A Fast Filter Optimization Scheme Based On Model Order Reduction

Klaus Krohne and Rüdiger Vahldieck

Swiss Federal Institute of Technology, Laboratory for Electromagnetic Fields and Microwave Electronics (IFH), Gloriastrasse 35, CH-8092, Zurich, Switzerland
kkrohne@ifh.ee.ethz.ch

Abstract — **This paper presents a novel, fast optimization technique that is especially suitable for the design of microwave filters. For the discretization of the structure, the well established Finite Integration Technique (FIT) is used. A recently published two step Model Order Reduction (MORe) technique computes the poles and residuals of the structure's impedance matrix as well as its sensitivity to the geometry parameters. The actual optimization is then performed in the reduced order space of the poles and residuals. This significantly reduces the number of full-wave simulation runs, that is usually associated with the optimization of filters.**

I. Introduction

The optimization of electromagnetic structures has received significant attention in recent years. This is seen in conjunction with the widespread availability of commercial EM simulators which have become an indispensable tool in the design of RF components and subsystems. However, optimizing or fine-tuning a given structure is often prohibited by the relatively long computation time of general-purpose EM simulators. This is true in particular for more complex structures with many variable geometry parameters, which require a large number of optimization steps.

The situation is impaired when highly resonant structures, like filters, are to be optimized, as they impose additional difficulties on the field simulators. Time domain codes require long time iterations to reach steady state and frequency domain codes suffer from reduced performance, as a fine resolution of the frequency axis is necessary, in order to capture sharp resonances, which can result in excessively long computation times. Thus, using general-purpose EM simulators directly in combination with optimization routines is difficult in general.

A possible solution to overcome the latter difficulty of long computation times is the use of the Model Order Reduction (MORe) technique. In this approach the poles and residuals of a structure are computed directly, which is particularly well suited for highly resonant structures. The reduced number of poles and residuals are expressed as a linearized function of the geometry parameters, with the help of a sensitivity analysis. The actual optimization is performed on the reduced order model. To ensure the validity of the reduced order model, the model itself, as well as its sensitivity with regard to the geometry parameters, is updated after each optimization step. This results in $(n_{\mathbf{x}}+1)\cdot p+1$ full-wave simulator runs for $n_{\mathbf{x}}$ geometry parameters and p optimization steps.

In previous work an alternative technique has been employed successfully. This technique is based on the use of a surrogate model to optimize filter structures. Also, here it has been shown that the number of necessary simulator runs can be significantly reduced if the electromagnetic filter structure is represented by a minimum prototype filter network [1]. The elements of this prototype are determined by matching its transfer function to that of the EM simulated structure. Geometry optimization is then performed on the network parameters with occasional model updates through EM simulation. A disadvantage of this approach is, however, that the surrogate model of the structure has to be determined in advance, which may not always be possible and, therefore, the approach in [1] lacks universality.

The Model Order Reduction technique presented in the following appears to be a very promising approach since it is general and, at the same time, computationally efficient. A combination of this method and the method in [1] is currently being investigated. The method requires the electromagnetic system to be described in its state-space. This is achieved with the help of the Finite Integration Technique (FIT, [2]); the procedure is briefly explained in section II. In section III. the MORe is explained in detail. Sections IV. and V. deal with the optimization being performed on the reduced parameter space of the poles and residuals. Finally, the efficiency of the method is demonstrated by optimizing a waveguide filter.

II. State-Space Representation

In the Finite Integration Technique, the first two Maxwell equations are discretized on two dual rectangular grids. With the help of the discrete *curl* operator $\mathbf{C}$, they are represented in the discrete domain according to [3] as

$$\mathbf{M}_{\varepsilon}\dot{\widehat{\mathbf{e}}} = -\mathbf{M}_{\kappa}\widehat{\mathbf{e}} + \mathbf{C}^{T}\widehat{\mathbf{h}} - \widehat{\mathbf{j}} \quad (1)$$

$$\mathbf{M}_{\mu}\dot{\widehat{\mathbf{h}}} = -\mathbf{C}\widehat{\mathbf{e}} \quad (2)$$

0-7803-7695-1/03/$17.00 © 2003 IEEE

with the electric and magnetic grid voltages $\widehat{\mathbf{e}}$ and $\widehat{\mathbf{h}}$, respectively, as well as the diagonal material matrices $\mathbf{M}_\varepsilon$, $\mathbf{M}_\mu$ and $\mathbf{M}_\kappa$. For a detailed introduction to the Finite Integration Technique see [2].

