Accelerated MR Physics Simulations on multi-GPU systems

Christos G. Xanthis, *Member, IEEE*, Ioannis E. Venetis, and Anthony H. Aletras

***Abstract* — A multi-GPU approach of MRISIMUL, a recently developed step-by-step comprehensive MR physics simulator of the Bloch equation, is presented in this study. The specific aim was to apply MRISIMUL on multi-GPU systems so as to achieve even shorter execution times. We hypothesized that such a simulation platform could achieve a scalable performance with the increasing number of available GPU cards on single node, multi-GPU computer systems. A parallelization strategy was employed using the MATLAB single-program-multiple-data (spmd) statement and an almost linear speedup was observed with the increasing number of available GPU cards on two separate systems: a single computer of 2 quad-core processors and two Tesla C2070 GPU cards and a single computer of 2 hexa-core processors and four Tesla C2075 GPU cards.**

I. Introduction

The acquisition and the interpretation of Magnetic Resonance (MR) images usually requires a higher level of understanding of the underlying physics. Towards that direction, MRI physics simulators have been used in the past for training purposes but also for tracking artifact sources and answering methodological problems during pulse sequence or imaging protocol development. However, current MRI simulators are usually confined to a handful of pulse sequences whereas important compromises are made due to the high computational power needed. Two previously developed MRI simulation platforms [1, 2] have been shown to address the long execution times of more complicated large-scale experiments by utilizing parallel processing on computer clusters based on the Message Passing Interface (MPI) communications protocol [2, 3]. Nevertheless, both systems require an advanced cluster setup of multiple nodes along with advanced technical knowledge.

Recently, a new MR physics simulation platform, by the name MRISIMUL, was developed and presented [4]. MRISIMUL is a step-by-step comprehensive Bloch equation simulator from signal acquisition to image formation which allows its application in large-scale analysis without model simplifications by employing GPU technology. Previous results [4] have already demonstrated high computational speedup of GPU-based simulations on a single board GPU personal computer compared to serially executed code on the CPU and OpenMP parallel executed code on the CPU with multi-threading (up to 8 threads).

In this study, the implementation of MRISIMUL on single node, multi-GPU computer systems is presented. The specific aim was to apply MRISIMUL on multi-GPU systems so as to achieve even shorter execution times. We hypothesized that such a simulation platform could be adapted and exploited on single node, multi-GPU computer systems achieving a scalable speedup compared to that of a single GPU system.

II. Methods

MRISIMUL is a step-by-step comprehensive Bloch equation simulator [4]. This simulation platform allowed the development of custom MRI pulse sequences and their application on 3D computer models of realistic objects. MRISIMUL was developed in MATLAB (The Mathworks Inc., Natick, MA) while the computationally demanding core services (kernel) were developed in CUDA-C (NVIDIA, Santa Clara, CA) and executed in parallel within the graphic processing unit (GPU) environment. MATLAB handled the basic MR imaging processes, such as pulse sequence programming, anatomical object development and reconstruction whereas CUDA-C handled the multiplication and summation of large matrices for the calculation of the object's magnetization vector.

For multi-GPU parallelism, the MATLAB single-program-multiple-data (spmd) statement was employed. The number of the on-system available GPU cards defined the size of the pool of processes for parallel computation. Each process was assigned a certain part of the object being analyzed and a GPU card that would be used for the computationally demanding CUDA-C kernel. The assignments were designed to preserve a balanced load for all GPUs. At the end, the raw datasets of each part of the object were summed so as to form the k-space of the object (figure 1).

To demonstrate the efficacy of the multi-GPU implementation of MRISIMUL, its performance was evaluated on two different single computers with multiple GPUs. The first system was a single computer of 2 quad-core processors and 48GB RAM with two Tesla C2070 GPU cards of 448 GPU cores and 6GB global memory each.

