System-Level Design Space Exploration for HMPSoC

Luigi Pomante

luigi.pomante@univaq.it

University of L'Aquila – ITALY

Center of Excellence DEWS

Overview

- Introduction
- System-Level Design Flow
- Target Architecture
- Application Specification
  - Internal Models
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- Conclusions

Overview

- **Introduction**
- System-Level Design Flow
- Target Architecture
- Application Specification
  - Internal Models
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- Conclusions

Introduction

- HMPSoC have been recently exploited for a wide range of application domains, for both the embedded and the general purpose products
  - Such systems can include several and different processors, memories, dedicated *ICs* and a set of interconnections between them
- They are so complex that the design methodology plays a major role in determining the success of a product

Introduction

- For this, in the past few years, a discrete number of research works has focused on system-level co-design of HMPSoC
  - Each of them has proposed a different approach to the design space exploration but all of them always rely on some designer experience to define some aspects of the target architecture
    - In particular, the definition of the communication architecture is often only an input to the design flow (typically imposed by a platform-based approach)

Introduction

- This talk presents a system-level design space exploration strategy that, starting from the application specification and related constraints, would be able to suggest to the designer
  - an HMPSoC reference architecture
  - an HW/SW partitioning of the given application
  - an allocation of the partitioned entities on the proposed HMPSoC
    - Moreover, by means of abstract modeling, the approach try to preserve general applicability and feasibility of the proposed solutions

