Profiling and controlling I/O-related memory contention in COTS heterogeneous platforms

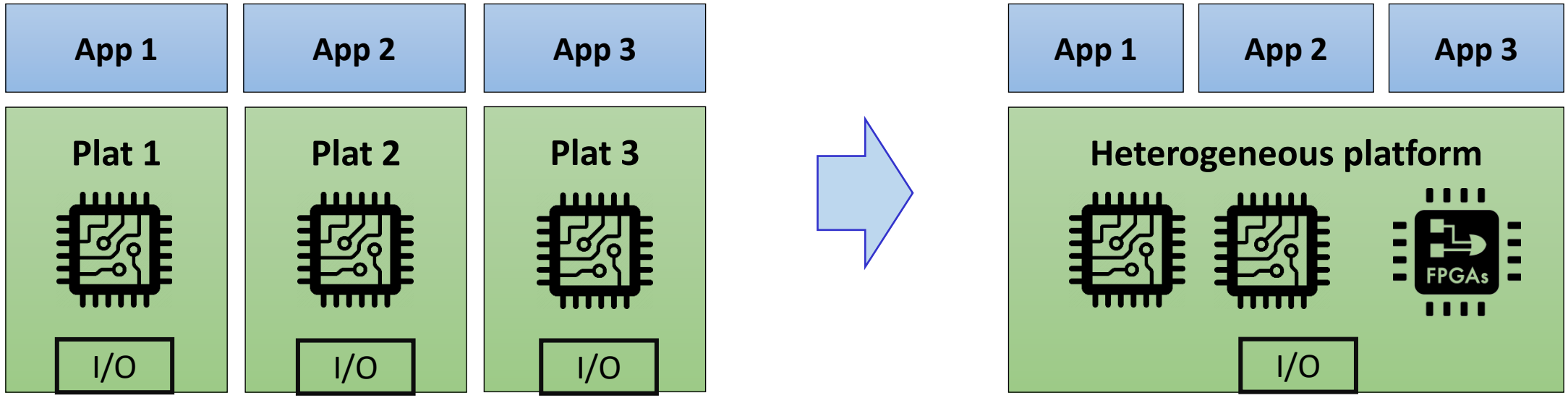
Matteo Zini, Giorgiomaria Cicero, Daniel Casini, Alessandro Biondi

