Efficient Stochastic FDTD Algorithm with Optimized GPU Acceleration

Athanasios N. Papadimopoulos, Georgios G. Pyrialakos, Nikolaos V. Kantartzis, and Theodoros D. Tsiboukis

Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
{papadian, pyrialak, kant, tsibukis}@auth.gr

Abstract — A 3-D curvilinear stochastic finite-difference time-domain (S-FDTD) technique on modern graphics processing units (GPUs) is introduced in this paper for complex media with high levels of statistically-variable heterogeneities. The novel accelerated methodology develops a robust covariant/contravariant dual-grid tessellation and estimates the mean value and standard deviation of field components during only a single run. In this way, notably accurate and stable estimations can be very rapidly and economically obtained, unlike the usual multiple-realization staircase Monte-Carlo FDTD schemes. These merits are successfully verified via realistic microwave setups with highly-varying media uncertainties, where the featured algorithm is shown to overwhelm the typical exceedingly resource consuming approaches.

Index Terms — Curvilinear coordinates, graphics processing units (GPUs), media uncertainties, Monte-Carlo schemes, statistical modeling, stochastic-FDTD method.

I. INTRODUCTION

Recently, an escalating research interest has arisen in the area of non-deterministic electromagnetic problems. Due to their abruptly-random media features, such applications are too complex for a conventional treatment; thus they are usually treated via the Monte-Carlo (MC) approach [1]. However, the large number of realizations and unduly resources actually prohibit its applicability to real-world 3-D (or even 2-D) arrangements. To mitigate these defects, several efficient alternatives have been proposed. Specifically, in [2] a stochastic finite-difference time-domain (S-FDTD) method is presented in Cartesian coordinates, while [3] follows a single-run approach for precise field statistics. Also, [4] launches an FDTD-based polynomial chaos expansion along with the proper basis functions and [5] derives a stochastic finite integration technique for electrokinetics. Despite the obvious profits, though, diverse issues are to be resolved, like the remarkable system demands and the manipulation of curved structures.

In this paper, a generalized S-FDTD method is developed via advanced graphics processing units (GPUs), for the accurate and fast analysis of electromagnetic problems with uncertainties in their constitutive and geometrical parameters. The new 3-D algorithm presents a covariant/contravariant metrics concept, based on a properly modified rendition of [6], for dual curved meshes to compute the mean value and standard deviation of electric and magnetic fields in just a single realization, unlike existing MC-FDTD techniques. For additional acceleration, the proposed method is programmed through the compute unified device architecture (CUDA) platform, which exploits the parallelized features of modern GPUs. Hence, very rapid and precise simulations are attained, whereas the detrimental lattice reflection errors, owing to staircase approximations, are drastically minimized. The prior advantages are substantiated via realistic microwave structures with complex curved parts, stochastic geometries as well as random electric permittivity, conductivity, and magnetic permeability. Numerical results reveal the accuracy and stability of the featured formulation, compared to the MC-FDTD outcomes, and its significant speedup over serialized implementations.

II. THE CURVILINEAR S-FDTD METHOD

The S-FDTD algorithm is derived for a general coordinate system, where we consider the covariant \( a_1 \), \( a_2 \), \( a_3 \) and contravariant \( a^1 \), \( a^2 \), \( a^3 \) bases for the electric, \( \mathbf{E} \), and magnetic field, \( \mathbf{H} \), vectors [6, 7]. This implies that in our formulation, there are four sets of components: (i) two covariant, i.e., \( (e_1, e_2, e_3) \), \( (h_1, h_2, h_3) \) and (ii) two contravariant i.e., \( (e^1, e^2, e^3) \), \( (h^1, h^2, h^3) \), for the \( \mathbf{E} \) and \( \mathbf{H} \) field components, respectively. In addition, for the case of unbounded computational domains, the convolution perfectly matched layer (CPML) absorbing boundary condition [8] is employed. Bearing in mind the above considerations, we apply operator \( \mathcal{L} \) (representing the mean value \( M \) or variance \( \sigma^2 \)) to Maxwell equations. For illustration, the \( e_1 \) covariant component is given by:
\[ L\{ e^{\text{up}1}_{i,j,k} \} = L\{ D_h e^{\text{up}1}_{i,j,k} \} + \frac{D_a}{\sqrt{g \Delta t}} \left( h^{\text{up}1/2}_{i,j,k} - h^{\text{down}1/2}_{i,j,k} + \zeta \right) - \frac{D_b}{\sqrt{g \Delta t}} \left( h^{\text{up}1/2}_{i,j,k} - h^{\text{up}3/2}_{i,j,k} + \zeta \right), \]

