Classifying and Evaluating Performance-relevant Parameters for Reconfigurable Processors

Lars Bauer, Muhammad Shafique, and Jörg Henkel

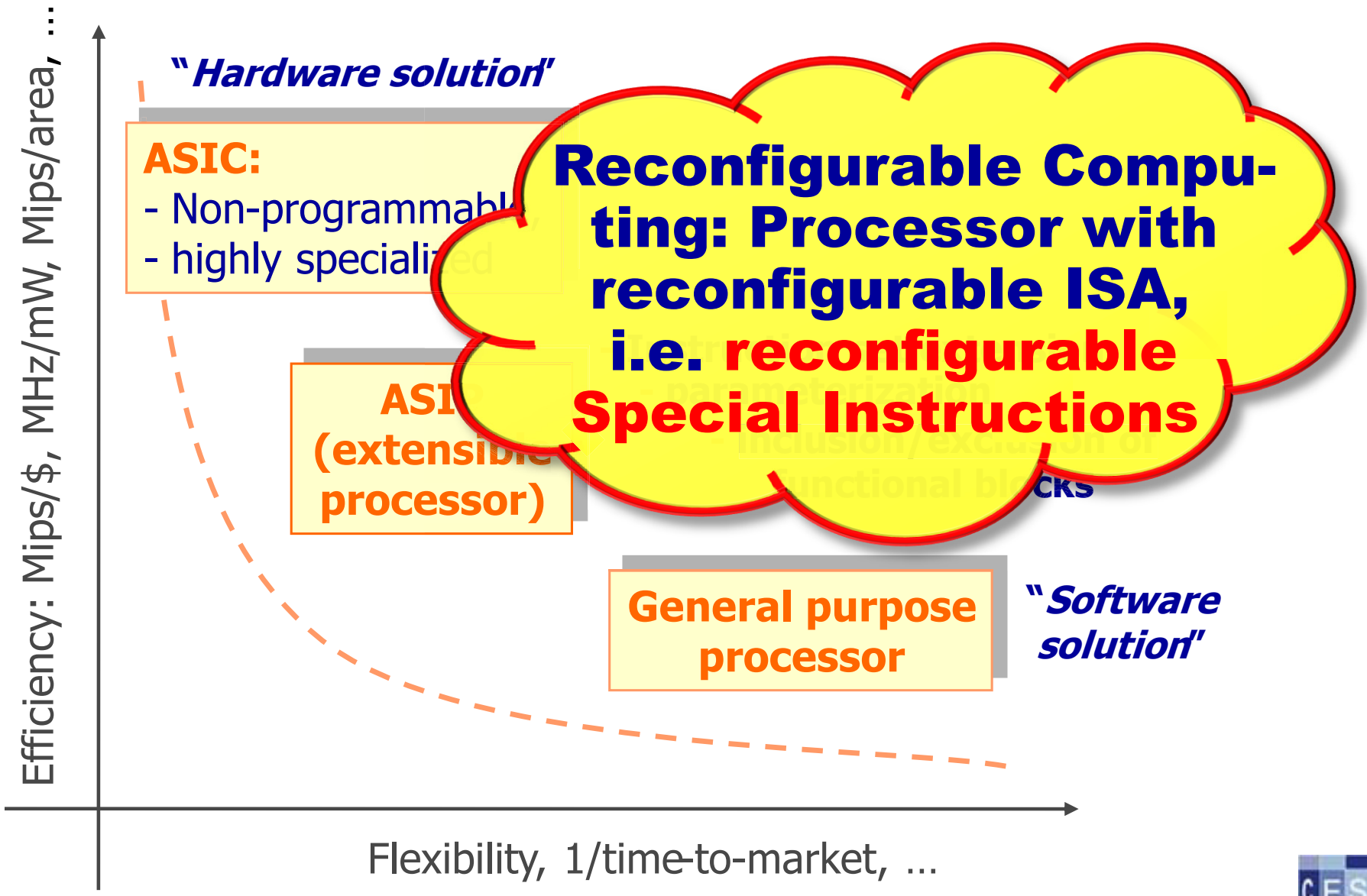
Chair for Embedded Systems (CES)

University of Karlsruhe

Development of Embedded Systems

- Typical:
  - Static analysis of hot spots
  - Building tightly optimized system
- Nowadays:
  - Increasing complexity
  - More functionality
- Problem:
  - Statically chosen design point has to match all requirements
  - Typically inefficient for individual components (e.g. tasks or hot spots)

Possible Solution: Extensible Processors



Related Work: Reconfigurable Processors

- [CoMPARE'98]:
  - Fine-grained reconfigurable fabric coupled to the core pipeline
  - Can implement a single Special Instruction (SI) at a time
- [CHIMAERA'00]:
  - Supports multiple SIs in the reconfigurable fabric at the same time
- [MOLEN'04]:
  - Can be configured to support only a single or multiples SIs at the same time
- [RISPP'07]:
  - Supports multiple SIs at the same time and allows multiple (hardware) implementations per SI (providing different performance/area trade-offs)
  - Partitions SIs into Data Paths which are reconfigured independently
- Coarse-Grained Reconfiguration: ADRES, etc. (not the focus of this talk)

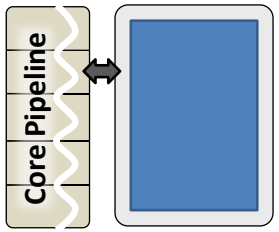
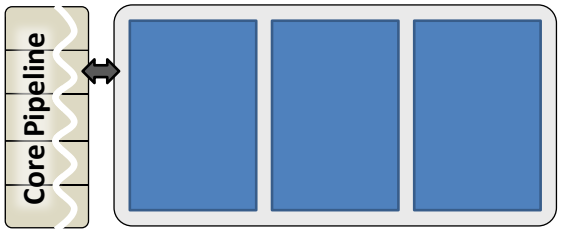
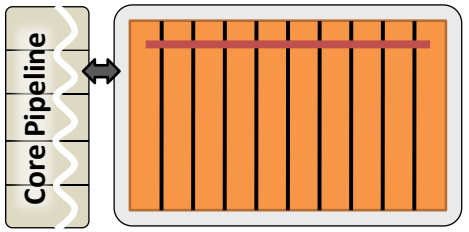
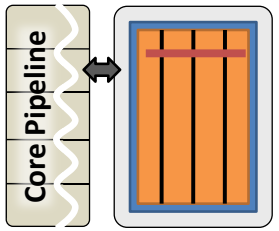
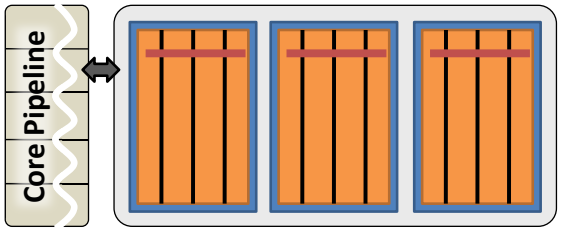
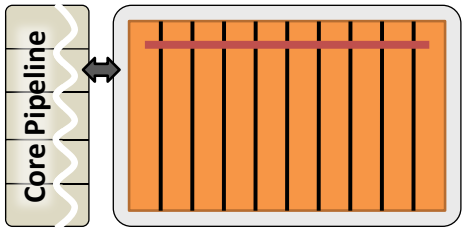
Outline

- Introduction
- Related Work
- Reconfigurable Processor Alternatives
  - Special Instruction-Based Categorization
  - Relevant Architectural Parameters
  - Design Space Exploration Tool
- Conclusion

SI-Based Categorization

- Providing Special Instruction (SI) Implementations:
  - How many SIs may be available at the same time?
  - How many implementation alternatives exist per SI?
  - Note: each SI may be executed by the ISA of the core pipeline
- Technical constraints: Rectangular implementation of a hardware description (e.g. an SI)
  - The typical shape for place & route tools
- These rectangular implementations cannot be placed at arbitrary positions within the reconfigurable fabric
  - They are typically aligned to dedicated communication ports that are provided at fixed positions

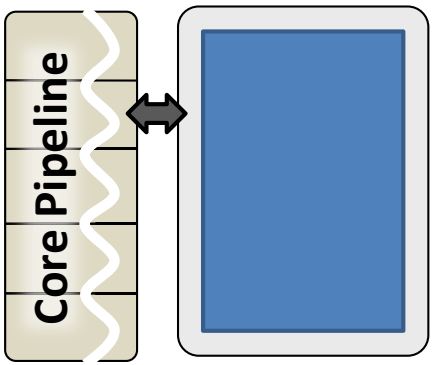
SI-Based Categorization Overview

How many Implementations per SI \ How many SIs at the same time	At most 1 SI with a fixed maximum *size*	At most n>1 SIs with a fixed maximum *size* per SI	Changing number of SIs with different sizes (e.g. 2 *big* SIs or 5 *small* SIs); maximum number of total data paths (DPs)
One Implementation per SI. SI has to be loaded completely before it is executable	**Category-1:** Single SI Container 	**Category-2:** Multiple SI Containers 	**Category-3:** Multiple *overlapping* SIs 
Multiple Implementations per SI. SIs are partitioned into Data Paths (DPs)	**Category-4:** Single Partitioned SI Container 	**Category-5:** Multiple Partitioned SI Containers 	**Category-6:** Multiple DP Containers 

Legend:

- Core Pipeline (scaled down):
- Reconfigurable area:
- Special Instruction Container (SIC):
- Communication System:
- Data Path Container (DPC):

