Simulation-Based Optimization – Myths and Reality

Przemysław Miazga [1]

[1] Faculty of Electronics and Information Technology,
Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland
p.miazga@ire.pw.edu.pl

Abstract: The present paper discusses the influence of simulation model accuracy on the convergence of electromagnetic structure simulation-based optimization. Neither response surface approximation method nor the algorithm of moving window filtering, commonly used for simulation error compensation, is not fully capable of guaranteeing proper convergence. The non-expensive device model with coarse meshing and a modified error compensation method can yield satisfactory results in a reasonable time.

1 Introduction

The simulation-based optimization method merges optimization techniques with full-wave 3D/V2D electromagnetic simulation technologies. The whole process is usually extremely time- and resource- consuming. The simulators also have limits of credibility, resulting from the model behavior and computer implementation. Prior to the optimization we are forced to make a tradeoff between reliability and accuracy of solution versus time and number of computer resources necessary to complete the project. The designing process with the use of a simulator is a struggle for a better performance of the device in acceptable time, with a limited number of computer resources. In practice, accurate simulation models with dense meshing, although extremely expensive, are more likely used. Simulation-based optimization has achieved growing interest in recent times [1,2]. Many computer packages for the design of microwave circuits, called electromagnetic solvers, are available on the market [1-4].

This paper is focused on those, using the finite-difference time-domain method FDTD [5]. In principle, the FDTD method is based on solving three-dimensional Maxwell equations with space (called meshing) and time discretisation. In the space domain, the considered area is decomposed into a set of small cuboids cells (the process called meshing). In the time domain, electromagnetic fields are evaluated iteratively with a time interval. The discrete and iterative character of the FDTD method causes some limits on the performance of the design process [6]. Simulation accuracy depends on the finite cell size and evaluation time. The cell size reduction by half may decrease the systematic space error, caused by discretisation, about 4 times. However, the computation time must grow almost 16 times, while the computer memory grows about 4 times. Also the location of meshing planes plays an important role.

Every circuit is composed of various materials with various shapes. In theory, meshing should match exactly the boundaries of each region. For some class of circuits (those with parallel-to-meshing boundaries) it is possible to compensate the meshing error, by applying various mesh snapping techniques [6]. The choice of meshing is a critical issue for timing and convergence [6,7]. Additionally, some problems are caused by data accuracy. Internal computer data are often coded as 32-bit floating numbers, which limits the result accuracy to four significant digits. What is more, input data are rounded, typically, to 4-6 significant digits. So the 0.01% change of an input variable may not affect the simulation result. The accuracy of float arithmetic, with an iterative character of the process, and input data rounding, seems to be not sufficient enough to guarantee the proper optimization convergence.

The present paper is organized as follows. Chapter II presents the extended results of a study on the influence of simulation space and meshing errors. Chapter III presents an improved moving window data (directional) filtering method to optimize the electromagnetic structure.

2 The Waveguide Transformer Case

The following example illustrates an influence of meshing and numerical problems on simulation optimization behavior. A simple rectangular waveguide impedance transformer is considered, with mesh snapping switched off (shown in Fig.1). The waveguide is built from a rectangular metal pipe, filled with air. The transformer, shown in Fig.1, consists of three waveguide sections: an input port (a=23mm, b=10mm), middle section (a=23mm, b=7.8mm) with a varying length l, and output port (a=23mm,b=6mm).

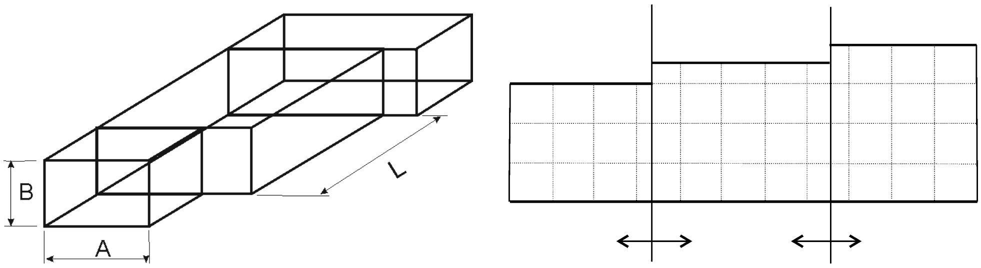

Figure 1 Rectangular waveguide impedance transformer and its cross-section with meshing

