CUDA Seminar at the University of Patras

For two days (19th and 20th March, 2009) CUDA was the focus of attention at the University of Patras, the third largest university in Greece. During that time, a seminar was held by Richard Membarth, a PhD student of the Hardware/Software Co-Design Chair (Prof. Teich) at the University of Erlangen-Nuremberg, in cooperation with Professor Stefanos Kaxiras, University of Patras. On Thursday, a seminar on CUDA, a new programming paradigm to leverage the potential of current highly parallel graphics cards hardware, was presented at the crowded conference centre. The architecture of current graphic cards as well as the programming model was discussed during the first day.

At the end of the first day, the students had the possibility to register for a hands-on workshop on Friday. Due to the high level of interest, the number of places for the workshop was limited. During the workshop the students learned how CUDA differs from other parallel programming models like OpenMP. They also had the chance to present their own work on CUDA and to experience CUDA for themselves. ■

Load Balancing for Mobile Devices (LoBaMo)

A new paradigm for future mobile devices

With the recent advances in wireless networks and the exponential growth in the usage of multimedia applications, multi-core platforms have emerged as the key for future mobile devices. A new paradigm has simultaneously been created, which we refer to in the ICT-eMuCo project as Load Balancer for Mobile Devices.

In general, a load balancer can be seen as a component that spreads the work between two or more entities such as: CPUs, clustered computer systems, network links or other resources, in order to get optimal resource utilization. Load balancing for mobile devices enables the concurrent and parallel execution of applications and control signals by the efficient execution of multiple threads on processors with multiple cores.

The load balancer for mobile devices enables:

- effective exploitation of the multi-core hardware platform
- reduction of the power consumption
- scalability
- and most importantly from users' perspective: assurance of the expected user experience.

Simultaneously meeting deadlines of real-time transactions and shortening the response time of online transactions is critical in mobile communications as it directly influences the user experience. As is shown in Figure 1, in a mobile device, there are threads generated in the modem subsystem and in the application subsystem. All these threads have native priorities by manual programming assignation or by an operating system assignation. Nevertheless, all the threads have to be distributed to be executed on the multiple cores. The role of the load balancer is to monitor thread execution and distribute the threads on the available cores based on: a) the capacity limit of the unit processing; b) the capacity limit of the communication channels;c) the profiling information of the processing stages; and d) approximate knowledge of the processing demand of the expected workload, which is in part provided by the protocol stack. Therefore, the scheduling policy of the load balancer depends not only on the performance metrics obtained from the hardware platform and applications, but by the

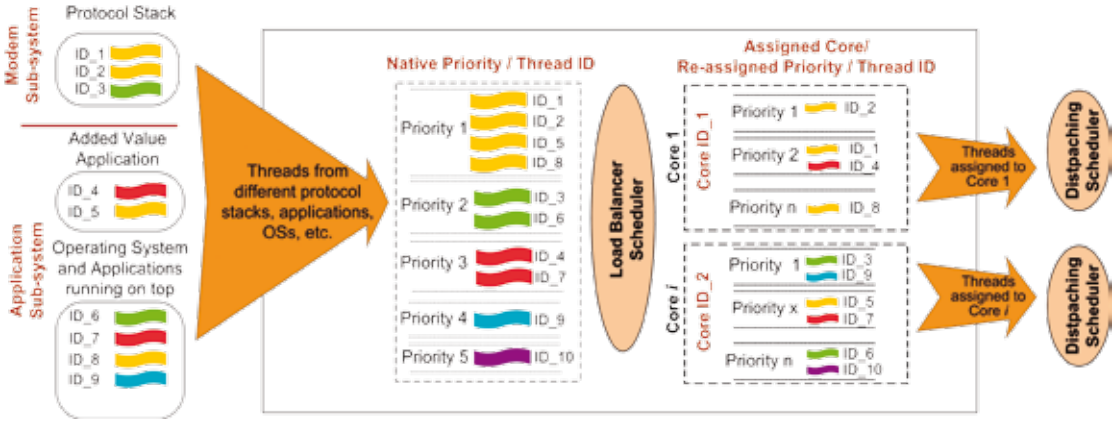
ICT-eMuCo is a European project supported under the Seventh Framework Programme (FP7) for research and technological development (www.emuco.eu)

