GEMSCLAIM

Greener Mobile Systems By Cross Layer Integrated Energy Management

Dimitrios S. Nikolopoulos, Queen's University of Belfast

Consortium:

- University of Innsbruck
- Queen's University of Belfast
- RWTH Aachen University
- Politenica University of Timisoara

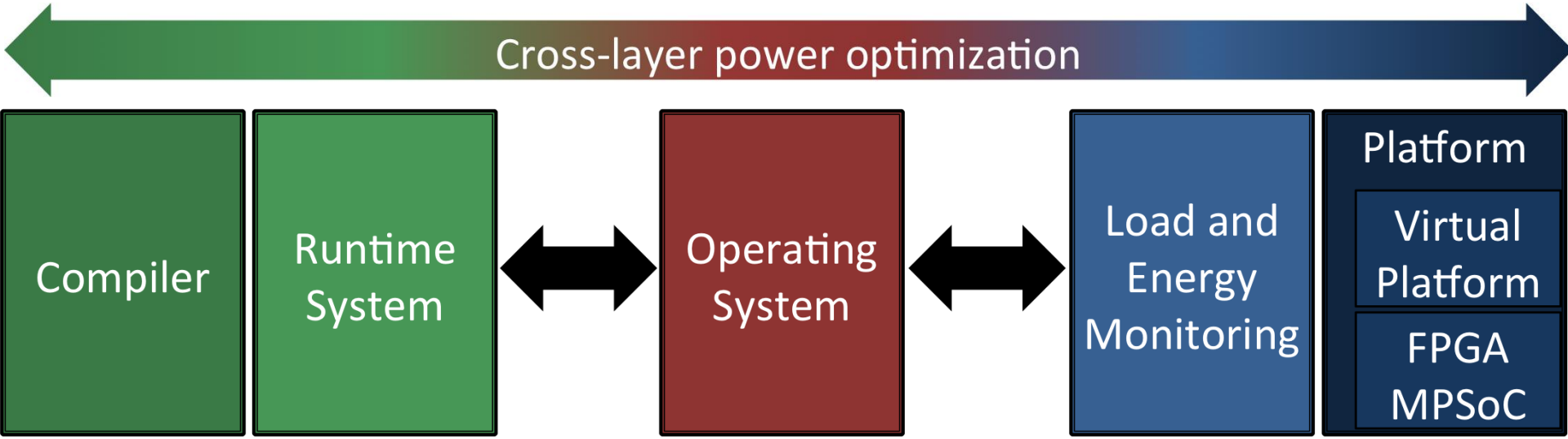
Start: September 1st, 2012

Duration: 3 years

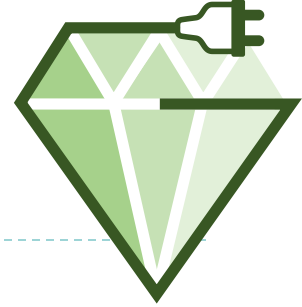
Scientific background

- Personal computing: desktop computers → mobile systems
- Mobile systems are becoming more energy demanding
  - Radio interfaces
  - High computational load (games, multimedia apps.)
- ``Major challenge for the mobile society in the next decade is that battery capacity is not keeping pace with Moore‘s Law.“ by M. Muller, CTO of ARM
- Trade-off between energy and performance for mobile devices and many other computing systems:
  - desktop computing, cloud computing, high performance computing

Vision in a nutshell



Key challenges and potential impact



Project duration: Sept. 2012 – Aug. 2015

Cross layer energy management and optimization for mobile devices: HW/simulator, OS, compiler

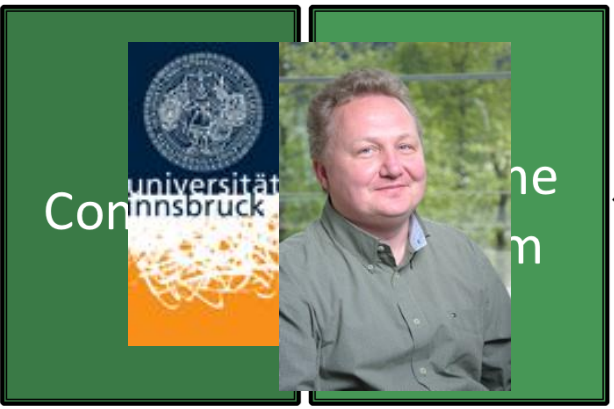
- Energy-aware optimizing and parallelizing compiler
- OS with energy-proportional resource management
- Customizable many-core HW with built-in energy monitoring and modeling facilities

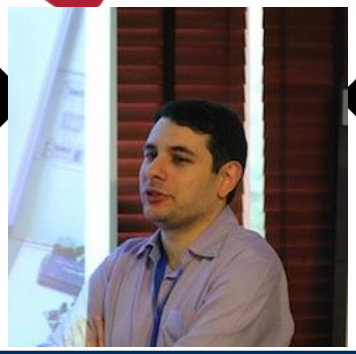
Potential impact

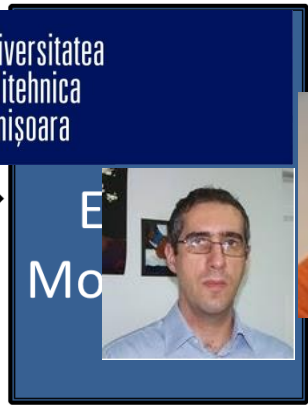
- control trade-off between energy optimization and performance
- 30% energy saving for mobile terminals over state-of-the-art

Consortium and synergies

Cross-layer power optimization







MPSoC

Project Coordination

Component-level energy monitors & sensors

Virtual prototyping & hardware design
Instruction-level power modelling

Compiler/RTL/OS interface, energy-aware optimisation
energy accounting and control abstractions

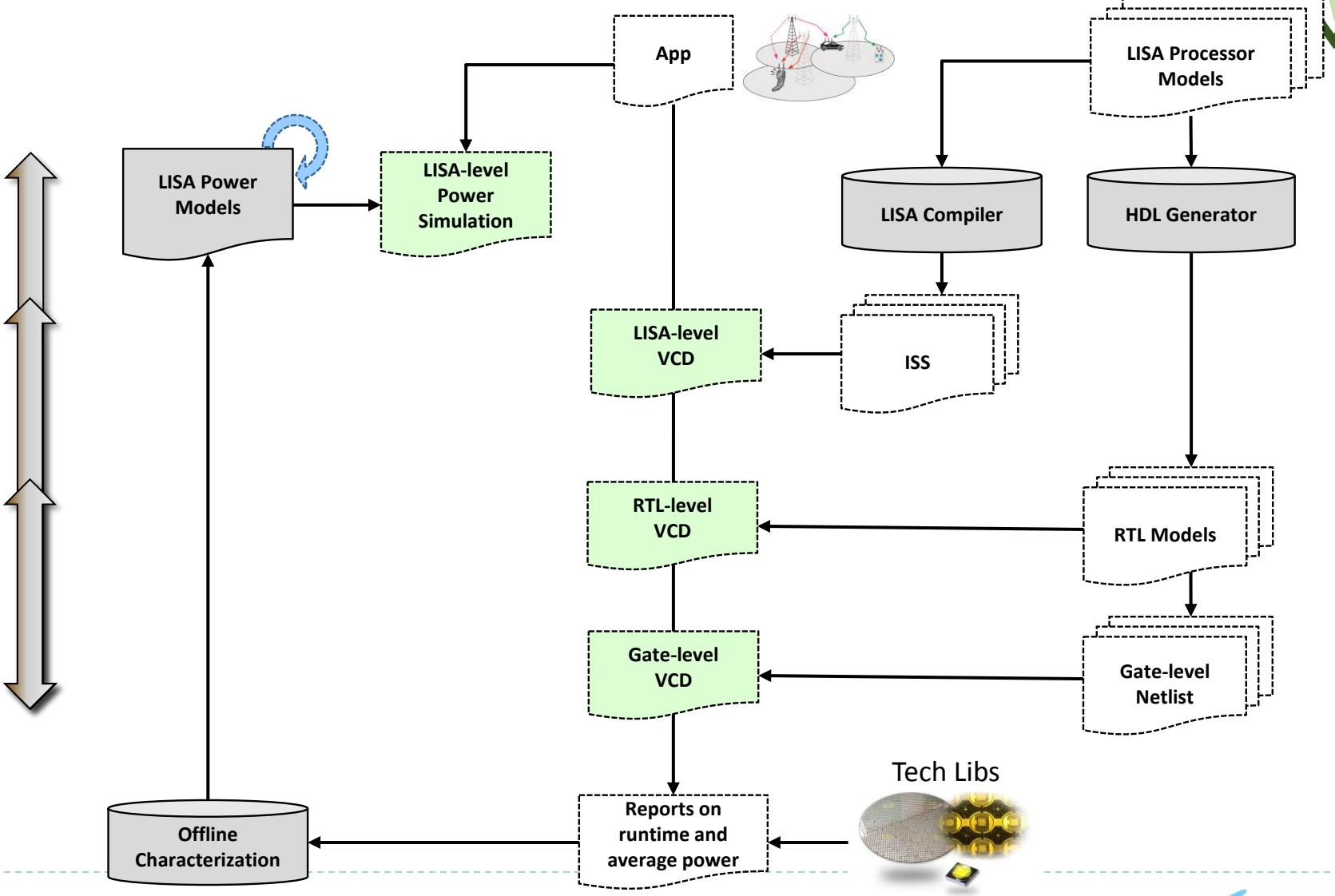
Application benchmarking with common smartphone workloads across all levels

chist-era

Work plan and milestones

Milestone	Delivery month	WP involved	Title
M1	6	WP1, WP2	Requirements specification and design of the GEMSCLAIM HW/SW environment
M2	15	WP1, WP3-WP7	Early prototype of the GEMSCLAIM HW/SW environment (able to simulate/execute benchmarks applications)
M3	30	WP1, WP3-WP7	Final prototype of the GEMSCLAIM HW/SW environment (able to perform energy savings benchmarking)
M4	36	WP1, WP7	Benchmarks fully optimized for energy (targeted savings: 30%) on FPGA with GEMSCLAIM SW; project completion

