinate system, in which the results are asymmetric. For a plane wave with a non-zero incidence angle, however, the results are symmetric in the rotated coordinate system, which is rotated by the incidence angle with respect to the global coordinate system. The asymmetry in the individual plane wave results is not reflected in the focused beam results, because for each (θ, ϕ) ray, there is a corresponding $(\theta, \phi+\pi)$ ray. The sum of these rays, each of which are asymmetric in the global x-z coordinate system, result in a symmetric distribution. Therefore, the asymmetry of the individual components shown in Figs. 7(e) and (f)

do not result in asymmetry in the focused beam results.
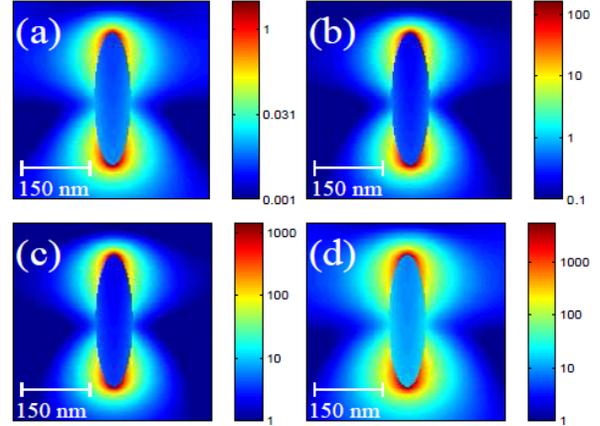
Fig. 8. Intensity distributions for a prolate spheroid. The distributions are given for various half-beam angles: (a) $\alpha = 15°$, (b) $\alpha = 30°$, (c) $\alpha = 45°$, and (d) $\alpha = 60°$.

In Fig. 8, the impact of the angular spectrum distribution of the incident beam on the near-field radiation of a prolate spheroidal nanoparticle is studied. In Fig. 8, the total intensity profile is plotted on the x-z plane, which passes through the center of a prolate spheroid particle with a major/minor axis ratio of 5. Intensity distributions for a prolate spheroid are given for various half-beam angles. In this figure, the prolate spheroids are illuminated with a radially focused beam with half-beam angles $\alpha = 15°$, $\alpha = 30°$, $\alpha = 45°$, and $\alpha = 60°$ The field distributions in Fig. 8 are normalized to the value of the incident intensity at the focus. The results suggest that the electric field distribution does not change as the half-beam angle is increased. The amplitude of the near field electric field distribution, however, increases as the half-beam angle is increased. The angular spectrum of the incident beam is tight for $\alpha = 15°$, becoming wider as the half beam angle is increased. Therefore, the incident wave amplitude onto the particle and resulting total field amplitude increases as the half-beam angle increases.

VI. CONCLUSION

In summary, the angular spectrum of the incident beam had a significant impact on the plasmon distribution of nanoparticles. Beams with

narrow and wide angular spectra interacted differently with a nanoparticle. For a focused beam with a small α, the results were similar to those of a plane wave. As α was increased, the results differed significantly from those of a plane wave. The results suggest that it is possible to manipulate plasmon distributions by adjusting the angular spectrum of an incident focused beam. On the other hand, for prolate spheroids the electric field distribution does not change as the half-beam angle is increased. The amplitude of the near-field electric field distribution, however, increases as the half-beam angle is increased.

ACKNOWLEDGMENT

REFERENCES

[1] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface Plasmon Subwavelength Optics," *Nature*, vol. 424, pp. 824-830, 2003.

[2] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary Optical Transmission through Sub-Wavelength Hole Arrays," *Nature*, vol. 391, pp. 667-669, 1998.

[3] P. Bharadwaj, B. Deutsch, and L. Novotny, "Optical Antennas," *Adv. Opt. Photon.*, vol. 1, pp. 438-483, 2009.

[4] L. Novotny and N. van Hulst, "Antennas for Light," *Nature Photon.*, vol. 5, pp. 83-90, 2011.

[5] H. Fischer and O. J. F. Martin, "Engineering the Optical Response of Plasmonic Nanoantennas," *Opt. Express*, vol. 16, pp. 9144-9154 (2008).

[6] L. Novotny, "Effective Wavelength Scaling for Optical Antennas," *Phys. Rev. Lett.*, vol. 98, pp. 266802, 2007.

[7] E. Rousseau, A. Siria, G. Jourdan, S. Volz, F. Comin, J. Chevrier, and J.-J. Greffet, "Radiative Heat Transfer at the Nanoscale," *Nature Photon.*, vol. 3, pp. 514-517, 2009.

[8] S. Shen, A. Narayanaswamy, and G. Chen, "Surface Phonon Polariton Mediated Energy Transfer between Nanoscale Gaps," *Nano Lett.*, vol. 9, pp. 2909-2913, 2009.

[9] M. Francoeur, M. P. Menguc, and R. Vaillon, "Near-Field Radiative Heat Transfer Enhancement via Surface Phonon Polaritons Coupling in Thin Films," *Appl. Phys. Lett.*, vol. 93, pp. 043109, 2008.

[10] A. Hartschuh, E. J. Sanchez, X. S. Xie, and L. Novotny, "High-Resolution Near-Field Raman Microscopy of Single-Walled Carbon Nanotubes," *Phys. Rev. Lett.*, vol. 90, pp. 095503, 2003.

[11] H. Atwater and A. Polman, "Plasmonics for Improved Photovoltaic Devices," *Nat. Mater.*, vol. 9, pp. 205-213, 2010.

[12] C. Peng, "Surface-Plasmon Resonance of a Planar Lollipop Near-Field Transducer," *Appl. Phys. Lett.*, vol. 94, pp. 171106, 2009.

[13] K. Sendur, C. Peng, and W. Challener, "Near-Field Radiation from a Ridge Waveguide Transducer in the Vicinity of a Solid Immersion Lens," *Phys. Rev. Lett.*, vol. 94, pp. 043901, 2005.

[14] X. Gu, T. Qiu, W. Zhang, and P. K. Chu, "Light-Emitting Diodes Enhanced by Localized Surface Plasmon Resonance," *Nanoscale Res. Lett.*, vol. 6, pp. 199, 2011.

[15] K. A. Willets and R. P. V. Duyne, "Localized Surface Plasmon Resonance Spectroscopy and Sensing," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 267-297, 2007.

[16] A. M. Gobin, M. H. Lee, N. J. Halas, W. D. James, R. A. Drezek, and J. L. West, "Near-Infrared Resonant Nanoshells for Combined Optical Imaging and Photothermal Cancer Therapy," *Nano Lett.*, vol. 7, pp. 1929-1934, 2007.

[17] N. Liu, M. L. Tang, M. Hentschel, H. Giessen, and A. P. Alivisatos, "Nanoantenna-Enhanced Gas Sensing in a Single Tailored Nanofocus," *Nat. Mater.*, vol. 10, pp. 631-636, 2011.

[18] K. Sendur, A. Şahinoz, E. Unlu, S. Yazici, and M. Gulhan, "Near-Field Radiation from Nano-Particles and Nano-Antennas Illuminated with a Focused Beam of Light," in *Materials for Nanophotonics — Plasmonics, Metamaterials and Light Localization*, edited by M. Brongersma, L. Dal Negro, J. M. Fukumoto, L. Novotny (Mater. Res. Soc. Symp. Proc. Volume 1182, Warrendale, PA, 2009), 1182-EE16-03.

[19] K. Sendur, "Optical Aspects of the Interaction of Focused Beams with Plasmonic Nanoparticles," *Proceedings of the International Workshop on Computational Electromagnetics*, Izmir, Turkey, August 2011.

[20] E. Wolf, "Electromagnetic Diffraction in Optical Systems I. An Integral Representation of the Image Field," *Proc. Roy. Soc. London Ser. A*, vol. 253, pp. 349-357, 1959.

[21] B. Richards and E. Wolf, "Electromagnetic Diffraction in Optical Systems II. Structure of the

Image Field in an Aplanatic System," *Proc. Roy. Soc. London Ser. A*, vol. 253, pp. 358-379, 1959.

[22] L. Novotny and B. Hecht, *Principles of Nano-Optics*, Cambridge University Press, New York, 2006.

[23] K. S. Youngworth and T. G. Brown, "Focusing of High Numerical Aperture Cylindrical-Vector Beams," *Opt. Express*, vol. 7, pp. 77-87, 2000.

[24] I. Ichimura, S. Hayashi, and G. S. Kino, "High-Density Optical Recording using a Solid Immersion Lens," *Appl. Opt.*, vol. 36, pp. 4339-4348, 1997.

[25] W. A. Challener, I. K. Sendur, and C. Peng, "Scattered Field Formulation of Finite Difference Time Domain for a Focused Light Beam in a Dense Media with Lossy Materials," *Opt. Express*, vol. 11, pp. 3160-3170 (2003).

[26] K. L. Kelly, E. Coronado, L. L. Zhao, and G. C. Schatz, "The Optical Properties of Metal Nanoparticles: The Influence of Size, Shape, and Dielectric Environment," *J. Phys. Chem. B*, vol. 107, pp. 668–677, 2003.

[27] S. J. Oldenburg, R. D. Averitt, S. L. Westcott, and N. J. Halas, "Nanoengineering of Optical Resonances," *Chem. Phys. Lett.*, vol. 288, pp. 243-247, 1998.

[28] E. Prodan, C. Radloff, N. J. Halas, and P. Nordlander, "A Hybridization Model for the Plasmon Response of Complex Nanostructures," *Science*, vol. 302, pp. 419-422, 2003.

[29] P. Nordlander, C. Oubre, E. Prodan, K. Li, and M. I. Stockman, "Plasmon Hybridization in Nanoparticle Dimers," *Nano Lett.*, vol. 4, pp. 899-903, 2004.

[30] K. Sendur, W. Challener, and O. Mryasov, "Interaction of Spherical Nanoparticles with a Highly Focused Beam of Light," *Opt. Express*, vol. 16, pp. 2874-2886, 2008.

[31] J. Lerme, G. Bachelier, P. Billaud, C. Bonnet, M. Broyer, E. Cottancin, S. Marhaba, and M. Pellarin, "Optical Response of a Single Spherical Particle in a Tightly Focused Light Beam: Application to the Spatial Modulation Spectroscopy Technique," *J. Opt. Soc. Am. A*, vol. 25, pp. 493-514, 2008.

[32] N. M. Mojarad, V. Sandoghdar, and M. Agio, "Plasmon Spectra of Nanospheres under a Tightly Focused Beam," *J. Opt. Soc. Am. B*, vol. 25, pp. 651-658, 2008.

[33] D. Khoptyar, R. Gutbrod, A. Chizhik, J. Enderlein, F. Schleifenbaum, M. Steiner, and A. J. Meixner, "Tight Focusing of Laser Beams in a λ/2-Microcavity," *Opt. Express*, vol. 16, pp. 9907-9917, 2008.

[34] A. V. Failla, H. Qian, H. H. Qian, A. Hartschuh, and A. J. Meixner, "Orientational Imaging of Subwavelength Au Particles with Higher Order

[35] N. M. Mojarad and M. Agio, "Tailoring the Excitation of Localized Surface Plasmon-Polariton Resonances by Focusing Radially-Polarized Beams," *Opt. Express*, vol. 17, pp. 117-122, 2009.

[36] K. Sendur, "An Integral Equation Based Numerical Solution for Nanoparticles Illuminated with Collimated and Focused Light," *Opt. Express*, vol. 17, pp. 7419-7430, 2009.

[37] K. Sendur and A. Sahinöz, "Interaction of Radially Polarized Focused Light with a Prolate Spheroidal Nanoparticle," *Opt. Express*, vol. 17, pp. 10910-10925, 2009.

[38] K. Sendur and E. Baran, "Near-Field Optical Power Transmission of Dipole Nano-Antennas," *Appl. Phys. B*, vol. 96, pp. 325-335, 2009.

[39] B. Alavikia and O. M. Ramahi, "Fundamental Limitations on the Use of Open-Region Boundary Conditions and Matched Layers to Solve the Problem of Gratings in Metallic Screens," *Appl. Comput. Electrom.*, vol. 25, pp. 652-658, 2010.

[40] S. S. Abdallah, A. Iolov, K. Bizheva, and O. M. Ramahi, "Effect of Numerical Dispersion in FDTD Simulations of Light Scattering from Photoreceptors," *Appl. Comput. Electrom.*, vol. 25, pp. 388-394, 2010.

[41] K. Sendur and P. Jones, "Effect of Fly Height and Refractive Index on the Transmission Efficiency of Near-Field Optical Transducers," *Appl. Phys. Lett.*, vol. 88, pp. 091110, 2006.

[42] W. E. Moerner and L. Kador, "Optical Detection and Spectroscopy of Single Molecules in a Solid," *Phys. Rev. Lett.*, vol. 62, pp. 2535–2538, 1989.

[43] W. E. Moerner and M. Orrit, "Illuminating Single Molecules in Condensed Matter," *Science*, vol. 283, pp. 1670-1676, 1999.

[44] S. Kühn, U. Hakanson, L. Rogobete, and V. Sandoghdar, "Enhancement of Single-Molecule Fluorescence using a Gold Nanoparticle as an Optical Nanoantenna," *Phys. Rev. Lett.*, vol. 97, pp. 017402, 2006.

[45] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, and A. A. G. Requicha, "Local Detection of Electromagnetic Energy Transport Below the Diffraction Limit in Metal Nanoparticle Plasmon Waveguides," *Nature Mater.*, vol. 2, pp. 229-232, 2003.

[46] S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. G. Requiche, and H. A. Atwater, "Plasmonics - Route to Nanoscale Optical Devices," *Adv. Mater.*, vol. 13, pp. 1501-1505, 2001.

[47] S. A. Maier and H. A. Atwater, "Plasmonics: Localization and Guiding of Electromagnetic

Energy in Metal/Dielectric Structures," *J. Appl. Phys.*, vol. 98, pp. 011101, 2005.

[48] T. H. Taminiau, F. D. Stefani, and N. F. van Hulst, "Enhanced Directional Excitation and Emission of Single Emitters by a Nano-Optical Yagi-Uda Antenna," *Opt. Express*, vol. 6, pp. 10858-10866, 2008.

[49] T. Shegai, V. D. Miljkovi, K. Bao, H. Xu, P. Nordlander, P. Johansson, and M. Kall, "Unidirectional Broadband Light Emission from Supported Plasmonic Nanowires," *Nano Lett.*, vol. 11, pp. 706-711, 2011.

[50] A. G. Curto, G. Volpe, T. H. Taminiau, M. P. Kreuzer, R. Quidant, and N. F. van Hulst, "Unidirectional Emission of a Quantum Dot Coupled to a Nanoantenna," *Science*, vol. 329, pp. 930-933, 2010.

Kürşat Şendur received his B.S., M.S., and Ph.D. degrees from Middle East Technical University, Bilkent University, and The Ohio State University, respectively. In 2002, he joined Seagate Technology Research Center, USA where he worked on emerging data storage technologies. Between 2005-2007 he worked at Advanced Microsensors, Boston, USA on thin-film sensor technologies. Since 2007, he has been a faculty member at Sabancı University, Istanbul, Turkey.

Kürşat Şendur received the European Community FP7 Marie Curie International Reintegration Grant in 2007, the Turkish Academy of Sciences Outstanding Young Scientist Award in 2009, the ODTÜ Parlar Foundation Young Researcher Award in 2010, and the İTÜ FABED Foundation Eser Tümen Outstanding Young Scientist Award in 2011.

