Diana Göhringer
Chair of Adaptive Dynamic Systems
TU Dresden

Methods and Tools for Accelerating Image Processing Applications on FPGA-based Systems

DATE 2022

Workshop W06: Data-driven applications for industrial and societal challenges: Problems, methods, and computing platforms

Outline

- Introduction and Motivation
- Reconfigurable FPGA Overlay Architecture
- Design / Programming Methodology
  - Image Processing Library: HiFlipVX
  - Application Partitioning and Mapping
- Application Examples in Relation to Funded Research Projects
- Conclusion and Outlook

Introduction and Motivation

Problem:

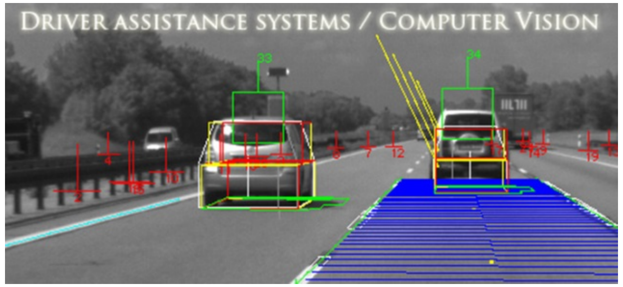
- ➢ Increasing performance demand
- ➢ Low power/energy consumption (batteries, cooling)
- ➢ Dynamic adaption to changing environments
- → Cannot be solved anymore by increasing the clock frequency, as $P_{dyn} \sim C \times V^2 \times f$

Possible solution:

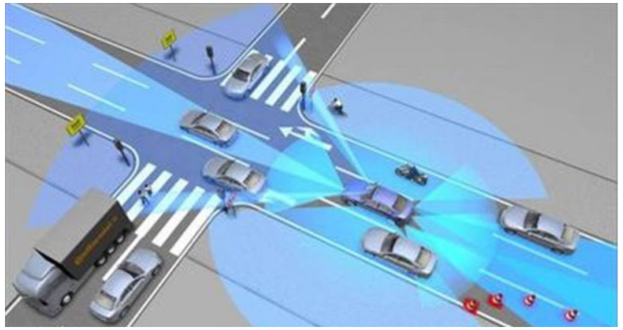
- ➢ Parallelization → Multi-Core Systems

Better solution:

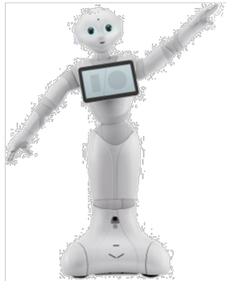
- ➢ Domain-specific Multi-Core System
- → Higher energy-efficiency



http://www.rcs.ei.tum.de/forschung/driver-assistance/

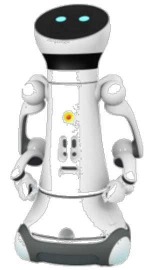


http:// www.asdreports.com/news-10595/key-players-advanced-driver-assistance-systems-adas-market-north-america-20152019



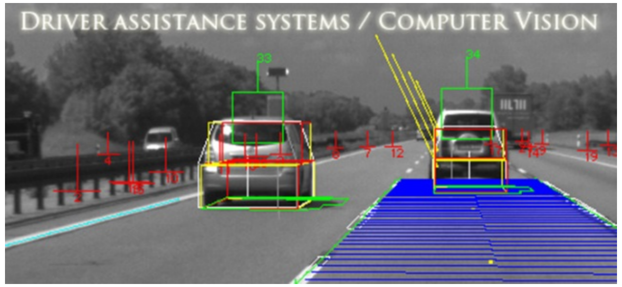
www.ald.softbankrobotics.com

asimo.honda.com

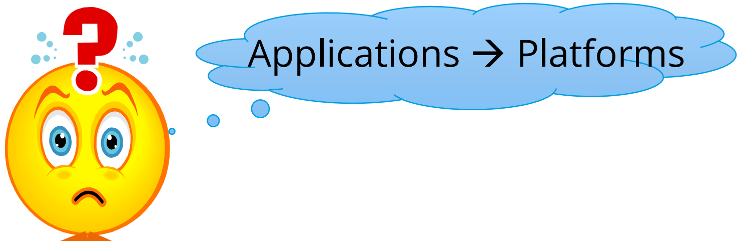


www.care-o-bot-4.de

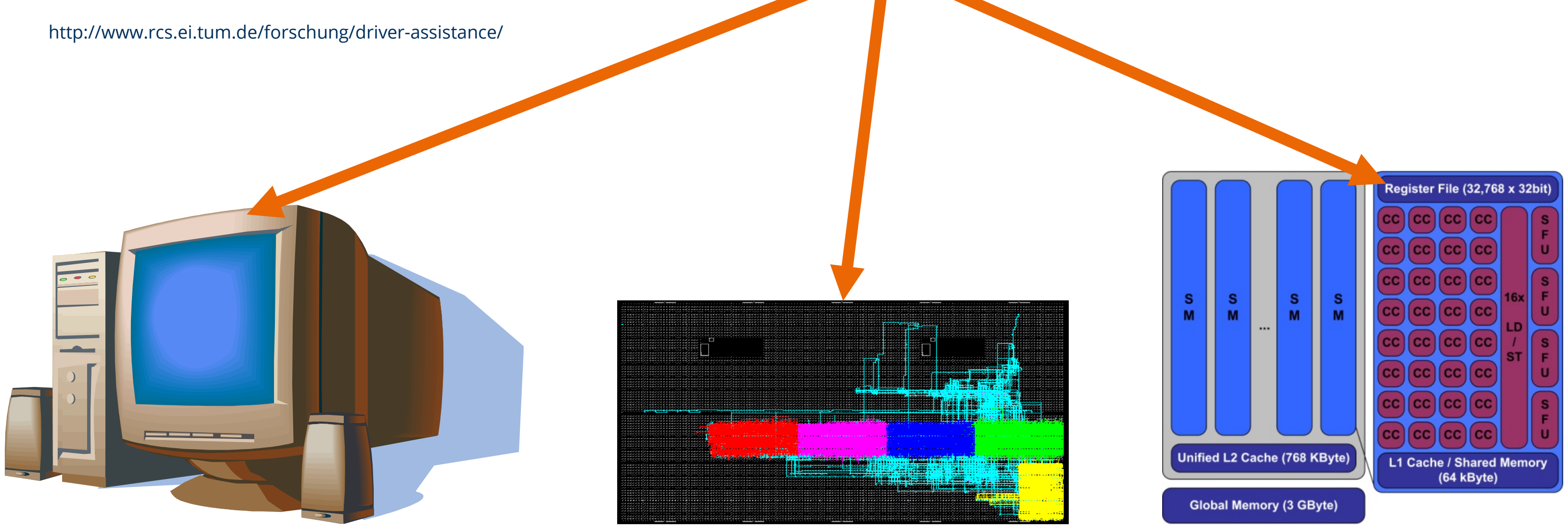
Which architecture to choose?