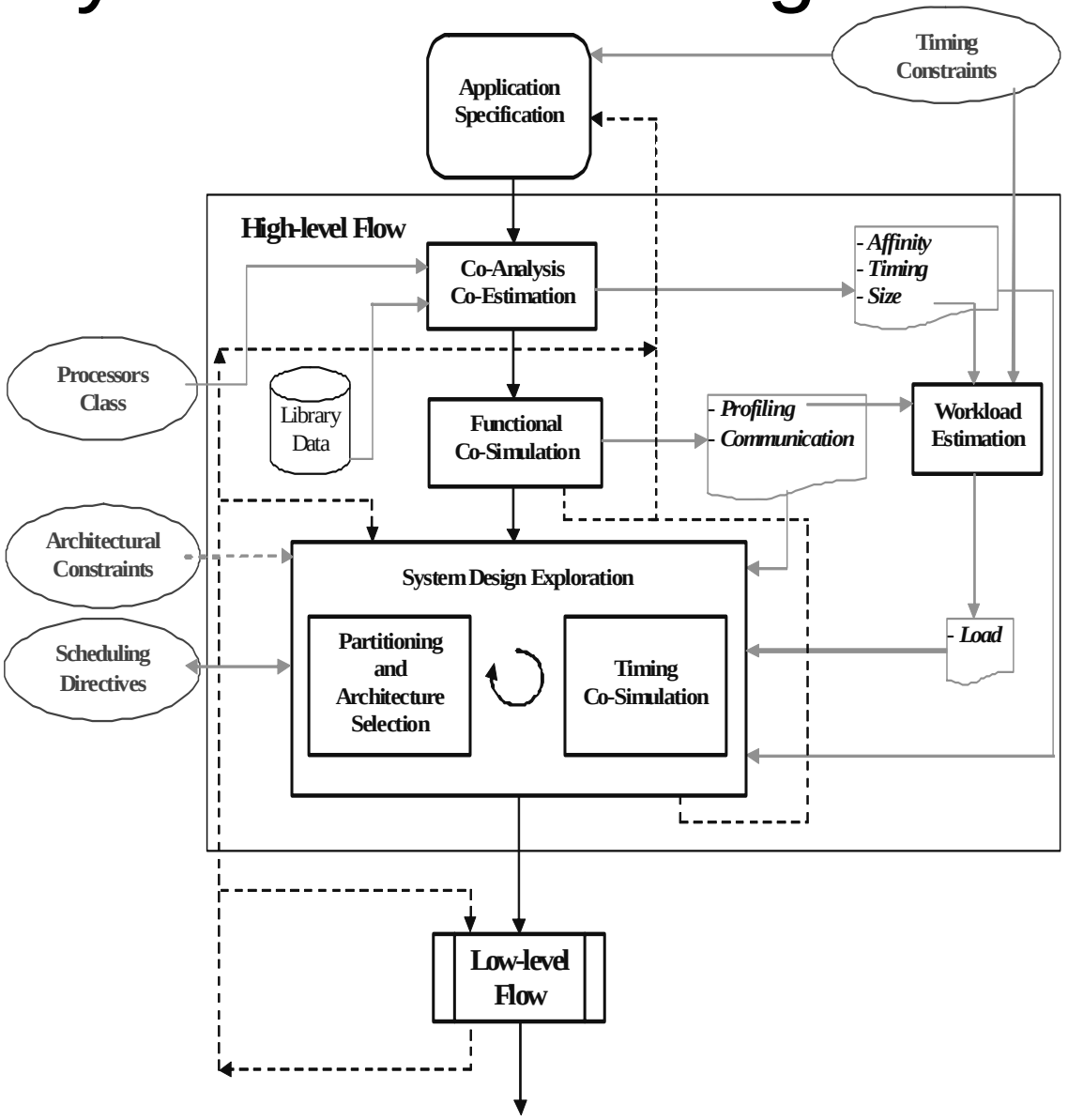
Introduction

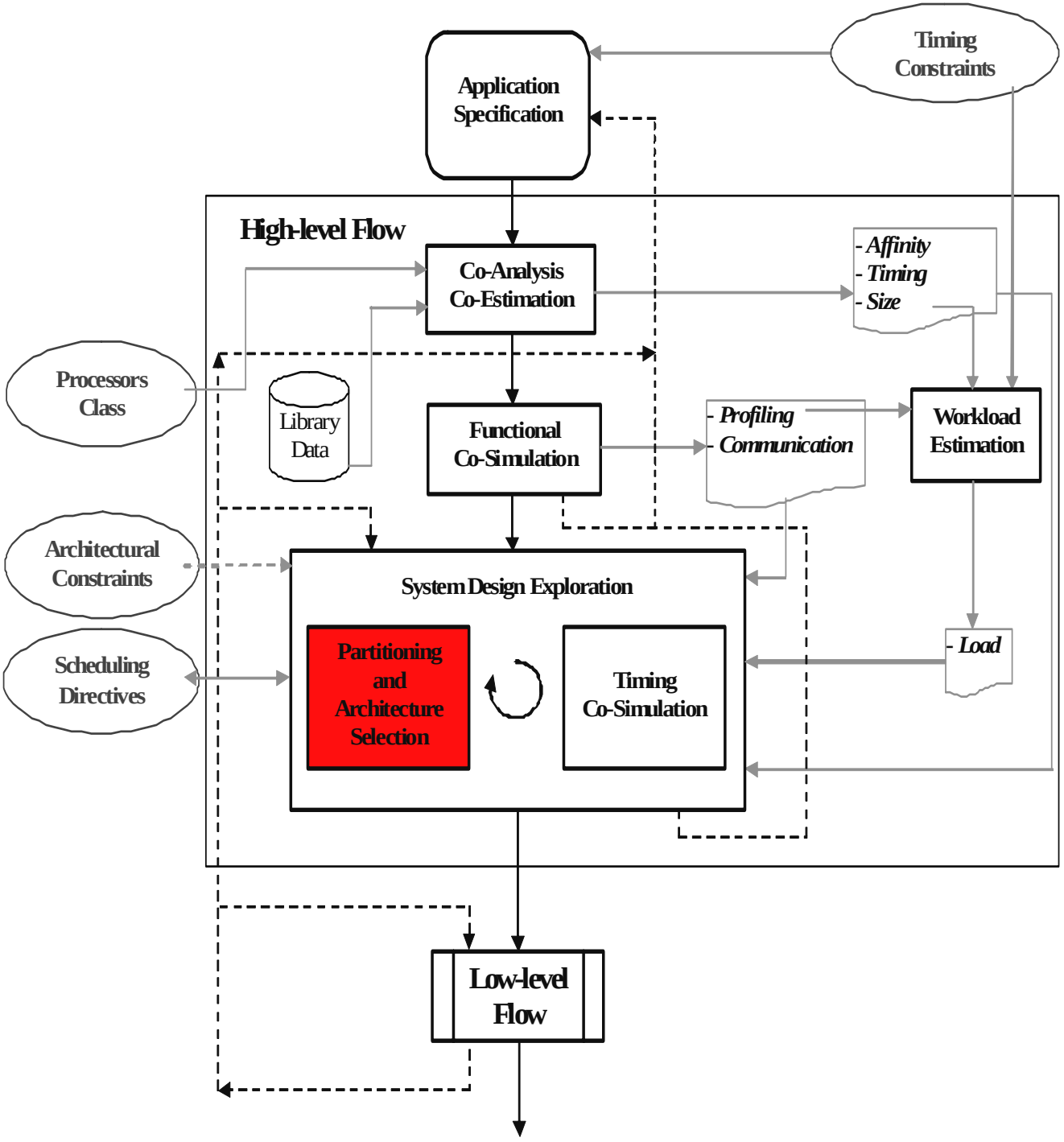
- In particular, this talk focuses on
  - *Modeling Issues*
    - with particular emphasis on model of computation and related “internal models” of representation
  - *System-Level Metrics*
    - used to identify suitable HMPSoC architectures by analyzing the application specification
  - *Design Space Exploration*
    - by means of a genetic algorithm that exploits metrics and some profiling/estimations
      - The approach has been partially validated by checking the consistency with respect to some case studies

Overview

- Introduction
- **System-Level Design Flow**
- Target Architecture
- Application Specification
  - Internal Models
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- Conclusions

System-Level Design Flow



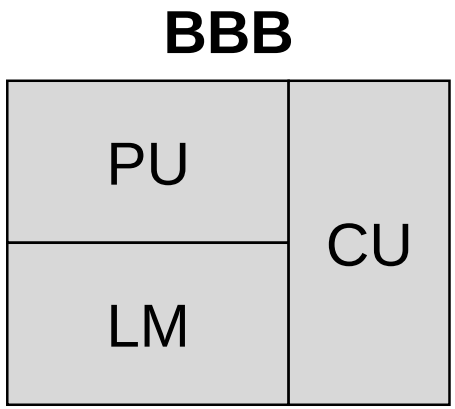

Timing Constraints
Application Specification
High-level Flow
Co-Analysis Co-Estimation
- Affinity
- Timing
- Size
Processors Class
Library Data
Functional Co-Simulation
- Profiling
- Communication
Workload Estimation
Architectural Constraints
System Design Exploration
Scheduling Directives
Partitioning and Architecture Selection
Timing Co-Simulation
- Load
Low-level Flow

Overview

- Introduction
- System-Level Design Flow
- **Target Architecture**
- Application Specification
  - Internal Models
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- Conclusions