Surface Integral Equation Solvers for Large-Scale Conductors, Metamaterials and Plasmonic Nanostructures

José M. Taboada[1], Marta G. Araújo[2], Javier Rivero[1], Luis Landesa[1], and Fernando Obelleiro[2]

[1] Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain
tabo@unex.es, jrivero@unex.es, llandesa@unex.es

[2] E.E. Telecomunicación, University of Vigo, 36310, Spain
martaga@com.uvigo.es, obi@com.uvigo.es

Abstract – Surface integral equation (SIE) approaches for the accurate solution of different problems in computational electromagnetics are addressed. First, an efficient message passing interface (MPI)/OpenMP parallel implementation of the multilevel fast multipole algorithm-fast Fourier transform (MLFMA-FFT) is presented for the solution of large-scale conducting bodies. By combining the high scalability of the fast multipole method-FFT (FMM-FFT) with the high efficiency of MLFMA, challenging problems up to one billion unknowns are solved using a parallel supercomputer. Second, looking for the extension of these rigorous approaches to new demanded areas, the SIE method is successfully applied to the solution of left-handed metamaterials and plasmonic nanostructures. Numerical examples are presented to confirm the validity and versatility of this approach for the accurate resolution of problems in the context of leading-edge nanoscience and nanotechnology applications.

Index Terms ─ Fast multipole methods, metamaterials, method of moments, parallel programming, plasmonics, supercomputing, surface integral equations.

I. INTRODUCTION

The rigorous solution of radiation and scattering problems using surface-integral-equation (SIE) formulations has commanded a great attention in computational electromagnetics for a long time. Integral-equation techniques, such as the method of moments (MoM) [1], provide accurate results when the surfaces of the problem are properly discretized in terms of the wavelength. Usually, the well-known Rao-Wilton-Glisson (RWG) [2] basis functions defined over planar triangles are applied to expand the unknown surface current density.

In the case of electromagnetic problems involving large-scale conducting bodies, the fast multipole method (FMM) [3] and its multilevel version, the multilevel fast multipole algorithm (MLFMA) [4, 5], are commonly applied in the framework of the iterative solution of the MoM matrix system. The development of these fast (low computational cost) algorithms has gone hand in hand with the constant advances in computer science and technology.

Because of this simultaneous growth, the new methods must be not only fast, efficient and with low complexity, but they also must be able to benefit from the computational capabilities of current high performance computing (HPC) computers and supercomputers. As a consequence, a great attention was focused on the parallelization of MLFMA over shared, distributed, and mixed memory computers. Important advances were achieved in this regard; see among others [6-14].

Another numerical technique that has gained interest in recent years, due to its natural high scalability propensity, is the FMM-fast Fourier transform (FMM-FFT) [15-17]. Using hybrid message passing interface (MPI)/OpenMP parallel implementations of this algorithm and other

proposed variants (such as nested FMM-FFT and MLFMA-FFT), the authors have demonstrated that this approach constitutes a good alternative to benefit from massively parallel distributed computers, achieving the solution of problems with more than 150, 620, and 1000 million unknowns [17-19].

Up to now, we have been able to analyze very large electromagnetic problems in high-frequency ranges, even reaching the terahertz (THz) region [19]. But growing up in frequency is not only a matter of increasing the number of unknowns. We must reformulate our codes in order to be able to extend their scope of application to emerging fields and areas, such as nanoscience and nanotechnology. In this context, the research in the optical plasmonic properties of metallic nanoparticles and artificial materials (metamaterials) has generated an increasing interest, due to their ability to route and manipulate light at nanometer length scale surpassing the classical diffraction limit.

The optical response of metals is quite different from the metallic conductivity observed at low frequencies. At optical frequencies the penetration of fields can no longer be neglected and the plasmonic optical properties of metals make it impossible to directly downscale the radio-frequency and microwave solutions. Otherwise, by properly engineering the underlying subunits, metamaterials can exhibit exotic optical properties that are not attainable in naturally occurring materials, such as prominent magnetic response or negative index of refraction. To date, the most common way to rigorously accomplish the resolution of such problems has been the use of differential-equation formulations.

Although not yet widespread in optics, the MoM integral-equation approach could bring important advantages for the analysis of homogeneous or piecewise homogeneous plasmonic and/or metamaterial bodies. According to this research interest, our recent efforts were headed to extend the SIE-MoM approach to the analysis of composite piecewise homogeneous metamaterial and plasmonic objects. So, in [20], we successfully applied this formulation to the solution of homogenized left-handed metamaterials. More recently, we generalized this approach to the simulation of arbitrary configurations of composite plasmonic nanoparticles in [21]. From these previous works, it is clear that the SIE-MoM approach can yield very efficient and highly accurate representations of these new electromagnetic problems, taking into account all the physical constraints as determined by the shape and the complex dielectric constants of the particles.

In this paper, we first make a review of the latter method we have proposed for the electromagnetic analysis of large-scale conducting bodies using the SIE formulation, namely, MLFMA-FFT. Then, we focus on the works we are doing presently to extend these techniques to the analysis of non-conventional (generalized) media, such as the left-handed metamaterials and plasmonic nanoparticles at optical frequencies.

II. MLFMA-FFT FOR THE ANALYSIS OF LARGE-SCALE CONDUCTING BODIES

The aim of MLFMA-FFT is to combine the best of MLFMA and FMM-FFT to obtain a highly scalable method with low computational cost, which is suitable to take advantage of existing HPC computers. Initially, the algorithm requires a multilevel octree decomposition of the geometry. The far-field interactions are obtained at the coarsest level of the geometry partition using a global distributed FMM-FFT algorithm [18]. A parallelization strategy based on the distribution of fields, by distributing the k-space samples among nodes, has been considered for this stage. Regarding the near-field contributions, they are obtained at the finer levels of the octree by using one or more local executions of MLFMA defined inside each shared-memory computing node.

For the parallel implementation, we have selected a hybrid parallel programming combining MPI with the OpenMP standard, which fits perfectly with mixed-memory computer systems. Thus, the parallel MLFMA-FFT approach provides a significant reduction of the computational complexity while maintaining a high scalability behavior for the solution of extremely large problems using supercomputers.

The main difference of MLFMA-FFT with regard to the original FMM-FFT algorithm is given by the procedure used for dealing with the near couplings. In the MLFMA-FFT method, this task is performed by MLFMA. The partition of work is based on a distribution by octree groups at

the coarsest level. This distribution is propagated downward to the finest level. The partial near contributions to the matrix-vector product (MVP) are then computed by several MLFMAs defined over the finer levels (from the finest to one before the coarsest) operating strictly inside each shared-memory computing node. In this way, MLFMA is never distributed among nodes, but it is only applied in shared-memory computations (so it does not suffer from poor parallel scaling).
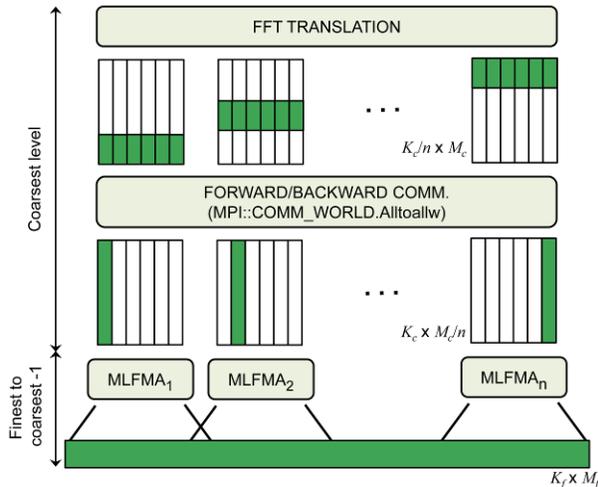
Fig. 1. MLFMA-FFT parallel algorithm.

Figure 1 illustrates the parallel operation of MLFMA-FFT. Let us define parameters K, N, M, and n as the number of k-space samples, the total number of unknowns of the problem, the number of non-empty groups, and the number of nodes, respectively. The subscripts c and f are added to indicate the coarsest and finest levels, respectively. It can be observed in the figure that MLFMA work in levels comprised between the finest and the preceding to the coarsest one. The near coupling contributions to the MVP at the coarsest level are computed at the finer levels by MLFMA inside each node. After the required interpolation of the outcoming fields to the coarsest level, each node has the complete set of directions, K_c, for its assigned M_c/n observation groups. At this point, an all-to-all communication is performed in order to obtain the partial K_c/n samples assigned to each node for all the M_c groups at the coarsest level. This switches from the group-driven distribution to the field-driven distribution.

The required inter-node communications are efficiently carried out during the MVP in a single step by using the asymmetric *MPI::Alltoallw* operation. The *Alltoallw* high-level command makes possible to carry out all the required communications without latency periods or explicit synchronization because of the efficient management provided by the MPI library (we are using the 2.3 version of HP-MPI).

When the 3D-FFT translation is done, the complementary all-to-all communication allows switching back to the group-driven distribution at the lower levels. After that, each node will have the complete K_c directions of the incoming fields for its M_c/n assigned observation boxes at the coarsest level. Starting from this level, these fields are recursively anterpolated, shifted, and combined at the centers of their child groups at the next lower level. Then, the collected incoming pattern of every observation group at the finest level is shifted and evaluated at each belonging testing function, hence providing the far contribution to the MVP for the set of M_c/n observation groups assigned to that node. As for the aggregation and interpolation stage, all the computations are efficiently performed inside each node without inter-node communications.

A. Numerical example

The NASA Almond is analyzed at a frequency of 3 THz with our parallel implementation of the MLFMA-FFT algorithm.

The electric-field integral equation (EFIE), the RWG basis functions, and the Galerkin testing procedure are used. No preconditioning is applied. An incident plane with horizontal polarization wave impinging on the back side of the Almond target is considered, as shown in the inset of Fig. 2. The analysis was done in the *Finis Terrae* supercomputer, consisting of 142 cc-NUMA HP Integrity rx7640 with 8 dual core Intel Itanium2 Montvale processors at 1.6 GHz with 18 MB L3 cache and 128 GB of memory each one. The nodes are interconnected through a high efficiency Infiniband network (4xDDR), and the operating system is Linux SLES 10. The solution of the 1,042,977,546 unknowns matrix system required the use of 64 nodes (involving a total of 1,024 processors) and 5 TB of memory. The setup time was about 105 minutes, while the iterative solution took less than 33 hours to attain a residual error of 0.023 (a total of 8 external GMRES iterations with restart 80 were required). The bistatic radar cross

section (RCS) of the Almond is shown in Fig. 2.
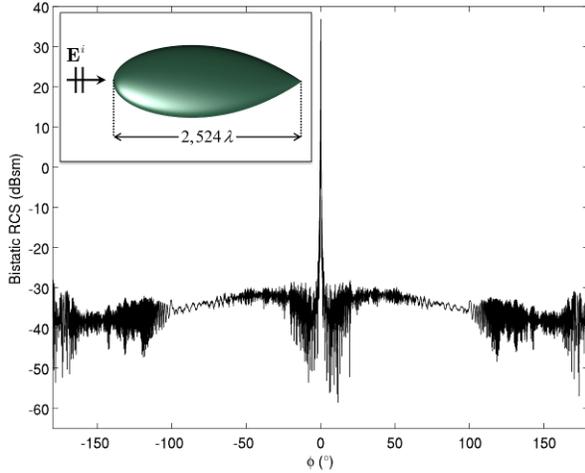
Fig. 2. Bistatic RCS of the NASA Almond target at a frequency of 3 THz.

III. SIE-MoM FOR LEFT-HANDED PLASMONIC MEDIA

When dealing with homogeneous dielectric materials, it is usual to consider the combination of normal (N-EFIE, N-MFIE) and tangential (T-EFIE, T-MFIE) equations derived from the boundary conditions imposed separately to the electric and magnetic fields. Among the multiple possibilities of combination, the following one is proven to be a stable proposal [22, 23]:

$$\sum_{l=1}^{2} a_l \frac{1}{\eta_l} \text{T-EFIE}_l + \sum_{l=1}^{2} b_l \text{N-MFIE}_l$$
$$- \sum_{l=1}^{2} c_l \text{N-EFIE}_l + \sum_{l=1}^{2} d_l \eta_l \text{T-MFIE}_l. \tag{1}$$

In equation (1), η_l is the intrinsic impedance in medium R_1 (R_1 and R_2 are the exterior and interior regions of the material, respectively). Different known formulations can be obtained depending on the selection of the complex combination parameters a_l, b_l, c_l and d_l, as shown in Table 1.

To extend the MoM formulation to the solution of piecewise homogeneous metamaterial and plasmonic penetrable objects we have opted to use the same formulation. Concretely, we concerned ourselves with the formulations known as Poggio-Miller-Chang-Harrington-Wu-Tsai (PMCHWT) [24], combined tangential

formulation (CTF) and combined normal formulation (CNF) [22], and the electric and magnetic current combined field integral equation (JMCFIE) [25]. PMCHWT and CTF combine tangential equations, CNF combines normal equations, while JMCFIE is the more general formulation combining both tangential and normal equations. For the discretization of the surface electric and magnetic currents, we opted again to use the RWG basis functions, as in the case of perfectly conducting bodies. Accurate calculation of the integrals involved in the evaluation of expressions in (1) using the RWG functions is crucial for the precision of the method. We have applied Gaussian quadrature rules for the numerical integration of smooth varying integrands, together with the analytical extraction procedures of [26-28] for the accurate evaluation of singular integrals.

Table 1: Combining parameters for different penetrable media formulations

Formulation	a_l	b_l	c_l	d_l
PMCHWT	η_l	0	0	$1/\eta_l$
CTF	1	0	0	1
CNF	0	1	1	0
JMCFIE	1	1	1	1

On the other hand, for the application of the above described formulation to the analysis of left-handed materials (LHM) and plasmonic nanoparticles, special care must be taken with the definition of the electromagnetic parameters and properties of the material. Concretely, if region defines a penetrable homogenized left LHM or a plasmonic particle, the following definitions should be applied for the wavenumber and the intrinsic impedance:

$$k_l = \omega \sqrt{\mu_l} \sqrt{\varepsilon_l}, \tag{2}$$

$$\eta_l = \frac{\sqrt{\mu_l}}{\sqrt{\varepsilon_l}}. \tag{3}$$

A. Numerical examples

We include in this section some examples to provide a verification of the validity and efficiency of the SIE-MoM approach for the solution of LHM and plasmonic nanostructures. All of them have been solved by matrix factorization. The first

one consists of the simulation of the Snell law experiment for a LHM to demonstrate the negative angle of refraction. A three-dimensional (3-D) prism was considered being $10\lambda_0$ high, $10\lambda_0$ wide, and $4\lambda_0$ deep in the largest dimension, with λ_0 the wavelength of the surrounding free-space medium. The prism is made of LHM with $\varepsilon_r = -1$ and $\mu_r = -1$ (so it is impedance matched to the surrounding space.) The angle of the second surface is 8º. An almost flat Gaussian beam, with a total angular spread of 25º and the waist at a distance of $5\lambda_0$ away is orthogonally impinging onto the first surface of the prism from the right hand side. The incident electric field is horizontally polarized.

(a)

(b)
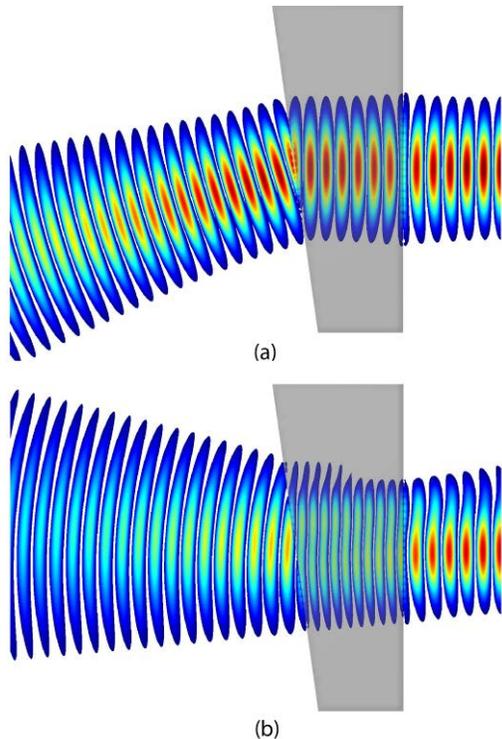
Fig. 3. (a) Incident flat Gaussian beam negatively refracted by impedance matched 3D LHM prism. (b) Conventional refraction from a prism made of Teflon.