Accurate power modeling



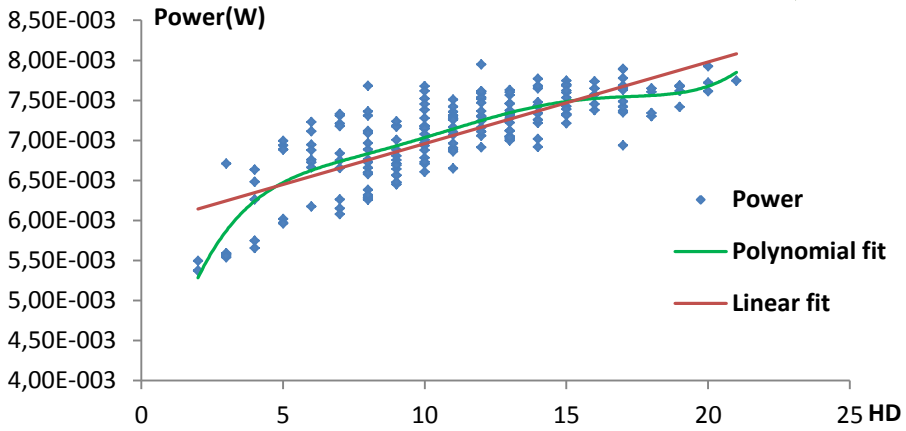
Instruction-level power models

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	0	opcode					1	src1				dst				imm16															

Instruction group of alu_rri

	LISA Resource (Pipeline Register)
opcode	opcode
imm16 (16-bit immediate operand)	op2
src1 (source register address)	rs1
op1 (source register value)	op1
dst (destination register address)	BPR

Resource Mapping



Power fitting for alu_rri

- **Linear Regression**

$P_{est} = 0.0001 \cdot HD + 0.0059$ $R^2 = 0.61$

- **Polynomial Regression**

$P_{est} = -2x10^{-10} \cdot HD^6 + 3x10^{-8} \cdot HD^5 - 10^{-6} \cdot HD^4 + 3x10^{-5} \cdot HD^3 - 0.0003 \cdot HD^2 + 0.0016 \cdot HD + 0.003$

$R^2 = 0.69$

- **Multivariate Linear Regression**

$P_{est} = 5.90x10^{-5} \cdot HD_{op1} + 8.4x10^{-5} \cdot HD_{imm16} + 1.97x10^{-4} \cdot HD_{src1} + 1.69x10^{-4} \cdot HD_{dst} + 5.51x10^{-3}$

$R^2 = 0.86$

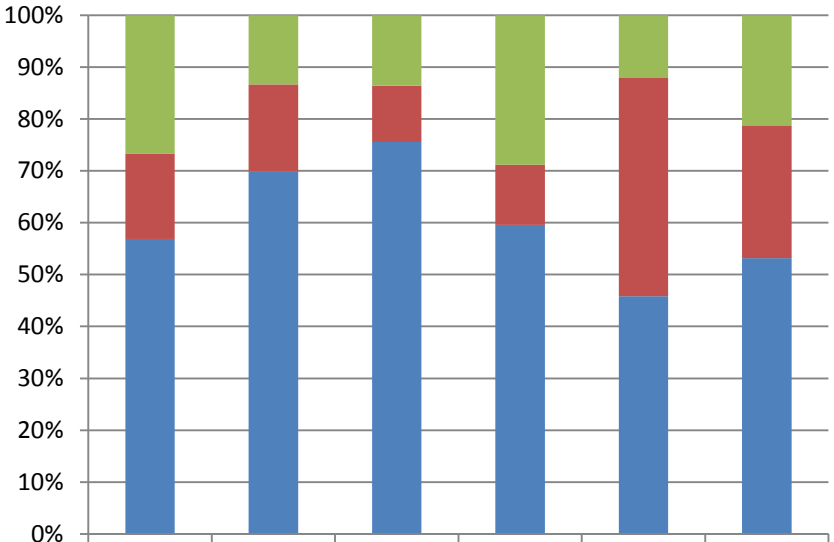
HD: hamming distance of instruction word
HD_{imm16}:hamming distance of immediate
HD_{src1}: hamming distance of source register address
HD_{op1}: hamming distance of source register value
HD_{dst}: hamming distance of destination register address

→R^2 increase: Improved estimation accuracy

chist-era

Alternative power modeling approaches

App	Average Power (mw)		Error	Speed (s)		Speed Improvement
	LISA power Simulation	Gate level Simulation		LISA power Simulation	Gate level Simulation	
LMS	8.70	8.80	**1.40%**	**0.41**	**40.57**	**X 98.95**
sobel	9.20	9.70	**5.04%**	**0.85**	**66.35**	**X 77.75**
cordic	8.20	8.40	**2.75%**	**0.26**	**32.59**	**X 123.81**
conv3x3	8.70	8.84	**1.58%**	**4.53**	**219.3**	**X 48.37**
sieve	8.62	9.61	**11.38%**	**3.8**	**215.8**	**X 56.79**
FFT	8.80	9.30	**5.38%**	**7.33**	**726.3**	**X 99.04**
Average			**5.02%**			**X 84.11**



	LMS	sobel	cordic	conv3x3	sieve	FFT
Branch	26,64%	13,36%	13,57%	28,83%	12,10%	21,25%
Load/Store	16,51%	16,74%	10,85%	11,60%	42,07%	25,59%
ALU	56,85%	69,90%	75,58%	59,57%	45,83%	53,16%

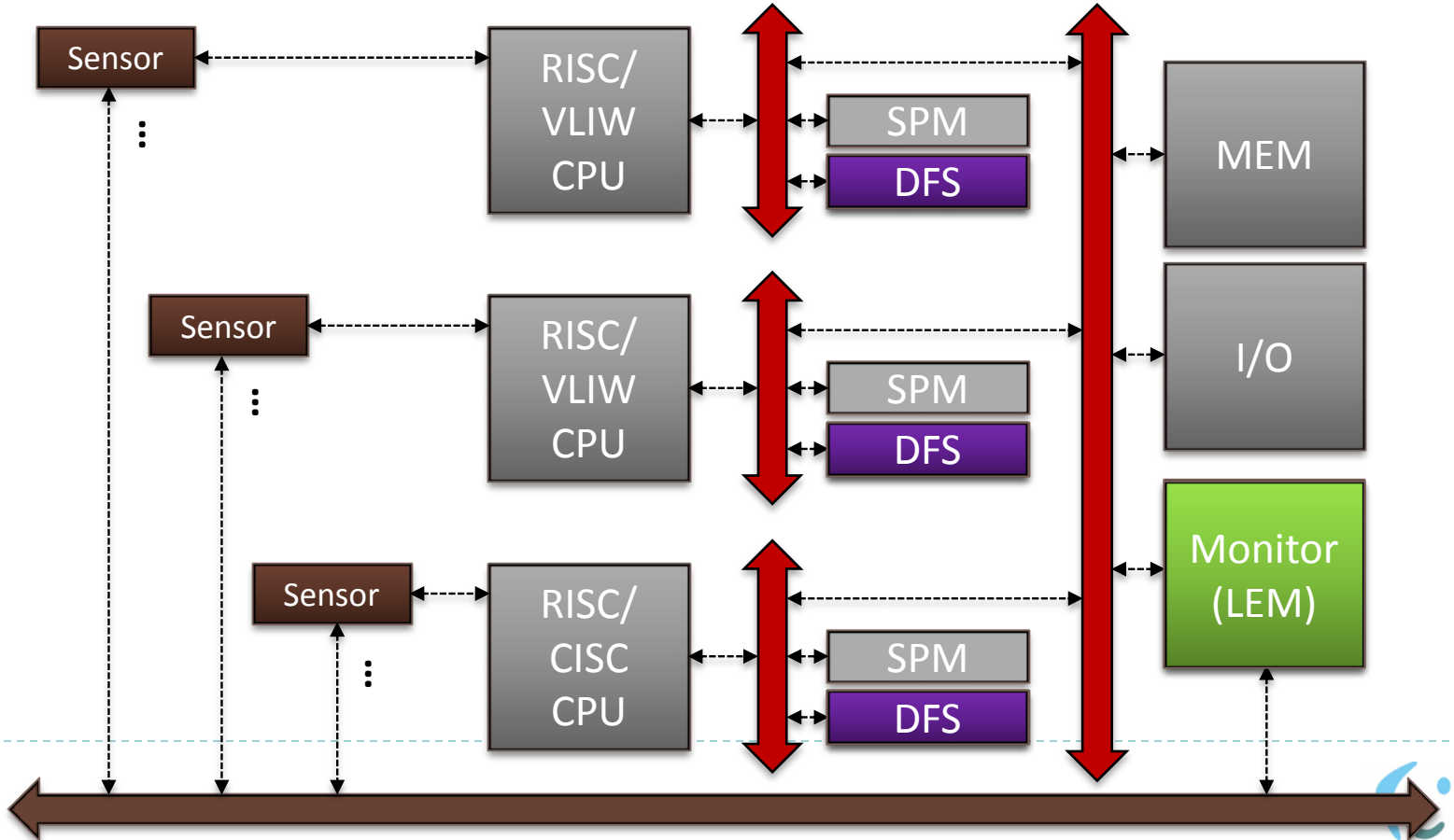
Proportion of Instructions

- **Accuracy**
  - **5% when portion of Load/Store instruction is low**
  - **An error of 11.38% for Sieve application related to the great proportion of memory access**
  - **Refined characterization for memory access instructions (on going work)**

