Energy Transparency from Hardware to Software

Energy-Aware COmputing:

The **EACO** Initiative at Bristol

Kerstin Eder

Design Automation and Verification, Microelectronics

University of BRISTOL

Department of COMPUTER SCIENCE

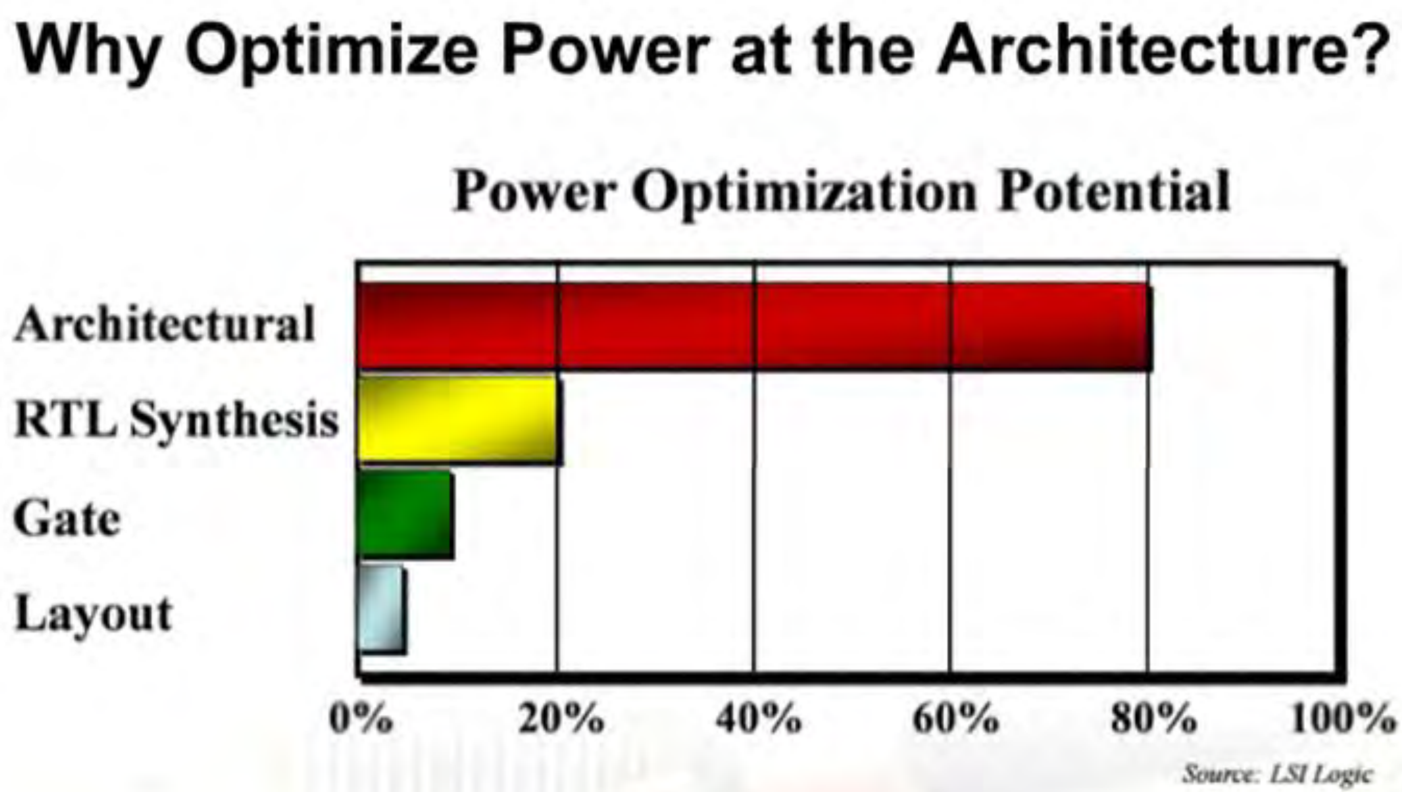
The Royal Academy of Engineering

XMOS®

“Energy-Aware System Design”

- Power management largely in domain of Hardware Design
  - Considerations to minimize/optimize
    - Dynamic (switching) and static (leakage) power
    - Energy consumption
  - On-chip power management
    - DVFS
    - Modes: on, standby, suspend, sleep, off
- Where can the greatest savings be made?

Greater Savings at Higher Levels



LOW POWER

Lack of software support marks the low power scorecard at DAC

One of the panels at the Design Automation Conference (DAC), which took place in California in early June, set out to get an idea of how well the industry is doing at delivering lower-power systems.

It is becoming clear, *writes Chris Edwards*, that the system level is currently the missing link.

Processes can deliver some gains – and Globalfoundries' Andrew Brotman was able to outline some of the features that the foundry has put into its recently launched low-power high-k, metal gate (HKMG) process.

FinFETs should bring power down as those processes become available, although they are not the only options. But if the software keeps cores active for no good reason, the lower switching power per bit processed won't deliver a realised saving.

In his keynote speech Gadi Singer, vice-president IAG and general manager of the SoC enabling group at Intel Corporation, said that with limited software support, dedicated low-power circuitry could save maybe 20% in a typical multimedia-oriented core.

Make the software controlling it better at controlling the power states and that difference could be three to five times.

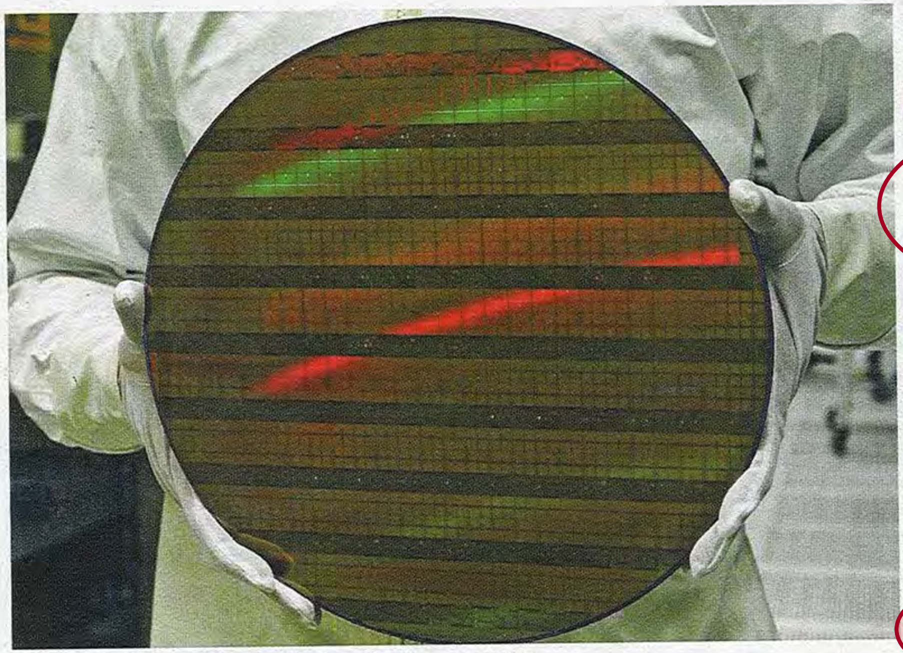
During an afternoon panel discussion Ambrose Low, director of design engineering at Broadcom said: "We have hundreds of knobs in the hardware to turn power down.

"The question is whether we can take the actual use-cases into consideration and optimise the software to power the logic circuits down. We still have a long way to go."

Ruggero Castagnetti of LSI argued that the desire to do more in software will grow.

"As we see power limits and targets becoming unachievable, customers will be willing to go to that extra step. There is a challenge that needs to be addressed and we have to do more on the systems side," Castagnetti said.

"We should put a challenge to the software designers to see how much power they can save," he added.



Intel waits for better low-power software control

Chris Edwards writes the Low-Power Design Blog (enabled by Mentor Graphics) on ElectronicsWeekly.com

www.electronicsweekly.com/ew-blogs/


Lack of software support marks the low power scorecard at DAC

Wasted Potential

Huge advances have been made in power-efficient hardware.

Various software-controllable energy-saving features available, such as DVFS, power modes.

BUT – potential energy savings are wasted by

- software that does not exploit energy-saving features of hardware;
- poor dynamic management of tasks and resources.

The Focus is on Software


news

Lack of software support marks the low power scorecard at DAC

- Software controls the behaviour of the hardware
  - Algorithms and Data Flow
  - Compiler (optimizations)
    - Traditional SW design goals:
      - performance, performance, performance
- Software engineers often **"blissfully unaware"**
  - Implications of algorithm/code/data on power/energy?
  - Power/Energy considerations
    - at best, secondary design goals
  - BUT the biggest savings can be gained from optimizations at the higher levels of abstraction in the system stack – this includes Algorithms, Data and SW

Aligning SW Design Decisions with Energy-Efficiency Design Goal

Key steps*:

- *"Choose the* ***best algorithm*** *for the problem at hand and make sure it* ***fits well with*** *the computational* ***hardware****. Failure to do this can lead to costs far exceeding the benefit of more localized power optimizations.*
- *Minimize* ***memory size*** *and expensive* ***memory accesses*** *through algorithm transformations, efficient mapping of data into memory, and optimal use of memory bandwidth, registers and cache.*
- *Optimize the* ***performance*** *of the application, making* ***maximum use of available parallelism.***
- *Take advantage of* ***hardware support for power management****.*
- *Finally, select instructions, sequence them, and order operations in a way that* ***minimizes switching*** *in the CPU and datapath."*

* Kaushik Roy and Mark C. Johnson. **1997**. "Software design for low power". In *Low power design in deep submicron electronics*, Wolfgang Nebel and Jean Mermet (Eds.). Kluwer Nato Advanced Science Institutes Series, Vol. 337. Kluwer Academic Publishers, Norwell, MA, USA, pp 433-460.

6.3. SOFTWARE DESIGN FOR LOW POWER

KAUSHIK ROY AND MARK C. JOHNSON
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, U.S.A.