Figure 3 (a) shows the computed electric field intensity for the experiment described above considering a refraction index of $n = -1$ for the LHM comprised prism. The problem was solved with the JMCFIE SIE-MoM formulation using 40,770 unknowns for the equivalent electric and magnetic currents induced on the surfaces of the prism. For the sake of comparison, Fig. 3 (b)

shows the same above experiment repeated for a prism made up of conventional material (we have considered Teflon, with $\varepsilon_r = 2.2$). In this case, the refraction is obtained in the conventional direction.
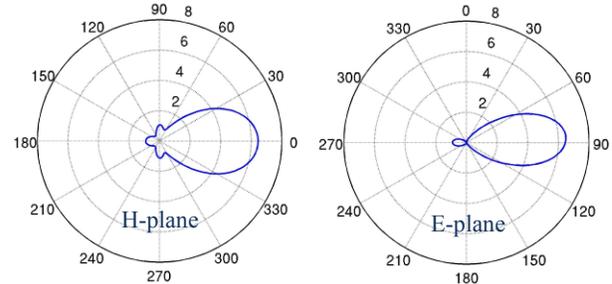
Fig. 4. Directivity of the Yagi-Uda antenna of [30] in dBi for a near-field coupled Hertzian dipole emitter: (left) H-plane ; (right) E-plane.
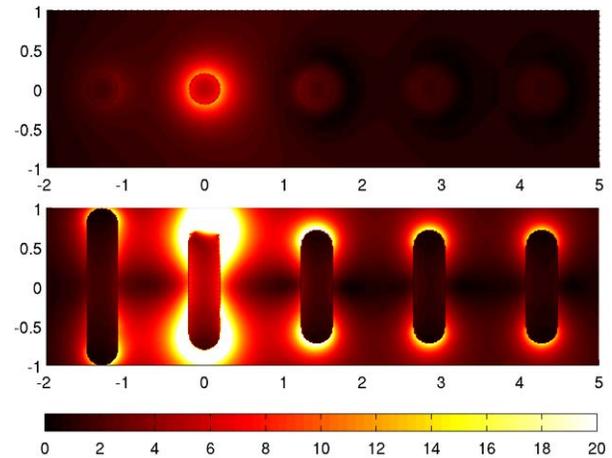
Fig. 5. Near-electric-field (V/m) distribution in the vertical and horizontal planes crossing the Yagi-Uda antenna. Dimensions are in hundreds of nanometers.

A thorough comparison of the accuracy of the usual formulations for problems involving LHMs was presented by the authors in [29].

In order to show the applicability of the method for general 3-D plasmonic nanostructures, the Yagi-Uda optical antenna designed in [30] has been analyzed with the JMCFIE SIE-MoM formulation. The antenna consists of five cylindrical elements made of aluminum, with radius 20 nm and terminated with hemispherical ends. It was optimized for an operating wavelength of $\lambda_0 = 570$ nm (in which the relative permittivity constant of aluminum is $\varepsilon_r = -38-10.9j$ and the relative permeability is $\mu_r = 1$). See [20]

and [30] for further details of the geometry. The antenna was analyzed using a total of 9,036 basis functions to represent the equivalent electric and magnetic currents on the surfaces of the nanodipoles. The computed directivity in the H- and E-planes for a near-field coupled Hertzian dipole emitter is depicted in Fig. 4. An excellent agreement with the results of [30] can be observed. Finally, Fig. 5 shows the near-electric-field distribution in the vertical and horizontal planes crossing the antenna for the Hertzian dipole excitation.

IV. CONCLUSION

In this work, we present some rigorous integral-equation solutions for different electromagnetic problems. In the case of large-scale metallic bodies, an efficient MPI/OpenMP parallel implementation of the MLFMA-FFT algorithm was shown to combine a low computational cost with a high scalability behavior, which make it an optimal choice to benefit from modern HPC computers and supercomputers. The MLFMA-FFT algorithm exploits the high scalability of FMM-FFT for the distributed computations, while the very efficient MLFMA is applied to expedite the local shared-memory computations. Since MLFMA is not distributed among nodes, but it is locally applied, it does not suffer from poor parallel scaling. In this way, we can say that the best of MLFMA and FMM-FFT algorithms is put together in MLFMA-FFT.

On the other hand, currently our work is headed to extend these rigorous integral-equation techniques to the electromagnetic simulation of problems involving homogenized left-handed metamaterials and plasmonic nanoparticles in the visible or near visible regimes. Different well-known SIE-MoM formulations, usually applied for dielectric bodies, where successfully applied to this kind of media. Numerical examples using the JMCFIE formulation were presented that confirm the validity and versatility of the integral-equation approach to the resolution of LHM and plasmonic problems in the context of leading-edge nanoscience and nanotechnology applications. Besides, it must be noticed that the application of MoM will bring the possibility of applying the latest breakthrough developments in fast integral-equation methods, such as MLFMA-FFT, for the solution of large-scale problems in metamaterials and plasmonics, which will be of great interest for the scientific community.

REFERENCES

[1] R. F. Harrington, *Field Computation by Moment Method*. New York: IEEE Press, 1993.

[2] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shape," *IEEE Trans. Antennas Propag.*, vol. 30, pp. 409–418, May 1982.

[3] R. Coifman, V. Rokhlin, and S. Wandzura, "The Fast Multipole Method for the Wave Equation: A Pedestrian Prescription," *IEEE Trans. Antennas Propagat. Mag.*, vol. 35, no. 6, pp. 7–12, June 1993.

[4] J. Song and W. C. Chew, "Multilevel Fast Multipole Algorithm for Solving Combined Eld Integral Equations of Electromagnetic Scattering," *Microw. Opt. Tech. Lett.*, vol. 10, pp. 14–19, 1995.

[5] J. Song, C. Lu, and W. C. Chew, "Multilevel Fast Multipole Algorithm for Electromagnetic Scattering by Large Complex Objects," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1488–1493, Oct. 1997.

[6] S. Velamparambil, W. C. Chew, and J. M. Song, "10 Million Unknowns: Is it That Big?," *IEEE Antennas Propagat. Mag.*, vol. 45, no. 3, pp. 43–58, Apr. 2003.

[7] S. Velamparambil and W. C. Chew, "Analysis and Performance of a Distributed Memory Multilevel Fast Multipole Algorithm," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 8, pp. 2719–2727, 2005.

[8] Ö. Ergül and L. Gürel, "Efficient Parallelization of the Multilevel Fast Multipole Algorithm for the Solution of Large-Scale Scattering Problems," *IEEE Trans. Antennas Propagat.*, vol. 56, no. 8, pp. 2335–2345, Aug. 2008.

[9] Ö. Ergül and L. Gürel, "A Hierarchical Partitioning Strategy for an Efficient Parallelization of the Multilevel Fast Multipole Algorithm," *IEEE Trans. Antennas Propagat.*, vol. 57, no. 6, pp. 1740–1750, Jun. 2009.

[10] H. Zhao, J. Hu, and Z. Nie, "Parallelization of MLFMA with Composite Load Partition Criteria and Asynchronous Communication," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 2, pp. 167 – 173, February 2010.

[11] H. Fangjing, N. Zaiping, and H. Jun, "An Efficient Parallel Multilevel Fast Multipole Algorithm for Large-scale Scattering Problems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 4, pp. 381 – 387, April 2010.

[12] M. Ujaldon, "Using GPUs for Accelerating Electromagnetic Simulations," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 4, pp. 294 – 302, April 2010.

[13] Z. N. Jiang, Z. H. Fan, D. Z. Ding, R. S. Chen, and K. W. Leung, "Preconditioned MDA-SVD-MLFMA for Analysis of Multi-Scale Problems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 25, no. 11, pp. 914 – 925, November 2010.

[14] N. Nakashima, S. Fujino, and M. Tateiba, "Performance Evaluation of State-of-the-Art Linear Iterative Solvers Based on IDR Theorem for Large Scale Electromagnetic Multiple Scattering Simulations," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 26, no. 1, pp. 37 – 44, January 2011.

[15] R. Wagner, J. Song, and W. C. Chew, "Monte Carlo Simulation of Electromagnetic Scattering from Two-Dimensional Random Rough Surfaces," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 2, pp. 235–245, Feb. 1997.

[16] C. Waltz, K. Sertel, M. A. Carr, B. C. Usner, and J. L.Volakis, "Massively Parallel Fast Multipole Method Solutions of Large Electromagnetic Scattering Problems," *IEEE Trans. Antennas Propagat.*, vol. 55, no. 6, pp. 1810–1816, 2007.

[17] J. M. Taboada, L. Landesa, F. Obelleiro, J. L. Rodriguez, J. M. Bertolo, M. G. Araujo, J. C. Mourino, and A. Gomez, "High Scalability FMM-FFT Electromagnetic Solver for Supercomputer Systems," *IEEE Antennas Propagat. Mag.*, vol. 51, no. 6, pp. 20–28, Dec. 2009.

[18] J. M. Taboada, M. Araujo, J. M. Bertolo, L. Landesa, F. Obelleiro, and J. L. Rodriguez, "MLFMA-FFT Parallel Algorithm for the Solution of Large Scale Problems in Electromagnetic (Invited Paper)," *Prog. Electromagn. Res.*, PIER, vol. 105, pp. 15–30, 2010.

[19] J. M. Taboada, L. Landesa, M. G. Araujo, J. Bertolo, F. Obelleiro, J. L. Rodriguez, J. Rivero, and G. Gajardo-Silva, "Supercomputer Solutions of Extremely Large Problems in Electromagnetics: from Ten Million to One Billion Unknowns," *IEEE Europ. Conf. Antennas Propagat. (EuCAP),* Rome, Apr. 2011.

[20] J. Rivero, J. M. Taboada, L. Landesa, F. Obelleiro, and I. García-Tuñon, "Surface Integral Equation Formulation for the Analysis of Left-Handed Metamaterials," *Opt. Express*, vol. 18, pp. 15876–15886, 2010.

[21] J. M. Taboada, J. Rivero, F. Obelleiro, M. G. Araujo, and L. Landesa, "Method of Moments Formulation for the Analysis of Plasmonic Nanooptical Antennas," *J. Opt. Soc. Am. A*, vol. 28, no. 7, July 2011.

[22] P. Ylä-Oijala, M. Taskinen, and S. Järvenpää, "Surface Integral Equation Formulations for Solving Electromagnetic Scattering Problems with Iterative Methods," *Radio Sci.*, vol. 40, no. 6, RS6002, 2005.

[23] Ö. Ergül and L. Gürel, "Comparison of Integral-Equation Formulations for the Fast and Accurate Solution of Scattering Problems Involving Dielectric Objects with the Multilevel Fast Multipole Algorithm," *IEEE Trans. Antennas Propag.,* vol. 57, no. 1, pp. 176 – 187, Jan. 2009

[24] J. R. Mautz and R. F. Harrington, "Electromagnetic Scattering from a Homogeneous Material Body of Revolution," *Arch. Elektron. Uebertraeg.*, vol. 33, pp. 71–80, 1979.

[25] P. Ylä-Oijala and M. Taskinen, "Application of Combined Field Integral Equation for Electromagnetic Scattering by Dielectric and Composite Objects," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 3, pp. 1168–1173, 2005.

[26] 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.*, vol. 32, no. 3, pp. 276–281, 1984.

[27] R. E. Hodges and Y. Rahmat-Samii, "The Evaluation of MFIE Integrals with the use of Vector Triangle Basis Functions," *Microw. Opt. Technol. Lett.*, vol. 14, no. 1, pp. 9–14, 1997.

[28] P. Ylä-Oijala and M. Taskinen, "Calculation of CFIE Impedance Matrix Elements with RWG and n×RWG Functions," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 8, pp. 1837–1846, 2003.

[29] M. G. Araújo, J. M. Taboada, J. Rivero, F. Obelleiro, "Comparison of Surface Integral Equations for Left-Handed Materials", *Prog. Electromagn. Res.*, vol. 118, pp. 425 – 440, 2011.

[30] T. H. Taminiau, F. D. Stefani, and N. F. van Hulst, "Enhanced Directional Excitation and Emission of Single Emitters by a Nano-Optical Yagi-Uda Antenna," *Opt. Express*, vol. 16, pp. 10858–10866, 2008.

Jose Manuel Taboada was born in Pontevedra, Galicia, Spain, in 1974. He received the M.S. and Ph.D. degrees in Telecommunication Engineering from the University of Vigo, Spain, in 1998 and 2001, respectively. From 1998 to 2002, he was working as a researcher at the University of Vigo. In 2002, he joined the University of Extremadura as an Associate Professor. He received the International PRACE Award and Intel Itanium Solutions Innovation Award in the year 2009 for their work in integration electromagnetics and supercomputing. His current research interests include fast integral-equation methods and supercomputing techniques in computational electromagnetics, electromagnetic compatibility, and analysis and design of metamaterial and nano-optical devices.

Marta G. Araújo was born in Vigo, Spain, in 1975. She received the M.S. and Ph.D. degrees in Telecommunication Engineering from the University of Vigo, Spain, in 1999 and 2008, respectively. From 1999 to 2000, she collaborated as a project engineer with the Department of Communications Technologies, University of Vigo, and from 2000 to 2005, she worked for a telecommunications company as a RF engineer. Currently, she is Assistant Professor at the University of Vigo and she is part of the researchers of the Antenna Group of the Signal Theory and Communications Department. Her current research interests include numerical methods in computational electromagnetics, electromagnetic compatibility, and radar systems.

F. Javier Rivero was born in Cáceres, Extremadura, Spain, in August, 1980. He received his M.S. degree in Telecommunication Engineering from the University Carlos III of Madrid, Spain, in 2008. He is currently working toward his Ph.D. with the Department of Tecnología de los Computadores y de las Comunicaciones at the University of Extremadura. His areas of interest include numerical methods in computational electromagnetics and supercomputing programming applied to metamaterials and the plasmonic behaviour of metals at optical frequencies.

Luis Landesa is full professor at the University of Extremadura. He received the M.S. and Ph.D. degree in Electrical Engineering from the University of Vigo in 1995 and 1998, respectively. He was a member of the University of Vigo and the Spanish Council of Scientific Research and, in 1999, he joined the University of Extremadura. He is the responsible of the bachelor degree in telecommunications engineering of the University of Extremadura where he teaches courses on Electromagnetics. He has more than 45 papers in refereed journals. He received the International PRACE Award and the Intel Innovation Award in the year 2009, both for their work in integration electromagnetics and supercomputing. Currently, his research is focused on fast algorithms in computational electromagnetics, supercomputing and antenna arrays.

Fernando Obelleiro Basteiro was born in Forcarei, Galicia, Spain, in 1968. He received the M.S. and Ph.D. degrees in Telecommunication Engineering from the University of Vigo, Spain, in 1991 and 1994, respectively. In 1993, he was a Visiting Scholar in the ElectroScience Laboratory, The Ohio State University. In 1991, he joined the faculty of the Telecommunication Engineering School, University of Vigo. He became an Associate Professor in 1994 and Full Professor in 2001. He is with the Spanish Navy as a Technical Consultant on radar cross-section measurements and electromagnetic compatibility, being a national representative in several NATO special working groups. He received the International PRACE Award in supercomputing in the year 2009 for their work in integration electromagnetics and supercomputing. His research interests include computational electromagnetics, antenna design, and RCS measurements.

Optimal Location of Multi-Antenna Systems - Influence of Noise-Corrupted Data

M. Zoppi [1], C. Dedeban [2], Ch. Pichot [1], S. Selleri [3], and G. Pelosi [3]

[1] Laboratoire d'Electronique, Antennes et Télécommunications (LEAT)
University of Nice Sophia-Antipolis, CNRS, Valbonne, 06560, France
mzoppi@unice.fr, christian.pichot@unice.fr

[2] Antenna R&D Unit
Orange Labs, La Turbie, 06320, France
claude.dedeban@orange-ftgroup.com

[3] Dipartimento di Elettronica e Telecomunicazioni
University of Florence, Florence, 50134, Italy
stefano.selleri@unifi.it, giuseppe.pelosi@unifi.it

Abstract ─ This paper describes the development of an antenna synthesis procedure for determining the optimal location of 2D array configurations. An inverse scattering algorithm based on a conjugate gradient method is used. The influence of noise-corrupted data on the results is investigated.