Expressing (1) and (2) in a single matrix equation directly yields the system's linear state-space representation

$$s\mathbf{X} = \mathbf{A}_1\mathbf{X} + \mathbf{R}_1\mathbf{i} \quad (3)$$

$$\mathbf{u} = \mathbf{L}_1\mathbf{X} \quad (4)$$

with the state vector $\mathbf{X} = [\ \widehat{\mathbf{E}}^T \quad \widehat{\mathbf{H}}^T\ ]^T$, where $\widehat{\mathbf{E}}$ and $\widehat{\mathbf{H}}$ denote the grid voltages' Laplace transforms and

$$\mathbf{A}_1 = -\begin{bmatrix} \mathbf{M}_\varepsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_\mu \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{M}_\kappa & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}.$$

The port operators $\mathbf{R}_1$ and $\mathbf{L}_1$ translate the port currents $\mathbf{i}$ and the port voltages $\mathbf{u}$ into grid currents and voltages, respectively. They are usually constructed by a 2D eigenvalue solution of the port region. For an m-port structure that is discretized with N mesh-nodes, $\mathbf{A}_1$ is $6N \times 6N$, $\mathbf{R}_1$ is $6N \times m$ and $\mathbf{L}_1$ is $m \times 6N$. Equations (3) and (4) allow easy expression of the system's impedance matrix

$$\mathbf{Z}(s) = \mathbf{L}_1 (s\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{R}_1 \quad (5)$$

by solving (3) for $\mathbf{X}$ and placing it into (4).

If the system is loss-free, a *curl-curl* formulation, presented in [4], can be applied. The result is a second degree system

$$s^2\mathbf{X} = \mathbf{A}_2\mathbf{X} + s\mathbf{R}_2\mathbf{i} \quad (6)$$

$$\mathbf{u} = \mathbf{R}_2^T\mathbf{X}, \quad (7)$$

halving the number of state variables, as only the electric grid voltages are relevant. The matrix $\mathbf{A}_2$ becomes real symmetric and is therefore at least semi definite, allowing the proof of stability and passivity [5]. In this case, the impedance matrix results in

$$\mathbf{Z}(s) = s\mathbf{R}_2^T \left(s^2\mathbf{I} - \mathbf{A}_2\right)^{-1} \mathbf{R}_2, \quad (8)$$

which can be brought into the canonical form

$$\mathbf{Z}(s) = s\mathbf{R}_2^T\mathbf{E} \left(s^2\mathbf{I} - \mathbf{P}\right)^{-1} \mathbf{E}^{-1}\mathbf{R}_2 \quad (9)$$

with the help of the eigenvalue decomposition $\mathbf{A}_2\mathbf{E} = \mathbf{E}\mathbf{P}$, where the columns of $\mathbf{E}$ are the eigenvectors of $\mathbf{A}_2$, and $\mathbf{P}$ is a diagonal matrix, the elements of which are the quadratic system poles. In (9), the inverse can then be computed easily, making it possible to express the impedance matrix in terms of the system poles and residuals.

$$\mathbf{Z}(s) = \sum_{k=1}^{n} \frac{s}{s^2 - p_k} \begin{bmatrix} r_{11k} & \cdots & r_{1mk} \\ \vdots & \ddots & \vdots \\ r_{m1k} & \cdots & r_{mmk} \end{bmatrix} \quad (10)$$

with the quadratic poles p_k and the residuals $r_{ijk} = \left[\mathbf{R}_2^T\mathbf{E}\right]_{ik} \left[\mathbf{E}^{-1}\mathbf{R}_2\right]_{kj}$. It is obvious that $p_k < 0$ has to yield for the system to be stable, as in this case, the poles become conjugate imaginary.

III. Model Order Reduction

The System in (6) and (7) can be transformed into an equivalent system with the help of a state-space coordinate transformation by any non singular real matrix $\mathbf{V}$ by simply replacing $\mathbf{X} = \mathbf{V}\mathbf{X}'$. If $\mathbf{V}$ does not have full column rank, certain state-space dimensions are neglected. This fact can be exploited to approximate the system with a reduced number of state variables. In the case of a second degree system, the order of the new system would be $2 \cdot rank(\mathbf{V})$. Therefore, it is straight forward to try and find a basis for the system's state-space in which as many dimensions as possible can be neglected, provided the system is only considered in a certain frequency band. It has been shown in [6] that the choice of a Krylov subspace $\mathcal{K}_n(\mathbf{B}, \mathbf{y}) = \left\{\mathbf{y}, \mathbf{B}\mathbf{y}, \mathbf{B}^2\mathbf{y}, \ldots, \mathbf{B}^{n-1}\mathbf{y}\right\}$, is especially suitable for this purpose. If the columns of $\mathbf{V}$ form an orthonormalized basis of $\mathcal{K}_n\left(\left(s_0^2\mathbf{I} - \mathbf{A}_2\right)^{-1}, \mathbf{R}_2\right)$, then this system

