Seminar Title: Programming Emerging Heterogeneous Computing: The HARP System
Presenter: Dr. Pedro C. Diniz, University of Southern California, Los Angeles, Calif., USA
Contact: pedro@isi.edu
Host: Prof. Eduardo Marques
Contact: emarques@icmc.usp.br
Dates: June 05 to June 07, 2017
Location: Instituto de Ciências Matemáticas e da Computação (ICMC),
Univ. São Paulo,
São Carlos, SP, Brasil

Participation: *Pending legal and administrative arrangements with Intel, Corp., this seminar is only open to graduate students and staff of the Univ. of São Paulo.*

Motivation:

Heterogeneous computing is seen as the only viable way towards energy-efficiency in the domains of high-performance and embedded computing. Multi-core heterogeneous architectures such as System-on-a-Chip or commercially available systems such as the Intel HARP accelerator-based system provide ample evidence of the industrial trend towards these systems. Still, and despite their obvious architectural benefits, there systems are hard to program as developers need to bridge the gap between traditional programming models (even with multi-core symmetric concurrency) and diverse systems where accelerators are tasked with very specialized tasks.

In this seminar, we will introduce the basic parallel programming models used in today's multicore systems as a natural progression towards the programming models geared towards heterogeneous systems. We will focus on the accelerator-based model used by the popular OpenCL programming environment geared towards GPU- and FPGA-based accelerators architectures. We will present a development environment used by the Intel-HARP system using OpenCL and include several programming projects using this environment. Most importantly, this seminar will provide a hands-on programming opportunity to its participants on how to use this accelerator-based programming paradigm to tackle important real-life problems.

Objectives:

The objectives of this seminar are manifold, namely:

- Introduce the participants to the fundamental aspects of parallel programming from classical multi-processor (symmetric) concurrent computing using traditional parallel systems such as OpenMP or MPI to emerging heterogeneous computing systems that use OpenCL as their main programming paradigm.
- Introduce the Intel HARP computing platform and its development environment as a specific case study of heterogeneous system supporting GPU- and FPGA-based acceleration capabilities.
- Provide the participants of a chance to develop their own hybrid application targeting the Intel-HARP architecture using Intel's OpenCL programming environment using a series of increasingly complex laboratories.

Tentative Schedule:

This seminar is tentatively scheduled for 3 afternoons, for a total of 18 hours thus leaving ample room for the participants to absorb the material from previous day's sessions and prepare for the following days.

The tentative schedule is as follows:

Day #1. **Introduction to Heterogeneous Architectures and Computing**

01:00 – 02:30 PM What are Heterogeneous Computing Architectures. Basic architecture organization and programming paradigms. Accelerator-based organizations and Embedded Multi-core Heterogeneous systems. The role of GPU-based and FPGA-based accelerator and its programming challenges.

02:30 – 04:00 PM Parallel Programming paradigms. Shared-memory and distributed-memory models and their support using OpenMP and MPI. Acceleration-based computing – OpenACC and OpenCL. Examples.

04:00 – 04:30 PM *Break*

04:30 – 06:30 PM Performance modeling and when to off-load computation to accelerators. Speed-up, Amdahl's Law and the cost of communication. Energy considerations and the use of custom or application-specific accelerators.

06:30 – 06:45 PM *Short break*

06:45 – 07:00 PM Recap and Review of basic concepts.

Day #2. **Programming HARP: Intel's OpenCL Programming Language**

01:00 – 03:00 PM The Intel HARP system. Architecture organization and programming model. Overview of OpenCL programming and simulation environment. Examples.

03:00 – 03:30 PM *Break*

03:30 – 04:30 PM Detailed Example 1: GPU acceleration with OpenCL in HARP. Kernel programming and performance analysis.

04:30 – 05:30 PM Detailed Example 2: Custom Acceleration using Arria 10 FPGA in HARP. High-level synthesis and performance analysis.

04:30 – 05:30 PM Detailed Example 2: Custom Acceleration using Arria 10 FPGA in HARP. High-level synthesis and performance analysis.

05:30 – 05:45 PM *Short break*

05:45 – 06:45 PM Project Set-up for each group.

06:45 – 07:00 PM Recap and Review of basic concepts.

Day #3. Group Projects (max. 3 participants each group)

01:00 – 03:30 PM	Presentation by each participant's group (max. 3 elements each, 20 min. each presentation) of its project and strategy to map it to Intel's HARP system. Analysis of Costs and Benefits and major challenges. Discussion.
03:00 – 03:15 PM	*Short Break*
03:15 – 06:15 PM	Open Lab.
06:15 – 06:45 PM	Presentation of results for each group.
06:45 – 07:00 PM	Conclusion.

Target Audience:

Graduate and undergraduate students from Computer Science and Computer engineering as well as graduate students from other domains such as physics or material sciences.

Anticipated Participation:

We expect a healthy audience of 30 participants from the University of São Paulo in São Carlos as well as from its main campus in the city of São Paulo. While traditional the vast majority of participants include Computer Science of Computer Engineering students, students from areas such as Physics and Material Science are also likely to engage in this seminar given their pressing need for high-performance computing and the lack of access to such facilities.

Resources and Facilities:

Participants of this seminar will be granted remote access during the sessions to an Intel HARP-2 where they will be developing their projects. Future access to this machine will be considered on a case-by-case basis and depending on legal arrangements between Intel and each participant's institution. Similarly, materials displayed are subject to NDA agreements.