SI-Based Categorization: Category-1: Single SI Container

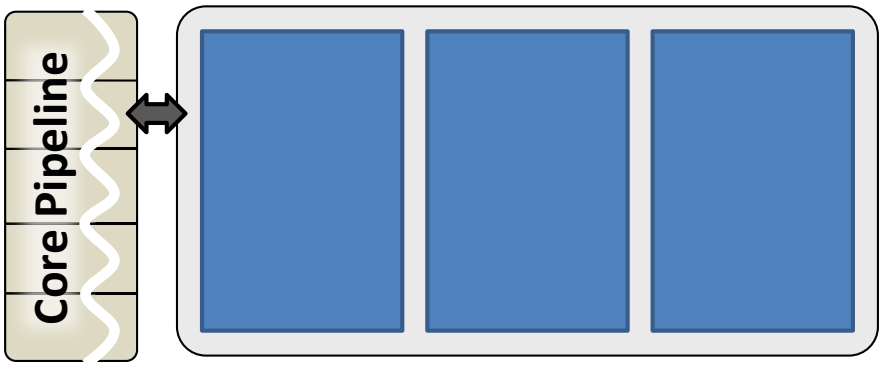


- Corresponds to [CoMPARE'98]

- At most one SI is available in hardware at a given time
- Relatively long reconfiguration time, depending on the size of the SI Container
- Depending upon the required amount of logic two SIs might fit into the Container, but it is not supported
  - → Internal fragmentation

Legend:	Core Pipeline (scaled down):	Reconfigu-rable area:	Special Instruction Container (SIC):

SI-Based Categorization: Category-2: Multiple SI Containers

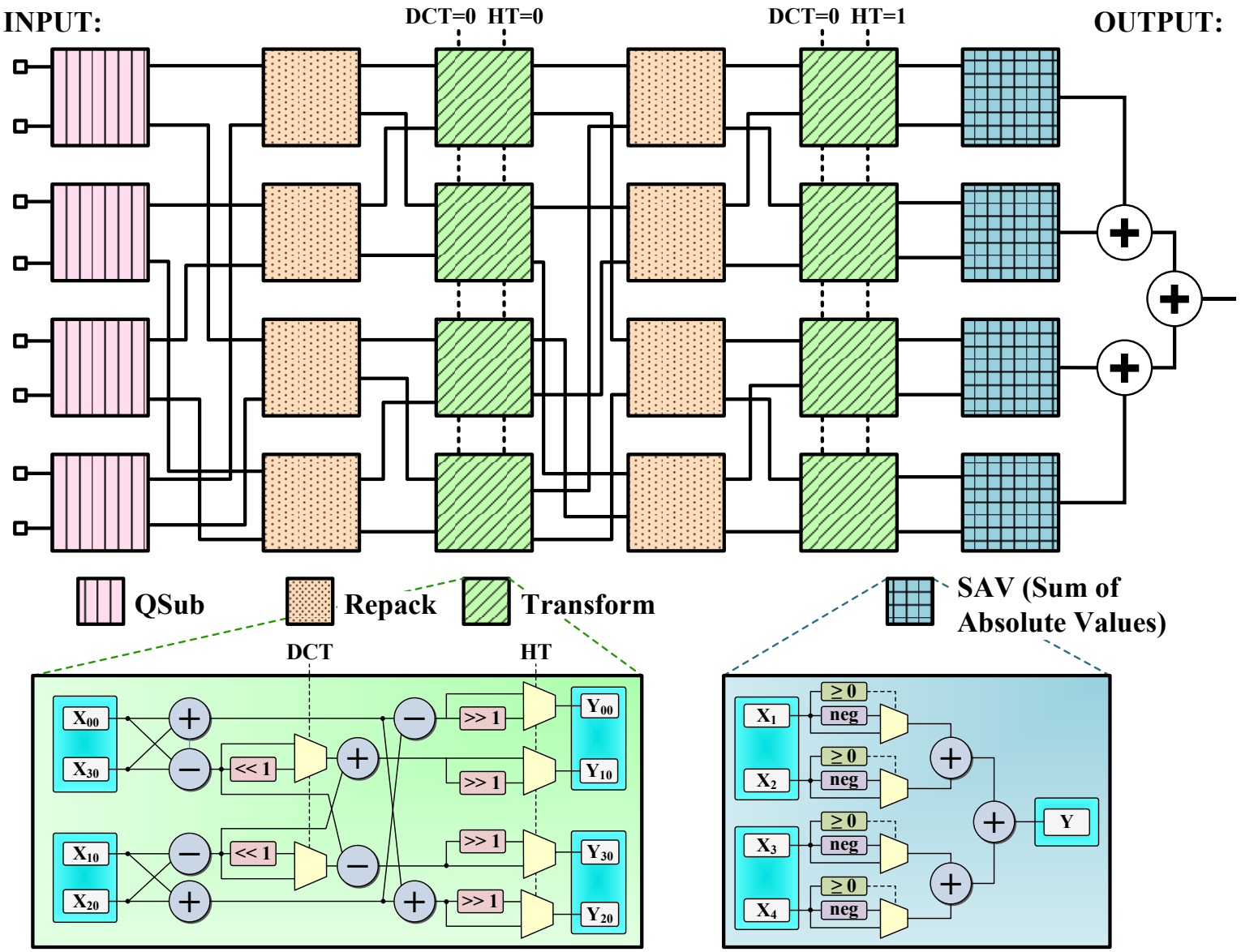


- Corresponds to [CHIMAERA'00] and [MOLEN'04]

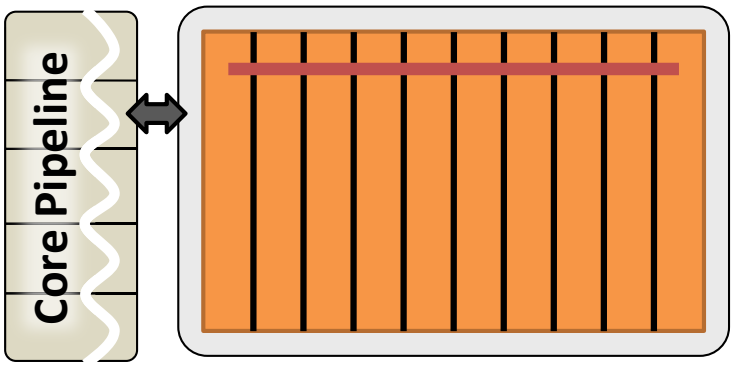
- An SI may be loaded into any free container
- SIs may not be bigger than the container, even if not all containers are demanded
  - → external fragmentation (in addition to the internal fragmentation per SI Container)

Legend:	Core Pipeline (scaled down):	Reconfigu-rable area:	Special Instruction Container (SIC):

Example: Modular SIs (using Data Paths)



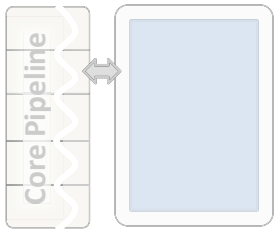
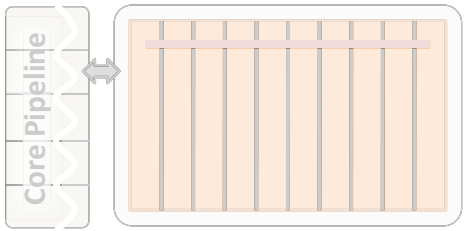
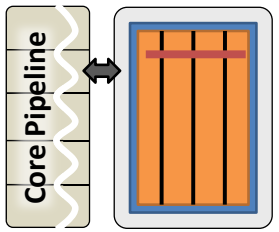
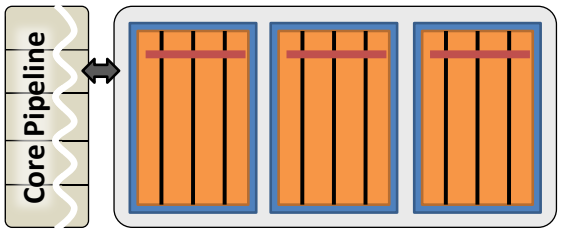
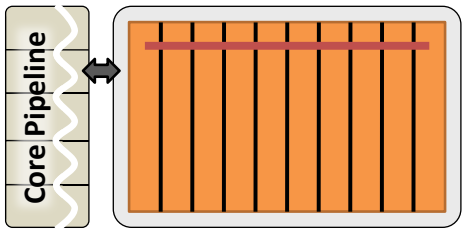
SI-Based Categorization: Category-3: Multiple *overlapping* SIs



- There is no predetermined maximum of supported SIs
- Multiple SIs may share common data paths (i.e. reuse them) because at most one SI is executed at a time.
- This addresses the internal and external fragmentation problem
- Demand for internal communication system

Legend:	Core Pipeline (scaled down):	Reconfigurable area:	Special Instruction Container (SIC):
		Communication System:	Data Path Container (DPC):