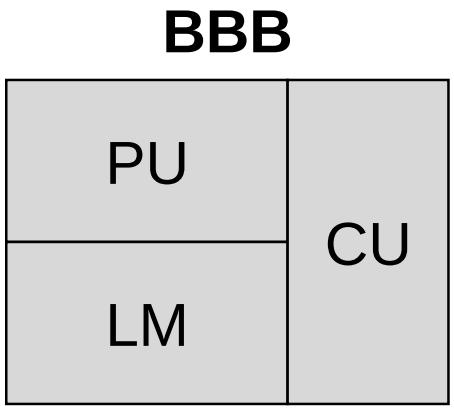
Target Architecture

- It is an interconnection of some instances of heterogeneous *Basic Building Block*
  - BBB represent the minimal computation, storage and communication unit in the system
    - Distributed memory architecture
  - Each BBB is composed of three main elements
    - The Processing Unit (PU)
    - The Local Memory (LM)
    - The Communication Unit (CU)
      - Number and type of possible isyances of interconnection links (IL)



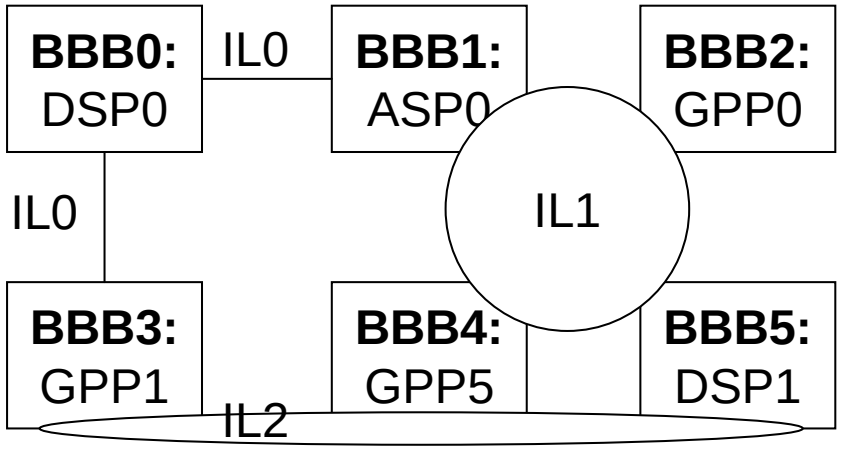
Target Architecture

- Basic Building Block
  - Processing Unit
    - GPP, DSP, [uC]: {€, L_{MAX}}
    - AS[I]P: {€, G_{eq_MAX}}
  - Local Memory
    - RAM, ROM: {€, KB_{MAX}}
  - Communication Unit
    - Possible Interconnection links
      - IL_i:{BW_{MAX}, N_{MIN}, N_{MAX}, €}



Target Architecture

- Given several instances of BBB and interconnecting them by means of some instances of the available IL it is possible to build a feasible application-specific architecture on which the application can be mapped on

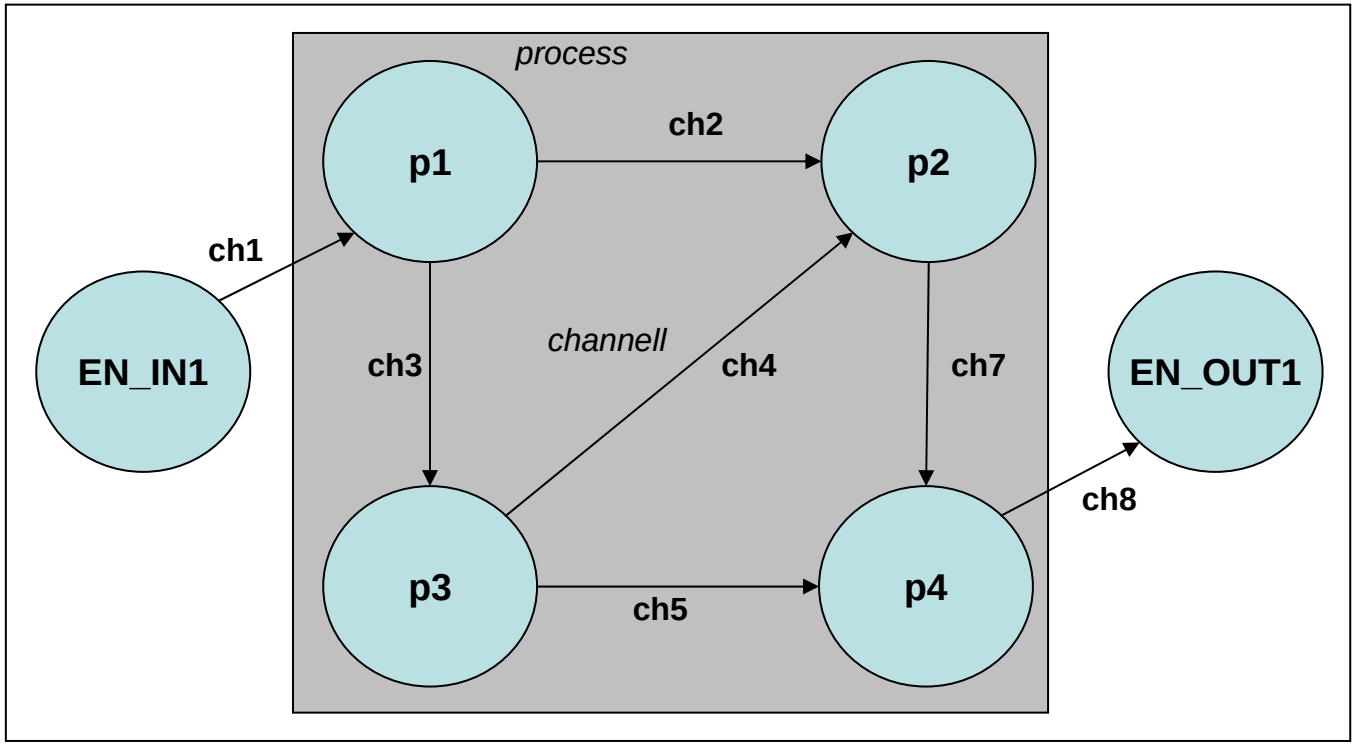


Overview

- Introduction
- System-Level Design Flow
- Target Architecture
- **Application Specification**
  - **Internal Models**
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- Conclusions

