MANCHESTER 1824
The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

Konstantinos Iordanou, Oscar Palomar, John Mawer, Cosmin Gorgovan, Andy Nisbet and Mikel Lujan.

{first}.{last} @manchester.ac.uk

FCCM 2019
Tuesday, April 30

APT

FCCM

The University of Manchester

Agenda

- Motivation.
- How can we design fast and accurate simulators?
- Simulator Infrastructure.
- Simulation of accelerated applications.
- Performance & Evaluation.
- Conclusions & Future Work.

The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

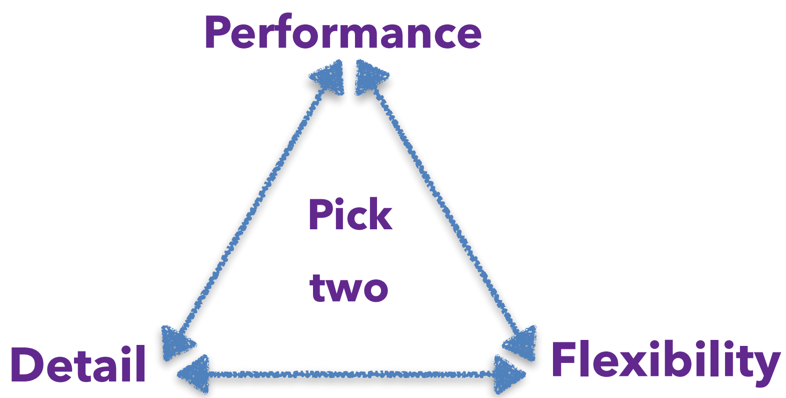
Motivation.

Motivation

In simulator's world, "good" and effective simulators are:

Fast, Accurate, Complete, Transparent, Inexpensive & Current.

However… many of the above properties conflict with each other.



Evaluate next generation processor and system architectures:

- software-based cycle-accurate simulators.
- These simulators are: **transparent, easy-to-use and can be cycle-accurate but are generally not fast or complete and often not current.**

MANCHESTER 1824
The University of Manchester

How can we design fast and accurate simulators?

- Proposing FPGA-based simulation which is based on the well known functional/timing model partitioning.
- This partition leverages two realisations:
  - functional model runs efficiently in parallel with the timing model.
  - the timing models because of the parallelism and silicon efficiency compared to a full implementation of an architecture can be implemented (easily) in a single FPGA.

MANCHESTER 1824
The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

Simulator Infrastructure.

Simulator Infrastructure

- The simulator consists of the following main components:
  - **Dynamic Binary Modification Tool (MAMBO).**
    - Scans the application and copies it to a software code cache.
    - It transforms the code to maintain correctness & control.
    - Doing other modifications (plugins via an API).
  - **MAST library.**
  - **Hardware Timing Models.**
    - Out of order pipeline Model (arm Cortex A9).
    - Cache Hierarchy Models.
      - Data cache, icache, L2 cache, Snoop Module.

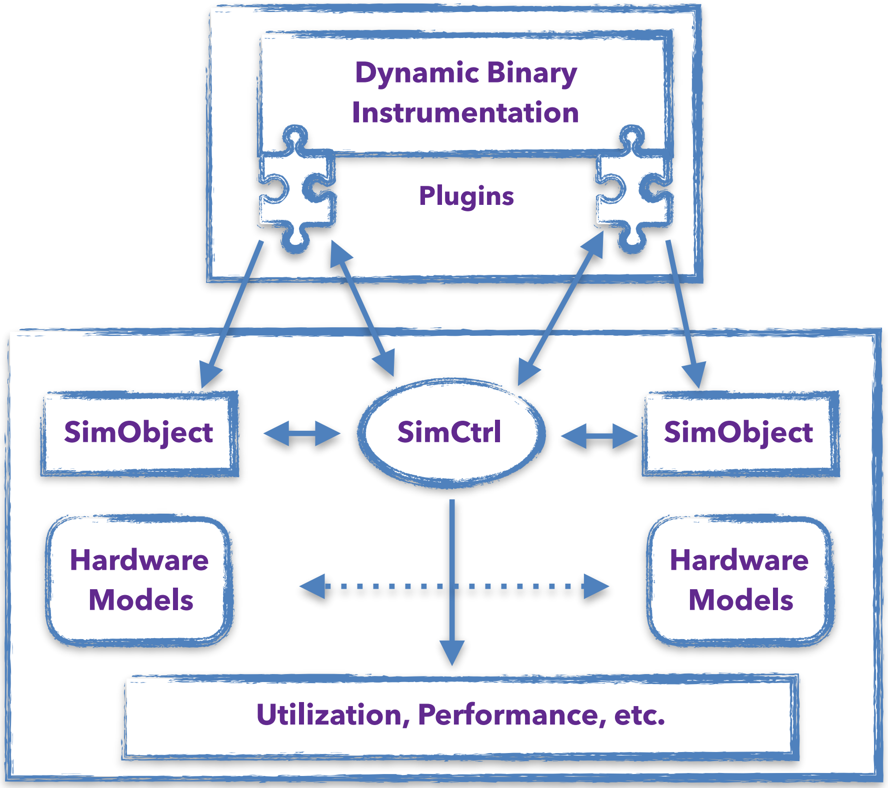
The University of Manchester

Simulator Infrastructure - MAST

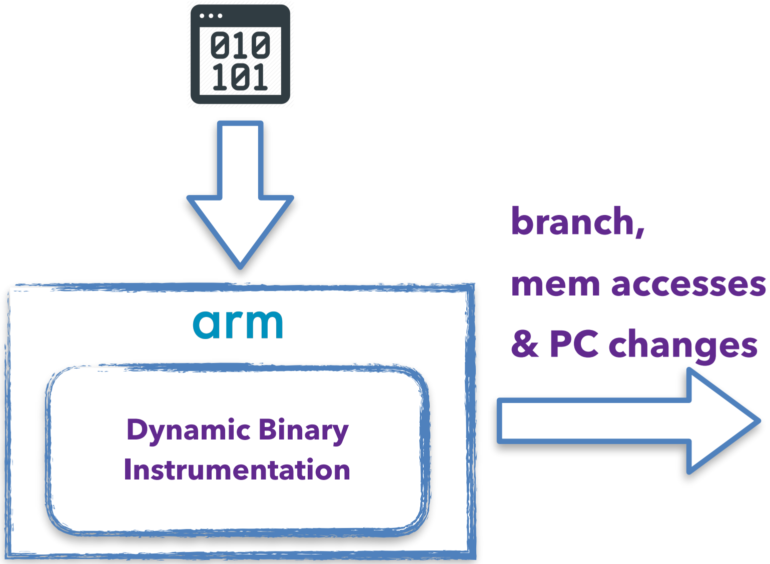
- A C++ library and hardware interface standard/ Bluespec library which:
  - Allows discovery of MAST compliant IP on an FPGA.
  - Allows management of compliant hardware e.g. locking blocks, reconfiguration of FPGA.
  - Provides a userSpace interface to hardware.
  - Enables IP accelerator blocks to be easily and efficiently integrated into applications.