The maximal value of the modulus of input reflection coefficient S_{11} in the frequency band from 8 to 9 GHz is used as an objective function. The optimization problem is stated as follows:

$$\min_{x \in D_n} \max_{f_i \in \langle f_{min}, \ldots, f_{max} \rangle} \left| S_{11}(x, f_i) \right| \tag{1}$$

where: x is a vector of optimized parameters ,
D_n is an n-dimensional problem domain,
f_i is a discretized frequency from <fmin,fmax> range,
S_{11} denotes a reflection coefficient (matching factor) on the input of the device.

The optimal value of the varying parameter - length l equals, approximately, 9.5 mm. With a change of the optimized parameter, the border plane, located on the step-in-width of the waveguide, is shifted. The upper border cell is then, more or less, filled with air, while the rest is assumed to be a metal. The FDTD method is based on the assumption that each cell is filled with a uniform medium. Some methods are used for compensation of the non-uniform meshing [6]. However, they are not fully capable of eliminating all associated phenomena and assuring a proper convergence of the optimization process [7].

The circuit was simulated with an FD-TD simulator [6], with four different cell sizes: 2, 1, 0.5 and 0.25 mm. The simulations results are shown in Fig.2. The theoretical shape of the objective is smooth and convex. The shape of the simulator evaluated objective, for the reference (cell size equal 0.25 mm) meshing, also seems to be smooth and convex, with one optimum. With a growing cell size the ripples on the curve are growing and the minimum of the function shifts sideways. Also the number of local optima is growing. One may expect that, more or less probably, the optimization process will end up at one of the local basins of attraction.

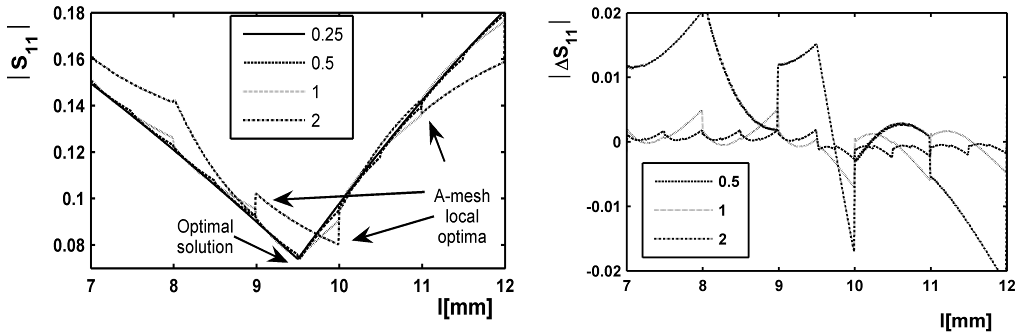

Figure 2 Plot of the objective function (left figure) and simulation error (right figure) versus the length of the matching section.

The systematic simulation error, shown in Fig.2, constitutes a quasi-periodic function with the period equal to the cell size, specified lengthwise the matching section, of a triangular shape. The phenomenon of the cell size dependent systematic error is caused by the fact that the boundaries of circuit regions do not coincide with the cell boundaries. Lets go further into detail. The shape of the objective curve in a neighborhood closer to the optimal solution is shown in Fig.3.

Even on the objective curve with higher simulation accuracy, some small ripples are observed. These ripples are relatively narrow and probability of falling into a “hole” during optimization is relatively small. However, they may cause premature stops of the process, even with very precise meshing. Another phenomena - a stepped like noise which is best visible on the upper (cell size equal 1mm) curve - seems to be caused by numerical errors, namely, the use of the float data type and input data rounding. Also some drift of the optima with a change of the mesh size can be observed.