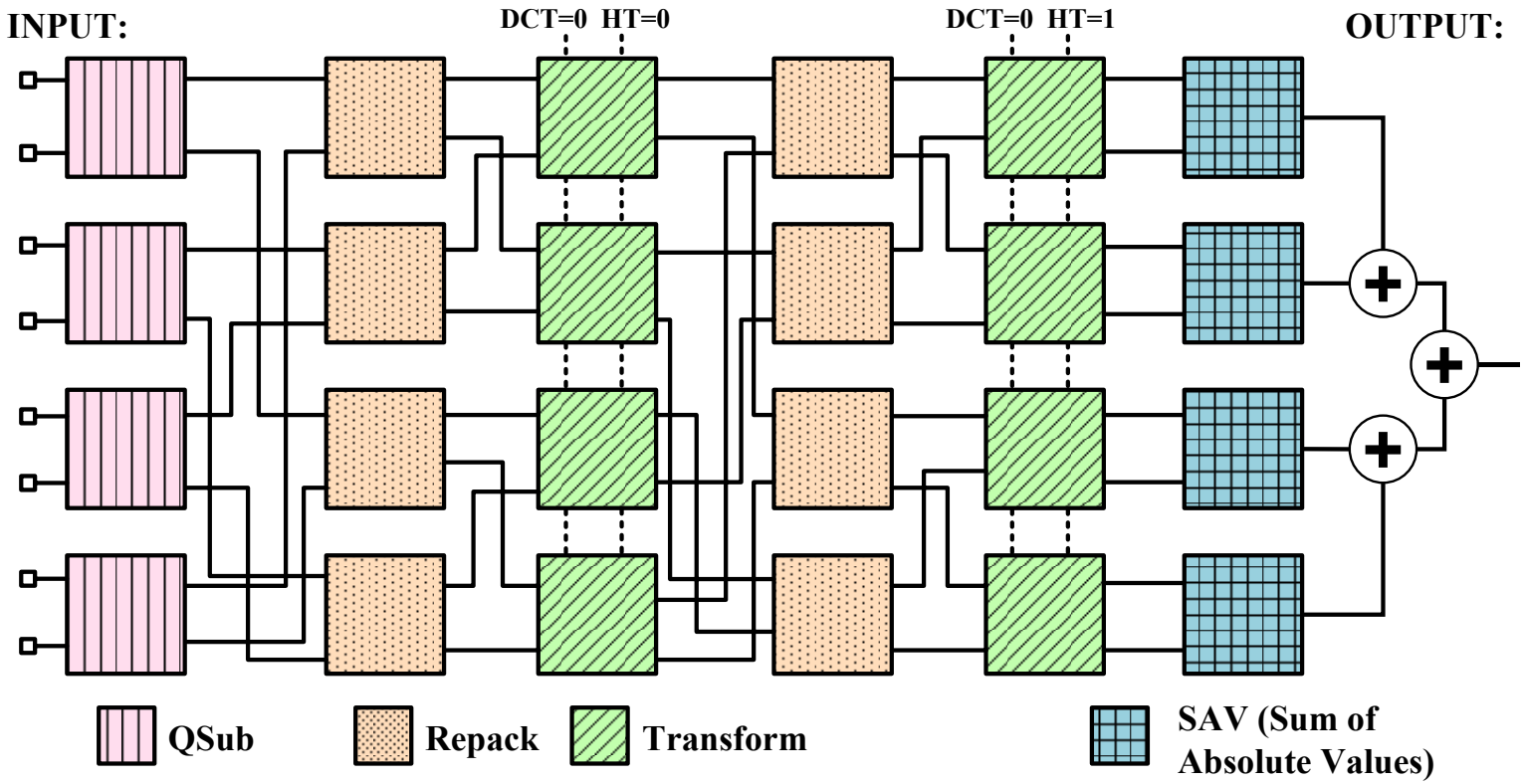
SI-Based Categorization Overview

How many Implementations per SI \ How many SIs at the same time	At most 1 SI with a fixed maximum *size*	At most n>1 SIs with a fixed maximum *size* per SI	Changing number of SIs with different sizes (e.g. 2 *big* SIs or 5 *small* SIs); maximum number of total data paths (DPs)
One Implementation per SI. SI has to be loaded completely before it is executable	**Category-1:** Single SI Container 	**Category-2:** Multiple SI Containers 	**Category-3:** Multiple *overlapping* SIs 
Multiple Implementations per SI. SIs are partitioned into Data Paths (DPs)	**Category-4:** Single Partitioned SI Container 	**Category-5:** Multiple Partitioned SI Containers 	**Category-6:** Multiple DP Containers 

Legend:

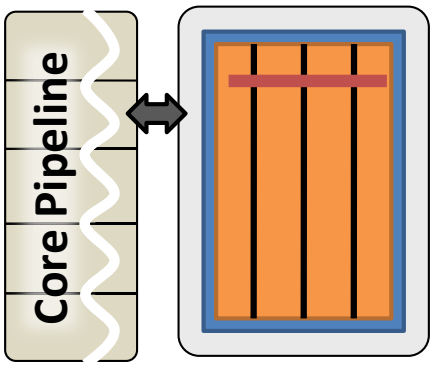
- Core Pipeline (scaled down):
- Reconfigurable area:
- Special Instruction Container (SIC):
- Communication System:
- Data Path Container (DPC):

Example: Modular SIs (allowing for multiple Implementations)



- The 'Transform' Data path might be available once (i.e. readily reconfigured) and used 8 times to realize the SI functionality
- Or it might be available twice and both instances are used 4 times, or …

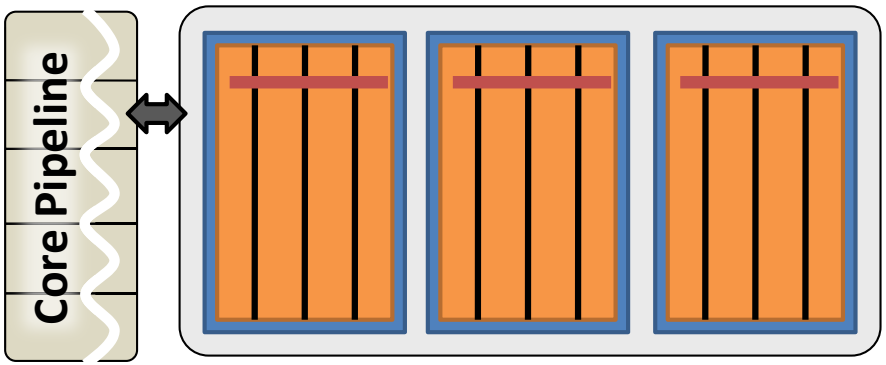
SI-Based Categorization:
Category-4: Single Partitioned SI Container



- 1 SI Container, partitioned into *n* DP Containers that are connected with a communication system
- When more DPs finish reconfiguration, then a faster implementation of an SI may become available
- Shorter reconfiguration time than Category-1 (only the demanded DPs need to be reconfigured)
- But still internal fragmentation (at most 1 SI supported; independent of it's size)

Legend:	Core Pipeline (scaled down):	Reconfigurable area:	Special Instruction Container (SIC):
		Communication System:	Data Path Container (DPC):

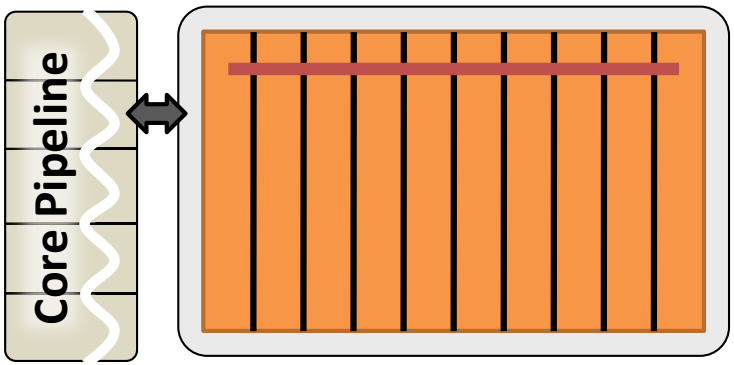
SI-Based Categorization:
Category-5: Multiple Partitioned SI Containers



- Shorter reconfiguration time
- Still internal fragmentation
- External fragmentation problems
- Additionally, if the DPs that are demanded by an SI are not in the same SI Containers, they can not be used together (e.g. to implement an SI)

Legend:	Core Pipeline (scaled down):	Reconfigurable area:	Special Instruction Container (SIC):
		Communication System:	Data Path Container (DPC):

SI-Based Categorization: Category-6: Multiple DP Containers



- Corresponds to [RISPP'07]

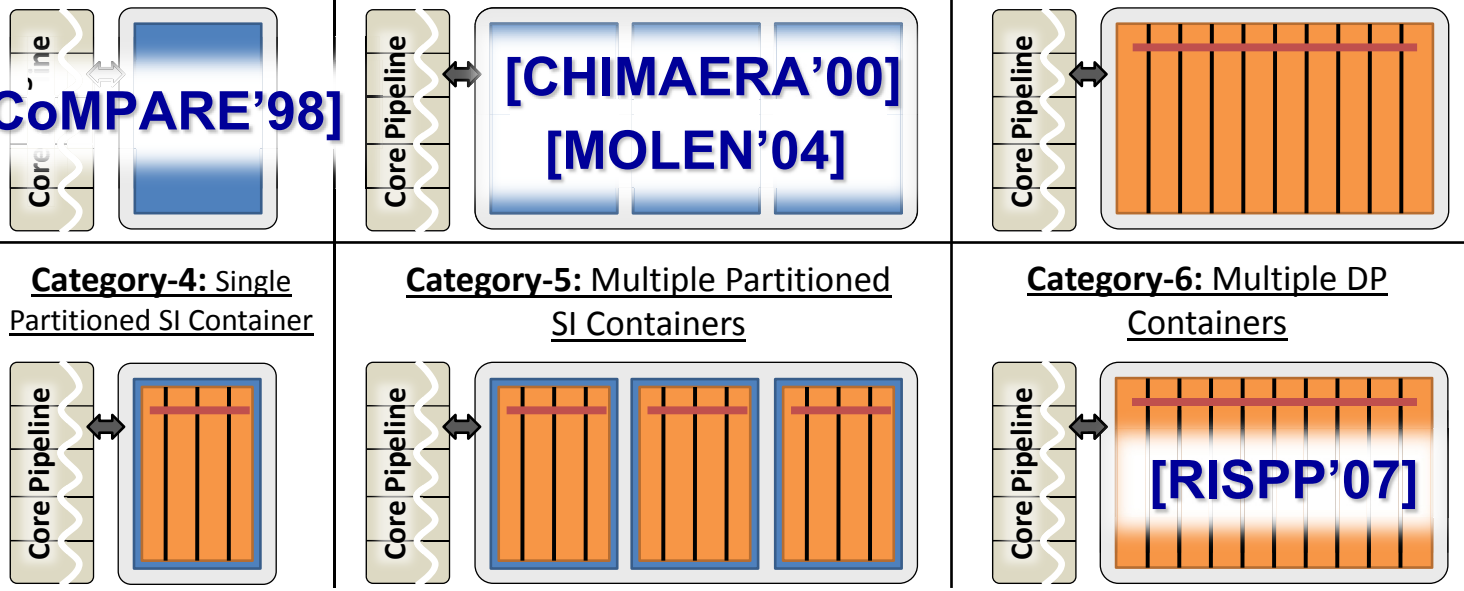
- Main differences to Category-3:
  - SIs can be *upgraded* (due to multiple available SI implementations; like in Category-4 & 5)
  - Decision how many DP Containers shall be spend for which SI can adapt at run time
  - →Demands a run-time system

