GPU Accelerated EM Modelling in Frequency Domain: Comparison of Performance of Various GPU Cards

Dusan P. Zoric [1], Dragan I. Olcan [2], [#] Branko M. Kolundzija [2]
[1] WIPL-D d.o.o.
Gandijeva 7/32, Belgrade, Serbia, dusan.zoric@wipl-d.com
[2] School of Electrical Engineering, University of Belgrade
King Alexander's Blvd 73, 11120 Belgrade, Serbia, olcan@etf.rs, kol@etf.rs

1. Introduction

There is a constant interest in solving electrically larger and more complex electromagnetic (EM) problems on present day computers. Method-of-moments (MoM) is one of general-purpose numerical techniques for time-harmonic numerical analysis of EM systems with isotropic, linear and piecewise homogeneous materials [1].

When solving EM problems using MoM, with the increase of electrical size and complexity of the EM structure, matrix fill time increases as N^2, and matrix inversion time increases as N^3 [1]. Presently, graphics processing units (GPU) offer more computational power for the same invested money than the conventional central processing units (CPU) [2]. NVIDIA CUDA technology exposes GPU hardware capabilities for higher level programming through libraries for C and Fortran [3].

Recently, CUDA technology was successfully used to accelerate the EM analysis in MoM based software WIPL-D [4] thus reducing the simulation times for one order of magnitude at desktop PC [5], [6], [7], [8]. The aim of this paper is to compare performance of the best GPUs from three NVIDIA GeForce series: GTX 480 (400 series), GTX 580 (500 series) and GTX 680 (600 series) in solving EM problems and to determine which hardware parameters have the most influence on computational power of given GPU.

2. Accelerating WIPL-D Numerical Kernel Using CUDA-enabled GPUs

Using CUDA technology, three operations in MoM analysis: matrix fill, matrix inversion and near field calculations were accelerated in software WIPL-D in last few years. In this section, only basic information about accelerating WIPL-D numerical kernel using CUDA will be given. More on this topic can be found in [5], [6], [7], [8].

GPU parallelized algorithm for MoM matrix calculations was developed starting from the original CPU parallelized WIPL-D algorithm. However, the existing algorithm was too complex for direct porting to GPU. Namely, for efficient acceleration the specific logic of GPU parallelization must be obeyed. Numerical results show that acceleration of up to 21 times was achieved on three GeForce GTX 480 GPUs compared to Intel's i7 CPU.

The most effort was invested to accelerate matrix inversion, the most time demanding part of MoM EM simulations. Algorithm for LU decomposition, that uses highly optimized subroutines from NVIDIA CUBLAS library [3], was developed. It enables usage of up to three GPUs in parallel. Also, many efforts have been made to extend size of problems that can be solved using GPUs. When the required memory for solving problem is larger than available random access memory, WIPL-D uses out-of-core solution, where hard-disks are used to store the MoM matrix. Hence, besides the GPU accelerated in-core matrix inversion, GPU accelerated out-of-core matrix inversion was developed. Recent results show that, for example, for project of 50000 unknowns, in-core matrix inversion can be done in 4 minutes, and out-of-core matrix inversion for project with about 450000 unknowns can be done in one day.

3. Comparison of Performance of Various GPU Cards

List of GPUs used in this paper with basic hardware specifications is shown in Table 1 [3]. GPUs from GeForce 400 series and GeForce 500 series are based on NVIDIA Fermi architecture, and GPU from GeForce 600 series is based on the new, Kepler architecture.

Table 1: The list of GPUs used in the paper with basic hardware specifications.

Name	GeForce GTX 480	GeForce GTX 580	GeForce GTX 680
# Streaming multiprocessors	15	16	8
# Cores	480	512	1536
Core clock [MHz]	1401	1544	1058
Memory bandwidth [GB/s]	177.4	192.4	192.2
Memory size [MB]	1536	1536	2048

In order to compare performances of given GPUs in solving real EM problems, several numerical examples are discussed. We discuss analyses of: (1) scattering from PEC sphere, (2) 14 λ Luneburg lens, (3) 4 by 4 microstrip patch antenna array placed at the bottom of a helicopter fuselage.

The same PC is used for all numerical experiments. The configuration of that PC is as follows: Intel Core i7 CPU 930 @2.8GHz, 24 GB of RAM, Windows 7 Professional 64-bit, and 2 hard-disk drives with I/O speed of about 100 MB/s.

3.1 PEC Sphere

As the first numerical example, we analyse PEC sphere, shown in Fig. 1. The sphere is illuminated with the linearly polarized plane TEM wave. The wave travels along x-axis and the electric field vector is along z-axis. The operating frequency and radius are adjusted so that the model has appropriate electrical size and number of unknown coefficients.

PEC spheres diameters take values from 10 to 70 λ, where λ is the wavelength at the operating frequency in a free-space. For models with diameters from 10 to 30 λ, the symmetry of the analysed structure is not used. The equivalent models have approximately 12000 to 77000 unknown coefficients.