chist-era

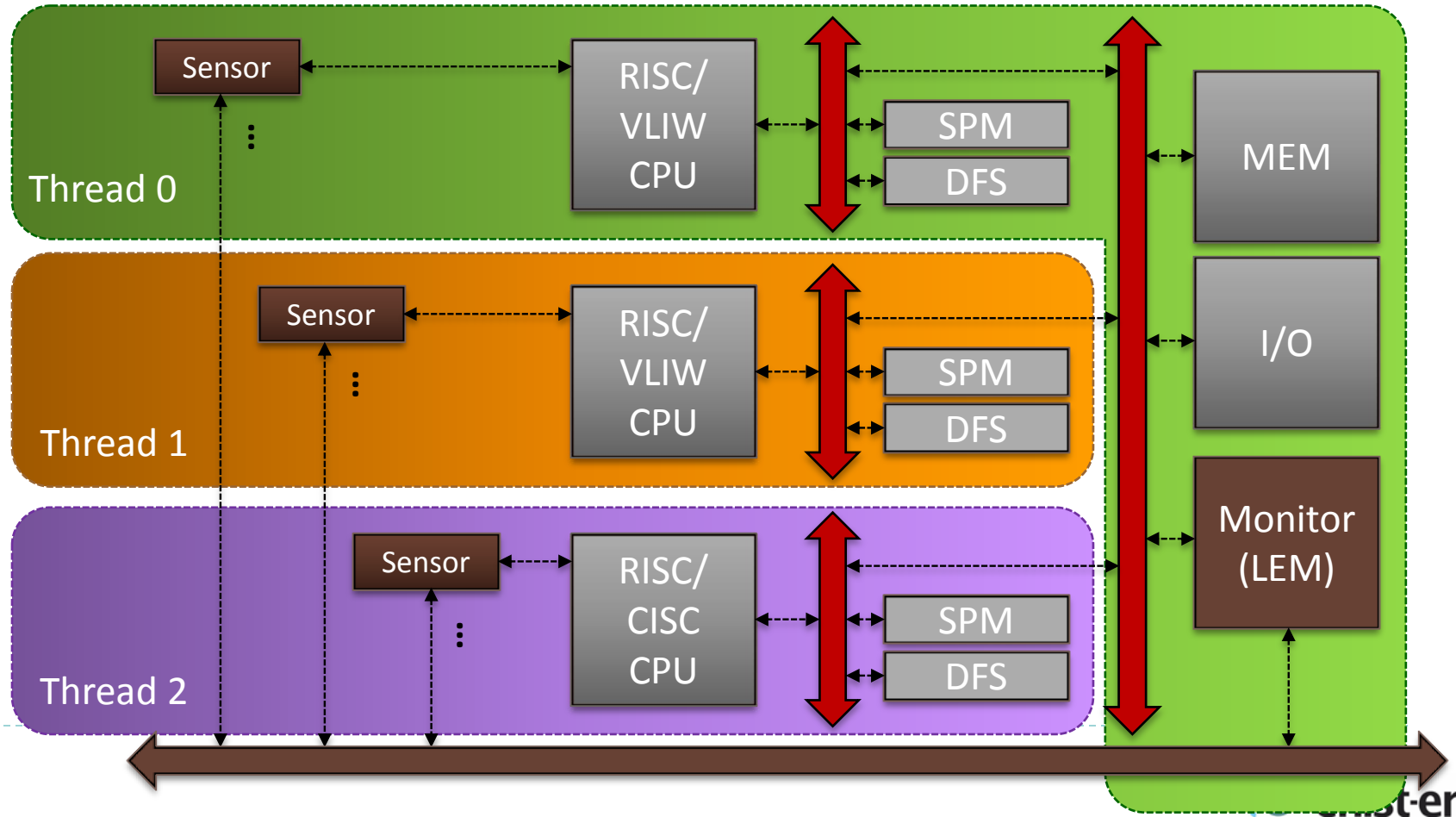
Reference many-core architecture

- Heterogeneous, customizable, power-aware architecture:
  1. Runtime power scaling for each core individually
  2. Fine-grain LEM: accounting modes + tunable sampling rate



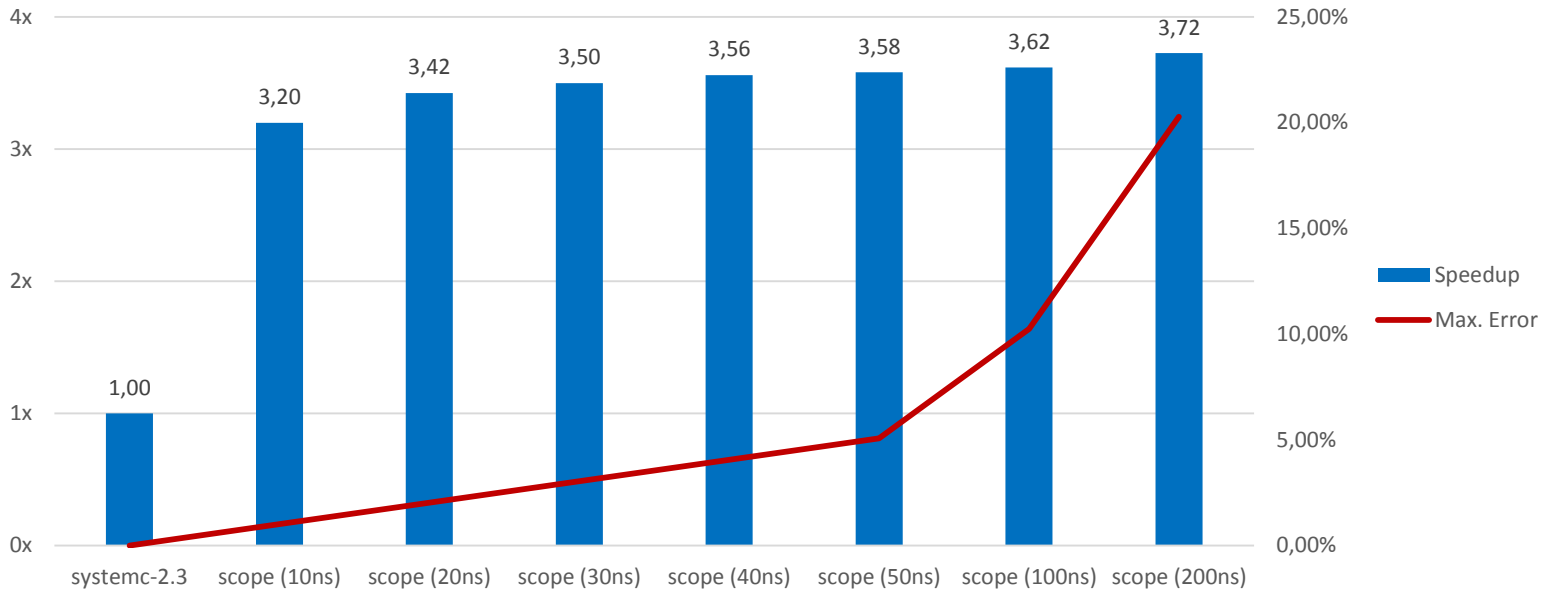
Parallel simulation framework

- Multithreaded GEMSCLAIM simulator
  - RISC/VLIW processors are simulated on different threads
  - LEM, shared memory and peripherals run on first thread



Simulation speed vs. accuracy

- Test setup (parallel simulator uses 4 threads):
  - 8 RISC + 8 VLIW system
  - „test_lem“ application (lem driver test)

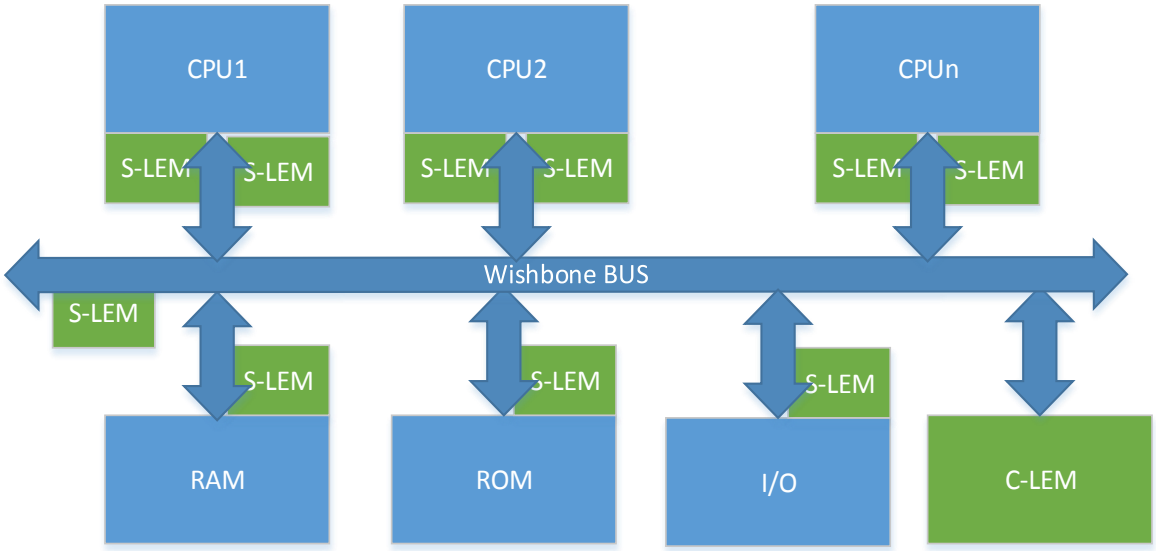


- ❑ Speedup between 3.2 and 3.7 using 4 threads
- ❑ Higher lookahead achieves higher speedup, but timing error increases linearly

chist-era

Hardware prototyping and energy monitor

- FPGA hardware prototype
  - Calibrates the power and energy measurements done in VP
  - Validates and quantifies total energy savings from both HW and SW optimizations, using physical measurements
- Load and Energy Monitor
  - Achieve accurate measurement of component-level energy consumption
  - Enable accurate accounting of energy at a fine granularity