Index Terms ─ MIMO, multi-antennas, position optimization.

I. INTRODUCTION

Research on multi-antennas systems has received a growing interest in the past few years. MIMO systems have demonstrated the potential for increasing capacity but many other applications are also using multi-antenna systems such as radar applications with inverse scattering including microwave imaging. However, the problem of determining their optimal location for each application, inside a noisy environment, remains of great interest. Previous works related to this subject are reported in [1] and [3]. Moreover, in [1], we have considered the reconstruction of the optimal location of multi-antenna system and obtained preliminary results using noiseless data. Here, we are studying the performance of the synthesis procedure versus noise in order to show the robustness of the algorithm. We are

considering two kinds of noise affecting the data, an uniform white noise (UWN) and an average white Gaussian noise (AWGN). Most of the inverse scattering algorithms based on gradient optimization are using forward and adjoint problems for calculating the cost function derivative. Here, in order to save computing time, we are calculating directly the derivative of the cost function. In this way, we have direct access to the sensitivity of the cost function versus the parameters we are interested in. The inverse problem is formulated in terms of an inverse scattering problem. We are interested in retrieving the location of N antenna elements modeled by monopoles located on a planar surface, illuminated successively by a certain number of plane waves.

II. THEORY

We are considering a 2D array of small antenna elements modeled by elementary sources non-regularly distributed on the surface of the *xy* plane (Fig.1) and described with an element factor *EF* which varies such as:

$$EF(\theta) = \cos(\theta). \qquad (1)$$

The inverse problem consists in retrieving the location of the sources from the knowledge of their radiation patterns when they are illuminated with L successive incident plane waves. The

position of a single antenna element n in the rectangular coordinate system is:

$$x_n = (x_{n1}, x_{n2}, 0). \quad (2)$$

Considering antenna elements as sources defined in the xy plane presenting a simple $cosine\theta$ dependance, the coupling effect between elements is supposed to be negligeable.
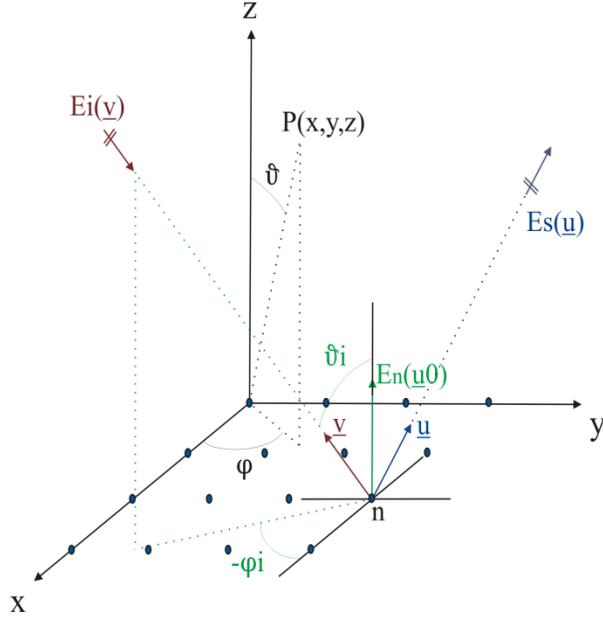
Fig. 1. Geometry of the problem.

Therefore, we can consider the total Field E as the sum of antenna elements. The scattered field for a single element n is:

$$E^S(M, x_n) = EF(\theta)m(x_n, u_0). \quad (3)$$

We can define the term m corresponding to the direction of incidence $\vec{u}_0$ of a plane wave illuminating an antenna element:

$$m(x_n, u_0) = ae^{jk\langle x_n, u_0\rangle}, \quad (4)$$

where $a \in \mathbb{R}$ is the plane wave amplitude. The dependance of scattered field with respect to the incident one is explicitly expressed as follows:

$$E^S(M, x_n) = acos(\theta)e^{jk\langle x_n, u_0\rangle}, \quad (5)$$

For N antenna elements, we have:

$$E^S(u, u_0) = acos(\theta) \sum_{n=1}^{N} e^{jk\langle x_n, (u-u_0)\rangle}, \quad (6)$$

Then we define:

$$\langle x_n, v_\vartheta \rangle = \langle x_n, (u - u_0)\rangle, \quad (7)$$

where $v_\vartheta = u - u_0$.

The derivative of E^S with respect to an antenna element position is:

$$\frac{\partial E^S}{\partial x_n} = acos(\theta)jke^{jk\langle x_n, v_\vartheta\rangle}\frac{\partial\langle x_n, v_\vartheta\rangle}{\partial x_n}. \quad (8)$$

We can now write taking into account the three Cartesian coordinates:

$$\frac{\partial\langle x_n, v_\vartheta\rangle}{\partial x_n} = \frac{\partial}{\partial x_n}\sum_{k=1}^{3}(x_n)_k(v_0)_k = \frac{\partial v_0}{\partial x_n}. \quad (9)$$

Then the final expression of the derivative becomes:

$$\frac{\partial E^S}{\partial x_n} = acos(\theta)jke^{jk\langle x_n, v_\vartheta\rangle}\frac{\partial v_0}{\partial x_n}. \quad (10)$$

E_{meas} is the given scattered field value obtained from numerical or experimental measurements, the cost function is defined such as:

$$J(x_n) = \sum_{l=1}^{L}|E^S(v_l) - E_{meas}|^2. \quad (11)$$

Its derivative with respect to the unitary element position is given by:

$$\frac{\partial J(x_n)}{\partial x_n} = 2\sum_{l=1}^{L}\Re\{E^S(v_l) \overline{- E_{meas}}\}\frac{\partial E^S}{\partial x_n}. (12)$$

Using the derivative of the cost function, an inverse algorithm is developed using a Polak-Ribière conjugate gradient method.

III. NOISE MODEL

Our previous [1] work was based on the study of a noise-free cost function defined in (11). Here, in order to study the robustness of the inversion procedure, the measured scattered field E_{meas} have been corrupted by an additive noise. As noise, we consider an ergodic stationary random process.

For simulating corrupted measured data, a signal noise has been added to synthetic data:

$$E_{meas} \approx E_{synt} + n, \quad (13)$$

where E_{synt} is the noise-free field, n the noise signal.

In order to compute the simulated measurement data, we assume to use an I-Q measurement model; it is well known that both I and Q measurements are afflicted by a noise signal. If we also assume to have additive noise we can easily write the I and Q signals expressions:

$$I = E_{Isynt} + n_I, Q = E_{Qsynt} + n_Q, \quad (14)$$

where n_I, n_Q are two random processes, with the same distribution [4]. So the simulated measurement data become:

$$E_{synt} = [E_{Isynt} + n_I] + j[E_{Qsynt} + n_Q]. \quad (15)$$

Once E_{synt} is calculated it is possible to obtain the noise n expression when fixing a desired SNR

level. In order to do this, the signal-to-noise ratio SNR equation is needed:

$$N = \frac{S}{SNR} = [|\,E_{synt}\,|^2/2]/[10^{SNR/10}], \quad (16)$$

where N is the noise power and S the signal power.

Two different kinds of noise are taken into account, i.e., an uniform white noise (UWN) and an average white Gaussian noise (AWGN). It is possible to find the noise signal contribution within n_I, n_Q (15) by simply using the variance value σ^2 of (16) and defining the appropriate expressions with respect to the chosen statistic process [4]. So for an UWN process, the expression for n_I and n_Q is:

$$n_{UNIF} = (x_{RAND} - 0.5)2\sqrt{3\sigma^2}, \quad (17)$$

while, for an AWGN process, the n expression is defined as:

$$n_{AWGN} = x_{RAND}\sqrt{\sigma^2}, \quad (18)$$

where $x_{RAND} \in [0,1]$ in (17) and (18) is a pseudo-random number generated by the numerical algorithm.

Finally, the cost function taking into account of computed noisy data is:

$$JN(x_n) = \sum_{l=1}^{L} |\,E^S(v_l) - E_{meas}\,|^2, \quad (19)$$

and for the derivative form of (19):

$$\frac{\partial J(x_n)}{\partial x_n} = 2\sum_{l=1}^{L} \Re\{E^S(v_l) - E_{meas}\}\frac{\partial E^S}{\partial x_n}. (20)$$

We use these last two expressions inside the optimization algorithm.

IV. NUMERICAL EXPERIMENTS

Different numerical experiments have been carried out in order to test the robustness of the inverse scattering algorithm with corrupted data. The working frequency is 2.45 GHz (λ_0= 12.24 cm). All the tests considered here start from a regular array 7x7 ($\lambda_0/3$ step on both x-axis and y-axis) for initial guess to retrieve an irregular array ($\lambda_0/6$ step on x-axis and $\lambda_0/9$ on y-axis). The knowledge of the radiation is related to a semi-hemisphere i.e. known over the upper semi-hemisphere ($\theta \in [0,90°]$ and $\varphi \in [0,180°]$). The tests have been done using a parametric sweep respectively over: the signal-to-noise ratio, the number of measurement points in θ and ϕ and the number of the incident plane waves. A comparison

between the two types of noise have also been done.

A. First test case: noiseless data

We consider a N = 7x7 array illuminated with L = 4 incident plane waves. The radiation pattern is known over the entire upper hemisphere. For this case, the considered noise level is zero in order to have a reference case for the corrupted-data tests. The value of the initial cost function is equal to 37.20 dB. After 2648 iteration steps, the final value of the cost function is -20.70 dB.
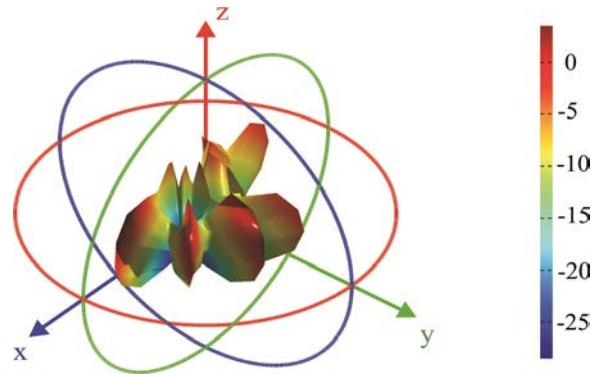
Fig. 2. Radiation pattern at initial guess (regular distribution): noiseless model.
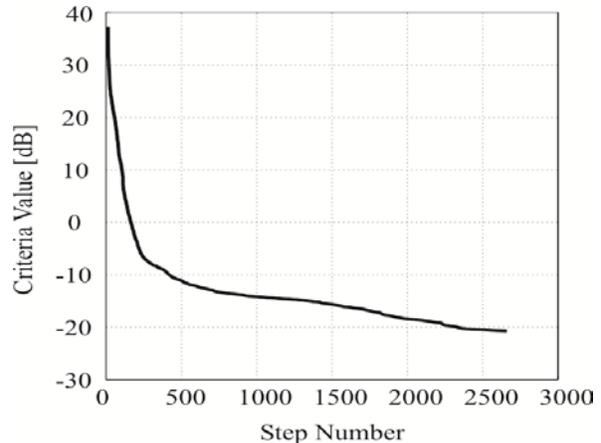
Fig. 3. Cost function (criteria values) with respect to the iteration step: noiseless model.
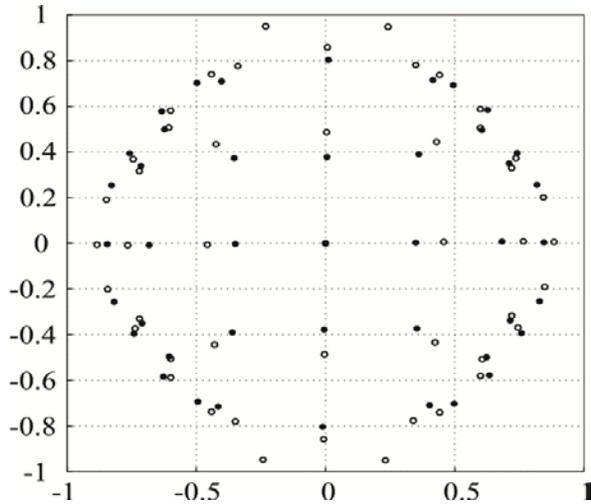
Fig. 4. Distribution of antenna elements at initial guess (white points) and final iteration (black points): noiseless model.
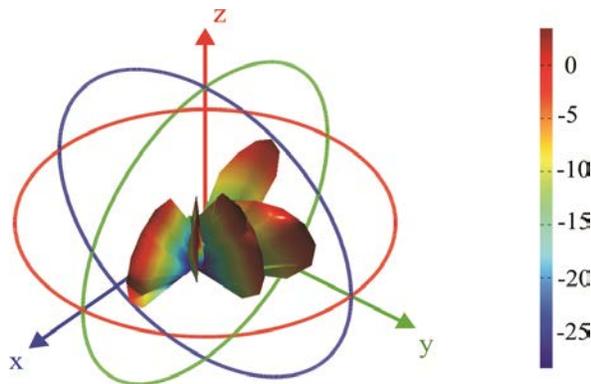
Fig. 5. Radiation pattern at final iteration (irregular distribution): noiseless model.
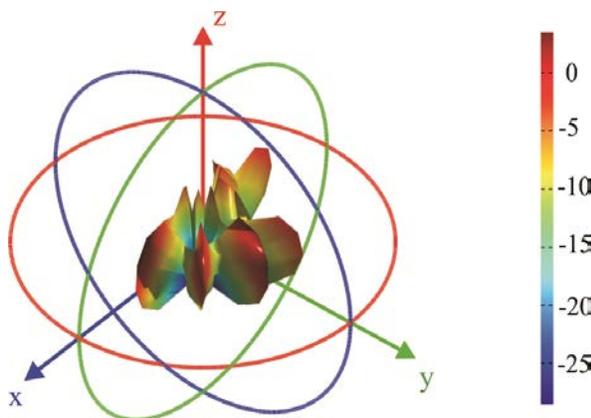
Fig. 6. Radiation pattern at initial guess (regular distribution): AWGN model with SNR 30dB.

B. Second test case: AWGN model with SNR=30dB

We consider a $N = 7 \times 7$ array illuminated with $L = 4$ incident plane waves. The radiation pattern is known over the entire upper hemisphere. For this case, the considered noise level of an AWGN model corresponds to SNR=30dB. The value of the initial cost function is equal to 37.15 dB. After 13 iteration steps, the final value of the cost function is 23.51 dB.
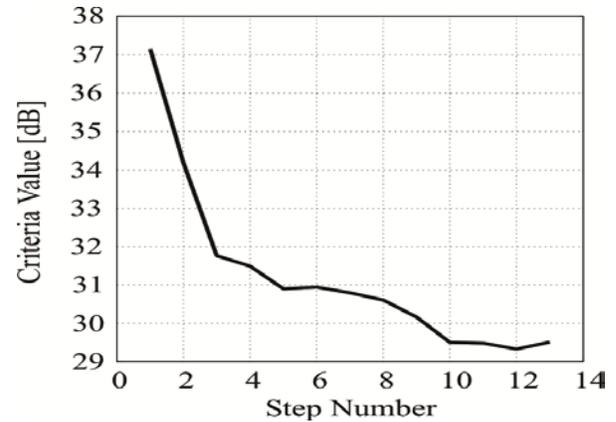
Fig. 7. Cost function with respect to the iteration step: AWGN model with SNR 30dB.
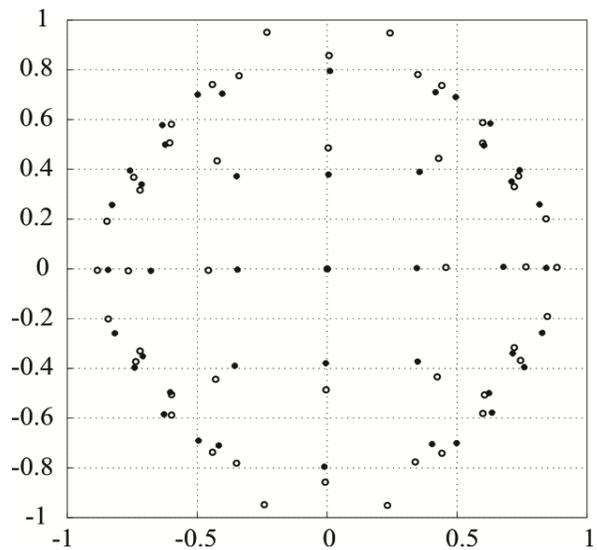
Fig. 8. Distribution of antenna elements at initial guess (white points) and final iteration (black points): AWGN model with SNR 30dB.