http://www.rcs.ei.tum.de/forschung/driver-assistance/



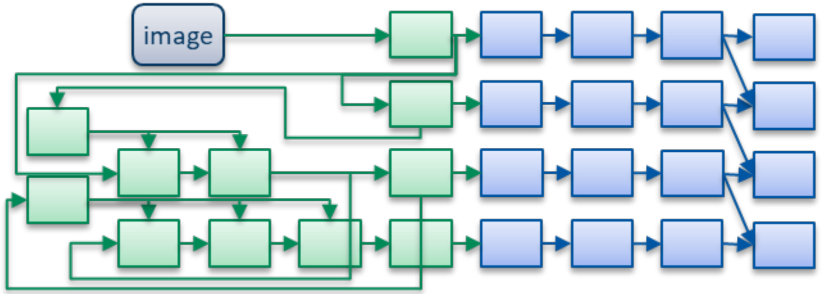
Applications → Platforms


S M
S M
...
S M
S M
Unified L2 Cache (768 KByte)
Global Memory (3 GByte)
Register File (32,768 x 32bit)
CC
16x LD / ST
SFU
L1 Cache / Shared Memory (64 kByte)

CPU **FPGA** **GPU**

Comparison of CPU, GPU and FPGA using OpenCL and AKAZE

- Evaluated for a 720p (1080p) resolution

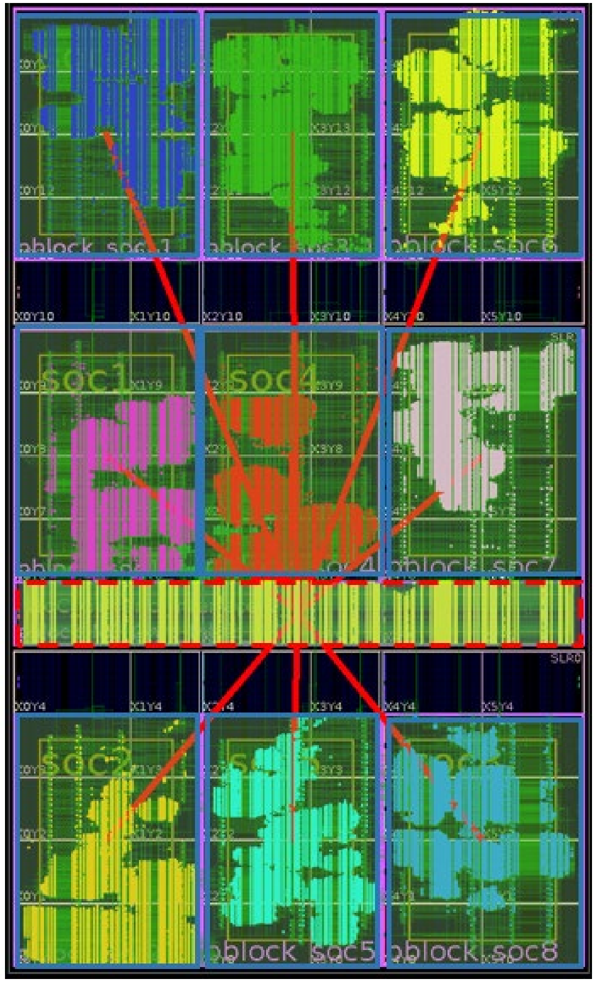


Device	Vendor	Model	Fab (nm)	Bandwidth (GB/s)
FPGA	Xilinx	Virtex-7 XC7VX690T	28	10.67
GPU	NVIDIA	GTX 780	28	288
CPU	Intel	Core-i7 4770k	22	25,6

	Speed Up	Power (watts)	Energy (mJ)
Single Thread	**1** (1)	23.8	1907
CPU OpenMP	**3.40** (3.42)	65.7	1528
CPU OpenCL	**5.15** (4.87)	66.9	1041
GPU OpenCL	**37.58** (42.77)	222.7	475
FPGA OpenCL	**55.14** (62.77)	30.5	45

Kalms L, Göhringer D (2017) Exploration of OpenCL for FPGAs using SDAccel and Comparison to GPUs and Multicore CPUs.In Proc. of the International Conference on Field Programmable Logic and Applications (FPL).

FPGA Architectures



Pros:

- ➢ Very flexible → You built your own hardware/processor!
- ➢ Can be adapted at design- and runtime → exploiting dynamic and partial reconfiguration
- ➢ Low-level parallelization
- ➢ Low power consumption compared to multi-cores

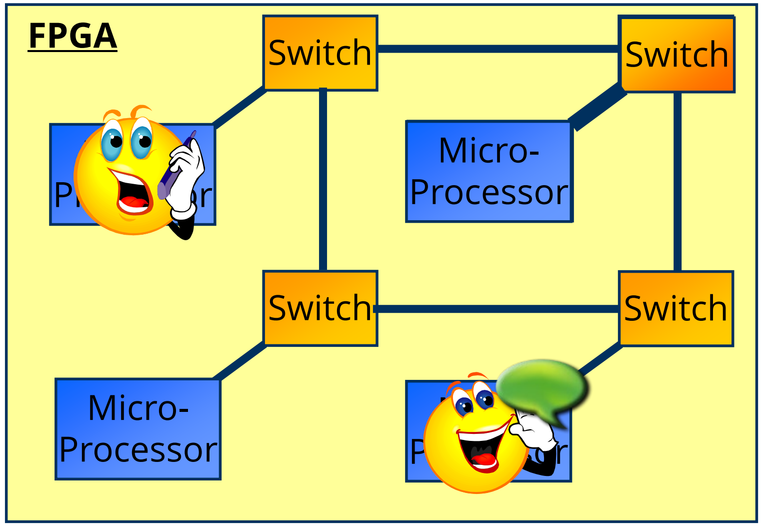
Cons:

- ➢ Difficult to program:
  - ➢ VHDL, Verilog → Not used by application engineers
  - ➢ C-to-FPGA tools (e.g. VivadoHLS)
    → only for accelerators
    → only a subset of ANSI-C, C++ is supported

Dynamic and Partial Reconfiguration

- ➢Manipulation of a fraction of the configuration data → the remaining hardware architecture stays operative and unaffected

- ➢For multi-cores → a processing element and its infrastructure can be substituted without disturbing the rest of the multi-core platform

