

COUPLER DESIGN AND OPTIMIZATION BY GPU-ACCELERATED DG-FEM

C. Potratz, H.-W. Glock and U. van Rienen, Universität Rostock, Rostock, Germany

Abstract

In this contribution we show how consumer graphic processors (GPUs) in conjunction with a suited numerical method can be used for the inexpensive design and optimization of small and medium sized rf components. The underlying scheme is implemented in an open source framework which is readily available. As an application example we present simulations of a possible CERN SPL [1] Higher Order Mode (HOM) coupler design.

INTRODUCTION

The numerical optimization of rf-components like couplers is a common task during the design phase of particle accelerators. Typically, these optimizations involve the simulation of a multitude of very similar structures with minor geometric variations. Nevertheless, this process is in its entire extend rather demanding on both the invested time and hardware budget. The most time consuming step of the optimisation procedure is the field simulation that has to be repeated for every geometrical variation investigated. Recent advancements in the field of numerical electromagnetic field simulation and consumer graphic processors open an interesting alternative for this task. In this contribution we show, how the Discontinuous Galerkin FEM method [3] in conjunction with consumer graphic cards can be used to build moderately prized cluster solutions for the parallel simulation of rf-components. The contribution will mainly focus on, but is not limited to, HOM couplers as a typical application example, where the DG-FEM method accelerated by a graphic processor might be used to significantly reduce the overall time necessary for the optimization. As an application example, we choose a possible HOM coupler for CERN-SPL. The proposed method is compared with a well established commercial code. Geometrical parameters for an optimized design are given.

THEORY

In this section we give a brief introduction into the used DG-FEM method. A more detailed overview can be found in [2]. For the application of the DG-FEM method [3], it is assumed that the computational domain Ω has been decomposed into a set of K non-overlapping elements with each element containing N_p allocation points. Furthermore, we assume that each element is homogeneously filled with a specific material with the permittivity ε and the permeability μ . Within each element, a set of orthogonal nodal basis

* Work supported by the German Ministry for Education and Research (BMBF) under contract 05K10HRC.

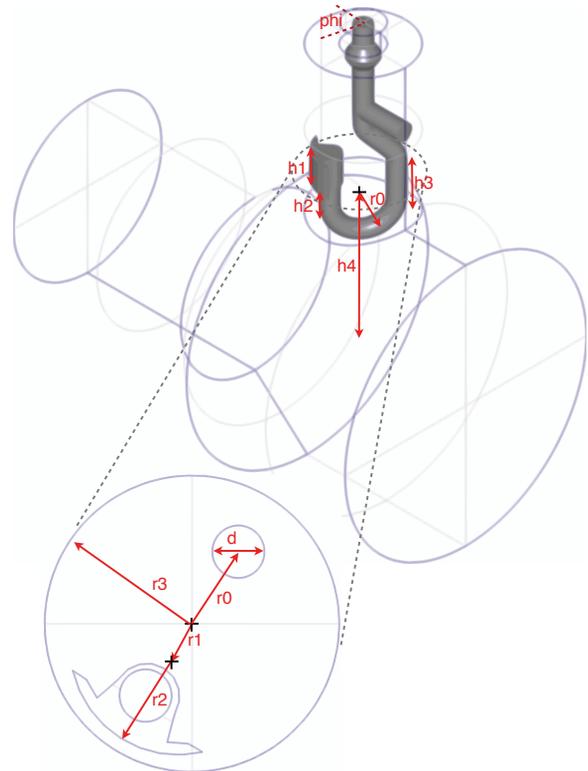


Figure 1: Model of the CERN-SPL HOM coupler (side without power coupler) with taper from 65 mm to 40 mm beam pipe (top). Cross section through the proposed HOM coupler (bottom). For a detailed description of the highlighted geometric parameters included in the optimization see table 1.