Simulator Infrastructure - MAST

- **SimCtrl: manages the system for an application.**
  - Discovers hardware.
  - Allows allocation/locking of a hardware to a process (or a thread).
  - Handles memory management.
  - Requirements of hardware e.g virtual -> physical lookups.
- **SimObject: is an interface to a hardware IP.**
  - Handles all interactions between application and IP.
  - Maintains information on system RAM used by hardware. (e.g. physical addresses)



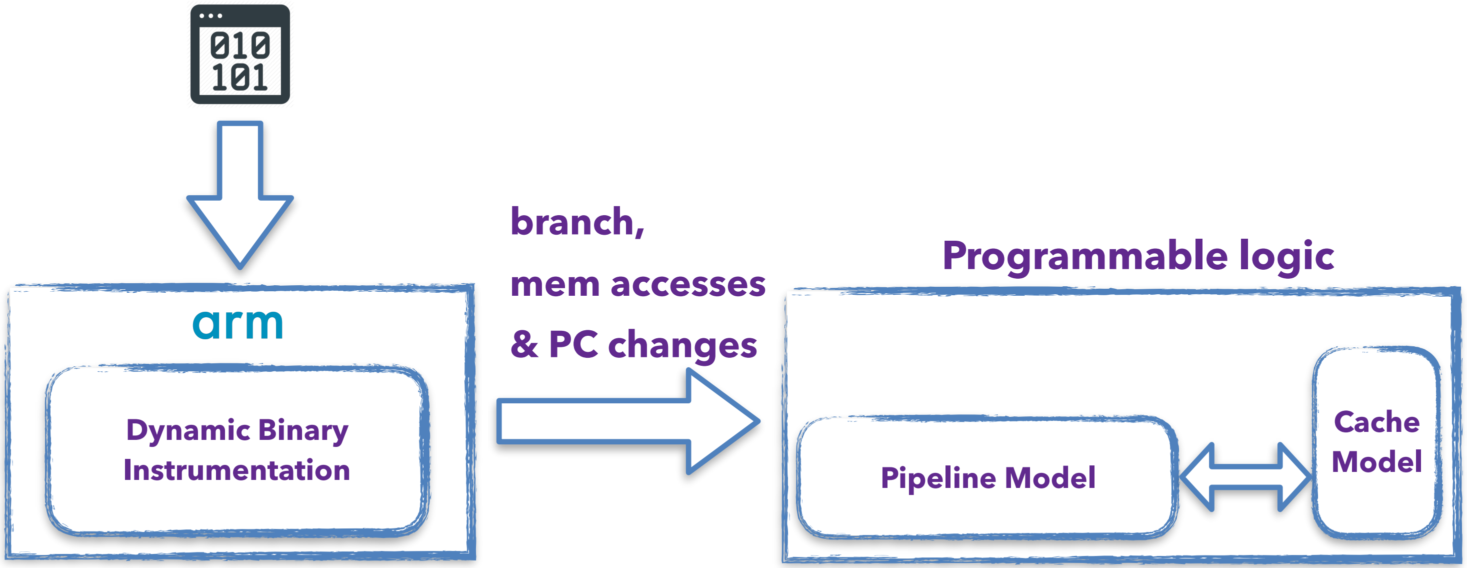
Simulator Infrastructure



- An efficient dynamic binary modification tool for arm architectures.
- Modifies the machine code of 32 bit and 16 bit instructions during execution.
- MAMBO plugins: driving our hardware models.
- MAMBO plugins: consist of a set of callbacks which are executed at various points of program execution.

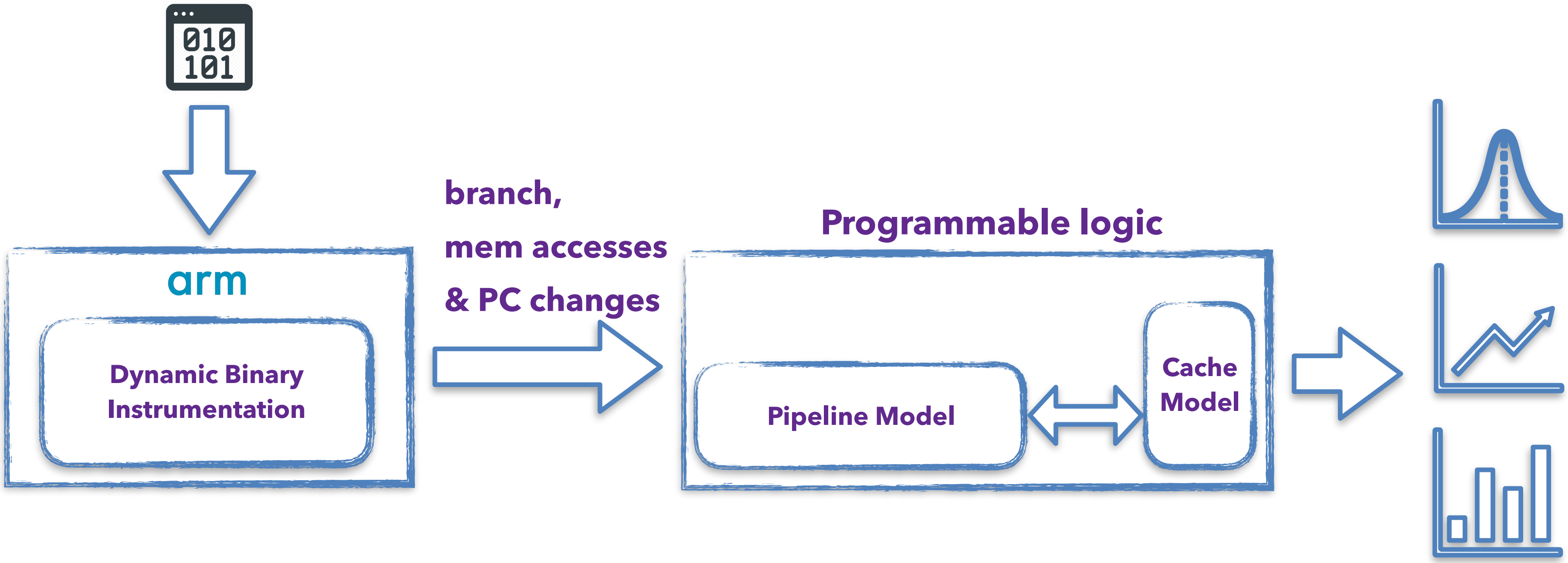
The University of Manchester

Simulator Infrastructure



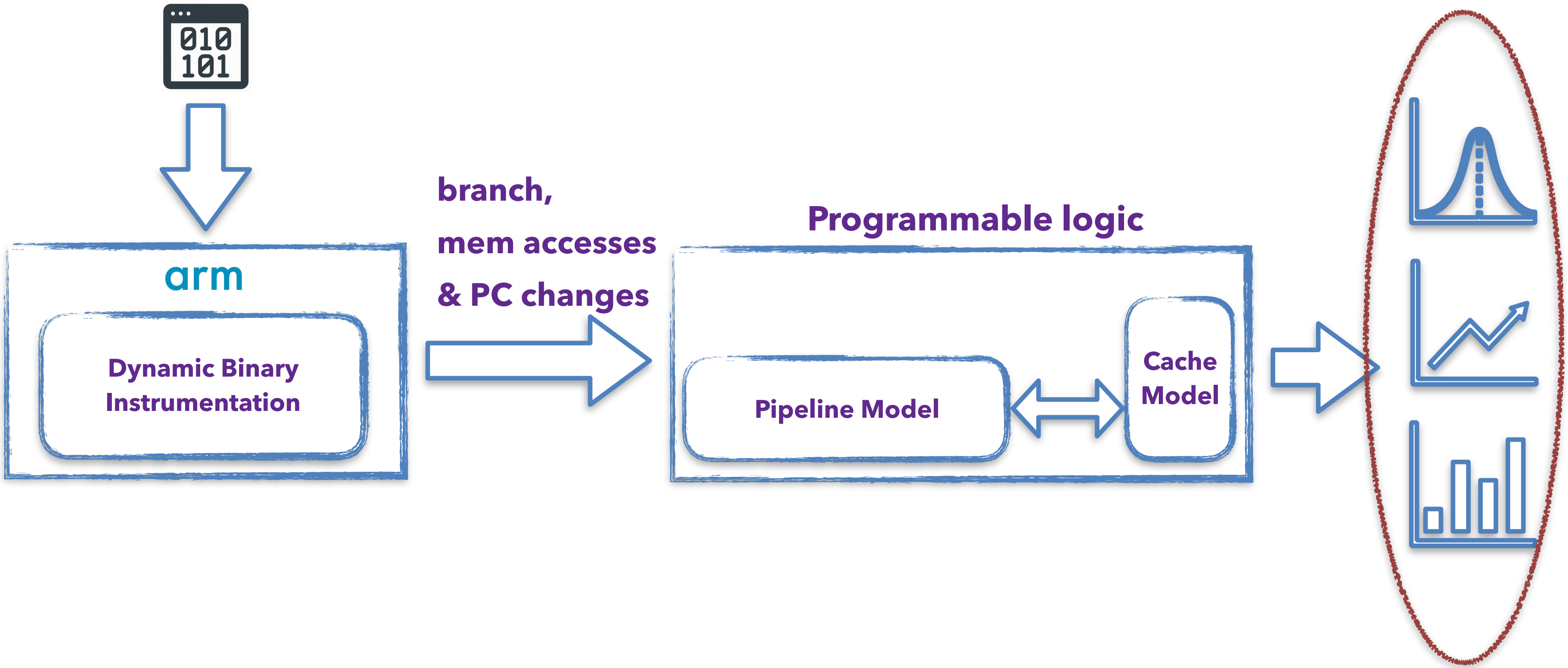
- Pipeline Model (inspired by arm Cortex-A9 dual-issue processor)
  - re-order buffer, register renaming module, branch prediction models, implementations of Branch Target Buffer (BTB), return address stack (RAS), and the load/store queue supports memory aliasing.
- Cache System Model:
  - It gathers statistics about the behaviour of a cache system. Only address tags and states for cache lines.