- Main diff. to Category-4 & 5:
  - No external fragmentation
  - Available DPs may be used for all SIs, i.e. not fixed to a certain SI Container

Legend:	Core Pipeline (scaled down):	Reconfigurable area:	Special Instruction Container (SIC):
		Communication System:	Data Path Container (DPC):

SI-Based Categorization Overview

How many Implementations per SI \ How many SIs at the same time	At most 1 SI with a fixed maximum *size*	At most n>1 SIs with a fixed maximum *size* per SI	Changing number of SIs with different sizes (e.g. 2 *big* SIs or 5 *small* SIs); maximum number of total data paths (DPs)
One Implementation per SI. SI has to be loaded completely before it is executable	**Category-1:** Single SI Container [CoMPARE'98]	**Category-2:** Multiple SI Containers [CHIMAERA'00] [MOLEN'04]	**Category-3:** Multiple *overlapping* SIs
Multiple Implementations per SI. SIs are partitioned into Data Paths (DPs)	**Category-4:** Single Partitioned SI Container	**Category-5:** Multiple Partitioned SI Containers	**Category-6:** Multiple DP Containers [RISPP'07]



Legend:

Core Pipeline (scaled down):

Reconfigurable area:

Special Instruction Container (SIC):

Communication System:

Data Path Container (DPC):

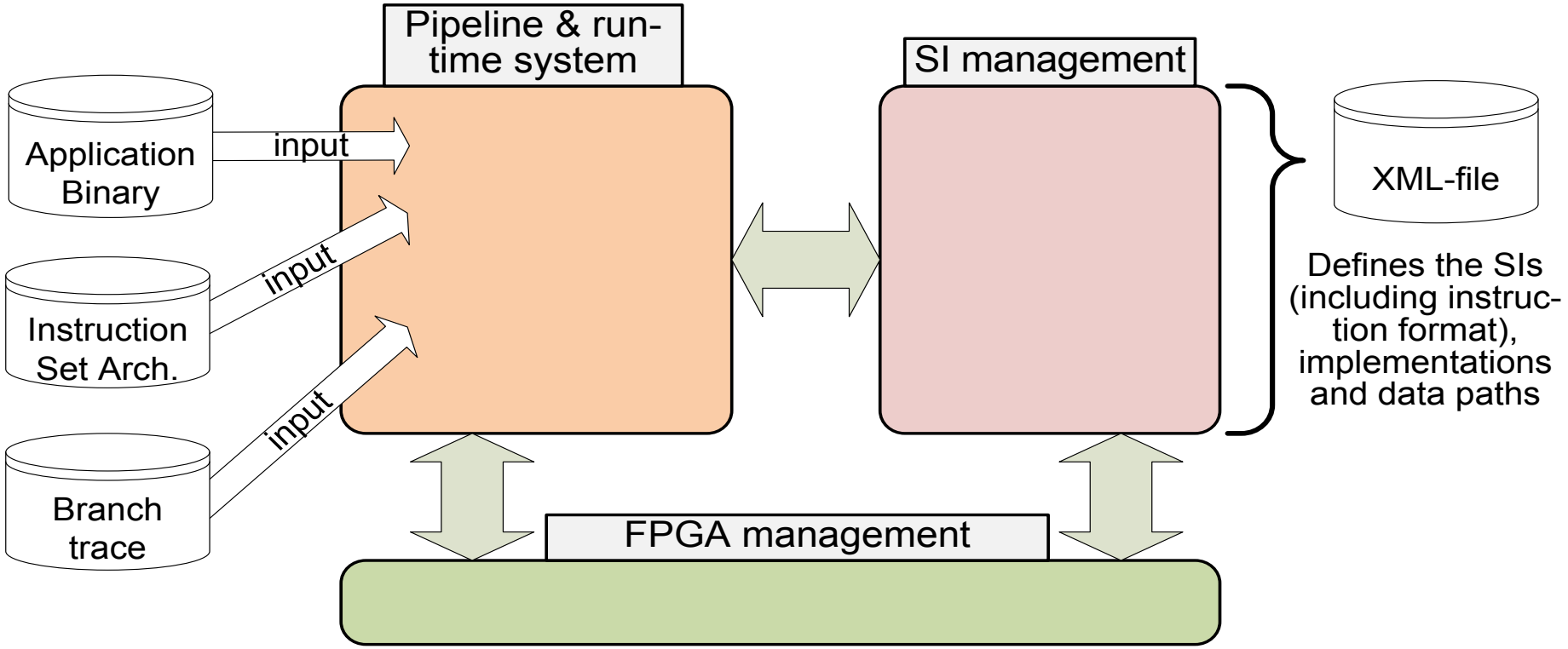
Outline

- Introduction
- Related Work
- Reconfigurable Processor Alternatives
  - Special Instruction-Based Categorization
  - **Relevant Architectural Parameters**
  - **Design Space Exploration Tool**
- Design Space Exploration
- Conclusion

Relevant Architectural Parameters

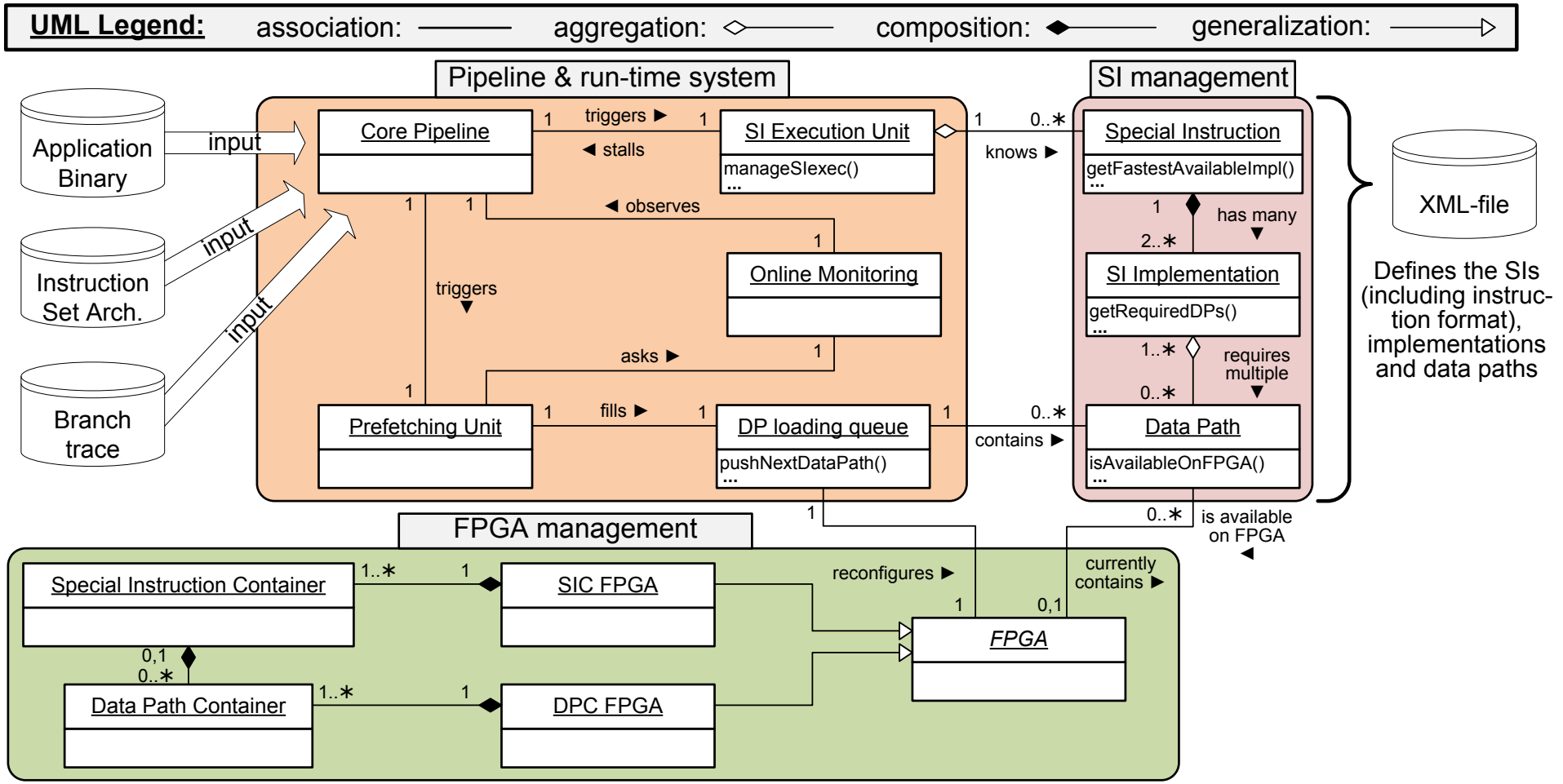
- Core Pipeline Frequency: f_{CPU} [MHz]
- FPGA Frequency: f_{FPGA} [MHz]
  - May differ, due to fabrication technology
- Data Memory Connection
  - Number of Memory Ports: P
  - Bit width per Memory Port: W [Bits]
- Reconfiguration Bandwidth: R [MB/s]
  - Determines time to reconfigure parts of the FPGA
  - Depends on the FPGA configuration port and the used memory

