MULTILEVEL SIMULATION-BASED CO-DESIGN OF NEXT GENERATION HPC MICROPROCESSORS

LILIA ZAOURAR, MOHAMED BENAZOUZ, AYOUB MOUHAGIR, FATMA JEBALI, TANGUY SASSOLAS, JEAN-CHRISTOPHE WEILL, CARLOS FALQUEZ, NAM HO, DIRK PLEITER, ANTONI PORTERO, ESTELA SUAREZ, POLYDOROS PETRAKIS, VASSILIS PAPAEFSTATHIOU, MANOLIS MARAZAKIS, MILAN RADULOVIC, FRANCESC MARTINEZ, ADRIÀ ARMEJACH, MARC CASAS, ALEJANDRO NOCUA, ROMAIN DOLBEAU

PMBS21: THE 12TH INTERNATIONAL WORKSHOP ON PERFORMANCE MODELING, BENCHMARKING AND SIMULATION OF HIGH-PERFORMANCE COMPUTER SYSTEMS

MULTILEVEL SIMULATION-BASED CO-DESIGN OF NEXT GENERATION HPC MICROPROCESSORS

Lilia Zaourar, Mohamed Benazouz, Ayoub Mouhagir, Fatma Jebali, Tanguy Sassolas, Jean-christophe Weill, Carlos Falquez, Nam Ho, Dirk Pleiter, Antoni Portero, Estela Suarez, Polydoros Petrakis, Vassilis Papaefstathiou, Manolis Marazakis, Milan Radulovic, Francesc Martinez, Adrià Armejach, Marc Casas, Alejandro Nocua, Romain Dolbeau

PMBS21: THE 12TH INTERNATIONAL WORKSHOP ON PERFORMANCE MODELING, BENCHMARKING AND SIMULATION OF HIGH-PERFORMANCE COMPUTER SYSTEMS

European Processor Initiative

cea

JÜLICH Forschungszentrum

FORTH Foundation for Research & Technology - Hellas

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

Atos

SIPEARL The Silicon Pearl

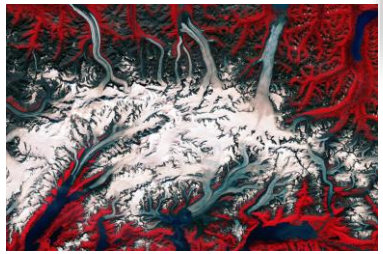
AGENDA

- Context : EPI Design Challenges
- State of the art
- Co-design methodology
  - Gem5
  - SESAM/VPSim
  - MUSA
- Reference Architecture & Design Concerns
- Experiments
  - SVE Register Length
  - Memory Bandwidth
  - NoC Dimensioning
- Conclusion

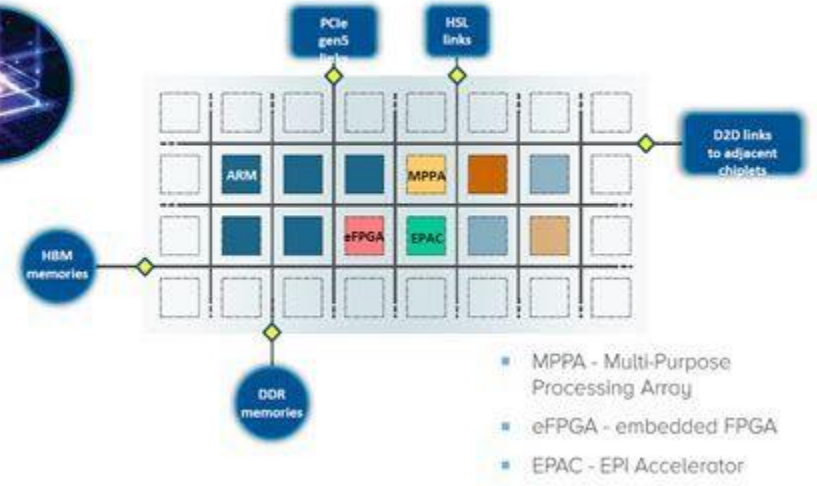
Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

CONTEXT

- European Processor Initiative (EPI)
  - Develop a complete EU designed high-end microprocessor, addressing Supercomputing and edge-HPC segments
  - Develop customized processors able to meet the performance needed
- Design high performance multicore processors
  - Various requirements : automotive, cryptography, AI, health, etc.
- Complex architectural trade off set up for General Purpose Processor (GPP) and accelerators
  - Arm architecture
  - Accelerator implementing RISCV
- Various design concerns
  - Cores (types, number), Processing Units, etc.
  - Communication bandwidth, on chip memories, etc.



GPP AND COMMON ARCHITECTURE



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

NEED FOR CO-DESIGN

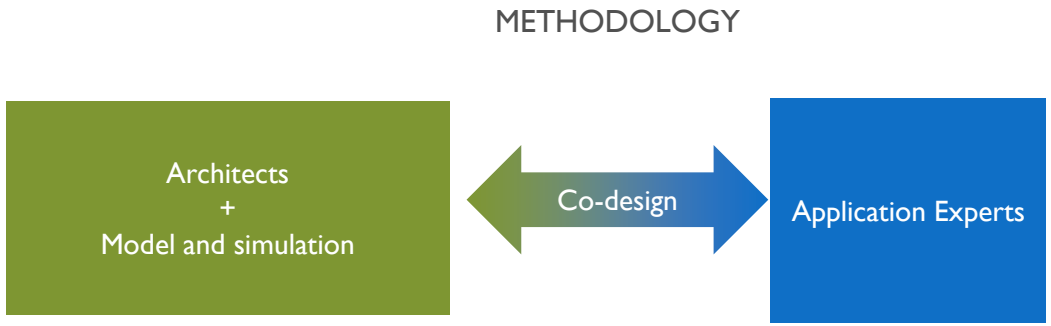
Why?

- Facilitate trade-off decisions to maximize performance and minimize costs under given technology boundary constraints
- Achieve best application performances

What?