functions is defined with $l_i^k(\mathbf{r})$ denoting the i th basis function ($i = 1, \dots, N_p$) on the k th element ($k = 1, \dots, K$) in Lagrange form [3]. Then the electric and magnetic field inside each element can be approximated as

$$\mathbf{E}(\mathbf{r}) \approx \sum_{i=1}^{N_p} \begin{pmatrix} E_{xi}^k \\ E_{yi}^k \\ E_{zi}^k \end{pmatrix} l_i^k(\mathbf{r}) \quad \mathbf{H}(\mathbf{r}) \approx \sum_{i=1}^{N_p} \begin{pmatrix} H_{xi}^k \\ H_{yi}^k \\ H_{zi}^k \end{pmatrix} l_i^k(\mathbf{r})$$

with $E_{xi}^k, \dots, H_{zi}^k$ denoting the coefficients of the nodal basis functions for each field component.

Inserting the approximation into Ampere's and Faraday's law, assuming constant material properties in each element and applying Galerkin's method of a weak formulation [3], two semi-discrete equations

$$\begin{aligned} \partial_t \underline{\mathbf{E}} &= \frac{1}{\varepsilon} \underline{\mathbf{M}}_k^{-1} \underline{\mathbf{S}}_k \underline{\mathbf{H}} + \frac{1}{\varepsilon} \underline{\mathbf{M}}_k^{-1} \underline{\mathbf{L}}_k \cdot (\mathbf{n} \times (\mathbf{F}_E - \mathbf{F}_E^*)) \\ \partial_t \underline{\mathbf{H}} &= -\frac{1}{\mu} \underline{\mathbf{M}}_k^{-1} \underline{\mathbf{S}}_k \underline{\mathbf{E}} - \frac{1}{\mu} \underline{\mathbf{M}}_k^{-1} \underline{\mathbf{L}}_k \cdot (\mathbf{n} \times (\mathbf{F}_H - \mathbf{F}_H^*)) \end{aligned} \quad (1)$$

for the nodal electric and magnetic field coefficients (\underline{E} , \underline{H}) for each element can be derived. The mass matrix \underline{M}_k , the discrete curl operator \underline{S}_k and the surface integration matrix \underline{L}_k , which connects adjacent elements are of the size $3N_p \times 3N_p$ (in the used formulation $N_p = 20$).

It is specific to the DG-FEM approach that the integration is performed only on the single element and not on the entire domain. This results in a strongly local nature of the mass matrices, expressed by their small size of $3N_p \times 3N_p$, which especially makes their immediate inversion feasible. This is the key fact to allow for the construction of an explicit time domain scheme best suited for the implementation on massive parallel structured hardware like GPUs.

For the time integration of (1) we use an explicit fourth-order low-storage Runge-Kutta method as suggested in [3]. The termination of the computational domain is realized by a set of one-dimensional waveguide models which are coupled to the three-dimensional domain. Each model terminates one corresponding waveguide mode at the boundary. Furthermore, the waveguide models provide a time series of the modal expansion coefficients of the corresponding mode which conveniently enables the computation of scattering parameters. A more detailed discussion on the specific boundary condition can be found in [2].

APPLICATION EXAMPLE

Performance Comparison

In this section, we present a comparison between the scattering parameters computed with DG-FEM on a GPU and CST Studio Suite 2011 [4]. For the actual implementation we used the NUDG++ framework [5]. This framework implements the basic DG-FEM functionality as described in [3] and also provides an implementation of a time iterator that integrates the equations (1) for graphic cards based on the NVIDIA CUDA library [6]. Within this framework, we implemented a waveguide boundary as GPU ready code (a detailed description of the waveguide boundary can be found in [2]). The reference solution was computed on a high end workstation with an Intel XEON 3.47 GHz CPU and CST Studio Suite 2011's transient solver using all six available cores in parallel.

The model shown in fig. 1 is a possible design for a SPL HOM coupler and is partially optimized (the notch frequency is not entirely tuned to 704.4 MHz). In both the CST Studio and the DG-FEM simulation, the model was excited at the coaxial port and the fields inside the structure where integrated in time until the total stored field energy in the three-dimensional structure was less than -40 dB of the total energy of the excitation signal.