Application Specification

- System-Level Model of Computation
  - *Communicating Sequential Processes*
    - e.g.



Application Specification

- System-Level Specification Languages
  - Some *homogeneous* languages able to support CSP
    - OCCAM
    - HandelC
    - SystemC
    - Simulink
    - …

Application Specification

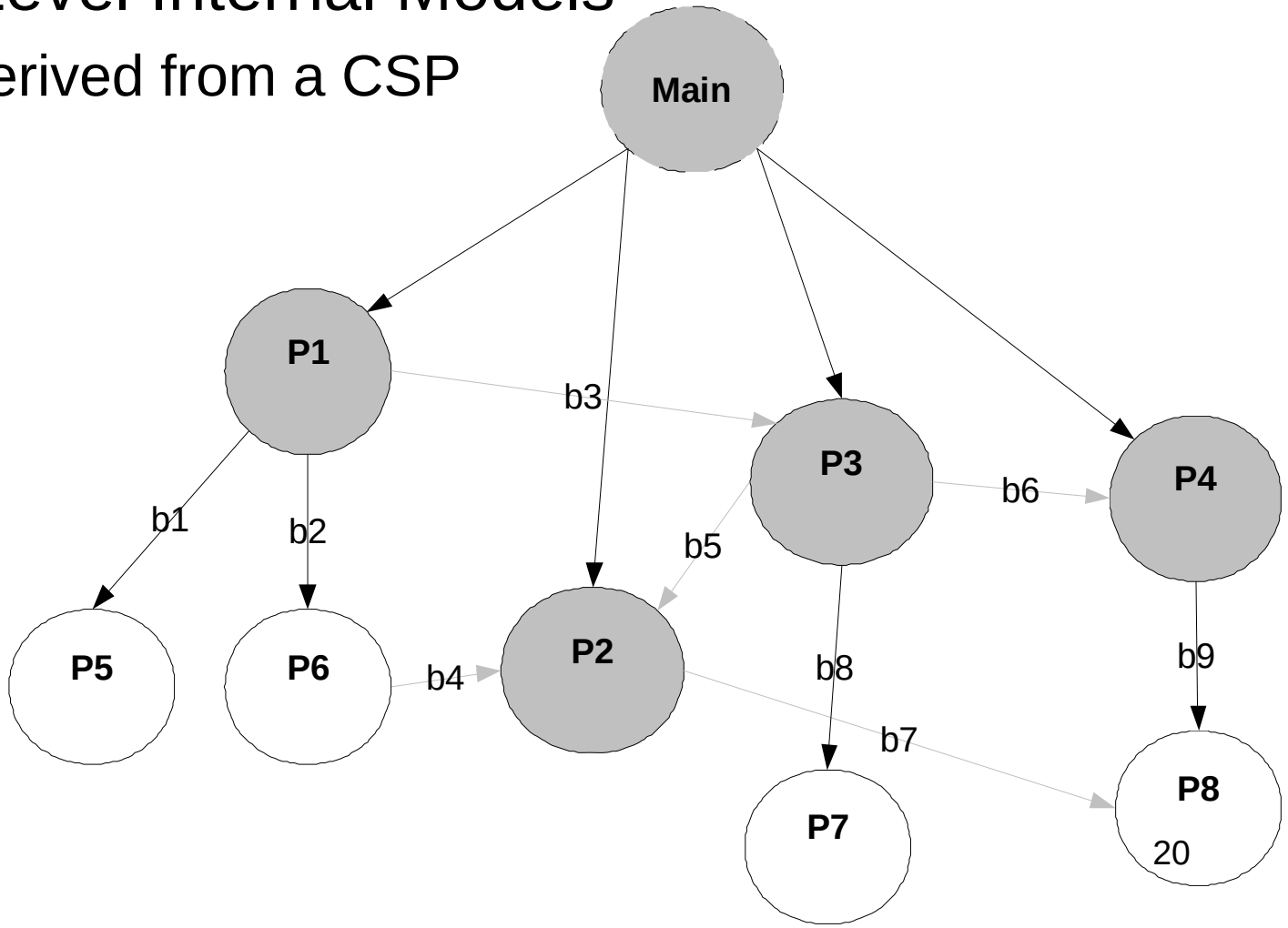
- System-Level Internal Models
  - Different levels of details for IM
    - Stament-level IM
      - Dependent on specification languages and parsers
    - Procedure-level IM
      - Independent from specification languages

Application Specification

- System-Level Internal Models
  - The procedure-level internal model used in this work during the design space exploration to represent the CSP is called *Procedural Interaction Graph (PING)*
  - The PING is
    - based on the well-known *Procedural Call Graph*
    - able to capture needed information from imperative, possibly *object-oriented*, specification

Application Specification

- System-Level Internal Models
  - PING derived from a CSP
    - e.g.