Scuola Superiore Sant'Anna, Pisa, Italy

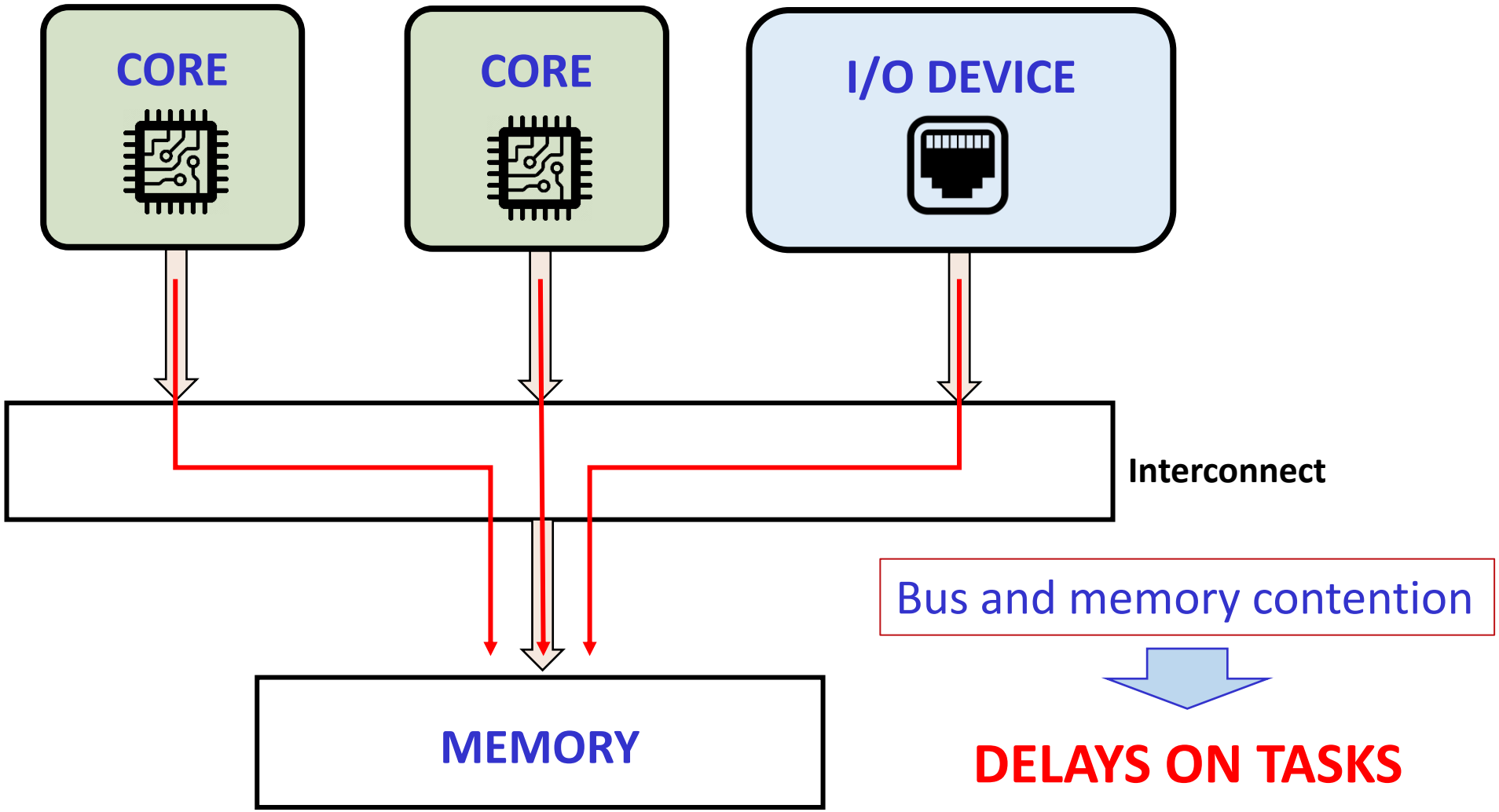
Sant'Anna
Scuola Universitaria Superiore Pisa

Introduction

- Increasing computational requirements of modern applications
- Next-generation embedded systems will move from single-processor to **multiprocessor and heterogeneous platforms**
- Need to design mechanisms to **isolate** single applications