$$s^2\mathbf{X}' = \mathbf{V}^T\mathbf{A}_2\mathbf{V}\mathbf{X}' + s\mathbf{V}^T\mathbf{R}_2\mathbf{i} \quad (11)$$

$$\mathbf{u} = \mathbf{R}_2^T\mathbf{V}\mathbf{X}' \quad (12)$$

has only a state-space dimension of $n \cdot m$ and the impedance matrix

$$\mathbf{Z}'(s) = s\mathbf{R}_2^T\mathbf{V} \left(s^2\mathbf{I} - \mathbf{V}^T\mathbf{A}_2\mathbf{V}\right)^{-1} \mathbf{V}^T\mathbf{R}_2 \quad (13)$$

is a Padé approximation of $\mathbf{Z}(s)$ in the sense of a Taylor series around s_0^2. If $n \cdot m \ll 3N$, the inverse in (13) can be computed with very little numerical effort.

For the construction of this basis, the common Arnoldi or Lanczos algorithms (see [7] for definition) can be applied but have to be implemented in a block-wise variant to be able do deal with the m column $\mathbf{R}_2$.

A major drawback of this approach is the necessity to invert or factorize the large matrix $\left(s_0^2\mathbf{I} - \mathbf{A}_2\right)$, which may become prohibitive for many realistic problems due to memory limits. Solving a linear system of equations in every Arnoldi/Lanczos step, however, severely impacts the computation time. To avoid these problems, $\mathbf{V}$ is constructed to be the basis of $\mathcal{K}_n(\mathbf{A}_2, \mathbf{R}_2)$, which corresponds to a Taylor series expansion around $s_0^2 = \infty$. Such an approach does not need the inversion of a large matrix, however, at the expense of a much larger rank of $\mathbf{V}$, i.e., a higher order system.

In order to avoid the computation of a large inverse but nevertheless benefit from a small order system, a combination of those two approaches has been introduced in [3]. In

a first step, the system order is reduced by a Lanczos process, constructing $\mathbf{V}_1$ as a basis of $\mathcal{K}_{n1}(\mathbf{A}_2, \mathbf{R}_2)$, resulting in a significantly smaller order system, which then is inverted and reduced again by an Arnoldi process, constructing $\mathbf{V}_2$ as a basis of $\mathcal{K}_{n2}\left(\left(s_0^2\mathbf{I} - \mathbf{V}_1^T\mathbf{A}_2\mathbf{V}_1\right)^{-1}, \mathbf{V}_1^T\mathbf{R}_2\right)$ to benefit from the Taylor expansion around s_0^2. In the first step, a symmetric variant of the Lanczos process is applied, as its short recursion limits the memory consumption. The second step is performed by an Arnoldi process, as for those relatively small matrices, memory consumption is no longer an issue and the Arnoldi process does not compute spurious eigenvalues.

With the help of the eigenvalue decomposition in (9), the poles and residuals can be computed for the reduced system. Those poles and residuals are a subset of the total number of poles and residuals neglecting the ones that have no or very little influence on the system's port behavior in a certain frequency range.

IV. Computation of Sensitivity

For the optimization in the reduced order space, the dependence of the system's port behavior on $n_\mathbf{x}$ geometry parameters is replaced by a linearized local model. Therefore, the sensitivity of the poles and residuals, with respect to each geometry parameter, must be computed. Let $\mathbf{x}_0$ be the vector of the considered geometry parameters for the initial model with its poles in the vector $\mathbf{p}_0$, and $\Delta\mathbf{x}_i = \Delta x_i \mathbf{e}_i$ the vector containing a slight variation Δx_i of the i-th parameter, then $\mathbf{p}_i = \mathbf{p}(\mathbf{x}_0 + \Delta\mathbf{x}_i)$ is the vector containing the poles of the model with the i-th parameter changed and $\Delta\mathbf{p}_i = \mathbf{p}_i - \mathbf{p}(\mathbf{x}_0)$ is the difference to the poles of the initial model. The sensitivity of the poles can be expressed in the matrix $\mathbf{M}_\mathbf{p} = \left[\begin{array}{cccc}\Delta\mathbf{p}_1 & \Delta\mathbf{p}_2 & \cdots & \Delta\mathbf{p}_{n_\mathbf{x}}\end{array}\right] diag(\Delta\mathbf{x})^{-1}$, where $\Delta\mathbf{x} = \sum_{i=1}^{n_\mathbf{x}} \Delta\mathbf{x}_i$. This allows $\mathbf{p}(\mathbf{x})$ to be replaced locally by the function

$$\tilde{\mathbf{p}}(\mathbf{x}) = \mathbf{M}_\mathbf{p}(\mathbf{x} - \mathbf{x}_0) + \mathbf{p}_0. \tag{14}$$