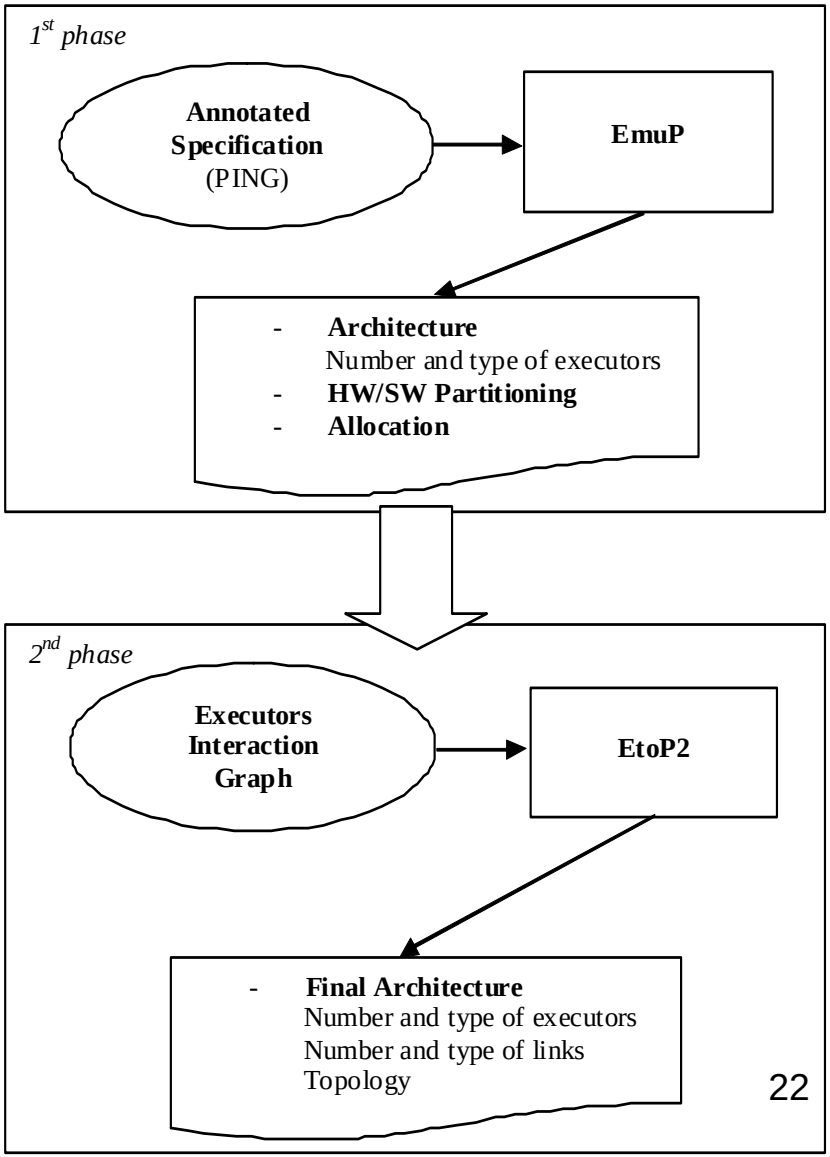


Overview

- Introduction
- System-Level Design Flow
- Target Architecture
- Application Specification
  - Internal Models
- **Design Space Exploration**
  - **Exploration Strategy**
  - **System-Level Metrics**
- Conclusions

Design Space Exploration

- The goal of DSE is the automatic identification of both the architecture and the mapping that optimizes a given CF

