

Measuring and modelling the energy consumption of multi-threaded, multi-core embedded software

Steve Kerrison, University of Bristol
steve.kerrison@bristol.ac.uk

Supervisor: Dr. Kerstin Eder
kerstin.eder@bristol.ac.uk

ICT-Energy Community Workshop, BSC

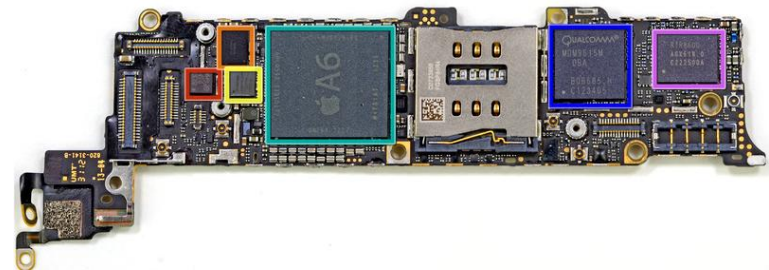
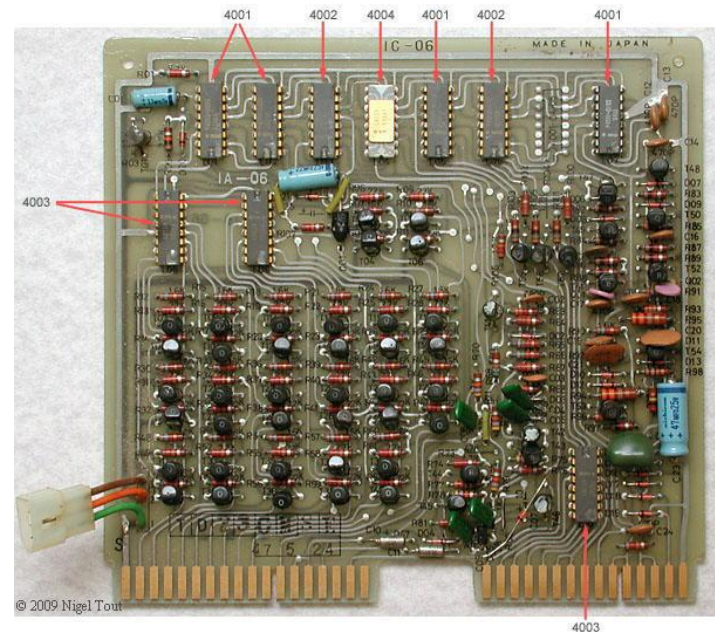
🔥 In this presentation

- Motivation behind my Thesis.
- Modelling a modern, embedded microprocessor.
- Demonstrating communication costs.
- Ongoing and future work.



Motivation

- Embedded systems are everywhere.
- Becoming more complex with every generation.
- Hardware and software increases in complexity.



Motivation

- Every generation of hardware strives for better energy efficiency.

Kepler – first introduced in desktop and notebook systems, and later brought to workstations and supercomputers – is the world's fastest and most energy-efficient GPU architecture.

