Progress In Electromagnetics Research, Vol. 136, 703–724, 2013

A MULTI-GPU SOURCES RECONSTRUCTION METHOD FOR IMAGING APPLICATIONS

Miguel López-Portugués, Yuri Álvarez*, Jesús A. López-Fernández, Cebrián García, Rafael G. Ayestarán, and Fernando Las-Heras

Universidad de Oviedo, Área de Teoría de la Señal y Comunicaciones, Campus Universitario, Edificio Polivalente, Gijón 33203, Spain

Abstract—A profile reconstruction method using a surface inverse currents technique implemented on GPU is presented. The method makes use of the internal fields radiated by an equivalent currents distribution retrieved from scattered field information that is collected from multiple incident fields. Its main advantage over other inverse source-based techniques is the use of surface formulation for the inverse problem, which reduces the problem dimensionality thus decreasing the computational cost. In addition, the GPU implementation drastically reduces the calculation time, enabling the development of real time and accurate geometry reconstruction at a low cost.

1. INTRODUCTION

Microwave imaging has been a key issue for decades due to its wide range of applications. Among the great variety of methods for profile reconstruction using scattered field information [1–10], some of them can be classified as inverse source problems [2–7], where an equivalent currents distribution is reconstructed from the acquired scattered field. Profile and material characteristics are extracted from the equivalent currents, taking into account the relationship existing between the placement of the physical currents induced by the incident fields and the placement of the maximum levels of the reconstructed equivalent electric currents. The linear problem formulation, introduced for inverse problems in [4], is revisited in [7] aiming to improve the method resolution. While the advantage is the simplicity, more scattered field information (more incident waves and/or frequencies) is required to

Received 21 December 2012, Accepted 23 January 2013, Scheduled 1 February 2013

* Corresponding author: Yuri Álvarez (yalopez@tsc.uniovi.es).

get the same resolution as those techniques that make use of non-linear optimisation methods such as Particle Swarm Optimization [11] or Genetic Algorithms [12–14].

Other profile reconstruction methods are based on profile parameterisation in Fourier series, being the unknowns the coefficients of the series. The methods presented in [8, 9] are based on a non-linear cost function minimisation where the zero tangential field condition is considered. In such case, the unknown is the metallic body contour (C'). An auxiliary equivalent current distribution inside the contour is introduced in [10] that improves the estimation of the scattered field.

One of the inverse source problem drawbacks is the fact that the resulting system of equations relating the scattered field data and the equivalent current distribution is, in general, ill-posed, being necessary the use of regularisation methods (e.g., Tikhonov [7, 10]). Also, those ones based on profile parameterisation require the solution of non-linear system of equations using global optimisation techniques.

Even when the given inverse problem is linear, the high computational cost, especially in cases where the observation and reconstruction domains are electrically large, represents a major drawback of the above-mentioned methods. As a consequence, the calculation time can be a bottleneck in those applications requiring real-time features such as security screening [15–17]. To address this problem, two strategies may be adopted: i) Introduction of acceleration methods in the inversion algorithms. ii) Hardware acceleration such as using Graphics Processing Units (GPUs).

The first strategy can be successfully applied to the volumetric sources reconstruction method described in [7]. In fact, the inverse Fast Multipole Method presented in [18], achieves two orders of magnitude speedup with respect to [7]. However, this acceleration scheme is based on the far field approximation of the translation operator, which might limit its scope of application.

Concerning hardware acceleration, those algorithms that entail a high computational cost may benefit from the parallel programming. In this way, if a problem is prone to parallelisation, the time-to-solution may be dramatically reduced when the sequential algorithm is carefully redesigned to fit the underlying parallel architecture. In the field of computational electromagnetics, there are many success cases where the parallel computing is a key factor, including [19–21].

GPUs show a strong growth in adoption among scientific programmers, mainly due to two reasons: i) Great floating-point performance. ii) Broad availability of this kind of hardware accelerators.

The double-precision floating-point peak performance of a high-

end GPU such as the NVIDIA Tesla K20X reaches 1310 GFLOPS (Giga FLoating-point Operations Per Second) [22], whereas a high-end CPU such as the Intel Xeon E5-2690 only reaches 185.6 GFLOPS [23]. Furthermore, the performance gap is even greater in single-precision.

In relation with the spread of GPU accelerators, it should be noted that virtually any new graphics card may accomplish general-purpose computation. In this way, a simple workstation equipped with a consumer grade GPU may have enough computational power to solve very demanding problems.

Moreover, the graphics processors have proved to be suitable for solving compute-bound problems with a high degree of parallelism and a low memory footprint [24]. Therefore, and taking into account our previous experience [25], the Sources Reconstruction Method (SRM) seems to be an algorithm able to fully exploit the massively parallel architecture of the GPUs.

In this contribution, we propose a new accelerator scheme for the SRM, implemented on a multi-GPU platform, on imaging applications. The speed up scheme combines inverse and forward problems, as shown in Figure 1, but overcoming the limitation introduced by Green's function far field approximation [18]. The idea is to recover an equivalent current distribution on a surface enclosing the object-under-test (Figure 1(b)). These equivalent currents are recovered without forcing zero internal field condition (Love's Equivalence Principle [31, 32]). Then, the fields inside this reconstruction surface are calculated (Figure 1(c)), proving that the higher field amplitude values will correspond to the placement of the metallic parts of the object.

2. METHODOLOGY

2.1. Surface Inverse Problem: The SRM

The Sources Reconstruction Method for antenna diagnostics [26–32] is based on the electromagnetic Equivalence Principle [33], allowing the replacement of the object-under-test (typically an antenna but also a scatterer) by an equivalent electric (J_{eq}) and magnetic (M_{eq}) currents distribution retrieved on a surface S' enclosing the radiating elements (see Figure 1(b)). The field in the outer region ($\vec{E}_{scatt}$) can be represented by means of the field generated by these equivalent currents distribution.

