PIRE 2009 Project Proposal

Student Name: Javier Delgado

Student's School: FIU

Student Email: Javier.delgado@fiu.edu

Student Home Page: http://www.javierdelgado.com

Student Rank: PhD

Student Expected Graduation Date: Fall 2010

Supervisor's Name and Title at FIU/FAU: Drs. Malek Adjouadi and Masoud Sadjadi

Name of the PIRE International Partner's Institution: Universidade Federal Fluminense

Supervisor's Name and Title at the PIRE International Partner's Institution: Dr. Esteban Walter Gonzalez Clua

Project Title: Towards GPU Enabled Weather Research and Forecasting

Problem Statement: *Weather simulations performed using the Weather Research and Forecasting (WRF) software are very computationally intensive. Even with modern clusters costing hundreds of thousands of dollars, it could take several days to perform a high-resolution forecast. Graphical processing units (GPU's) have been shown to be effective at the kinds of computations that are used in WRF. The intention of this project is to port a portion of WRF to work on GPU's using nVidia corporation's Compute Unified Device Architecture (CUDA).*

Motivation and Impact: *There are two motivations for this work. The first is to enhance WRF by allowing it to exploit the computational power of nVidia GPU's. These processors have been shown to be very cost effective to purchase and maintain, which will make running WRF simulations feasible for many institutions that currently do not have the computational infrastructure to run high-resolution simulations and/or require several days to analyze results. The second motivation, which has more to do with my main line of research, is to learn about GPU programming and architecture. I intend to eventually be able to implement a methodology for predicting execution time or applications that run on GPU's. At UFF, I will be interacting with people who can help me acquire the background knowledge that I need for this. In turn, I will be able to show them how to work with WRF, and share information about its inner workings.*

Current Status: *My host advisor, Dr. Clua, has have experience working with GPU's. GPU's have become popular in HPC due to their cost-effectiveness and scalability. Last year, the main developer of WRF, John Michalakes, co-authored and published a paper describing the use of GPU's for executing a specific module of WRF. Very compelling speed benefits were obtained. They are currently looking for contributors to port more WRF models to work on GPU's. One of Dr. Clua's other students, Gabriel Gazolla, and I, are interested in performing this task. I have been working remotely with Gabriel. I have been supporting him with WRF-related issues. He has helped me understand CUDA. We have been looking at the source code of a WRF module that is already CUDA-enabled and studying Michalakes' publication. This will help us in the porting process of the new module(s). In regards to the performance prediction, I have been working on this for approximately 18*

months, including during my PIRE internship last year at Barcelona Supercomputing Center. One paper has been published related to this performance prediction work and another is in the works. So far, our results have been good.

Research Roadmap:

- *July 31 – Familiarized with University operations and work schedule/ethics of all collaborators; distribution of tasks*
- *August 14 – Working CUDA implementation of the WRF SWRAD module*
- *August 21 – Sorted out bugs with the implementation and ensured that simulations execute correctly*
- *August 28 – Series of benchmarks created after running the SWRAD kernel on different resources, including Lincoln cluster at NCSA*
- *September 3 – Did preliminary brainstorming on GPU performance modeling and discussed Gabriel's ideas for his Master's thesis*

Relation to PIRE Core Research Projects: *This project fits into* the CI Application Layer->Hurricane Mitigation Apps *box. This work is directly enhancing WRF by allowing it to run faster on certain hardware platforms. This will make running weather simulations possible for many researchers. For example, a meteorology department with little IT budget and knowledge can deploy a simple GPU cluster, consisting of one or two nodes, which requires minimal maintenance and fits in a relatively small space, with no special cooling needs. Thus, the meteorologists can focus on their own work. The task of predicting execution time on different CUDA-enabled platforms will provide additional CI improvements. It will provide better resource allocation possibilities to components in* the CI Enablement layer *(e.g. metaschedulers).*