NVIDIA K1 UK press release,
CES January 2014

- **What about software?**

Motivation



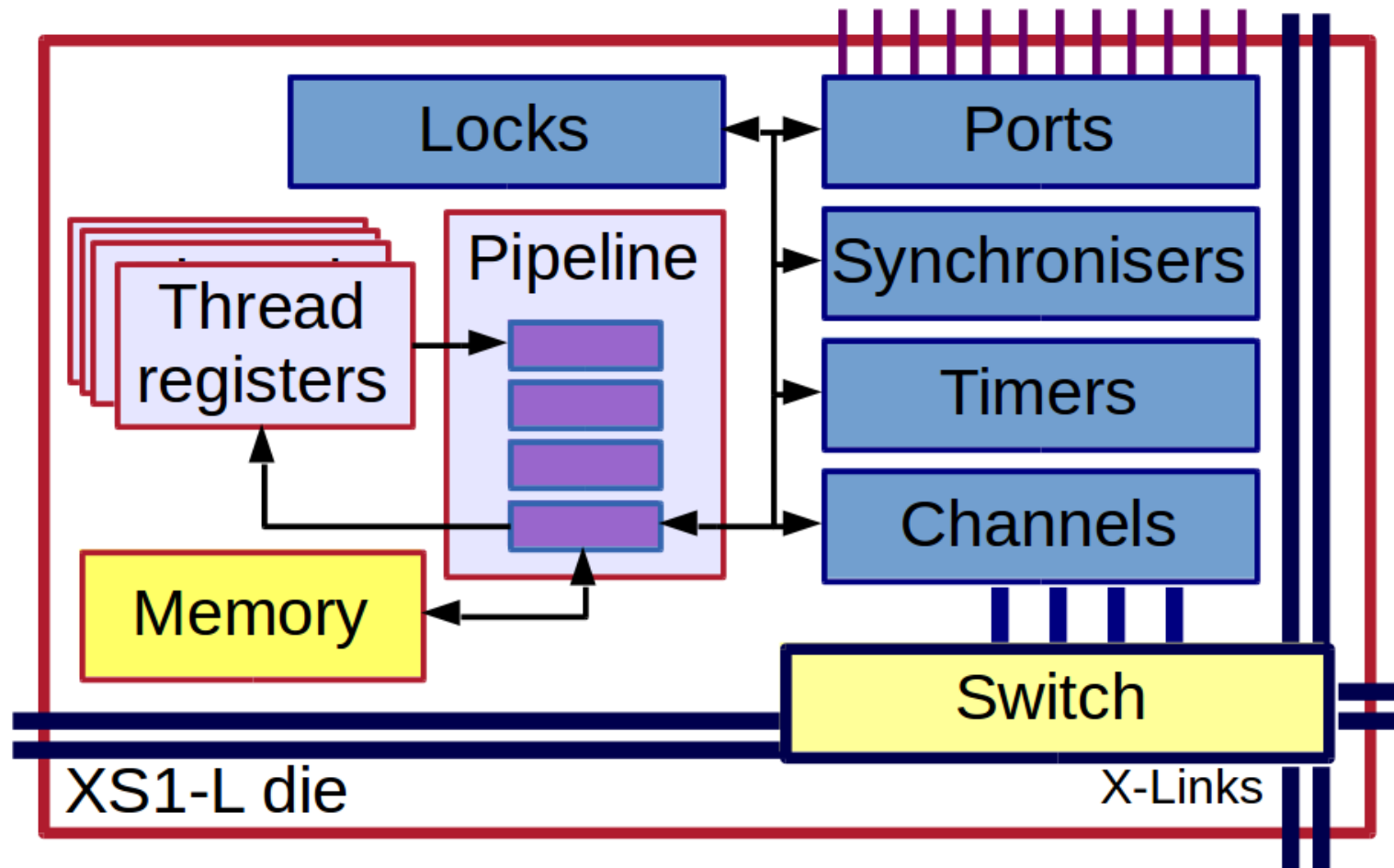
- Helping software engineers understand the energy cost of the code that they write.
- Encouraging a **better understanding** of the **behaviour** and **exploitable features** of an embedded processor and the **system** around it.