Design Space Exploration Tool: Overview Input Data and Connections

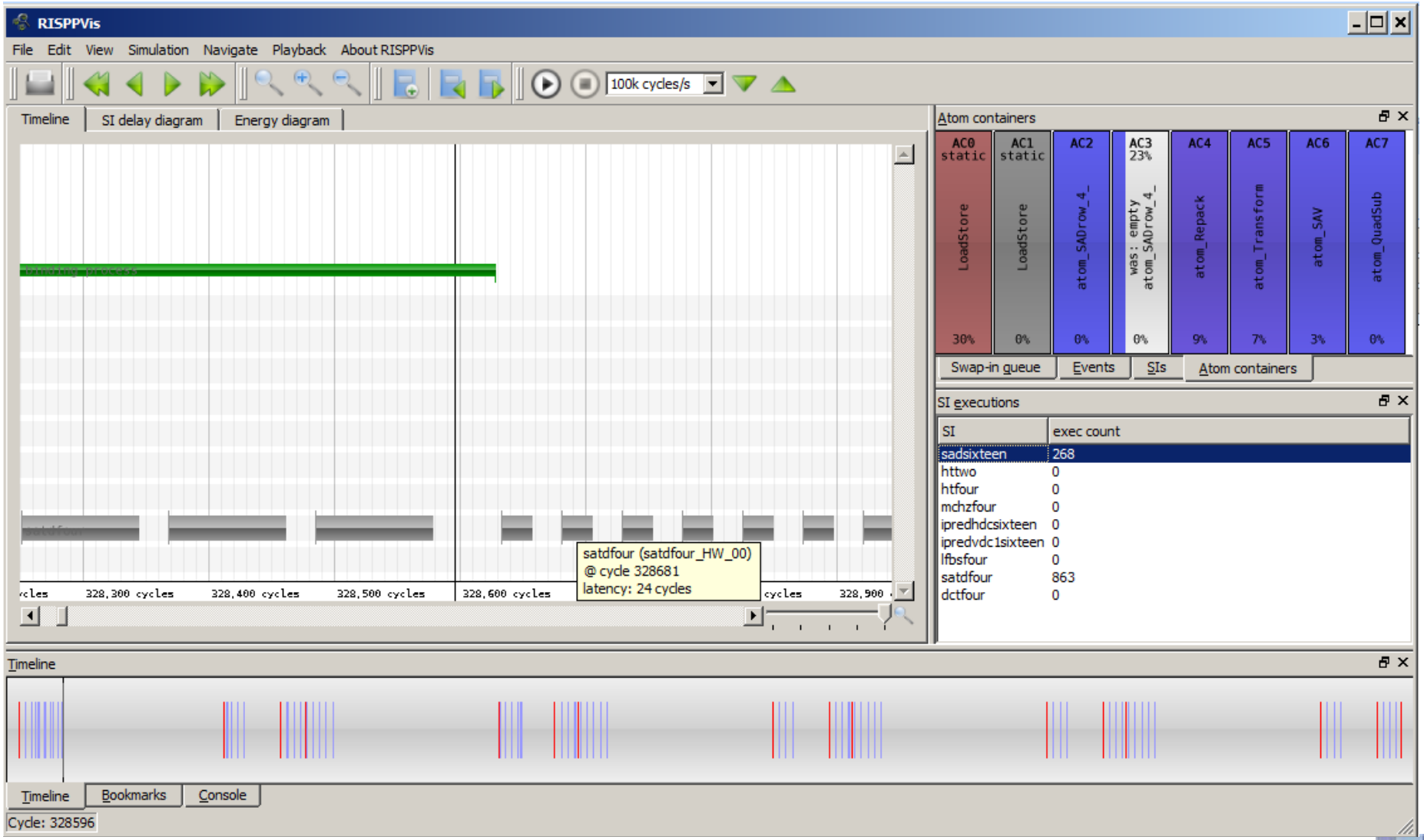


- System C based simulator
- Input for pipeline is obtained from Instruction Set Simulator (ArchC)
- SI information is semi-automatically derived at compile time

Design Space Exploration Tool: Internal Composition



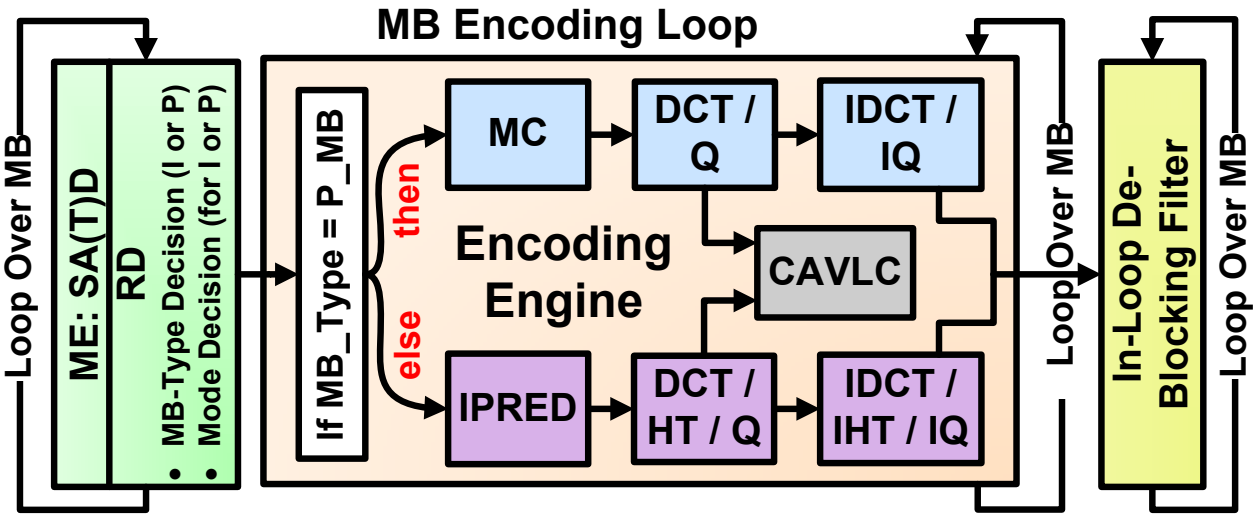
Design Space Exploration Tool: Detailed Run-time Analysis



Outline

- Introduction
- Related Work
- Reconfigurable Processor Alternatives
  - Special Instruction-Based Categorization
  - Relevant Architectural Parameters
  - Design Space Exploration Tool
- **Design Space Exploration**
- Conclusion

Benchmark Application: H.264 Video Encoder



- Challenging Application with many computational Hot Spots
- Benchmarking 20 frames in QCIF resolution (176x144)
- The GPP (i.e. a Sparc-V8 without reconfigurable hardware) requires 10.6 seconds @ 100 MHz → 1.89 fps
  - 2.1 seconds @ 500 MHz → 9.52 fps