Bi-directional and iterative interaction process

- Application experts
- Hardware/System Software developers

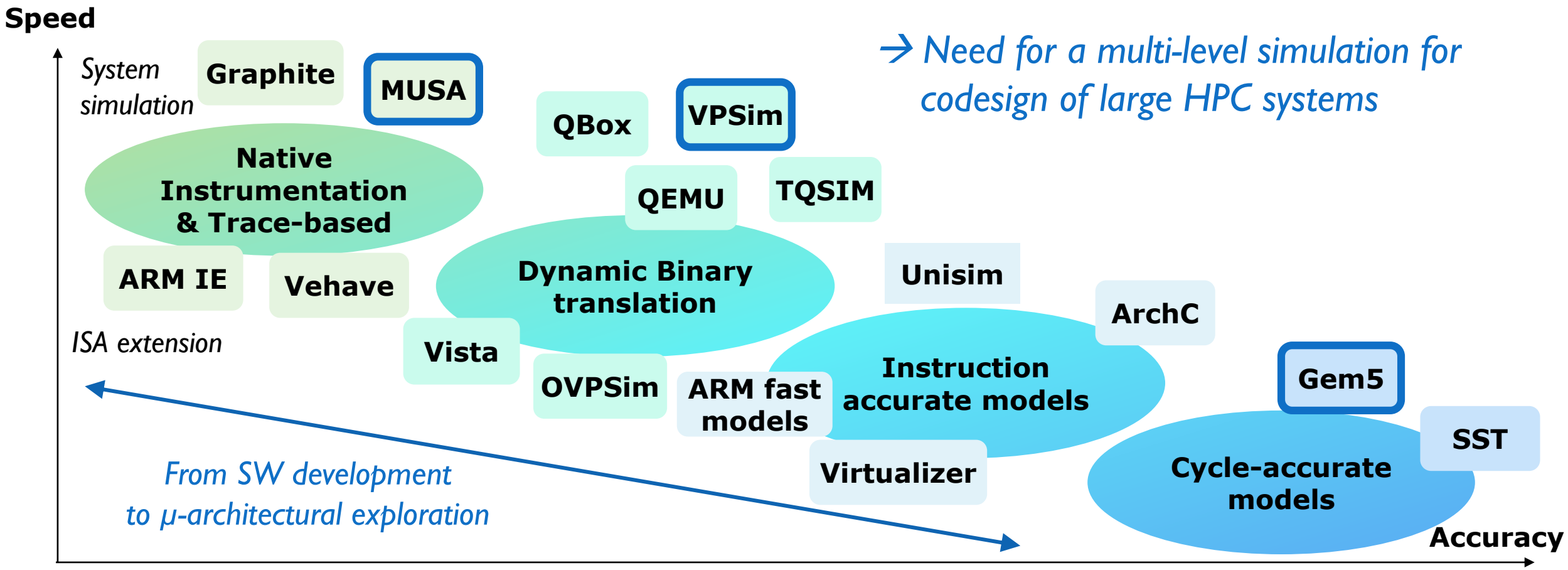


How?

- Identify specific application's needs
- Multilevel hierarchy of models and simulators
- Feed insights/suggestions into EPI's Hardware/Software technologies and developments

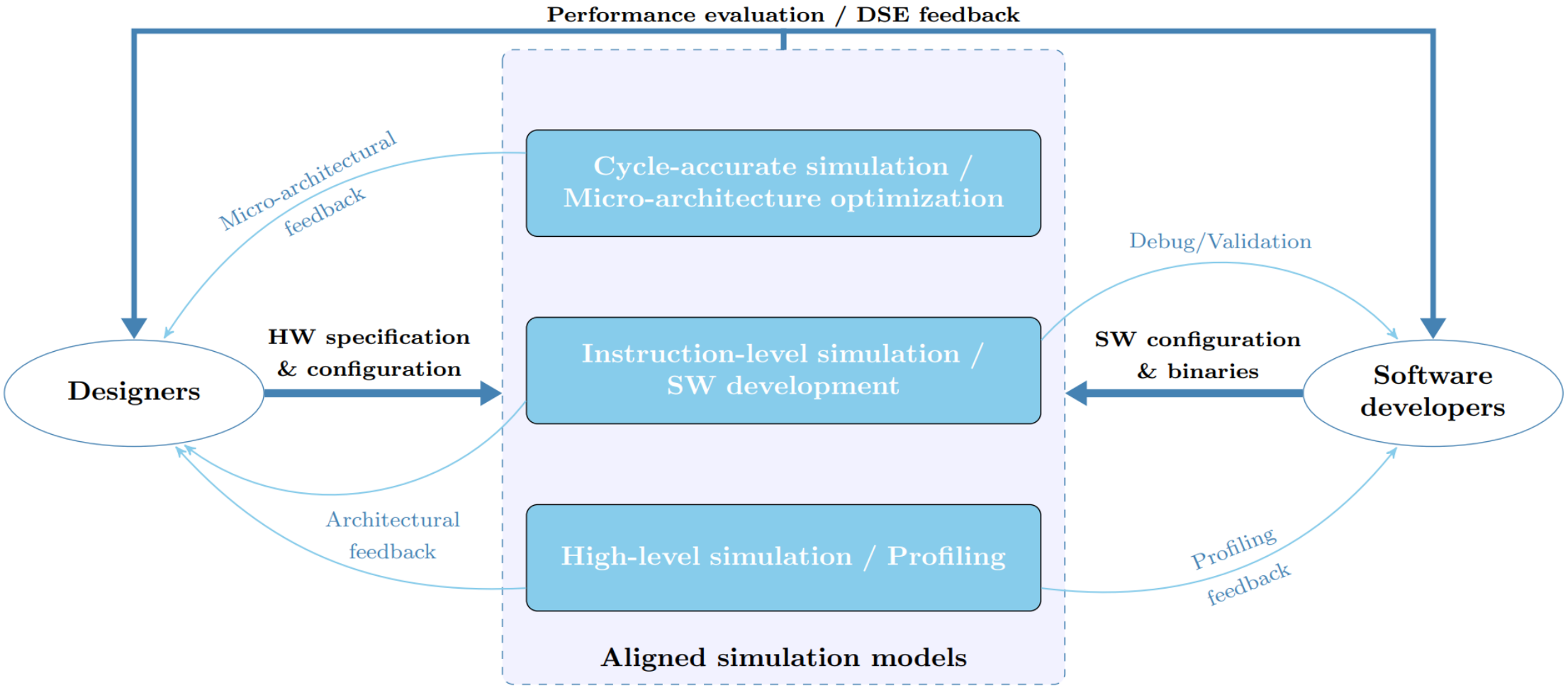
Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