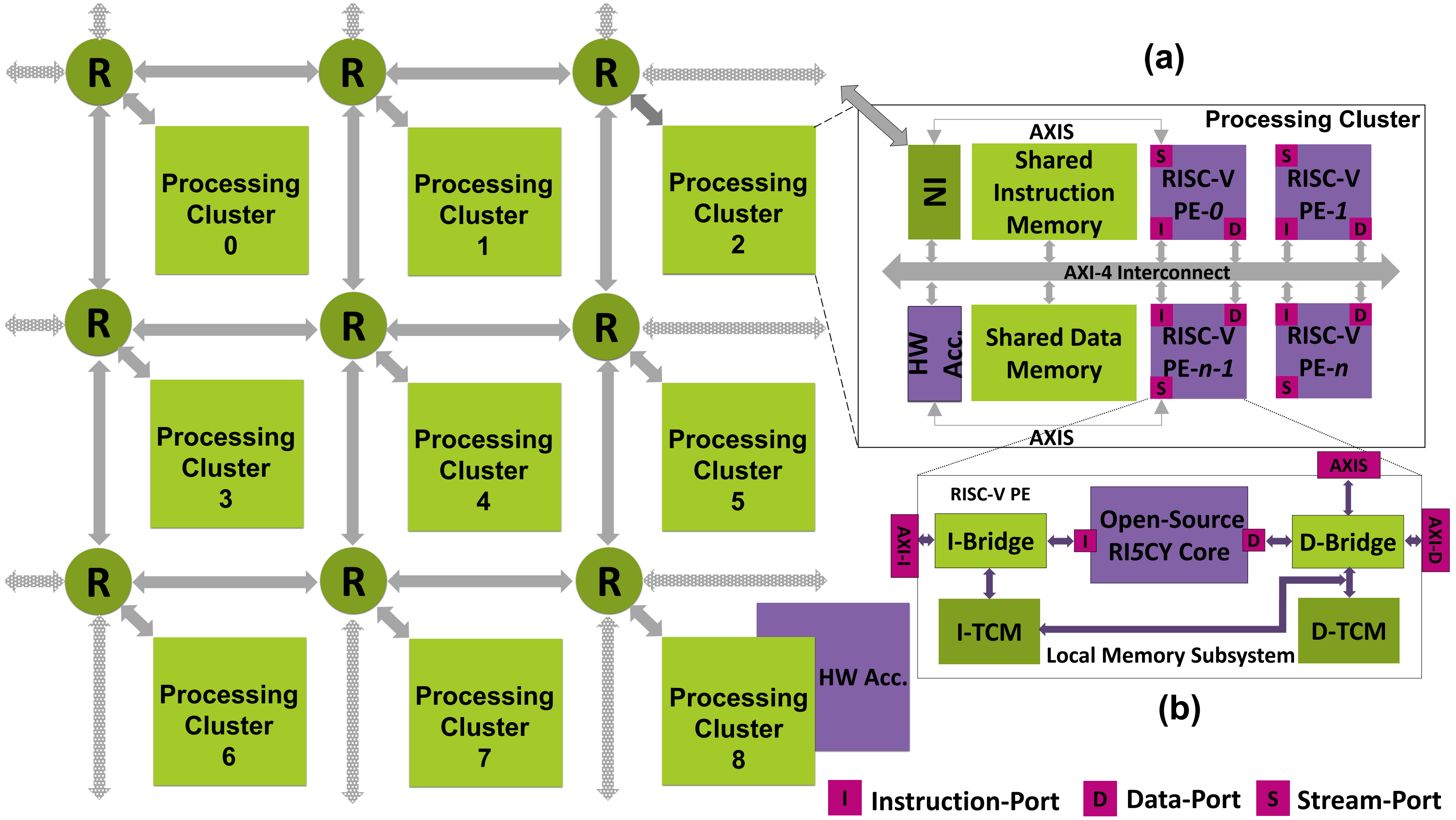


→ "Computing in time and space" : area utilization as well as the time variant content of the hardware device is run-time adaptive

Outline

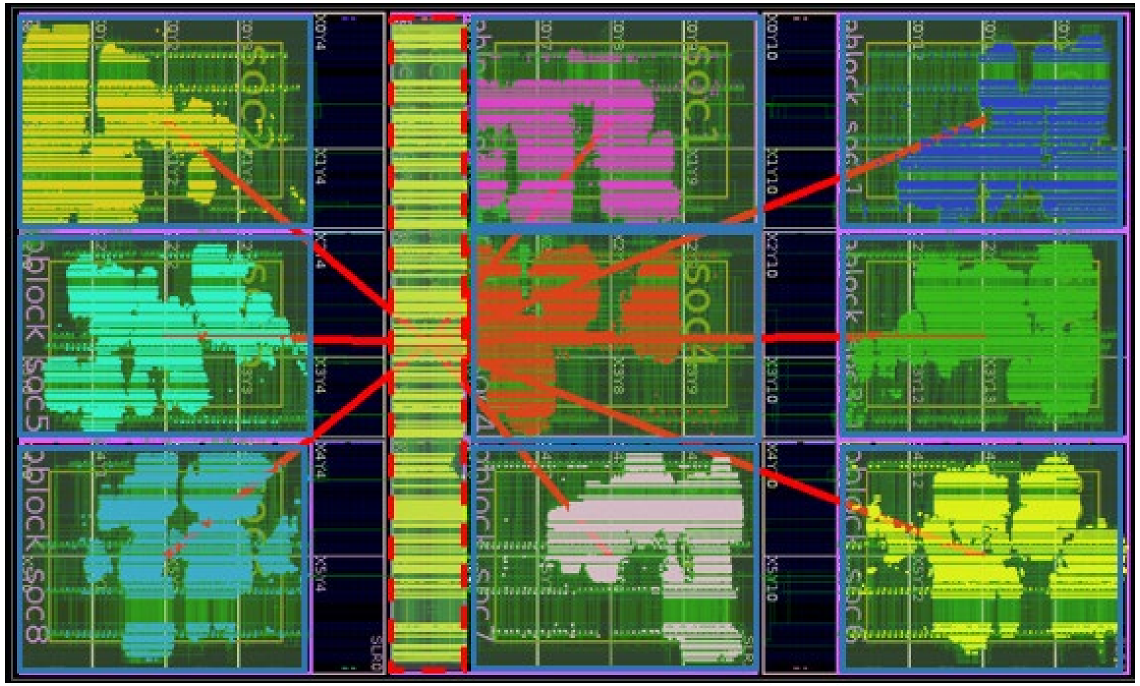
- Introduction and Motivation
- Reconfigurable FPGA Overlay Architecture
- Design / Programming Methodology
  - Image Processing Library: HiFlipVX
  - Application Partitioning and Mapping
- Application Examples in Relation to Funded Research Projects
- Conclusion and Outlook

Reconfigurable FPGA Overlay Architecture

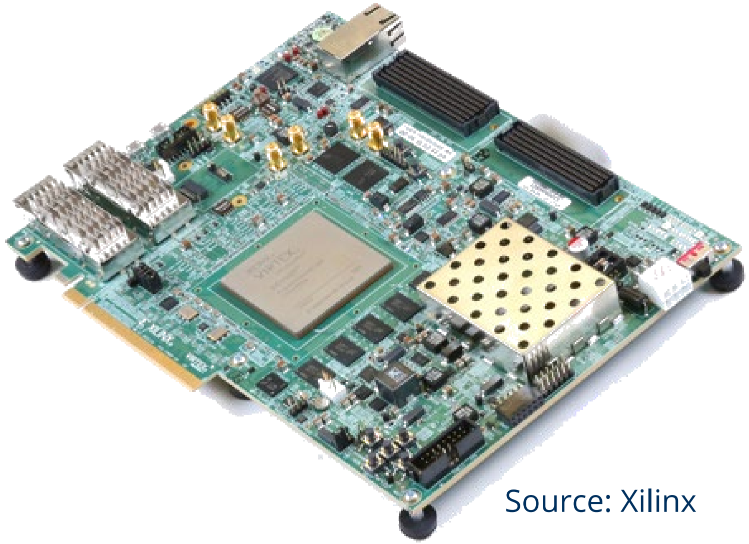


Kamaleldin A, Hesham S, Göhringer D (2020) Towards a Modular RISC-V Based Many-Core Architecture for FPGA Accelerators. IEEE Access.