Memory contention



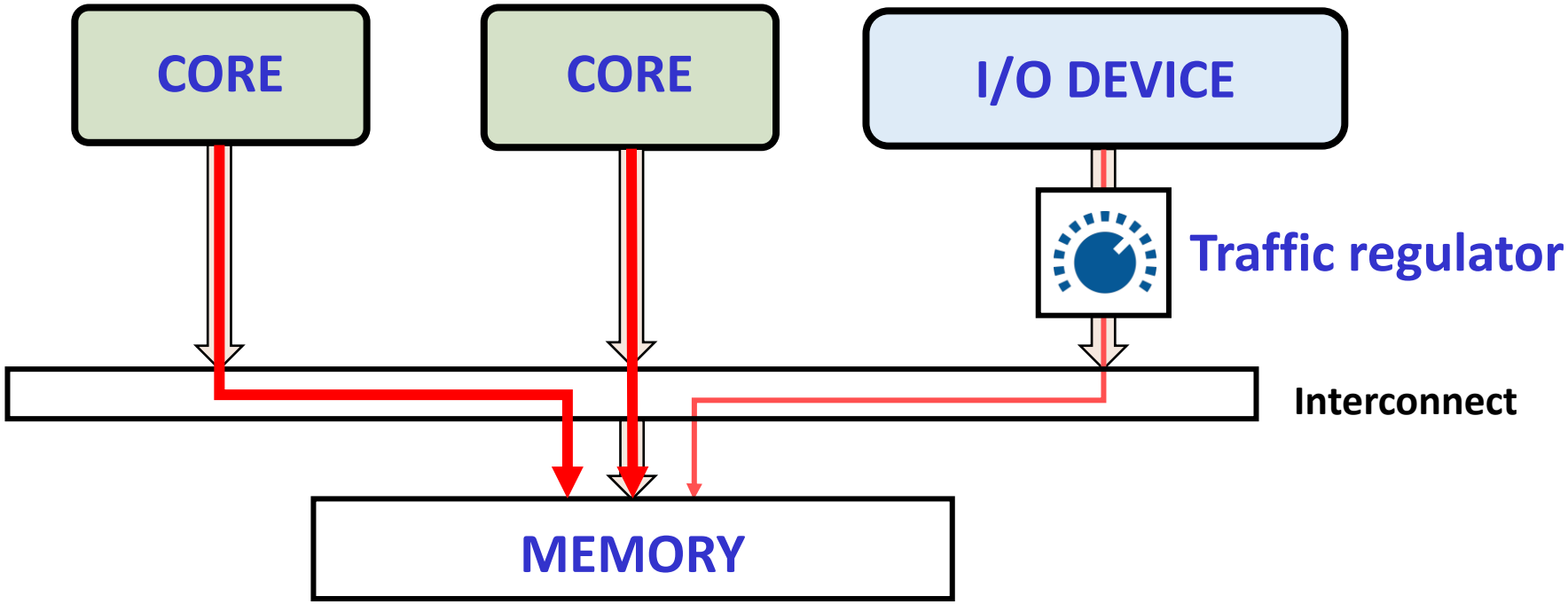
Bus and memory contention

DELAYS ON TASKS

This Paper

Zini M, Cicero G, Casini D, Biondi A. Profiling and controlling I/O-related memory contention in COTS heterogeneous platforms. *Softw Pract Exper*. 2021;1-19. doi: 10.1002/spe.3053

OBJECTIVE: evaluate the effect of the memory interference generated by I/O devices and its regulation on tasks and on I/O devices themselves



This Paper

Zini M, Cicero G, Casini D, Biondi A. Profiling and controlling I/O-related memory contention in COTS heterogeneous platforms. *Softw Pract Exper*. 2021;1-19. doi: 10.1002/spe.3053

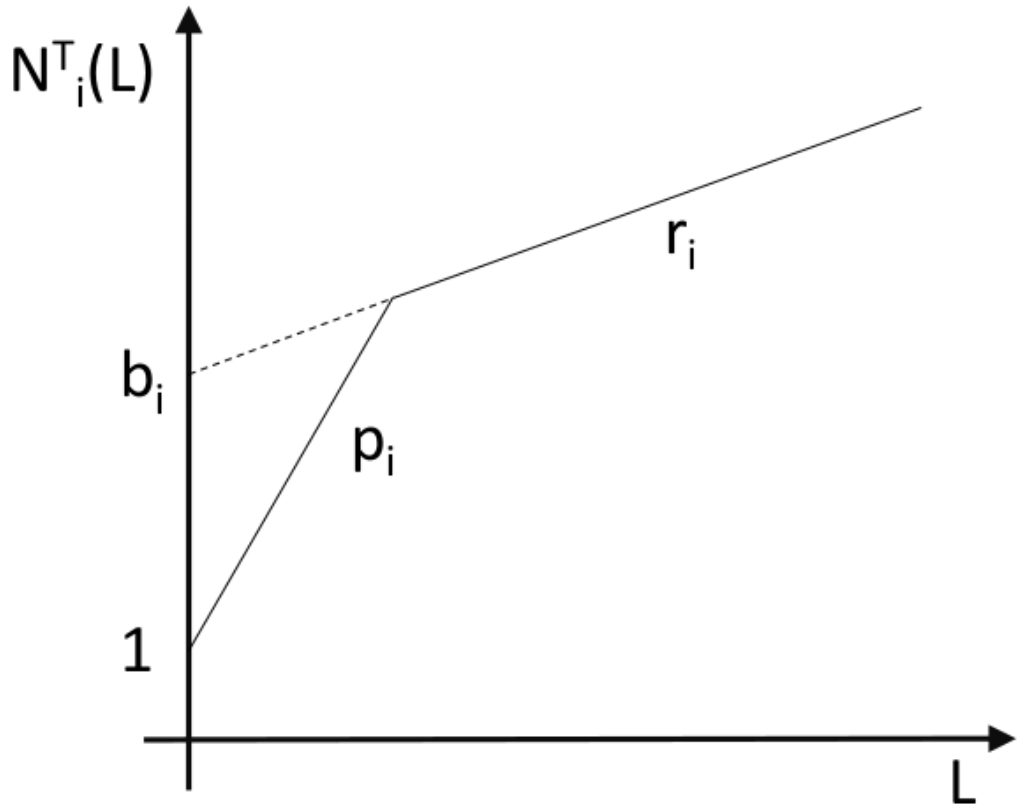
OBJECTIVE: evaluate the effect of the memory interference generated by I/O devices and its regulation on tasks and on I/O devices themselves