Most of the response surface algorithms [7,10,11] preserves higher moments. The consequences of this phenomena is illustrated in Fig.3. The simulation points, marked as black dots, were chosen randomly. The approximation error, even for a large number of samples, is comparable with the objective function. Combined with the simulation error, it causes an appearance of some new parasitic local optima. An algorithm to eliminate the notches of the objective function is welcome, otherwise the convergence of the optimization procedure cannot be guaranteed. The exploitation of fuzzy systems [12] may reduce the sensitivity of the optimization process due to the simulation error, but at a high cost

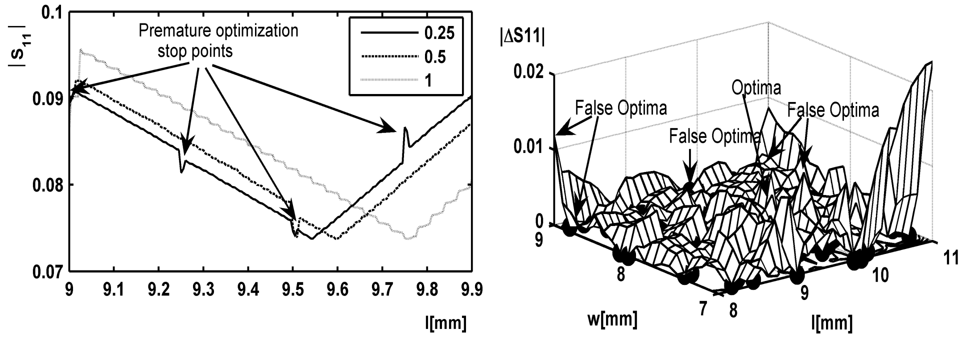

Figure 3 Magnified plot of the objective function (left figure) versus length of the matching section and error chart of the response surface approximation (right figure), evaluated for the cell size equal to 0.5 mm and 80 simulation samples, versus the length and width of the matching section.

For a multi-modal objective function and lots of optimization parameters, space sampling should be dense enough in order to reduce the approximation error. This raises the cost of the design to an unacceptably high level. With a growing number of samples the execution time increases exponentially. Recently published examples [12] deal with 2-3 optimization variables, and therefore are capable of fine-tuning the network. You can forget about solving multiple-variable global optimization in reasonable time.

The amplitude of the meshing error decreases with increased simulation accuracy. However model improvement causes an unreasonable growth of execution time and greater need for more operational memory. Also with a growing amount of necessary computer resources the simulation process is more likely to hang up, crash or become unstable, due to hardware problems. So in practice, we may never be certain about the simulation result. Anyway, from the numerical point of view, 0.25 mm meshing seems to be an acceptable choice for structure optimization. However, the required computer memory is, approximately, 64 times greater than for the 2-mm meshing, while the simulation time grows about 4000 times. In practice, the programming overhead (input/output processing system procedures), and the limited bandwidth of memory transfer reduces this ratio several times. But still the opportunity for fast exploration - hundreds or even thousands of rough characteristic evaluations in the same time as the very precise one, is very encouraging.

3 Simulation Error Compensation

With a large number of simulation samples available, we may try to filter the parasitic periodic peak out of the shape of the response function. Smoothing the function must flatten the ripples, creating a convex function with one optimum. The commonly used way of objective function smoothing is called “moving window filtering” [10]. For uniform approach, the search space is discretized into a regular grid and the simulation result is evaluated for every node of the grid. Every grid point constitutes the center point of a window used for filtering. Finally, the filtered objective value is evaluated as the average of values of all points located inside the window, and the minimum is found. Therefore with a growing dimensionality and sampling rate the cost increases exponentially. The filtration range and quality of smoothing is determined by the window size and grid density. Finally, the algorithm assumes that simulation results are corrupted by random variations. But in general, the simulator specific error is deterministic. So the method will not work as we expect. First, filtering accurate and "quite" good models is too expensive and therefore impractical. The peaks of the objective function are so narrow that with any global optimization method we can quite easily escape from their neighborhood. Moving window filtering may help only for a coarse model when the meshing error is high, like one with the 2 mm cell size. Surprisingly, a given accuracy is assumed as the best compromise between result accuracy and evaluation time, from a point of view of field theory specialists. The results of the application of moving window filtering (curve F) with rough meshing (cell size = 2 mm) are presented in Figure 4. It should be mentioned that in the aforementioned example the optimized variable vector is orthogonal to the meshing plane. In general, moving window filtering method applied to the media with more complicated shapes, like e.g. circular resonator, behaves differently and the considered phenomena need further investigation.