C. Third test case: UWN model with SNR=30dB

We consider a N = 7x7 array illuminated with L = 4 incident plane waves. The radiation pattern is known over the entire upper hemisphere. For this case, the considered noise level of an uniform noise model corresponds to SNR=30dB. The value of the initial cost function is equal to 37.32 dB. After 13 iteration steps, the final value of the cost function is 31.24 dB.
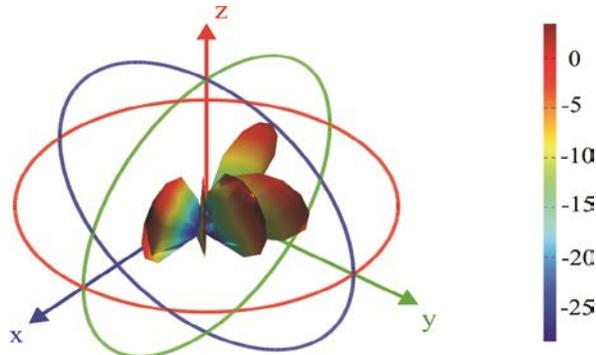
Fig. 9. Radiation pattern at final iteration (irregular distribution): AWGN model with SNR 30dB.
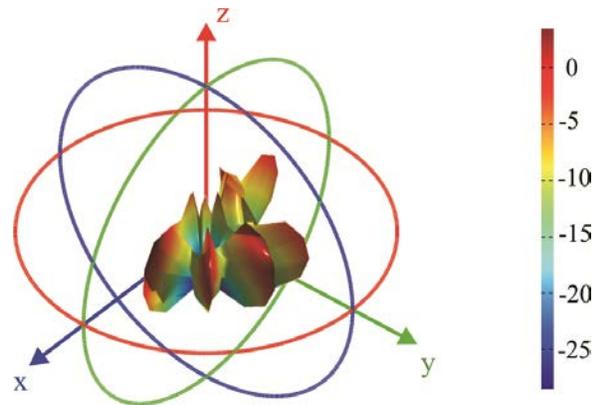
Fig. 10. Radiation pattern at initial guess (regular distribution): UWN model with SNR 30dB.

D. Fourth test case: AWGN vs. UWN

We consider a N = 7x7 array illuminated with L = 4 incident plane waves. The radiation pattern is known over the entire upper hemisphere. With this case, we compare the results obtained using an UWN model and an AWGN model, keeping in mind that the AWGN model is normally a better approximation of the real signal noise than the uniform noise.
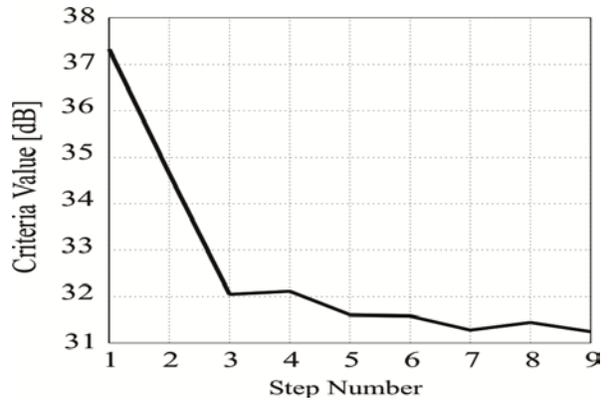
Fig. 11. Cost function with respect to the iteration step: UWN model with SNR 30dB.
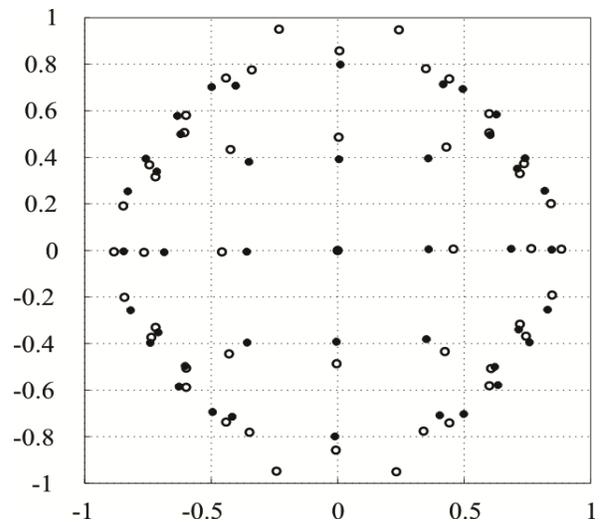
Fig. 12. Distribution of antenna elements at initial guess (white points) and final iteration (black points): UWN model with SNR 30dB.
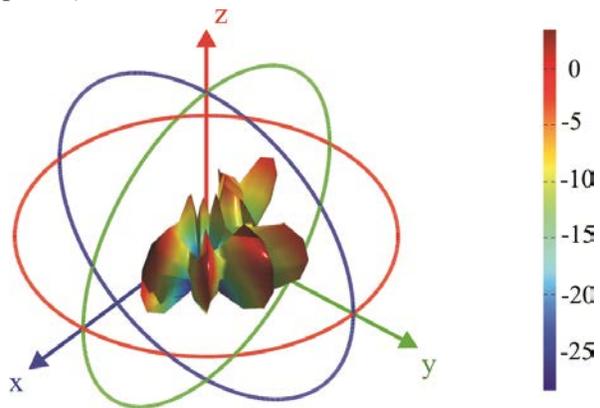
Fig. 13. Radiation pattern at final iteration (irregular distrib.): UWN model with SNR 30dB.
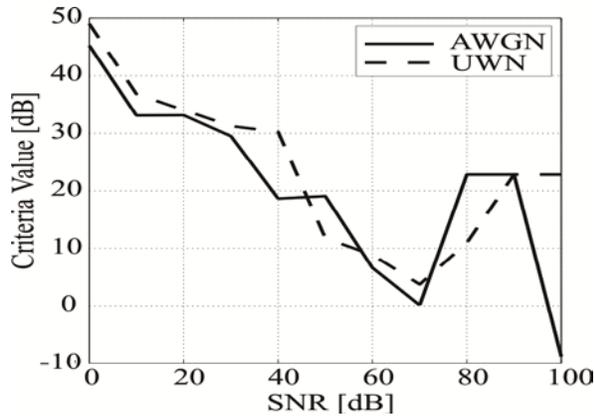
Fig. 14. Cost function (criteria values) with respect to the SNR value: AWGN vs UWT model.
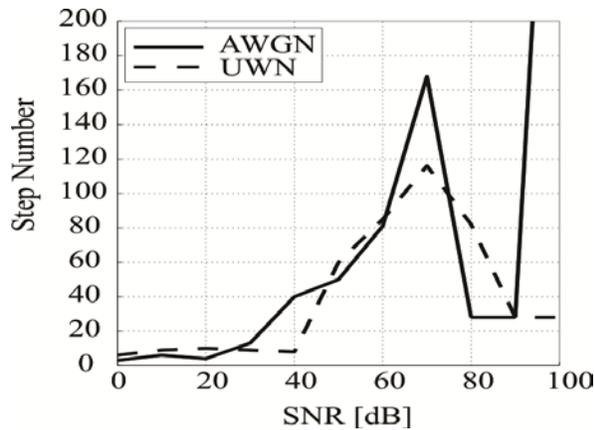
Fig. 15. Number of iterations with respect to the SNR value: AWGN vs UWT model.

E. Fifth test case: measurement points with AWGN model

We consider a N = 7x7 array illuminated with L = 4 incident plane waves. The radiation pattern is known over the entire upper hemisphere. For this case, we compare five different sets of measurement points using the AWGN model. The Figure 16 shows the convergence of the cost function (criteria values) with respect to the SNR value for different sets of measurement points. The Figure 17 shows the number of iterations with respect to the SNR value for different sets of measurement points. When examining the figures 16 and 17, the optimum set of measuring points is found to be 4x9=36. Although the continuous black curve may seem to the best one, we have to be careful before concluding. As the final radiation pattern has many local minima, we have a certain

risk that the measuring points be close or correspond to these minima.
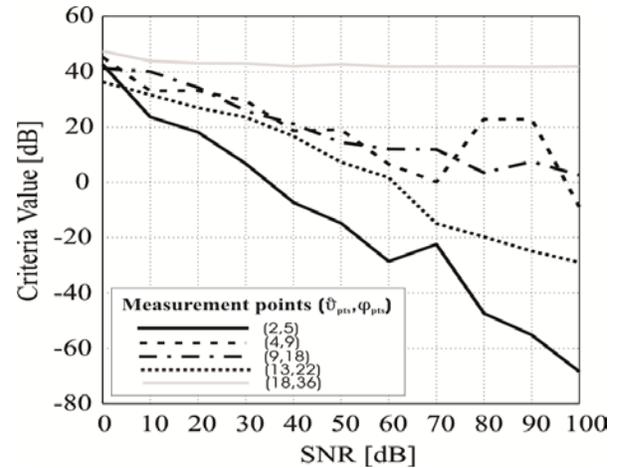
Fig. 16. Cost function (criteria values) with respect to the SNR value for different sets of measurement points: AWGN model.
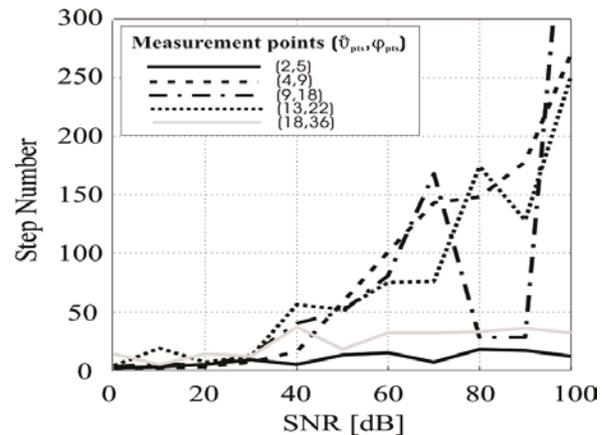
Fig. 17. Number of iterations with respect to the SNR value: AWGN model.

That may lead to a rapid convergence but to a wrong solution. In conclusion, some precaution has to be done in decreasing the number of measurement points. For the continuous gray curve, corresponding to the set of 18x36=648 measurement points, we may expect to obtain the best convergence due to the high number of measurement points but we still observe a strong non-convergence. This is simply due to an excess of information. In conclusion, we chose the dashed black curve corresponding to 4x9=36 as the optimum set of measuring points.

F. Sixth test case: number of plane waves with AWGN model

We consider a N = 7x7 element array illuminated with L = 4 incident plane waves. The radiation pattern is known over the entire upper hemisphere. For this case, we compare the results for three different numbers of incident plane waves using the AWGN mode. The Figure 18 shows the convergence of cost function (criteria values) with respect to the SNR value for different sets of measurement points and number of incident plane waves.
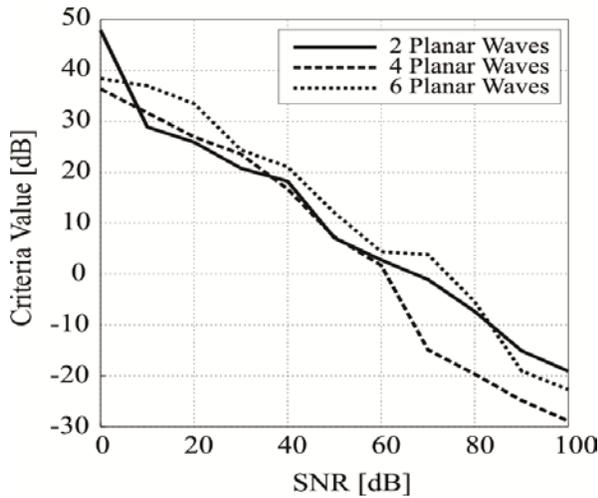
Fig. 18. Cost function (criteria values) with respect to the SNR value: AWGN model.
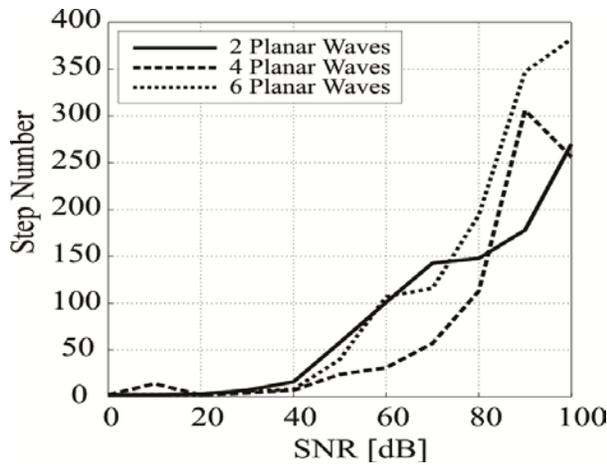
Fig. 19. Number of iteration steps with respect to the SNR value: AWGN model.

Figure 18 shows the number of iteration steps with respect to the SNR value for different number of plane waves and sets of measurement points. We can observe a similar behavior between the cases using two, four, and six incident plane waves.

Nevertheless, we prefer to choose results having a stronger convergence results despite of results obtained with less number of iteration steps. Therefore, we choose for the optimal configuration, the results obtained with four incident waves (corresponding to the dashed black line) which represents the best convergence with respect to SNR value with only few additional iterations steps.
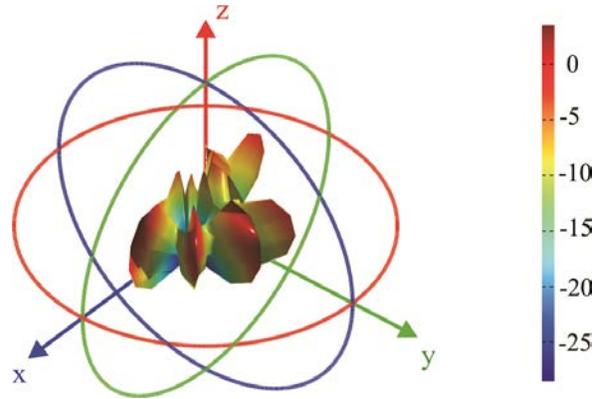
Fig. 20. Radiation pattern at initial guess (regular distribution): AWGN model with SNR 70dB.
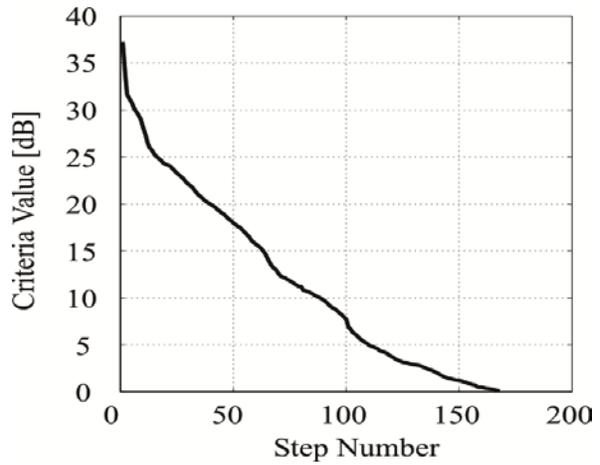
Fig. 21. Cost function (criteria values) with respect to the iteration step: AWGN model with SNR 70dB.