🔥 In this presentation

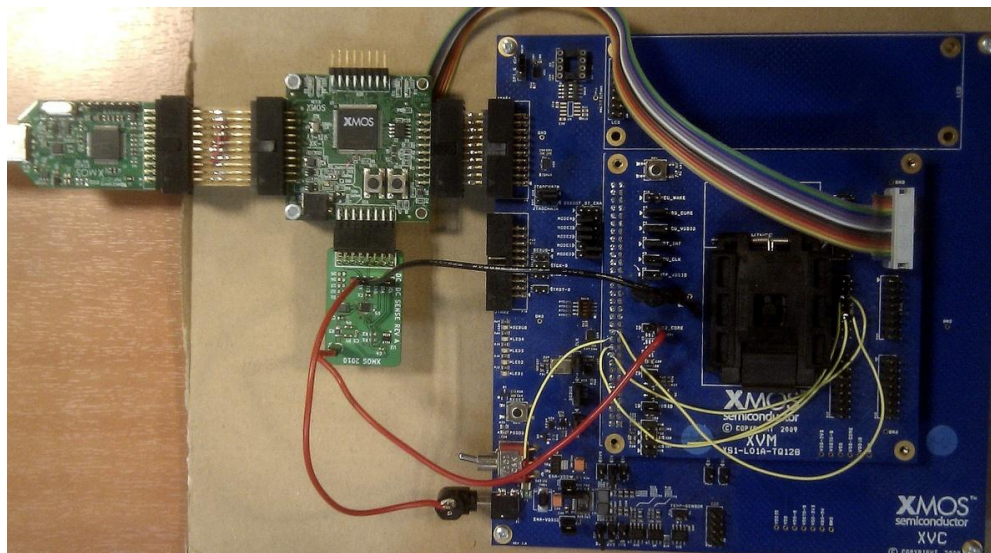
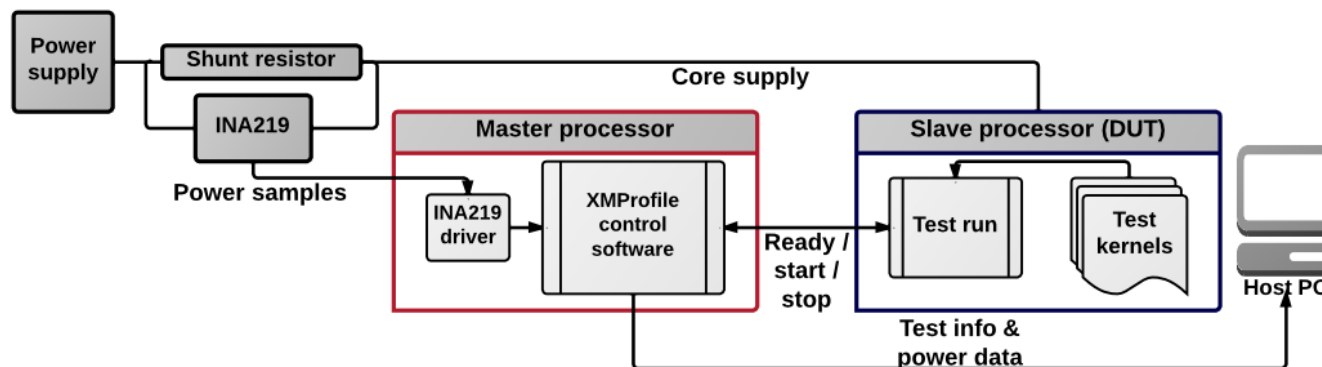
- Motivation behind my Thesis.
- **Modelling a modern, embedded microprocessor.**
- Demonstrating communication costs.
- Ongoing and future work.



XMOS XS1-L



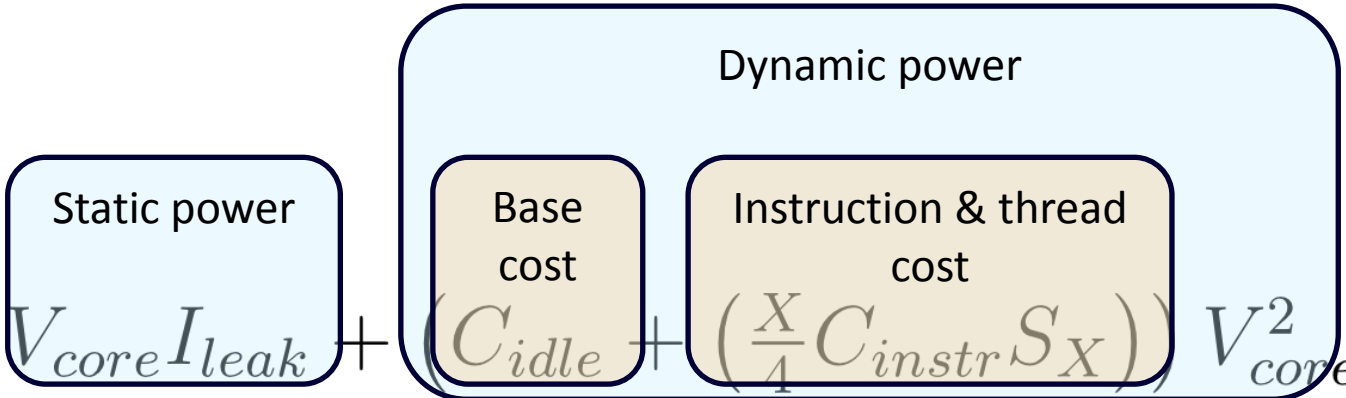
The set up...



-
- Heatmap visualization showing the power consumption (mW) for various instructions (Y-axis) against other instructions (X-axis). The color scale ranges from 100 mW (blue) to 200 mW (red). The X-axis labels are instruction names and their encodings, and the Y-axis labels are instruction names and their encodings.