Maxwell equations solution for free-space conditions (finite source distribution in homogeneous medium) gives the Integral Equations (1) and (2), relating the fields radiated by electric and magnetic currents

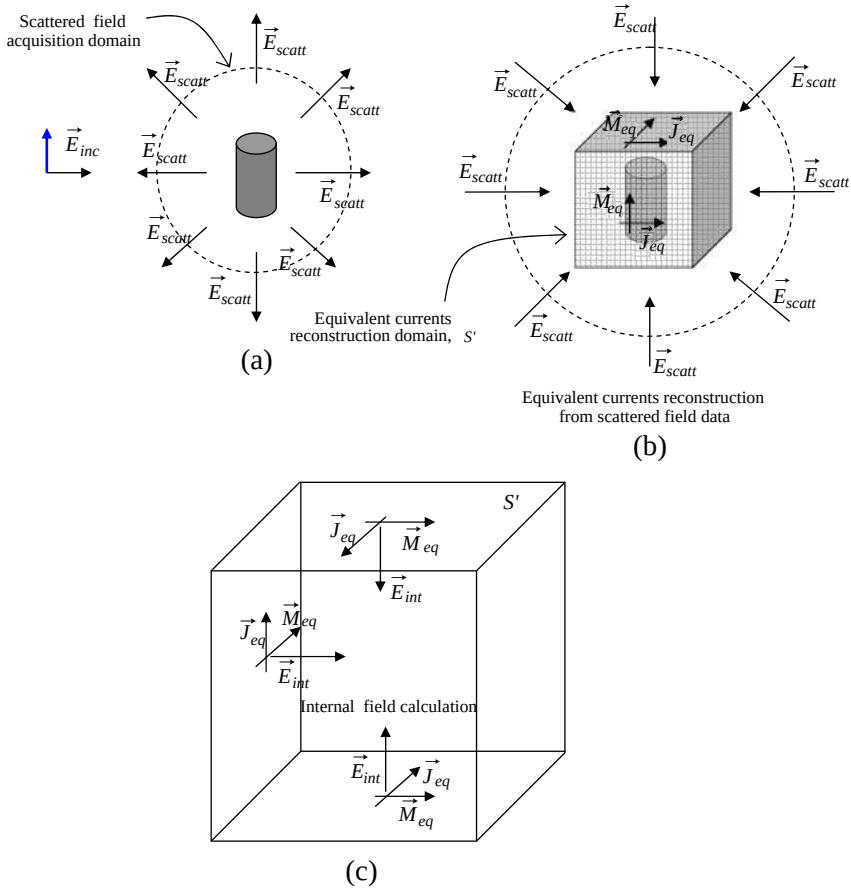

Figure 1. Profile reconstruction method. (a) Direct scattering problem. (b) Surface inverse problem. (c) Internal field calculation.

distribution:

$$\vec{E}_{scatt,J_{eq}}(\vec{r}) = -\frac{j\eta}{4\pi k_0}\int_{S'}\left\{\left(k_0^2+\nabla^2\right)\left[\frac{e^{-jk_0R(\vec{r};\vec{r}\,')}}{R(\vec{r};\vec{r}\,')}\vec{J}_{eq}(\vec{r}\,')\right]\right\}dS', \quad (1)$$

$$\vec{E}_{scatt,M_{eq}}(\vec{r}) = -\frac{1}{4\pi}\nabla\times\int_{S'}\left[\frac{e^{-jk_0R(\vec{r};\vec{r}\,')}}{R(\vec{r};\vec{r}\,')}\vec{M}_{eq}(\vec{r}\,')\right]dS', \quad (2)$$

where k_0 is the wavenumber and η is the intrinsic impedance of the medium. $\vec{r}(x, y, z)$ is the position vector for observation points, and $\vec{r}\,'(x', y', z')$ is the vector defining the position of the sources. $R(\vec{r};\vec{r}\,') = |\vec{r}-\vec{r}\,'|$. Integral Equations (1) and (2) can be expressed on

a particular coordinate system, as for example, Cartesian coordinates.

Solving the integral Equations (1) and (2) to calculate the equivalent currents given the scattered fields requires the discretisation of the previous integral Equations (1) and (2) by means of the Method-of-Moments (MoM), yielding the following linear system of equations:

$$\begin{bmatrix} E_{s,x} \\ E_{s,y} \\ E_{s,z} \end{bmatrix} = \begin{bmatrix} Z_{E_{s,x};J_{eq,x}} & Z_{E_{s,x};J_{eq,y}} & Z_{E_{s,x};J_{eq,z}} & Z_{E_{s,x};M_{eq,x}} & Z_{E_{s,x};M_{eq,y}} & Z_{E_{s,x};M_{eq,z}} \\ Z_{E_{s,y};J_{eq,x}} & Z_{E_{s,y};J_{eq,y}} & Z_{E_{s,y};J_{eq,z}} & Z_{E_{s,y};M_{eq,x}} & Z_{E_{s,y};M_{eq,y}} & Z_{E_{s,y};M_{eq,z}} \\ Z_{E_{s,z};J_{eq,x}} & Z_{E_{s,z};J_{eq,y}} & Z_{E_{s,z};J_{eq,z}} & Z_{E_{s,z};M_{eq,x}} & Z_{E_{s,z};M_{eq,y}} & Z_{E_{s,z};M_{eq,z}} \end{bmatrix} \begin{bmatrix} J_{eq,x} \\ J_{eq,y} \\ J_{eq,z} \\ M_{eq,x} \\ M_{eq,y} \\ M_{eq,z} \end{bmatrix} \tag{3}$$

$E_{s,x}$, $E_{s,y}$, $E_{s,z}$ are the cartesian components of the scattered field on the measurement surface. $J_{eq,x}$, $J_{eq,y}$, $J_{eq,z}$ and $M_{eq,x}$, $M_{eq,y}$, $M_{eq,z}$ are, respectively, the cartesian components of the electric and magnetic equivalent current on the reconstruction surface. $Z_{X;Y}$ are the Z-matrix terms relating the X-scattered field component to the Y-equivalent current component. As an example, the matrix term that relates the x-component of the electric field with the y-component of the equivalent magnetic currents is:

$$Z_{E_{s,x};M_{eq,y}}(n,m) = \frac{-1}{4\pi}\left(z_n - z'_m\right)\left[\frac{1+jk_0R\left(\vec{r}_n;\vec{r}'_m\right)}{R\left(\vec{r}_n;\vec{r}'_m\right)^3}\right] e^{-jk_0R(\vec{r}_n;\vec{r}'_m)}\Delta S'_m, \tag{4}$$

n index is related to the scattered field observation domain, and m index is related to the equivalent currents reconstruction domain. $\Delta S'_m$ denotes a reconstruction surface patch.

Sometimes, instead of using the Cartesian components of the fields and currents, a coordinate system transformation is applied so only two tangential components are needed, as [28, 29]. Thus, the system of equations to be solved is:

$$\begin{bmatrix} E_{s,t_1} \\ E_{s,t_2} \end{bmatrix} = \begin{bmatrix} Z_{E_{s,t_1};J_{eq,t_1}} & Z_{E_{s,t_1};J_{eq,t_2}} & Z_{E_{s,t_1};M_{eq,t_1}} & Z_{E_{s,t_1};M_{eq,t_2}} \\ Z_{E_{s,t_2};J_{eq,t_1}} & Z_{E_{s,t_2};J_{eq,t_2}} & Z_{E_{s,t_2};M_{eq,t_1}} & Z_{E_{s,t_2};M_{eq,t_2}} \end{bmatrix} \begin{bmatrix} J_{eq,t_1} \\ J_{eq,t_2} \\ M_{eq,t_1} \\ M_{eq,t_2} \end{bmatrix} \tag{5}$$

where subindexes t_1 and t_2 denote the tangential components of the fields and equivalent currents in a base on an orthogonal coordinate system.