1. Introduction

It is tempting to suppose that only hardware dissipates power, not software. However, that would be analogous to postulating that only automobiles burn gasoline, not people. In microprocessor, micro-controller, and digital signal processor based systems, it is software that directs much of the activity of the hardware. Consequently, the software can have a substantial impact on the power dissipation of a system. Until recently, there were no efficient and accurate methods to estimate the overall effect of a software design on power dissipation. Without a power estimator there was no way to reliably optimize software to minimize power. Since 1993, a few researchers have begun to crack this problem. In this chapter, you will learn

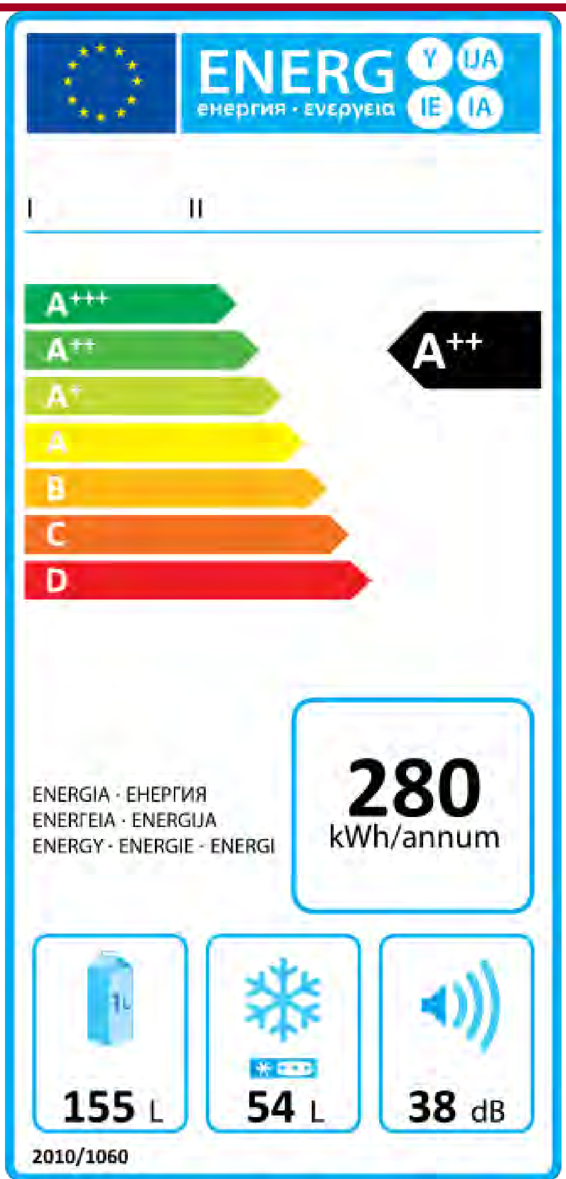
Energy Transparency

Information on energy usage is available for programs:

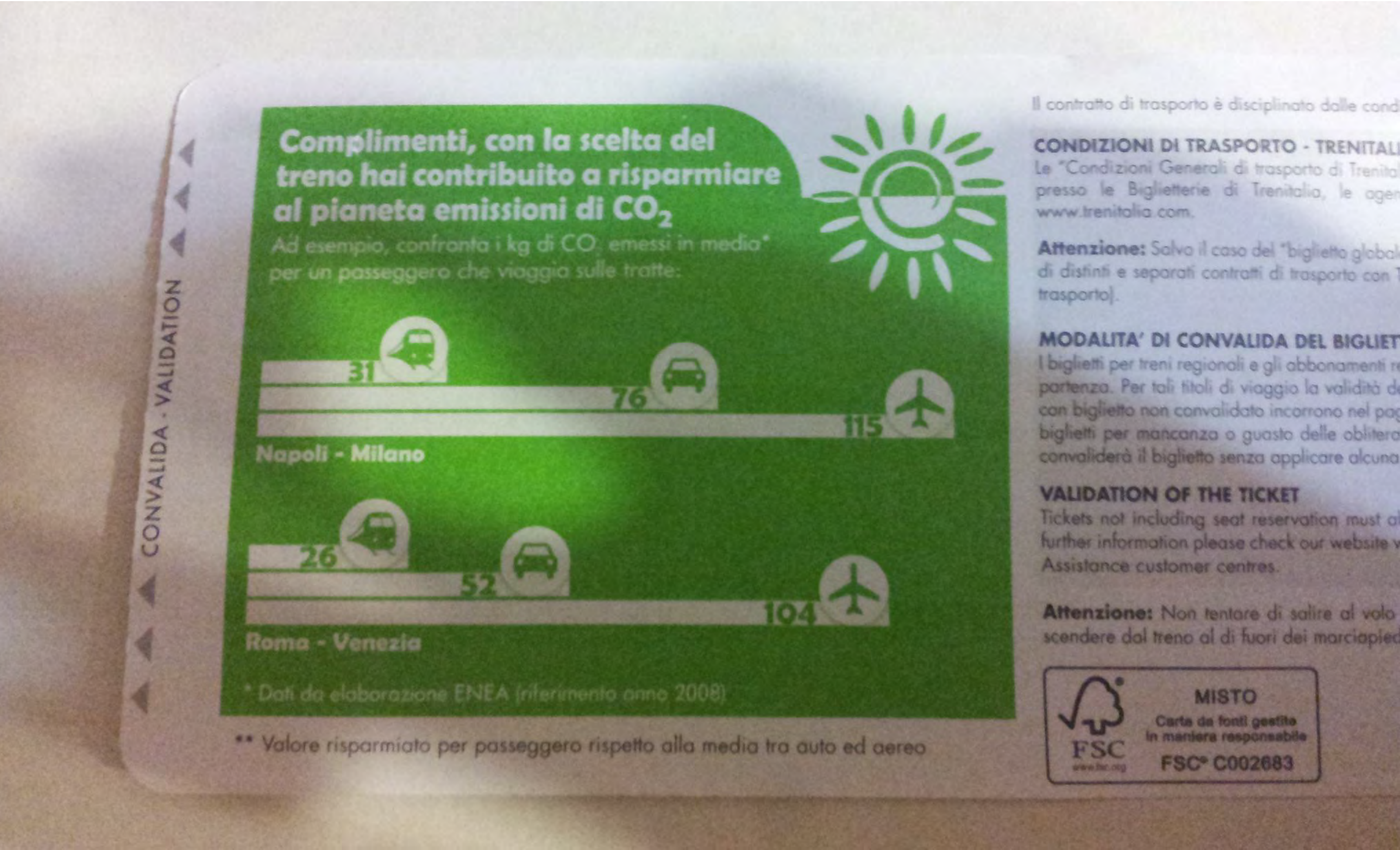
- (ideally) without executing them, and
- at all levels from machine code to high-level application code.

Transparency

Transparency



Transparency



Why Energy Transparency?

Energy transparency enables a deeper understanding of how algorithms and coding impact on the energy consumption of a computation when executed on hardware.

The EACOF

A simple Energy-Aware
COmputing Framework

Comparing Sorting Algorithms

- Sorting of integers in [0,255]

		Data Type											
		uint8_t			uint16_t			uint32_t			uint64_t		
Algorithm	Num Elements	Total Time (s)	Total Energy (J)	Average Power (W)	Total Time (s)	Total Energy (J)	Average Power (W)	Total Time (s)	Total Energy (J)	Average Power (W)	Total Time (s)	Total Energy (J)	Average Power (W)
Bubble Sort	50,000	5.53	66.66	12.03	5.39	65.29	12.09	5.66	69.05	12.19	5.78	71.83	12.41
Insertion Sort	200,000	7.98	■102.18	12.75	7.98	■103.00	12.85	7.46	■98.81	13.21	7.54	■105.03	13.89
Quicksort	2,000,000	5.51	61.73	11.20	5.53	61.90	11.19	5.52	61.60	11.15	5.51	62.90	★11.42
Merge Sort	60,000,000	•6.06	•72.33	11.93	6.07	72.46	11.93	6.12	75.65	12.36	•5.93	•76.98	★12.98
qsort	100,000,000	•5.84	•72.39	12.37	6.15	76.90	12.48	6.79	86.29	12.69	•5.69	•73.25	12.86
Counting Sort	200,000,000	0.23	♦2.92	12.75	0.24	♦3.16	13.23	0.25	♦3.58	14.15	0.35	♦5.12	14.44

■ Insertion Sort: 32 bit version more optimized

♦ Counting Sort:
75% more energy for 64 bit compared to 8 bit values

• Sorting 64 bit values takes less time than sorting 8 bit values, but consumed more energy

★ Average power variations between algorithms

H. Field, G. Anderson and K. Eder, 2013. "EACOF: A Framework for Providing Energy Transparency to enable Energy-Aware Software Development", under review for publication in ACM SAC 2014.