Reconfigurable FPGA Overlay Architecture



▭ NoC PR ▭ Processing Cluster PR



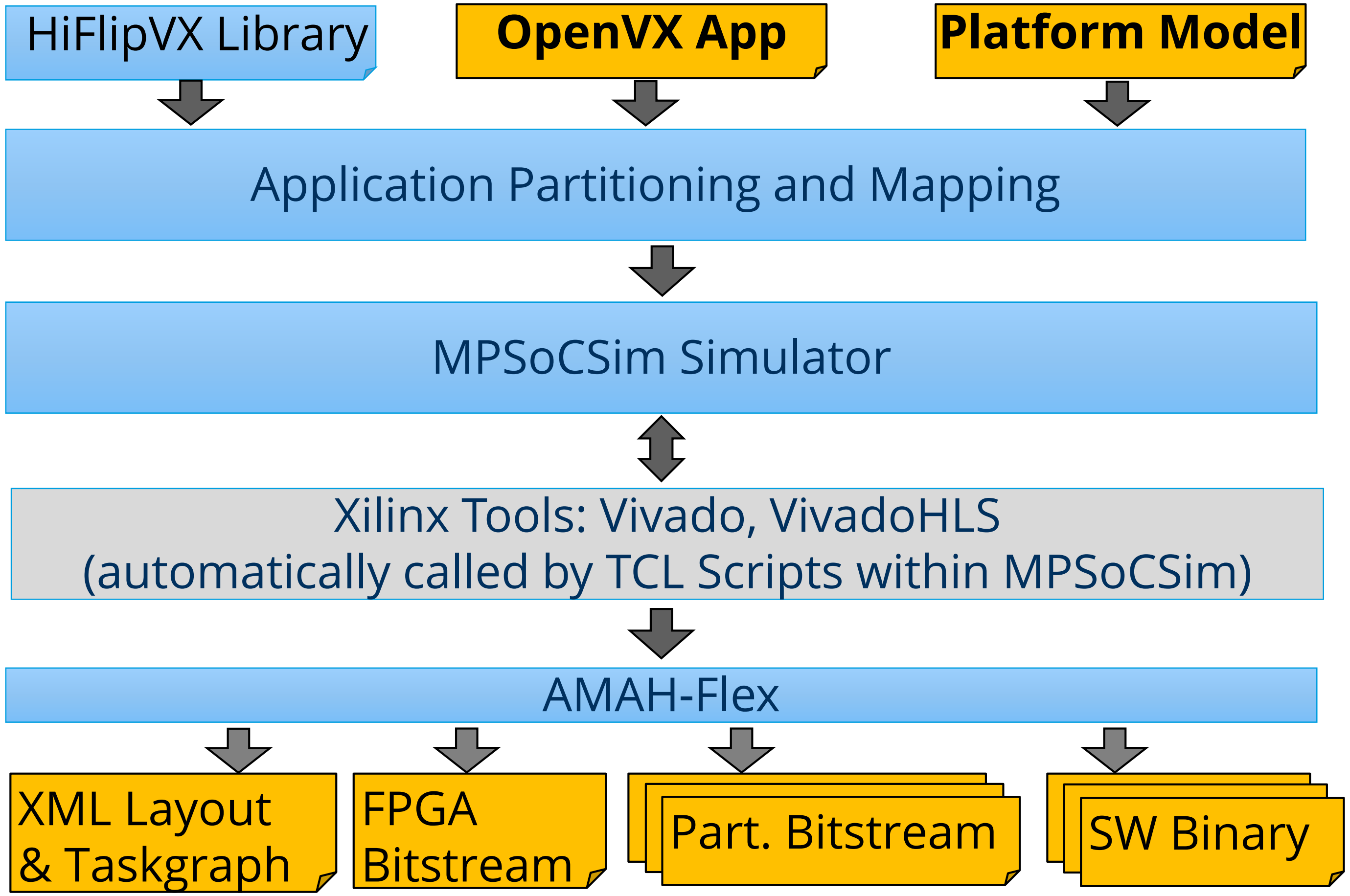
Source: Xilinx

Kamaleldin A, Hesham S, Göhringer D (2020) Towards a Modular RISC-V Based Many-Core Architecture for FPGA Accelerators. IEEE Access.

Outline

- Introduction and Motivation
- Reconfigurable FPGA Overlay Architecture
- Design / Programming Methodology
  - Image Processing Library: HiFlipVX
  - Application Partitioning and Mapping
- Application Examples in Relation to Funded Research Projects
- Conclusion and Outlook

Design and Programming Methodology



- Kalms L, Rad PA, Ali M, Iskander A, Göhringer D (2021) A Parametrizable High-Level Synthesis Library for Accelerating Neural Networks on FPGAs. Journal of Signal Processing, pp. 1-17.
- Kalms L, Göhringer D (2018) Scalable Clustering and Mapping Algorithm for Application Distribution on Heterogeneous and Irregular FPGA Clusters. Journal of Parallel and Distributed Computing (JPDC).
- Wehner P, Rettkowski J, Kalb T, Göhringer D (2016) Simulating Reconfigurable Multiprocessor Systems-on-Chip with MPSoCSim. ACM Transactions on Embedded Computing Systems (TECS), pp. 1-24.
- Charaf N, Tietz C, Raitza M, Kumar A,, Göhringer D (2021) AMAH-FLEX: A Modular and Highly Flexible Tool for Generating Relocatable Systems on FPGAs. In Proc. of the International Conference on Field-Programmable Technology (FPT). pp.1-6.

Image Processing Library - HiFlipVX

Open Source High-Level Synthesis FPGA Library for Image Processing

- ➢ Includes image processing and neural network functions
- ➢ Parametrizable and highly optimized for High-Level Synthesis (HLS)
- ➢ Functions are based on the OpenVX specification with some extensions
- ➢ Supports auto-vectorization & more data types
- ➢ Filters support different kernel sizes

Pixel-wise Operations

Bitwise AND	Min	Arithmetic Addition
Bitwise XOR	Max	Arithmetic Subtraction
Bitwise OR	Data Copy	Pixel-wise Multiplication
Bitwise NOT	Absolute Difference	Magnitude

Image Filter Functions

Gaussian Filter	Median Filter	Box Filter	Sobel
Custom Convolution	Dilate Image	Erode Image	Scharr 3x3

Image Conversion and Analysis Functions

Convert Bit Depth	Integral Image	Scale Image	Color Convert
Channel Combine	Channel Extract	Histogram	TableLookup

- Dávila-Guzmán M A, Kalms L, Gran Tejero R, Villarroya-Gaudo M, Suárez Gracia D, Göhringer D (2022) A Cross-Platform OpenVX Library for FPGA Accelerators. Journal of System Architecture.
- Kalms L, Rad PA, Ali M, Iskander A, Göhringer D (2021) A Parametrizable High-Level Synthesis Library for Accelerating Neural Networks on FPGAs. Journal of Signal Processing, pp. 1-17.
- Kalms L, Göhringer D (2020) Accelerated High-level Synthesis Feature Detection for FPGAs using HiFlipVX. In Towards Ubiquitous,Low-power Image Processing Platforms, pp. 61-78, Springer.
- Kalms L, Podlubne A, Göhringer D (2019) HiFlipVX: an Open Source High-Level Synthesis FPGA Library for Image Processing, 15th Int. Symposium on Applied Reconfigurable Computing. Architectures, Tools, and Applications (ARC), pp. 1-15.