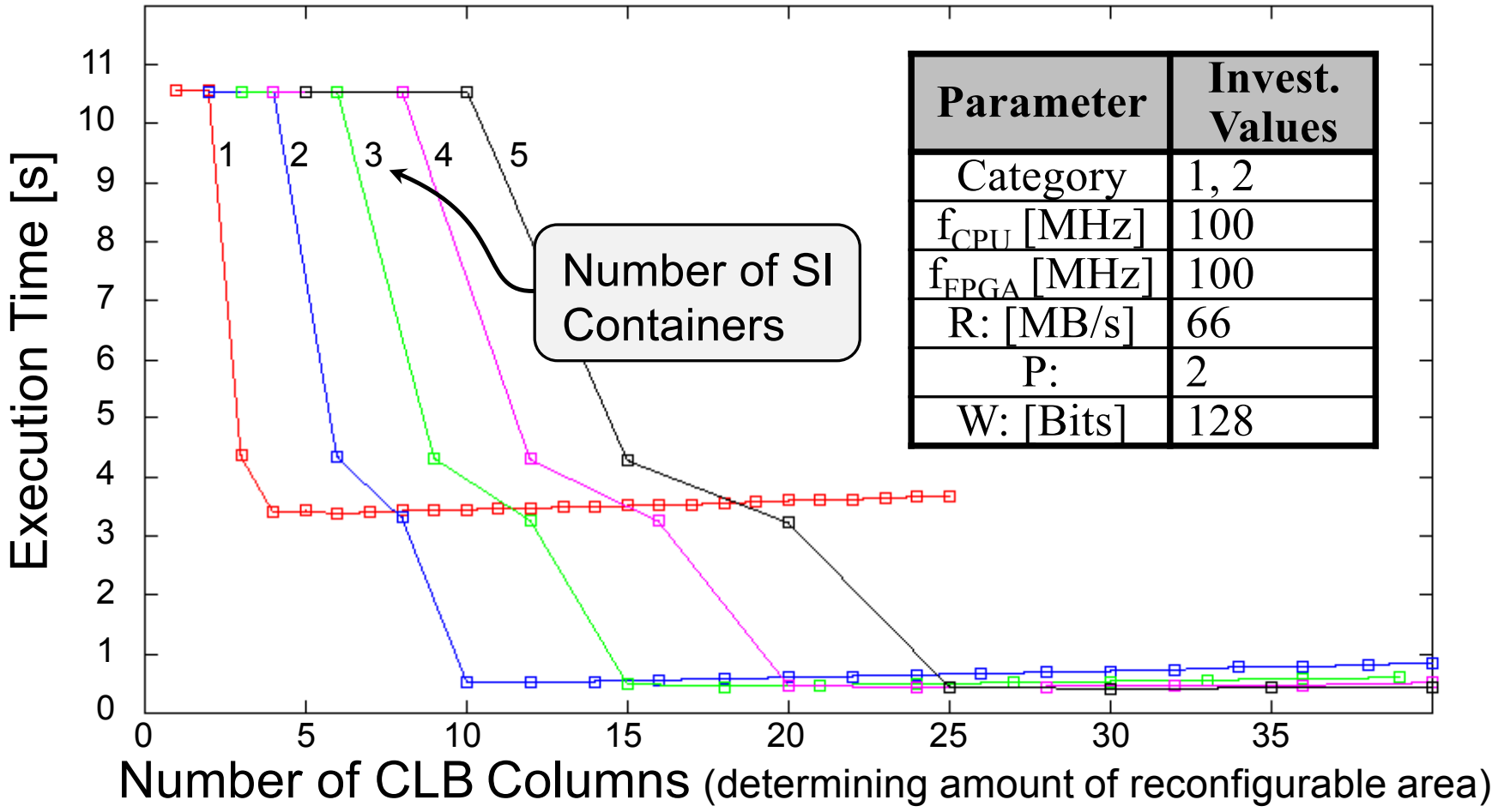
Special Instruction Overview

Parameter	Value	Comment
# SIs	9	4/1 in/out register (e.g. for memory addresses)
# Data Paths	10	2/2 32-bit in/out values
SI composition	1 - 4 DPs	Utilizing multiple instances per DP
SI memory accesses	0 – 128 words	For some SIs the input from register file is sufficient, others work on data memory (using up to 2 ports á 128 bit)
DP Bitstream	42,719 - 43,638 Byte	Bitstream for partial reconfiguration on Xilinx Virtex-II xc2v6000 FPGA
DP logic requirements	16 – 192 slices	Note: these readings correspond to the pure computational logic without the necessary interconnection overhead

Special Instruction Overview (cont'd)

	Special Instr.	Implemented Data Paths
Motion Estimation	SAD	SAD_16
	SATD	QSub, Transform/HT_4, Repack, SAV
(Inverse) Transform	(I)DCT	Transform/DCT_4, Repack, (QSub)
	(I)HT_2x2	Transform/HT_2
	(I)HT_4x4	Transform/HT_4, Repack
Motion Compensation	MC_Hz_4	PointFilter, Repack, Clip3
Intra Prediction	IPred_HDC	CollapseAdd, Repack
	IPred_VDC	CollapseAdd
Loop Filter	LF_BS4	Cond, LF_4

Evaluating Category-1 and 2



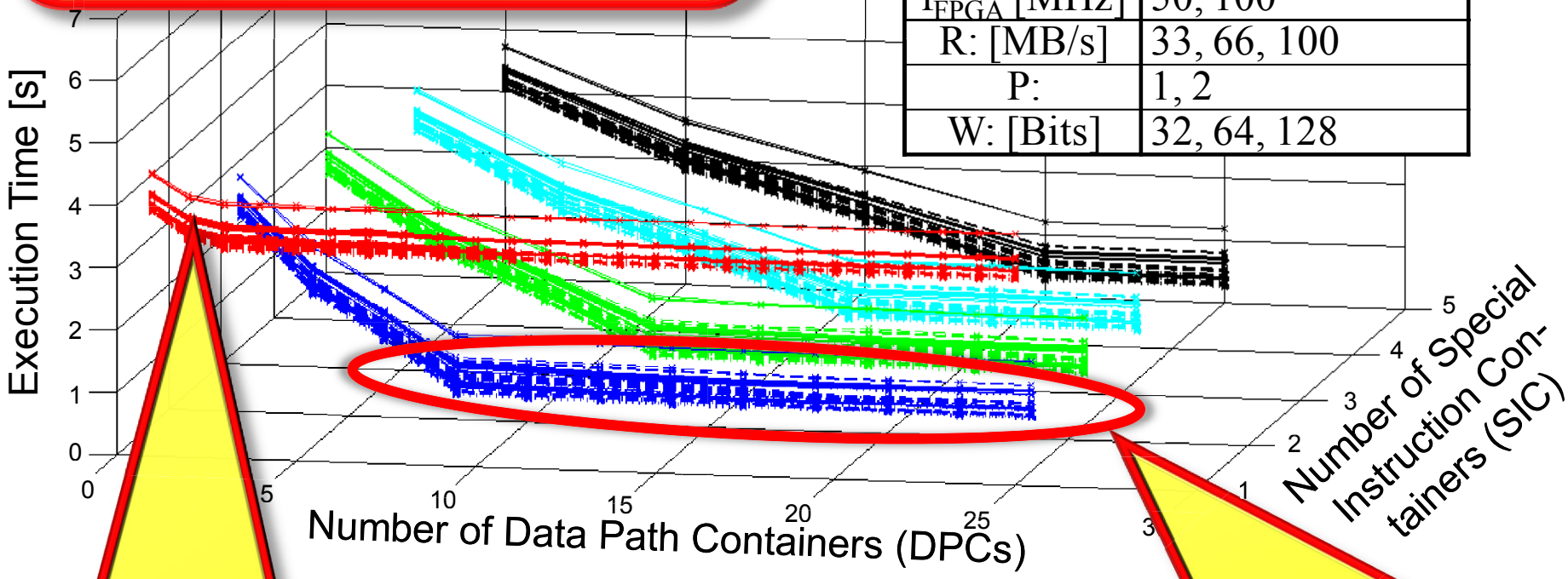
Parameter	Invest. Values
Category	1, 2
f_{CPU} [MHz]	100
f_{FPGA} [MHz]	100
R: [MB/s]	66
P:	2
W: [Bits]	128

Evaluating Category-4 and 5

Evaluating many different parameters, but #SICs and #DPCs dominate the other readings

Summarizing 2016 Measurements

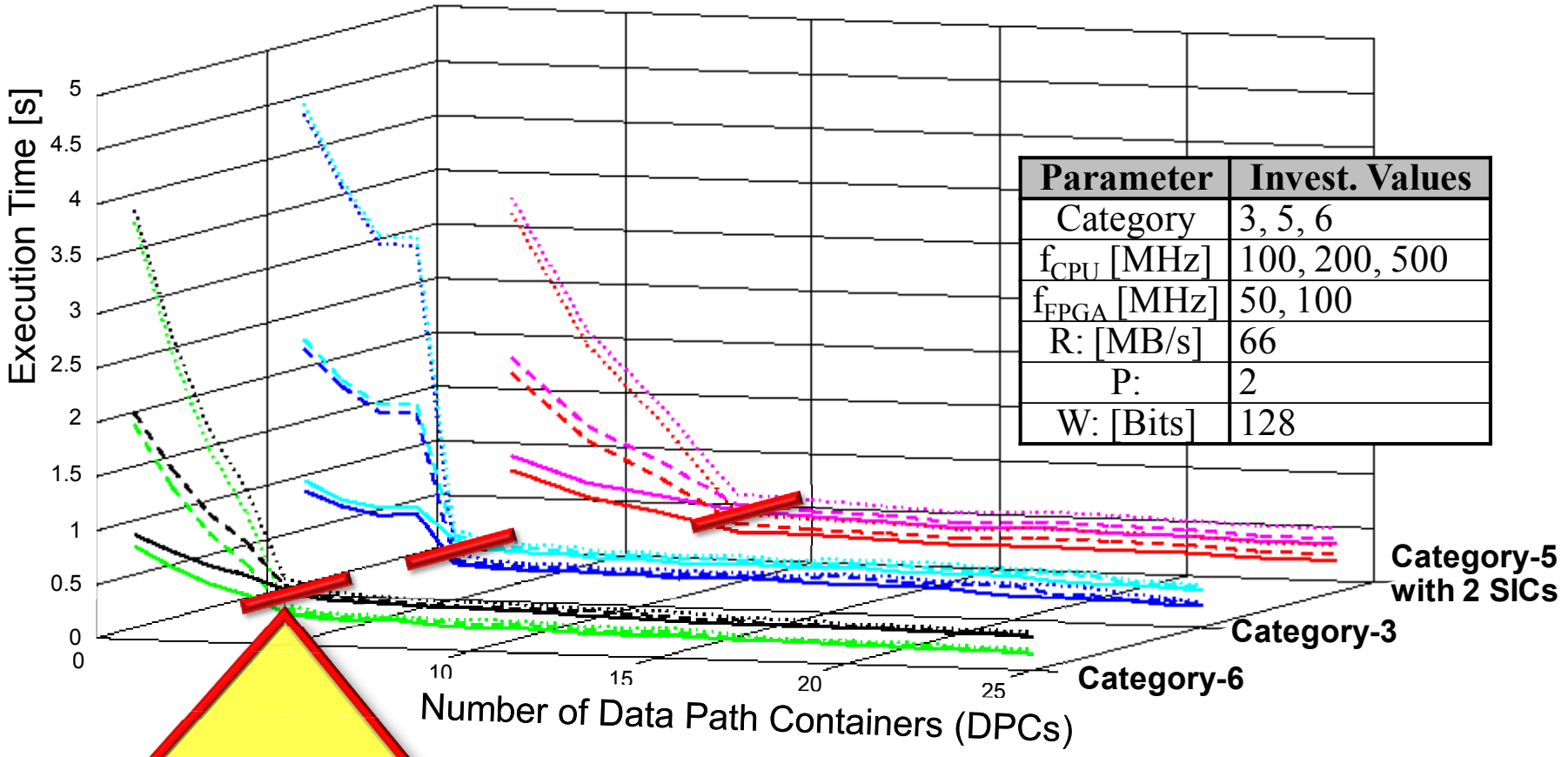
Parameter	Invest. Values
Category	4, 5
f_{CPU} [MHz]	100
f_{FPGA} [MHz]	50, 100
R: [MB/s]	33, 66, 100
P:	1, 2
W: [Bits]	32, 64, 128