ENTRA

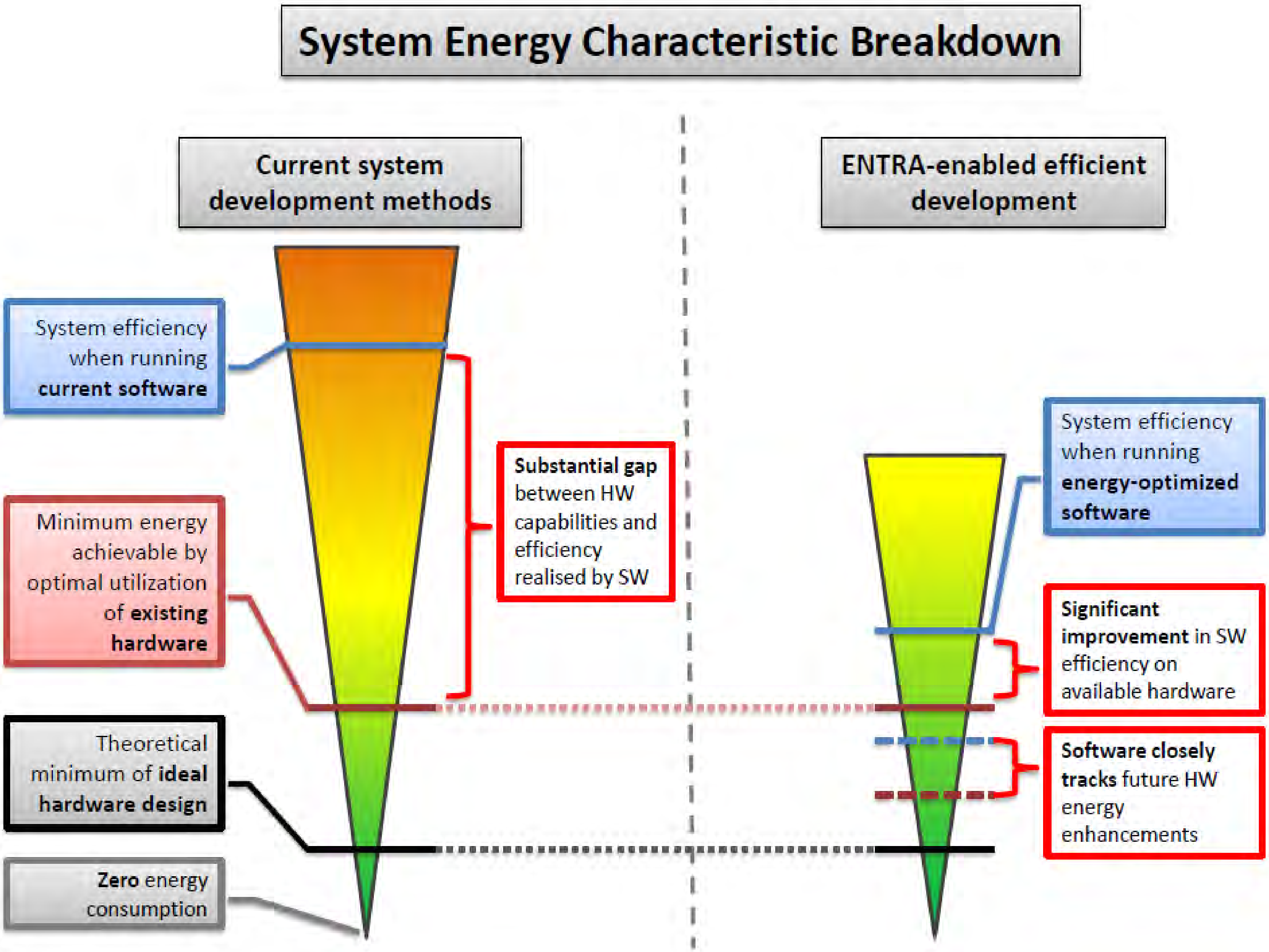
Whole-Systems
Energy Transparency

The ENTRA Project

- Whole Systems ENergy TRAnsparency
  - 1.10.2012 - 30.9.2015
  - EC FP7 FET MINECC:

"Software models and programming methodologies supporting the strive for the energetic limit (e.g. energy cost awareness or exploiting the trade-off between energy and performance/precision)."

Towards the Limit



Energy Modelling

Data to enable

Energy Transparency

Energy Modelling

Energy Cost *(E)* of a program *(P)*:

$$E_P = \sum_i (B_i \times N_i) + \sum_{i,j} (O_{i,j} \times N_{i,j}) + \sum_k E_k$$

Instruction Base Cost, B_i, of each instruction i

Circuit State Overhead, $O_{i,j}$, for each instruction pair

Other Instruction Effects (stalls, cache misses, etc)

V. Tiwari, S. Malik and A. Wolfe. "Instruction Level Power Analysis and Optimization of Software", Journal of VLSI Signal Processing Systems, 13, pp 223-238, 1996.

XCore Energy Modelling

Energy Cost *(E)* of a multi-threaded program *(P)*:

$$E_{\mathrm{p}} = P_{\mathrm{base}} N_{\mathrm{idle}} T_{\mathrm{clk}} + \sum_{t=1}^{N_t} \sum_{i \in \mathrm{ISA}} \left(\left(M_t P_i O + P_{\mathrm{base}} \right) N_{i,t} T_{\mathrm{clk}} \right)$$

Idle base power and duration

Concurrency cost, instruction cost, generalised overhead, base power and duration

- Use of execution statistics rather than execution trace.
- Fast running model with an average error margin of less than 7%

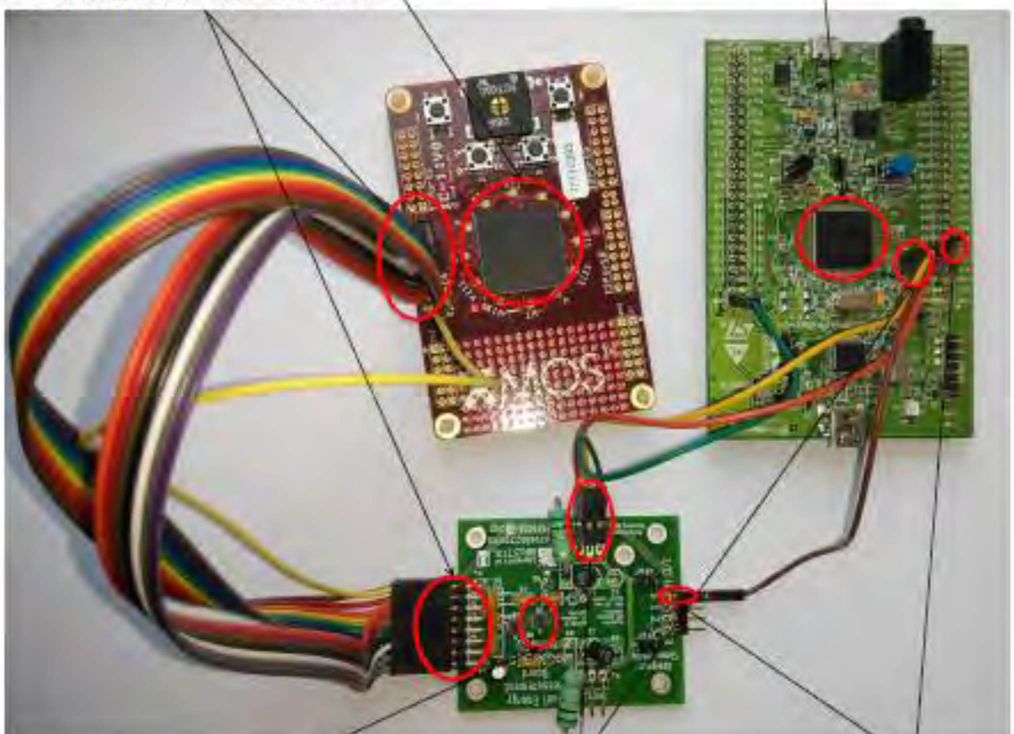
S. Kerrison and K. Eder, 2013. "Energy Modelling of Software for a Hardware Multi-threaded Embedded Microprocessor ", submitted to ACM TECS.

The set up...

32-bit XC-1 core to calculate the power consumption

32-bit ARM Cortex-M4F core with 1MB Flash and 192KB RAM, low power features included