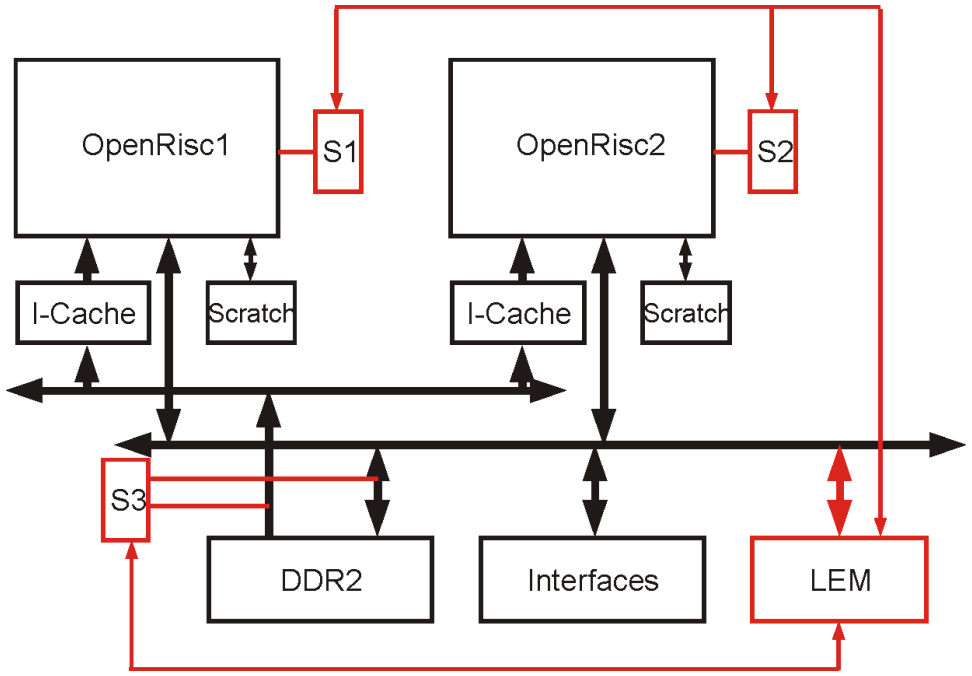


Distributed sensors
Monitoring of cores, memories, interconnect
Monitoring of private and shared components
Automatic buffering, averaging, accumulation

chist-era

Initial FPGA prototype - OpenRISC

- Dual Core OpenRISC
- Instruction Cache
- DDR2
- LEM infrastructure
- LEM sensors
- Scratchpad

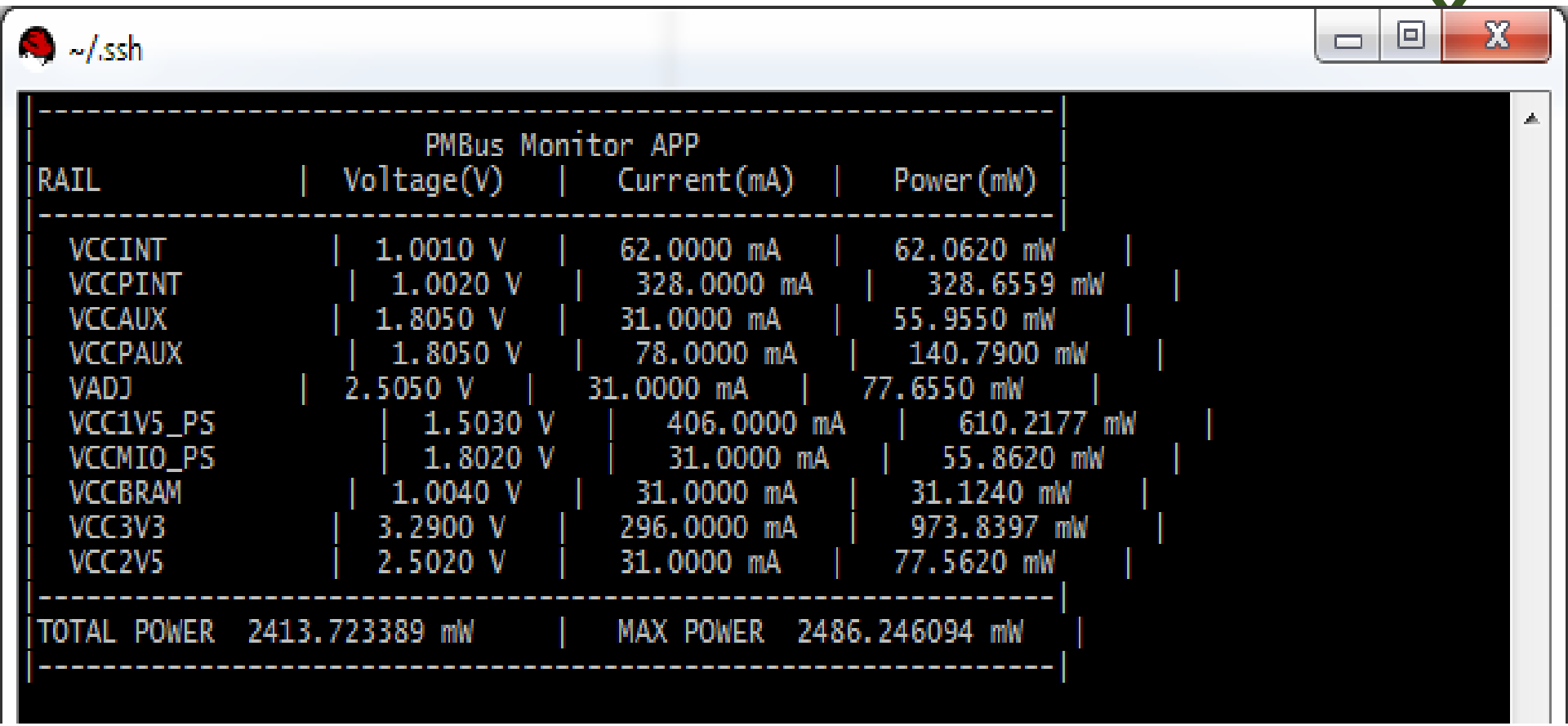


Physical power instrumentation

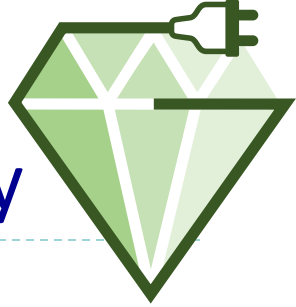
- FPGA test bench setup
  - DC power meter
  - Built-in power meter (Atlys board)
    - INA219 devices for the current monitoring on 2.5V line (FPGA), 1.2V line (FPGA), 1.8V (DDRAM) and 3.3V lines (I/O).
    - 2mA accuracy for the current measurement.

Atlys Power Supplies			
Supply	**Circuits**	**Device**	**Amps (max/typ)**
3.3V	FPGA I/O, video, USB ports, clocks, ROM, audio	IC16: LT3501	3A / 900mA
2.5V	FPGA aux, VHDC, Ethernet PHY I/O, GPIO	IC15: LTC3546	1A / 400mA
1.2V	FPGA core, Ethernet PHY core	IC15: LTC3546	3A / 0.8 – 1.8A
1.8V	DDR & FPGA DDR I/O	IC16: LT3501	3A / 0.5 -- 1.2A
0.9V	DDR termination voltage (V_{TT})	IC14: LTC3413	3A / 900mA