STATE OF THE ART



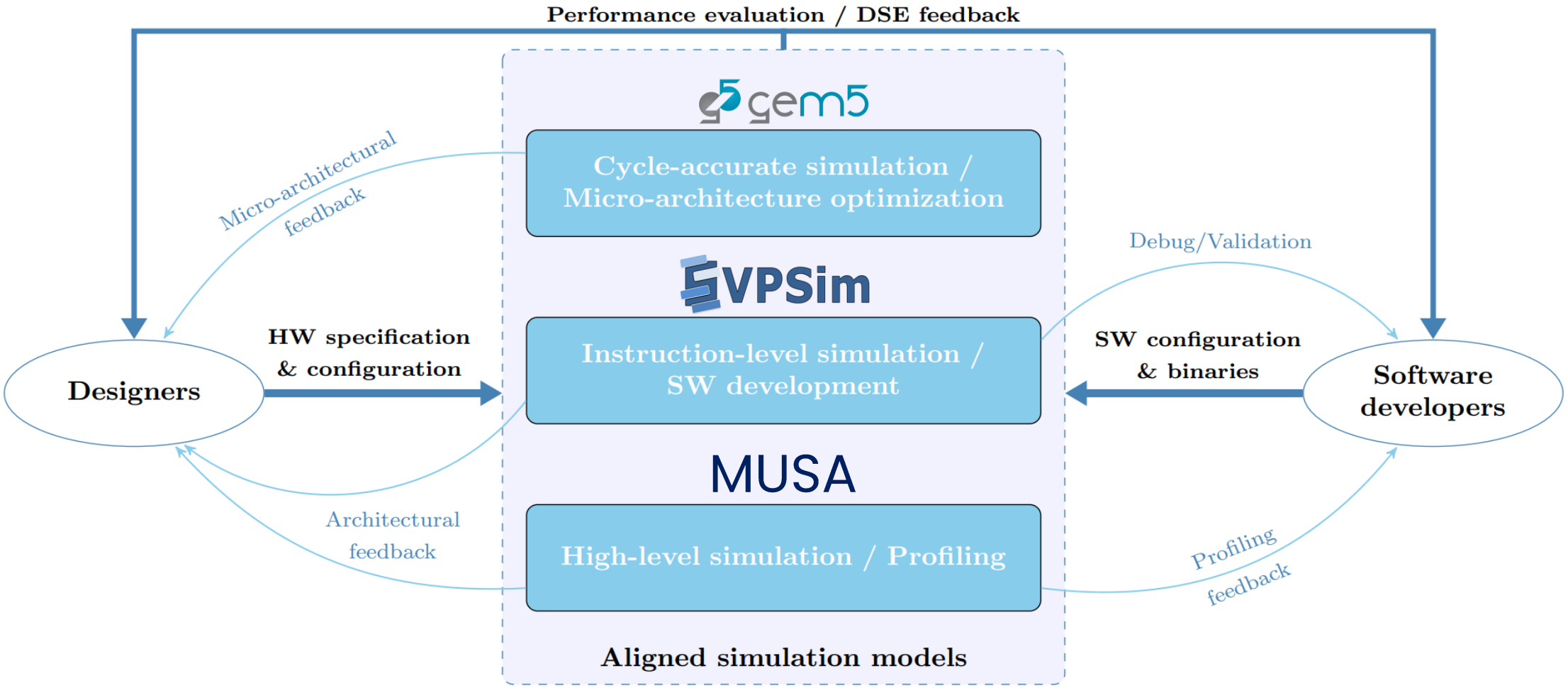
Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

CO-DESIGN METHODOLOGY



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

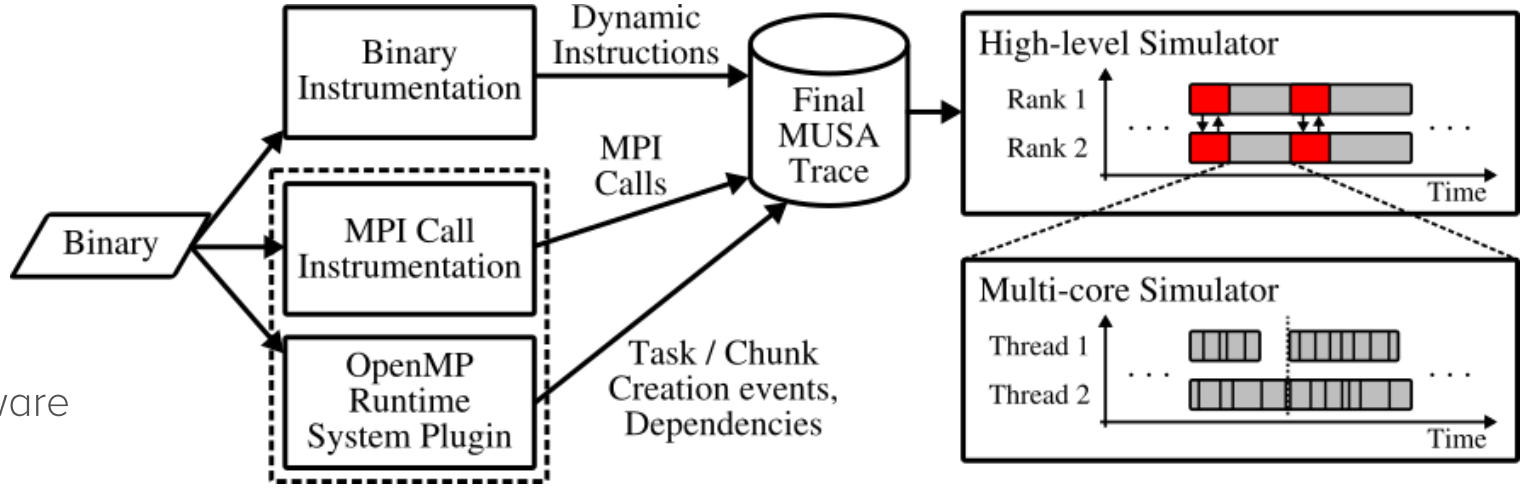
CO-DESIGN METHODOLOGY



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

MUSA OVERVIEW

- Enables large-scale simulations
  - Different communication networks
  - Core counts per node
  - Micro-architectural parameters
  - Detailed memory models (DRAMSim2 and RAMULATOR)
  - Considers the effects of system software