I²C interface connection

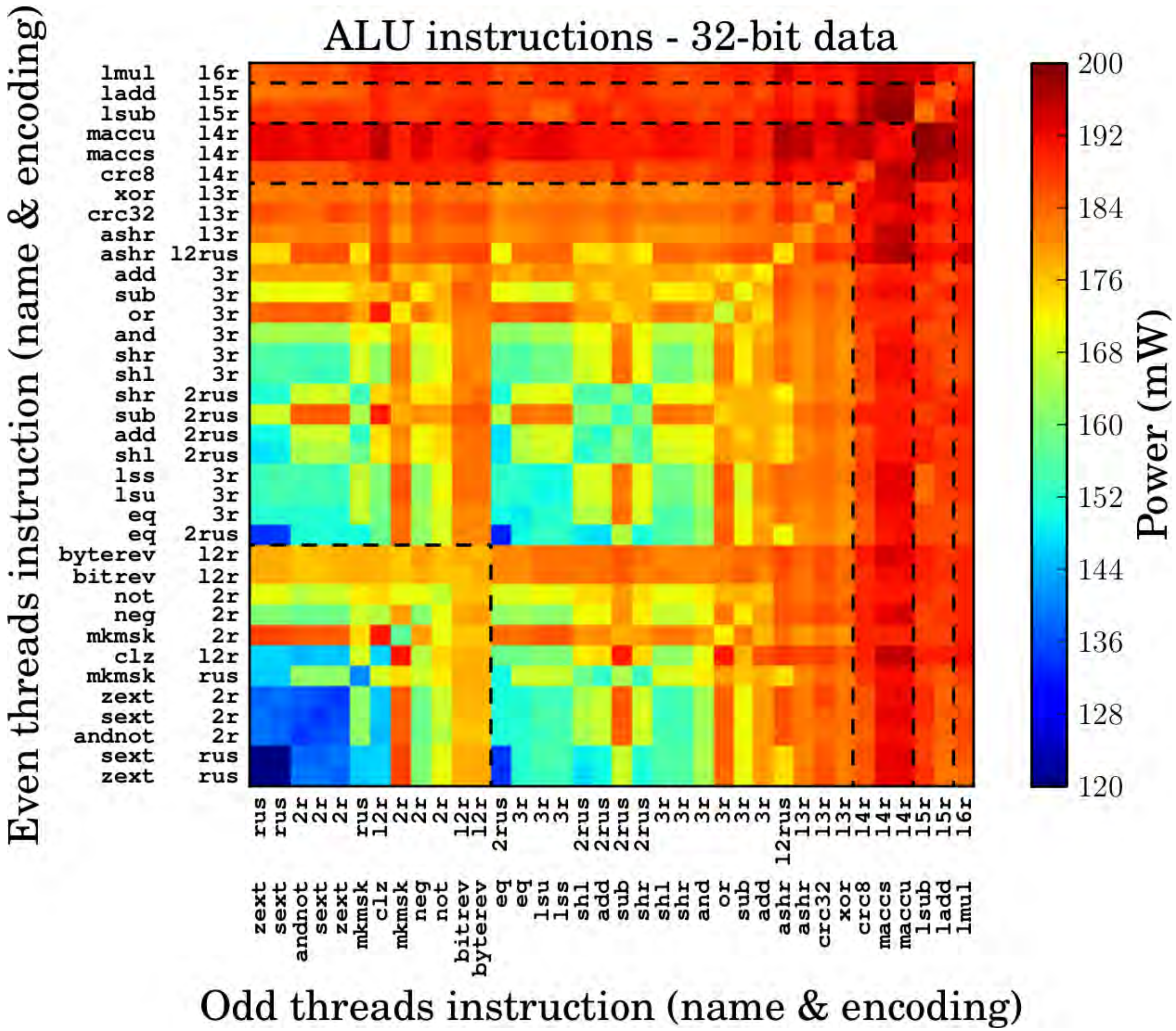


INA219 current/power monitor

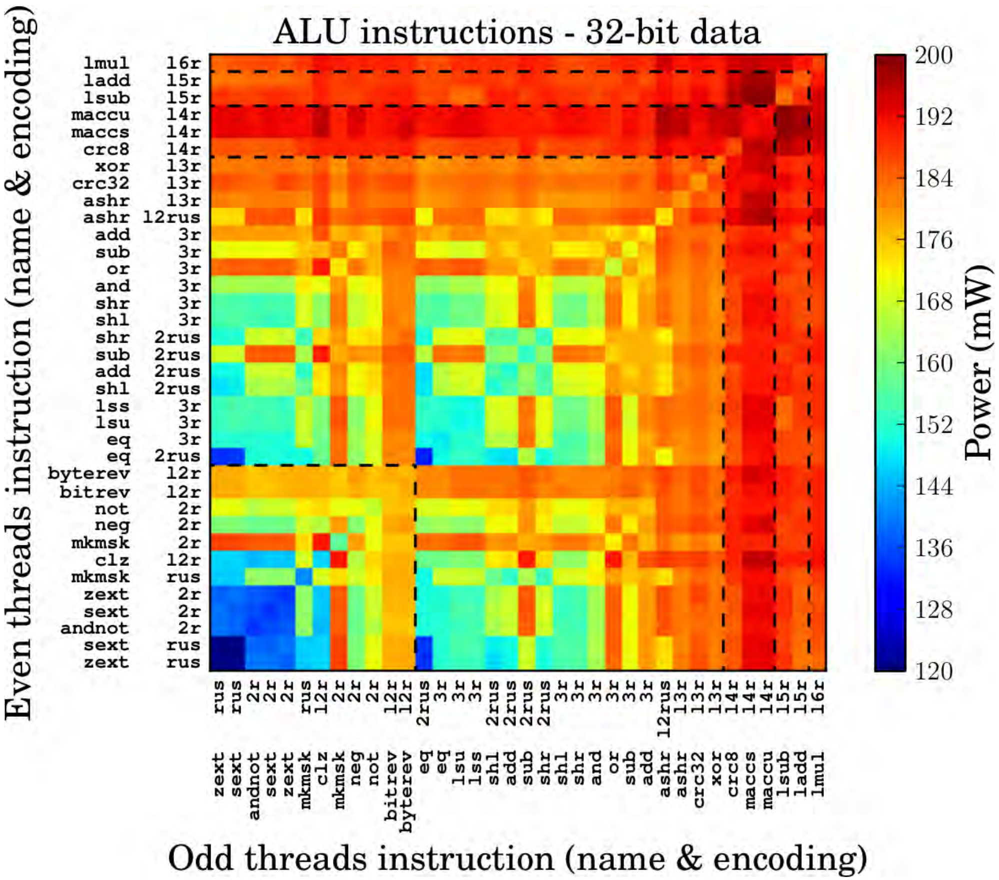
Jumper connection to the supply voltage to the core

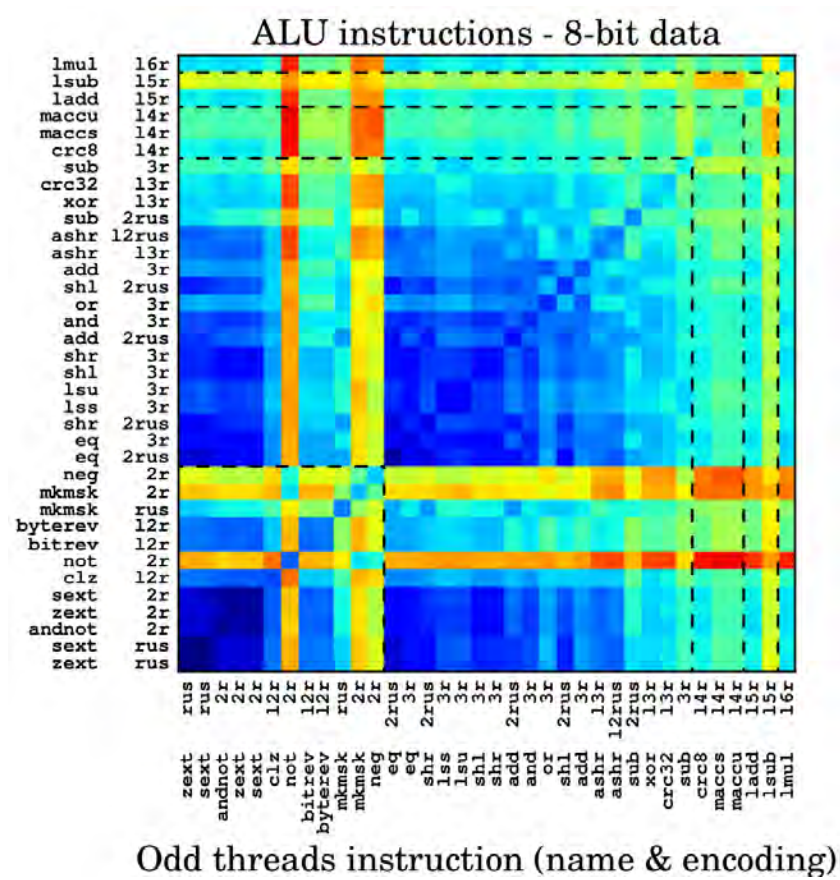
I/O control pin

ISA Characterization

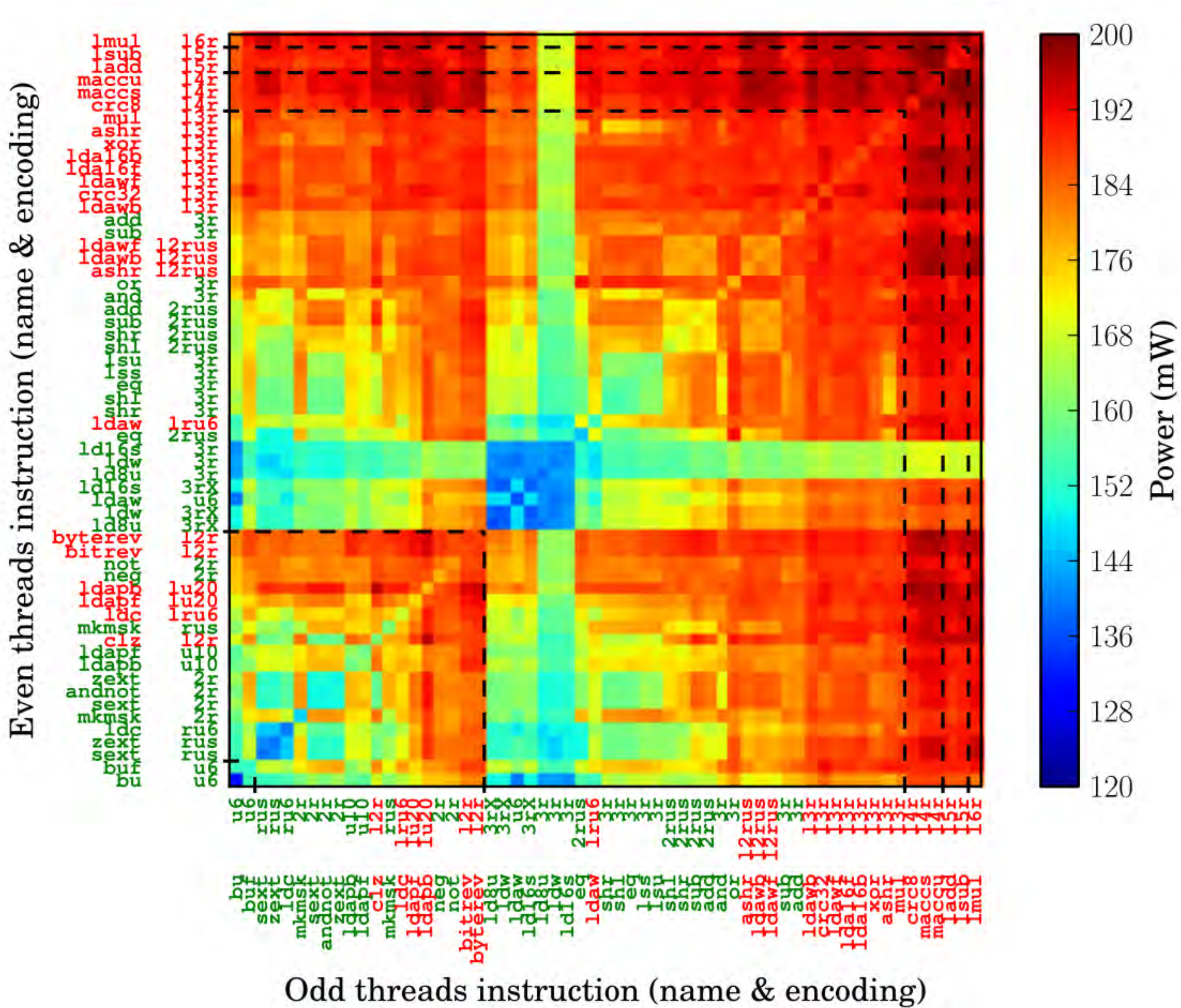


ISA Characterization

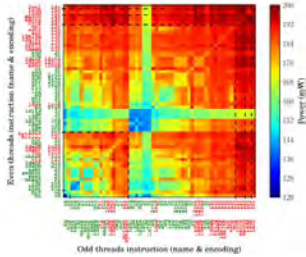




ISA Characterization



New Energy Models



Parameterised models

- Effective capacitance in Farad*
- Average execution time for each instruction in clock cycles
- Core supply voltage
- Clock frequency

Allows exploration of different voltage/frequency operating modes.

Cost of data communication

* C. Brandolese, W. Fornaciari. “Software Energy Optimization Through Fine-Grained Function-Level Voltage and Frequency Scaling”, International Conference on Hardware/Software Co-design and System Synthesis, pages 539-546, New York, USA, 2012.

The Communication "Overhead"

COMPUTE

Operation	Energy
Integer Add	1pJ

Interconnects dominate energy consumption in nVIDIA's 28nm chips

INTERCONNECT

Fetching operands for

3D can shorten these ...nnects

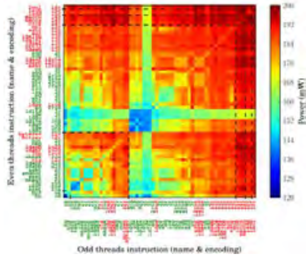
...ained 3D,
...ize TSVs,
...-chip wire problems

...DRAM 3D stacks,
micron-scale TSVs,
less off-chip wire problems

Ref.: W. Dally (nVIDIA), Supercomputing 2010

Communication is a core part of computation, not just an overhead.

New Energy Models



Parameterised models

- Average execution time for each instruction in clock cycles
- Effective capacitance in Farad*
- Core supply voltage
- Clock frequency

Allows exploration of different voltage/frequency operating modes.

Cost of data communication

- Communication instructions IN, OUT are part of XMOS ISA
- Uniform representation of energy consumption model for all instructions, including communication.

* C. Brandolese, W. Fornaciari. “Software Energy Optimization Through Fine-Grained Function-Level Voltage and Frequency Scaling”, International Conference on Hardware/Software Co-design and System Synthesis, pages 539-546, New York, USA, 2012.