Figure 2 shows the transmission S_{32} computed with CST Studio's Transient Solver (red, dashed) and with the DG-FEM code (blue, continuous) and the fundamental notch frequency F_n for different discretizations (expressed by the number of degrees of freedom). The computation of the DG-FEM solution with $3.7 \cdot 10^6$ dofs (leftmost point)

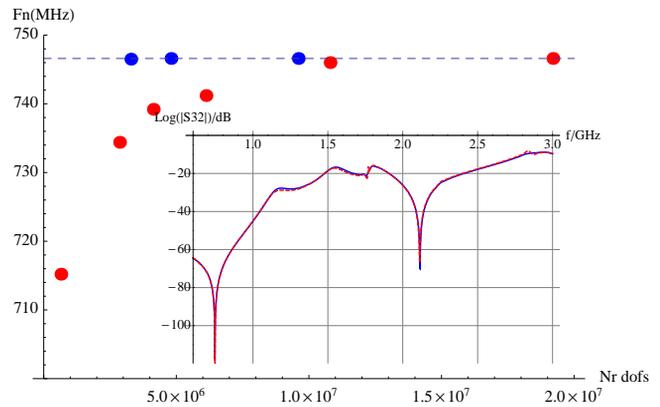


Figure 2: Transmission S_{32} computed with the DG-FEM on a GPGPU (blue solid line) and computed using the transient solver of CST Studio Suite 2011 and fundamental notch filter frequency (F_n , blue dots DG-FEM and red dots CST Transient Solver) as function of the degrees of freedom (dofs).

took 1920 s. The computation of the CST Studio solution $1.9 \cdot 10^7$ dofs that reaches a similar precision was found with significantly higher hardware effort in 1900 s. The notch frequency extracted from the simulated model converges to 746.5 MHz for both programs. As expected, the DG-FEM code needs far less dofs for the same geometric accuracy which is related to the unstructured mesh and the higher order of the used polynomial basis functions. Especially, the unstructured mesh allows for a good shape representation of small geometric features (e.g. the capacitive element) without distributing local refinements throughout the rest of the structure.

On the other hand, the DG-FEM code suffers from a drawback which is related to the maximum stable time step that can be used for temporal integration. For the same simulated time span of 14 ns approximately $3.5 \cdot 10^5$ time steps are needed in the DG-FEM code while CST Studio Suite can use a considerably larger time step which leads to a total of $1.1 \cdot 10^5$ integration steps. Therefore, while the unstructured mesh helps to reduce the number of dofs significantly, the higher approximation order of the underlying polynomial basis function enforces a smaller global time step and thus increases the number of necessary evaluations.

This limitation can be compensated by a local time stepping approach. Here, different regions of the discretized model are integrated with different time steps according to the corresponding element sizes. An approach tailored to DG-FEM was presented by Gödel et al. [7]. We implemented a similar procedure which is based on an Adams-Bashforth scheme that reduced the simulation time for the model to 1300 s.

In conclusion, the DG-FEM code in conjunction with consumer GPUs is well suited for the time consuming field computations in coupler optimization. However care is advised if the geometry features very small geometric details

which enforce small time steps and might drastically reduce the total performance. These situations can be partly avoided using a local time stepping for the temporal integration.

Optimized Higher Order Mode Coupler for CERN-SPL

In this section we present an optimized HOM coupler for CERN SPL which was optimized using the method described above. For the optimization, we used a cluster of 12 NVidia GTX 470 cards, each card simulating one model at the time. For the optimization, the following optimization goals were defined:

1. tune the fundamental notch filter to 704.4 MHz;
2. minimize the transmission at the fundamental notch frequency 704.4 MHz;
3. avoid additional frequency notches at multiple integers of the fundamental frequency;
4. maximize the coupling at all other modes/frequencies.

For these goals approximately 10^4 geometrical parameters of the HOM coupler design were simulated on the GPU cluster.

Table 1: Geometric Parameters of the optimized HOM coupler for the design shown in figure 1.