Manuscript received July 1, 2013. This project is funded by the European Research Council (PIRG06-GA-2009-256569 FP7 MARIE CURIE IRG) and by the "Alexander S. Onassis public benefit foundation". Two NVIDIA Tesla C2070 GPU computing cards were donated from NVIDIA through the "Professor Partnership" program. The University of Thessaly funded hardware for this project. The United States National Library of Medicine (Bethesda, MD) for the Visible Human Project (VHP) dataset.

C. G. Xanthis and A. H. Aletras are with the Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece and with the Department of Clinical Physiology, Skåne University Hospital Lund, Lund University, Sweden (e-mail: cxanthis@gmail.com; aletras@hotmail.com).

I. E. Venetis is with the Department of Computer Engineering and Informatics, University of Patras, Greece (e-mail: venetis@capsl.udel.edu).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

The second system was a single computer of 2 hexa-core processors and 32 GB RAM with four Tesla C2075 GPU cards of 448 GPU cores and 6GB global memory each.

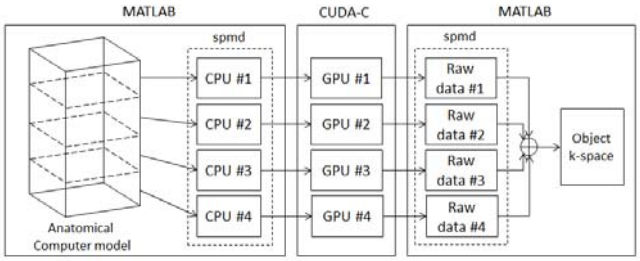

Fig. 1. Each CPU core is assigned a certain part of the anatomical computer model and a certain GPU card. A balanced load among the GPUs has been considered. At the end of the kernel execution, the resulting signals were summed and the final k-space of the object was formed.

The execution times were recorded for all the possible combinations of the available GPUs in both computers for the application of a Gradient Echo (GE) pulse sequence on a 3D computer model of the anatomy of the human heart [4]. The total number of timesteps was 98816 while the anatomical computer model consisted of a total of 4947708 tissue isochromats. The GE pulse sequence was applied with TE = 4.5msec and TR = 50msec, the RF pulse was a three lobe sinc-shaped pulse with a flip angle of 15^{o} whereas a gradient crusher pulse was applied along the slice selection direction at the end of each TR. The temporal resolution of the pulse sequence was 10μsec, the bandwidth of the receiver was 50 kHz, the k-space matrix was of size 128x128 and the main magnetic field was set to 1.5 Tesla. The heart computer model had a voxel size of 0.33mm x 0.33mm x 0.99mm and it was based on the segmentation of a dataset of high resolution images provided by the Visible Human Project (VHP) of the United States National Library of Medicine (Bethesda, MD) [5].

TABLE I
KERNEL EXECUTION TIMES

Single Computer with 2 Tesla C2070 GPUs

No. of GPUs	Execution time (hours)	Speedup
1	3.10	-
2	1.56	1.99

Single Computer with 4 Tesla C2075 GPUs

No. of GPUs	Execution time (hours)	Speedup
1	3.15	-
2	1.58	1.99
3	1.09	2.89
4	0.82	3.84

Execution times of the kernel were recorded for the application of a pulse sequence of 98816 timesteps on a computer model of the human heart of 4947708 tissue isochromats.

III. RESULTS

The execution times of MRISIMUL's kernel on different single node, multi-GPU computer configurations were recorded and are shown in Table I. It can be seen that a speedup of about 2 times when compared to a single Tesla C2070 GPU was observed on the first computer system with the parallel use of two Tesla C2070 GPUs whereas a speedup of almost 4 times when compared to a single Tesla C2075 GPU was observed on the second computer system with the parallel use of four Tesla C2075 GPUs. To illustrate this better, figure 2 depicts an almost linear scalable performance along with the increasing number of available GPU cards on both single node, multi-GPU computers. The resulting simulated MR image of the above-mentioned simulations, which was acquired from the same slice location of the Visible Human Project (figure 3A), is illustrated in figure 3B.