Zynq MPSoC physical power instrumentation



RAPMI: Abstracting component-level energy



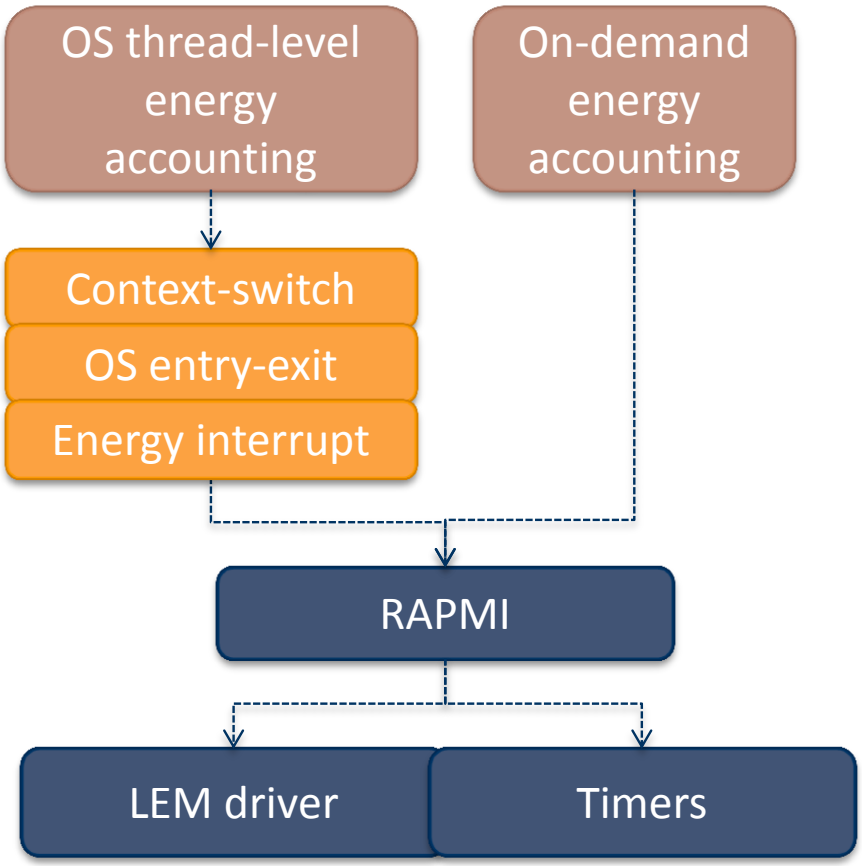
- **`rapmi_ctrl(mode, rate, power_level)`**
  - sensor mode (INSTANT, ACCUM, AVG)
  - rate (Not yet VP/HW impl.)
  - power_level (Not yet VP/HW impl.)
- **`rapmi_start()`**
  - Start sensor measuring
- **`rapmi_stop()`**
  - Stop sensor measuring
- **`rapmi_read(rapmi_t *pm)`** ⟶
  - Read sensor values
- **`rapmi_set_freq(freq)`**
  - Set core frequency (VP)
- **`rapmi_get_freq()`**
  - Returns core frequency (VP)

```
typedef struct {
  uint32_t core;
/*uint32_t mem, bus...*/
  uint32_t timestamp;
} rapmi_t
```

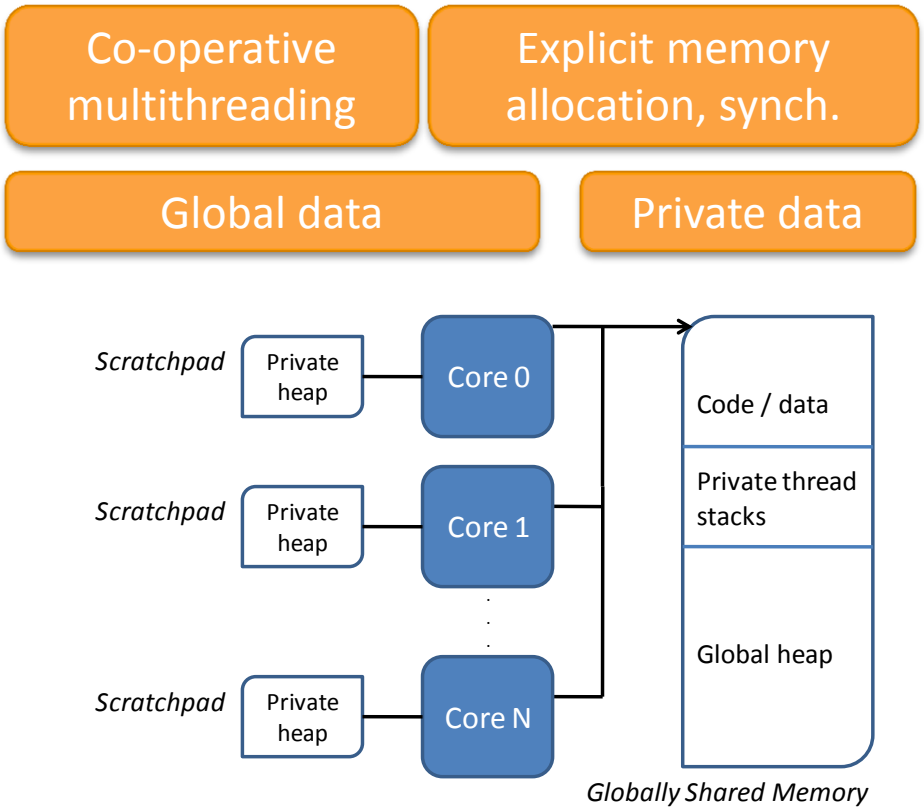
chist-era

GEMSCLAIM OS components

Monitoring infrastructure



FPGA/VP lightweight runtime/OS



Rethinking the OS

Energy slices (first-class resource)

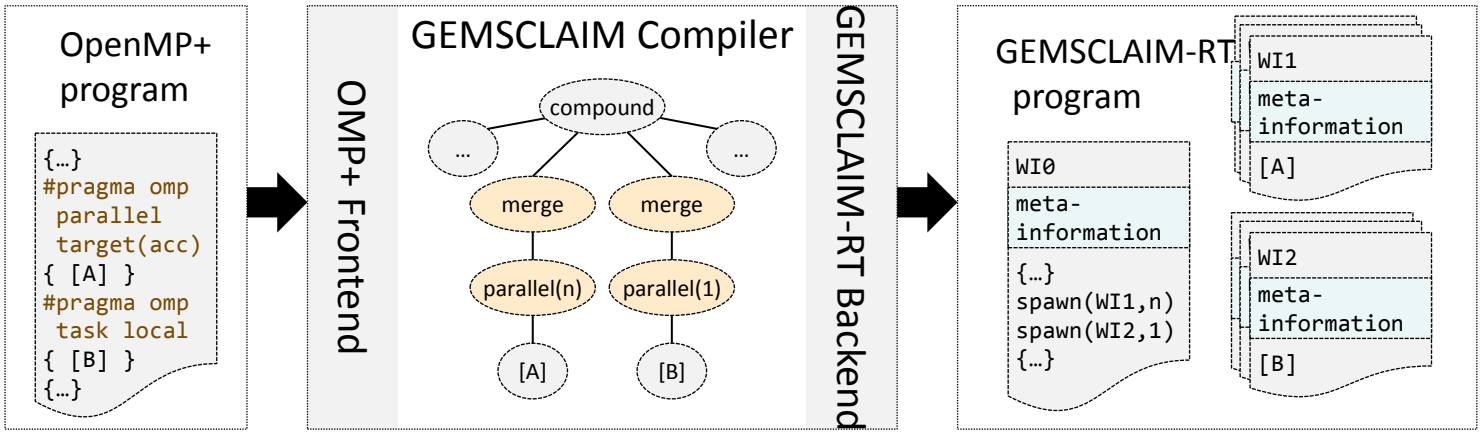
Processor time slices (within energy budget)

Application A

Application B

chist-era

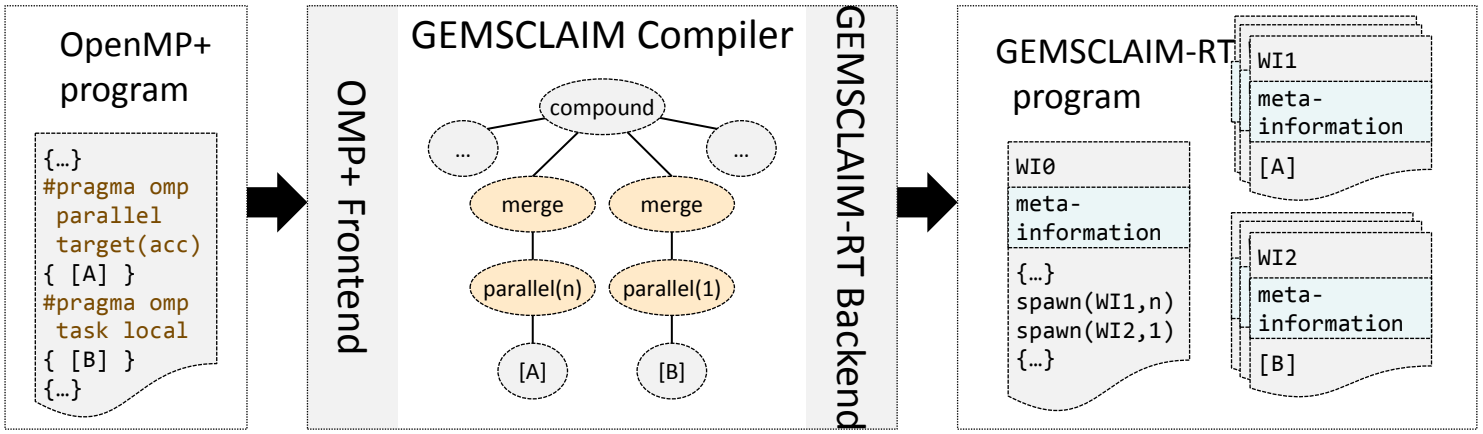
Compiler and runtime system



- Source-to-source C compiler which supports C programs annotated with OpenMP+ energy-aware extension to OpenMP
- Programs analyzed and sub-divided into delineated code regions, individually tunable and annotated with additional metadata
- Execution of the C program orchestrated by the runtime system which handles dynamic decision making and scheduling, and communicates with the GEMSCLAIM OS.