Simulator Infrastructure

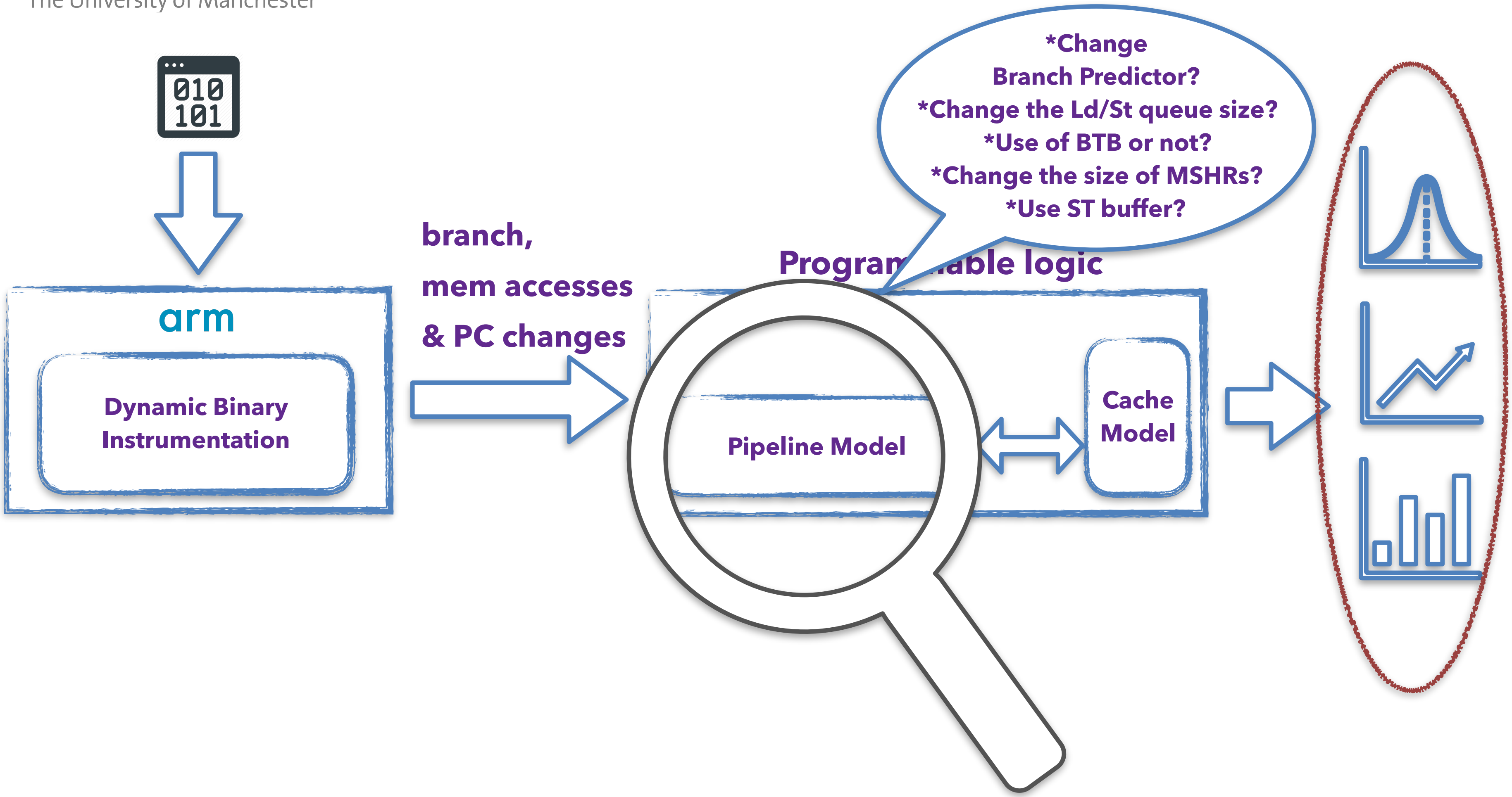


- Every model includes registers which count events like read/write hits or misses per cache level, instruction counters, simulated cycles, branch hits and misses.