Resource Usage Analysis

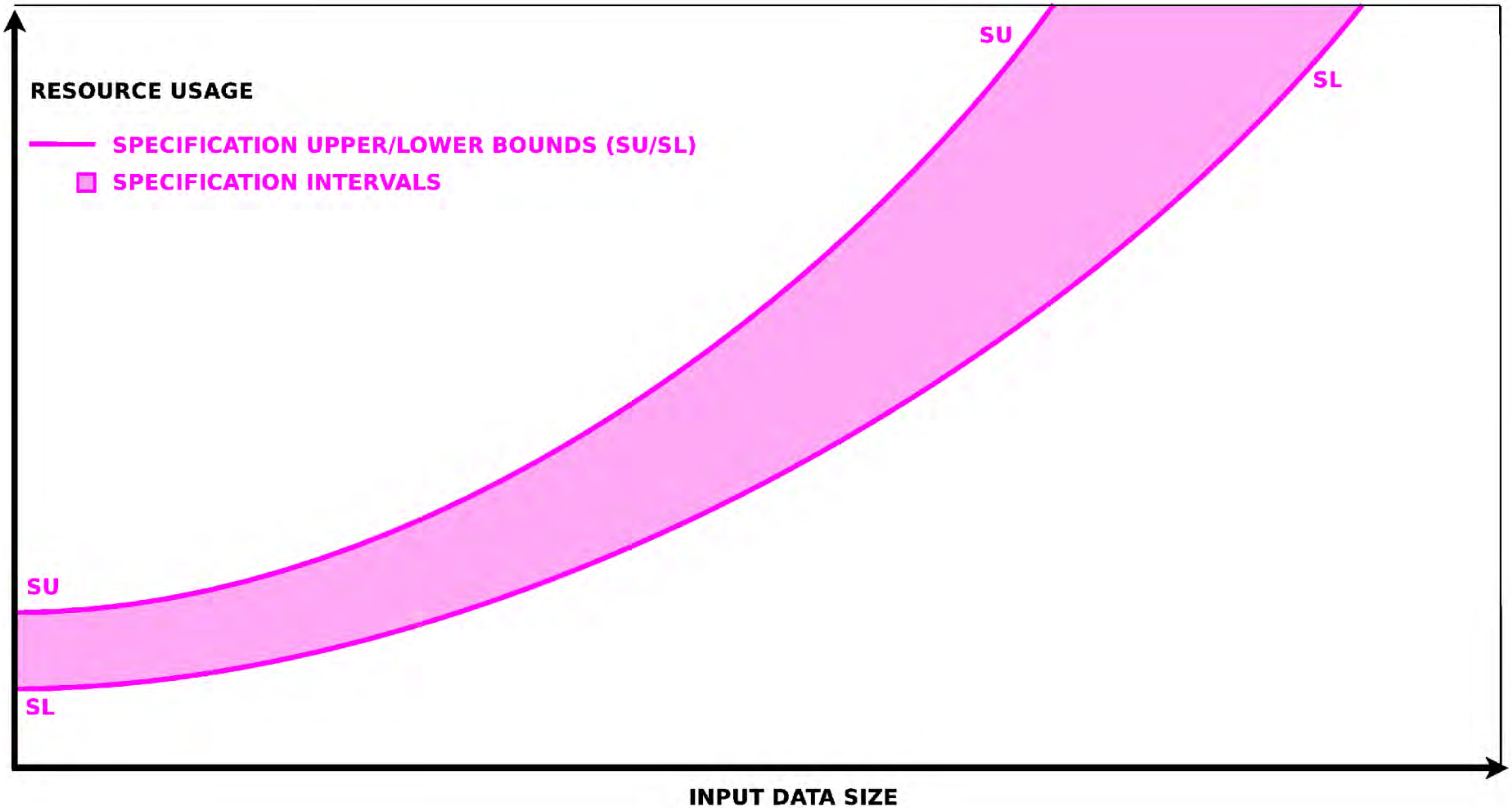
Static analysis for Energy Transparency and Optimization

Resource Usage Analysis

Resource usage analysis framework: CiaoPP

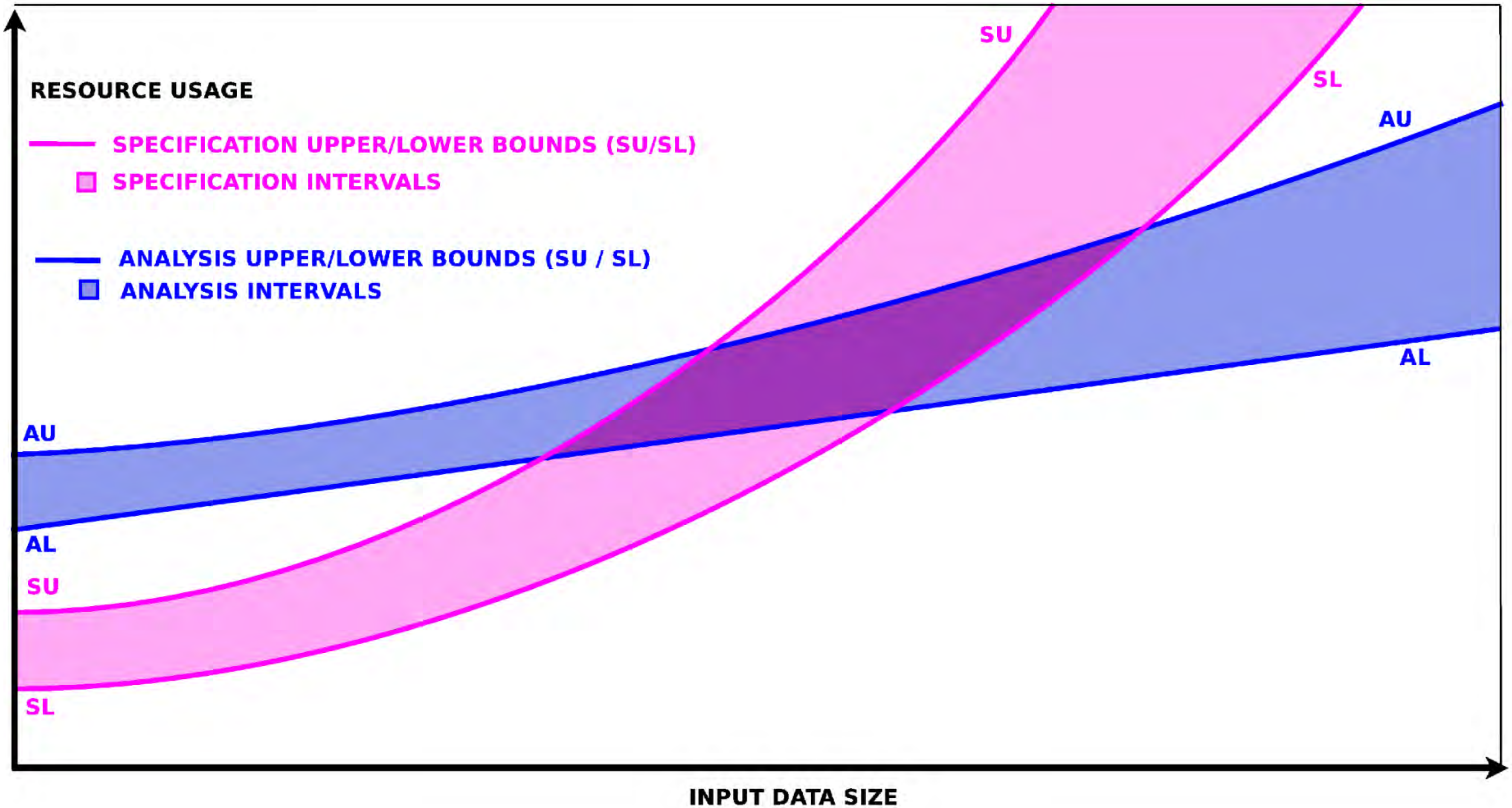
- Adaptation of advanced cost and WCET analysis techniques to energy consumption
- Techniques for automatically inferring both upper and lower bounds on energy usage of a procedure
- Bounds expressed using monotonic arithmetic functions per procedure depending on input data sizes or any other properties (e.g. location)
  - Bounds are safe and also as accurate as possible
- Verification can be done statically by checking if the upper and lower bounds on energy usage and any other resource defined in the specifications hold

Specified Resource Usage



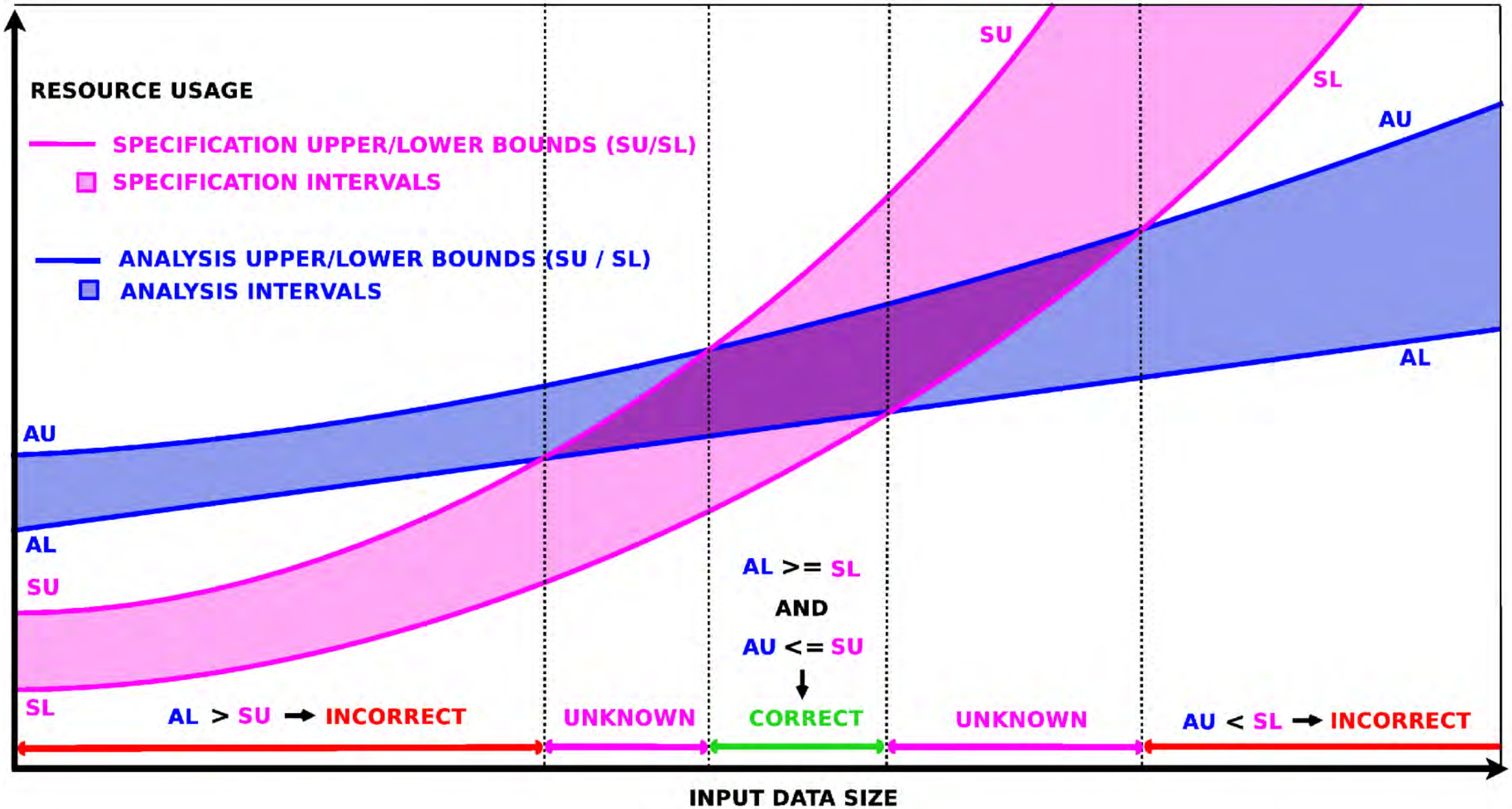
Source: Pedro Lopez Garcia, IMDEA Software Research Institute

