System-Level Design Space Exploration for Heterogeneous Parallel Dedicated Systems

Dott. Ric. Ing. Luigi Pomante
luigi.pomante@univaq.it

Università degli Studi dell’Aquila
Center of Excellence DEWS
ITALY
Overview

• Introduction
• System Specification
• Target Architecture
• Design Space Exploration
Introduction
Introduction

• Electronic digital systems based on heterogeneous parallel architectures have been recently exploited for a wide range of application domains
  – They often include a combination of several heterogeneous single/multi-core processors, memories, and a set of interconnections among them

• Moreover, such systems are often dedicated systems
  – Digital electronic systems with application-specific HW/SW architecture

• When dedicated systems are also heterogeneous parallel ones they are so complex that the HW/SW co-design methodology plays a major role in determining the success of a product
  – Electronic Design Automation (Electronic System Level)
Diagram of the design flow for system-level and algorithm-level analysis.

**System-level Flow**
- Specification
- Functional Simulation
- Co-Analysis Co-Estimation
- Design Space Exploration
  - HW/SW Partitioning and Architecture Definition
  - Timing Co-Simulation
- Workload & Bandwidth Estimation
- Load & Bandwidth Constraints
- Profiling
- Communication
- Concurrency

**Legend:**
- Flow steps
- Designer Interaction
- Exchanged Data

**Algorithm-level Flow**
- DHPS
Introduction

• The focus of this talk is on a Design Space Exploration (DSE) approach that, starting from the system specification and related requirements, would be able to automatically suggest to the designer
  – an HW/SW partitioning of the given system specification
  – an heterogeneous parallel architecture
  – a mapping of the partitioned entities onto the proposed architecture able to satisfy imposed requirements
System Specification
System Specification

• The entry point of the proposed co-design flow is the behavioral specification of the system functionalities
  – The system specification is based on CSP MoC (*Communicating Sequential Processes*) and described by means of an executable specification language (e.g. OCCAM, HandelC, **SystemC**, etc…)

• An internal model of representation is used to allow a proper tool-chain to make automatic analysis and transformations
  – *Procedure Interaction Graph (PING)*
    • It is a formalism that provides information about the relationships among procedures (communication, synchronization and concurrency issues)
      – It is based on the well-known *Procedural Call Graph*
System Specification

[Diagram showing a network of processes and channels labeled P1, P2, P3, P4, EN_IN1, and EN_OUT1.]
Target HW Architecture
Target HW Architecture

- The target HW architecture is an heterogeneous parallel one with shared and distributed memory
  - It is composed of proper interconnections of some instances of different basic elements called *Basic Block* that represent the minimal computation, storage and communication units in the system

PUi
- GPP, DSP: $\epsilon$, $L_{MAX}$
- SPP: $\epsilon$, $G_{eq_{MAX}}$, $C_{Cell_{MAX}}$ / $LUT_{MAX}$

LM
- $KBD_{MAX}$
- $KBC_{MAX}$

IIL
- IIL: $BW_{MAX}$, $N_{MAX}$

ECU
- $EIL_{K}$: $BW_{MAX}$, $N_{min}$, $N_{MAX}$, $CC_{MAX}$, $\epsilon$
Target HW Architecture

- Given some instances of BBs and interconnecting them by means of some instances of EILs it is possible to define a feasible dedicated heterogeneous parallel architecture on which the system functionalities can be mapped to
  - Such an architecture is represented by means of a hierarchical architecture graph
Design Space Exploration
Design Space Exploration

1st Phase

- Annotated Specification (PING)
  - Partial Architecture
    Number and type of processors/cores
  - HW/SW Partitioning
  - Mapping

PAM1

2nd Phase

- BB Interaction Graph
  - Final Architecture
    Number and type of processors/cores
    Number and type of interconnection links
    Topology
  - HW/SW Partitioning
  - Mapping

PAM2
Design Space Exploration

- The input is a PING annotated by several metrics
  - load imposed by each procedure to a single GPP under a time-to-completion constraint: $l$
  - bandwidth needed to communicate with other procedures while fulfilling a time-to-completion constraint: $b$
  - size for HW/SW implementations: $s$ (KBD and KBC bytes, and Geq/cells/LUT)
  - affinity of each procedure towards a set of processor classes (GPP, DSP and SPP): $a$
Design Space Exploration

• 1\textsuperscript{st} phase
  – From Annotated PING to determine number/type of BB/PU trying to minimize a cost function by means of a genetic approach
  • Each individual represents a possible mapping/architecture item
Design Space Exploration

• 2\textsuperscript{nd} phase
  – The starting point of the second phase is an internal model used to represent the partial system obtained at the end of the first phase
  • BB Interaction Graph (BING)
    – Such a model is used to determine number and type of EILs between BBs that minimizes a proper cost function by means of a genetic approach
Design Space Exploration

- **2\(^{nd}\) phase**
  - The approach is very similar to the 1\(^{st}\) phase but with different system-level metrics, individuals structure and cost function
  - Each individual of the population represents a possible interconnections/topology item
Design Space Exploration

• Starting from a CSP Specification, let’s follow the design flow…
Thanks!