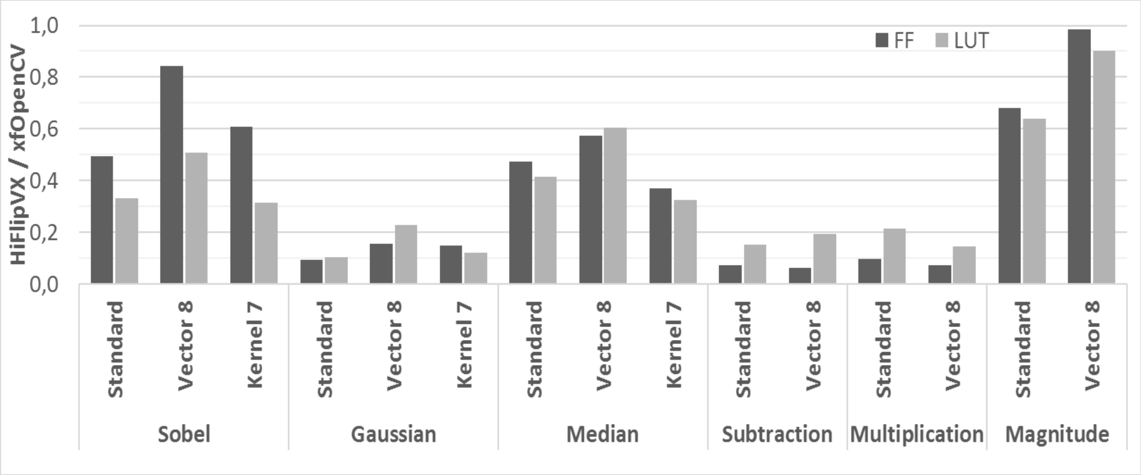
Image Processing Library - HiFlipVX

Features:

- Good Scalability
- Low resource usage
- Good comparison to related work

Standard configuration is:

- Vectorization: 1
- Kernel size: 1
- Data type: 8-bit



Comparison of HiFlipVX and xfOpenCV

Kalms L, Podlubne A, Göhringer D (2019) *HiFlipVX: An Open Source High-Level Synthesis FPGA Library for Image Processing*. In Proc. of Applied Reconfigurable Computing (ARC).

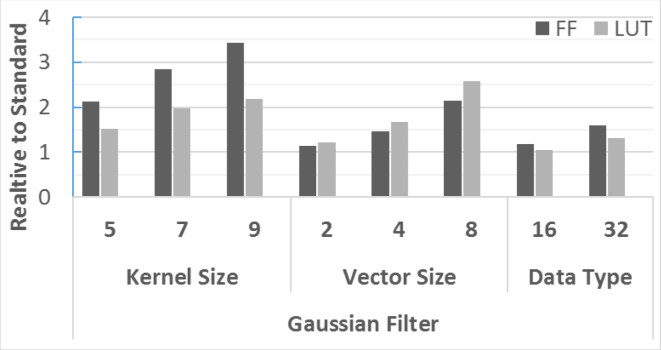
Image Processing Library - HiFlipVX

Features:

- Good Scalability
- Low resource usage
- Good comparison to related work

Standard configuration is:

- Vectorization: 1
- Kernel size: 1
- Data type: 8-bit



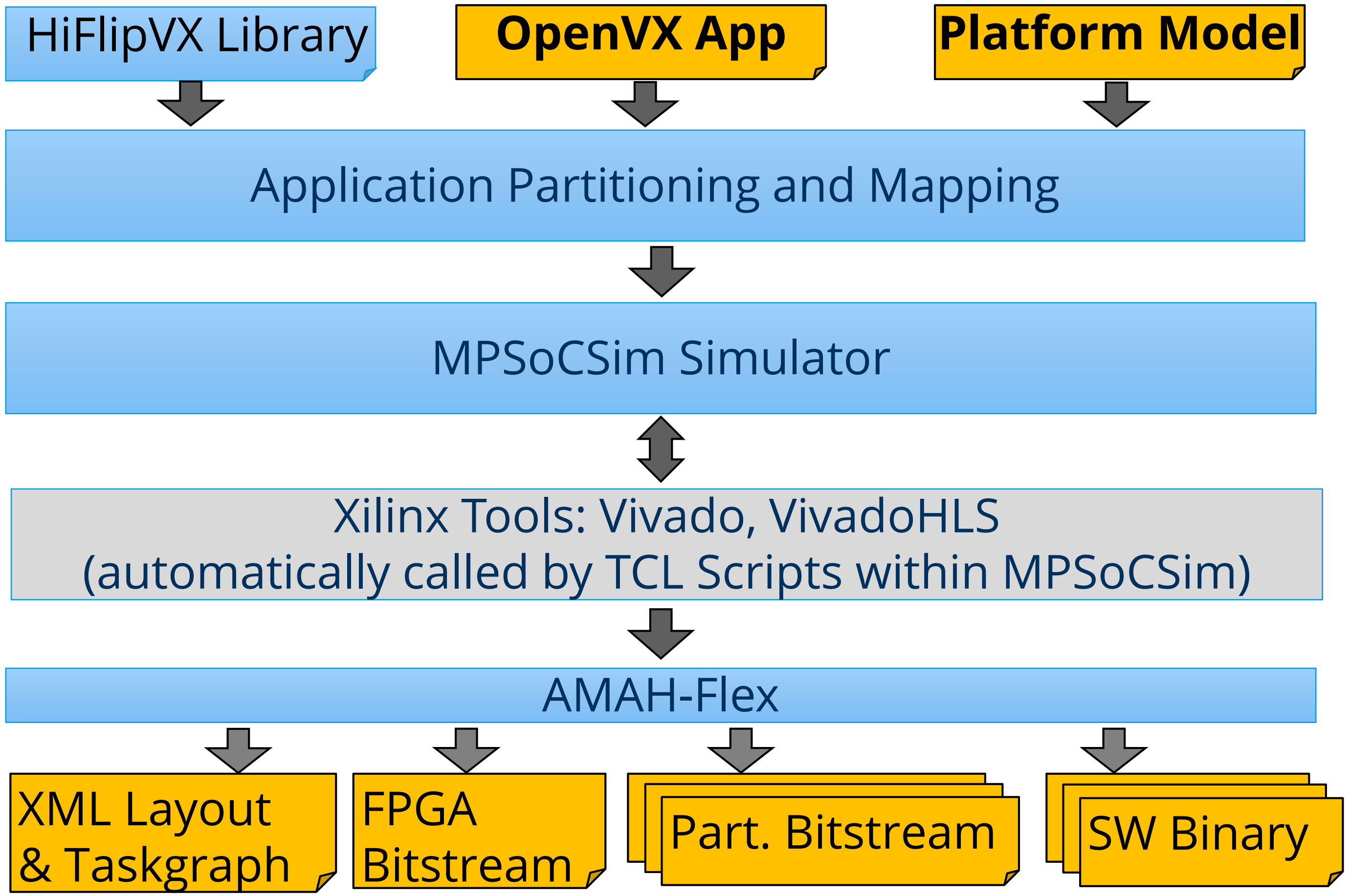
Scalability for different parameters

Function	FF	LUT	DSP	BRAM
Box Filter	257	536	2	2
Gaussian Filter	257	624	0	2
Sobel Filter	292	758	0	2
Median Filter	490	1180	0	2
AND, XOR,OR, Add, Subtract	27	171	0	0
Pixel-wise Multiplication	27	156	1	0
Magnitude	345	1106	0	0
Color Convert (RGBX to Gray)	50	258	2	0
Integral	82	389	0	4
Table Lookup (8-bit)	52	293	0	1
Histogram (8-bit)	113	593	0	2

Low resource usage

Kalms L, Podlubne A, Göhringer D (2019) *HiFlipVX: An Open Source High-Level Synthesis FPGA Library for Image Processing*. In Proc. of Applied Reconfigurable Computing (ARC).