Analysis Result



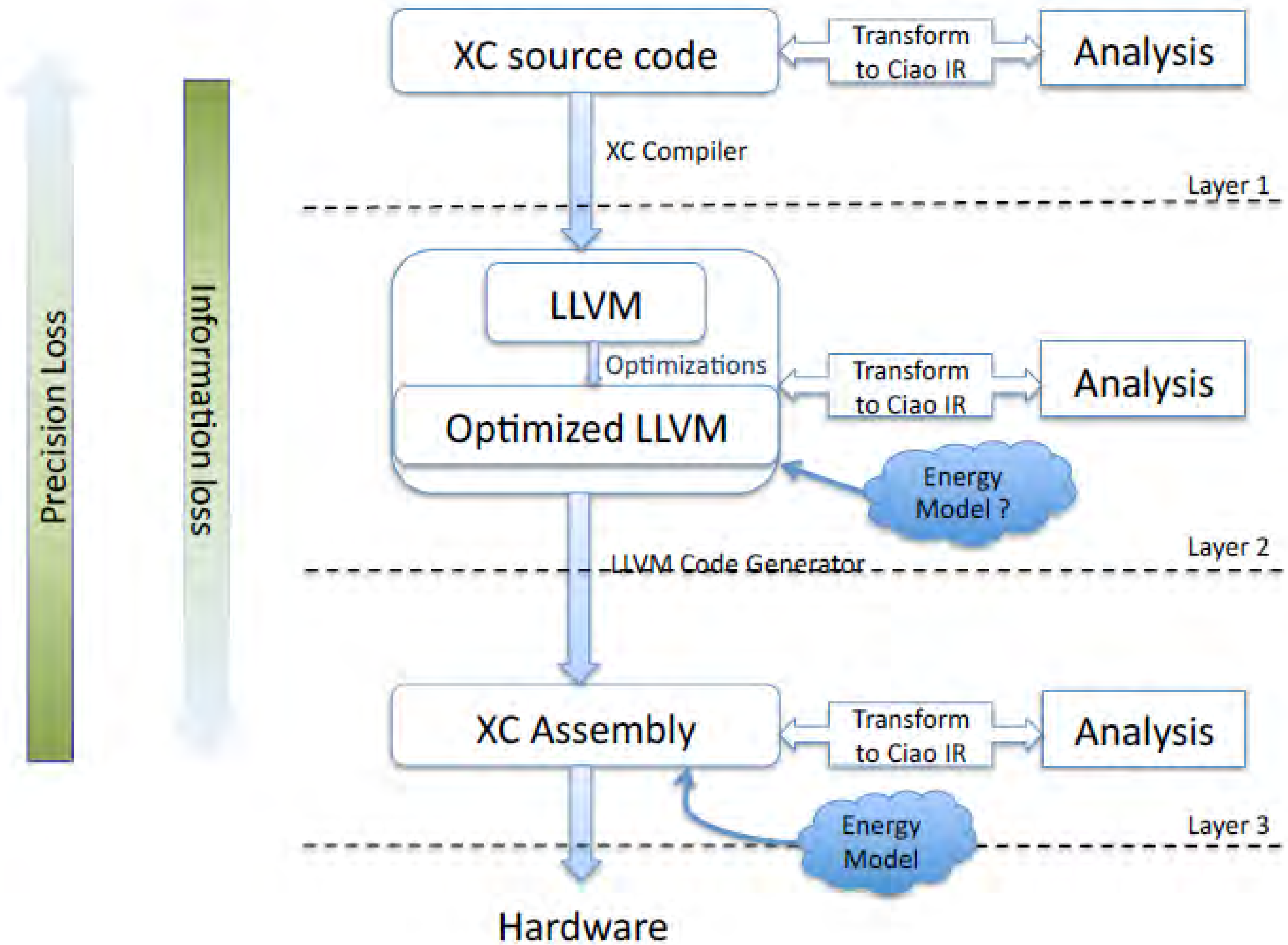
Source: Pedro Lopez Garcia, IMDEA Software Research Institute

