Technology Transfer in Computing Systems

D3.15: Individual TTP15 abstract

Project no.:	609491
Funding scheme:	Collaborative project
Start date of the project:	1st September 2013
Duration:	36 months
Work programme topic:	FP7-ICT-2013-10
Deliverable type:	Report
Deliverable reference number:	ICT-609491 / D3.15
WP and tasks contributing:	WP 3 / all
Due date:	01/06/2015
Actual submission date:	21/12/2015
Responsible Organization:	INRIA
Dissemination Level:	Public
Revision:	1.0

TETRACOM D3.15: GOMPPA – GNU OpenMP 4.0 for the Kalray MPPA manycore processor

Albert Cohen, Mircea Namolaru, Romaric Jodin (INRIA)

Benoît Dupont de Dinechin, Pierre de Massas (Kalray SA)

Context of the project

Kalray (http://www.kalray.eu) is the developer of the low-power, high-performance MPPA manycore processor solutions. It was incorporated in July 2008 as a spinoff of CEA and industry, founded after 2 years of incubation. It builds on more than 40 patents and a unique and important know-how base. The company is based in Orsay (Saclay area, south of Paris) and in Montbonnot (Grenoble area).

The MPPA 256 manycore processor implements a distributed on-chip memory architecture, with 16 compute clusters of 17 processors and 4 IO clusters with 4 processors each, connected with a 2D mesh. Two of the IO clusters are directly connected with external DRAM resources. Each compute clusters has access to a small local memory, shared among processors in the cluster.

Please note that this TTP covers partially the period of the whole collaboration project between INRIA and Kalray.

Goal and evolution of the TTP from the initial proposal

The programming models and software environment are key to the success of an hardware architecture. Kalray currently distributes a GCC-based OpenMP implementation for intra-cluster thread-level parallelism. The initial goal of the TTP was to contribute compiler and runtime system technology developed at INRIA to support OpenMP as a programming interface to the entire manycore processor. Special emphasis has been put on OpenMP 4.0 evolutions for heterogeneous computing and the management of multiple memory spaces.

Early on in the technology transfer implementation, the partners of the project realized that a higher customer demand was pushing for OpenCL developments, where the application code base is better estabished than in the still emerging area of high-level accelerator programming with OpenMP 4.0. This was not expected at the time the TTP was drafted, given the dominance of OpenMP in high-performance computing. But the trends happened to be different in the embedded and acceleration usage scenarios of the MPPA. At the same time, difficult choices had to be made to prioritize the evolutions of Kalray's Distributed Shared Memory (DSM) interface underlying both OpenMP and OpenCL, and distribted with the Accesscore 2.x tool suite.

To maximize the business potential of the TTP, it was decided to realign INRIA's contribution to OpenCL rather than OpenMP 4.0, synchronized with Kalray's internal developments. Note that this move does not undermine the general direction of supporting OpenMP 4.0 on the MPPA. It implements a different path, through OpenCL and the associated compiler and runtime system developments as an intermediate stage in the development. It also motivates

further collaborations between INRIA and Kalray on the transfer of parallelizing compilation technology and tools, to provide performance portability to Kalray's future OpenMP 4.0 flow.

Outcome of the TTP

INRIA and Kalray collaborated on retargeting LLVM to th three VLIW core architectures (k1a, k1b 32-bit, k1b 64-bit) of the MPPA manycore processor. In parallel, Kalray contributed out-of-order OpenCL queues to the POCL project, the open source host-side OpenCL runtime upon which its OpenCL flow is based. Based on this collaborative development effort, Kalray was able to evolved its prototype OpenCL environment from a "native task" with GNU-C as the kernel language to OpenCL's comprehensive data parallel and task parallel models and the standard OpenCL-C kernel language. The figure below summarizes the flow and mapping from OpenCL 1.2 (host and kernel source code) to the manycore platform.
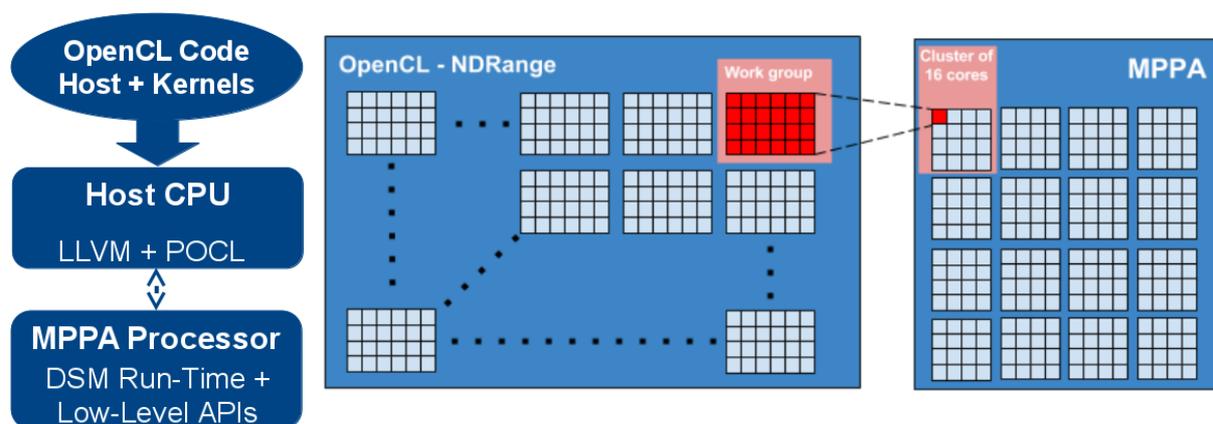
In addition, INRIA provided PENCIL-C benchmarks (including the Polybench-4.1 suite and the SLAMBench k-fusion application ported to PENCIL), and translated these to OpenCL using its PPCG parallelizing compiler. This effort provided a first step towards the support of OpenMP 4.0 performance portability, building upon established compilation technology and on the results of the OpenCL development part of the project. The experiment was also useful to validate many corner cases of the OpenCL implementation and to demonstrate the feasibility of a fully automated flow from sequential code to hardware-accelerated OpenCL running on the MPPA.

Full conformance with OpenCL 1.2 could not be validated during the course of the TTP, but the work was continued at Kalray towards this end using MPPA I/O cores as host CPU. Romaric Jodin, expert engineer at INRIA during the course of the TTP, has been subsequently hired by Kalray to maintain its OpenCL flow.

In parallel, and as explained earlier, the collaboration continues on the TTP's original goal of supporting OpenMP 4.0 on the MPPA manycore processor, using Kalray's ongoing work on the DSM (software cache) and transferring INRIA's expertise on GCC and on the compilation of OpenMP.