For a good sampling rate the optimum of a filtered function is located very close to the "real" optimal point. The shift of optima with 2 mm mesh is better than 0.8%. The calculation time of the whole process is, approximately, the same as that of the simulation with the 0.5 mm mesh. It must be also emphasized that the accuracy of optima location for the 0.5 mm cell size (with no filtering) is about 1.2%. To achieve the same result with any other optimization method, at least, 10-20 objective evaluation is required. So the estimated convergence improvement is of about ten times. The simulation cost, thanks to coarse meshing, is relatively low.

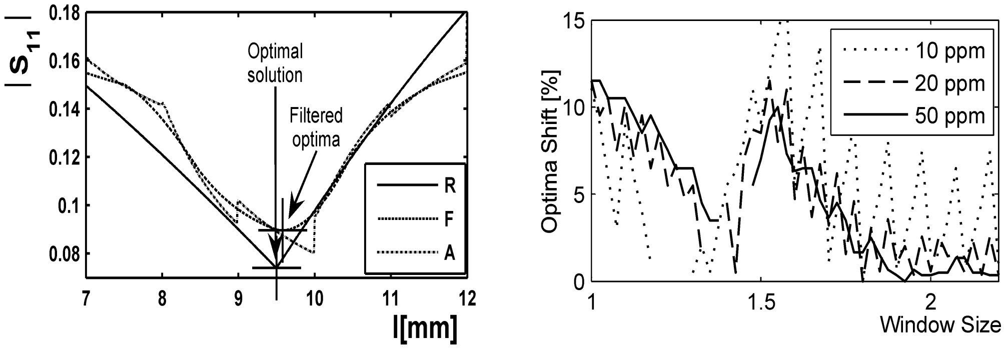

Figure 4 Results of "moving window" filtering: objective function - left figure, and accuracy of optima localization versus the size of moving window (in millimeters) for various samples (ppm – points per mesh size) – right figure.

The performance chart, shown in Fig. 4, illustrates that sufficient accuracy can be achieved only for relatively dense probing. More than 20 points are necessary to locate optima with ±2% accuracy. In one-dimensional case, 20 evaluations is an acceptable cost. But it makes 400 for 2 variables and 8000 for 3 variables. So by increasing dimensionality, we are losing all the profit resulting from coarse meshing. Also in a multi-dimensional case we must keep in mind that the simulation error depends on the orientation of an optimized variable vector according to the meshing plane. When some vectors are parallel, the application of moving window filtering may lead to a misleading result as the simulation errors are correlated.

Anyway, the method was also tested in a two-dimensional convex case, with varying length l and width w of the matching section. These vectors are orthogonal, so their simulation errors are non-correlated. From the initial number of 6 local optima, only one global remains after filtering. The shift of optima location was about 2.5%, which is sufficient enough in the considered case, but worse than before. The only advantage of moving window filtering in a multi-dimensional case stems from the fact that the objective function after filtering is smooth, so we can effectively use direct search methods [7,9] for local exploitation. Concluding, the application of multi-dimensional moving window filtering is neither safe nor efficient. The only reasonable solution is to use one-dimensional moving window algorithm, called later, directional filtering. The filtering along the axes of the problem domain space is the most suitable choice for the steepest descent Gauss-Seidel algorithm. It may overcome a commonly used asynchronous parallel pattern search algorithm [8]. We can also apply other non-gradient algorithms and evaluate the filtered objective in line search procedure along the current direction.