The number of unknowns for the largest sphere is 72710. However, (A)symmetry option is used, which means that two analysis are actually performed [4]. Table 2 shows the comparison of the measured times for analysing spheres using three different GPUs.

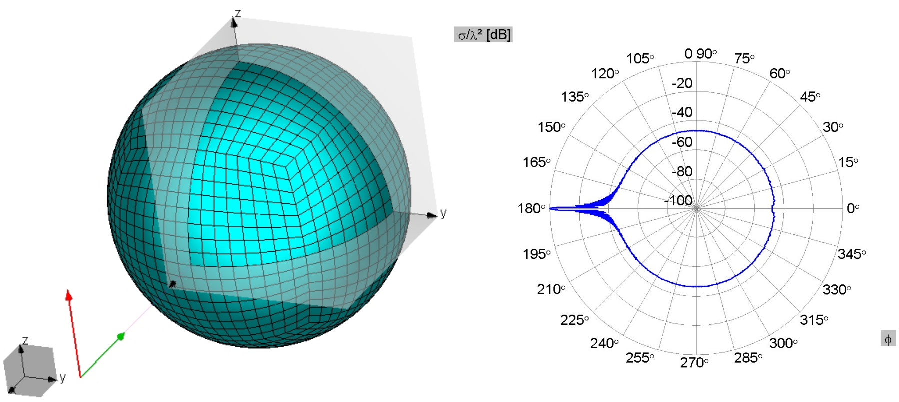

Figure 1: WIPL-D model of PEC sphere and RCS of 70 λ sphere.

It can be seen that on all GPUs similar simulation times were achieved. If we compare only GTX 480 and GTX 580, we can see almost constant difference of about 10% in matrix inversion times and more or less similar difference in matrix fill times. This difference is expected knowing that both GPUs are based on same architecture and that GTX 580 has about 10% higher number of cores and memory bandwidth. If we now compare GTX 580 with GTX 680, we can see that matrix inversion is only up to 9% shorter on GTX 680. On the other hand, matrix fill times achieved on GTX 680 are about 15% higher compared to GTX 580. From Table 1 we can see that GTX 680 has three times more cores, and according to that fact, we will expect to obtain much better results on this card. However, the memory bandwidth is practically the same on both GPUs. Thus, it seems that memory bandwidth is the most critical parameter of GPU hardware specification when using it to accelerate MoM based EM simulation code.

Table 2: Comparison of the measured times for analysing PEC spheres using three different GPUs.

N	GeForce GTX 480			GeForce GTX 580			GeForce GTX 680		
	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]
12288	12	10.2	22.2	11.9	9.6	21.5	13.6	10.2	23.8
27648	49	67	116	48	61	109	53	60	113
49152	143	299	442	126	271	397	141	260	401
76800	348	1027	1375	327	923	1250	368	866	1234
110592	756	2739	3495	689	2435	3124	784	2222	3006
*72710	3294	1764	5058	3080	1590	4670	3710	1500	5210

3.2 Luneburg Lens Excited with Half-wavelength Dipole

The second analysed EM structure is a spherical Luneburg lens exited by a half-wavelength dipole. The dipole is placed quarter of the wavelength away from the lens surface. The relative permittivity of the Luneburg lens is continuous function, given as

$$\varepsilon_{\mathrm{r}}(r) = 2 - \left(\frac{r}{a}\right)^2, \tag{1}$$

where r is the distance from the lens center and a is the lens radius. The diameter of the analysed lens is 14 λ. The lens is modeled with ten concentric spherical layers. The layers have equal thickness and constant relative permittivity. The relative permittivity of a layer is set to the relative permittivity at the layer midpoint radius given by (1). Due to the symmetry of the structure it is sufficient to analyze only one quarter of it. One quarter of the lens WIPL-D model is shown in Fig. 2. The measured analysis times on three different GPUs are given in Table 3.

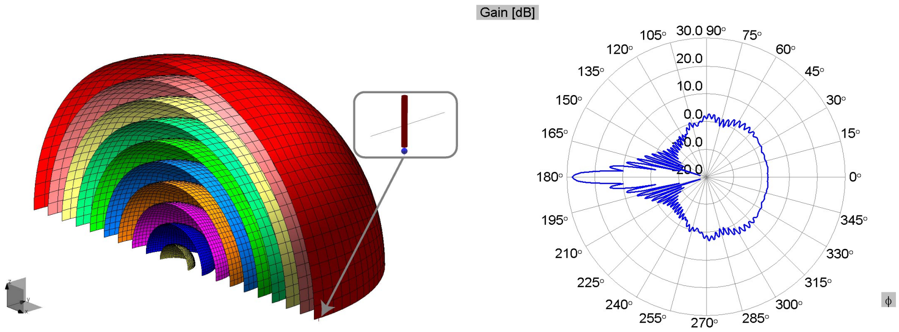

Figure 2: One quarter of the 14 λ Luneburg lens excited with a half-wavelength dipole and radiation pattern of the lens.