G. Seventh test case: optimal configuration

As we can see from Figs. 14 and 15, the best results, in terms of convergence of the cost

function, are given from the AWGN model with a SNR level of 70dB. Therefore, the optimal configuration in terms of number of measurement points and incident plane waves, when observing the results achieved at points D. F. G., corresponds to the following one: Gaussian signal noise model (AWGN); four measurement points in θ, nine measurement points in ϕ over the upper hemisphere; and four incident plane waves.
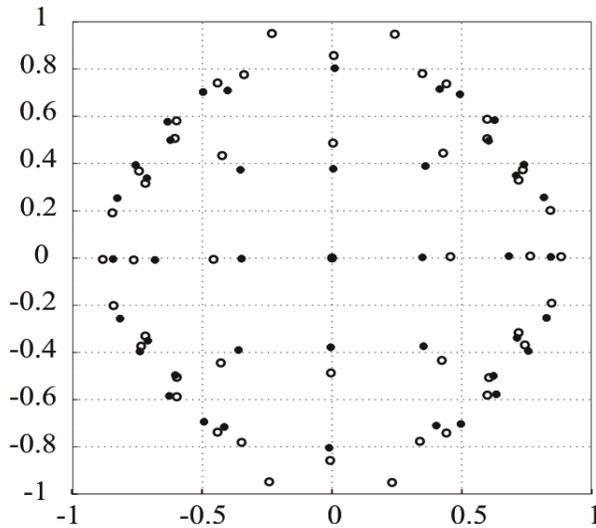
Fig. 22. Distribution of antenna elements at initial guess (white points) and final iteration (black points): AWGN model with SNR 70dB.
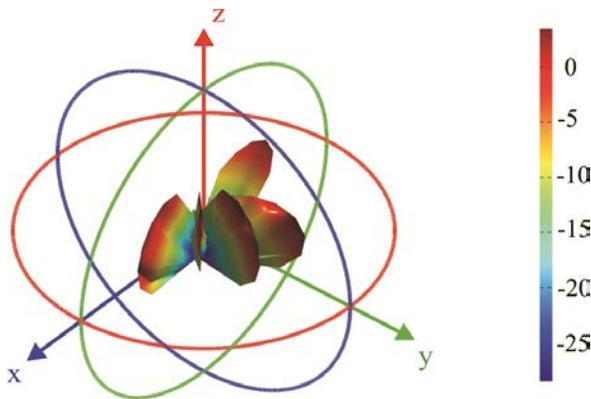
Fig. 23. Radiation pattern at final iteration (irregular distribution): AWGN model with SNR 70dB.

We consider a N = 7x7 array and the radiation pattern is known over the upper hemisphere. The value of the initial cost function is 37.30 dB. After 168 iteration steps, the final value of the cost function is 0.13 dB.

V. CONCLUSION

An optimization technique has been developed for solving an inverse scattering problem in order to retrieve the location of N antenna elements modeled by sources located on a planar surface, illuminated by plane waves. We have investigated the robustness of the algorithm with noise-corrupted data, using average Gaussian noise (AWGN) and uniform white noise (UWN) models. Different numerical results for testing the performance of the optimization technique have been presented in terms of the noise model, the number of measurement points number and number of incident planar waves. The radiation pattern is assumed to be known over the upper hemisphere.

REFERENCES

[1] M. Zoppi, C. Dedeban, Ch. Pichot, S. Selleri, and G. Pelosi, "Optimal Location of Multi-Antenna System using a Conjugate Gradient Method," *2011 Computational Electromagnetics International Workshop (CEM'11)*, Izmir, Turkey, August 10-13, 2011.

[2] M. Zoppi, C. Dedeban, Ch. Pichot, S. Selleri, and G. Pelosi, "Antenna Shape Synthesis of Various Planar Configurations using a Hybrid Conjugate Gradient Method," *2010 IEEE International Conference on Wireless Information Technology and Systems (IEEE ICWITS 2010)*, Honolulu, USA. Paper WITS1258.pdf, 28 August-3 September 2010.

[3] M. Farmahini-Farahani, R. Faraji-Dana, and M. Shahabadi, "Fast and Accurate Cascaded Particle Swarm Gradient Optimization Method for Solving 2-D Inverse Scattering Problems," *Applied Computational Electromagnetics Society (ACES) Journal*, vol. 24, no. 4, pp. 511 – 517, October 2009.

[4] R. E. Ziemer and W. H. Tranter, "Principles of Communication Systems, Modulation and Noise," *Haughton Mifflin Company*, Boston 1976.

CUDA-OpenGL Interoperability to Visualize Electromagnetic Fields Calculated by FDTD

Veysel Demir[1] and Atef Z. Elsherbeni[2]

[1]Department of Electrical Engineering
Northern Illinois University, DeKalb, IL 60115, USA
vdemir@niu.edu

[2]Department of Electrical Engineering
The University of Mississippi, University, MS 38677, USA
atef@olemiss.edu

Abstract ─ In this contribution, a compute unified device architecture (CUDA) implementation of a two-dimensional finite-difference time-domain (FDTD) program is presented along with the OpenGL interoperability to visualize electromagnetic fields as an animation while an FDTD simulation is running. CUDA, which runs on a graphics processing unit (GPU) card, is used for electromagnetic field data generation and image manipulation, while OpenGL is used to draw field distribution on the screen. Since CUDA and OpenGL both run on GPU and share data through common memory in the framebuffer, the CUDA-OpenGL interoperability is very efficient in visualization of electromagnetic fields. Step by step details of implementation of this interoperability are demonstrated.

Index Terms ─ FDTD, GPGPU, visualization.

I. INTRODUCTION

Recently, graphics processing units (GPUs) have become a viable alternative to multi-core central processing units (CPUs) for parallel processing architectures to perform high performance scientific computing. Due to the increasing demand from the scientific community, vendors have been improving both the hardware and the required software platforms, thus introducing a new generation of general purpose computing on graphic processing unit (GPGPU) cards.

Initially, the GPUs were not designed for general purpose computing and programming these cards required the use of programming platforms such as OpenGL, Brook [1], and High Level Shader Language (HLSL), which require a steep learning curve. Recently, NVIDIA introduced the Compute Unified Device Architecture (CUDA) [2] development environment as a general purpose parallel computing architecture which makes GPU computing much easier. Developers can use C language to write functions that can achieve high performance on a CUDA enabled graphics processor.

The computational electromagnetics community as well has started to utilize the computational power of these cards, and in particular, several implementations of finite-difference time-domain (FDTD) method [3,4,5] have been reported by academic researchers and commercial software vendors including the implementations based on CUDA [6-11].

CUDA and OpenGL are two software platforms both of which operate on the GPU hardware, while their intended use are different; CUDA is suitable for improving the performance of data parallel computations, while OpenGL is for producing 2D and 3D computer graphics. While running a FDTD simulation, it is possible to capture electromagnetic fields and visualize them as an on-the-fly animation. If the FDTD calculations are performed on a graphics card, which is used also to perform OpenGL operations

to display the fields, one can copy the field data from the graphics card memory (device memory) to the computer's main memory (host memory) that is processed by the CPU, process the data to create an image, and copy the image back to the GPU memory to display via OpenGL. It is possible to avoid the back and forth data transfer between the host and device memories, and perform all the processing required for the display on the graphics card by employing CUDA-OpenGL interoperability provided by CUDA. Performing field calculations and processing the fields to create images for visualization simultaneously can considerably slow down the FDTD simulations and hinder the efficiency. The goal is to balance the tradeoff between a fast simulation and a high quality and smooth visualization. In this context, as presented in this contribution, CUDA-OpenGL interoperability improves the simultaneous calculation and visualization efficiency significantly. An implementation of CUDA-OpenGL interoperability is presented in the subsequent sections.

II. CUDA-OPENGL INTEROPERABILITY IN FDTD

A. Integration of FDTD with GLUT

The FDTD method is an iterative method in which the progressions of electromagnetic fields in time are simulated in a time marching loop. The time marching loop typically consists of functions to update sources, update electric and magnetic fields, apply boundary conditions, and capture fields. Many times the captured fields can be displayed on the fly as an animation of the fields. Such a FDTD algorithm is illustrated in Fig. 1. It is usually straightforward to program this algorithm in a programming language where a programmer can simply call a built-in function to display the fields. An example is Matlab [12], in which several plotting functions, such as `plot` and `imagesc`, are provided to the programmer to display data while the program is running. However in many other languages, such as C++ and Fortran, such functions are not available and one has to program the details of visualization code as well. OpenGL has been one of the most popular platforms to facilitate programming with visualization.
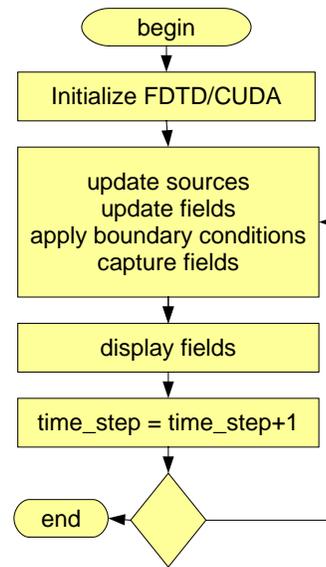
Fig. 1. FDTD algorithm.
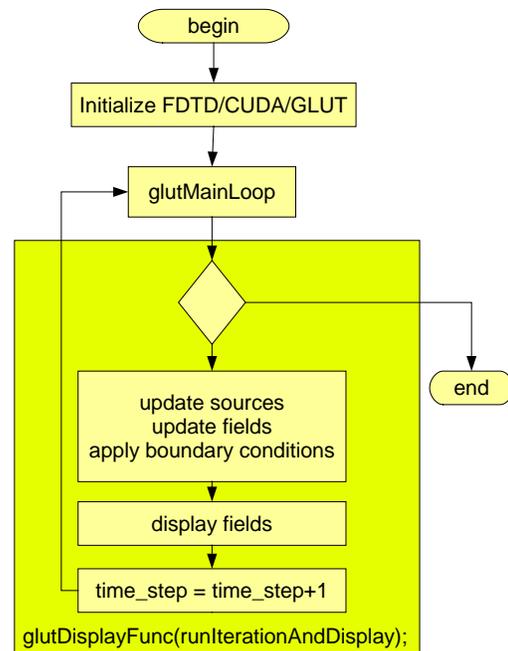
Fig. 2. FDTD algorithm integrated with GLUT.

Both CUDA and OpenGL can be programmed in C language. There are methods to integrate OpenGL in a program developed for an event driven operating system such as Microsoft Windows or Linux. One of the methods is to use GLUT. GLUT is the OpenGL Utility Toolkit, a window system independent toolkit for writing OpenGL programs. It implements a simple

windowing application programming interface (API) for OpenGL [13]. In this contribution, CUDA/OpenGL interoperability is presented through the use of GLUT.

GLUT makes OpenGL programming simple yet platform independent, however, GLUT implements its own event loop. Therefore, mixing GLUT with an algorithm that demands its own event handling structure may be difficult. In order to integrate the FDTD algorithm with GLUT, the algorithm in Fig. 1 is modified as the one in Fig. 2. In this algorithm, first of all, FDTD, CUDA, and GLUT are initialized. CUDA-OpenGL interoperability also requires additional initialization at this stage as will be discussed. Then GLUT loop is started. Whenever GLUT loop triggers a display event, first a single iteration (or a number of iterations) of FDTD time marching loop is performed, and then results are displayed on a window.

B. Initialization of OpenGL with CUDA

CUDA is GPU programming platform developed and introduced by Nvidia. Nvidia provides extensive support to CUDA programmers. An article titled as "What Every CUDA Programmer Should Know about OpenGL" [14] is a good reference for beginners who want to learn CUDA-OpenGL interoperability. In this contribution, guidelines in [14] are followed to achieve interoperability between OpenGL and CUDA in an FDTD code. The details are presented in the steps below.

Listing 1. Initialization of FDTD, CUDA, and OpenGL
```
// global parameters
GLuint pbo_destination;
struct cudaGraphicsResource
*cuda_pbo_destination_resource;
GLuint cuda_result_texture;

bool runFdtdWithFieldDisplay
      (int argc, char** argv)
{
// Initialize CUDA context
cudaGLSetGLDevice
(cutGetMaxGflopsDeviceId());

// Initialize GL context
initializeGL(argc, argv);
```

```
// Initialize GL buffers
initializeGLBuffers();

// colormap used to map field
intensity
createColormapOnGpu();

// Display list of objects in problem
space
createDisplayListForObjects();

// copy data from CPU RAM to GPU
global memory
copyFdtdArraysToGpuMemory();

glutMainLoop();    // GLUT loop
}
```

Listing 2. Creating GL context
```
void initializeGL
      (int argc, char **argv )
{
setImageAndWindowSize();

// Create GL context
glutInit(&argc, argv);
glutInitDisplayMode(GLUT_RGBA |
      GLUT_ALPHA | GLUT_DOUBLE |
      GLUT_DEPTH);
glutInitWindowSize(window_width,
      window_height);
iGLUTWindowHandle =
      glutCreateWindow
            ("CUDA OpenGL FDTD");

// initialize necessary OpenGL
extensions
glewInit();

// Initialize GLUT event functions
glutDisplayFunc
      (runIterationAndDisplay);
glutKeyboardFunc(keyboard);
glutReshapeFunc(reshape);
glutIdleFunc(idle);
}
```

Initialize CUDA

Listing 1 shows a function in which several functions are called to initialize FDTD, CUDA, and OpenGL and then FDTD simulations are started through GLUT. The first step is to initialize CUDA: the GPU device with maximum Gflops is set as the active device to run the FDTD calculations by a call to the function `cudaGLSetGLDevice()`.

Initialize GL and Create a Window

The next step is initialization of OpenGL and GLUT and creation of a window to display captured electromagnetic fields. GL initialization is performed in the function `initializeGL()` shown in Listing 2. The first part of Listing 2 initializes the GLUT. Next is an important step for CUDA-OpenGL interoperability in which OpenGL extensions are loaded to support buffers by calling the function `glewInit()` in Listing 2. Then event functions for GLUT are defined, thus GL initialization is completed. Here one should notice that `glutDisplayFunc()` is defined as `runIterationAndDisplay()`. As will be discussed later, when GLUT triggers a display event, the `runIterationAndDisplay()` will be executed, which will perform an iteration of FDTD time marching loop and display the electromagnetic field distribution in the problem space.

Listing 3. Initializing GL buffers
```
void initializeGLBuffers()
{
// create pixel buffer object
createPixelBufferObject
(&pbo_destination,
&cuda_pbo_destination_resource);

// create texture that will receive
the result of CUDA
createTextureDestination
      (&cuda_result_texture,
      image_width, image_height);
}
```

Listing 4. Creating pixel buffer object
```
void createPixelBufferObject(GLuint*
pbo, struct cudaGraphicsResource
**pbo_resource)
{
unsigned int texture_size =
sizeof(GLubyte) * image_width *
image_height * 4;

void *data = malloc(texture_size);

// create buffer object
glGenBuffers(1, pbo);
glBindBuffer(GL_ARRAY_BUFFER, *pbo);
glBufferData(GL_ARRAY_BUFFER,
texture_size, data, GL_DYNAMIC_DRAW);
free(data);
```

```
glBindBuffer(GL_ARRAY_BUFFER, 0);

// register this buffer object with
CUDA
cudaGraphicsGLRegisterBuffer
(pbo_resource, *pbo,
cudaGraphicsMapFlagsNone);
}
```

Listing 5. Create texture
```
void createTextureDestination
(GLuint* cuda_result_texture,
unsigned int size_x,
unsigned int size_y)
{
// create a texture
glGenTextures(1,cuda_result_texture);
glBindTexture(GL_TEXTURE_2D,
*cuda_result_texture);

// set basic parameters
glTexParameteri(GL_TEXTURE_2D,
GL_TEXTURE_WRAP_S, GL_CLAMP_TO_EDGE);
glTexParameteri(GL_TEXTURE_2D,
GL_TEXTURE_WRAP_T, GL_CLAMP_TO_EDGE);
glTexParameteri(GL_TEXTURE_2D,
GL_TEXTURE_MIN_FILTER, GL_NEAREST);
glTexParameteri(GL_TEXTURE_2D,
GL_TEXTURE_MAG_FILTER, GL_NEAREST);

glTexImage2D(GL_TEXTURE_2D, 0,
GL_RGBA8, size_x, size_y, 0, GL_RGBA,
GL_UNSIGNED_BYTE, NULL);
}
```

Create an OpenGL Buffer

CUDA and OpenGL will use common resources on GPU for interoperability. Basically, these resources are buffers on the GPU's memory space. These buffers shall be created and initialized. In Listing 1, `initializeGLBuffers()` function, shown in Listing 3, is called for buffer initialization. Implementation of `initializeGLBuffers()` is shown in Listing 3. First, `createPixelBufferObject()` function, shown in Listing 4, is called, which creates a pixel buffer object and allocates memory for this buffer. The buffer will hold image data. The image is a field distribution in a two dimensional problem space composed of $n_{xx} \times n_{yy}$ cells. The field in each cell will be displayed with a single pixel, thus the image size is `image_width` * `image_height`, where `image_width` =

nxx; image_height = nyy;. Each pixel will hold red, green, blue, and alpha (RGBA) value of the pixel, thus each pixel uses four bytes of memory. Thus, the allocated memory is 4 * image_width * image_height. CUDA will create the image and write to this buffer through the pixel buffer object, and then OpenGL will access to the same memory space and process it as a texture and display the image.

