

#### **Open Hardware**

#### Jürgen Becker – KIT – EDABarCamp @ IBM

Institute for Information Processing Technologies (ITIV)



#### KIT - The Research University in the Helmholtz Association

#### www.kit.edu



### **Context and Motivation**

- Current embedded systems are subject to challenging requirements:
  - Increased performance is necessary to facilitate the execution of computationally intensive algorithms (machine learning, big data, ...)
  - Power and energy consumption must be minimized to facilitate the constraints of mobile and wireless devices (internet of things, ...)
  - A sufficient degree of **dependability** is necessary to employ digital systems in safety-critical environments (autonomous driving, ...)



Existing technology must evolve in order to meet these requirements ⇒ Open hardware can play a key role in this research process



## Agenda

#### Motivation

- OpenHardware as enabler in researching novel heterogeneous architectures
- EPI: European Processory Initiative
  - EPAC Accelerator, eFPGA
- InvasIC: Invasive Computing
  - *i*-Core: LEON3 Extension for reconfigurable near-memory computing

# ARoMA: Adaptive Redundancy for Manycore Architectures LEON3 Extensions for Adaptive Redundancy





InvasIC

#### ARoMA







#### **EPI – European Processor Initiative**

- Dec. 2018 to End of 2021
- Target markets: HPC and Automotive
- Proposal drivers:
  - Create a competitive European HPC and Automotive platform



#### → Mission: undependable EU Exascale machine by 2023

More Info: https://european-processor-initiative.eu



#### **EPI** Roadmap





Institute for Information Processing Technologies (ITIV)

### **EPI Common Architecture**



- MPPA Multi-Purpose Processing Array
  - eFPGA embedded FPGA
  - EPAC EPI Accelerator



#### **EPAC – EPI RISC-V Accelerator**



8 RISC-V based Vector Processing Units (VPUs)

- Stencil/Tensor Accelerators (STX), controlled by RISC-V cores
- RISC-V based Variable Precision Co-Processor (VRP)
- Coherent L2 Cache Banks
- NoC Interconnect



Copyright © European Processor Initiative 2019.

Institute for Information Processing Technologies (ITIV)



## **EPI** automotive

- Autonomous driving systems
- Connected mobility
- EPI: A powerful data fusion platform –the automotive embedded HPC platform
- EPI heterogeneous multicore architecture can provide enough performance and low power consumption in parallel



# **EPAC – Variable Precision Processor (VRP)**

- Efficient Computation in Scientific Domains
  - e.g., Finite Element Simulations
- Embedded as Functional Unit in RISC-V Core





### **EPAC – STX Accelerator**

- Domain-Specific acceleration of both Deep Learning and HPC workloads
- Up to 4 clusters of STX blocks controlled by RISC-V cores

Local Scratchpad Access



#### EPI – eFPGA



- Provided by Menta S.A.S.
- Optimized for GP-HPC and automotive applications
  - e.g. image processing using machine learning

Run-time reconfigurable crypto and general purpose accelerators



# EPI – eFPGA automotive application scenarios

- Face detection for access control = unlock and engine start
- Object detection can be used as an early stage for an ADAS use case
- Use of state-of-the-art machine learning algorithms





# **INVASIC ARCHITECTURE**

**14** February 19, 2020 J. Becker - edaBarCamp Lightning Talk

Institute for Information Processing Technologies (ITIV)



### **InvasIC Heterogeneous Architecture**

- Tiled many-core architecture
  - One or more CPUs per tile
  - Network-on-Chip (NoC) interconnect
  - Tile Local Memory (TLM)
- PGAS memory architecture
- LEON3 cores and GRLIB from Cobham Gaisler
- NoC router and network adapter offering guaranteed service connections
- FPGA Prototype



# **Near Memory Computing (NMC)**

- Bandwidth is limiting factor of data-centric workloads
  - Neural Networks, Big Data
  - Moving data > processing data
- In-Memory Computing
  - Processing inside memory chips
  - Dependant on data locality
  - High parallel bandwidth
- Close-to-Memory Computing
  - Accelerators next to memory controller
  - Less restricted by data locality



M. Gao and C. Kozyrakis, "HRL: Efficient and flexible reconfigurable logic for near-data processing," in 2016 IEEE International Symposium on High Performance Computer Architecture (HPCA), March 2016

# NMC in InvasIC

#### Networks-on-Chip (NoC)

- Complex memory hierarchy
- More layers of data locality (Global Memory/Tile-Local Memory)

#### i-Core: Runtime-adaptive processor

- Integrated reconfigurable FPGA fabric
- Close to TLM with high bandwidth
- Near-memory computations







#### **AROMA - ADAPTIVE REDUNDANCY FOR MANYCORE ARCHITECTURES**

### ARoMA

- Current mixed criticality systems lacking fault-tolerance
- Adaptive Redundancy for Manycore Architectures (ARoMA) targets a high computing performance coupled with high safety integrity to close the gap
- Cobham Gaisler's LEON3 as enabler for research on processor architectures
  - Open source integer unit
  - Open source L1-cache
  - Free ecosystem (compiler, debug tools)
- ARoMA supports different fail-operational modes
  - Runtime-adaptive lockstep architecture
  - Runtime-adaptive cache-based check-pointing



Institute for Information Processing Technologies (ITIV)

### **ARoMA – Pipeline Extension**



- Based on the LEON3's open source 7-stage pipeline implementation
- The integer unit (IU) is extended by a commit stage performing the following tasks
  - Compares the results of the execution stages of two pipelines to detect faults
  - Determine PC and NPC for the rollback
  - Load of a processor state to realize an adaptive runtime lockstep cluster
  - Save current processor state
  - Restore saved processor state to release the lockstep cluster and execute the previous software
- The fetch stage is modified to support the rollback mechanism



#### **ARoMA – Cache Extension**



- Based on the LEON3 open source Level-1 cache controller
- Modification of the Level-1 data cache to realize an adaptive write strategy
  - Write-through with cache coherence
  - Write-back without cache coherence
- Utilizing the incoherent write-back caches to realize cache-based checkpoints. Valid data will be written back to the memory. Faulty data initiate an rollback