In order to solve the linear system of Equation (5), a cost function that relates the measured fields to the fields radiated by the equivalent currents must be minimised:

$$f(I) = \frac{1}{2}\left\|[E_s] - \begin{bmatrix} Z_{E;J_{eq}} & Z_{E;M_{eq}} \end{bmatrix}\begin{bmatrix} J_{eq} \\ M_{eq} \end{bmatrix}\right\|^2 = \frac{1}{2}\left\|[E] - [Z][I]\right\|^2 \quad (6)$$

The Conjugate Gradient Normal Residual (CGNR) [34], also known as Conjugate Gradient on the Normal Equations [26], is used for minimising the cost function (6). The first CGNR step is the initialisation of the algorithm by fixing the solution vector $[I] = 0$. The initial residue $[r_0]$ and the initial descendent direction $[p_1]$ are calculated as follows:

$$\begin{aligned} [r_0] &= -[E] \\ [p_1] &= -[Z]^H[r_0] = [num_0] \end{aligned} \quad (7)$$

where $[Z]^H$ is the Hermitian matrix of $[Z]$.

At the k-iteration, the parameters α_k and β_k of the CGNR algorithm as well as the solution vector, the descent direction and the residue are updated:

$$\begin{aligned} \alpha_k &= -\frac{\left\|[Z]^H[r_{k-1}]\right\|}{\|[Z][p_k]\|} = -\frac{\|[num_{k-1}]\|_2}{\|[den_k]\|_2} \\ [I_k] &= [I_{k-1}] + \alpha_k[p_k] \\ [\hat{E}_k] &= [Z][I_k] \\ [r_k] &= [\hat{E}_k] - [E] \;=\; [r_{k-1}] + \alpha_k[den_k] \\ \beta_k &= -\frac{\left\|[Z]^H[r_k]\right\|}{\|[Z]^H[r_{k-1}]\|} = -\frac{\|[num_k]\|_2}{\|[num_{k-1}]\|_2} \\ [p_{k+1}] &= -[num_k] + \beta_k[p_k] \end{aligned} \quad (8)$$

The accuracy of the solution in the k-iteration is estimated using the Root Mean Square Error (RMSE) between the value estimated of the independent term $[\hat{E}_k]$ and the true value of the independent term $[E]$ as it is indicated below:

$$RMSE_k = \frac{1}{\sqrt{N}}\left\|[E] - [\hat{E}_k]\right\| \quad (9)$$

As the CGNR exhibits a monotonic decrease of the error, the stop condition of the CGNR is fixed from the $RMSE_k$ with a threshold to reach.

2.2. Internal Field Calculation

Some implementations of the SRM [31, 32] are based on the Love's Equivalence Principle, in which the field inside the reconstruction domain S' is forced to be zero. However, in the proposed methodology, the internal field is needed to extract information about the geometry of the scatterer, so the Love's Equivalence Principle is not applied.

The internal electric field inside S' is calculated, from the reconstructed equivalent currents on the reconstruction surface S', as follows:

$$\begin{bmatrix} E_{i,u_1} \\ E_{i,u_2} \\ E_{i,u_3} \end{bmatrix} = \begin{bmatrix} Z_{E_{i,u_1};J_{eq,t_1}} & Z_{E_{i,u_1};J_{eq,t_2}} & Z_{E_{i,u_1};M_{eq,t_1}} & Z_{E_{i,u_1};M_{eq,t_2}} \\ Z_{E_{i,u_2};J_{eq,t_1}} & Z_{E_{i,u_2};J_{eq,t_2}} & Z_{E_{i,u_2};M_{eq,t_1}} & Z_{E_{i,u_2};M_{eq,t_2}} \\ Z_{E_{i,u_3};J_{eq,t_1}} & Z_{E_{i,u_3};J_{eq,t_2}} & Z_{E_{i,u_3};M_{eq,t_1}} & Z_{E_{i,u_3};M_{eq,t_2}} \end{bmatrix} \begin{bmatrix} J_{eq,t_1} \\ J_{eq,t_2} \\ M_{eq,t_1} \\ M_{eq,t_2} \end{bmatrix} \quad (10)$$

E_{i,u_p} is the internal field p-component defined according to an orthonormal coordinate system (typically, it will be Cartesian components, with $p = 1, 2, 3$ corresponding to x, y, and z components respectively). $Z_{X;Y}$ are, again, the Z-matrix elements relating the X-internal field component with the Y-equivalent current component, as defined in (4).

The analysis of these internal fields reveals a correspondence between the placement of the radiating elements (e.g., the induced currents on a metallic surface) and the highest amplitude values of the internal fields. The internal field amplitude is given by the root mean square value (RMS):

$$|E_i| = \sqrt{\frac{1}{3}\sum_{p=1}^{3} \left|E_{i,u_p}\right|^2} \quad (11)$$

Concerning numerical issues, the inverse problem to be solved has M equations (number of scattered field acquisition points $\times$ 2 tangential field components), and N unknowns (number of subdomains in which the reconstruction surface S' is discretised $\times$ 2 types of currents $\times$ 2 tangential components). And for the forward internal problem, M_i values are calculated (number of internal points $\times$ 3 field components).

Finally, the object-under-test can be illuminated using multiple frequencies and incident directions to increase the amount of information about the scatterer. Combination of the internal fields recovered for every incident direction and working frequency improves

the profile reconstruction accuracy. Results are combined again using the RMS for all the L-incidences and F-frequencies:

$$|E_i| = \sqrt{\frac{1}{F \cdot L} \sum_{l=1}^{L} \sum_{f=1}^{F} |E_i|_{f,l}^2} \tag{12}$$

The volumetric Sources Reconstruction Method [7] requires solving a linear inverse problem having M equations and M_i unknowns. The main drawback of this formulation is that the reconstruction domain (a volume) is one dimension higher than the scattered field observation domain (a surface). Let us assume that, given a particular physical size of observation and reconstruction domain, we have M equations and M_i unknowns for a frequency f_0. Now, if the working frequency is increased to $2f_0$, while keeping the same problem discretisation (subdomain size) in terms of electric size (i.e., wavelength), the number of equations would be $4M$, but the number of unknowns would increase to $8M_i$.

By means of the methodology presented in this contribution, the inverse problem to be solved relates two domains (surfaces) with the same dimensionality. In the previous example, the number of equations would be M, and the number of unknowns, N, for f_0. In this case, doubling the frequency means that the number of unknowns would be $4N$. Thus, for electrically large problems, the number of unknowns on a surface domain S' will be significantly smaller than the number of unknowns in the volume enclosed by S'.