chist-era

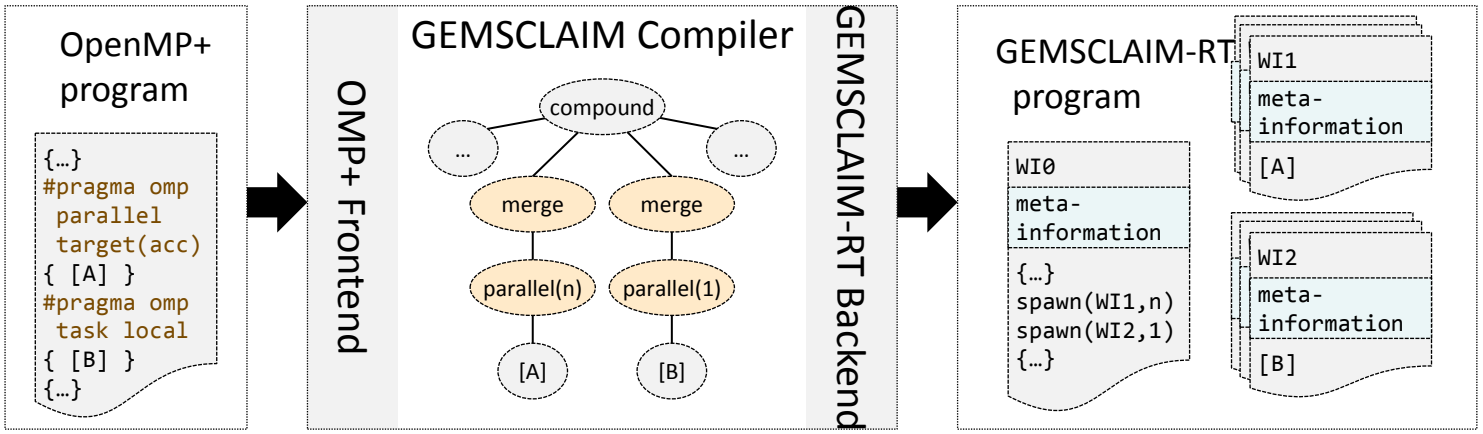
Compiler and runtime system



- Fully ported to the GEMSCLAIM platform
  - Complies with backend compiler restrictions (minimal preprocessor functionalities, ...)
  - Supports VP RISC architecture (context switching assembly code). VP VLIW architecture still to be addressed
  - Supports GEMSCLAIM OS (scratchpad memory allocation, LEM API)
  - Integrates LEM measurements (generic region instrumentation framework)

chist-era

OpenMP extensions (OpenMP+)



- OpenMP+ extension fully supported by the Insieme compiler and partially by the runtime system
  - Region construct (handling code outside OpenMP regions)
  - Target clause (assignment of work to heterogeneous cores)
  - Local clause (mapping data to scratchpads)
  - Objective clause (define optimisation objectives)
  - Param clause (define tunable parameters for compiler optimisation)

chist-era

Elements of management

- Four plenary project meetings
  - Innsbruck, 28/9/12
  - Belfast, 12/3/13
  - Aachen, 12/9/13
  - Timisoara, 11/2/14
  - Several longer co-development meetings among partners
- Monthly teleconferences
- Web page (http://www.gemsclaim.eu)
- All deliverables on time
  - D1.1 (CA), D2.1-D2.3 (requirements), D3.1, D4.1 D5.1, D6.1, D7.1 (early prototypes of compiler, OS, architecture, LEM, integrated HW/SW)
  - Large volume of hardware and software development
- Resources spent during first year
  - 25% out of theoretical target 33%
- 10 scientific publications
  - top conferences (e.g. SC, PACT), 1 best paper award (COSMIC)

Conclusions and way forward

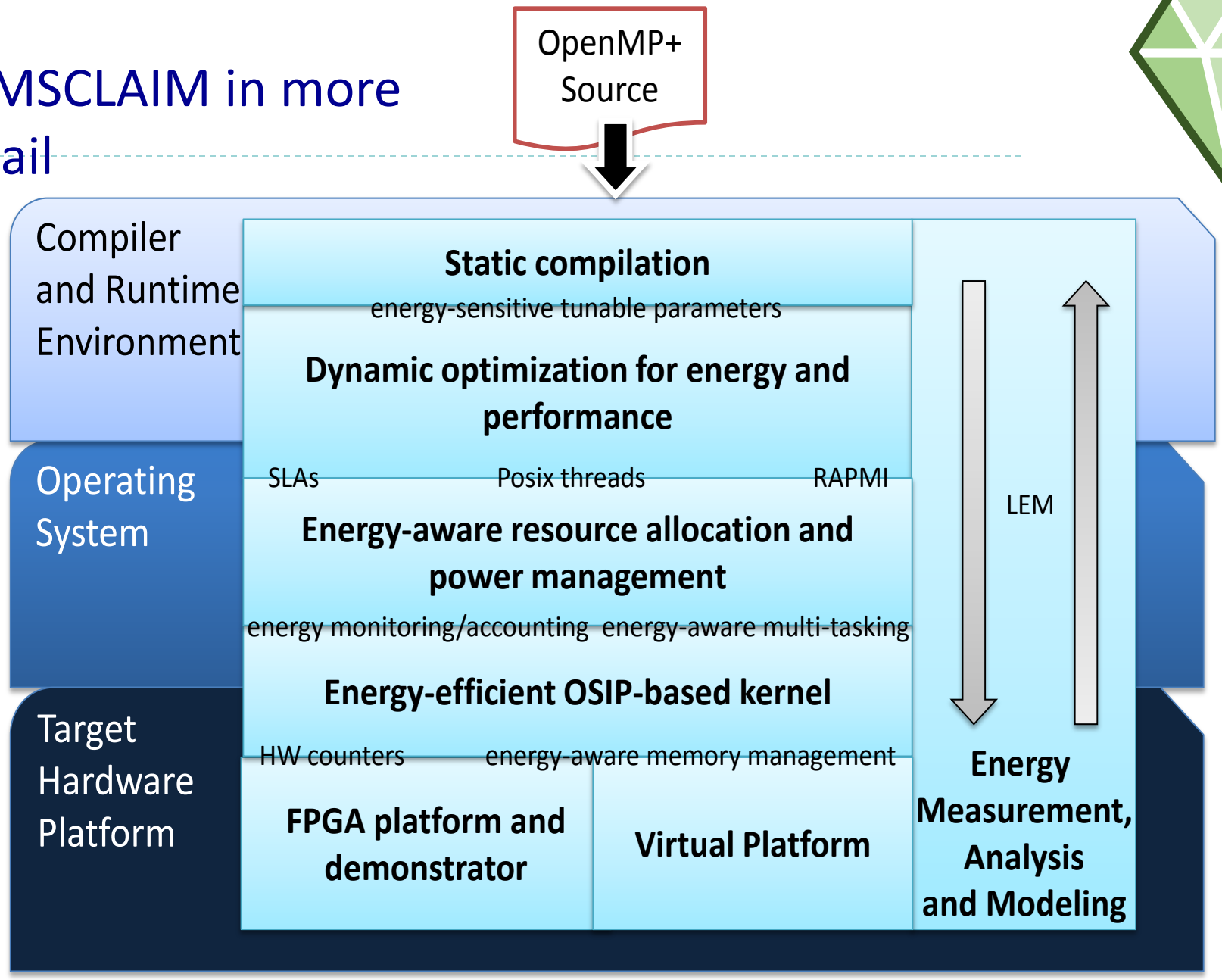
- Cross-layer energy optimization requires breaking barriers between layers
  - Common abstractions, metrics and models
  - Highly synergetic optimization approaches
- Fine-grain, component-level energy monitoring remains a challenge
  - Component power measurements are machine-specific, coarse-grain
  - Software power modelling is intrusive
  - Hybrid approaches are viable
- *GEMSCLAIM hopes to break the barriers via several innovative pathways*
  - Common power monitoring/measurement substrate
  - Low-power, power-scalable many-core architecture
  - Explicit energy accounting, allocation and optimization the software stack (compiler, runtime, OS)

Thank you!

Q&A

Backup

GEMSCLAIM in more detail



chist-era

UIBK: University of Innsbruck, Austria

- Group leader: Thomas Fahringer
- Involvement in other related projects:
  - Hipeac, En-act, AutoCore, OpenCore
- Background
  - parallelizing and optimizing compilers
  - programming languages
  - energy measurement and optimization for parallel programs
  - multi-objective optimization (energy, runtime, costs, etc.)
- Added-value
  - extend OpenMP to program heterogeneous mobile systems
  - optimize OpenMP+ for energy and runtime
  - extend runtime system for dynamic resource allocation, de-allocation and adjustment

QUB: Queen‘s University Belfast, N. Ireland

- Group leader: Dimitrios S. Nikolopoulos
- Involvement in other related projects:
  - ALEA, ENPOWER, NanoStreams, SCoRPiO, NovoSoft
- Background
  - Parallel programming languages and runtime systems
  - Operating systems and virtualization
  - Software-controlled energy-efficiency
- Added-value
  - Component-level, fine-grain, per-task energy accounting
  - Energy-aware SLAs
  - Holistic energy optimization
  - Disruptive technologies for computing with uncertainty