1 SI Container is not sufficient

Rather many DPCs no longer lead to a slower execution time

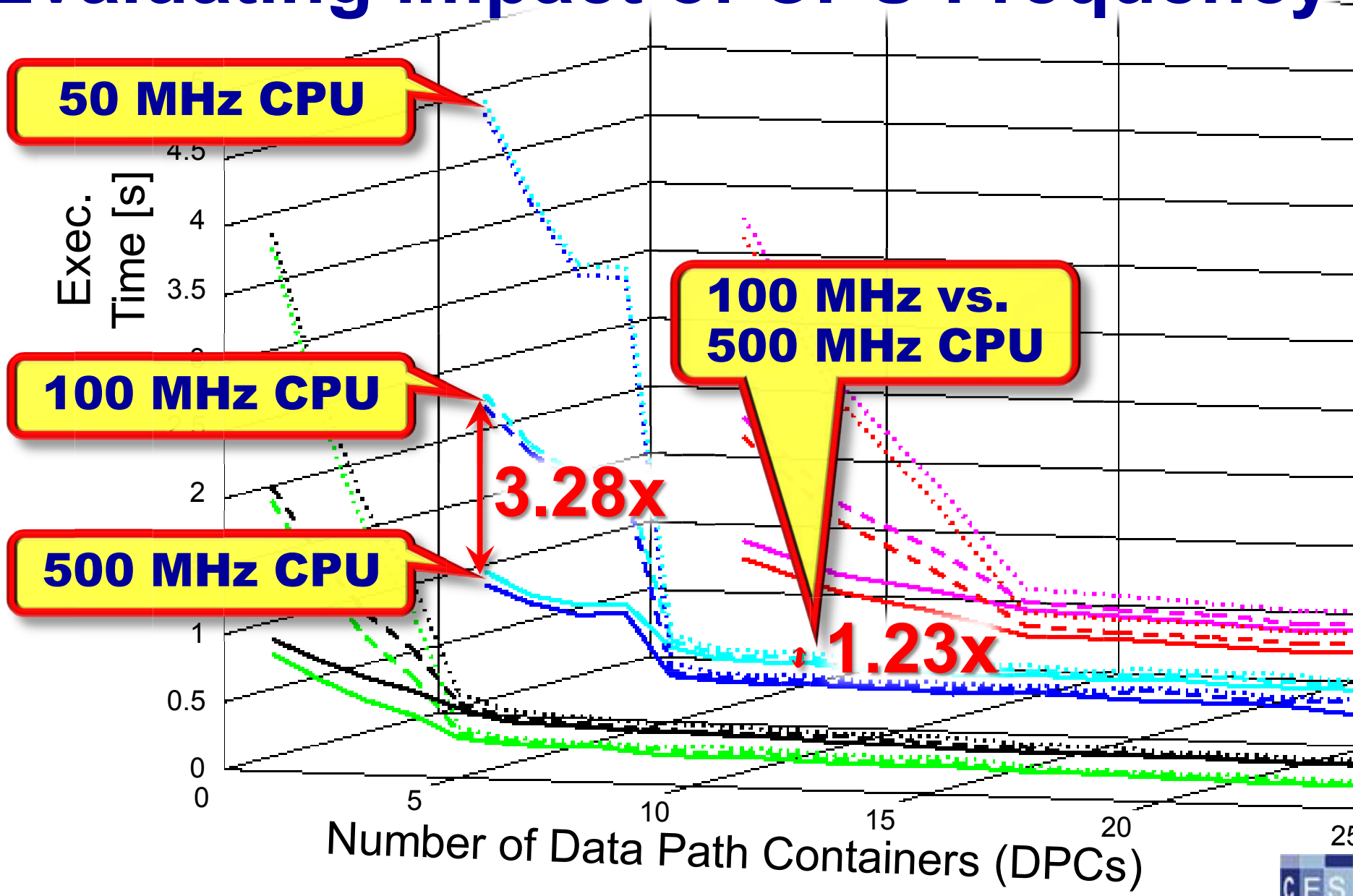
Evaluating Category-3, 5, and 6



Parameter	Invest. Values
Category	3, 5, 6
f_{CPU} [MHz]	100, 200, 500
f_{FPGA} [MHz]	50, 100
R: [MB/s]	66
P:	2
W: [Bits]	128

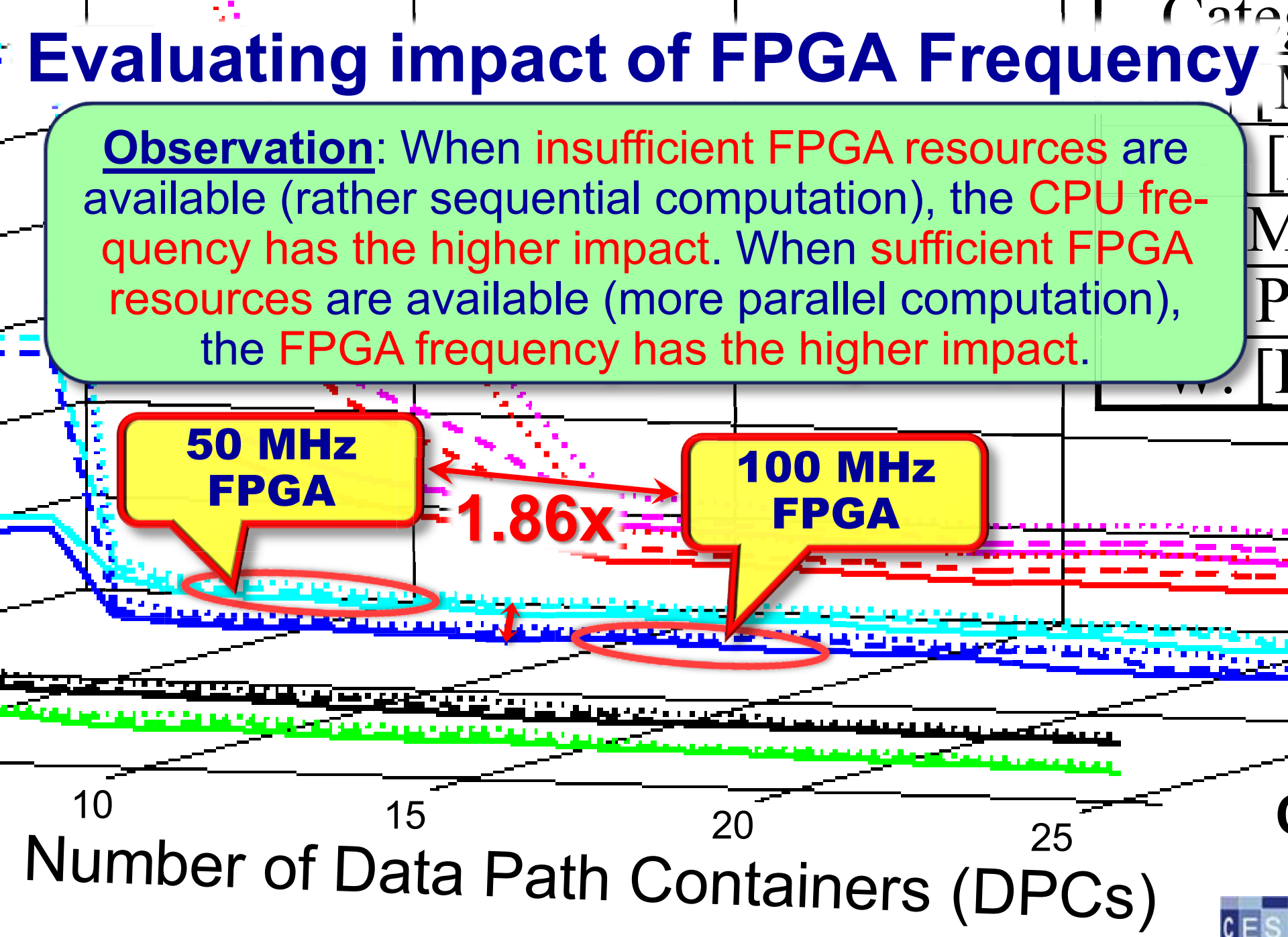
'Critical' Amount of DP Containers:
Category-5: 8 DPCs
Category-3: 5 DPCs
Category-6: 5 DPCs