3. PARALLEL ALGORITHM FOR MULTIPLE GPUS

In this work, we deal with a quadratic problem whose computational complexity is $O(MN)$ in terms of time and $O(\max(M, N))$ in memory. In order to achieve such a linear memory cost, though, the storage of the whole Z-matrix must be avoided. Thus, the time cost associated to the problem is mainly due to the calculation of the elements of the Z-matrix and the Matrix-Vector Products (MVPs) of each CGNR iteration. Once the equivalent currents are reconstructed, the calculation of the internal field requires only two additional MVPs, one for the field due to the electric currents and one for the field due to the magnetic currents.

In order to exploit the intrinsic parallelism of the problem, we chose a partitioning scheme comprising two levels. The different angles of incidence represent the first level of parallelism, whereas the MVPs that are performed for each angle of incidence constitute the second level.

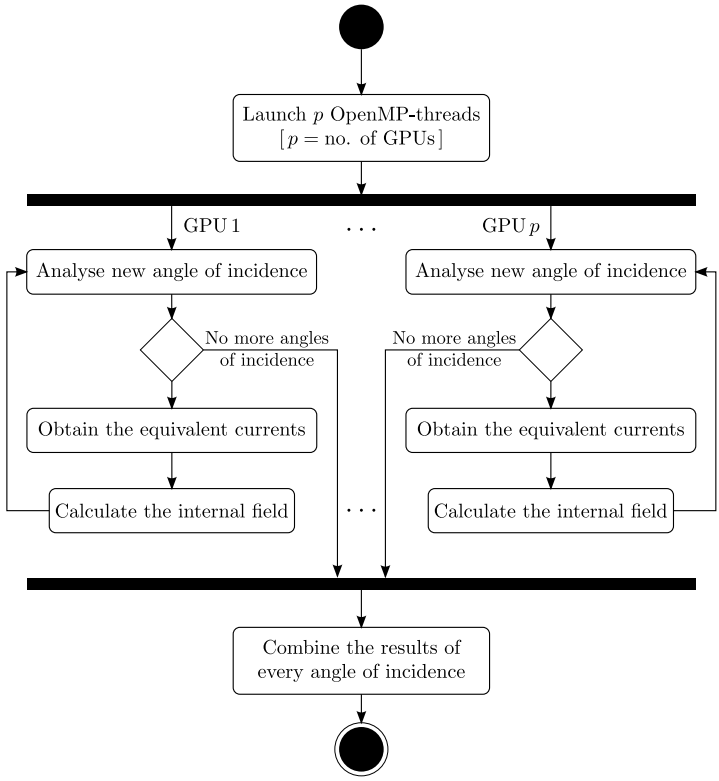

Figure 2. Parallel implementation using multiple GPUs.

Figure 2 is a flowchart that shows the strategy used for the first level of parallelism (i.e., the coarse level). In this level, we take advantage of OpenMP multi-threading [35] to use multiple graphics cards concurrently. The loop that iterates through the different angles of incidence was replaced with a parallel construct where the number of created threads matches the number of available GPUs. In this manner, it is possible to analyse more than one angle of incidence at a time. Furthermore, since a given computer may have different graphics cards installed, we decided to use the OpenMP dynamic scheduling to achieve a good load-balancing among the GPUs.

The second level of parallelism (i.e., the fine level) tackles the MVPs required to calculate the internal field for each angle of incidence. In this level, we use the CUDA (Compute Unified Device Architecture) C programming language [36] to turn the initial MVP into many simpler sub-problems that may be computed concurrently

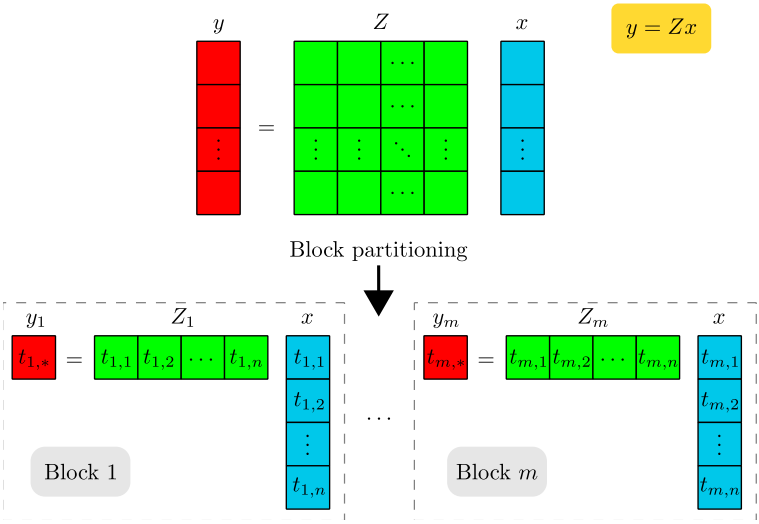

Figure 3. MVP parallelisation technique avoiding the Z-matrix storage. Example using a Z-matrix with m rows and n columns, m blocks and n threads per block.

using thousands of threads running on the GPU.

The parallelisation technique of the MVPs is depicted in Figure 3. This technique is based on block partitioning, in such a way that matrix rows are assigned to CUDA blocks [36] and the elements within a row are assigned to the threads within a given block. First, each thread computes on the fly its corresponding element of the Z-matrix. In this manner, the matrix is not permanently stored, neither in CPU nor in GPU memory, yielding a very low memory footprint. Then, each thread multiplies its element of the Z-matrix and its element of the x vector. Finally, all the threads that pertain to the same block contribute to the corresponding entry in the y vector by means of a parallel reduction.

Our multi-GPU implementation of the SRM is made up of ten kernels (i.e., functions that are executed in parallel by many CUDA threads) for reconstructing the equivalent currents and two kernels for calculating the internal field, using half of them to deal with electric currents and the other half to deal with magnetic currents. Those kernels that deal with the reconstruction of the equivalent currents were first used by the authors in [25]. Since then, we found several ways for improving the performance of the above mentioned kernels. The most relevant changes are related to the reduction in the register

Table 1. CUDA kernels execution parameters for both Fermi and Kepler architectures.

Fermi architecture	
Compute capability	2.1
Shared memory : L1 cache	16 kB : 48 kB
Number of blocks $*$ threads per block	$8192 * 128$
Shared memory per block	1.5–2.5 kB
Registers per thread	34–50
Multiprocessor occupancy	42–58%
Kepler architecture	
Compute capability	3.0
Shared memory : L1 cache	32 kB : 32 kB
Number of blocks $*$ threads per block	$8192 * 128$
Shared memory per block	1.5–2.5 kB
Registers per thread	35–52
Multiprocessor occupancy	56–75%

usage per thread, just as the better usage of the shared memory (for further details on CUDA terminology, please see [36]).

Table 1 summarises the execution parameters of the developed kernels for both Fermi [37] and Kepler [38] architectures. After doing benchmarking, we found that the shown parameters are the ones that deliver the best performance on each GPU architecture. However, some parameters vary from kernel to kernel, so they are shown as a range of values. It is worth noting the high occupancy (defined as the ratio of the number of active threads to the maximum possible number of active threads) is achieved due to the moderate usage of both registers and shared memory. The occupancy has gone from 33% in [25] to 42–58% in the case of the Fermi GPUs and 56–75% in the case of the Kepler GPUs. In order to compare the current occupancy to that of our previous work [25], the occupancy calculator that is part of the CUDA toolkit [39] has been used.