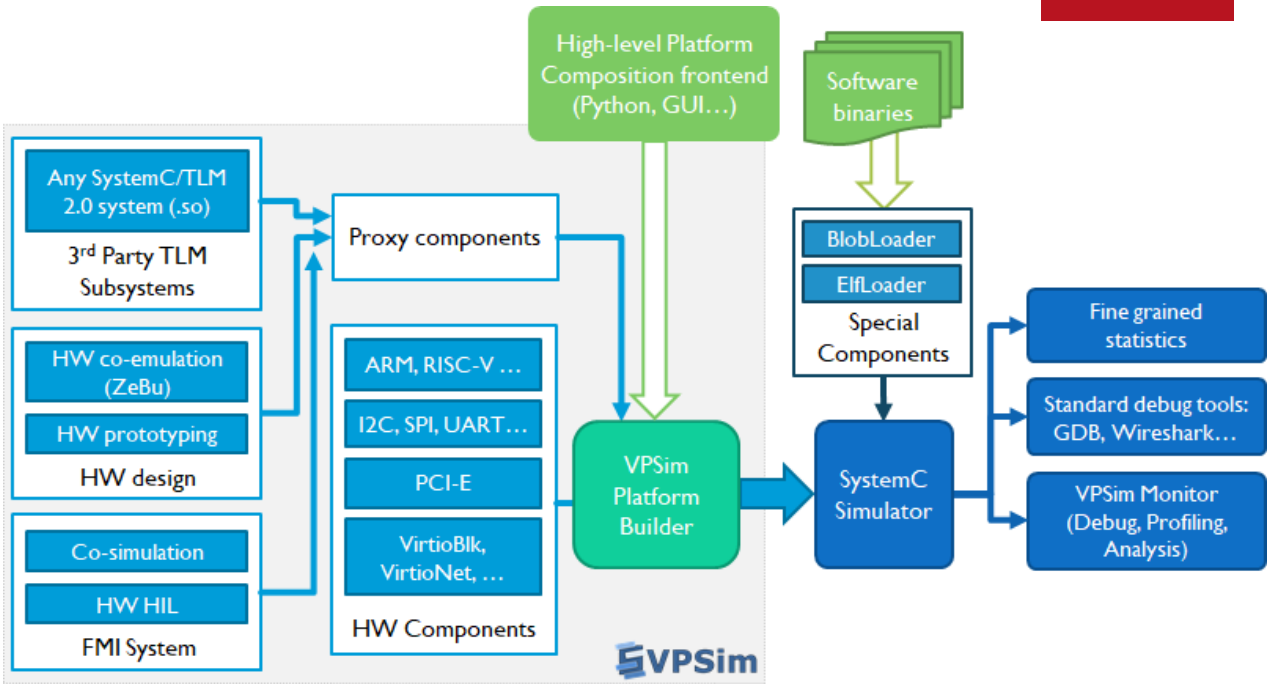


- Combine high-level event (MPI, OpenMP) traces with detailed instruction traces
- Detailed instruction traces supported for both Arm-v8 and RISC-V binaries
  - Out-of-order core model with renaming and memory dependencies

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

SESAM/VPSIM

- Virtual prototyping Environment
  - Early Software development
  - Hardware/Software Co-Design
  - Performance profiling and debug
  - Support all levels from BIOS to OpenMPI
- Easy interfacing thx to SystemC/TLM 2.0 and FMI
- Fast platform description with Python
  - Large and flexible IP portfolio (Arm, RiscV)
- Rapid simulation able to run full software stacks
  - From hypervisor, to full-fledged applications with standard debugging features



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

GEM5 SIMULATOR

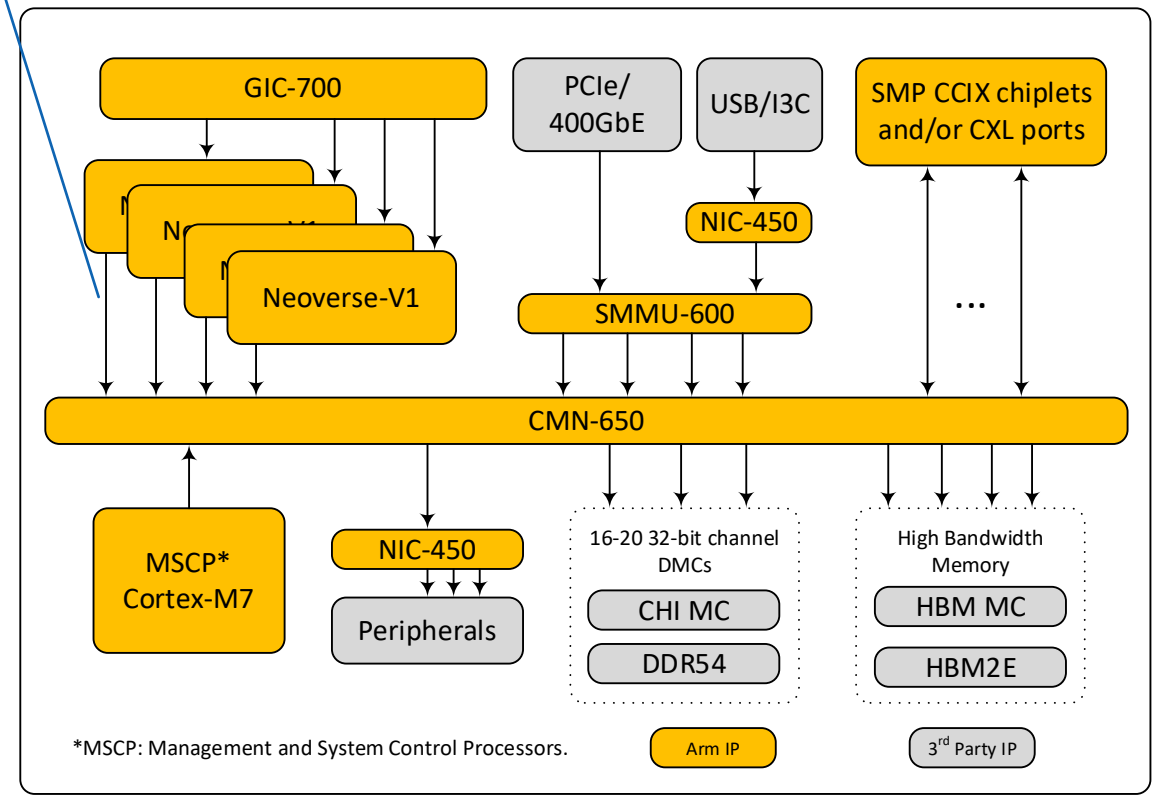
- Modular simulation framework
  - Cycle accurate
  - Supports the modeling of various HW platforms
- Used for computer-system architecture research
- Community-led Open source project (Open governance model)
- Suitable
  - Processor microarchitecture
  - Cache Coherent NoC modeling (Ruby subsystem)
  - Full System Simulation (with OS)
- Combines {O3 Processor, ARM SVE, CC NoC, Detailed Memory models (HBM2)} in a single framework

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

REFERENCE ARCHITECTURE

Scalable Vector Extension (SVE): two 256 bits wide units (16 double-precision FLOP-cycle)