Project Coordination:
Attila Bilgic
Ruhr-University Bochum
Attila.bilgic@is.rub.de

rich amount of information provided by the protocol stack and the predictable interdependency of the protocol stack processes, which mark the difference between a general purpose load balancer and a load balancer for mobile devices.

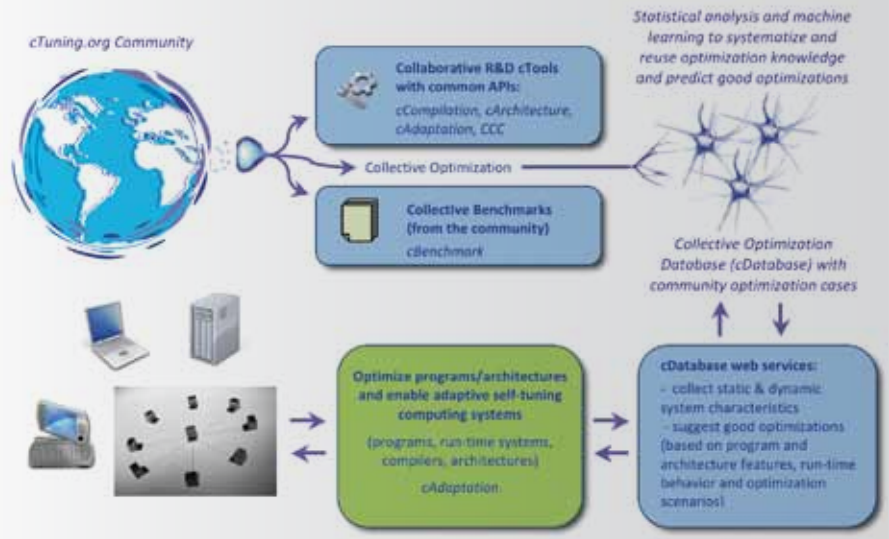
The research group involved with the "load balancer for mobile devices" is part of the eMuCo project and its research provides a natural answer to the requirements of the system platform for future convergence mobile devices. It is expected that the output of this research group will have an impact on market trends for mobile devices. ■

Maria Elizabeth Gonzalez
Institute for Integrated Systems
Ruhr Universität Bochum
D-44780 Bochum, Germany
Elizabeth.gonzalez@is.rub.de



Functionality view of the Load Balancer

Access to the Collective Optimization Database



Public access to the Collective Optimization Database (COD)* has been arranged to provide a common repository for sharing knowledge about program and architecture optimizations. Optimization cases, including program transformations or architecture configurations to improve execution time, code size, power consumption, etc can be submitted either manually through the web form or automatically using Continuous Collective Compilation Framework** through COD web-services.

The COD is intended to improve the quality of academic research by avoiding costly duplicate experiments and providing replicable referable results. It can be combined with the Collective Benchmark/MiDataSets** (collection of publicly available benchmarks and datasets) and provide detailed performance analysis and comparison of different programs, datasets, compilers and architectures.

The Collective Optimization Database is also intended to help end-users and companies optimize their computing systems. The Collective Optimization Database is currently used in the EU FP6 MILEPOST project (http://www.milepost.eu) to automate compiler and architecture design and program optimization based on statistical and machine learning techniques. ■

Project leader: Grigori Fursin,
INRIA Saclay, France
grigori.fursin@inria.fr

(*) Website: http://ctuning.org/cdatabase/
(**) Tools: http://ctuning.org/wiki/index.php/CTools