Verification

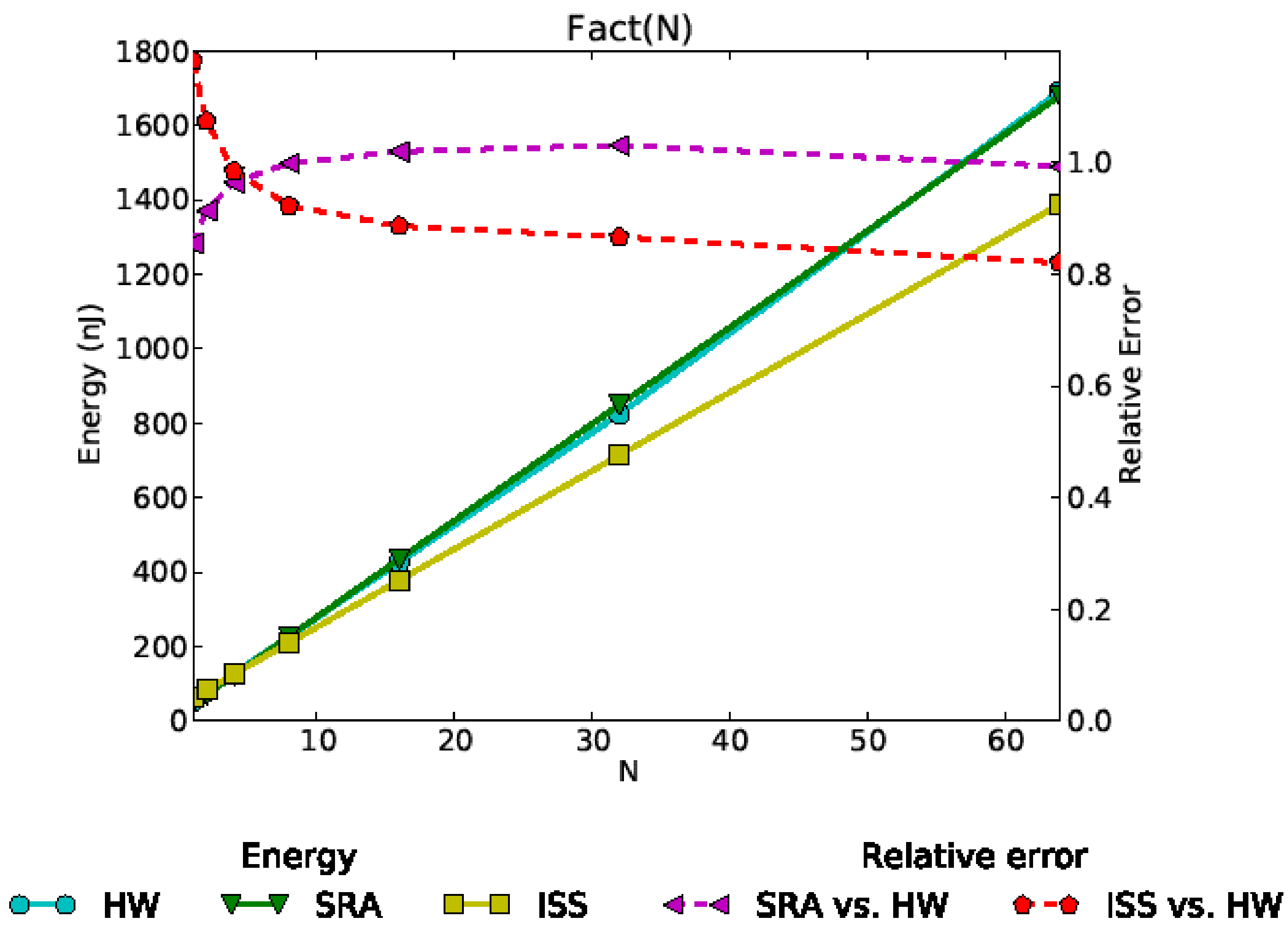


Source: Pedro Lopez Garcia, IMDEA Software Research Institute

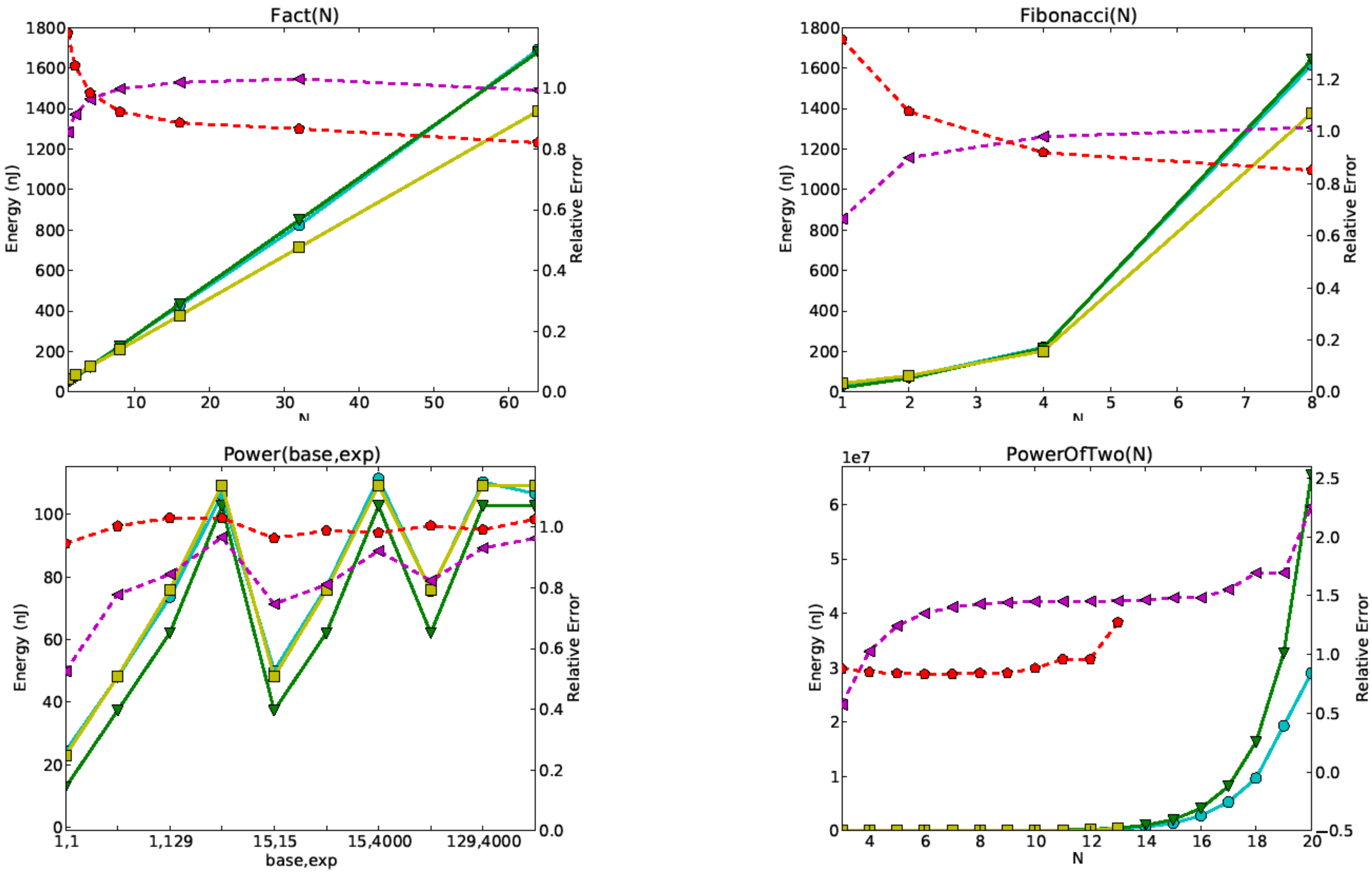
Analysis Options



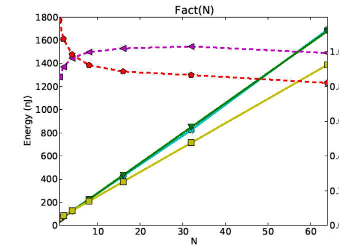
First Results



First Results



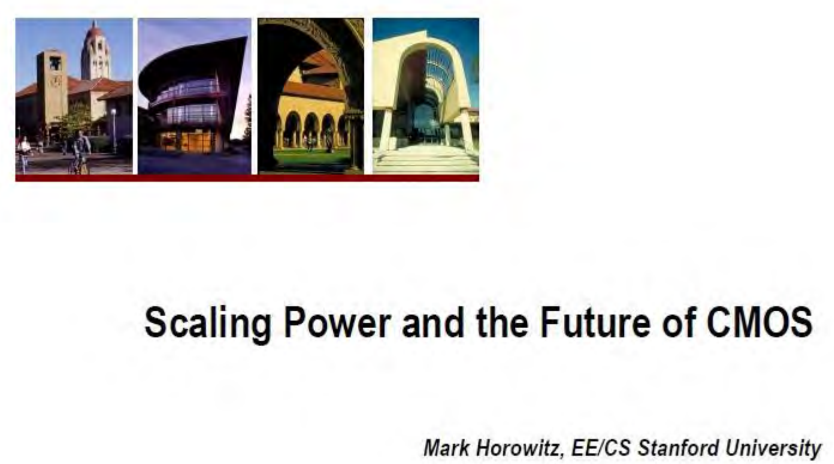
U. Liqat, S. Kerrison, A. Serrano, K. Georgiou, N. Grech, P. Lopez-Garcia, M.V. Hermenegildo and K. Eder. "Energy Consumption Analysis of Programs based on XMOS ISA-Level Models". LOPSTR 2013.



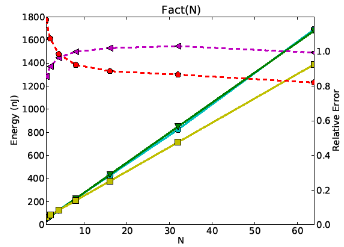
Vision

- For HW designers:
  "Power is a 1st and last order design constraint."
  [Dan Hutcheson, VLSI Research, Inc., E3S Keynote 2011]

- "Every design is a point in a 2D plane."
  [Mark Horowitz,E3S 2009]

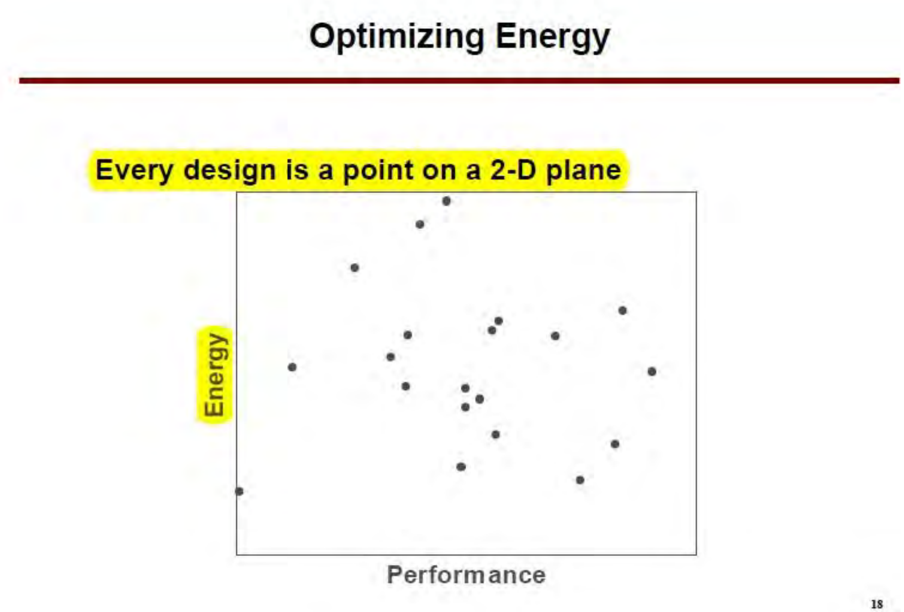


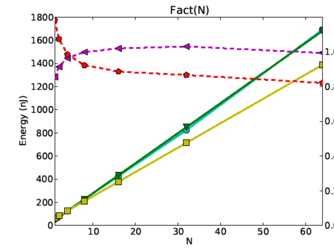
Vision



- For HW designers:
  "Power is a 1st and last order design constraint."
  [Dan Hutcheson, VLSI Research, Inc., E3S Keynote 2011]

- "Every design is a point in a 2D plane."
  [Mark Horowitz,E3S 2009]

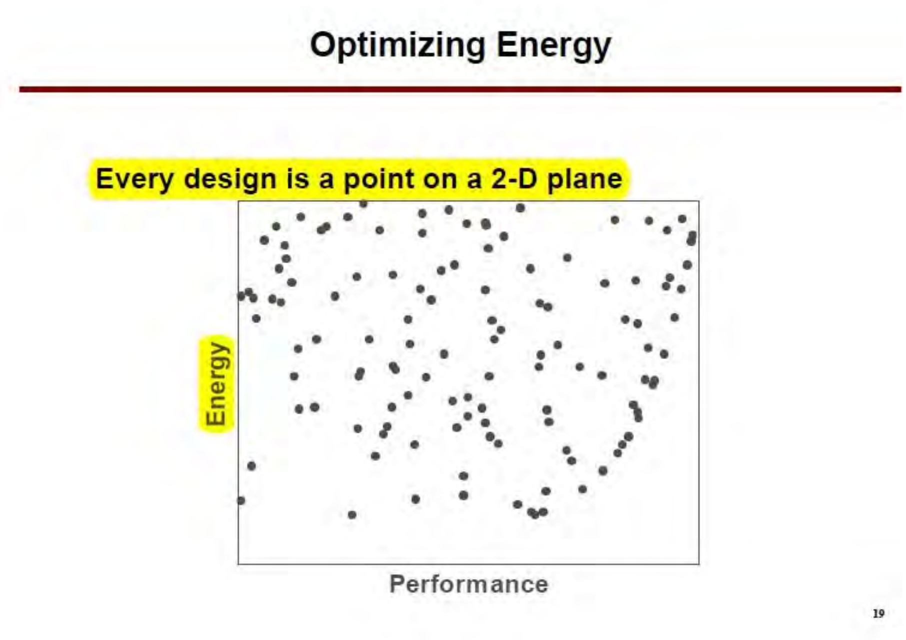


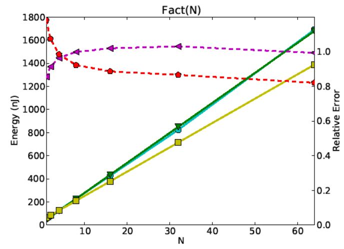


Vision

- For HW designers:
  "Power is a 1st and last order design constraint."
  [Dan Hutcheson, VLSI Research, Inc., E³S Keynote 2011]

- "Every design is a point in a 2D plane."
  [Mark Horowitz,E³S 2009]

