Massively Parallel CUDA Simulations of Cardiac and Embryonic MRI on a Cloud-Based Cluster

George Kantasis[1, 3], Christos G Xanthis[1, 3], Kostas Haris[2, 1], Anthony H. Aletras[2, 1]

[1]Department of Clinical Physiology, Skåne University Hospital in Lund, Lund University, Lund, Sweden
[2]Laboratory of Medical Informatics, School of Medicine, Aristotle University of Thessaloniki, Greece
[3]Department of Computer Science and Biomedical Informatics, School of Sciences, University of Thessaly, Greece

Abstract

Embryonic and cardiovascular Magnetic Resonance Imaging (MRI) in both clinic and research environments is confronted by challenges like motion, flow and low resolution. MRI simulations can be used as a valuable tool against these challenges but they are characterized by high computational intensity. In this study we implemented cardiac and embryonic MR simulations on a cloud-based cluster. For this purpose, an anatomical model of a human embryo was split over a variable number of computer nodes on the cloud. Simulation of a Gradient-Echo pulse sequence was performed and the partial results were aggregated in a central node to produce the simulated MR image, following the MapReduce paradigm. The measured execution times and speedups demonstrated the benefits of this cloud-based approach.

1. Introduction

Magnetic Resonance Imaging simulations have been used for pulse sequence optimization and training purposes. Recently, complex experimental setups [1, 2] became feasible using CUDA-technology which resulted in faster cardiac and embryonic MRI simulations. However, in cases of more complex experiments (cardiac motion, blood flow) the execution time is increased which leads to a need for advanced hardware configurations. In this study, we hypothesized that a cloud-based application of advanced MR simulations could address that need. The aim was to distribute a previously developed MR simulator on a scalable cloud-based infrastructure

2. Methods

The MapReduce algorithm was used for the distribution of the recently developed MR simulation platform MRISIMUL on the cloud. In this study, a Gradient-Echo pulse sequence was simulated on a 3D anatomical model of a human embryo. The model was evenly shared among the nodes of the cluster where the simulation was subsequently performed. Then the results were collected on a central node, where the final images was formed.

To evaluate scalability, the execution times of the simulations were recorded for an increasing number of available nodes on the cloud (1-16 nodes). The corresponding speedup, GPU speedup and overhead were calculated.

Speedup was defined as the single node execution time divided by the total execution time for the corresponding number of nodes. GPU speedup was defined in a similar manner; however taking into account only the execution time corresponding to the GPU part of the simulation. The overhead was defined as the remaining non-GPU part of the execution time which included data transfers, slave preparation and the Fourier transform.

2.1. Simulation platform

MRISIMUL [1] is an advanced GPU-based (Graphics Processing Unit), comprehensive Bloch equation simulation platform of MR physics that employs CUDA technology. It allows for pulse sequence simulations on 3D computer anatomical models of isochromats (collection of spins with common MR characteristics) and production of realistic MR images. Since MRISIMUL makes no assumptions regarding the imaging technique or the imaged structure, it can be used in a large variety of pulse sequences and anatomical models.

2.2. Anatomical model

In this study, an anatomical model of a human embryo (Carnegie 23rd stage) was developed based on images from the Multi-Dimensional Human Embryo dataset [3]. The dataset consisted of 38 transverse MR slices of a fixed embryo acquired from a 9.4T MR scanner. The images were T_1, T_2 and diffusion weighted whereas the voxel size was equal to 156.3μm^3. Using a semi-automatic segmentation algorithm, a set of tissue masks were extracted for the 3D anatomical model and T_1, T_2, and proton density values were assigned to the masks. The anatomical computer model of the human embryo consisted of 1,101,824 isochromats.

2.3. Pulse sequence

The simulated pulse sequence was a Gradient-Echo (GRE) with a repetition time (TR) of 8 msec, RF duration of 2 msec, flip angle of 15 degrees, main magnetic field (B0) of 1.5 Tesla, acquisition matrix of size 400 x 400 and field of view (FOV) of 200mm x 200mm. The simulation was performed in 241,601 discreet time-steps.

2.4. MapReduce

The high-parallelizable nature of MR simulations allowed for partitioning the anatomical model and performing the simulation for each partition on different nodes. The MapReduce paradigm was utilized in this study and the problem was distributed evenly across a number of computer nodes (slaves). One of these nodes was also assigned the role of job manager (master node). All nodes were initialized by loading the simulation software and the entire pulse sequence. The Mapper step was implemented by executing the simulation of the pulse sequence on the corresponding part of the model. The Reducer step was then implemented on the master node by collecting the resulting k-space matrices, summing them up and performing the Fourier-transform.