The sensitivity of the residuals can be computed in the same way. In general, this is required for every one of the m^2 elements of the impedance matrix if the system is not reciprocal, resulting in

$$\tilde{\mathbf{r}}_{ij}(\mathbf{x}) = \mathbf{M}_{\mathbf{r}ij}(\mathbf{x} - \mathbf{x}_0) + \mathbf{r}_{0ij}. \tag{15}$$

Computing only a subset of the original poles and residuals, the Arnoldi/Lanczos processes used in the model order reduction do not guarantee that the same subset is found for each set of parameters. In addition, the eigenvalue solver used in (9) generally returns the eigenvalues in an arbitrary order. Those two facts can make it difficult to find the poles $\mathbf{p}_i$ for one set of parameters that correspond to those for another set of parameters $\mathbf{p}_j$ $(i \neq j)$. In other words, $\mathbf{p}_j$ is not only different from $\mathbf{p}_i$ because it has been computed for a different set of parameters, but also because it lacks some poles, contains additional ones and is reordered. In order to assure consistent poles and residuals, the spaces spanned by the eigenvectors have to be projected onto each other, exploiting the fact that the eigenvectors belonging to corresponding poles are almost parallel.

V. Optimization

With the help of (14) and (15), the reduced number of poles and residuals are expressed as a function of the geometry parameters. Placing those into (10) yields the structure's impedance matrix computed with very little numerical effort. After the conversion to a scattering matrix $\mathbf{S}$, the objective function $e(\mathbf{S}, \mathbf{S}_d)$ can be computed. The desired target function $\mathbf{S}_d$ may be the result of a standard filter synthesis or simply a set of specifications. The objective function e combines the computed scattering parameters with the desired ones and returns a real scalar that can be minimized with the help of commercially available optimizers such as a Sequential Quadratic Programming (SQP) method [8], yielding an optimum set of parameters $\mathbf{x}_1$.

As the poles' and residuals' – in general nonlinear – dependence on the geometry parameters has been replaced by a local model, one can not expect the set $\mathbf{x}_1$ to be the global optimum. Therefore, the process described above is restarted with the parameters in $\mathbf{x}_1$ as the initial set.

The optimization algorithm can be described as follows:

compute $\mathbf{p}_0$ and $\mathbf{r}_0$ for the parameters $\mathbf{x}_0$
for $k = 1$ until goal attained
  for $i = 1$ to number of parameters
    compute $\mathbf{p}_i^{(k)}$ and $\mathbf{r}_i^{(k)}$ for the parameters in $\mathbf{x}_{k-1} + \Delta\mathbf{x}_i$
  endfor
  find corresponding poles and residuals
  compute sensitivity matrices $\mathbf{M}_\mathbf{p}$ and $\mathbf{M}_\mathbf{r}$
  minimize $e(\mathbf{S}, \mathbf{S}_d)$ and obtain optimum parameters $\mathbf{x}_k$
  compute $\mathbf{p}_k$ and $\mathbf{r}_k$ for the parameters $\mathbf{x}_k$
endfor

VI. Iris Coupled Waveguide Filter

In order to demonstrate the efficiency and validity of the presented approach, a coaxial fed waveguide filter with iris coupling presented in [9] shall be optimized. The considered geometry parameters are the inset and the height of the coaxial probe as well as the lengths of the cavities and the dimensions of the coupling irises as shown in Fig. 1.