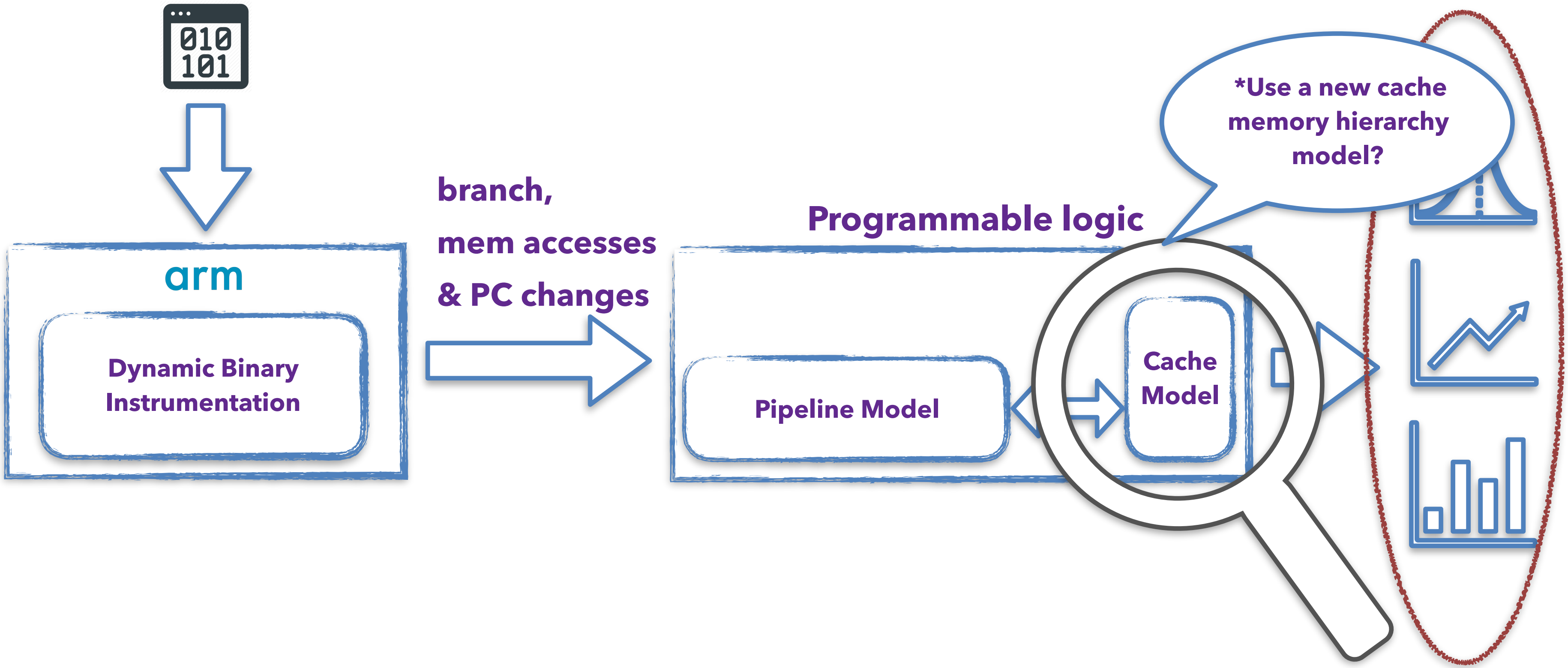
Simulator Infrastructure



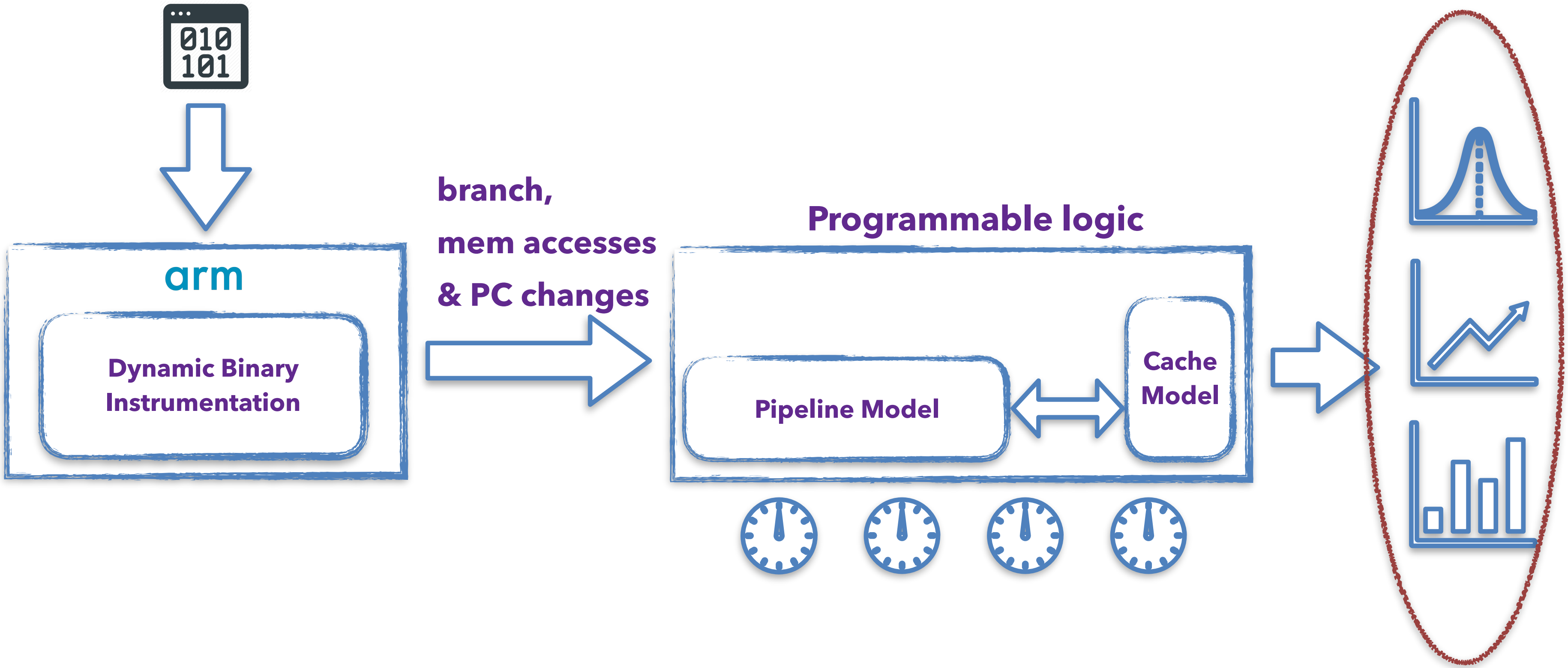
Simulator Infrastructure



Simulator Infrastructure



Simulator Infrastructure



The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

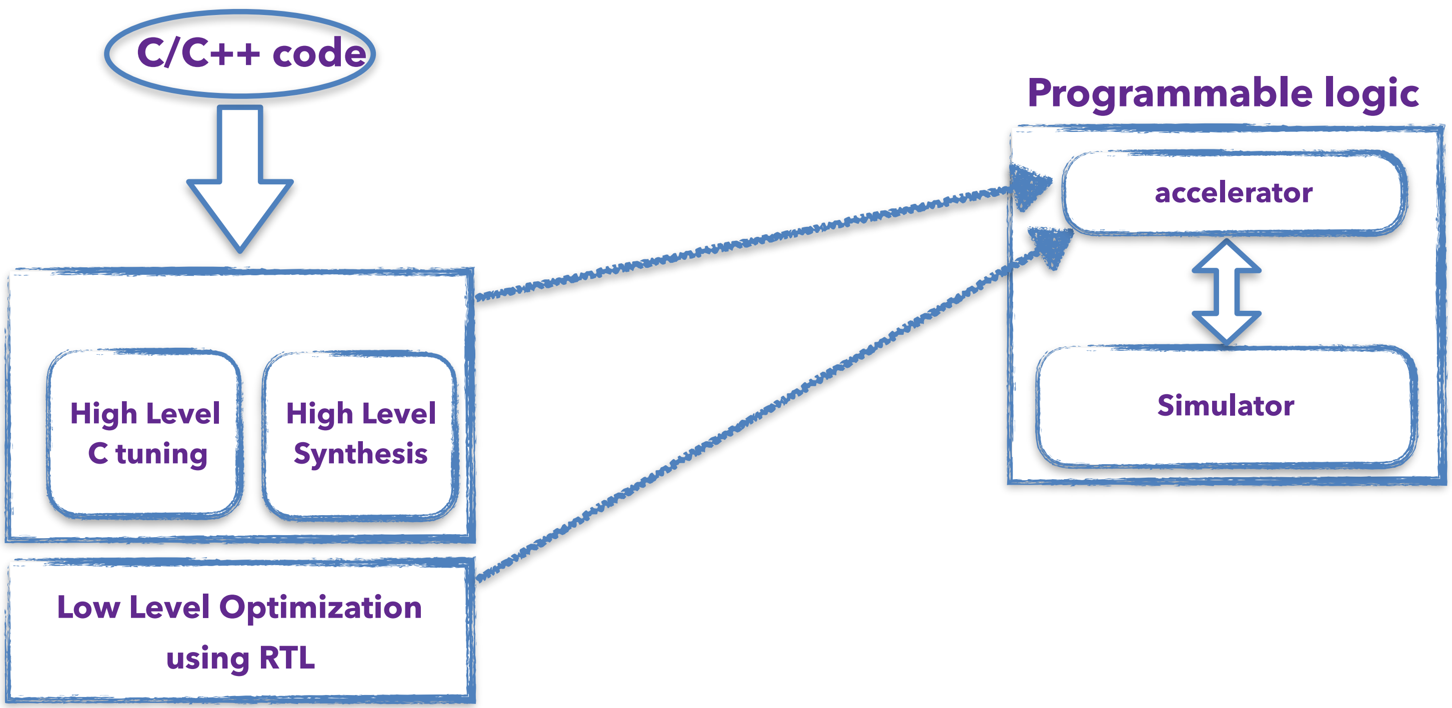
Simulation of accelerated applications.