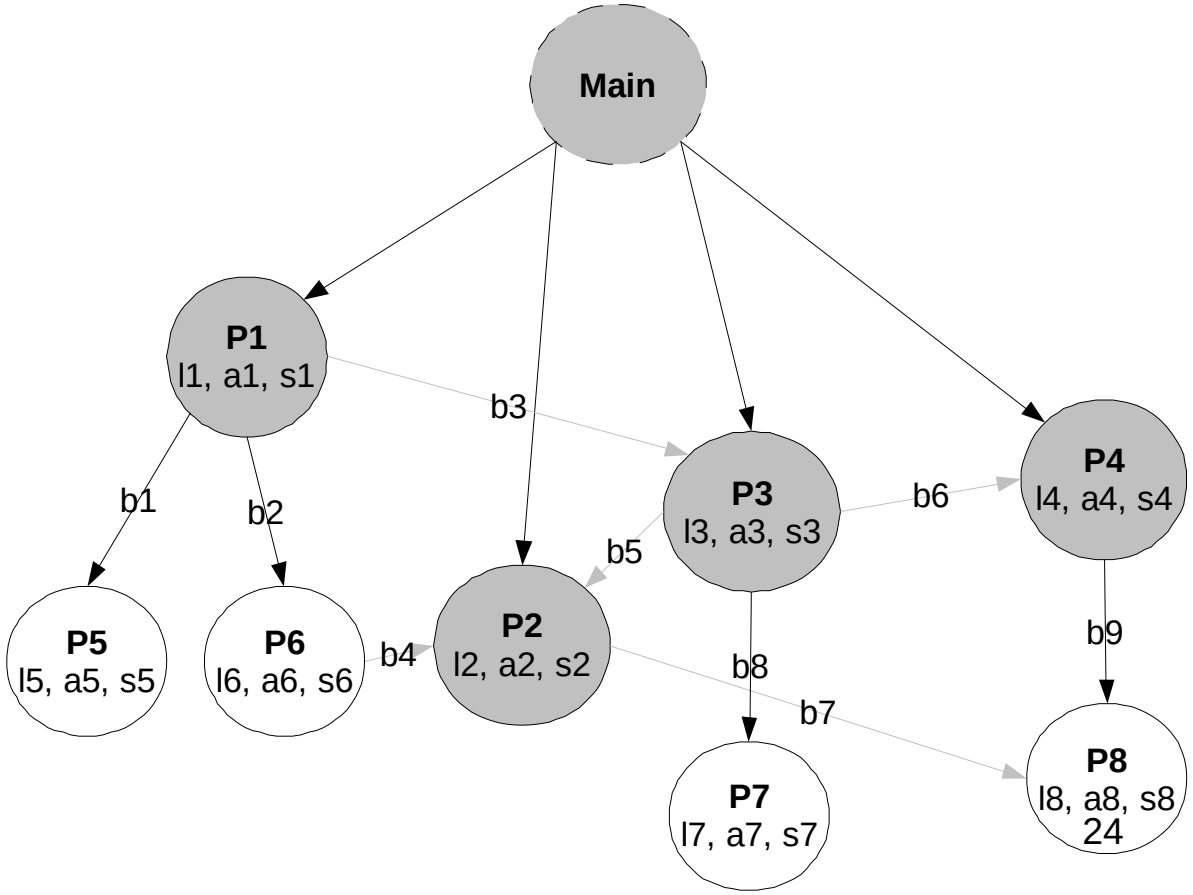


Design Space Exploration

- Each procedure in the PING is supposed to be annotated by different metrics
  - Load imposed to a single GPP
    - l_i
  - Bandwidth needed to communicate with other procedures
    - $b_{i,}$
  - Size
    - s_i: KB or G_{eq}
  - Affinity towards GPP, DSP and ASP
    - a_i

Design Space Exploration

- System-Level Internal Models
  - Annotated PING derived from CSP
    - e.g.



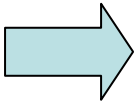
Design Space Exploration

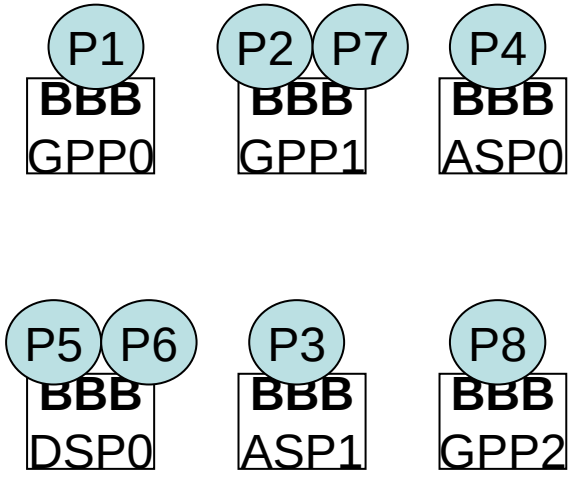
- 1st phase
  - Starting from PING to determine number and type of BBB/PU…
    - Goals
      - Minimize the cost of the set of BBB/PU
        - » Max number for each kind of PU could be provided by the designer
      - Exploit the potential parallelism expressed in the PING
      - Keep the load of each PU near but under its L_{MAX}
      - Minimize communications between different BBB/PU
      - Keep the used size near but under KB_{MAX} or G_{eq_MAX}
      - Exploit affinity between BBB/PU and the procedures

Design Space Exploration

- 1st phase
  - …minimizing a cost function by means of a genetic approach

P1	P2	P3	P4	P5	P6	P7	P8
GPP	GPP	ASP	ASP	DSP	DSP	GPP	GPP
0	*1*	*1*	*0*	*0*	*0*	*1*	*2*





Design Space Exploration

- System-Level Metrics (1st phase)
  - *Affinity Index* (I_A) [0, 1]
    - Matching between the functionalities and the processing elements on which they have been allocated
  - *Load Indexes* (I_{Lsw}, I_{Lhw}) [0, 1]
    - Balancing of the workload over the available HW and SW processing elements
  - *Communication Index* (I_C) [0, 1]
    - Exchanged data size between functionalities allocated on different processing elements
  - *Physical Cost Index* ($I_\$$) [0, 1]
    - Cost of the solution with respect to the most expensive one