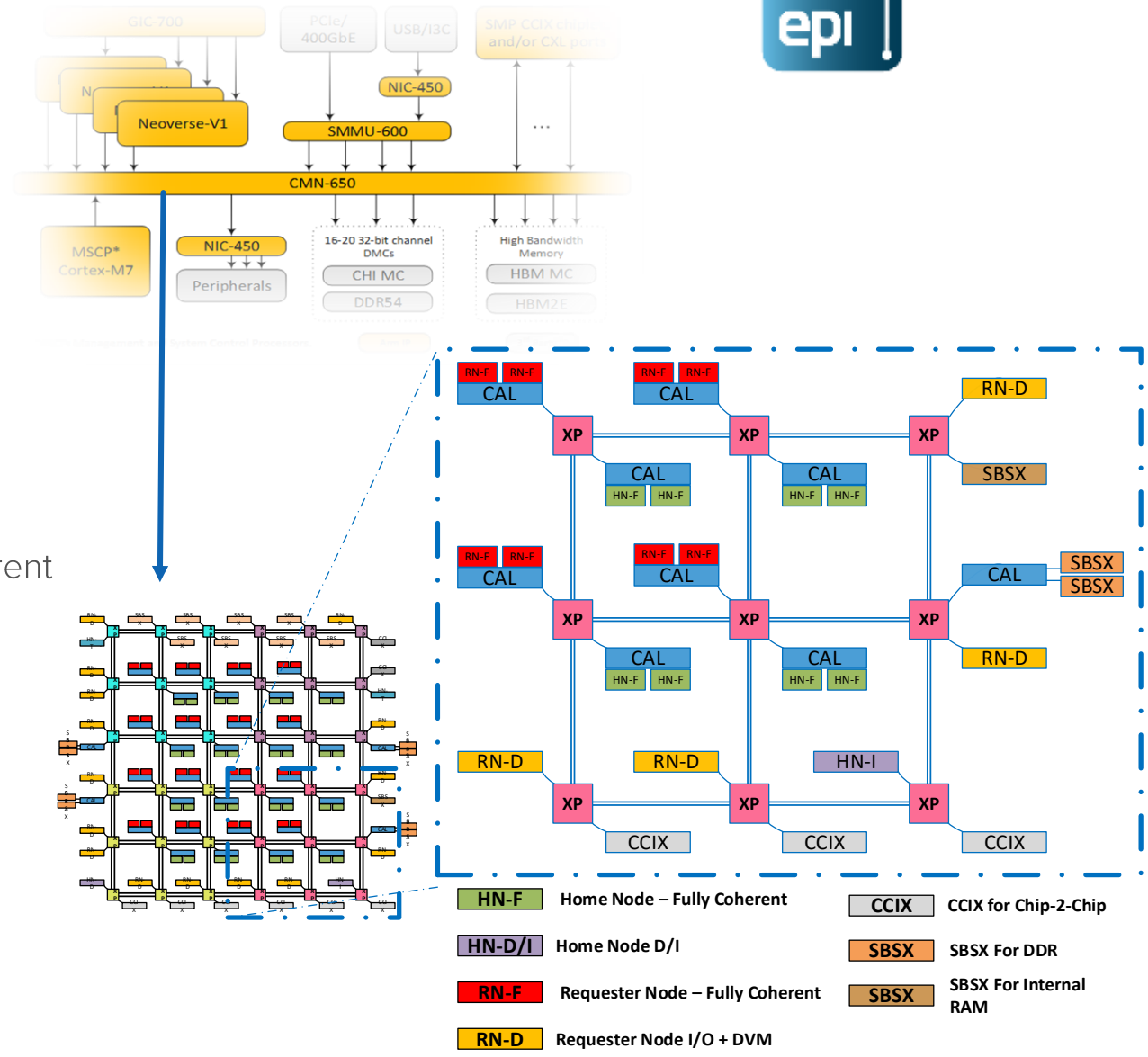
- NEOVERSE V1 REFERENCE DESIGN (RD-V1)
  - **Processing Element:** High-Performance Armv8,4A Arm Neoverse V1 Cores. One core contain, 1MB private L2 Cache and supports DVFS
  - **Interconnect Element:** Includes a coherent mesh network (CMN-650), an interrupt controller (GIC-700) & a system memory management unit (MMU-600)
  - **MSCP Element:** The Manageability and System Control Processor, implements the Cortex-M7 based SCP & MCP. The SCP manages the overall power, clock, reset, and system control of RD-V1, while the MCP manages communication with external Baseboard
  - **Memory Element:** Memory controllers that implements AMBA AXI data path to the CMN-650, such as DDR54 or HBM memories



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

REFERENCE ARCHITECTURE

- Coherent Mesh Network (CMN)
  - Highly scalable mesh network (up to 10x10)
  - Custom mesh size and placement
  - Programmable System Address Map
  - Up to 256 RN-F interfaces for CHI-Based compute clusters, accelerators, etc.
- Mesh Structure
  - **Requesting Masters:** Number and type, such as RN-F for the Neoverse-V1 cores, or RN-D for masters without internal coherent caches.
  - **Home Nodes:** The SLC & Snoop Filter Size requirements determines the HN-F instances
- Configurations
  - **Config-M:** 3x5 Mesh Network, 16 Neo-V1 cores., 16MB SLC, 32 MB SnoopFilter.
  - **Config-L:** 6x6 Mesh Network. 32 Neo-V1, 32MB SLC, 64 MB SnoopFilter.



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

DESIGN CONCERNS

- Vector processing resources
  - impact of SVE register length of Neoverse V1 cores
- Cache and main memory system performance
  - memory access performance
  - on-chip memory size dimensioning
  - external I/O bandwidth requirements
- NoC topology and IP block placement
  - Consider various design scale

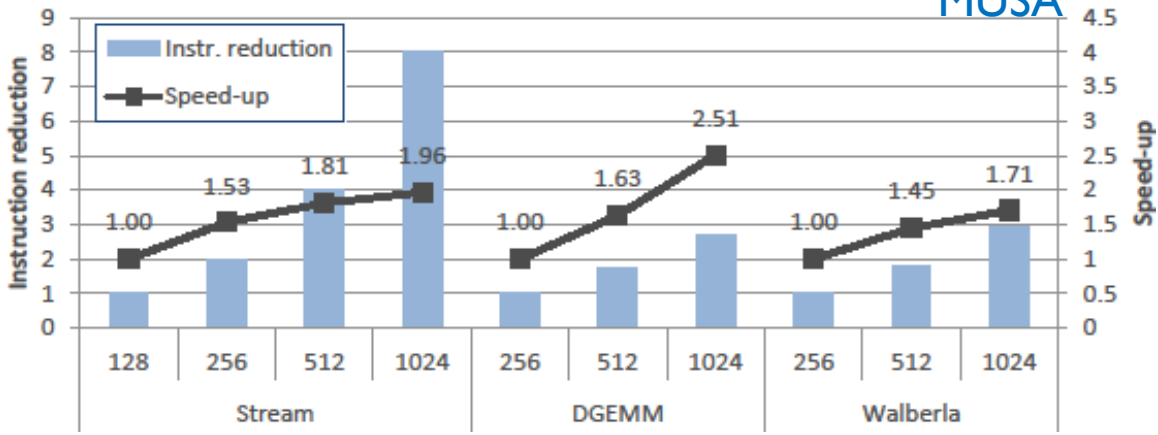
Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