2.5. Cloud infrastructure

Amazon Web Services (aws.amazon.com) was used as the cloud service provider for hosting the cluster infrastructure. The cloud-based experiments were performed on g2.2xlarge computer instances equipped with one NVIDIA GRID K520 GPU of 4GB global memory and 1536 cores.

3. Results

Figure 1 shows the total execution time of the experiment for a variable number of nodes whereas the dashed line represents the measured overhead. In figure 2, the speedups versus the number of nodes are presented. The dashed line is the identity line, the squared line is the GPU speedup and the crossed line is the total speedup for this experiment configuration.

The total simulation time was described by the following formula: TotalTime = 1232 / Nodes + 92 (time is in sec), resulting in a speedup of about 8 times on a sixteen-node cluster compared to the single-node setup. The GPU speedup was almost linear with a slope close to 1 whereas a roughly constant overhead of 92sec, associated with processing and collecting the data, was measured.

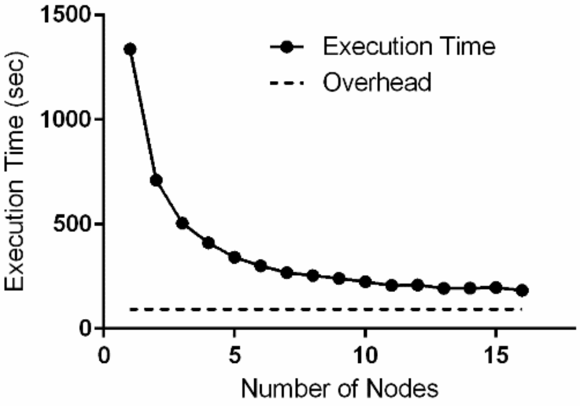

Figure 1: Experiment execution times versus number of cluster nodes (1 – 16). The dashed line shows the overhead.

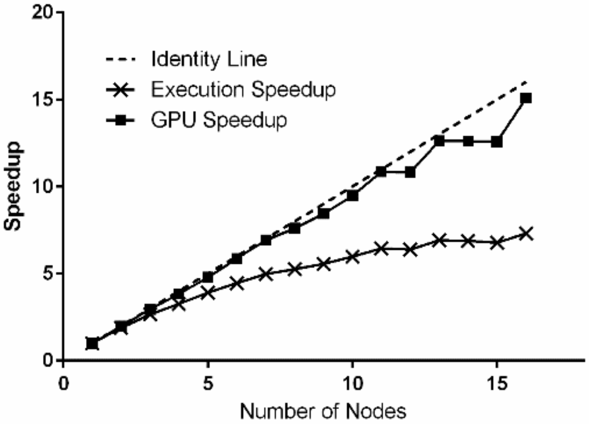

Figure 2: Effective speedup achieved for a variable number of cluster nodes (1 – 16). The squared line shows the GPU speedup, the crossed line shows the total speedup whereas the dashed line shows the identity line

4. Discussion

In this study, we have shown the first implementation of an MRI simulator with cloud services on embryonic data. The GPU speedup was roughly linear with unity slope. The total speedup demonstrated the benefit of distributing an MRI simulation problem across a cloud-based cluster. The main limitation of the total speedup is the overhead which for larger experiment sizes would represent a small fraction of the overall execution time.

Acknowledgements

The cloud-based infrastructure was provided by Amazon web services via a two-year Education Research Grant Award.

This work was funded by an Excellence Grant from the Greek General Secretariat for Research and Development.

References

[1] Xanthis CG, Venetis IE, Chalkias AV, Aletras AH. MRISIMUL: a GPU-based parallel approach to MRI simulations. IEEE Trans Med Imag. 2014;3:607–17.

[2] Xanthis CG, Venetis IE, Aletras AH. High performance MRI simulations of motion on multi-GPU systems. J Cardiovasc Magn Reson 2014;4:16–48.

[3] Smith BR, Huff DS, Johnson GA. Magnetic resonance imaging of embryos: an Internet resource for the study of embryonic development. Computerized Medical Imaging and Graphics 1999;23.1:33-40

Address for correspondence.

Anthony H. Aletras
Aristotle University of Thessaloniki
School of Medicine, Box 323
Thessaloniki, GR-54124
Greece
tel: +30 (2310) 999245, +46 (76) 2822988
fax: +30 (2310) 999263
email: aletras@auth.gr