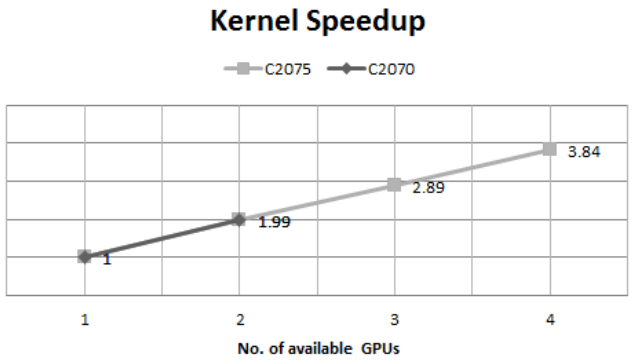

Fig. 2. Speedup observed with increasing number of available GPU cards on the single node, multi-GPU computer systems.

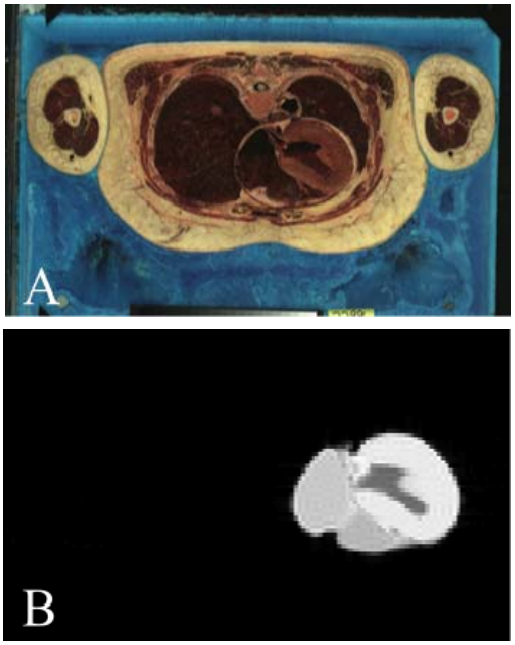

Fig. 3. (A) Thorax image of the Visible Human Project dataset. (B) Simulated MR image of the human heart resulted from the application of the pulse sequence on the computer model of the anatomy of the human heart at the same slice location as image (A).

IV. DISCUSSION

A multi-GPU approach of MRISIMUL was presented in this study. The almost linear scalable performance that MRISIMUL demonstrated with the increasing number of available GPU cards on single-node, multi-GPU computer systems enhances its application in large-scale complicated experiments without model simplifications. Moreover, it suggests that an even better performance may be achieved in multiple nodes, multi-GPU computer systems. However, for those cases, a further investigation of the inter-node communication protocol shall be performed in the future. A web-based version of MRISIMUL, along with its latest releases, is available online under mri.dib.uth.gr.

REFERENCES

[1] H. Benoit-Cattin, G. Collewet, B. Belaroussi, H. Saint-Jalmes, and C. Odet, "The SIMRI project: a versatile and interactive MRI simulator," *J Magn Reson,* vol. 173, pp. 97-115, Mar 2005.

[2] T. Stocker, K. Vahedipour, D. Pflugfelder, and N. J. Shah, "High-performance computing MRI simulations," *Magn Reson Med,* vol. 64, pp. 186-93, Jul 2010.

[3] K. G. Baum, G. Menezes, and M. Helguera, "Simulation of High-Resolution Magnetic Resonance Images on the IBM Blue Gene/L Supercomputer Using SIMRI," *International Journal of Biomedical Imaging,* vol. 2011, 2011.

[4] C. G. Xanthis, I. E. Venetis, and A. H. Aletras, "MRISIMUL: A GPU-based Parallel Approach to MRI Simulations " *Manuscript submitted for publication,* 2013.

[5] M. J. Ackerman, "The Visible Human Project: a resource for anatomical visualization," *Stud Health Technol Inform,* vol. 52 Pt 2, pp. 1030-2, 1998.