Odd threads instruction (name & encoding)

Model



The diagram illustrates the components of the energy model. A large light blue rounded rectangle labeled 'Dynamic power' contains two smaller yellow rounded rectangles: 'Base cost' and 'Instruction & thread cost'. To the left of this rectangle is a light blue rounded rectangle labeled 'Static power'. The equation below shows these components as terms in a sum, all divided by $N_{threads}$ and multiplied by T .

$$E = \frac{V_{core} I_{leak} + (C_{idle} + (\frac{X}{4} C_{instr} S_X)) V_{core}^2 F}{N_{threads}} T$$

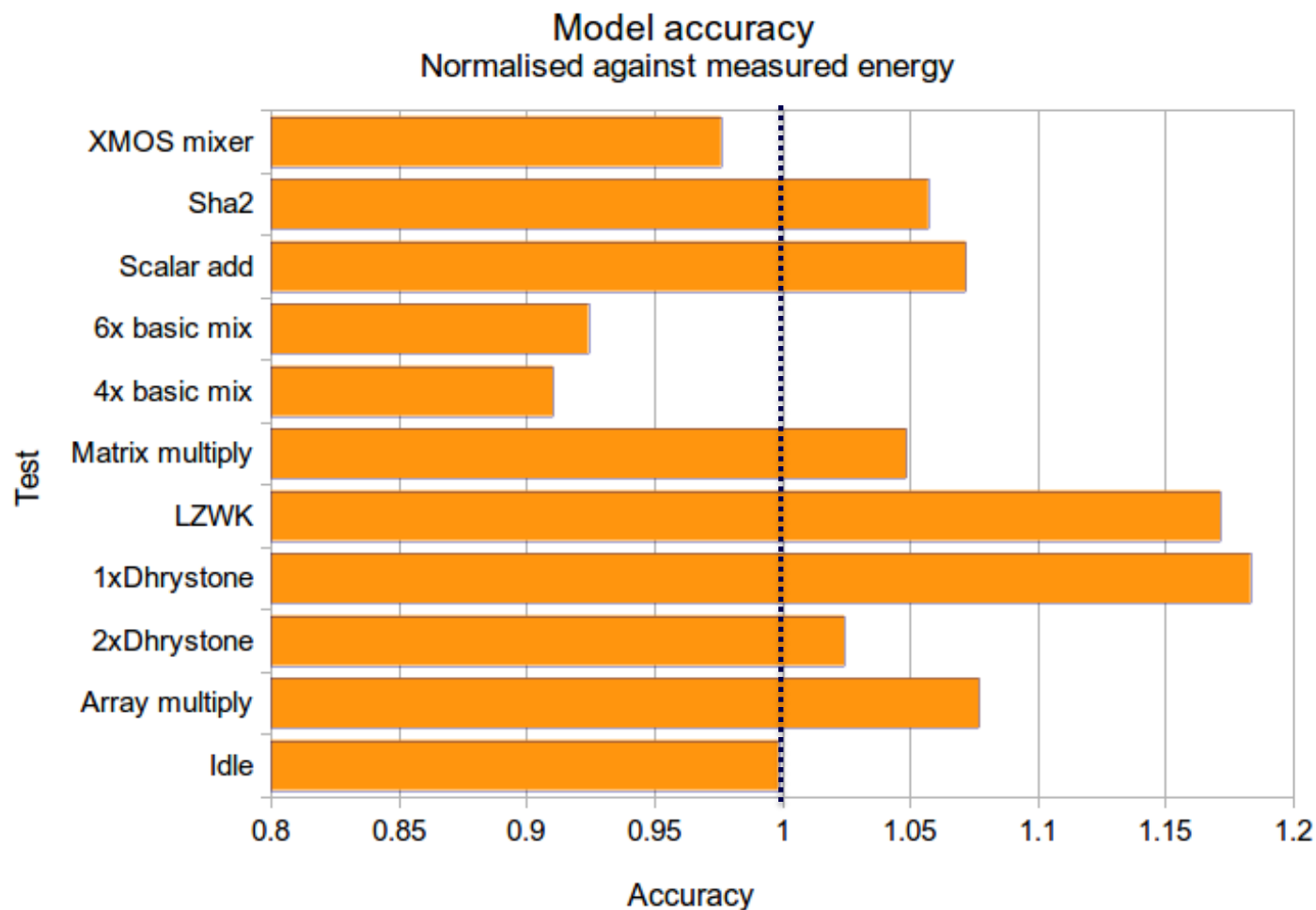
$$X = \min(N_{threads}, 4)$$

- Express frequency, voltage, thread activity, instruction costs and time consumption.