Design and Programming Methodology

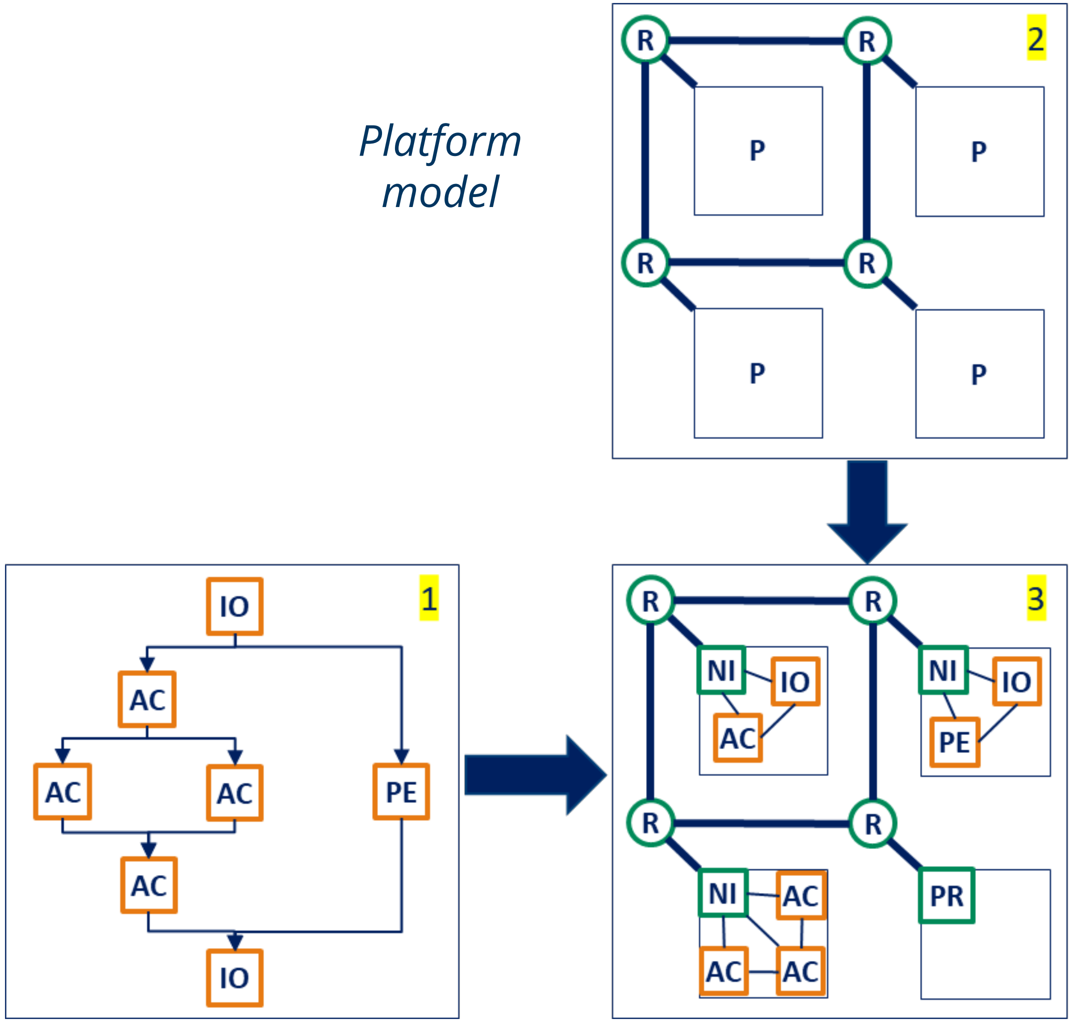


- Kalms L, Rad PA, Ali M, Iskander A, Göhringer D (2021) A Parametrizable High-Level Synthesis Library for Accelerating Neural Networks on FPGAs. Journal of Signal Processing, pp. 1-17.
- Kalms L, Göhringer D (2018) Scalable Clustering and Mapping Algorithm for Application Distribution on Heterogeneous and Irregular FPGA Clusters. Journal of Parallel and Distributed Computing (JPDC).
- Wehner P, Rettkowski J, Kalb T, Göhringer D (2016) Simulating Reconfigurable Multiprocessor Systems-on-Chip with MPSoCSim. ACM Transactions on Embedded Computing Systems (TECS), pp. 1-24.
- Charaf N, Tietz C, Raitza M, Kumar A,, Göhringer D (2021) AMAH-FLEX: A Modular and Highly Flexible Tool for Generating Relocatable Systems on FPGAs. In Proc. of the International Conference on Field-Programmable Technology (FPT). pp.1-6.

Application Partitioning and Mapping

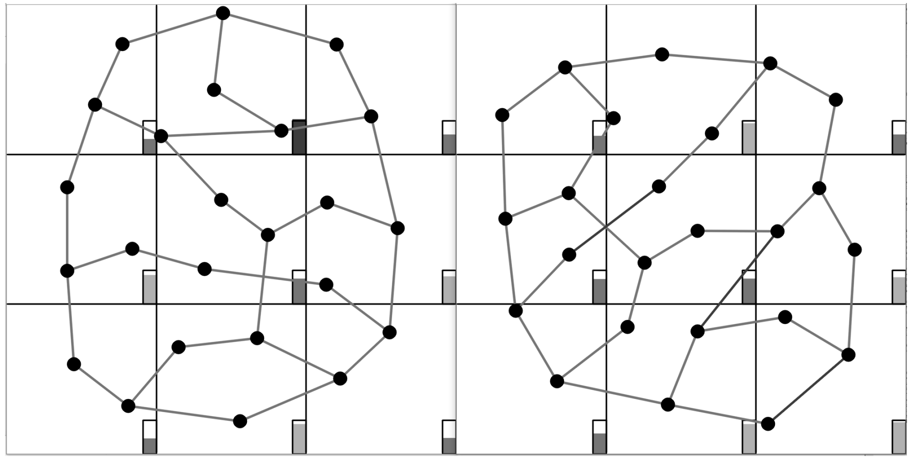
- Clustering & placement in 1 step
- Partitioning (task creation) is done by the developer
- Clustering and mapping are known as NP-hard for irregular graphs
- E.g. 9 partitions and 24 tasks would result in 9^{24} different mapping solutions
- Therefore, we use load balancing and heuristics

Kalms L, Göhringer D (2018) *Scalable Clustering and Mapping Algorithm for Application Distribution on Heterogeneous and Irregular FPGA Clusters*. Journal of Parallel and Distributed Computing (JPDC).


