Séminaire Informatique Haute Performance @ Campus Teratec

Séminaire n°55 du Jeudi 23 Février 2017, 11h, Ter@tec.

Conciliating programmability and efficiency by vectorizing SPMD programs at microarchitecture level

Jeudi 23 Février 2017, Sylvain Collange, Chargé de Recherche à l'INRIA Rennes, accompagné de André Seznec, Directeur de Recherche à l'INRIA Rennes et ACM Fellow, nous présentera ses travaux intitulés *Conciliating programmability and efficiency by vectorizing SPMD programs at microarchitecture level.*

Voici le résumé de cette présentation qui aura lieu dans la salle Paul Gauguin à Ter@tec, à 11h.

Conciliating programmability and efficiency by vectorizing SPMD programs at microarchitecture level

GPUs are now established parallel accelerators for high-performance computing applications and machine learning. Part of their success is based on their so-called SIMT execution model. SIMT binds together threads of parallel applications so they perform the same instruction at the same time, in order to execute their instructions on energy-efficient SIMD units. Unfortunately, current GPU architectures lack the flexibility to work with standard instruction sets like x86 or ARM. Their implementation of SIMT requires special instruction sets with control-flow reconvergence annotations, and they do not support complex control flow like exceptions, context switches and thread migration.

In this talk, I will present how we can generalize the SIMT execution model of GPUs to general-purpose processors and multi-thread applications, and then use it to design new CPU-GPU hybrid cores. These hybrid cores will form the building blocks of heterogeneous architectures mixing CPU-like cores and GPU-like cores that all share the same instruction set and programming model. Besides improving programmability, generalized SIMT enables key improvements that were not possible in the traditional SIMD model, such as simultaneous execution of divergent paths. It opens the way for a whole spectrum of new architectures, hybrids of latency-oriented superscalar processors and throughput-oriented SIMT GPUs.

Sylvain Collange is a Research Scientist at Inria Rennes. His research interests include the architecture of throughput processors and Graphics-Processing Units, compiler optimizations for GPUs and computer arithmetic. His key contributions to these areas include scalarization for GPU architectures, dynamic SPMD vectorization and the Barra GPU simulator.
