2PARMA Project
PARallel PARadigms and Run-time MANagement techniques for Many-core Architectures

HiPEAC Clusters Meeting, Chamonix, April 7th, 2011

Cristina SILVANO
Politecnico di Milano
silvano@elet.polimi.it

List of Project Partners

1. Politecnico di Milano (POLIMI) – Italy (Coordinator)
2. STMicroelectronics (STM) – Italy
3. Fraunhofer Institut for Telecommunications / Heinrich-Hertz Institut (HHT) – Germany
4. Interuniversitair Micro-Electronica Centrum (IMEC) – Belgium
5. Institute of Communication and Computer Systems (ICCS) - Greece
6. RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN (RWTH) - Germany
7. Synopsys (CoWare) - Belgium

Project Coordinator
prof. Cristina SILVANO, POLIMI

Project Technical Manager
prof. William Fornaciari, POLIMI

Start date: January 1st, 2010
Project duration: 3 years
Total effort: 408 PM
EC Contribution: 2.74 M€

www.2parma.eu
Back-ground & side-ground projects

2PARMA

Scientific and Technical Objectives

Main Goals

- Programmability of Many-core Computing Fabrics
- Virtualisation and Continuous Adaptation
- Design Space Exploration
- Runtime Adaptivity

Project Outcomes

- Integrated Compiler Toolchain and OS Layer
- DSE Toolchain
- Run-time Resource Manager

The 2PARMA project focuses on the definition of suitable parallel programming models, instruction set virtualisation, run-time energy/power and resource management policies and mechanisms as well as design space exploration methodologies for Many-core Computing Fabrics.

April 7th, 2011
The 2PARMA project focuses on the flexible family of parallel and scalable computing processors, which we call Many-core Computing Fabric Template, composed of many processing cores interconnected by an on-chip network.

P2012 Cluster

- Cluster Architecture evolutions supporting many applications.
- Heterogeneous Cluster RTL implementation(s):
  - ENCore<\(N\)>, Cluster Controller,
  - Cluster Control, DMA Control,
  - Debug logic, Stream-flow logic,
  - User-defined HWPEs.

User-defined HWPEs

Stream-flow logic
- Inter-HWPE communication,
- L3 or shared Memory \(\leftrightarrow\) HWPEs communication.

Cluster Control
- ENCore Boot, HWPE control, etc…

DMA Control
- L3 \(\leftrightarrow\) Sh. Mem., L3 \(\leftrightarrow\) Stream, Sh. Mem. \(\leftrightarrow\) Stream

Configurable
- (N, EFUs, banking factor, …)
- “Computing Farm”
  - (supplied by “Computing”)
  - For SW PEs and/or intensive control

Debug Multicore Debug

Source: STMicroelectronics

IMEC COBRA Platform

Data Plane

Control Plane

April 7th, 2011

Source: STMicroelectronics
Synopsys Virtual Platforms

1. Abstract model mimicking P2012 in support of CBSE methodology

2. ARM based multi-core platform supporting Linux OS for design tool validation
   - Integration with DSE tools
   - Investigate RTRM support

April 7th, 2011

2PARMA Target Applications

- Scalable Video Coding (SVC) – HHI
- Cognitive Radio
  - Physical Layer – RWTH-ISS
  - MAC Layer – RWTH-INETs
  - Reconfigurable Radio – IMEC
- Multi-view Image Processing - IMEC
Applications/Architecture Integration

Run-time Management and Virtualisation

OS Integration Layer

The 2PARMA Run-Time Resource Manager

Overall view

Legend
- SW Interface (API)
- SW/HW Meta-data

April 7th, 2011
WP2: Programmability of Parallel Computing Systems to Exploit Task/Data Parallelism

Partners role

**RWTH:** Component-based SE toolchain to support applications design.

**POLIMI:** OpenCL compilation toolchain. Dynamic compilation to adapt parallelism to system resources.

**STM:** Operating System Integration. Development of device drivers to support runtime management.

Applications Description

Component-Based Toolchain

Parallel Functional Specification

Native Code

Portable Bytecode

Instruction Set Virtualisation

Component Interfaces and SW Channels

Parallelism to System Resources

Computing Fabric Device Drivers

Many-Core Computing Fabric

STM: Operating System Integration. Development of device drivers to support runtime management
Mapping OpenCL to Computing Fabrics