Platform model

Task Interaction Graph (TIG)

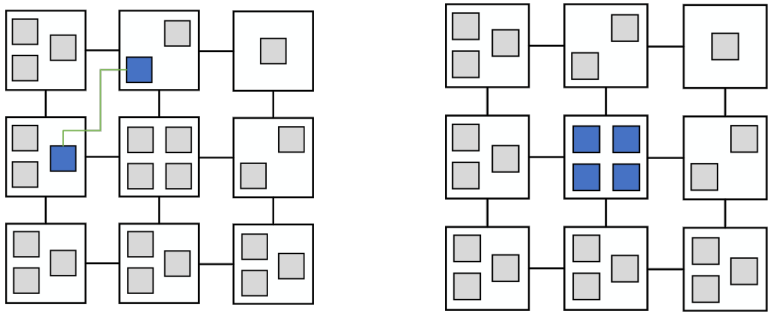
Application Partitioning and Mapping



Guest graph load balanced into host graph for two threads. Bars show maximum resource usage of partitions.

Algorithm overview:

- Load Balancing
  - Connected and unweighted edges are forced together like a rubber
  - All tasks are forced apart from each other based on their capacity utilization
- Optimization
  - Reduce maximum dilation between tasks
  - Reduce maximum partition capacity utilization of clusters
- Used optimization heuristics:
  - Gradient Descent, Taboo Search, Simulated Annealing, Parallelization



Dilation Optimization

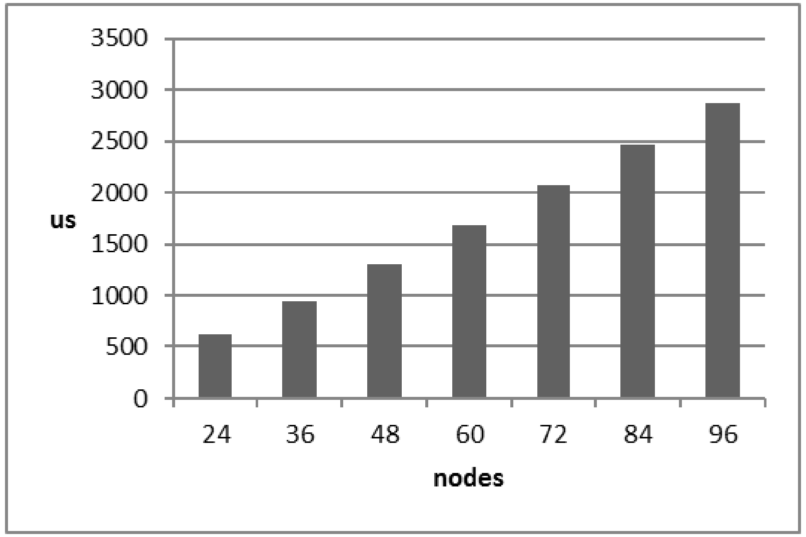
Capacity Utilization Optimization

Kalms L, Göhringer D (2018) *Scalable Clustering and Mapping Algorithm for Application Distribution on Heterogeneous and Irregular FPGA Clusters*. Journal of Parallel and Distributed Computing (JPDC).

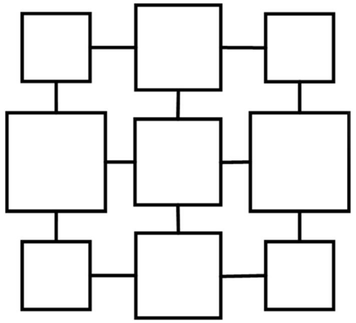
Application Partitioning and Mapping

Evaluation:

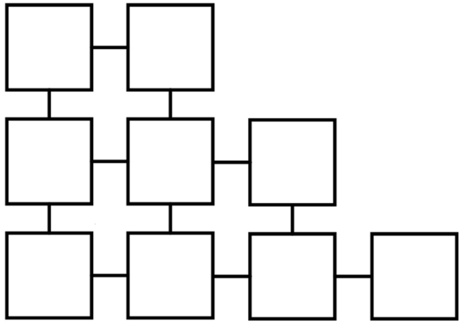
- ➢ Support different topologies
- ➢ Fast computation for mapping



Computation time of different TIGs for 3 × 3 Partitions that has a homogeneous and regular topology



Heterogeneous



Irregular

Kalms L, Göhringer D (2018) *Scalable Clustering and Mapping Algorithm for Application Distribution on Heterogeneous and Irregular FPGA Clusters*. Journal of Parallel and Distributed Computing (JPDC).

Outline

- Introduction and Motivation
- Reconfigurable FPGA Overlay Architecture
- Design / Programming Methodology
  - Image Processing Library: HiFlipVX
  - Application Partitioning and Mapping
- Application Examples in Relation to Funded Research Projects
- Conclusion and Outlook

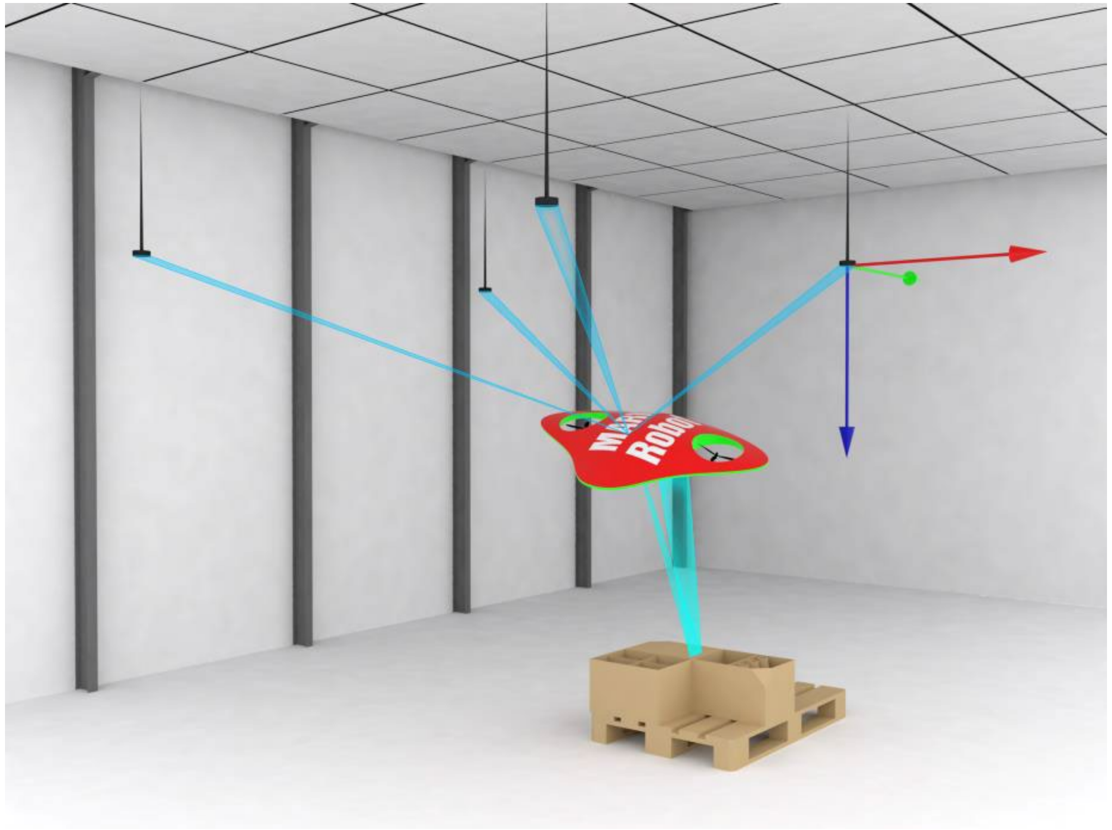
DFG SFB/Transregio MARIE (2017 – 2024)