Design Space Exploration

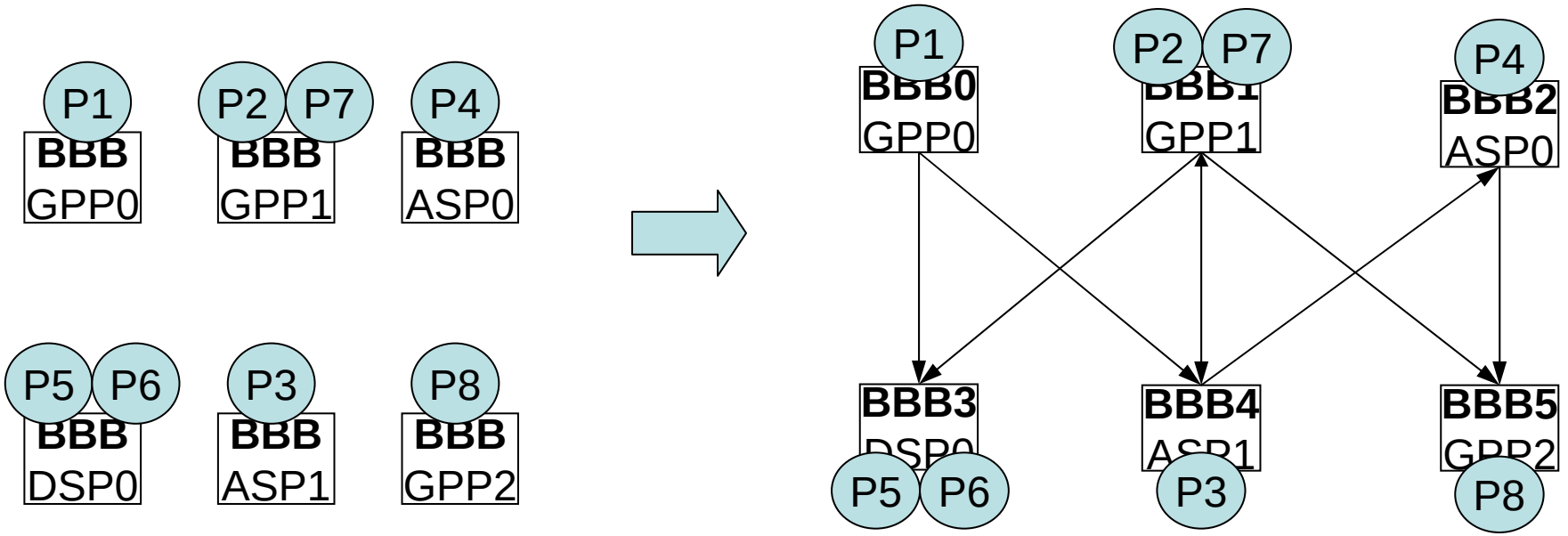
- System-Level Metrics (1st phase)
  - By combining the metrics a cost function has been built to compare different solutions

$$CF = w_A \cdot I_A + w_{Lsw} \cdot I_{Lsw} + w_{Lhw} \cdot I_{Lhw} + w_C \cdot I_C + w_\$ \cdot I_\$$$

  - Affinity and load parameters
    - Tend mainly to separate the functionalities to balance the load and exploit the processing elements features
  - Communications and physical cost
    - Tend to keep together the functionalities to minimize the number of processing elements

Design Space Exploration

- 2nd phase
  - From the results of the 1st phase and the PING it is possible to build a BBB Interaction Graph (BING)
    - The BING is a model used to represent the partial system at the end of the first phase
      - Each edge will present a proper B=$f(b_i)$



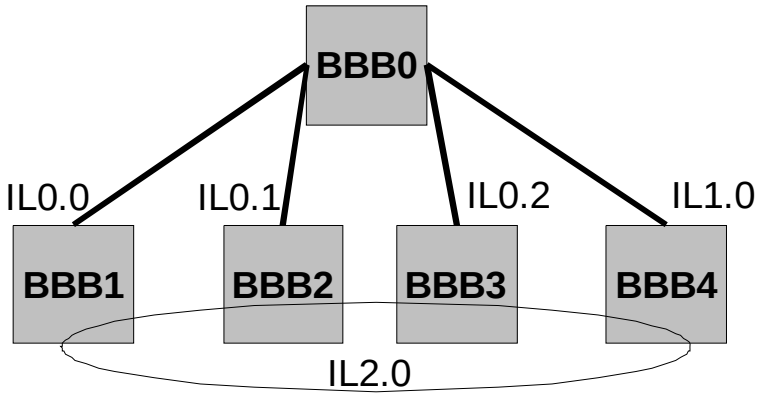
Design Space Exploration

- 2nd phase
  - Starting from BING to determine number and type of IL between PU…
    - Goals
      - Minimize cost of Il_i
        - » Max number of instances for each IL can be specified by the designer
      - Keep the bandwidhth of each IL_i under but near BW_{MAX}
      - Keep the number of executors of each IL_i under but near N_{MAX} (and > N_{MIN})
      - Satisfy eventual constraints on the latency
      - Keeping feasibility while respecting CU characterization

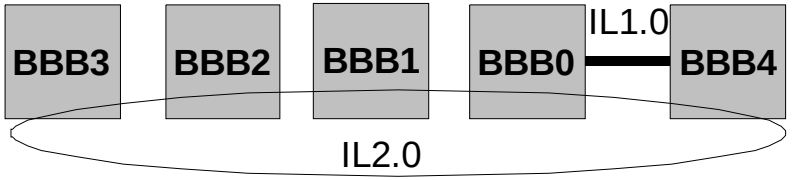
Design Space Exploration

- 2nd phase
  - …minimizing a cost function by means of a genetic approach

BBB0					
BBB1	**IL0.0**				
BBB2	**IL0.1**	**IL2.0**			
BBB3	**IL0.2**	**0**	**0**		
BBB4	**IL1.0**	**IL2.0**	**IL2.0**	**0**	
	BBB0	*BBB1*	*BBB2*	*BBB3*	*BBB4*



BBB0					
BBB1	**IL2.0**				
BBB2	**IL2.0**	**IL2.0**			
BBB3	**IL2.0**	**0**	**0**		
BBB4	**IL1.0**	**IL2.0**	**IL2.0**	**0**	
	BBB0	*BBB1*	*BBB2*	*BBB3*	*BBB4*



Design Space Exploration

- System-Level Metrics (2nd phase)
  - *Saturation Index* (I_B) [0, 1]
    - Respect of max bandwidth offered by the IL
  - *Exploitation Index* (I_E) [0, 1]
    - Respect min/max number of BBB that can use a single IL instance
  - *Physical Cost Index* ($I_€$) [0, 1]
    - Cost with respect to the expensive solution
  - *Feasibility Index* (I_F) [0, 1]
    - A pair of CU should be able to manage at least a common IL in order to allow the related BBB to directly communicate

Design Space Exploration

- System-Level Metrics (2nd phase)
  - CU characterization for feasibility
    - e.g.