RWTH: Rheinisch-Westfälische Technische Hochschule Aachen, Germany

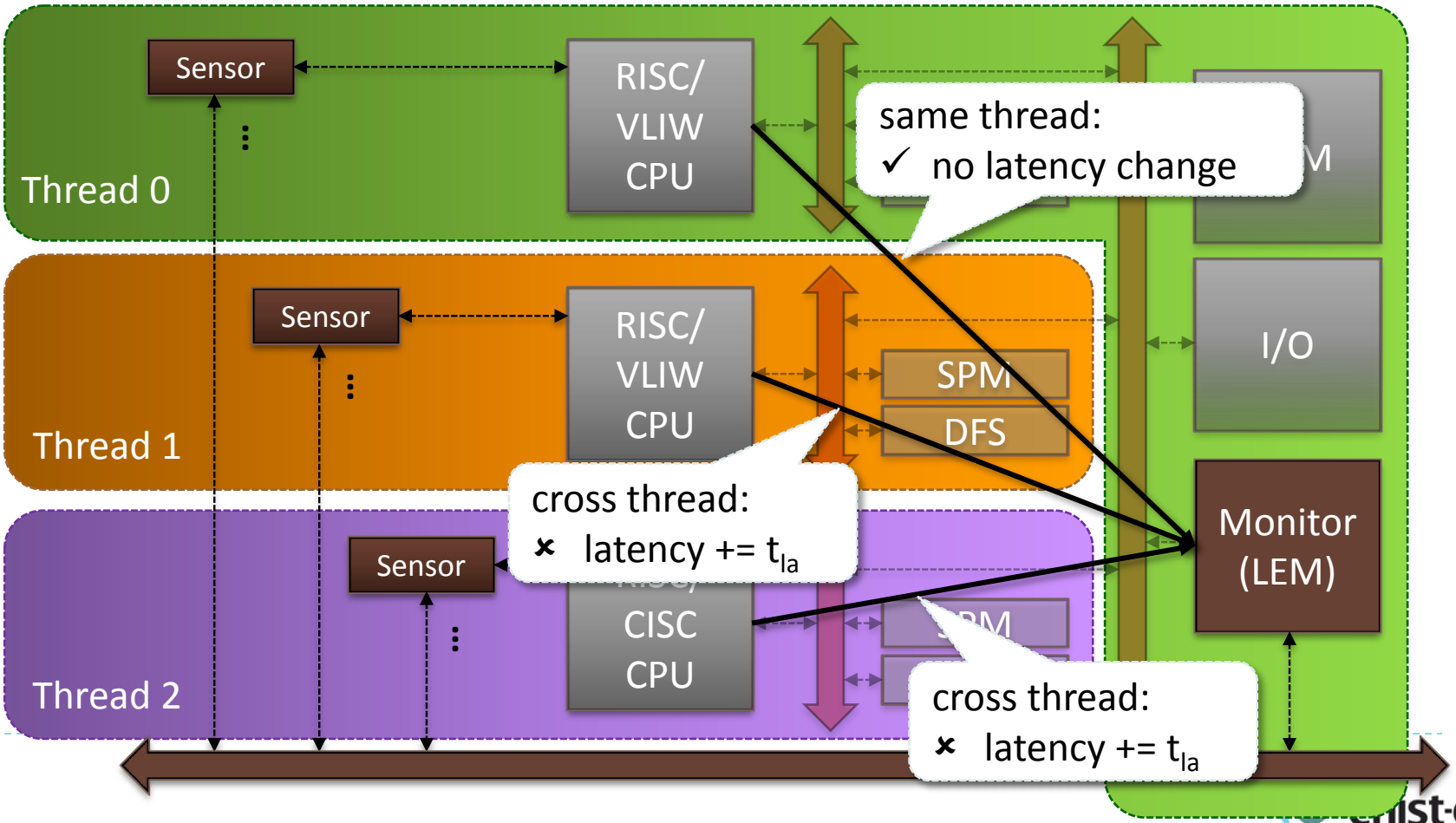
- Group leaders: Rainer Leupers, Anupam Chattopadhyay
- Involvement in other related projects
  - HiPEAC, UMIC, DSPACE, EURETILE, METIS, TETRACOM
- Background
  - SoC platform design
  - Automated Processor Synthesis and Customization
  - Processor and System Simulation
  - Compilation flow for heterogeneous MPSoC
- Added-value
  - System simulation/Virtual Platforms
  - High-level power estimation

UPT: Politenica University of Timisoara, Romania

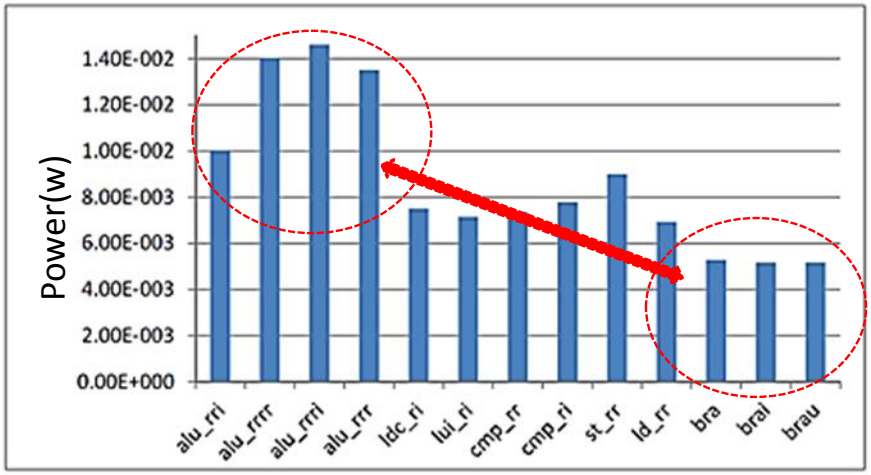
- Group leader: Marius Marcu
- Involvement in other related projects:
  - ICT-eMuCo
- Background
  - Energy monitoring and estimation for physical components
  - Thermal and power benchmarking
  - Energy-aware mobile processing and communication
- Added-value
  - Component-level, fine-grain, run-time energy measurement
  - Memory, caches, bus energy estimation
  - High-level control interfaces for low-level power management mechanisms
  - Hardware validation of GEMSCLAIM platform

Parallel simulation timing errors

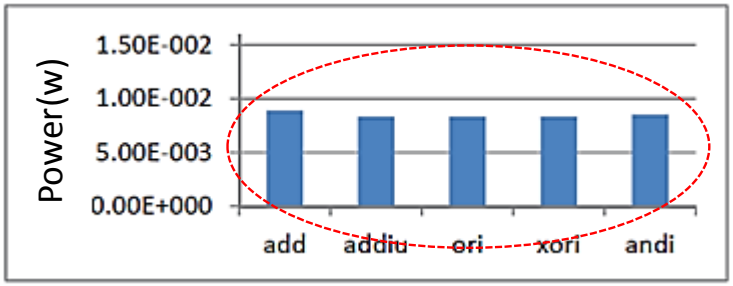
- Parallel simulation will induce timing errors
  - Local times of all threads may differ by the lookahead t_{la}
  - Cross thread communication has t_{la} as a minimum latency



GEMSCLAIM: instruction-level power characterisation

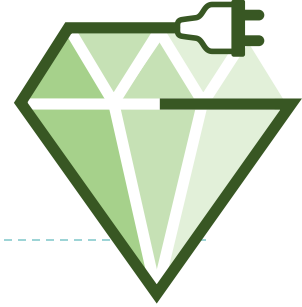


Average Power and Relative Deviation among Instruction Groups



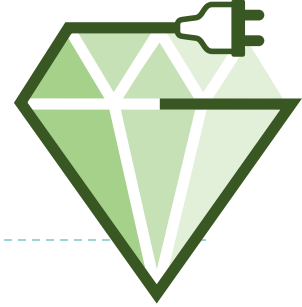
Average Power and Relative Deviation of Instruction Group alu_rri

- **The effect of modes within instruction group on power is relatively small**
  - Reduced modeling complexity
  - Significant implication on runtime power monitoring (CAD-on-Chip)
  - Applied for GEMSCLAIM processors



LEM design API

- Start monitor
  - Component
  - Rate
  - Mode
  - Reset
- Stop monitor
  - Returns the requested and collected values
- get/set parameter
  - Component
  - Power mode



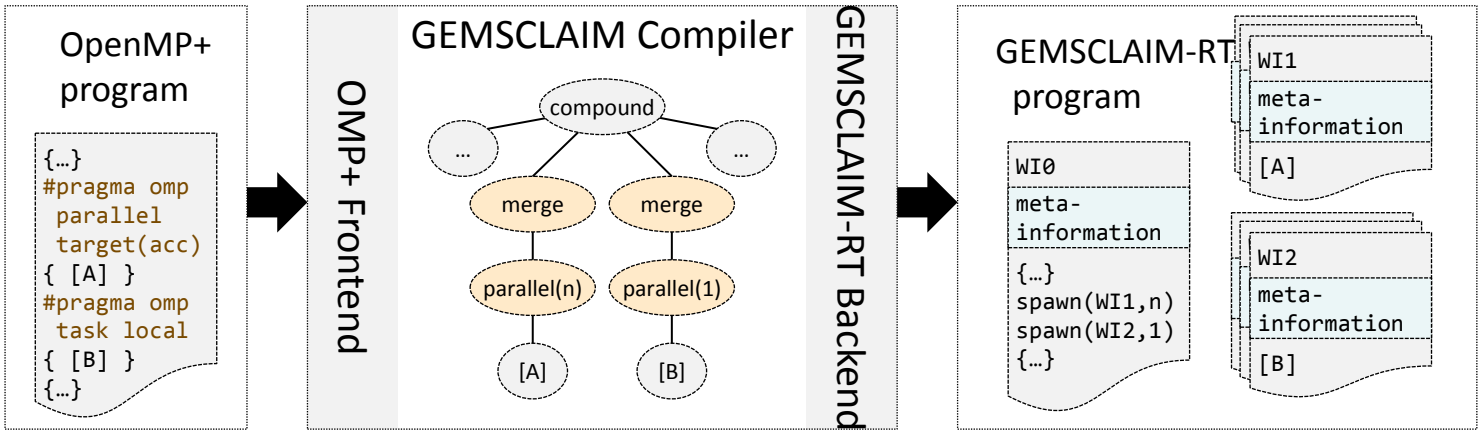
LEM design API

- Start monitor
  - Component
  - Rate
  - Mode
  - Reset
- Stop monitor
  - Returns the requested and collected values
- get/set parameter
  - Component
  - Power mode