The texture as well needs to be initialized. The createTextureDestination(), shown in Listing 5, is used to initialize the texture.

Register Buffers for CUDA

The last step in initialization of the pixel buffer object is to register the created buffer for CUDA. This is done in the last line of Listing 4 by calling the cudaGraphicsGLRegisterBuffer(). This command simply informs the OpenGL and CUDA drivers that this buffer will be used by both.

Listing 6. Display function of GLUT
```
void runIterationAndDisplay()
{
// run an FDTD iteration on GPU using
CUDA
for (int i=0; i< plotting_step; i++)
if (time_step<number_of_time_steps)
    fdtdIterationOnGpu();
    else
    {
    fetchResultsFromGpuMemory();
    deallocateArrays();
    saveSampledFieldsToFile();
    Cleanup(EXIT_SUCCESS);
    }

// Create image of field using CUDA
unsigned int* image_data;
// map the GL buffer to CUDA
cudaGraphicsMapResources(1,
&cuda_pbo_destination_resource, 0);
cudaGraphicsResourceGetMappedPointer
((void **)&image_data,
&number_of_bytes,
cuda_pbo_destination_resource);

// execute CUDA kernel
createImageOnGpu(image_data);
// unmap the GL buffer
cudaGraphicsUnmapResources(1,
&cuda_pbo_destination_resource, 0);
```

```
// Create a texture from the buffer
glBindBuffer(GL_PIXEL_UNPACK_BUFFER_A
RB, pbo_destination);
glBindTexture(GL_TEXTURE_2D,
cuda_result_texture);
glTexSubImage2D(GL_TEXTURE_2D, 0, 0,
0, image_width, image_height,
GL_RGBA, GL_UNSIGNED_BYTE, NULL);
glBindBuffer(GL_PIXEL_PACK_BUFFER_ARB
, 0);
glBindBuffer(GL_PIXEL_UNPACK_BUFFER_A
RB, 0);

// draw the image
displayTextureImage
     (cuda_result_texture);

cudaThreadSynchronize();

// swap the front and back buffers
glutSwapBuffers();
}
```

Listing 7. Updating electric fields
```
__global__ void
update_electric_fields_on_kernel_TMz
(float* Ceze, float* Cezhy, float*
Cezhx, float* Hx,  float* Hy,
float* Ez, int nxx)
{
__shared__ float
sHy[TILE_SIZE][2*TILE_SIZE+1];

int tx = threadIdx.x;
int ty = threadIdx.y;
int i = blockIdx.x * blockDim.x + tx;
int j = blockIdx.y * blockDim.y + ty;

int ci = (j+1)*nxx+i;

sHy[ty][tx+TILE_SIZE] = Hy[ci];
sHy[ty][tx] = Hy[ci-TILE_SIZE];

__syncthreads();
Ez[ci] = Ceze[ci] * Ez[ci] +
Cezhy[ci] * (sHy[ty][tx+TILE_SIZE]-
sHy[ty][tx+TILE_SIZE-1]) + Cezhx[ci]
* (Hx[ci]-Hx[ci-nxx]);
}
```

Listing 8. Launch kernel to create the image
```
extern "C" void
createImageOnGpu
(unsigned int* image_data)
{
dim3 block(TILE_SIZE, TILE_SIZE, 1);
```

```
dim3 grid(nxx/block.x,
      nyy/block.y, 1);
createImageOnKernel
<<< grid, block>>>(image_data, dvEz,
nxx, min_value, max_value);
}
```

Listing 9. The kernel to create the image

```
__global__ void
createImageOnKernel(unsigned int*
image_data, float* Ez, int nxx,
float minval, float maxval)
{
int i = blockIdx.x * blockDim.x +
threadIdx.x;
int j = blockIdx.y * blockDim.y +
threadIdx.y;
int color_ind; float F;
int ci = j*nxx+i;
int ti = (j+1)*nxx+i;

F = Ez[ti] - minval;
color_ind = floor(255 * F/(maxval-
minval));
image_data[ci] = dvrgb[cind];
}
```

Listing 10. Displaying the image using OpenGL

```
void displayTextureImage
(GLuint texture)
{
glBindTexture(GL_TEXTURE_2D,
texture);
glEnable(GL_TEXTURE_2D);
glTexEnvf(GL_TEXTURE_ENV,
GL_TEXTURE_ENV_MODE, GL_REPLACE);

glMatrixMode(GL_PROJECTION);
glLoadIdentity();
glOrtho(domain_min_x,
domain_min_x+nxx*dx, domain_min_y,
domain_min_y+nyy*dy, -1.0, 1.0);

glMatrixMode( GL_MODELVIEW);
glViewport(0, 0, window_width,
window_height);

glBegin(GL_QUADS);
      glTexCoord2f(0.0, 0.0);
      glVertex3f(domain_min_x,
      domain_min_y, 0.0);
      glTexCoord2f(1.0, 0.0);
      glVertex3f(domain_min_x+nxx*dx,
      domain_min_y, 0.0);
```

```
      glTexCoord2f(1.0, 1.0);
      glVertex3f(domain_min_x+nxx*dx,
      domain_min_y+nyy*dy, 0.0);
      glTexCoord2f(0.0, 1.0);
      glVertex3f(domain_min_x,
      domain_min_y+nyy*dy, 0.0);
glEnd();

glDisable(GL_TEXTURE_2D);

glCallList(objects_display_list);
}
```

Copy Colormap to GPU Memory

After OpenGL and CUDA initializations are completed in Listing 1, a colormap is constructed on the GPU constant memory in `createColormapOnGpu()`. The colormap is basically a one dimensional array of 4 bytes integer including RGB values of colors that begin with blue, and pass through cyan, yellow, orange, and red. The colormap will be used to reflect the intensity of the displayed fields. The colormap array will be accessed at every iteration of the FDTD time marching loop in a CUDA kernel to generate the image of field distribution; thus, its access by CUDA has to be fast. To achieve fast access, the colormap array is stored in the constant memory of GPU.

Create a Display List of Objects

It is possible to display an outline of objects that exist in the problem space as polygons together with the field distribution. An OpenGL display list is created in the function `createDisplayListForObjects()`. This display list is drawn on the field distribution image at every iteration, so it is more efficient to create it once at the beginning and use it during the time marching loop.

Copy FDTD Arrays to GPU Memory

A CUDA program is a hybrid code which mainly runs on CPU, while parallel processing sections run on GPU. Therefore, the FDTD problem space, i.e. coefficient arrays, and field arrays, are initially constructed and allocated on the CPU RAM. These arrays need to be copied to GPU global memory to have them available for CUDA computations on GPU. These arrays are copied to GPU global memory in the

`copyFdtdArraysToGpuMemory()` function in Listing 1.

C. Field calculations using CUDA and visualization of fields

As discussed before, whenever the display event (`glutDisplayFunc()`) is triggered in the GLUT loop, the associated function `runIterationAndDisplay()` is executed. Implementation of this function is shown in Listing 6. In Listing 6, a check is performed to see if the FDTD iterations are complete. If the iterations are complete, results are copied from the GPU global memory to the CPU memory, other post-processing operations are performed, and FDTD simulation is ended. If iterations are not complete, a number of iterations of the FDTD time marching loop are performed on GPU using CUDA by a call to the `fdtdIterationOnGpu()` function. An iteration includes usual steps such as an update of sources, an update of electric fields and magnetic fields, application of boundary conditions, and capture of electromagnetic fields. For instance, Listing 7 shows the kernel function that updates the electric field for the TM_z case.
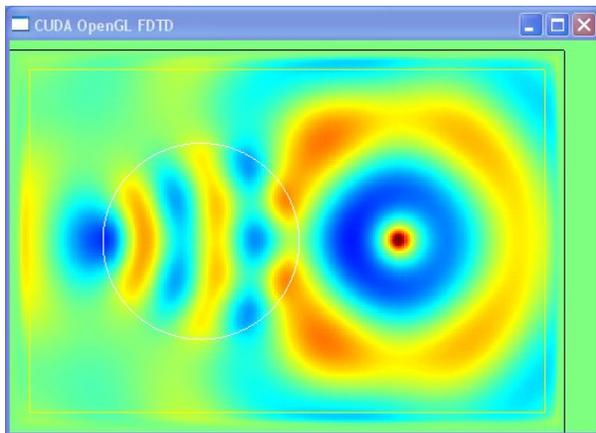
Fig. 3. A snapshot from the 2D FDTD program display.

After the field calculations, the new field distribution can be displayed on the created window, however, the field data cannot be displayed directly; data needs to be converted to an image first and the image needs to be stored in a texture in a form to be used by OpenGL. First, the GL buffer is mapped to CUDA, as shown in Listing 6, such that CUDA can process the field data and create image data. A GPU kernel function needs to be called to process the field data, thus the `createImageOnGpu()` function is called to launch the kernel function. Implementation of `createImageOnGpu()` is shown in Listing 8. The kernel function is `createImageOnKernel()` and is shown in Listing 9. This kernel basically maps the field value at each cell from a range between a minimum and a maximum to one of the 256 colors in the colormap, and stores to the image buffer.

Once the image is created in the buffer, the buffer is unmapped and released from CUDA. Then a texture is created from this buffer as shown in Listing 6. Once the texture is created, it is ready to get displayed by OpenGL. The function `displayTextureImage()`, shown in Listing 10, is called to perform the final display operations. Then `glutSwapBuffers()` is executed to show the image on the screen.

Figure 3 shows a snapshot from an animation of a two-dimensional FDTD simulation. The image is generated through the CUDA-OpenGL interoperability.

III. PERFORMANCE OF CUDA-OPENGL INTEROPERABILITY

Two parametric sweep tests are performed by running the presented FDTD code in different modes to assess the performance improvement provided by CUDA-OpenGL interoperability. The following four modes are considered:
1. program is run on CPU only without field visualization
2. program is run on GPU using CUDA without field visualization
3. program is run on GPU using CUDA with field visualization and with CUDA-OpenGL interoperability
4. program is run on GPU using CUDA with field visualization, but without CUDA-OpenGL interoperability (the image data is transferred from device memory to host memory and displayed using OpenGL)

The analyses are performed on an NVIDIA® Tesla™ C1060 Computing Processor installed on a 64 bit Windows XP computer. This card has 240 streaming processor cores operating at 1.3 GHz.

The CPU results obtained using an Intel Xeon processor at 2 GHz.

Here, it should be noted that it is not necessary to display the images at every time step of the FDTD loop. Often it is sufficient to display a frame after a few time steps. A parameter denoted as `plotting_step`, shown in Listing 6, is used to control how often the images are displayed.

As the first parameter sweep test, the problem size of the two-dimensional space is increased, and each time the simulations are performed for the four modes with 1 frame per 5 iterations rate. Then, the throughputs of the simulations are calculated as [15]

$$NMCPS = \frac{n_{steps} \times n_{xx} \times n_{yy}}{t_s} \times 10^{-6} , \qquad (1)$$

where *NMCPS* is the number of million cells processed per second, n_{steps} is the total number of time steps the program has been run, and t_s is the total computation time in seconds. Here, n_{xx} and n_{yy} are the number of cells in an FDTD problem space in *x* and *y* directions, respectively. The results are shown in Fig. 4. Throughput is a measure of how fast the computations are performed, thus it can be used to assess the efficiency of the codes. The computation on GPU without any field visualization is much faster, as expected. Computation on CPU, even without visualization, is not comparable with GPU computation. In the presented Fig. 4, the computation with visualization is faster by 30% with CUDA-OpenGL interoperability. The results verify the efficiency improvement achieved by interoperability.

In the second parameter sweep test, the frame display rate is reduced (the image frames are displayed less often) and each time the simulations are performed for the four modes for a problem size of 6 million cells. The throughput results are shown in Fig. 5. Results again verify that visualizations are considerably faster with CUDA-OpenGL interoperability. The data transfers from the device memory to the host memory take considerable time when CUDA-OpenGL interoperability is not utilized. Moreover, as image display rate is reduced, the computations on GPU with visualizations converge to that of GPU without visualization.
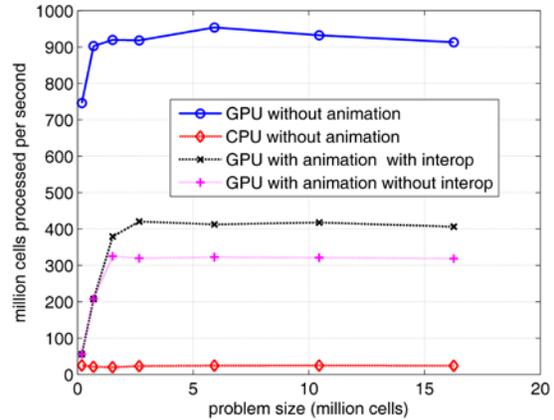
Fig. 4. Throughputs of different modes vs. problem size. 1 frame displayed per 5 time steps.
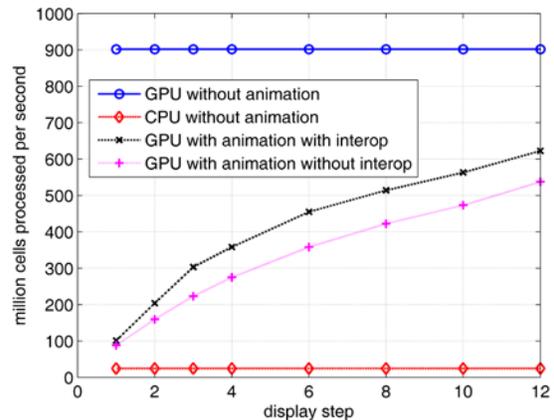
Fig. 5. Throughputs of different modes vs. image display rate. Problem size is 6 million cells.

The results in Fig. 5 also show that displaying the frames more often slows down the computations significantly. It is not necessary to display the frames very often. It has been observed that it is sufficient to display one frame per 5-10 iterations to achieve a smooth visual animation simultaneously with fast computation.

IV. CONCLUSION

An implementation of CUDA-OpenGL interoperability to visualize electromagnetic fields in a two-dimensional FDTD simulation is developed and presented. It is shown that interoperability can improve the visualization efficiency significantly. Interoperability can be extended to three-dimensional FDTD and more complicated field visualizations can be achieved.

REFERENCES

[1] I. Buck, *Brook Spec v0.2*, Stanford Univ. Press, 2003.

[2] NVIDIA CUDA ZONE: www.nvidia.com/object/cuda_home.html.

[3] K. S. Yee, "Numerical Solution of Inital Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Transactions on Antennas and Propagation*, vol. 14, pp. 302–307, May 1966.

[4] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd edition, Artech House, 2005.

[5] A. Elsherbeni and V. Demir, *The Finite Difference Time Domain Method for Electromagnetics: with MATLAB Simulations*, *SciTech Publishing*, 2009.

[6] P. Sypek, A. Dziekonski, and M. Mrozowski, "How to Render FDTD Computations More Effective Using a Graphics Accelerator," *IEEE Transactions on Magnetics*, vol. 45, no. 3, pp. 1324-1327, 2009.

[7] V. Demir and A. Z. Elsherbeni, "Compute Unified Device Architecture (CUDA) Based Finite-Difference Time-Domain (FDTD) implementation," *Journal of the Applied Computational Electromagnetics Society (ACES)*, vol. 25, no. 4, pp. 303-314, April 2010.