Evaluating impact of CPU Frequency

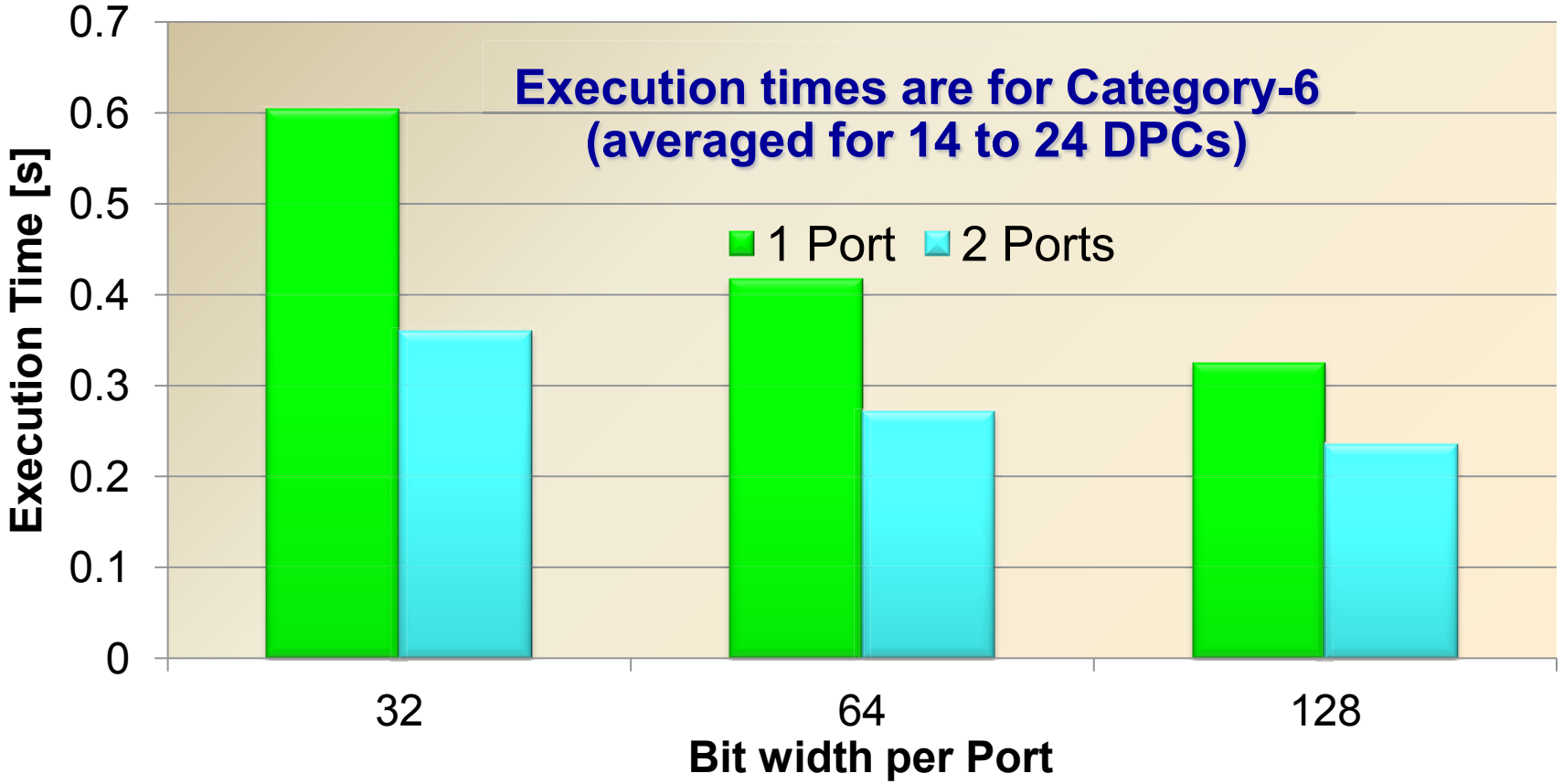


Evaluating impact of FPGA Frequency

Observation: When insufficient FPGA resources are available (rather sequential computation), the CPU frequency has the higher impact. When sufficient FPGA resources are available (more parallel computation), the FPGA frequency has the higher impact.

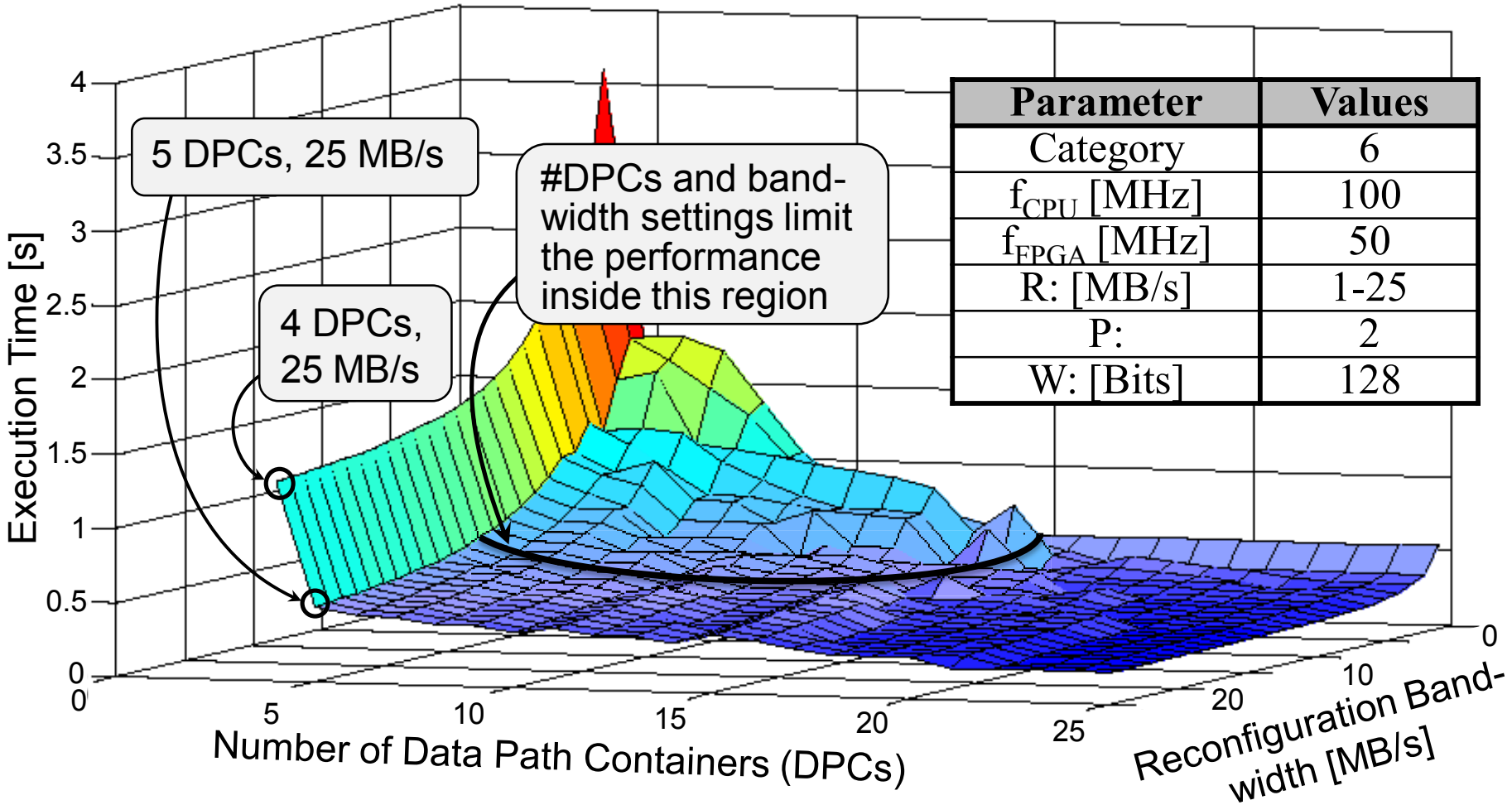


Evaluating Data Memory Connection



- For a given total bit width the 2-port data memory always outperforms the 1-port memory
- Data Memory Connection affects potential parallelism
  ➔ affects the relevance of CPU and/or FPGA frequency

Evaluating Reconfiguration Bandwidth



Parameter	Values
Category	6
f_{CPU} [MHz]	100
f_{FPGA} [MHz]	50
R: [MB/s]	1-25
P:	2
W: [Bits]	128

Summary & Conclusion

- Reconfigurable Processors are a promising approach for challenging and/or dynamically changing applications
- They can be categorized according their implementation of Special Instructions
  - How many SIs may be available at the same time?
  - How many implementation alternatives exist per SI?
  - Covering existing architectures and unveiling further ones
- Furthermore, different architectural parameters affect the performance of the system
- These settings interfere with each other (e.g. data memory connection and CPU/FPGA frequency)
  - This talk highlighted which parameters are relevant in which situation, based on an exhaustive design space exploration

Classifying and Evaluating Performance-relevant Parameters for Reconfigurable Processors

Lars Bauer, Muhammad Shafique,
and Jörg Henkel

Thank you for your attention !

Chair for Embedded Systems (CES)
University of Karlsruhe