As the filter consists of six cavities, a 6th order Chebyshev filter with a 0.5dB ripple in the passband, a center frequency at 15.2GHz and a bandwidth of 0.6GHz has been chosen as a target function. Fig. 2 shows the transmission

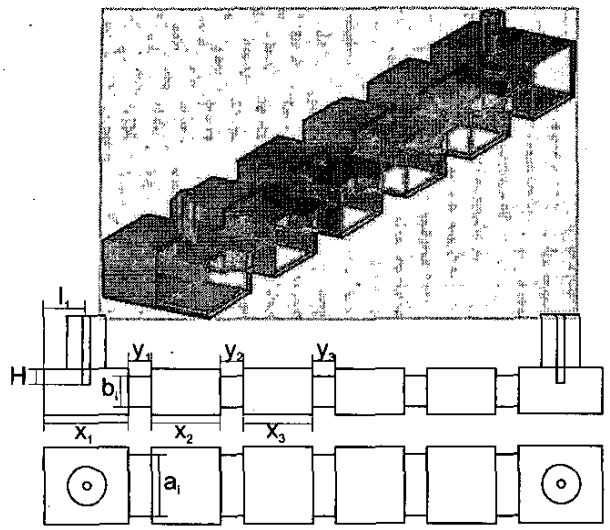

Fig. 1. Iris coupled waveguide filter

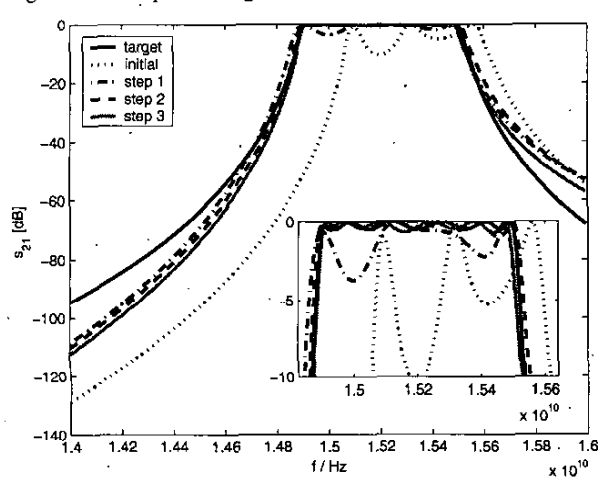

Fig. 2. Transmission coefficient in three optimization steps

coefficient s_{21} over frequency after each optimization step. The objective function in this case has simply been the sum over the quadratic differences between the absolute values of the computed response and the designed filter response at each frequency sample. Objective functions that also consider the phase or weigh certain frequency bands differently are expected to yield even better results and higher stability. It can be seen already in the first step that the center frequency and bandwidth requirements have been met and the passband ripple is only 2dB. After the second step, the ripple has been decreased down to 1dB, and in the third step, the requirements are fulfilled. To achieve this goal, the filter has been discretized with roughly 60000 mesh-nodes, and 34 Model Order Reduction runs were necessary each taking roughly two minutes on an off the shelf PC. The objective function has been evaluated about 500 times, which means that reaching this goal with the help of a conventional optimizer would have taken over 16 hours.

VII. Conclusions

A fast filter optimization scheme has been presented. It has been shown, that a Model Order Reduction approach is highly suitable for this kind of resonant structures. Combined with the optimization in the reduced order space, it can be applied to the optimization of filters in a very efficient way. An example has shown that the approach is capable of significantly reducing the number of simulator runs that is necessary to obtain the optimum result.

References

[1] P. Harscher, E. Ofli, R. Vahldieck, and S. Amari, "EM-simulator based parameter extraction and optimization technique for microwave and millimeter wave filters," in *Microwave Symposium Digest*, pp. 1113–1116, 2002.

[2] T. Weiland, "Time domain electro-magnetic field computations with finite difference methods," *Int. J. Num. Modeling*, vol. 9, pp. 295–319, 1996.

[3] T. Wittig, I. Munteanu, R. Schuhmann, and T. Weiland, "Two step lanczos algorithm for model order reduction," *IEEE Transactions on Magnetics*, vol. 38, pp. 673–676, March 2002.

[4] T. Wittig, I. Munteanu, R. Schuhmann, and T. Weiland, "Model order reduction and equivalent circuit generation for a FIT curl-curl formulation," in *Proc. of ACE Conference*, pp. 256–272, 2002.

[5] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

[6] P. Feldmann and R. W. Freund, "Interconnect-delay computation and signal-integrity verification using the SymPVL algorithm," in *Proc. 1997 Eur. Conf. Circuit Theory and Design*, pp. 408–413, 1997.

[7] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford: Oxford University Press, 1965.

[8] R. Fletcher and M. Powel, "A rapidly convergent descent method for minimization," *Computer Journal*, vol. 6, pp. 163–168, 1963.

[9] J.-F. Liang, H.-C. Chang, and K. A. Zaki, "Coaxial probe modeling in waveguides and cavities," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, pp. 2172–2180, December 1992.