Initial FPGA prototype - Issues

- linker scripts (stack, program for multi core, multiple stack addresses, multiple boot addresses) - SOLVED
- stack location (bram, global memory, size etc.) - SOLVED
- multiple cpu identification (implemented & verified CPU ID register) - SOLVED
- atomic read-modify-write instructions – OS provided

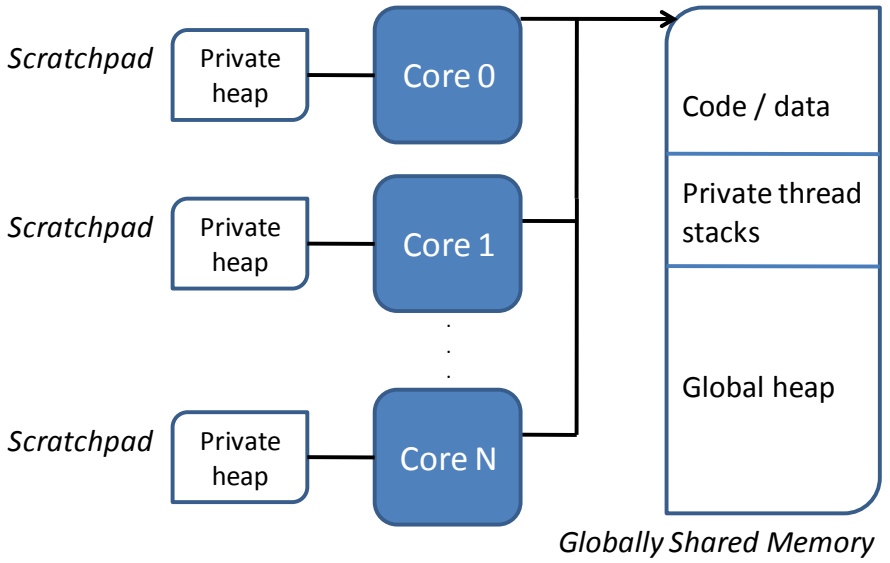
Compiler and runtime System



- Extended to:
  - support all C language features needed by the benchmark applications (recursive structures, variadic arguments …)
  - deal with backend compiler restrictions (interception of unavailable functions, GCC built-ins, ...)
  - Emulate file I/O (not supported on FPGA board)
  - Simplify integration with LLRT OS

OS memory address spaces

- Global shared memory
  - Code and data
  - Global heap
  - Synchronized allocation in the kernel
  - Per-thread stacks
- Private heap in scratchpad

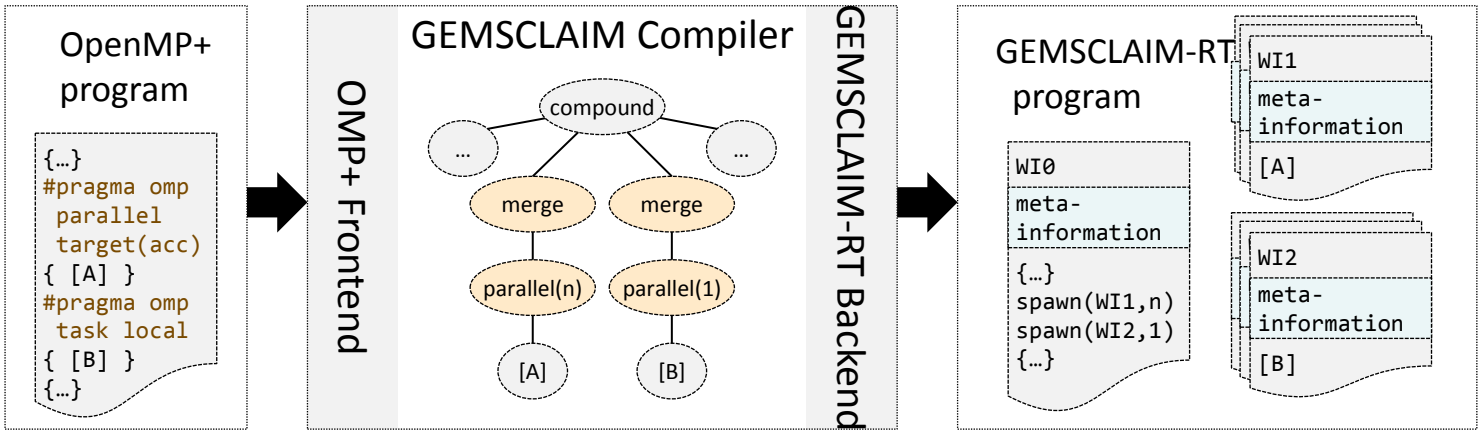


FPGA physical power instrumentation

- FPGA test bench setup
  - DC power meter
  - Built-in power meter (Atlys board)
    - INA219 devices for the current monitoring on 2.5V line (FPGA), 1.2V line (FPGA), 1.8V (DDRAM) and 3.3V lines (I/O).
    - 2mA accuracy for the current measurement.

Atlys Power Supplies			
Supply	**Circuits**	**Device**	**Amps (max/typ)**
3.3V	FPGA I/O, video, USB ports, clocks, ROM, audio	IC16: LT3501	3A / 900mA
2.5V	FPGA aux, VHDC, Ethernet PHY I/O, GPIO	IC15: LTC3546	1A / 400mA
1.2V	FPGA core, Ethernet PHY core	IC15: LTC3546	3A / 0.8 – 1.8A
1.8V	DDR & FPGA DDR I/O	IC16: LT3501	3A / 0.5 -- 1.2A
0.9V	DDR termination voltage (V_{TT})	IC14: LTC3413	3A / 900mA

Application benchmarking



- Bzip2
  - Bzip2 compression process implementation
- LTE Uplink simulator
  - Implementation of the baseband processing for an LTE mobile base station
- Jasper
  - Free reference implementation of the Jpeg2000 codec

chist-era

Publications with preliminary results http://www.gemsclaim.eu

1. A Multi-Objective Auto-Tuning Framework for Parallel Codes Herbert Jordan, Peter Thoman, Juan J. Durillo, Simone Pellegrini, Philipp Gschwandtner, Thomas Fahringer, and Hans Moritsch . SC '12
   November 11 - 15, 2012
2. Adaptive Granularity Control in Task Parallel Programs using Multiversioning. Peter Thoman, Herbert Jordan, Thomas Fahringer. Euro-Par 2013 - Aug 26 Aachen, Germany
3. INSPIRE: The Insieme Parallel Intermediate Representation. Herbert Jordan, Simone Pellegrini, Peter Thoman, Klaus Kofler, Thomas Fahringer. PACT 2013, September 7-11, 2013.
4. Deterministic Scale-Free Pipeline Parallelism with Hyperqueues, Vandierendonck, H., Chronaki, K. & Nikolopoulos, D. Nov 2013, Proceedings of Supercomputing'13: International Conference for High Performance Computing, Networking, Storage and Analysis (SC). New York, NY, USA: ACM
5. BDDT: Block-level Dynamic Dependence Analysis for Task-Based Parallelism (full version), Tzenakis, G., Papatriantafyllou, A., Vandierendonck, H., Pratikakis, P. & Nikolopoulos, D. Sep 2013 Proceedings of the 2013 International Conference on Advanced Parallel Processing Technology. Springer, (Lecture Notes in Computer Science)
6. Programming the Energy Efficiency of High Performance Computing Systems: Keynote Talk: Proceedings of the 4th International Conference on Energy-Aware High Performance Computing (Dresden, Germany) Nikolopoulos, D. Sep 2013 1 p.
7. Prefetching and Cache Management using Task Lifetimes, Papaefstathiou, V., Katevenis, M. G. H., Nikolopoulos, D. & Pnevmatikatos, D. Jun 2013 27th ACM International Conference on Supercomputing . New York, NY,USA: ACM, p. 325-334 10 p.
8. Fast Dynamic Binary Rewriting to Support Thread Migration in Shared-ISA Asymmetric Multicores: Best Paper Award, Georgakoudis, G., Nikolopoulos, D. & Lalis, S. Feb 2013 Proceedings of the First ACM International Conference on Code Optimisation for Multi- and Many-Cores (COSMIC). ACM New York, NY: ACM, p. 4:1 4 p. **Best Paper Award**
9. A Case Study of Automated Testing Implementation in the Automotive Industry, D. Ganea, R. Bogdan, V. Ancusa, M. Popa. 14th IEEE International Symposium on Computational Intelligence and Informatics, CINTI 2013, Budapest, November 19-21, 2013.
10. A. Power Modeling and Estimation during ADL-driven Embedded Processor Design, Wang, Z., Wang, L., Xie, H., and Chattopadhyay, in 4th International Conference on Energy Aware Computing Systems & Applications (ICEAC), Dec 2013, Istanbul, Turkey

Overall progress through M17

- Virtual prototyping
  - Existing multi-core (4 & 4) RISC/VLIW prototype, simulated, integrated with runtime/OS
  - Memory power modeling ongoing
  - Compiler integration ongoing
- Physical prototyping
  - Preliminary OpenRISC multi-core prototype, including LEM, runtime/OS support
  - VP platform porting to FPGA ongoing
- Operating system
  - Lightweight runtime/OS for VP & FPGA based on co-operative multi-threading
  - Energy accounting and allocation framework for ARM Linux & lightweight OS
- Compiler and runtime system
  - Integrated with VP RISC platform and OS
  - Several realistic smartphone apps ported
  - OpenMP+ language implementation ongoing

OS task queues

- Global task list
- Per-core task lists
- RR scheduling
  - Intra-core co-operative scheduling
- Syscall interface
  - Kernel locking
  - Private kernel stack