BENCHMARKS

- DGEMM (Double-precision, GEneral Matrix-Matrix multiplication)
  - Developed within the BLIS framework. The binaries were compiled with fixed SVE vector length
  - Compute-bound benchmark for assessing CPU performance
- STREAM Triad
  - Representative for HPC applications sensitive to the available system bandwidth
  - Performance measure : the utilization percentage of memory bus
- WaLBerla : Example of a stencil kernel
  - It contains efficient, hardware specific compute kernels
  - Performance measure: the attained number of Million Lattice Updates Per Second (MLUPS)
- PARSEC and SPLASH-2 suites
  - HPC-targeted applications that mimic large-scale emerging programs
  - NoC & shared memory

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

NUMERICAL RESULTS AND ANALYSIS



Do wider SVE SIMD instructions improve performance?

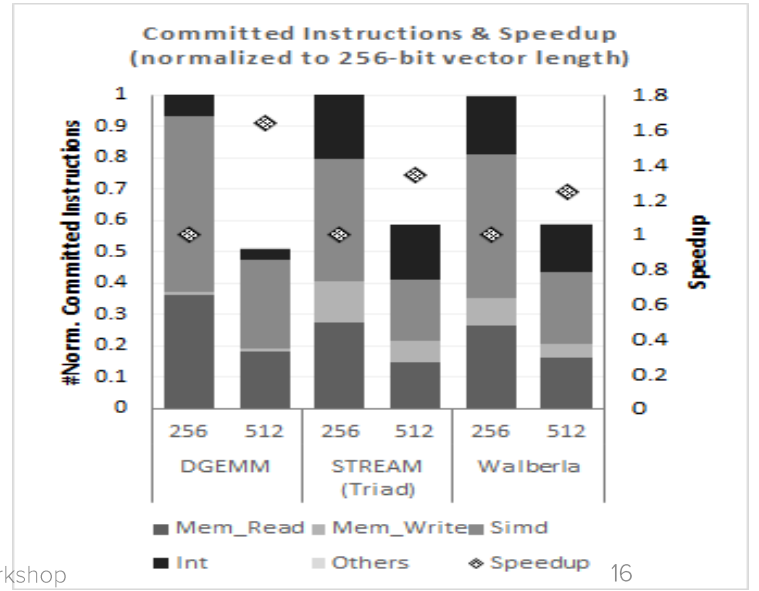
- Evaluation with 32 cores
- Early SVE register length (MUSA)
  - SVE lengths of 128, 256, 512 and 1024 bits

the instruction reduction achieved as the SVE vector length increases

SVE units of 256 or 512 bits are good design points

- Refined SVE register length (gem5)
  - scaling vector length from 256 to 512 bits
  - STREAM : reduction in memory operation usage from 3 LDs and 1 ST, to 2 LDs and 1 ST
  - Speedups DGEMM (1.34×), STREAM (1.63×), and WaLBerla (1.24×)

reducing instruction when scaling vector length leads to reducing throughput on register file allocation, and thus fewer stalls in the pipeline execution



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

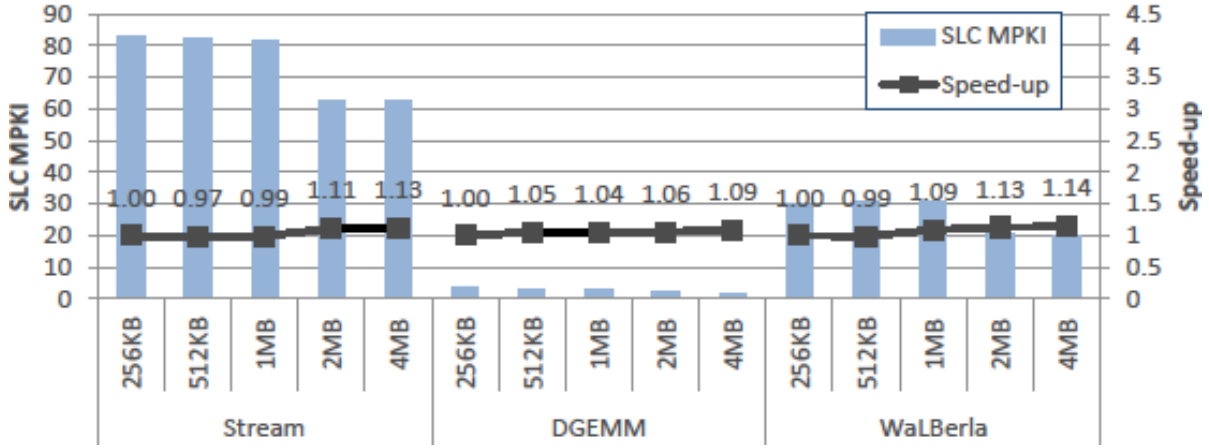
NUMERICAL RESULTS AND ANALYSIS : ON CHIP MEMORY

How large should the System Level Cache (SLC) be to maximise performance?