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

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

Preliminary results

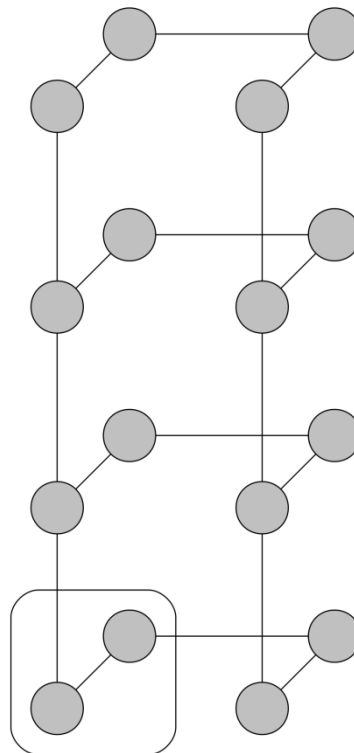
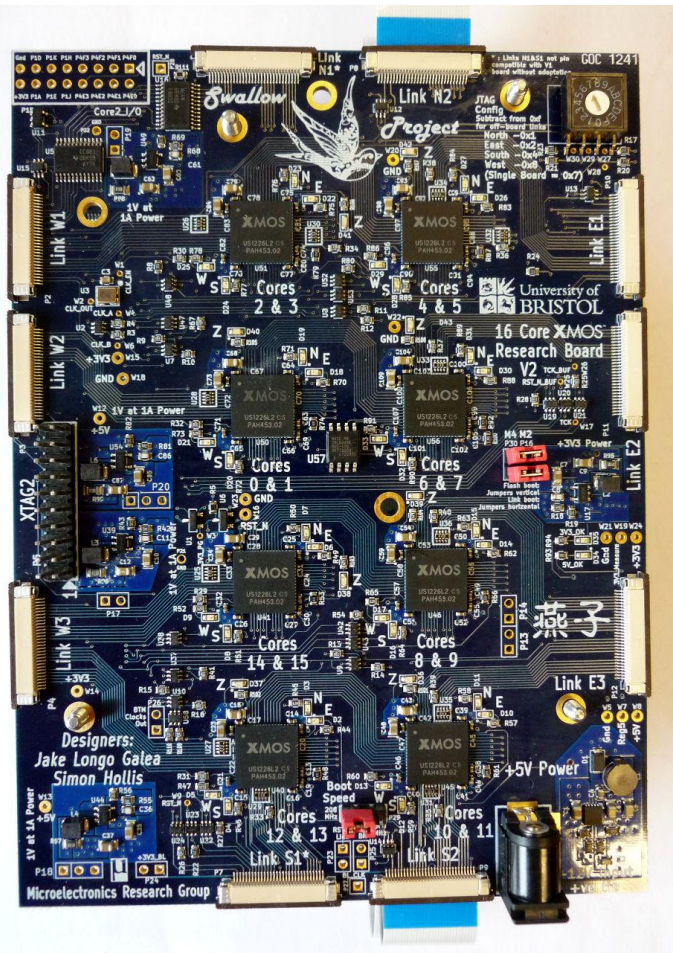


🔥 In this presentation

- Motivation behind my Thesis.
- Modelling a modern, embedded microprocessor.
- **Demonstrating communication costs.**
- Ongoing and future work.