Table 3: Comparison of the measured times for analyzing Luneburg lens using three different GPUs.

N	GeForce GTX 480			GeForce GTX 580			GeForce GTX 680		
	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]
102962	927	3861	4788	872	3449	4321	968	2995	3963

3.3 Microstrip Patch Antennas Placed on Helicopter

The last numerical example considered is the radiation of a 4 by 4 microstrip patch antenna array placed at the bottom of a helicopter fuselage, as shown in Fig. 3. The array is adjusted to operate at a frequency of 1.15 GHz. Total length of helicopter is 15 m, so that its electrical length at this frequency is 57.5 λ. The measured analysis times on three different GPUs are given in Table 4.

Table 4: Comparison of the measured times for analysing microstrip patch antenna array on helicopter fuselage using three different GPUs.

N	GeForce GTX 480			GeForce GTX 580			GeForce GTX 680		
	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]	Fill [s]	Inv. [s]	All [s]
99741	878	3550	4428	815	3096	3911	895	2794	3689

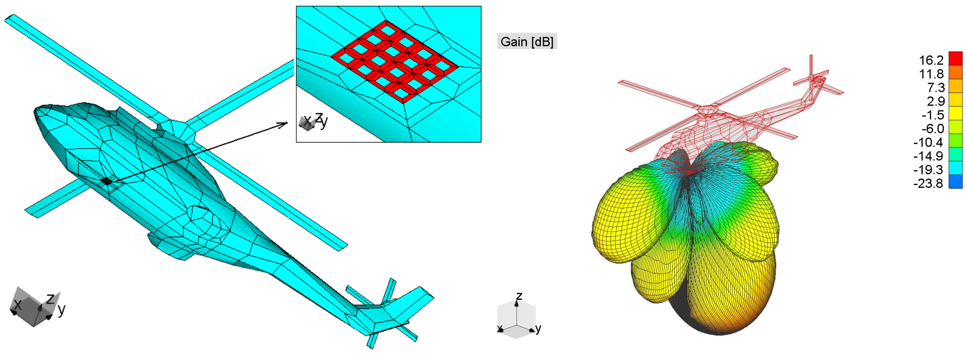

Figure 3: Geometrical model of an array of 4 by 4 microstrip patch antennas placed on helicopter and obtained radiation pattern.

As in the case of analysing PEC cube scattered, the similar performances were achieved in last two examples, and the same conclusion can be made. Besides GPUs stated in this paper, the similar test were performed on the following GPUs: GTX 460, GTX 470, GTX 570 and TESLA C2070. In all tests, the difference in achieved performances more or less followed the difference in GPU memory bandwidths.

5. Conclusion

In this paper, we compared performances of different NVIDIA CUDA-enabled GPUs in accelerating MoM based software for EM modelling in frequency domain, WIPL-D. Three different GeForce GPUs were used: GTX 480, GTX 580 and GTX 680. The performance was measured on three EM problems: PEC sphere, Luneburg lens, and microstrip patch antenna array placed at the helicopter fuselage. The best performance in matrix fill was achieved on GTX 580, and the best performance in matrix inversion was achieved on GTX 680. Also, it was shown that the most critical parameter for achieving better performance is GPU memory bandwidth.

References

[1] B.M. Kolundzija, A.R. Djordjevic, *Electromagnetic modelling of composite metallic and dielectric structures*, Artech House, Boston, 2002.

[2] V. Volkov, J.W. Demmel, "Benchmarking GPUs to tune dense linear algebra," ACM/IEEE Conference on Supercomputing SC2008, Austin, Texas, 2008.

[3] http://www.nvidia.com/object/cuda_home.html

[4] WIPL-D v9.0, Software and Users' Manual, www.wipl-d.com

[5] D.P. Zoric, D.I. Olcan, B.M. Kolundzija, "Solving electrically large EM problems by using out-of-core solver accelerated with multiple graphical processing units," IEEE AP-S 2011, Spokane, WA, USA, pp. 3309-3312, 2011.

[6] Branko M. Kolundzija, Dragan I. Olcan, Dusan P. Zoric, and Sladjana M. Maric, "Accelerating WIPLD Numerical EM Kernel by using Graphics Processing Units", Proc. of TELSIKS 2011, Nis, Serbia, pp. 413-419, 2011.

[7] D.P. Zoric, D.I. Olcan, B.M. Kolundzija, "On Increasing Capabilities of GPU In-Core Solver Applied to Method of Moments," Proc. of EuCAP 2012, Prague, Czech Republic, 2011

[8] D.P. Zoric, D.I. Olcan, B.M. Kolundzija, "Efficient Evaluation of MoM Matrix Elements Using CPU and/or GPU," Proc. of EuCAP 2012, 2012, Prague, Czech Republic.