Parameter	Value	Description
r0	13 mm	Hook bend radius
r1	4 mm	Cap. Offset radius
r2	16.5 mm	Cap. Element radius
r3	22.5 mm	HOM inner radius
d	8 mm	Hook diameter
phi	40°	Rotation around axes
h1	14 mm	Height of cap. element
h2	10.5 mm	Height of cap. element support rod 1
h3	17.5 mm	Height of cap. element support rod 2
h4	67.5 mm	distance to beam axis

Table 1 summarizes the optimized geometrical parameters for the design shown in figure 1. Figure 3 shows the dependency of the first and second notch frequency (see figure 2) of the parameters h_1 and ϕ extracted from the simulated geometric variations. Parameter combinations of h_1 and ϕ resulting in a fundamental notch frequency of 704.4 MHz or 2113.2 MHz ($3 \cdot 704.4$ MHz) are marked in red and green respectively. Care must be taken when choosing a specific parameter combination to avoid a second notch filter at an integer multiple of the fundamental frequency of $3 \cdot 704.4$ MHz.

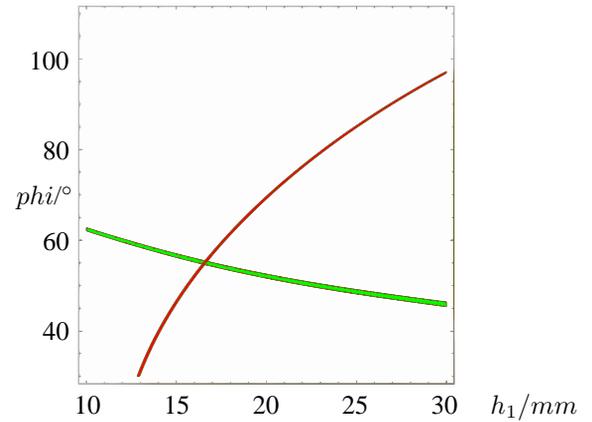


Figure 3: Results from a sweep of the parameter h_1 and ϕ of figure 1. Combinations resulting in a fundamental notch frequency of 704.4 MHz and a second notch filter frequency of 2113.2 MHz ($3 \cdot 704.4$ MHz) are marked in red and green, respectively.

CONCLUSIONS

In this contribution we have shown how moderately priced consumer GPUs in conjunction with a suited numerical scheme can be used to accurately compute scattering parameters of lowly resonant structures like the HOM couplers. An application example from an actual computation performed for the CERN-SPL HOM coupler was given and the results were compared to a standard commercial software package. Based on the presented method, the HOM coupler for CERN SPL was optimized and a preliminary set of geometrical parameters is given. Furthermore, a secondary notch filter frequency was found for the proposed design and it was shown, that this notch can occur at an integer multiple of the fundamental frequency if the geometrical parameters are not carefully chosen.

REFERENCES

- [1] M. Baylac et. al, Conceptual design of the SPL II : A high-power superconducting H linac at CERN, Report (CERN-2006-006)
- [2] C. Potratz, H.-W. Glock and U. van Rien, Time Domain Field and Scattering Parameter Computation in Waveguide Structures by GPU-Accelerated Discontinuous-Galerkin Method, accepted for publication in IEEE Transactions on Microwave Theory and Techniques.
- [3] J. Hesthaven and T. Warburton, Nodal Discontinuous Galerkin Methods", 1st ed., Springer, 2007.
- [4] CST AG, Bad Nauheimer Str. 19, 64289 Darmstadt, Germany
- [5] N. Nunn and T. Warburton, "Nudg++: a Nodal Unstructured Discontinuous Galerkin framework", 2007.
- [6] NVIDIA CUDA Programming Guide 3.0, 2010.
- [7] N. Gödel, S. Schomann, T. Warburton, and M. Clemens, "GPU Accelerated Adams-Bashforth Multirate Discontinuous Galerkin FEM Simulation of High-Frequency Electromagnetic Fields", IEEE Transactions on Magnetics, vol. 46, no. 8, pp. 2735–2738, 2010.