4. APPLICATION EXAMPLE

In order to illustrate the capabilities of the proposed technique for geometry reconstruction, we chose two metallic scatterers as the objects-under-test. The first one is a cylinder, centred in $x = -0.75\lambda$, $y = 0.25\lambda$, $z = 0\lambda$, having 0.5λ diameter and 2λ height (along z-axis).

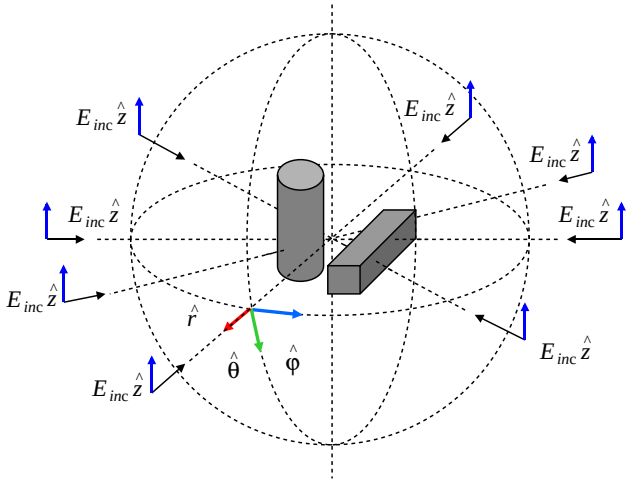

Figure 4. Incident plane waves.

The second one is a rectangular prism centred in $x = 1\lambda$, $y = 0\lambda$, $z = -0.5\lambda$, with length in x, y, and z of 0.5λ, 2λ, and 1λ, respectively.

First, the direct scattering problem is evaluated by means of a commercial MoM software. The object-under-test is illuminated by 8 z-polarised plane waves placed in the $z = 0$ plane. Incidences are equally spaced from $\varphi = 0°$ to 315° in 45°-steps, as shown in Figure 4. For every incident wave, the field scattered by the objects is acquired on a 10λ radius sphere. The acquisition sphere is sampled every 3° in θ and 6° in φ, resulting in 3600 acquisition points.

Second, for each one of these 8 direct problems, an inverse problem is solved, consisting on an equivalent currents retrieval on a $Lx = Ly = Lz = 4\lambda$ box which encloses the domain-under-study. The surface of the box is discretised into 0.25λ-edge square patches, yielding 1734 square patches. The inverse problem (i.e., the equivalent currents calculation) requires solving a linear system that consists of $M = 7200$ Equations (3600 acquisition points $\times$ 2 tangential field components, see Equation (5)) with $N = 6936$ unknowns (1734 subdomains $\times$ 2 tangential components $\times$ 2 kinds of equivalent currents, see Equation (5)). The linear system of Equation (5) is solved by minimising a cost function using the Conjugate Gradient method [26]. Convergence is considered to be reached when the root mean square error between the scattered field and the field radiated by the reconstructed equivalent currents is less than 2%, which is achieved in 11–12 iterations.

Then, from the reconstructed equivalent currents, the internal

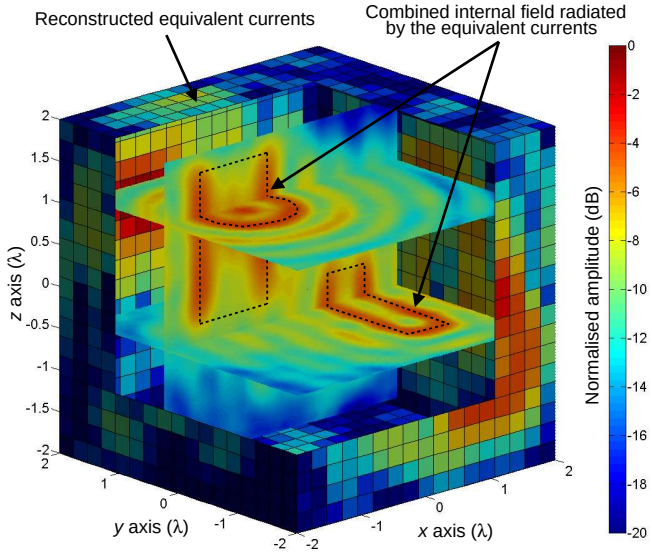

Figure 5. Reconstructed equivalent currents on the box enclosing the object-under-test, and calculated internal fields on several slices. Combination of 8 inverse problems. The dashed black line represents the objects-under-test profiles.

fields inside the box are evaluated in 19683 equally-spaced points (27 points per x, y, z dimension, so $\Delta x = \Delta y = \Delta z = 0.14\lambda$), so that $M_i = 19683$ points $\times$ 3 internal field components $= 59049$ (see Equation (5)). Finally, the field components and the 8 solutions are combined as indicated in Equations (11) and (12), respectively.

4.1. Accuracy of the Method

The combined internal fields in several slices, as well as the reconstructed equivalent currents on the box enclosing the domain-under-test, are represented in Figure 5. It is possible to appreciate that the maximum of the combined internal field amplitude fits the metallic scatterers geometry.

For a better representation of how the combined internal fields are able to provide geometry information, the -3 dB amplitude isosurfaces are plotted in Figure 6. It can be observed that the -3 dB isosurface fits the objects contour. It should be noted, however, that the upper and lower surfaces of the objects are not reconstructed because they are not directly illuminated by any incident plane wave.

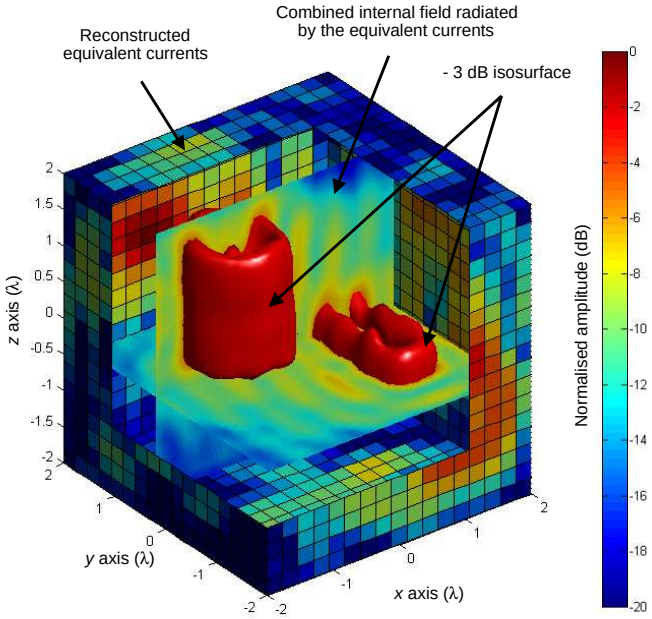

Figure 6. Reconstructed equivalent currents on the box enclosing the object-under-test, calculated internal fields on several slices, and −3 dB field amplitude isosurfaces.