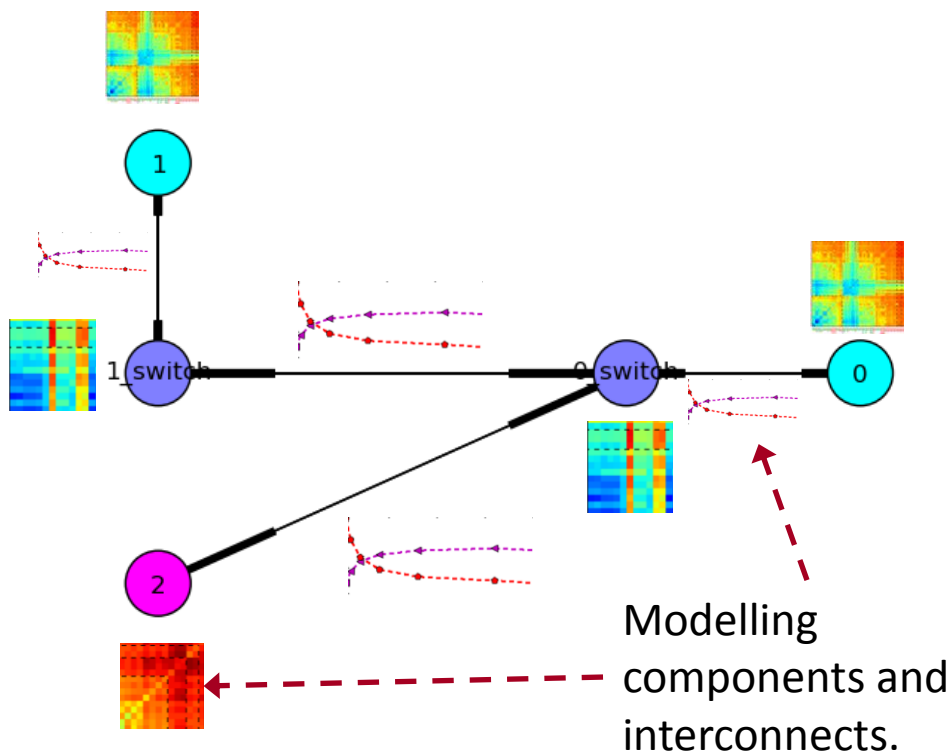
Swallow



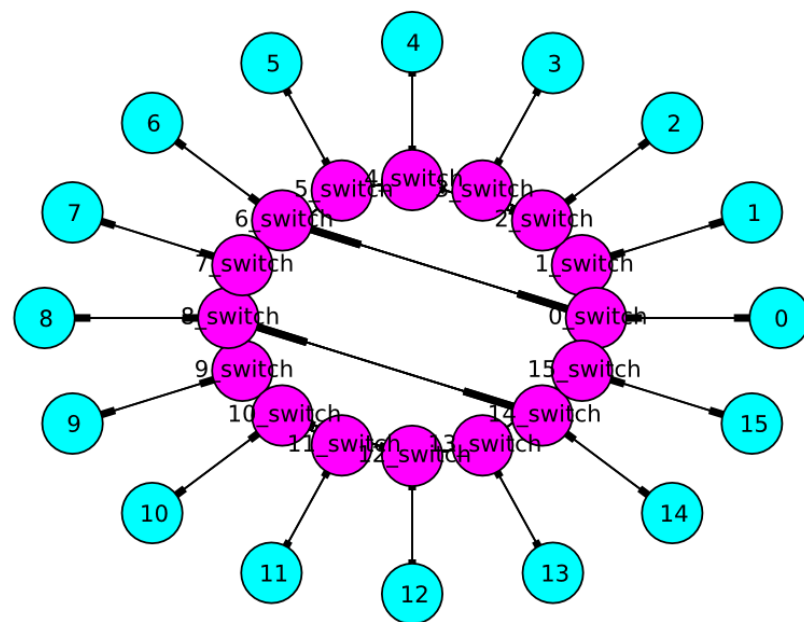
- 16 cores
- 8 chips
- < 24 watts
- 8 threads per core
- Low-latency network
- ISA provides network access

Comms & peripheral hardware

2-tile XMOS network with USB

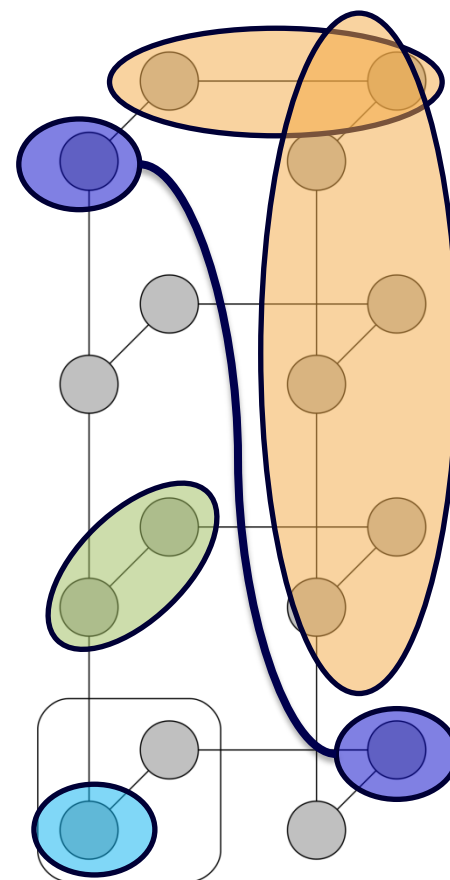


16-tile XMOS board (UoB project Swallow)