The University of Manchester

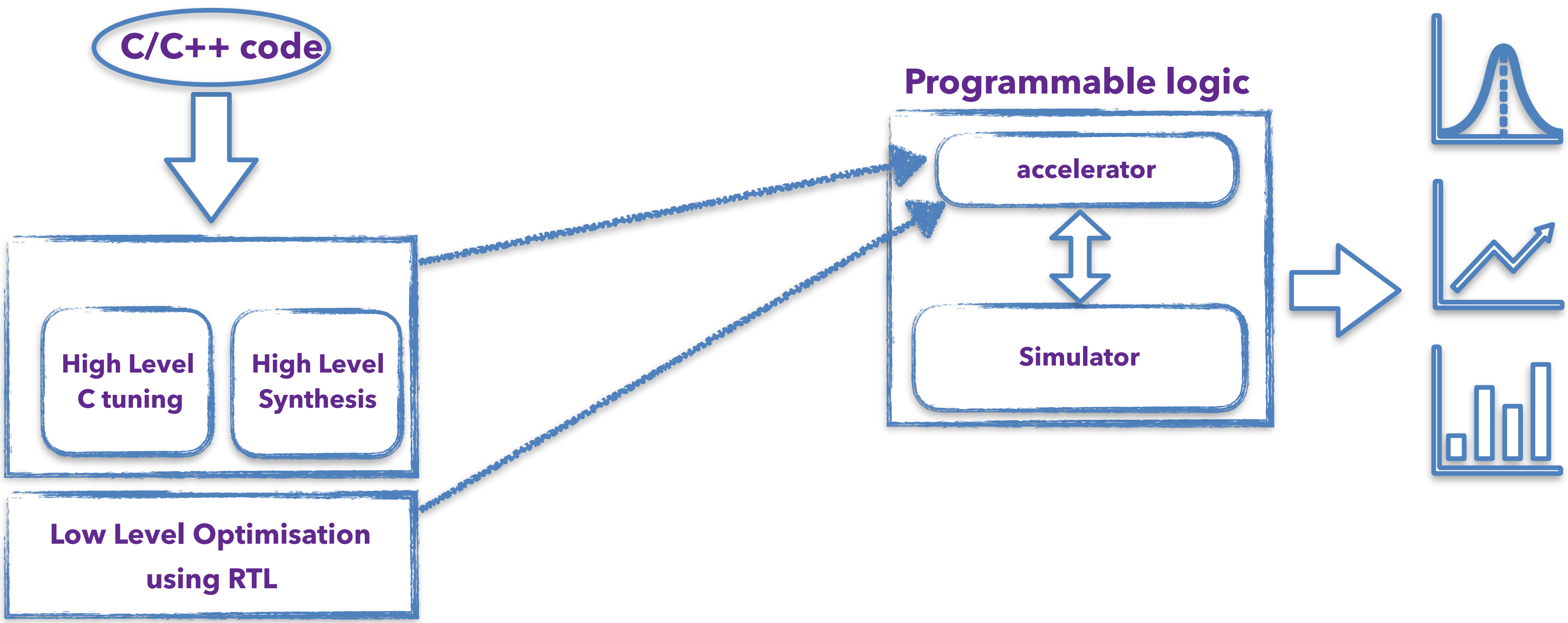
Simulation of accelerated applications

- A methodology for including in the simulation the interaction of processors and accelerators.
- Using unmodified IP blocks, which are wrapped with logic to capture the memory accesses they initiate and simulate them.
- Implementing conventional FPGA-based accelerators, such as image processing filters, and access these from regular applications (Bluespec System Verilog or HLS).