Table 2. Comparison between both CPU and GPU implementations of the algorithm.

Comparison	Root Mean Square Error
CPU — Tesla GPU	0.014%
CPU — Kepler GPU	0.013%

Concerning the accuracy of the implemented algorithm on GPU, comparison between the combined internal fields for all the incidences is shown in Table 2. It can be concluded that differences between CPU and GPU implementations are negligible, as the RMSE is smaller than 0.015% in all the cases. As a reference, the inverse method it is considered to converge when the field RMSE is smaller than 2%.

4.2. Speedup by Means of Multiple GPUs

To obtain the results presented in this section, we used two workstations with different GPU architectures. The first workstation consists of 1 CPU (Intel Core i5-2500T with 4 cores at 2.3 GHz), 16 GB

of RAM, and 2 NVIDIA Fermi GPUs (1 GTX 460 with 336 cores at 1.43 GHz, and 1 GTX 560 with 336 cores at 1.62 GHz). The second workstation consists of 1 CPU (Intel Core i7-3820 with 4 cores and 8 threads at 3.6 GHz), 64 GB of RAM, and 2 NVIDIA Kepler GPUs (2 GTX 680 with 1536 cores at 1.07 GHz).

With the aim of proving the performance of our multi-GPU SRM, we compared it to our own parallel CPU implementation. On the one hand, the multi-GPU implementation is written in C and CUDA C. On the other hand, the reference CPU implementation is written in C (the SRM) and MATLAB (the calculation of the internal field). All of the C codes were compiled using Intel icc 11.1, whereas the GPU codes were compiled with NVIDIA nvcc 4.2. Moreover, MATLAB 2012a was used to deal with the CPU implementation. It should also be noted that all codes are based on single precision arithmetic.

Table 3 shows the time cost of the multi-GPU SRM compared to a reference parallel CPU implementation, using the first workstation. The most noticeable achievement of the multi-GPU implementation is the ability to provide a solution in quasi real-time, since it only takes 4.1 s when using one GPU and 2.2 s with two GPUs. On the contrary, the reference CPU implementation takes 350 s to do all the calculations, despite it is using 4 threads (1 thread per CPU core).

Table 3. Runtime of the CPU implementation vs the multi-GPU implementation, using the first workstation.

Workstation with Fermi GPUs			
CPU	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	$1.1 \cdot 10^{1}$ s
		Overall calculation time	$1.4 \cdot 10^{2}$ s
	Internal field	Avg. time per incidence	$1.8 \cdot 10^{1}$ s
		Overall calculation time	$2.1 \cdot 10^{2}$ s
CPU + 1 GPU	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	$3.2 \cdot 10^{-1}$ s
		Overall calculation time	3.8 s
	Internal field	Avg. time per incidence	$2.7 \cdot 10^{-2}$ s
		Overall calculation time	$3.2 \cdot 10^{-1}$ s
CPU + 2 GPUs	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	$1.7 \cdot 10^{-1}$ s
		Overall calculation time	2.0 s
	Internal field	Avg. time per incidence	$1.3 \cdot 10^{-2}$ s
		Overall calculation time	$1.6 \cdot 10^{-1}$ s

Given the data shown in Table 3, one may also calculate how many times the multi-GPU implementation is faster than the reference CPU implementation. When using two GPUs, the SRM is 69 times faster than the CPU-only version, whereas the computation of the internal field is 1313 times faster. Although, a remark should be done on the overall calculation time of the internal field. Since the performance of MATLAB is usually low even when compared to a code written in C, the speed of the kernels that deal with the internal field should not be excessively overrated.

It is also worth mentioning the parallel efficiency, which is a performance metric that may be defined as follows:

$$E_p = \frac{T_1}{p * T_p}, \tag{13}$$

where p is the number of GPUs, T_1 the execution time with 1 GPU, and T_p the execution time when using p GPUs. The parallel efficiency allows to know if it may be convenient to add more processors (in this case, GPUs) for solving a given problem. If the efficiency is low (close to 0), it means that there is almost no reduction in the runtime when using more GPUs. On the contrary, if the efficiency is high (close to 1), there is almost a linear reduction in the runtime when using more GPUs. In this case, E_2 reaches 95.8%. Such a high efficiency shows that the developed algorithm works well when using multiple GPUs to reduce even more the runtime.

In Table 4, we summarise the run time of the multi-GPU SRM compared to the CPU implementation when using the second workstation. Once more, the multi-GPU implementation is able to provide a solution in quasi real-time, as it only takes 1.6 s when using one GPU and 0.9 s when using two GPUs. As a reference, the CPU implementation takes 246 s to obtain the solution on this workstation.

In this case, when using two GPUs, the SRM is 88 times faster than the reference (CPU) version, whereas the calculation of the internal field is 2621 times faster. Once again, the low performance of MATLAB results in a huge speedup in the internal field calculation.

Finally, it is also worth noting the parallel efficiency when we use two GPUs, that is E_2, which reaches 93.6%.

5. EXPERIMENTAL VALIDATION

Once the effectiveness in terms of calculation time and accuracy of the imaging technique has been proved by means of simulations, next step is validation with experimental data. Aiming to provide a comparison

Table 4. Runtime of the CPU implementation vs the multi-GPU implementation, using the second workstation.

Workstation with Kepler GPUs			
CPU	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	6.3 s
		Overall calculation time	$7.6 \cdot 10^1$ s
	Internal field	Avg. time per incidence	$1.4 \cdot 10^1$ s
		Overall calculation time	$1.7 \cdot 10^2$ s
CPU + 1 GPU	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	$1.3 \cdot 10^{-1}$ s
		Overall calculation time	1.6 s
	Internal field	Avg. time per incidence	$1.1 \cdot 10^{-2}$ s
		Overall calculation time	$1.3 \cdot 10^{-1}$ s
CPU + 2 GPUs	SRM	RMS error	$\leq 2\%$
		Avg. time per incidence	$7.2 \cdot 10^{-2}$ s
		Overall calculation time	$8.6 \cdot 10^{-1}$ s
	Internal field	Avg. time per incidence	$5.5 \cdot 10^{-3}$ s
		Overall calculation time	$6.5 \cdot 10^{-2}$ s

with other imaging methods, an example described in Section 5.2 of [40] has been selected.