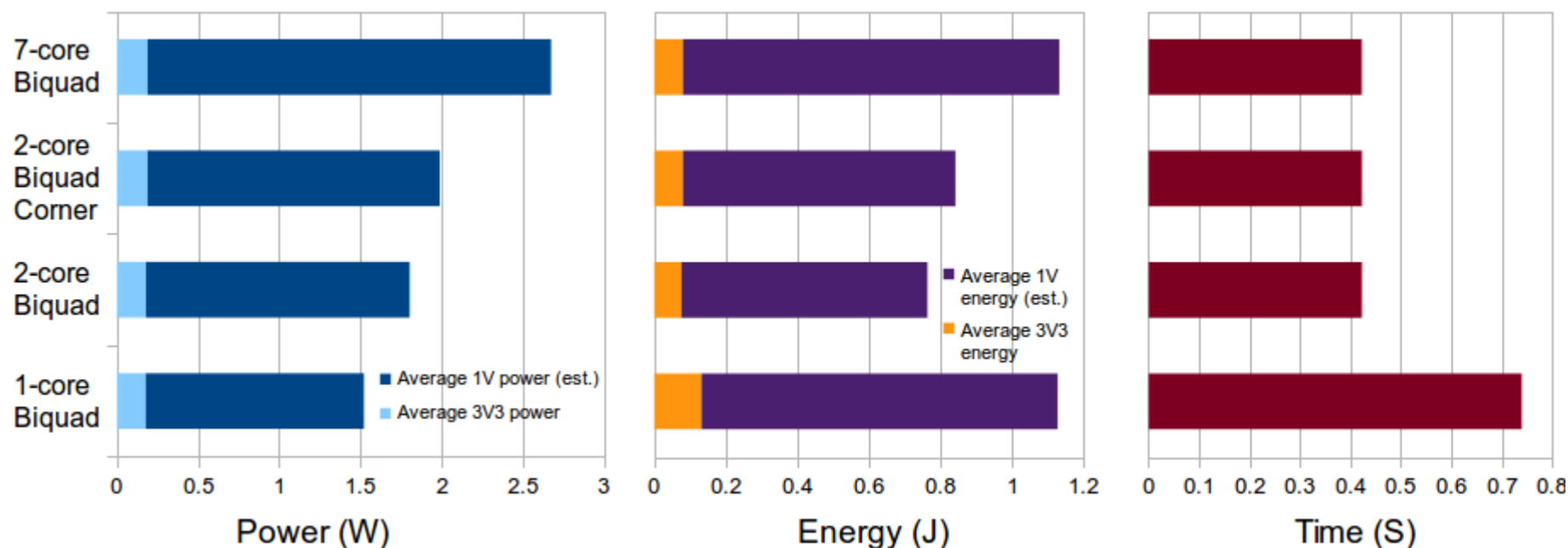
Biquad filter

- 7-stage biquad filter implemented in various configurations on Swallow.
 - 7 threads on 1 core.
 - 7 threads across two cores.
 - Good spatial locality.
 - Bad spatial locality.
 - 7 threads across 7 cores.



Comms example: Biquad filter

- Active cores, latency, contention and under/over-allocation all affect total energy.
- Power, energy & time a valuable triple.



Communication model



- Simpler than core model.
 - Just model the behaviour based on the amount of data carried.
- Several components can be modelled in this way:
 - Switches
 - Links
 - Peripherals

$$E_{com} = \left(V_{com} I_{leak} + N F_{com} V_{com}^2 \right) T$$

Simulation



```
pigz -dc fir-20140416.axe.trace.gz |  
./xmtracem2.py --xn swallow-1x1.xn -m model-20140307.pkl
```

Model stats

=====

Base power: 90.44e-03 W

Time (clocks/wall): 1e06 clk / 3.11e-03 S

Core 0:

Energy (static	dynamic	comms	total):
0.28e-03 J	0.28e-03 J	0.00e-03 J	0.55e-03 J
Power (static	dynamic	comms	total):
88.65e-03 W	88.71e-03 W	0.00e-03 W	177.36e-03 W

Core 1:

Energy (static	dynamic	comms	total):
0.18e-03 J	0.14e-03 J	0.00e-03 J	0.33e-03 J
Power (static	dynamic	comms	total):
59.36e-03 W	45.80e-03 W	0.00e-03 W	105.16e-03 W

...