[8] C. Y. Ong, M. Weldon, S. Quiring, L. Maxwell, M. C. Hughes, C. Whelan, and M. Okoniewski, "Speed it Up," *IEEE Microwave Magazine*, vol. 11, no. 2, pp. 70-78, April 2010.

[9] M. Ujaldon, "Using GPUs for Accelerating Electromagnetic Simulations," *Journal of the Applied Computational Electromagnetics Society (ACES)*, vol. 25, no. 4, pp. 294-302, April 2010.

[10] N. Takada, T. Shimobaba, N. Masuda, and T. Ito, "Improved Performance of FDTD Computation Using a Thread Block Constructed as a Two-Dimensional Array with CUDA," *Journal of the Applied Computational Electromagnetics Society (ACES)*, vol. 25, no. 12, pp. 1061-1069, December 2010.

[11] M. R. Zunoubi and J. Payne, "Analysis of 3-Dimensional Electromagnetic Fields in Dispersive Media using CUDA," *Progress In Electromagnetics Research M*, vol. 16, pp. 185-196, 2011.

[12] http://www.mathworks.com.

[13] M. J. Kilgard, "The OpenGL Utility Toolkit (GLUT) Programming Interface, API Version 3". Silicon Graphics, Inc. November 13, 1996.

[14] J. Stam, "What Every CUDA Programmer Should Know About OpenGL," in GPU Technology Conference, San Jose, CA, October 1, 2009.

[15] Acceleware: www.acceleware.com

Veysel Demir is an Assistant Professor at The Department of Electrical Engineering, Northern Illinois University. He received his B.Sc. degree in Electrical Engineering from Middle East Technical University, Ankara, Turkey, in 1997. He studied at Syracuse University, New York, where he received both a M.Sc. and Ph.D. in Electrical Engineering in 2002 and 2004, respectively. During his graduate studies, he worked as research assistant for Sonnet Software, Inc., Liverpool, New York. He worked as a visiting research scholar in the Department of Electrical Engineering at the University of Mississippi from 2004 to 2007. He joined Northern Illinois University in August 2007. His research interests include numerical analysis techniques as well as microwave and radiofrequency (RF) circuit analysis and design. Dr. Demir is a member of IEEE and ACES and has coauthored more than 20 technical journal and conference papers. He is the coauthor of the books *Electromagnetic Scattering Using the Iterative Multiregion Technique* (Morgan & Claypool, 2007) and *The Finite Difference Time Domain Method for Electromagnetics with MATLAB Simulations* (SciTech 2009).

Atef Z. Elsherbeni is a Professor of Electrical Engineering and Associate Dean for Research and Graduate Programs, the Director of The School of Engineering CAD Lab, and the Associate Director of The Center for Applied Electromagnetic Systems Research (CAESR) at The University of Mississippi. In 2004, he was appointed as an adjunct Professor, at The Department of Electrical Engineering and Computer Science of the L.C. Smith College of Engineering and Computer Science at Syracuse University. On 2009, he was selected as Finland Distinguished Professor by the Academy of Finland and TEKES. Dr. Elsherbeni is a Fellow member of the Institute of Electrical and Electronics Engineers (IEEE) and a Fellow member of The Applied Computational Electromagnetics Society (ACES). He is the Editor-in-Chief for ACES Journal and an Associate Editor to the Radio Science Journal.

2012 INSTITUTIONAL MEMBERS

DTIC-OCP LIBRARY
8725 John J. Kingman Rd, Ste 0944
Fort Belvoir, VA 22060-6218

AUSTRALIAN DEFENCE LIBRARY
Northcott Drive
Canberra, A.C.T. 2600 Australia

BEIJING BOOK CO, INC
701 E Linden Avenue
Linden, NJ 07036-2495

DARTMOUTH COLLEGE
6025 Baker/Berry Library
Hanover, NH 03755-3560

DSTO EDINBURGH
AU/33851-AP, PO Box 830470
Birmingham, AL 35283

SIMEON J. EARL – BAE SYSTEMS
W432A, Warton Aerodome
Preston, Lancs., UK PR4 1AX

ENGINEERING INFORMATION, INC
PO Box 543
Amsterdam, Netherlands 1000 Am

ETSE TELECOMUNICACION
Biblioteca, Campus Lagoas
Vigo, 36200 Spain

GA INSTITUTE OF TECHNOLOGY
EBS-Lib Mail code 0900
74 Cherry Street
Atlanta, GA 30332

TIMOTHY HOLZHEIMER
Raytheon
PO Box 1044
Rockwall, TX 75087

HRL LABS, RESEARCH LIBRARY
3011 Malibu Canyon
Malibu, CA 90265

IEE INSPEC
Michael Faraday House
6 Hills Way
Stevenage, Herts UK SG1 2AY

INSTITUTE FOR SCIENTIFIC INFO.
Publication Processing Dept.
3501 Market St.
Philadelphia, PA 19104-3302

LIBRARY – DRDC OTTAWA
3701 Carling Avenue
Ottawa, Ontario, Canada K1A OZ4

LIBRARY of CONGRESS
Reg. Of Copyrights
Attn: 407 Deposits
Washington DC, 20559

LINDA HALL LIBRARY
5109 Cherry Street
Kansas City, MO 64110-2498

MISSOURI S&T
400 W 14th Street
Rolla, MO 56409

MIT LINCOLN LABORATORY
Periodicals Library
244 Wood Street
Lexington, MA 02420

NATIONAL CHI NAN UNIVERSITY
Lily Journal & Book Co, Ltd
20920 Glenbrook Drive
Walnut, CA 91789-3809

JOHN NORGARD
UCCS
20340 Pine Shadow Drive
Colorado Springs, CO 80908

OSAMA MOHAMMED
Florida International University
10555 W Flagler Street
Miami, FL 33174

NAVAL POSTGRADUATE SCHOOL
Attn:J. Rozdal/411 Dyer Rd./ Rm 111
Monterey, CA 93943-5101

NDL KAGAKU
C/0 KWE-ACCESS
PO Box 300613 (JFK A/P)
Jamaica, NY 11430-0613

OVIEDO LIBRARY
PO BOX 830679
Birmingham, AL 35283

DAVID PAULSEN
E3Compliance
1523 North Joe Wilson Road
Cedr Hill, TX 75104-1437

PENN STATE UNIVERSITY
126 Paterno Library
University Park, PA 16802-1808

DAVID J. PINION
1122 E Pike Street #1217
SEATTLE, WA 98122

KATHERINE SIAKAVARA
Gymnasiou 8
Thessaloniki, Greece 55236

SWETS INFORMATION SERVICES
160 Ninth Avenue, Suite A
Runnemede, NJ 08078

YUTAKA TANGE
Maizuru Natl College of Technology
234 Shiroya
Maizuru, Kyoto, Japan 625-8511

TIB & UNIV. BIB. HANNOVER
DE/5100/G1/0001
Welfengarten 1B
Hannover, Germany 30167

UEKAE
PO Box 830470
Birmingham, AL 35283

UNIV OF CENTRAL FLORIDA
4000 Central Florida Boulevard
Orlando, FL 32816-8005

UNIVERSITY OF COLORADO
1720 Pleasant Street, 184 UCB
Boulder, CO 80309-0184

UNIVERSITY OF KANSAS –
WATSON
1425 Jayhawk Blvd 210S
Lawrence, KS 66045-7594

UNIVERSITY OF MISSISSIPPI
JD Williams Library
University, MS 38677-1848

UNIVERSITY LIBRARY/HKUST
Clear Water Bay Road
Kowloon, Honk Kong

CHUAN CHENG WANG
8F, No. 31, Lane 546
MingCheng 2nd Road, Zuoying Dist
Kaoshiung City, Taiwan 813

THOMAS WEILAND
TU Darmstadt
Schlossgartenstrasse 8
Darmstadt, Hessen, Germany 64289

STEVEN WEISS
US Army Research Lab
2800 Powder Mill Road
Adelphi, MD 20783

YOSHIHIDE YAMADA
NATIONAL DEFENSE ACADEMY
1-10-20 Hashirimizu
Yokosuka, Kanagawa,
Japan 239-8686

INFORMATION FOR AUTHORS

PUBLICATION CRITERIA

Each paper is required to manifest some relation to applied computational electromagnetics. **Papers may address general issues in applied computational electromagnetics, or they may focus on specific applications, techniques, codes, or computational issues.** While the following list is not exhaustive, each paper will generally relate to at least one of these areas:

1. **Code validation.** This is done using internal checks or experimental, analytical or other computational data. Measured data of potential utility to code validation efforts will also be considered for publication.

2. **Code performance analysis.** This usually involves identification of numerical accuracy or other limitations, solution convergence, numerical and physical modeling error, and parameter tradeoffs. However, it is also permissible to address issues such as ease-of-use, set-up time, run time, special outputs, or other special features.

3. **Computational studies of basic physics.** This involves using a code, algorithm, or computational technique to simulate reality in such a way that better, or new physical insight or understanding, is achieved.

4. **New computational techniques** or new applications for existing computational techniques or codes.

5. **"Tricks of the trade"** in selecting and applying codes and techniques.

6. **New codes, algorithms, code enhancement, and code fixes.** This category is self-explanatory, but includes significant changes to existing codes, such as applicability extensions, algorithm optimization, problem correction, limitation removal, or other performance improvement. **Note: Code (or algorithm) capability descriptions are not acceptable, unless they contain sufficient technical material to justify consideration.**

7. **Code input/output issues.** This normally involves innovations in input (such as input geometry standardization, automatic mesh generation, or computer-aided design) or in output (whether it be tabular, graphical, statistical, Fourier-transformed, or otherwise signal-processed). Material dealing with input/output database management, output interpretation, or other input/output issues will also be considered for publication.

8. **Computer hardware issues.** This is the category for analysis of hardware capabilities and limitations of various types of electromagnetics computational requirements. Vector and parallel computational techniques and implementation are of particular interest.

Applications of interest include, but are not limited to, antennas (and their electromagnetic environments), networks, static fields, radar cross section, inverse scattering, shielding, radiation hazards, biological effects, biomedical applications, electromagnetic pulse (EMP), electromagnetic interference (EMI), electromagnetic compatibility (EMC), power transmission, charge transport, dielectric, magnetic and nonlinear materials, microwave components, MEMS, RFID, and MMIC technologies, remote sensing and geometrical and physical optics, radar and communications systems, sensors, fiber optics, plasmas, particle accelerators, generators and motors, electromagnetic wave propagation, non-destructive evaluation, eddy currents, and inverse scattering.

Techniques of interest include but not limited to frequency-domain and time-domain techniques, integral equation and differential equation techniques, diffraction theories, physical and geometrical optics, method of moments, finite differences and finite element techniques, transmission line method, modal expansions, perturbation methods, and hybrid methods.

Where possible and appropriate, authors are required to provide statements of quantitative accuracy for measured and/or computed data. This issue is discussed in "Accuracy & Publication: Requiring, quantitative accuracy statements to accompany data," by E. K. Miller, *ACES Newsletter,* Vol. 9, No. 3, pp. 23-29, 1994, ISBN 1056-9170.

SUBMITTAL PROCEDURE

All submissions should be uploaded to ACES server through ACES web site (http://aces.ee.olemiss.edu) by using the upload button, journal section. Only pdf files are accepted for submission. The file size should not be larger than 5MB, otherwise permission from the Editor-in-Chief should be obtained first. Automated acknowledgment of the electronic submission, after the upload process is successfully completed, will be sent to the corresponding author only. It is the responsibility of the corresponding author to keep the remaining authors, if applicable, informed. Email submission is not accepted and will not be processed.

EDITORIAL REVIEW

In order to ensure an appropriate level of quality control, papers are peer reviewed. They are reviewed both for technical correctness and for adherence to the listed guidelines regarding information content and format.

PAPER FORMAT

Only camera-ready electronic files are accepted for publication. The term **"camera-ready" means that the material is neat, legible, reproducible, and in accordance with the final version format listed below.**

The following requirements are in effect for the final version of an ACES Journal paper:

1. The paper title should not be placed on a separate page.

The title, author(s), abstract, and (space permitting) beginning of the paper itself should all be on the first page. The title, author(s), and author affiliations should be centered (center-justified) on the first page. The title should be of font size 16 and bolded, the author names should be of font size 12 and bolded, and the author affiliation should be of font size 12 (regular font, neither italic nor bolded).

2. An abstract is required. The abstract should be a brief summary of the work described in the paper. It should state the computer codes, computational techniques, and applications discussed in the paper (as applicable) and should otherwise be usable by technical abstracting and indexing services. The word "Abstract" has to be placed at the left margin of the paper, and should be bolded and italic. It also should be followed by a hyphen (─) with the main text of the abstract starting on the same line.

3. All section titles have to be centered and all the title letters should be written in caps. The section titles need to be numbered using roman numbering (I. II. ….)

4. Either British English or American English spellings may be used, provided that each word is spelled consistently throughout the paper.

5. Internal consistency of references format should be maintained. As a guideline for authors, we recommend that references be given using numerical numbering in the body of the paper (with numerical listing of all references at the end of the paper). The first letter of the authors' first name should be listed followed by a period, which in turn, followed by the authors' complete last name. Use a coma (,) to separate between the authors' names. Titles of papers or articles should be in quotation marks (" "), followed by the title of journal, which should be in italic font. The journal volume (vol.), issue number (no.), page numbering (pp.), month and year of publication should come after the journal title in the sequence listed here.

6. Internal consistency shall also be maintained for other elements of style, such as equation numbering. Equation numbers should be placed in parentheses at the right column margin. All symbols in any equation have to be defined before the equation appears or right immediately following the equation.

7. The use of SI units is strongly encouraged. English units may be used as secondary units (in parentheses).

8. Figures and tables should be formatted appropriately (centered within the column, side-by-side, etc.) on the page such that the presented data appears close to and after it is being referenced in the text. When including figures and tables, all care should be taken so that they will appear appropriately when printed in black and white. For better visibility of paper on computer screen, it is good to make color figures with different line styles for figures with multiple curves. Colors should also be tested to insure their ability to be distinguished after black and white printing. Avoid the use of large symbols with curves in a figure. It is always better to use different line styles such as solid, dotted, dashed, etc.

9. A figure caption should be located directly beneath the corresponding figure, and should be fully justified.

10. The intent and meaning of all text must be clear. For authors who are not masters of the English language, the ACES Editorial Staff will provide assistance with grammar (subject to clarity of intent and meaning). However, this may delay the scheduled publication date.

11. Unused space should be minimized. Sections and subsections should not normally begin on a new page.

ACES reserves the right to edit any uploaded material, however, this is not generally done. It is the author(s) responsibility to provide acceptable camera-ready files in pdf and MSWord formats. Incompatible or incomplete files will not be processed for publication, and authors will be requested to re-upload a revised acceptable version.

COPYRIGHTS AND RELEASES
Each primary author must execute the online copyright form and obtain a release from his/her organization vesting the copyright with ACES. Both the author(s) and affiliated organization(s) are allowed to use the copyrighted material freely for their own private purposes.

Permission is granted to quote short passages and reproduce figures and tables from and ACES Journal issue provided the source is cited. Copies of ACES Journal articles may be made in accordance with usage permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. The reproduction of multiple copies and the use of articles or extracts for commercial purposes require the consent of the author and specific permission from ACES. Institutional members are allowed to copy any ACES Journal issue for their internal distribution only.

PUBLICATION CHARGES
All authors are allowed for 8 printed pages per paper without charge. Mandatory page charges of $75 a page apply to all pages in excess of 8 printed pages. Authors are entitled to one, free of charge, copy of the printed journal issue in which their paper was published. Additional reprints are available for $ 50. Requests for additional re-prints should be submitted to the managing editor or ACES Secretary.

Corresponding author is required to complete the online form for the over page charge payment right after the initial acceptance of the paper is conveyed to the corresponding author by email.

ACES Journal is abstracted in INSPEC, in Engineering Index, DTIC, Science Citation Index Expanded, the Research Alert, and to Current Contents/Engineering, Computing & Technology.