- SLC size : 256 KiB, 512 KiB, 1 MiB, 2 MiB and 4 MiB per core (MUSA)
  - STREAM : 2 MiB SLC slice configuration attains a 1.12× speed-up over the 1 MiB slice
  - DGEMM and WaLBerla the 512 KiB and 1 MiB slice sizes already capture most of the benefits

→ Choose a moderate size that captures most of the performance

sizes between 512 KiB and 1 MiB provide a good speed-up



MUSA

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

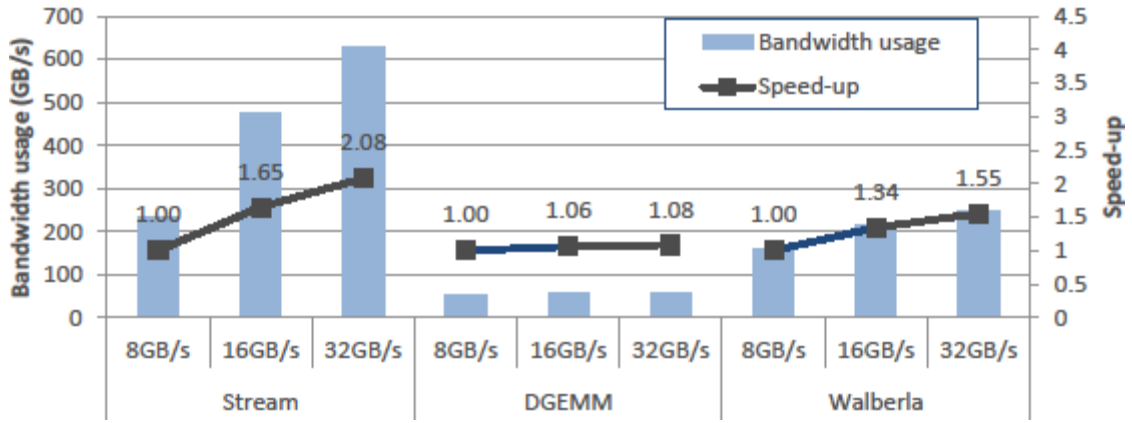
MEMORY BANDWITH

What is the best amount to satisfy HPC excution needs?

- Memory bandwidth per core : 8 , 16 , 32 GB/s (MUSA)
  - STREAM : 16 GB/s per core, bandwidth usage of 93%, 32 GB/s per core it drops to 61%
  - WaLBerla, : 16 GB/s per 1.34× improvements
  - DGEMM : 8 GB/s per core is sufficient to feed the functional units

→ increasing the available bandwidth leads to significant performance improvements

Bandwidth of around 20 GB/s bandwidth per core capture most of the performance benefits seen on the 32 GB/s configurations



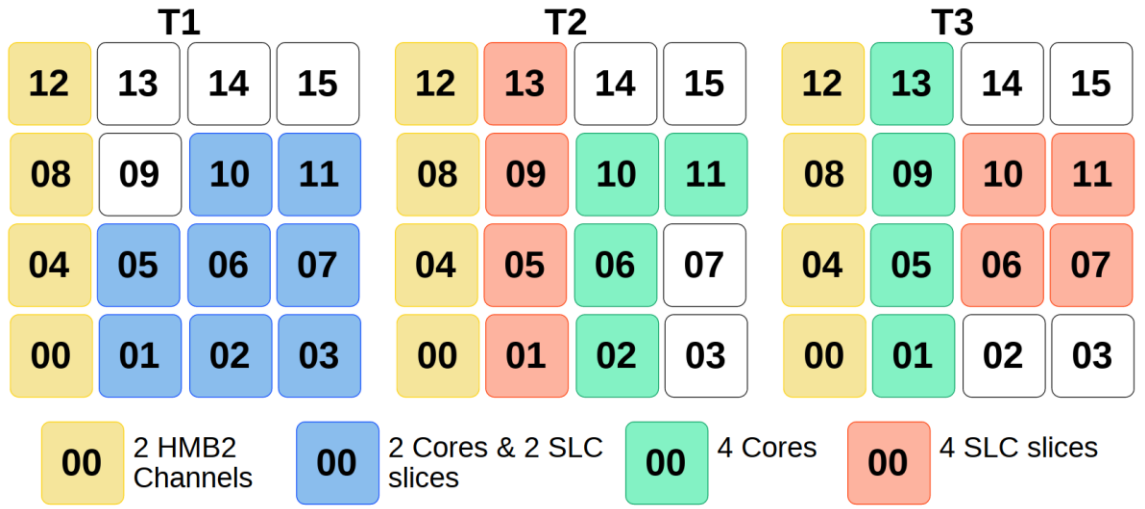
MUSA

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

MEMORY BANDWITH

- Refined bandwidth evaluation with NoC routers (gem5)
  - 16 threads for each of the topology layouts T1, T2 and T3
  - Single- and multi-channel VNET support
  - 256 and 512 SVE vector length, 3 MiB SLC size

→ increasing the available bandwidth leads to significant performance improvements



- All topologies benefit from increased bandwidth
- the multiple link usage significantly reduces the network latencies of the Request, Response and Data VNETs the queueing latencies for the Request VNET

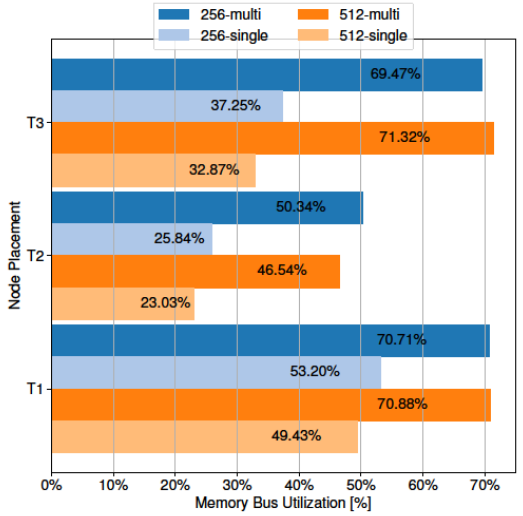
- Increasing NoC bandwidth increased benchmark performance for memory sensitive kernels
- The magnitude of the speedup depends strongly on NoC placement and SVE register length

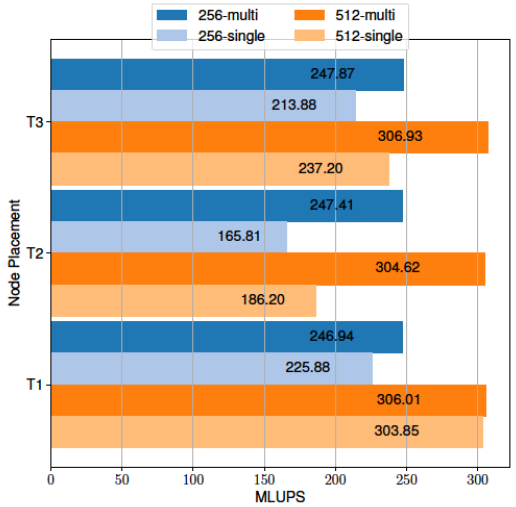
Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

NOC DIMENSIONING AND IP PLACEMENT

- Node placement (gem5)
  - STREAMTRIAD (SVE-256, 16 threads, 3MiB SLC, multi-link), measured by percentage of Memory Bus usage
  - T1 has the lowest NoC Queueing and Network latencies, and is the best performing Topology
  - T3 presents slightly higher NoC Queueing and Network latencies than T1, and achieves comparable bandwidth performance