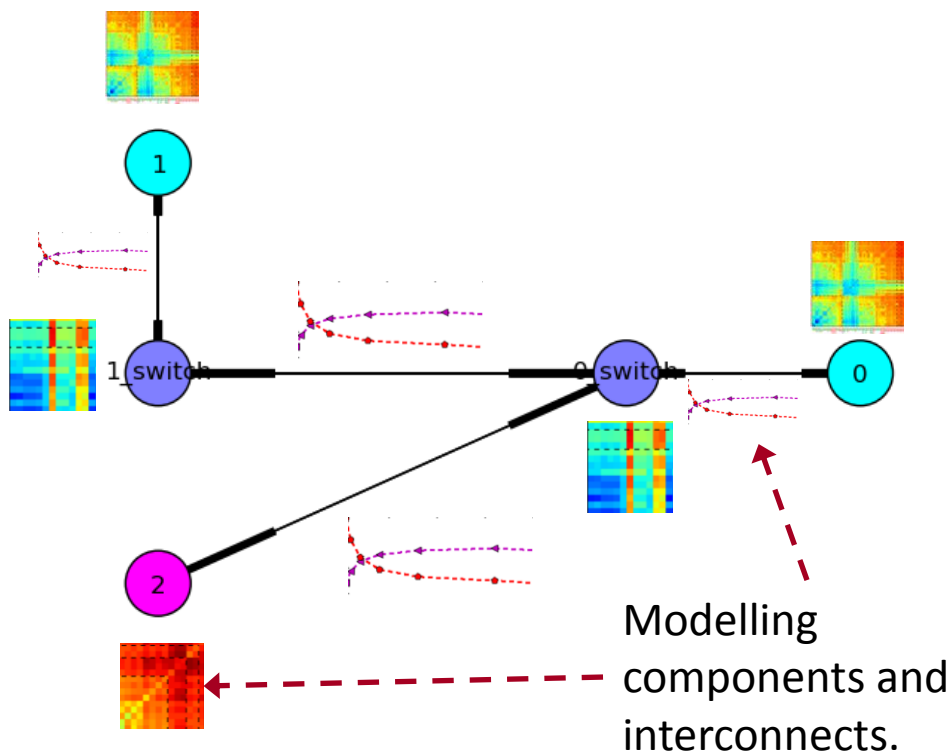
🔥 In this presentation

- Motivation behind my Thesis.
- Modelling a modern, embedded microprocessor.
- Demonstrating communication costs.
- **Ongoing and future work.**

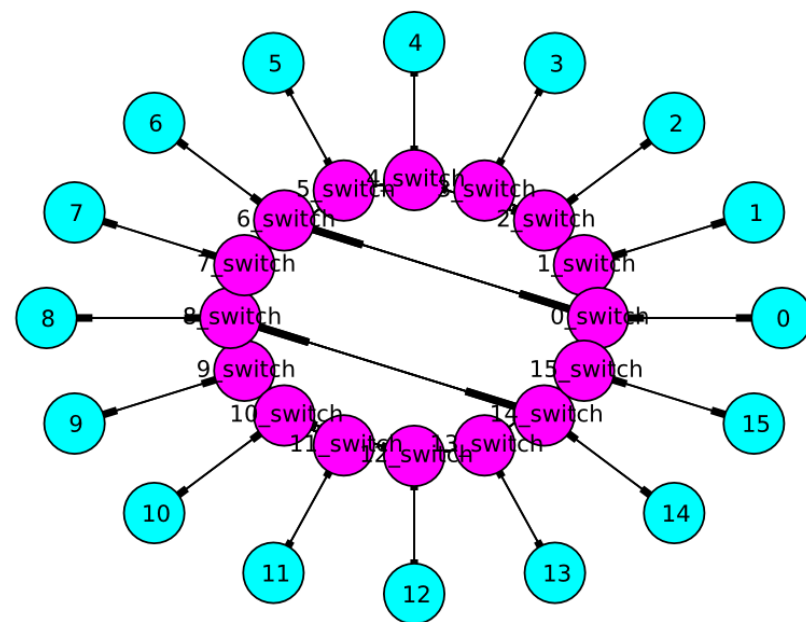


Comms & peripheral hardware

2-tile XMOS network with USB



16-tile XMOS board (UoB project Swallow)



Ongoing and future work

- Tying together with *Dynamic Voltage & Frequency Scaling*.
- Applying similar techniques to other architectures.
 - ARM, MIPS.
- Modelling peripherals.
- Higher-level modelling, not simulation based.

The research presented in this paper has received funding from the EU FP7 framework agreement 318337 – ENTRA, Whole-Systems Energy Transparency.
<http://entraproject.eu>

Thank you

Steve Kerrison, University of Bristol
steve.kerrison@bristol.ac.uk

Supervisor: Dr. Kerstin Eder
kerstin.eder@bristol.ac.uk

ICT-Energy Community Workshop, BSC