Simulation of accelerated applications

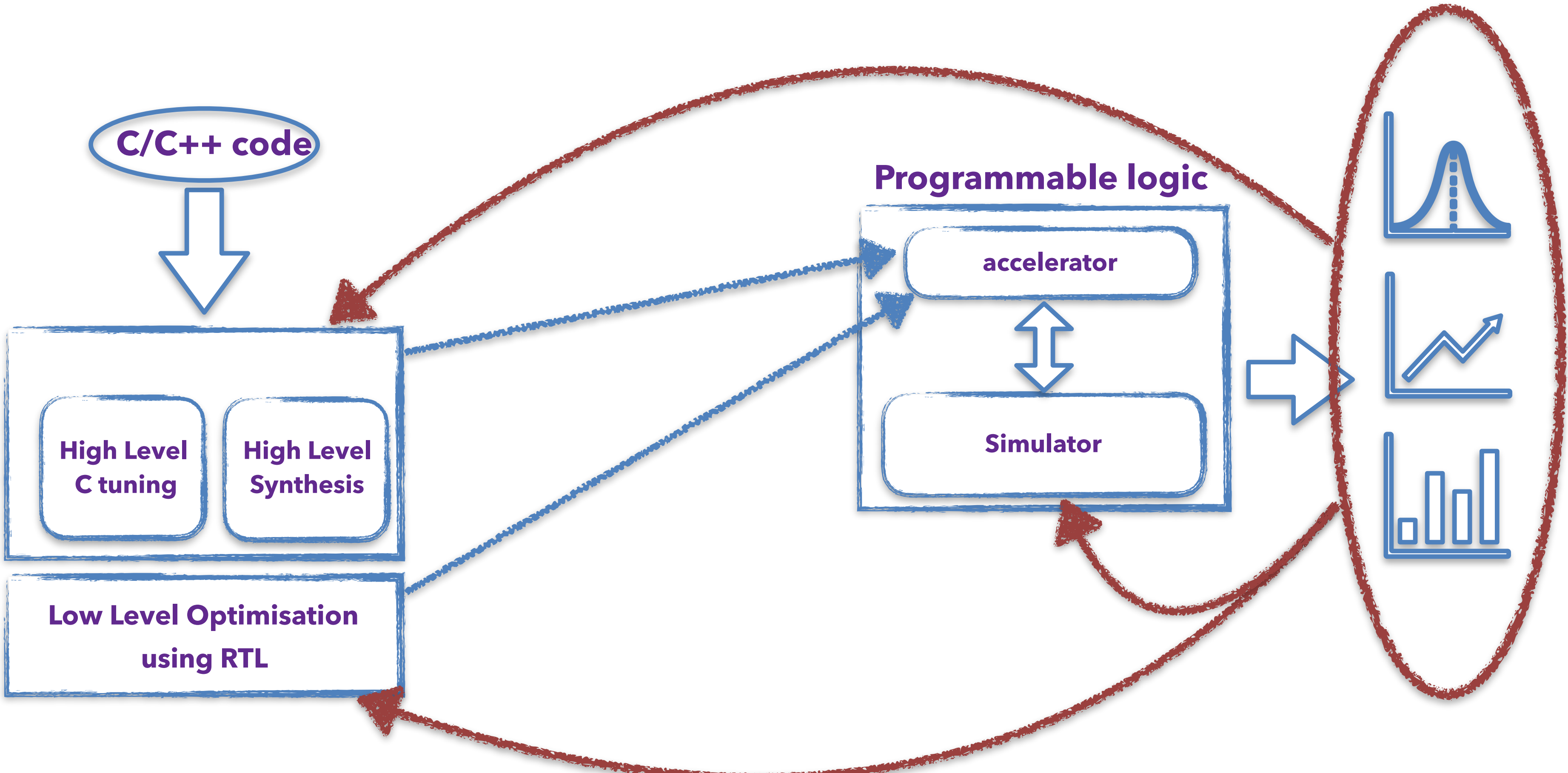


Simulation of accelerated applications



MANCHESTER 1824
The University of Manchester

Simulation of accelerated applications



APT

Simulation of accelerated applications

- **ACP Snoop:**

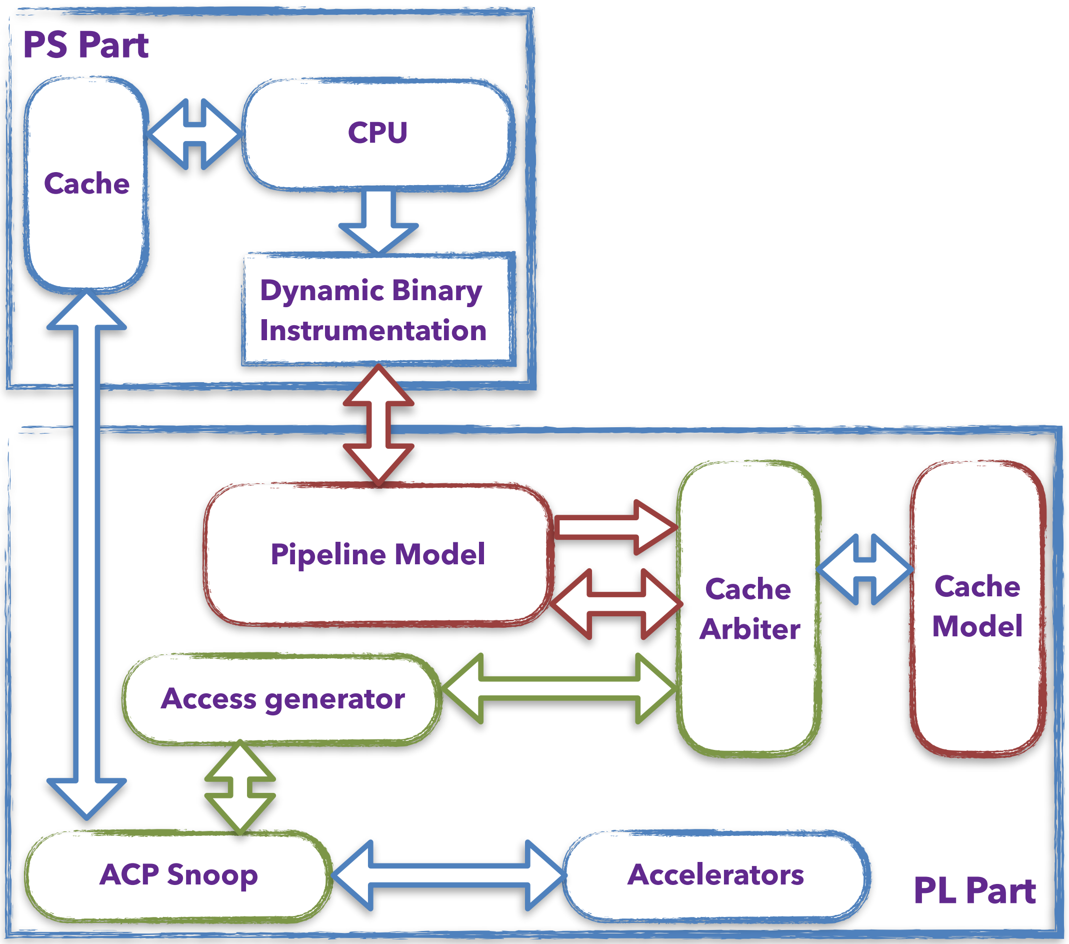
It intercepts the ACP accesses, creates a descriptor for them and sends the original request to the actual ACP port.

- **Access generator:**

It is responsible of retrieving descriptors from the ACP Snoop, and break them into individual memory accesses to send to the cache model.

- **Cache Arbiter:**

It handles requests from the pipeline model and from the accelerator and send them to the Cache Hierarchy model



The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

Performance & Evaluation.

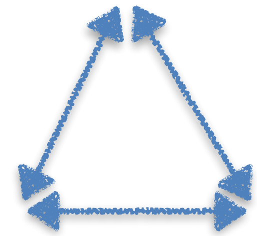
The University of Manchester

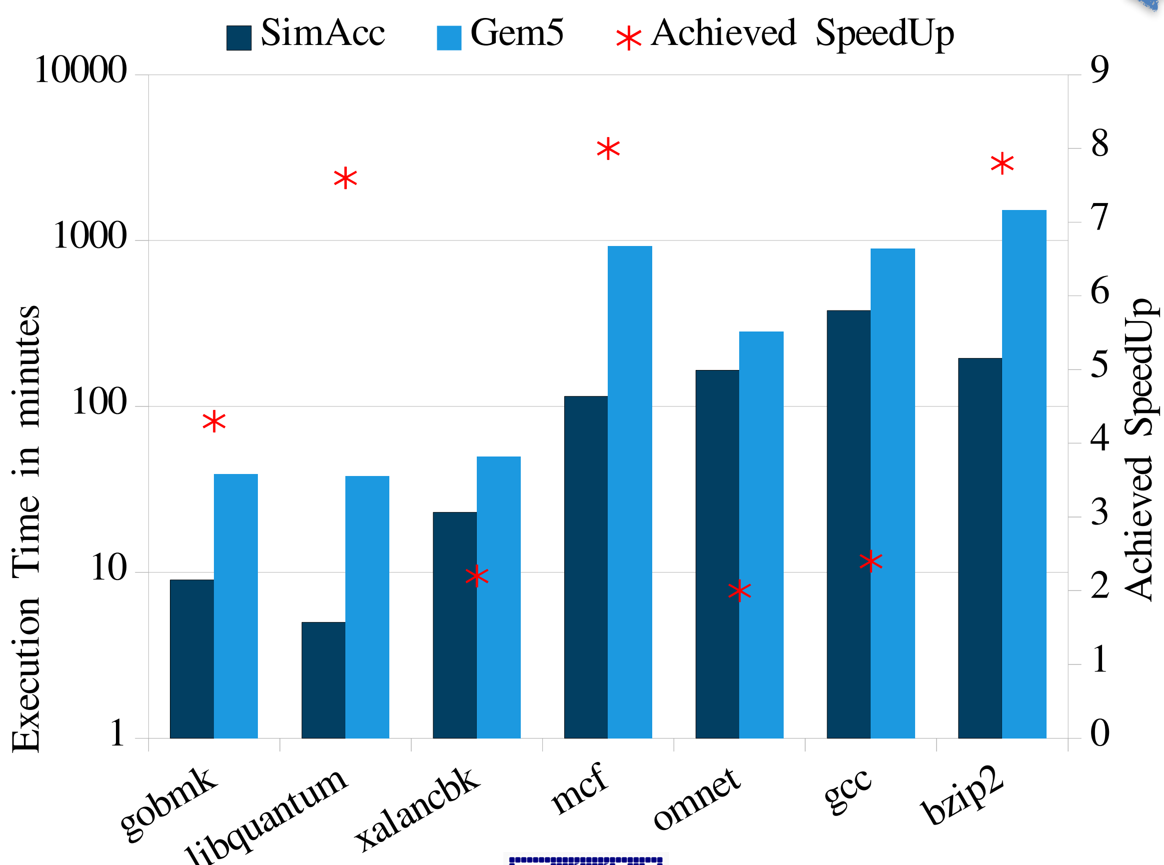
Experimental Evaluation

- SimAcc uses:
  - Xilinx Zynq-7000 XC7Z045 evaluation board running Ubuntu 14.04 with 1GB DRAM (no swap), and dual 667MHz arm Cortex-A9 processors.
- For the experimental evaluation, we are using gem5:
  - DerivO3 CPU model.
- Benchmarks:
  - SPEC CPU 2006.
  - Mach Benchmark Suite, Computer Vision Applications.

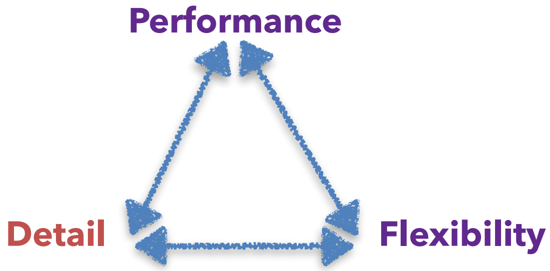
MANCHESTER 1824
The University of Manchester