- Analysis of OpenCL programming model (strengths and weaknesses) vs P2012 Computing Fabric
- Specification of Computing Fabric-specific extensions to OpenCL
- Proposed OpenCL extensions targeting Computing Fabrics
- OpenCL front-end compilation toolchain for LLVM to be publicly released soon
- Dynamic compilation to adapt parallelism to system resources

April 7th, 2011  FP7-248716-2PARMA Project

Nucleus Methodology

Transceiver Description

Nuclei

PEs

HW Platform

Nucleus Project within UMIC research cluster at RWTH Aachen University
Nucleus Methodology

Transceiver Description

Transceiver Description

Nucleus Library

Compiles

Nucleus Library

N1 N2 N7 Non N Tasks N5 N1 N2 N7 N5 NN

Tasks

Nucleus Library

N1 N2 N7 N5 NN

Board Support Package

HW Platform

Nucleus Project within UMIC research cluster at RWTH Aachen University

Component-based Software Engineering Toolchain

User inputs

Application domain specific knowledge required

Config & constraints

Forem description

Libraries

Nucleus Library

Board Support Package

Flavor Library

Platform Description

2PARMA IP - WP4

2PARMA IP - WP4

STM P2012 Platform

STM P2012 SDK required

Synopsys Virtual Platform

Synopsys IP required

PA, VPU library, ...

Nucleus Mapper

Code Generator

2PARMA IP - WP4

STM P2012 Platform

STM P2012 SDK required

Application domain specific knowledge required

Nucleus Project within UMIC research cluster at RWTH Aachen University
WP3: Co-exploration of Architectural Platforms and Programming Models

Partners role

SYNOPSYS: EDA tools and virtual platform provider
IMEC: Support for run-time task mapping and scheduling
POLIMI: Design Space Exploration for supporting run-time system management

HHI NoCTrace Architecture

- NoCTrace backend
  - collects data
- NoCTrace frontend
  - processes data
  - manages data (persistency)
  - presents data (GUI)

April 7th, 2011
FP7-248716-2PARMA Project
HHI NoCTrace Backend

- Extract/record program counter and transaction data in SystemC kernel
- Perform automatic design analysis to get to know where to place the probes

Automatic Multi-Objective Design Space Exploration

Source: Politecnico di Milano - MPEG2 decoder - MIPS-based multiprocessor - 70nm tech. node - 10 frames
Automatic Multi-Objective Design Space Exploration

Design-time Exploration to support Run-time Resource Management

- Based on the **design-time exploration**, we derive a set of Pareto **operating points** corresponding to power, resources (number of cores) and QoS (average time per frame).
- The operating points will be used by the **Run-time Resource Manager** to achieve QoS requirements (average time per frame) while meeting overall resources (number of cores) and minimizing power consumption.
Run Time Management of multi-core platform

- The system state can change due to some events:
  - A new application executed or
  - QoS requirements modified
Run-time resource management (RTRM): Overview

RTRM offers an integrated solution and the mechanisms to perform optimal WM selection at run-time.

What is a working mode?

What is a working mode?

Platform Parameters

Application Parameters

Adaptive DMM Parameters (example)

Design Space Exploration

Platform Params

App. Params

Adaptive DMM Params

<table>
<thead>
<tr>
<th>ATXMS</th>
<th>Stream size</th>
<th>Stream width</th>
<th>Multi-core</th>
<th>ASR</th>
<th>Deep arch.</th>
<th>Loading</th>
<th># lists</th>
<th>P1 policy</th>
</tr>
</thead>
</table>

April 7th, 2011
Example of Intra-Heap MTh-DMM Design Space

Intra-Thread Level Design Space

E. Block Structure Decisions

- Block Sizes:
  - One Size
  - Many Sizes

- Block Recorded Info:
  - Size
  - Status
  - Pointers
  - SLL
  - DLL

- Block Tags:
  - Non-Specific
  - Header
  - Boundary Tags

- Block Structure:
  - SLL
  - DLL
  - Dynamic Array

F. Pool Organization Decisions

- Pool Structure Based on Block Size:
  - One Pool per Size
  - Single Pool

- Pool Structure Based on Blocks Order:
  - Single Pool

- Pool Structure Based on Block Address:
  - Single Pool

www.2parma.eu