then the covariance in (1) - (3), we deduce from (6) that variance terms include covariance components of two random variables, given by \( \text{Cov}(y_1, y_2) = \rho \sigma_{y_1} \sigma_{y_2} \), for \( \rho \) the correlation coefficient. In fact, \( \rho \) varies between 0 and 1, with values near unity revealing a strong correlation. Observing the covariances in (1)-(3), we may derive that \( \rho \) must be practically equal to unity. This is evident for field components and \( \zeta \) terms, which are highly correlated, as they exist at very proximate time intervals. Thus, an instructive linear expansion of these terms, i.e., \( \sigma(y_1 + y_2) = \sqrt{\sigma(y_1)^2 + \sigma(y_2)^2} \) is extracted. In this paper, the standard deviation of all stochastic parameters does not exceed the 10% of the threshold which is deemed very realistic for our simulations. In this context, we acquire (6) for the standard deviation of \( e_t \) (bottom of page) and

\[ M\{ f(y_1, y_2, \ldots, y_n) \} \approx f(m_y, m_y, \ldots, m_y), \quad \sigma^2 \{ f(y_1, y_2, \ldots, y_n) \} \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f}{\partial y_j} \frac{\partial f}{\partial y_i} \text{Cov}(y_i, y_j), \]

for \( m_y, m_y, \ldots, m_y \) the mean values of \( y_1, y_2, \ldots, y_n \), which, herein, are the \( e_t, h_t, e_t, h_t \) components and the four media parameters, i.e., electric permittivity \( \varepsilon \), magnetic permeability \( \mu \), conductivity \( \sigma \), and magnetic losses \( \bar{\rho} \).
\[
\sigma \left( \zeta_{rs}^{m+1/2} \right) = R \sigma \left( \zeta_{rs}^{m-1/2} \right) + S \left[ \sigma \left( h^{m+1/2} \right) - \sigma \left( h^{m-1/2} \right) \right], \tag{8}
\]

with similar formulas for the other quantities. In (6) and (7), \( m \), \( m \sigma_\varepsilon \), \( \sigma_\varepsilon \), \( \sigma_\sigma \) are the mean value and standard deviation of \( \varepsilon \) and \( \sigma \), while \( \rho_{\varepsilon,\varepsilon}, \rho_{\sigma,\sigma} \) are the cross-correlation coefficients between component \( P \) (here \( P = h^3, h^3, \zeta_{rs}^3, \zeta_{rs} \sigma \)) and material parameter. Moreover, (7) shows that to obtain \( \sigma \), one has to evaluate the mean value of \( E \) and \( H \) components via (4) at the previous time-step. This does not permit the extraction of either the mean value or standard deviation alone, due to the coupling of (4) and (6). So, the formulation requires twice the usual FDTD system resources. In contrast, the MC-FDTD scheme focuses on the computed fields, whose uncertainties are found by samples over the total number of simulations. As a result, its overhead is analogous to this (usually very large) number, yet obligatory for a viable approximation of the ideal normal distribution. Finally, since the typical Courant criterion in Cartesian coordinates is affected by the curvilinear formalism, the stability throughout our paper is guaranteed by the generalized Courant condition [7]:

\[
\Delta t \leq c \sup \left( \sqrt{\sum_{i} \sum_{j} \sum_{k} g_{ij}^{n-1}} \right), \tag{9}
\]

pertinent to handle non-uniform lattices, as the ones implemented herein. In (9), \( \sup(.) \) denotes the maximum value in the entire \( i,j,k \) domain and \( g_{ij}^{n} \) is a metric that is the inverse of \( g_{ij} \). Notice that all of our studies have been proven completely stable, even for very detailed grids or very long simulation times.