→ The average hops per cycle for the three topologies are 19.6, 15.2 and 22.7





- T3 requires higher NoC hops for STREAM TRIAD benchmark, it does not get congested and maintains a high hops/cycle rate
- T3 performing similarly to T1

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

DDR MEMORY CONTROLLERS PLACEMENT

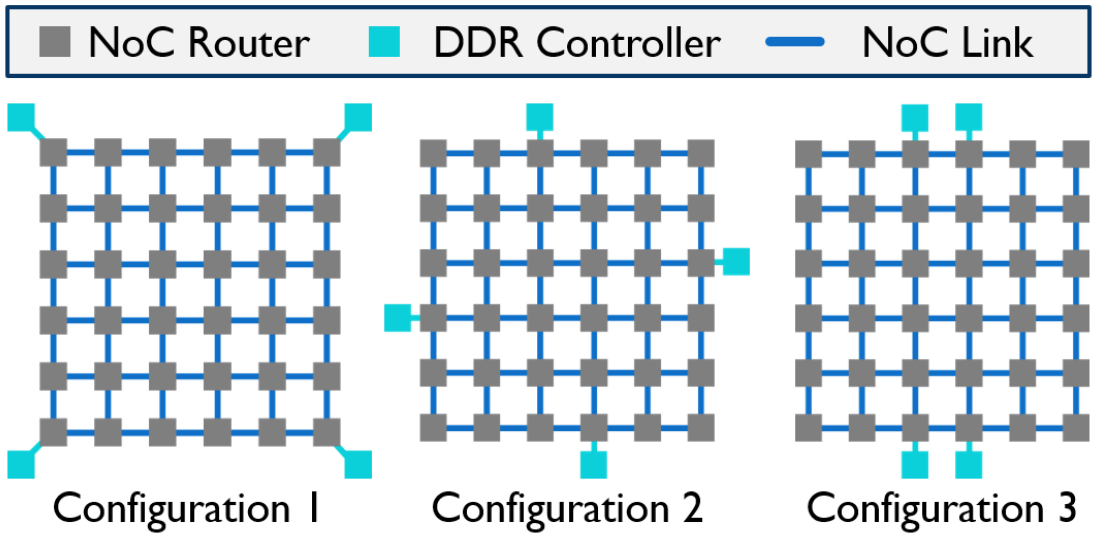
What is the best placement for DDR?

- Evaluation using packets average latency : SESAM/VPSim

➡ Configuration 3 better than Configuration 2, which is better than Configuration 1

Placement far from corners in order to reduce the total distance of packets

- No significant impact on the average packet queuing delay

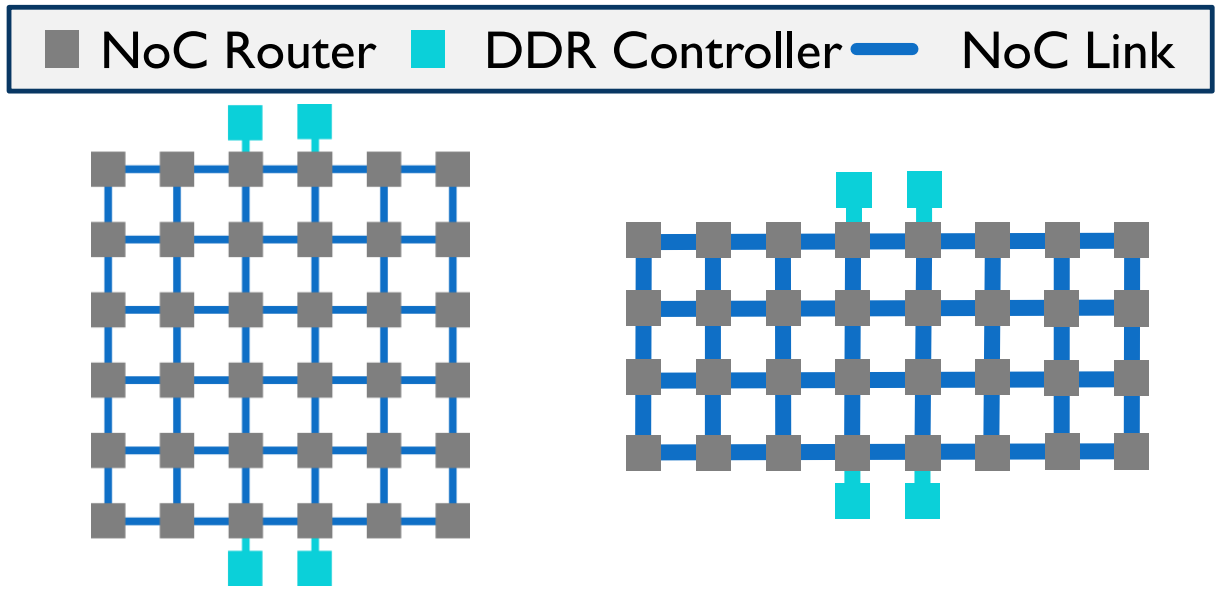


Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

NOC DIMENSIONING

What is the best NoC form factor?

- SESAM/VPSim : square (6 × 6 or 8 × 8), rectangle (4 × 8 or 6 × 12), 32, 64 cores
- Average Packet Latency (cycles)
  - Average packet distance : square shape induces an increase in the number of routers traversed by packets
  - Congestion is managed better by the 4 × 8 NoC than by the 6 × 6 for 32-core architecture (likewise 64-core architecture)



Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

CONCLUSION

- Multilevel co-design methodology
  - Impact of alternative chip design parameters onto application performance and system efficiency
- Trade-offs between simulation speed, accuracy and model abstraction level
  - Cycle-accurate microarchitecture simulator : gem5
  - Transaction-level simulator/emulator: SESAM/VPSim
  - Trace-based simulator full co-design methodology: MUSA
- Performance analysis is carried out with a number of representative benchmarks
- Several system design concerns, real HW RDV1
  - Dimensioning of SVE register length, on-chip memory requirements and external memory bandwidth (MUSA)
  - NoC topology and memory controller positioning for large scale HPC design (SESAM/VPSim)
  - Full insight on the impact of SVE register length, NoC bandwidth, and components placement strategies (gem5)

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop

THANKS

27 PARTNERS FROM 10 EU COUNTRIES

- www.european-processor-initiative.eu
- @EuProcessor
- European Processor Initiative
- European Processor Initiative

Copyright © European Processor Initiative 2021. L. Zaourar et al. Multilevel simulation-based co-design of next generation HPC microprocessors. PMBS 21 workshop