Mobile Material Characterization and Localization by Electromagnetic Sensing

Marie
SFB / TRR 196

Challenges:

1) THz wave propagation measurement, analyzation and modeling
2) Small sub-mm-wave transceivers
3) Material characterization
4) Material localization



Our Contributions:

- ➢ Domain-specific computing architecture
- ➢ Design and programming methodology

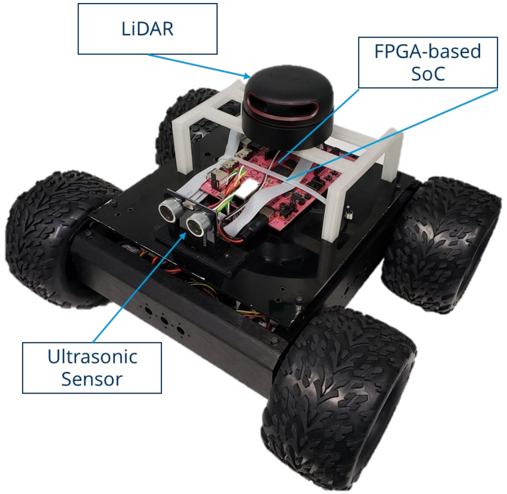
→ For enabling a real-time material map

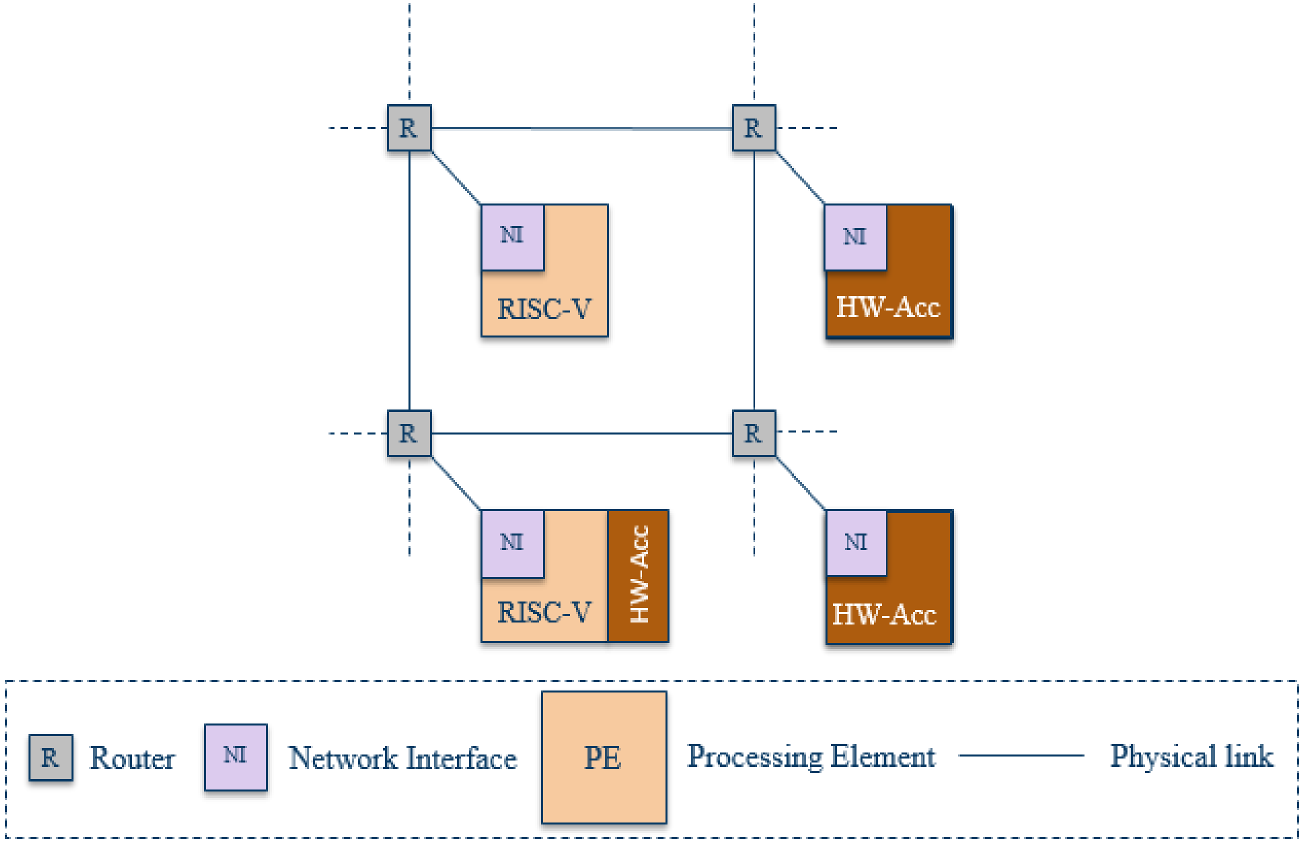
CORNET-AITIA (2019 – 2021)

Embedded AI Techniques for Industrial Applications

AI at the cutting edge

- Reconfigurable FPGA Overlay
  - Application-specific processors
  - Hardware accelerators
- Toolchains for ML/AI optimization
- Application partitioning and mapping tools
- Use case: Automotive and mobile robotics





Outline

- Introduction and Motivation
- Reconfigurable FPGA Overlay Architecture
- Design / Programming Methodology
  - Image Processing Library: HiFlipVX
  - Application Partitioning and Mapping
- Application Examples in Relation to Funded Research Projects
- Conclusion and Outlook

Summary and Outlook

Summary: Reconfigurable FPGA Overlay Architecture

- High flexibility → HW and SW can be adapted at design- and runtime to the application requirements, such as data throughput, real-time, safety
- High performance → Domain-specific architecture
- High energy efficiency → On demand functionality, better area utilization by hardware task multiplexing
- Programming support → Tool support for application partitioning and mapping, HiFlipVX

Outlook:

- Development and support of other/new:
  - Architecture components
  - Design methods and tools, new library functions for HiFlipVX
  - Runtime management systems
- Evaluation with further applications

Thank you! Questions?

Contact:

Prof. Dr.-Ing. Diana Göhringer

Chair of Adaptive Dynamic Systems
TU Dresden, Germany

Email: diana.goehringer@tu-dresden.de
www: http://www.tu-dresden.de/inf/ads