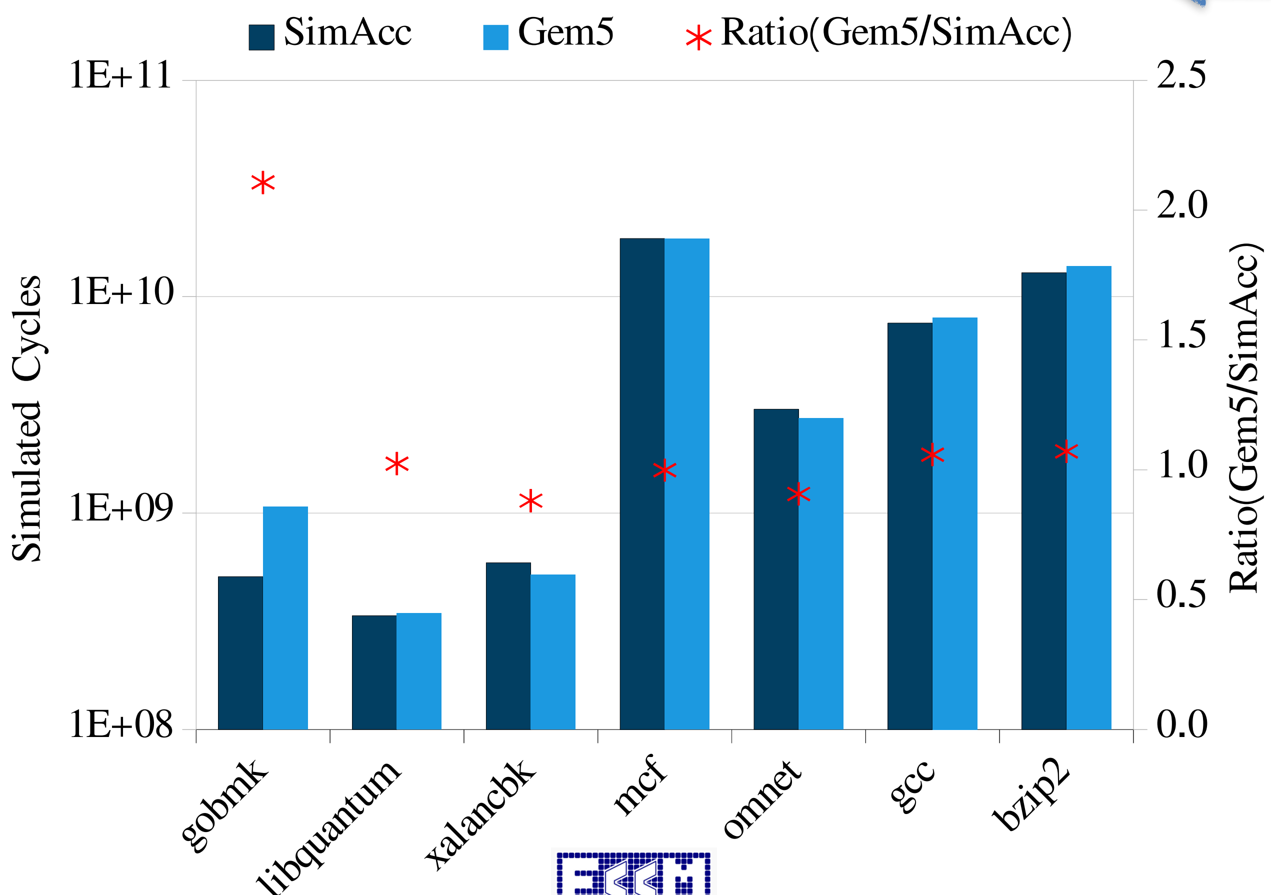
Simulator Results - Execution Time





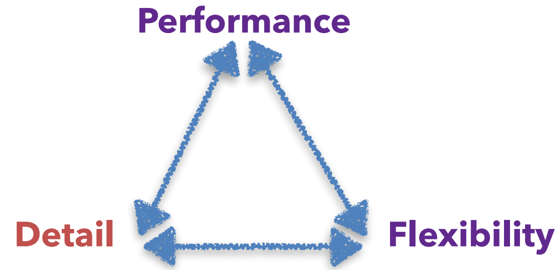
Simulator Results - Simulated Cycles

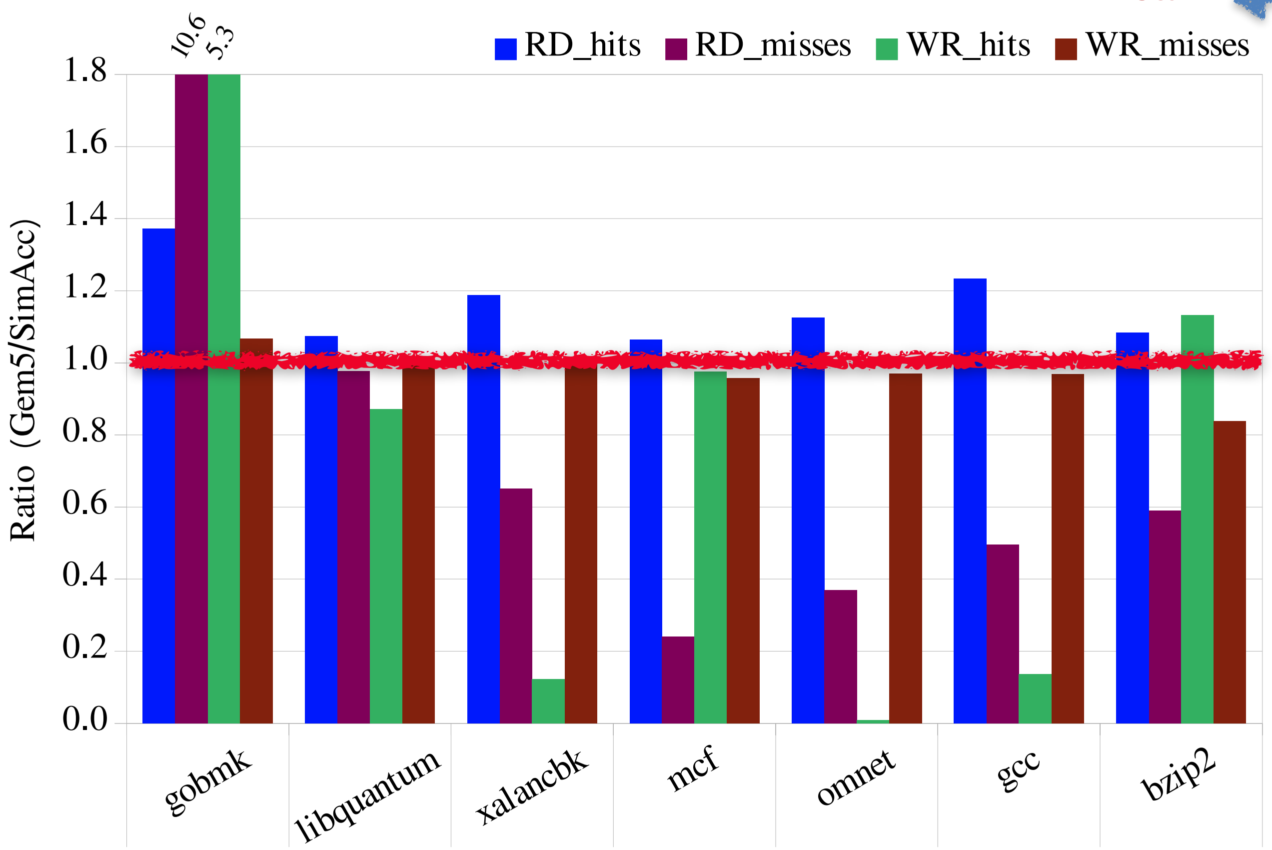




Simulator Results - L2 cache

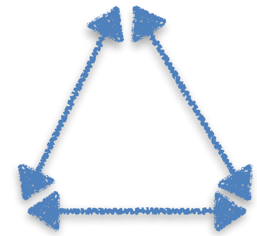
The University of Manchester

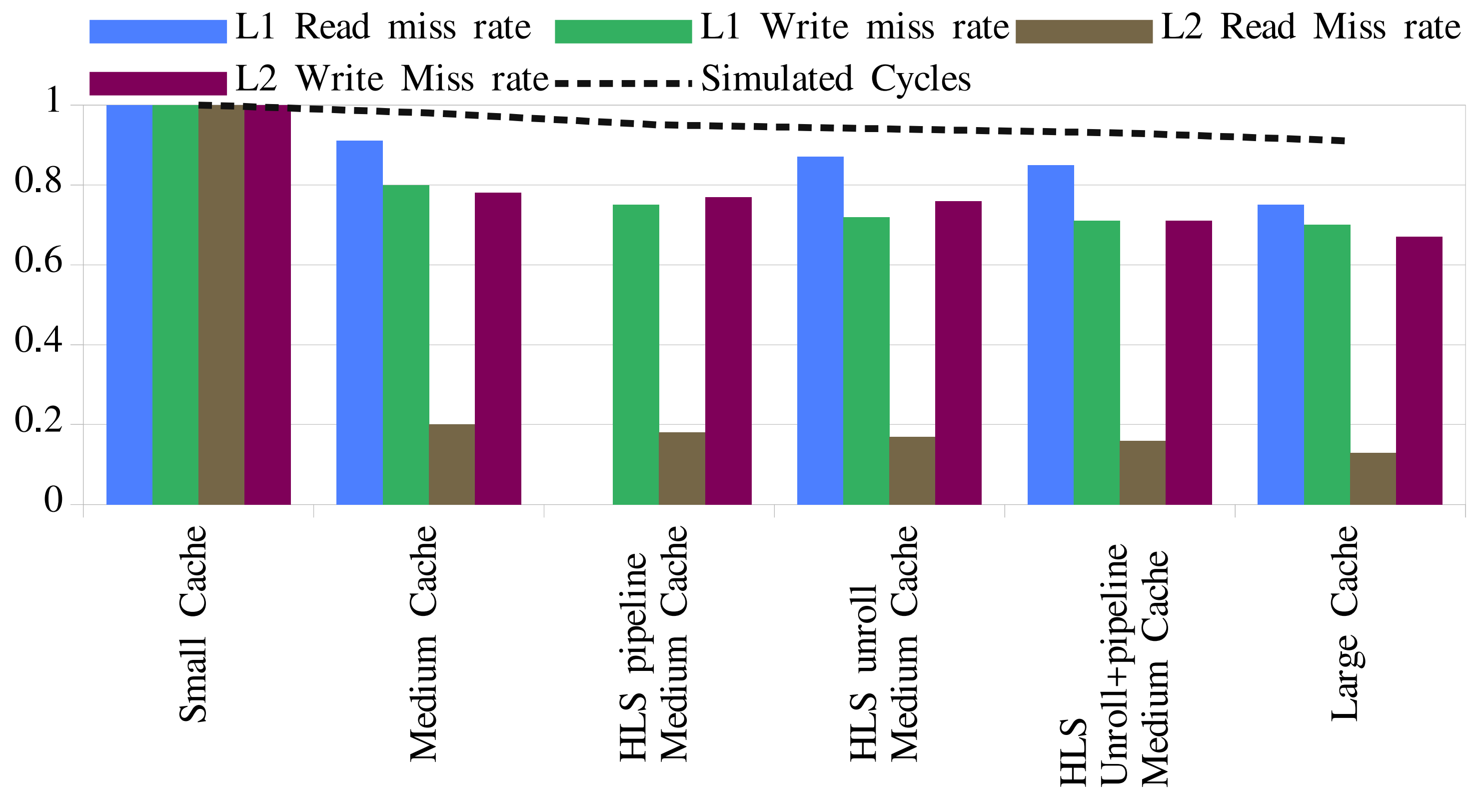




MANCHESTER 1824
The University of Manchester

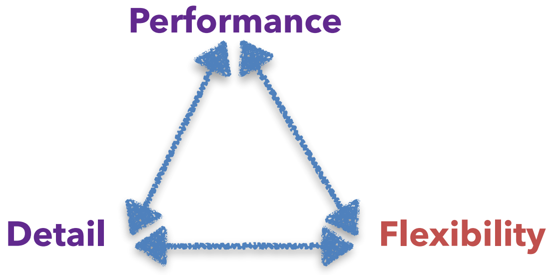
SimAcc Results - SVO (HLS)

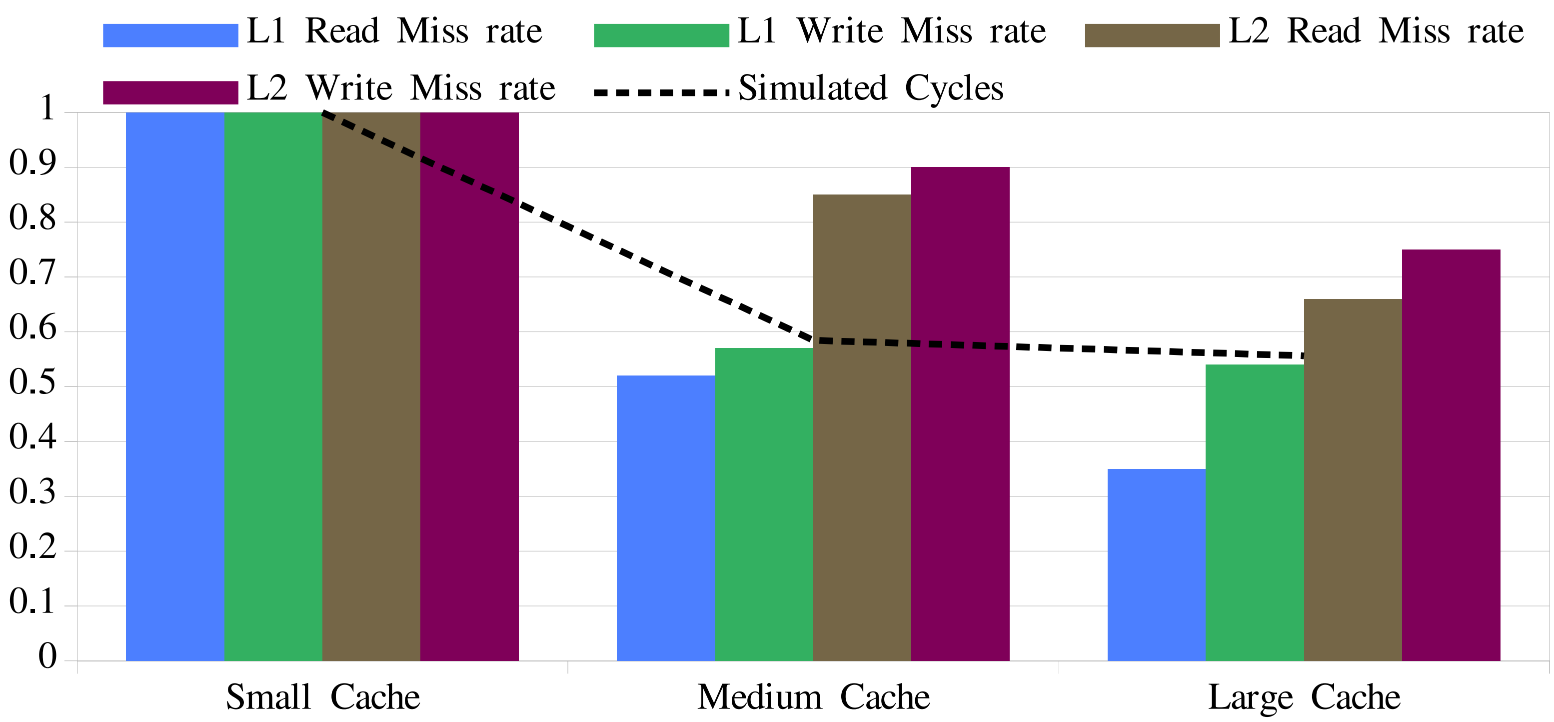




APT

SimAcc Results - ORB-slam (Bluespec SystemVerilog)





The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

Conclusions & Future Work.

Conclusions & Future Work

- **Conclusions:**
  - We have demonstrated the potential of combining a flexible IP hardware library, a user-level driver library and dynamic binary instrumentation for microarchitecture simulation and prototyping.
  - We exploit the advantages of an FPGA SoC to accelerate at a very fine granularity (instructions).
  - We can benefit from the accuracy and speed of FPGA-based modelling and the ability to run binaries.
  - It is the first FPGA-based simulator for arm combined with accelerators, significantly extending the options for simulating arm processors.
- **Future work:**
  - Use of an Ultrascale+ Zynq board. (e.g. faster CPUs, improved memory interface)
  - Extend the models to include more micro-architectural features and alternatives.
  - Multi-core version of the Simulator Infrastructure.

MANCHESTER 1824
The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

Thank you.

The University of Manchester

SimAcc: A Configurable Cycle-Accurate Simulator for customized accelerators on CPU-FPGAs SoCs

BackUp Slides

Simulator Results - data & L2 cache