The object-under-test is a dielectric target made of $d_{dielectric} = 3\,\text{cm}$ diameter twin cylinders (cross-section centered at $(x, y) = (0, 4.5)$ and $(0, -4.5)\,\text{cm}$ respectively, see Figure 7(a)), with relative permittivity of $\epsilon_r = 3 \pm 0.3$. The working frequency band ranges from 1 to 8 GHz, sampled in 1 GHz steps. The scattered field is acquired on a $R = 76\,\text{cm}$ circumference, placed at $z = 0\,\text{cm}$. The sampling interval is from $\varphi = 60°$ to 300° in 5°-step, thus resulting in 49 field samples. For this example, only one incidence (source placed at $x = -72\,\text{cm}$, $y = 0\,\text{cm}$, $z = 0\,\text{cm}$) is considered. Field samples can be downloaded from [41] (file twodielTM_8f.exp).

For the inverse problem, a $Lx = Ly = Lz = 20\,\text{cm}$ box open on the upper and lower sides is chosen. The surface of the box is discretised into square patches (0.33λ edge at the highest frequency, 8 GHz), resulting in 1156 patches. From the equivalent currents retrieved on this surface, the internal field is evaluated in ∼40000 points inside this box. Calculation time for the inverse problem plus internal field calculation on our workstation with 2 Kepler GPUs is 0.32 s, which proves the real-time capabilities of the proposed imaging system.

Results combining all the frequencies are depicted in Figure 7. Due to the dielectric nature of the cylinders, the slower wave speed propagation inside a dielectric medium, and the fact

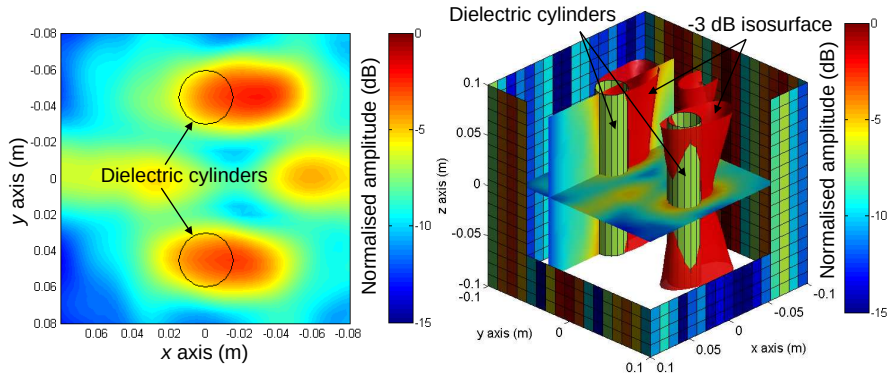

Figure 7. (a) Internal field evaluated on the $z = 0\,\text{cm}$ cut plane. (b) Reconstructed equivalent currents on the box enclosing the object-under-test, calculated internal fields on $z = 0$ and $y = -0.45\,\text{cm}$ cut planes, and $-3\,\text{dB}$ field amplitude isosurfaces.

that the reconstruction technique assumes free-space medium, the reconstructed shape is displaced $d = d_{dielectric} \cdot (\sqrt{\epsilon_r} - 1) = 2.2\,\text{cm}$ [16] backwards to the true position (that is, $-2.2\,\text{cm}$ in x direction), as observed in Figure 7(a). $-3\,\text{dB}$ internal field isosurfaces are plotted in Figure 7(b).

6. CONCLUSIONS

An accurate and fast method for geometry reconstruction has been presented. The novelties are both theoretical and hardware-related improvements. First, the dimensionality of the inverse problem is reduced from a volume to a surface reconstruction domain, using a forward step to evaluate the internal field. Second, our efficient multi-GPU implementation allows reducing the computational cost in almost two orders of magnitude when compared to our parallel CPU implementation. In consequence, the resulting technique provides real-time capabilities for problems that would require up to 4 minutes calculation time on a modern CPU with a parallel-CPU approach.

ACKNOWLEDGMENT

This work has been partially supported by the European Union under COST action IC1102 (VISTA); by "Ministerio de Ciencia e

Innovación" from Spain/FEDER under the projects TEC2011-24492/TEC (iScat) and CONSOLIDER CSD2008-00068 (TeraSense), and grant BES-2009-024060; by "Gobierno del Principado de Asturias" (PCTI)/FEDER-FSE under the projects IPT-2011-0951-390000 (Tecnigraf), EQUIP08-06, FC09-COF09-12, EQUIP10-31, and grant BP11-166.

REFERENCES

1. Donelli, M., I. J. Craddock, D. Gibbins, and M. Sarafianou, "A three dimensional time domain microwave imaging method for breast cancer detection based on an evolutionary algorithm," *Progress In Electromagnetics Research M*, Vol. 18, 179–195, 2011.
2. Van Den Berg, P. M. and R. E. Kleinman, "A contrast source inversion method," *Inverse Problems*, Vol. 13, No. 6, 1607–1620, 1997.
3. Lin, C. Y. and Y. W. Kiang, "Inverse scattering for conductors by the equivalent source method," *IEEE Trans. Antennas Propag.*, Vol. 44, No. 3, 310–316, 1996.
4. Caorsi, S., G. L. Gragnani, and M. Pastorino, "Two-dimensional microwave imaging by a numerical inverse scattering solution," *IEEE Trans. Microw. Theory Techn.*, Vol. 38, No. 8, 981–989, 1990.
5. Catapano, I., L. Crocco, and T. Isernia, "On simple methods for shape reconstruction of unknown scatterers," *IEEE Trans. Antennas Propag.*, Vol. 55, No. 5, 1431–1436, 2007.
6. Devaney, A. J. and G. C. Sherman, "Nonuniqueness in inverse source and scattering problems," *IEEE Trans. Antennas Propag.*, Vol. 30, No. 5, 1034–1037, 1982.
7. Álvarez, Y., B. A. Casas, C. García, and F. Las-Heras, "Geometry reconstruction of metallic bodies using the sources reconstruction method," *IEEE Antennas Wireless Propag. Lett.*, Vol. 9, 1197–1200, 2010.
8. Çayören, M., I. Akduman, A. Yapar, and L. Crocco, "A new algorithm for the shape reconstruction of perfectly conducting objects," *Inverse Problems*, Vol. 23, No. 3, 1087–1100, 2007.
9. Farmahini-Farahani, M., R. Faraji-Dana, and M. Shahabadi, "Fast and accurate cascaded particle swarm gradient optimization method for solving 2-D inverse scattering problems," *Appl. Comput. Electrom.*, Vol. 24, No. 5, 511–517, 2009.
10. Qin, Y. M. and I. R. Ciric, "Inverse scattering solution with current modeling and Tikhonov regularization," *Proc. IEEE*

Antennas Propag. Soc. Int. Symp., Ann Arbor, Michigan, USA, 1993.