III. OPTIMAL GPU IMPLEMENTATION

For our curvilinear simulations, we develop the full 3-D code based on the CUDA 6.0 platform [11]-[13] to exploit modern GPUs. This decision is favored from the parallelization potential of the S-FDTD method, which offers substantial acceleration. Our algorithm allows the fully independent execution of update equations at each grid node during a single time-step. Thus, we assign one or more such nodes to the various independent execution flows of the hardware, to get the maximum performance via a comprehensive optimization process that will be discussed. In CUDA independent processes called threads, are arranged in an algorithmic 3-D “grid”. Such a structure provides these threads with unique coordinates and allows the manipulation of the actual space coordinates in the domain. To this aim, we assign specific memory addresses from our 3-D matrix (i.e., the electromagnetic field space) to specific thread coordinates, for the entire space and connect nearby memory addresses with equally adjacent threads.

Emphasis should be given to the handling of wraps, groupings of 32 threads that although, by their nature, perform 32 individual parallel executions, they are issuing a common command for all the involved execution flows at the same time. This leads to inefficient algorithms, where incorporation of branching below a 32 thread interval lead to serialization in the execution flow. Threads are also organized into larger structures called blocks that may have up to three dimensions in a local grid of thread coordinates. All wraps that reside inside each block are assigned to the same streaming multiprocessor (SM) in the hardware; so making use of the same local memories (shared memory, L1 cache, etc.) and the same schedulers. Lastly, the concept of “kernel”, a CUDA function executed on the GPU, must be considered. Through kernels we can create our grid and block structure, control the hardware, and use the different local memories available per block. A flow chart of our algorithm is given in Fig. 1, while its key realization features are summarized as follows.

**A. Kernel size and register usage**

According to our analysis, a grouping of 32×16 threads-per-block has been found to guarantee very good performance in Cartesian grids. The complexity of the curvilinear S-FDTD algorithm did, however, require extensive treatment of kernel variables, yielding an optimal 24×16 grid. Although in the simple case of fully orthogonal meshes, the use of one kernel for each of the FDTD updates suffices, we observed that a multi-kernel implementation (exploiting streams as discussed later) is more beneficial for the curvilinear S-FDTD algorithm. Regarding the surrounding CPML layers, four different kernels for each side (two for electric and two for magnetic components), running simultaneously via streams, are necessary for the additional calculations (due to the extra CPML terms) to be completed.
B. Use of streams

Aiming at the highest parallelization, we resort – at various points – to the use of streams. They refer to independent flow sequences, defined outside a kernel and passed as an argument when the latter is called. Also, they allow kernels, likely not to interfere or share cross-dependences, to be executed concurrently: something impossible in their default state, so avoiding unnecessary serializations. Thanks to them, we can partially achieve the simultaneous execution of the main routine and the CPML kernels, resulting in about 72% less overall computational time, as illustrated in Fig. 2.

![Single- and multiple-kernel implementations](image)

Fig. 2. Efficiency differences between single- and multiple-kernel realizations for the CPML update.

C. Memory considerations

In our realization, the global memory is used for the storage of the main components and the CPML variables. Special attention has been paid to the suitable matrix alignment in memory, which ensures that adjacent threads in the kernels access similarly-placed elements. Only when this occurs, transfers of 32 elements from global memory (in the case of floats) are conducted in a single memory access cycle (i.e., “coalesced access”). For the CPML, specific grid alignments are selected, because a simplistic implementation of a unique kernel for all CPML areas would render coalesced access unfeasible. This is due to the use of a small grid along the perpendicular (with respect to the CPML) directions, which does not allow the assign of the first grid dimension to the first dimension of CPML auxiliary field matrices. However, the latter action is required for memory coalescing. A simple, yet inefficient, solution could use a full size grid, which would introduce large numbers of idle threads. So, the division of the algorithm into individual parts with different grids, that exploit streams, appears as the optimal choice.

D. Use of atomics

 Atomics are necessary for the simultaneous update of elements at the grid corners. Figure 3 clarifies the correction provided in our algorithm by this idea, when the race between concurrent threads is likely to produce undesirable miscalculations (the case of top/bottom CPMLs is shown in Fig. 3). So, correct updating is ensured without causing any performance degradation.

IV. NUMERICAL RESULTS

The performance of the proposed algorithm is validated in terms of complicated microwave structures with real-world stochastic media parameters. Such uncertainties are normally encountered at their substrate, copper parts and geometry, possibly created during the manufacturing process. Furthermore, comparisons are conducted with an MC-FDTD scheme, which needs the large number of almost $10^4$ FDTD realizations, whereas all infinite domains are terminated by an 8-cell CPML.