What about global optimization? Very low single evaluation cost is a promising and encouraging factor. Evolutionary algorithms are implemented in a couple wave simulation solvers [13]. However, the efficiency of the EA is poor due to high single evaluation cost. The application of directional filtering changes the situation dramatically. Within the same time we are able to generate ten to hundred times solutions more. Some tests confirm this thesis. But a complete problem investigation, with still an expensive objective function, requires a lot of computation time. The case is currently analyzed.

The distribution of a directional filtering algorithm is easy and effective. It can be realized via asynchronous parallel simulations. With a large number of “real simulation” samples the method is not sensitive to the uncertainty about the simulation result, the loss of data or delay in data transmission evaluated from some sample points. So it can even be safely used on a computer grid with various processors.

4 Conclusions

Many heuristic techniques have been evaluated by experimenting with test cases. Several efforts have been made to prove the convergence of such, intuitively created methods. Few of them were successful. Anyway, the proof of convergence is not necessary for efficient functioning of the algorithm. Directional filtering of simulation data seems to perform faster and be more reliable than the response surface algorithm, particularly in the case of the QuickWave FDTD simulator. The phenomena of the space error under consideration is common for all finite-difference simulators. So, intuitively thinking, we may expect that the method can perform well, in general. The problem needs further investigation though. Simulation result redundancy not only lets us explore a wider space range, increasing the probability of finding a new and better solution but it also enables us to neglect the uncertainty and unreliability of the simulation as well. The filtering method is simple and easy to implement with an evolutionary algorithm and ideal for concurrent asynchronous processing.

References

[1] Y. Carson, A. Maria, “Simulation optimization: methods and applications”, Proc. of 1997 Winter Simulation Conference, pp.118-126.

[2] P.Miazga, W.Gwarek, “Improved design of passive coaxial components using electromagnetic 2-D solver in an optimization loop”, IEEE Trans on MTT, vol 45, May, 1997, pp. 858—860.

[3] J.W.Bandler et al., “Electromagnetic optimization of 3D structures”, IEEE Trans on MTT, vol 45, May 1997, pp. 770—779.

[4] P. Miazga, M. Kamola, "Global and local optimization algorithms in automated waveguide design", in Evolutionary Computation And Global Optimization, Prace Naukowe Politechniki Warszawskiej, Elektronika, Vol. 139, pp. 123-135, Oficyna Wydawnicza Politechniki Warszawskiej, 2002

[5] A.Taflove, “Computational electrodynamics: The Finite-Difference Time-Domain Method”, Boston-London: Artech House, 1995.

[6] “QuickWave 3D electromagnetic simulator users manual”, QWED Ltd. Warszawa

[7] P. Miazga, “Simulation optimization of electromagnetic structure–case study”, Proc. of 16th Int. Conf. on Microwaves, Radar and Wireless Communications, MIKON-2006, Kraków, Poland, May 2006, pp.820-823.

[8] T.K.Kolda, V.J.Torczon, “Understanding asynchronous parallel pattern search” in “High Performance Algorithms and Software for Nonlinear Optimization”, G.Di Pillo and A.Murli, Editors, 2003 Kluwer Academic Publishers B.V. pp. 316-335.

[9] P.Miazga, “Users Manual of QW-Optimizer”, QWED Ltd. Warszawa, 2005.

[10] D.R.Jones, “A Taxonomy of Global Optimization Methods Based on Response Surfaces”, Journal of Global Optimization 21, 2001, pp.345-383.

[11] P.Miazga, “Global optimization of expensive simulator-evaluated function based on response surface interpolation”, Proc. of 7th Conf. Evolutionary Algorithms and Global Optimization, Korbielów, May 2005, pp.151-156.

[12] S.Koziel, J.W.Bandler, “A Space-Mapping Approach to Microwave Device Modeling Exploiting Fuzzy Systems”, IEEE Trans on MTT, vol 55, Dec. 2007, pp. 2539-2547.

[13] J.W.Bandler, Q.S.Cheng, D.M.Hailu, N.K.Nikolova, “A Space-Mapping Design Framework”, IEEE Trans on MTT, vol 55, Nov. 2004, pp. 2601-2610.