11. Donelli, M. and A. Massa, "Computational approach based on a particle swarm optimizer for microwave imaging of two-dimensional dielectric scatterers," *IEEE Trans. Microw. Theory Techn.*, Vol. 53, No. 5, 1761–1776, 2005.
12. Massa, A., D. Franceschini, G. Franceschini, M. Pastorino, M. Raffetto, and M. Donelli, "Parallel GA-based approach for microwave imaging applications," *IEEE Trans. Antennas Propag.*, Vol. 53, No. 10, 3118–3127, 2005.
13. Caorsi, S., A. Massa, M. Pastorino, and M. Donelli, "Improved microwave imaging procedure for nondestructive evaluations of two-dimensional structures," *IEEE Trans. Antennas Propag.*, Vol. 52, No. 6, 1386–1397, 2004.
14. Benedetti, M., M. Donelli, and A. Massa, "Multicrack detection in two-dimensional structures by means of GA-based strategies," *IEEE Trans. Antennas Propag.*, Vol. 55, No. 1, 205–215, 2007.
15. Álvarez, Y., B. González-Valdés, J. Ángel Martínez, F. Las-Heras, and C. M. Rappaport, "3D whole body imaging for detecting explosive-related threats," *IEEE Trans. Antennas Propag.*, Vol. 60, No. 9, 4453–4458, 2012.
16. Martínez-Lorenzo, J. A., F. Quivira, and C. M. Rappaport, "SAR imaging of suicide bombers wearing concealed explosive threats," *Progress In Electromagnetics Research*, Vol. 125, 255–272, 2012.
17. Cooper, K. B., R. J. Dengler, N. Llombart, B. Thomas, G. Chattopadhyay, and P. H. Siegel, "THz imaging radar for standoff personnel screening," *IEEE Trans. THz Sci. Technol.*, Vol. 1, No. 1, 169–182, 2011.
18. Álvarez, Y., J. A. Martínez, F. Las-Heras, and C. M. Rappaport, "An inverse fast multipole method for geometry reconstruction using scattered field information," *IEEE Trans. Antennas Propag.*, Vol. 60, No. 7, 3351–3360, 2012.
19. Zhang, Y. and T. Sarkar, *Parallel Solution of Integral Equation-based EM Problems in the Frequency Domain*, Wiley-IEEE Press, New Jersey, 2009.
20. Araújo, M. G., J. M. Taboada, F. Obelleiro, J. M. Bértolo, L. Landesa, J. Rivero, and J. L. Rodríguez, "Supercomputer aware approach for the solution of challenging electromagnetic problems," *Progress In Electromagnetics Research*, Vol. 101, 241–256, 2010.
21. Taboada, J. M., M. G. Araújo, J. M. Bértolo, L. Landesa,

F. Obelleiro, and J. L. Rodríguez, "MLFMA-FFT parallel algorithm for the solution of large-scale problems in electromagnetics," *Progress In Electromagnetics Research*, Vol. 105, 15–30, 2010.

22. NVIDIA Corporation, "Tesla Kepler GPU accelerators," 2012, Available online at: http://www.nvidia.com/content/tesla/pdf/Tesla-KSeries-Overview-LR.pdf.
23. Intel Corporation, "Intel Xeon processor E5-2600 series," 2012, Available online at: http://download.intel.com/support/processors/xeon/sb/xeon_E5-2600.pdf.
24. Owens, J. D., M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, "GPU computing," *Proc. IEEE*, Vol. 5, No. 96, 879–899, 2008.
25. López-Fernández, J. A., M. López-Portugués, Y. Álvarez, C. García, D. Martínez-Álvarez, and F. Las-Heras, "Fast antenna characterization using the sources reconstruction method on graphics processors," *Progress In Electromagnetics Research*, Vol. 126, 185–201, 2012.
26. Álvarez, Y., F. Las-Heras, and M. R. Pino, "Reconstruction of equivalent currents distribution over arbitrary three-dimensional surfaces based on integral equation algorithms," *IEEE Trans. Antennas Propag.*, Vol. 55, No. 12, 3460–3468, 2007.
27. Persson, K. and M. Gustafsson, "Reconstruction of equivalent currents using a near-field data transformation — With radome applications," *Progress In Electromagnetics Research*, Vol. 54, 179–198, 2005.
28. Eibert, T. F. and C. H. Schmidt, "Multilevel fast multipole accelerated inverse equivalent current method employing Rao-Wilton-Glisson discretization of electric and magnetic surface currents," *IEEE Trans. Antennas Propag.*, Vol. 57, No. 4, 1178–1185, 2009.
29. Jørgensen, E., P. Meincke, C. Cappellin, and M. Sabbadini, "Improved source reconstruction technique for antenna diagnostics," *Proc. 32nd ESA Antenna Workshop on Antennas for Space Applications*, Noordwijk, Netherlands, 2010.
30. Cano, F., M. Sierra-Castañer, S. Burgos, and J. L. Besada, "Applications of sources reconstruction techniques: Theory and practical results," *Proc. 4th Europ. Conf. Antennas Propag. (EuCAP)*, Barcelona, Spain, 2010.
31. Quijano, J. L. A. and G. Vecchi, "Field and source equivalence in source reconstruction on 3D surfaces," *Progress In Electromagnetics Research*, Vol. 103, 67–100, 2010.

32. Leonardo, J., J. L. A. Quijano, and G. Vecchi, "Removal of unwanted structural interactions from antenna measurements," *IEEE Antennas Propag. Soc. Int. Symp.*, 2009.
33. Balanis, C. A., *Advanced Engineering Electromagnetics*, John Wiley & Sons, New York, 1989.
34. Wang, H.-C. and K. Hwang, "Multicoloring of grid-structured PDE solvers on shared-memory multiprocessors," *IEEE Trans. Parallel Distrib. Syst.*, Vol. 6, No. 11, 1195–1205, 1995.
35. The OpenMP ARB, "OpenMP," 2004, Available online at: www.openmp.org.
36. NVIDIA Corporation, "NVIDIA CUDA C Programming Guide," 2012, Available online at: http://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf.
37. NVIDIA Corporation, "NVIDIA's next generation CUDA compute architecture: Fermi," 2009, Available online at: http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf.
38. NVIDIA Corporation, "NVIDIA GeForce GTX 680," 2012, Available online at: http://www.geforce.com/Active/en_US/en_US/pdf/GeForce-GTX-680-Whitepaper-FINAL.pdf.
39. NVIDIA Corporation, "CUDA Toolkit 4.2," 2012, Available online at: https://developer.nvidia.com/cuda-toolkit-42-archive.
40. Belkebir, K., A. Baussard, and D. Prémel, "Edge-preserving regularization scheme applied to the modified gradient method for the reconstruction of two-dimensional targets from laboratory-controlled data," *Progress In Electromagnetics Research*, Vol. 54, 1–17, 2005.
41. Belkebir, K. and M. Saillard, "Special section: Testing inversion algorithms against experimental data — Guest editors' introduction," *Inverse Problems*, Vol. 17, No. 6, 1565–1571, 2001, Available online at: http://iopscience.iop.org/0266-5611/17/6/301/media.