A. Koch-shaped fractal microstrip

Let us study the Koch fractal microstrip filter of Fig. 4, which suppresses particular frequencies with a reduced size. Its dimensions are given in [14] and all statistical constitutive/geometric parameters of its copper parts and substrate are obtained from the Ventec™ and Isola™ Corporation datasheets. Therefore, for the substrate: $m_s = 4.4$, $\sigma(\varepsilon) = 0.088$, $m_p = 3.427$ mS/m, $\sigma(\overline{\sigma}) = 0.07$ mS/m, while for its height $h$: $m_h = 15$ mm and $\sigma(h) = 0.3$ mm. On the other hand, for the copper parts, we have $m_p = 58.6$ S/m and $\sigma(\overline{\sigma}) = 0.17$ S/m. Figures 5 and 6 present the $E_x$ mean value and standard deviation, as derived from the MC-FDTD technique (mesh: 203x407x117 cells and total CPU computational time: 62.57 hours) and our algorithm (mesh: 101x201x51 cells and total CPU computational time: 22.92 min). Note that all calculations are performed across a straight horizontal line at the center of filter’s top side. As discerned, the agreement between the two methods is promising, whereas the convergence of the standard deviation outcomes (Fig. 6) is very satisfactory. However, it is stated that as $\rho \to 1$, the S-FDTD technique tends to overestimate the solution, as also discussed in [2], [3]. Furthermore, in the inlet snapshot of Fig. 5, we illustrate the mean value of the current distribution atop the filter, while Fig. 7 summarizes the significant GPU acceleration of our algorithm for various lattices via two NVIDIA™ models. Similar deductions can be drawn for the $S$-parameters of Fig. 8.
B. Compact wideband resonator filter

The second application is the compact wideband filter of Fig. 9, with a folded multiple-mode resonator [15]. For the same stochastic constitutive/geometric parameters as in the previous problem, the basic dimensions are selected as: \( L_1 = 18.8 \) mm, \( L_2 = 5.8 \) mm, \( L_3 = 9 \) mm, \( L_4 = 5 \) mm, \( w_1 = 2.27 \) mm, \( w_2 = 5.5 \) mm, \( w_3 = 0.4 \) mm, \( g_1 = 1 \) mm, \( g_2 = 0.4 \) mm, \( g_3 = 0.21 \) mm, and \( H = 4 \) mm. This realistic setup (mesh: \( 273 \times 10^7 \times 49 \) cells and total CPU computational time: 36.27 min) leads to the single-run current distribution mean value of Fig. 10, at the filter’s top side, which is very smooth, unlike the one of the MC-FDTD method (mesh: \( 511 \times 263 \times 105 \) cells and total CPU computational time: 71.34 hours) through an extremely large number of multiple realizations. Finally, the accuracy of our algorithm are successfully substantiated through the \( S \)-parameters of Fig. 11, again evaluated during only a single realization.
V. CONCLUSION

A 3-D curvilinear S-FDTD method has been presented in this paper for statistically heterogeneous materials regarding their constitutive/geometric parameters. The new scheme launches a covariant/contravariant concept, while its single-run GPU rendition is more than 100× faster than existing techniques. So, accurate outcomes can be very rapidly and cost-effectively derived.

ACKNOWLEDGMENT

This research has been co-financed by the EU and Greek funds through Research Program: Aristeia.

REFERENCES


Athanasios N. Papadimopoulos is pursuing his Ph.D. degree in the Aristotle University of Thessaloniki, Greece, since 2014. His research focuses on numerical techniques, graphene forms and semi-analytical Green’s function methods.

Georgios G. Pyrialakos is pursuing his Ph.D. degree in the Aristotle University of Thessaloniki, Greece, since 2013. His research aims at computational methods, hardware acceleration, advanced microwave components and metamaterials.
Nikolaos V. Kantartzis serves as an Associate Professor in the Aristotle University of Thessaloniki, Greece, since 2010. His research interests include EMC modeling, numerical methods, graphene, metamaterials, contemporary microwave applications, and enhanced antenna systems.

Theodoros D. Tsiboukis is a Professor Emeritus in the Aristotle University of Thessaloniki, Greece. His interests involve computational electromagnetics, energy methods, EMC applications, metamaterials, antennas, and waveguides. Tsiboukis has been the recipient of several awards and is a member of various societies and chambers.