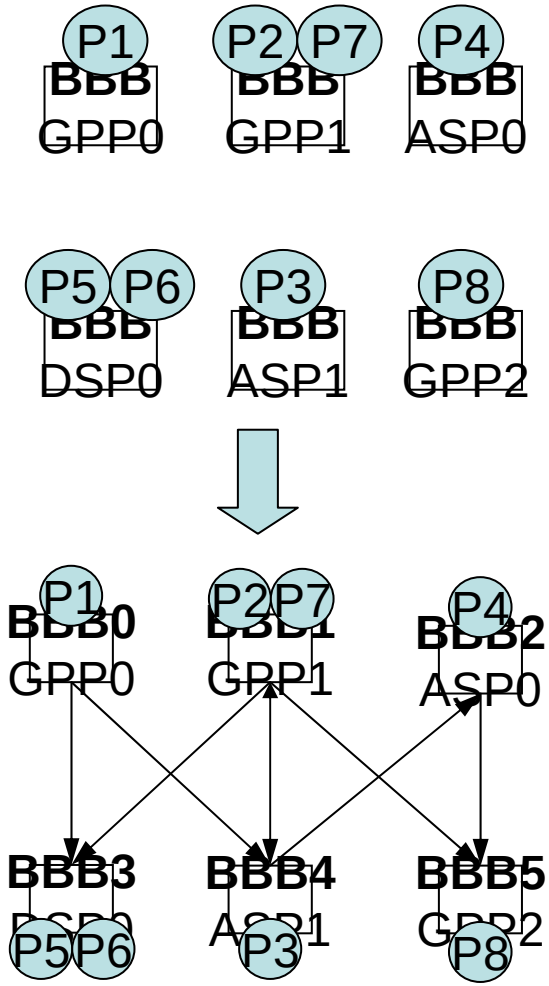
	IL0	IL1	IL2	IL3	IL4
BBB0-CU	X	X	X		X
BBB1-CU	X	X	X	X	X
BBB2-CU	X		X		X
BBB3-CU	X		X	X	
BBB4-CU		X	X	X	X

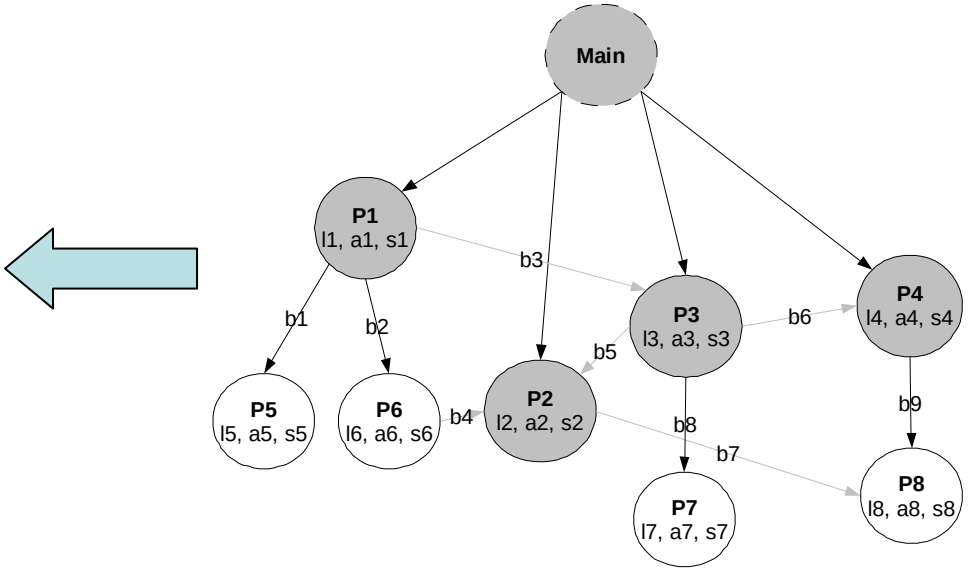
  - Cost Function
    - Used to compares different proposed solutions identifying the one that better tradeoffs different parameters

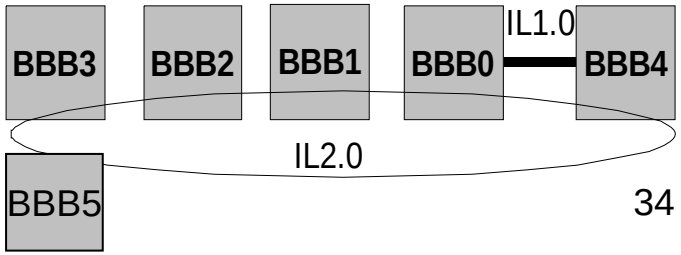
$$CF = w_B \cdot I_B + w_E \cdot I_E + w_€ \cdot I_€ + I_F$$

Design Space Exploration

- Summary







Overview

- Introduction
- System-Level Design Flow
- Target Architecture
- Application Specification
  - Internal Models
- Design Space Exploration
  - Exploration Strategy
  - System-Level Metrics
- **Conclusions**

Conclusions

- This talk has presented a System-Level Design Space Exploration strategy for HMPSoC
  - It has presented a methodology able to propose an HW/SW partitioning of the specification, mapping this one onto an automatically selected architecture
    - The methodology has been only partially validated
- The experimental results are encouraging and justify further research efforts in this direction may be also thanks to your comments…