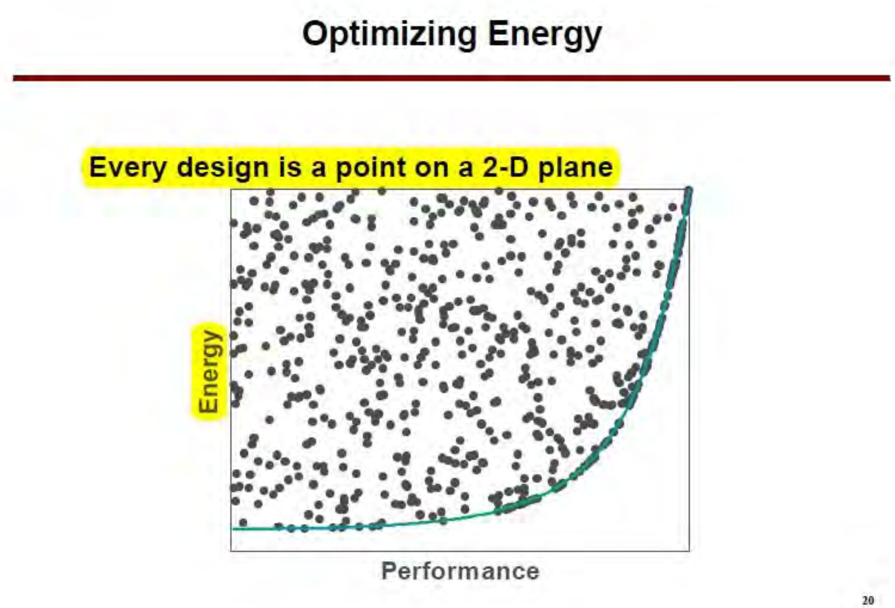




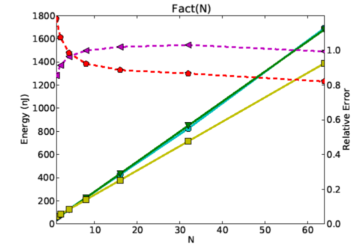
Vision

- For HW designers:
  "Power is a 1st and last order design constraint."
  [Dan Hutcheson, VLSI Research, Inc., E³S Keynote 2011]

- "Every design is a point in a 2D plane."
  [Mark Horowitz,E³S 2009]



Vision



More POWER to software developers!

```
in 15ms do
{...}
```

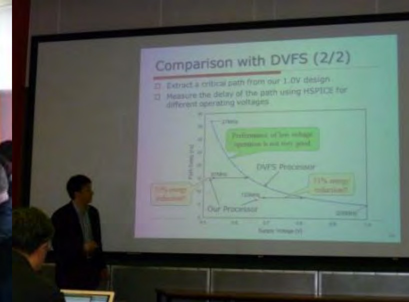
- Energy Transparency from HW to SW
- Location-centric programming model

```
in 29000uJ do
{...}
```

- **"Cool" code**

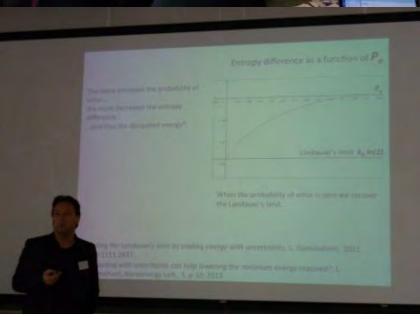
  A cool programming competition!

Promoting energy efficiency to a 1st class SW design goal is an urgent research challenge.



EACO @ Bristol

The Energy-Aware COmputing Initiative, Research Agenda and Workshops



Energy-Aware COmputing

- So far 6 dedicated EACO workshops at Bristol:
  - http://www.bristol.ac.uk/ias/workshops/current-workshops/energy-aware-computing.html
  - http://www.cs.bris.ac.uk/Research/Micro/eaco.jsp

- Intellectual Challenges (pdf online)
  - Inclusive (up/down system stack)
  - Radically new innovative approaches
  - **“We strive to go significantly beyond the soa.”**
  - Strong industrial engagement: Industrial Partners

Energy-Aware Computing (EACO) workshop
18 Apr 2012

University of Bristol

Initiating Industry-Academic Collaborations

Purpose: To bring together researchers and engineers with interests in energy-aware computing for discussions to identify intellectual challenges that can be developed into collaborative research projects. We strive to go significantly beyond the state of the art.

The main objective of this EACO workshop will be to identify and begin organising a number of projects where industrial and academic partners can work together to address some of the Intellectual Challenges put forward in previous EACO workshops.

Preliminary Agenda

18 April 2012 (all times are approximate)

9:00	Registration
9:30	Welcome and Introduction Kerstin Eder and David May, University of Bristol
9:45	Preliminary Project Presentations (I) by current project leads followed by a discussion
	Short Coffee Break
11:00	Presentation of potential funding instruments from funding bodies
11:30	Preliminary Project Presentations (II) by current project leads followed by a discussion
	Lunch including Networking and Intellectual Challenges Discussion
13:30	Round Table Sessions one per project in project working groups
	Short Coffee Break
15:00	"Seed" Project Presentations
16:00	Discussion and Conclusion
	Wine Reception and Networking
17:00	END

Attendance at EACO workshops is FREE. Places are limited.
Please register online at: http://eaco-w4.eventbrite.co.uk/

University of Bristol

Contact
Dr Kerstin Eder
Department of Computer Science
Merchant Venturers Building
University of Bristol
Bristol BS81UB, United Kingdom
Phone: +44(0) 117 954 5146
E-mail: Kerstin.Eder@bristol.ac.uk

Industrial Partners

XMOS, ARM, TVS, PICOCHIP, NVIDIA, nmi, ST, EMBECOSM, cadence, ST ERICSSON, IBM, Imagination

Past Events

EACO Workshop 3 (Oct '11): Alternative Models for Energy-Aware Computation

EACO Workshop 2 (July '11): Beyond the state of the art

EACO Workshop 1 (May '11): Intellectual Challenges

Sponsored by

iNets South West Microelectronics

European Regional Development Fund

Institute for Advanced Studies University of Bristol

μ microelectronics research group
www.cs.bris.ac.uk/Research/Micro

www.cs.bris.ac.uk/Research/Micro/eaco.jsp

Energy-Aware COmputing

- So far 6 dedicated EACO workshops at Bristol:
  - http://www.bristol.ac.uk/ias/workshops/current-workshops/energy-aware-computing.html
  - http://www.cs.bris.ac.uk/Research/Micro/eaco.jsp

Institute for Advanced Studies University of Bristol

iNets South West Microelectronics

European Regional Development Fund: Investing in your Future

- Intellectual Challenges
  - Inclusive (up/down system stack)
  - Radically new innovative approaches
  - **"We strive to go significantly beyond the soa."**
  - Strong industrial engagement: Industrial Partners
- EACO W7: Q1 2014 (~Easter 2014)
  - Dates reserved: 8-10 April and 15 & 16 April 2014
  - Interested collaborators please contact: Kerstin.Eder@bristol.ac.uk

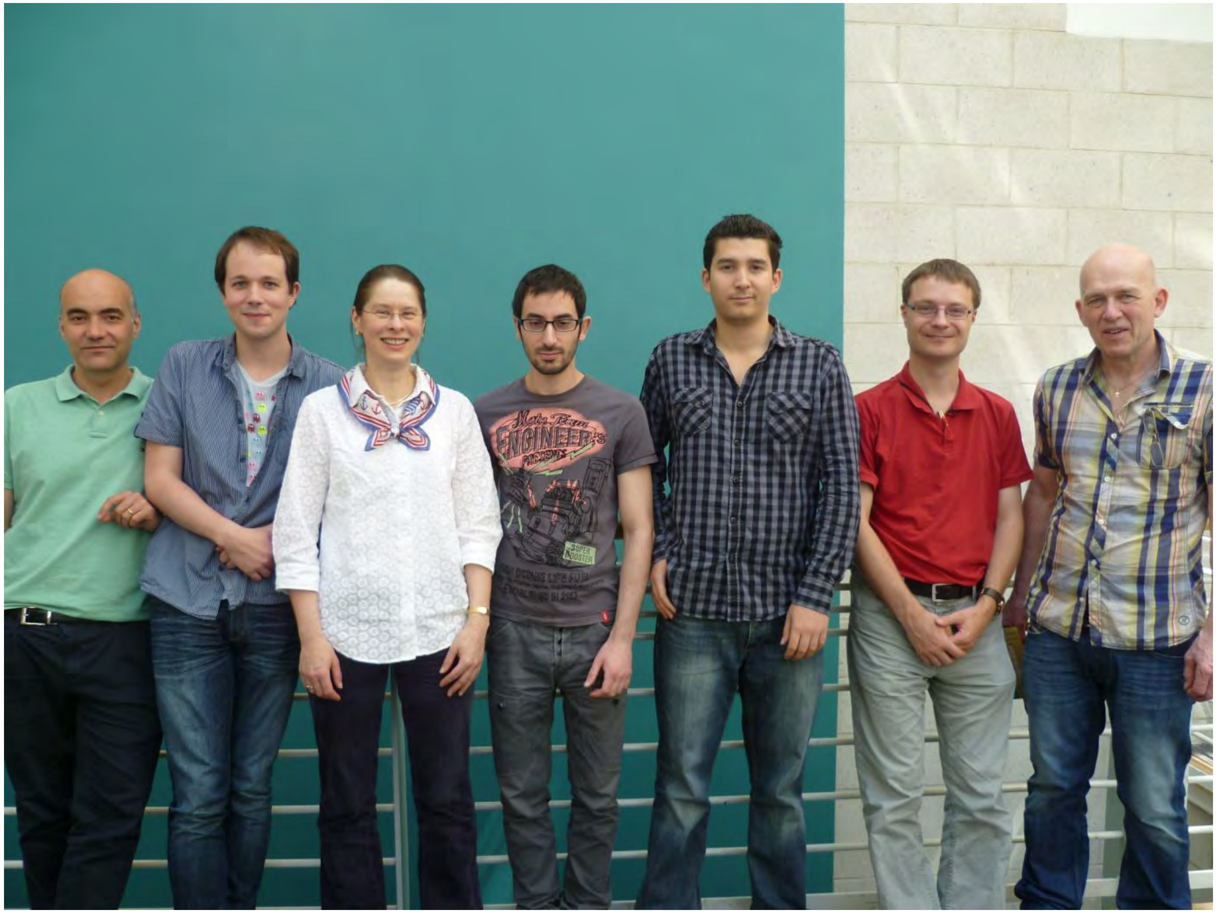
Thank you for your attention

COOPERATION

XMOS®

The Royal Academy of Engineering

University of BRISTOL



Kerstin.Eder@bristol.ac.uk

http://entraproject.eu/

and

http://www.cs.bris.ac.uk/Research/entra/home.html

Watch this space. ☺

Please contact Kerstin.Eder@bristol.ac.uk to find out more or to explore opportunities for collaboration.