- Derive a mathematical model of the regulation
- Build a software infrastructure to enable the regulation from Linux
- Evaluation of the effectiveness of the regulation using benchmarks

ARM QoS-400 regulators

- The ARM **QoS-400** traffic regulators can limit the traffic that is traversing them. Based on a variant of TSPEC specifications (RFC 2215).
- The Xilinx's **Ultrascale+** MPSoC has a dedicated regulator for almost every device in the system.
- **3 operating modes**:
  - Outstanding transaction regulation
  - Transaction latency regulation
  - Transaction rate regulation

QoS-400 parameters

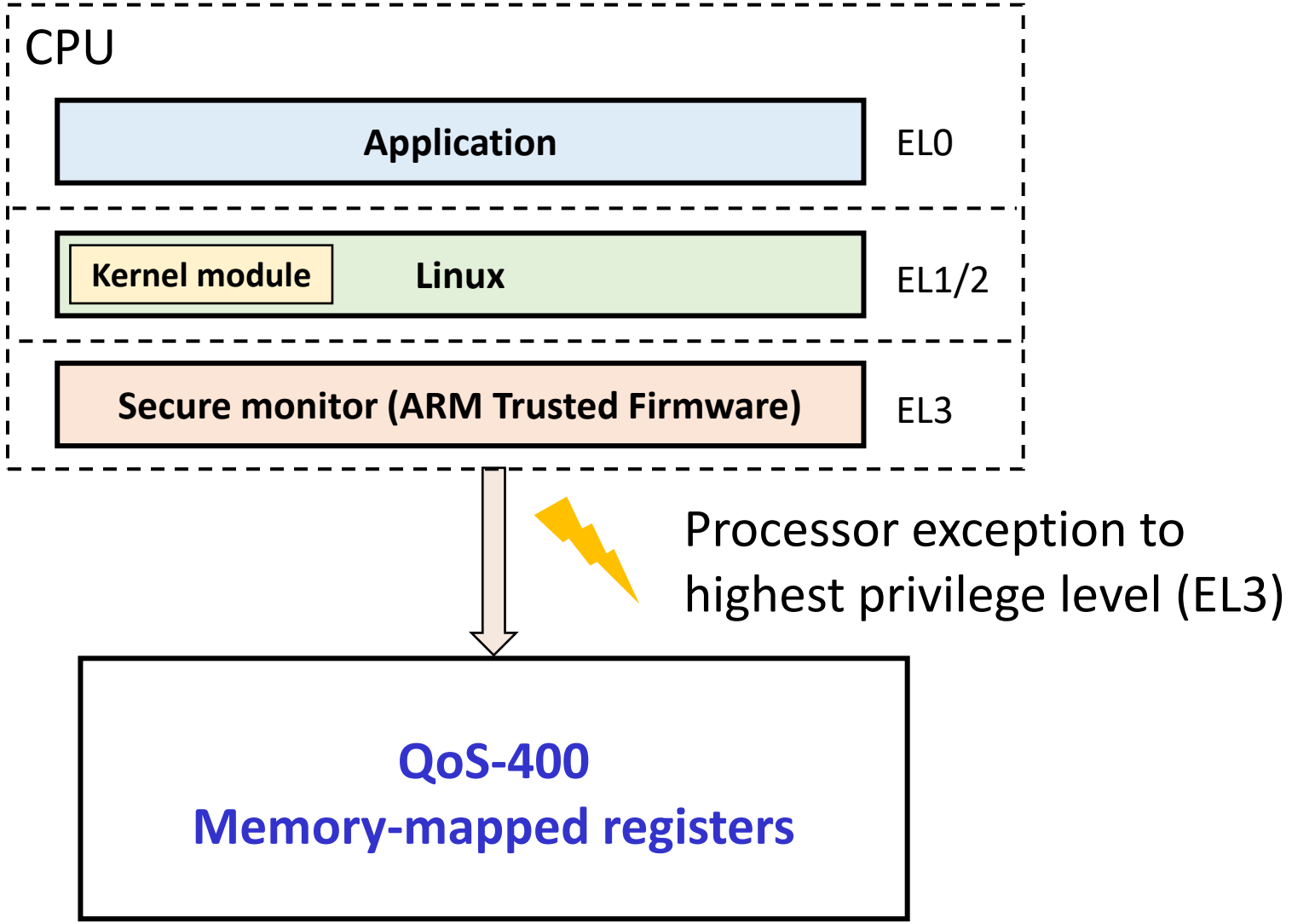


- **3 control parameters**:
  - r_i → Average-rate
  - p_i → Peak-rate
  - b_i → Burstiness allowance

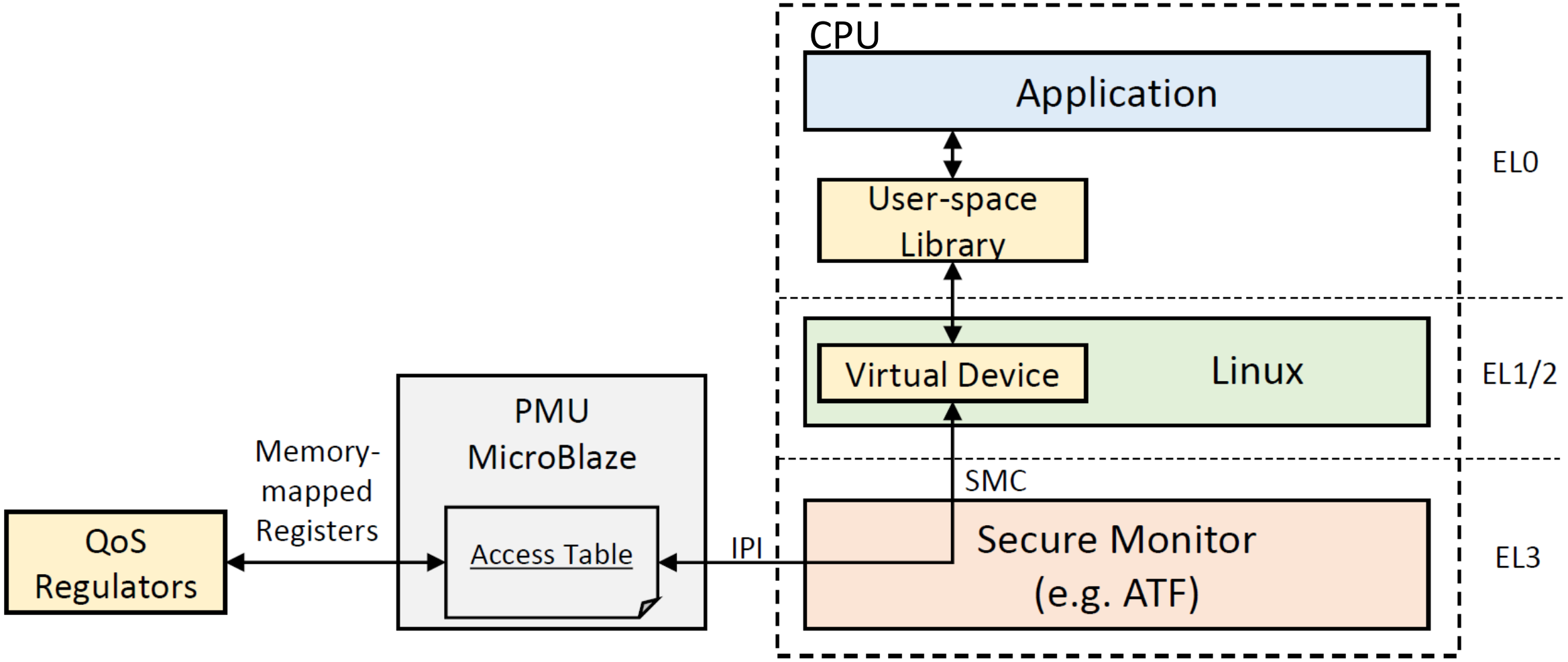
$$N_i^T(L) = \min\{1 + p_i \cdot L,\ b_i + r_i \cdot L\}$$

QoS-400 configuration

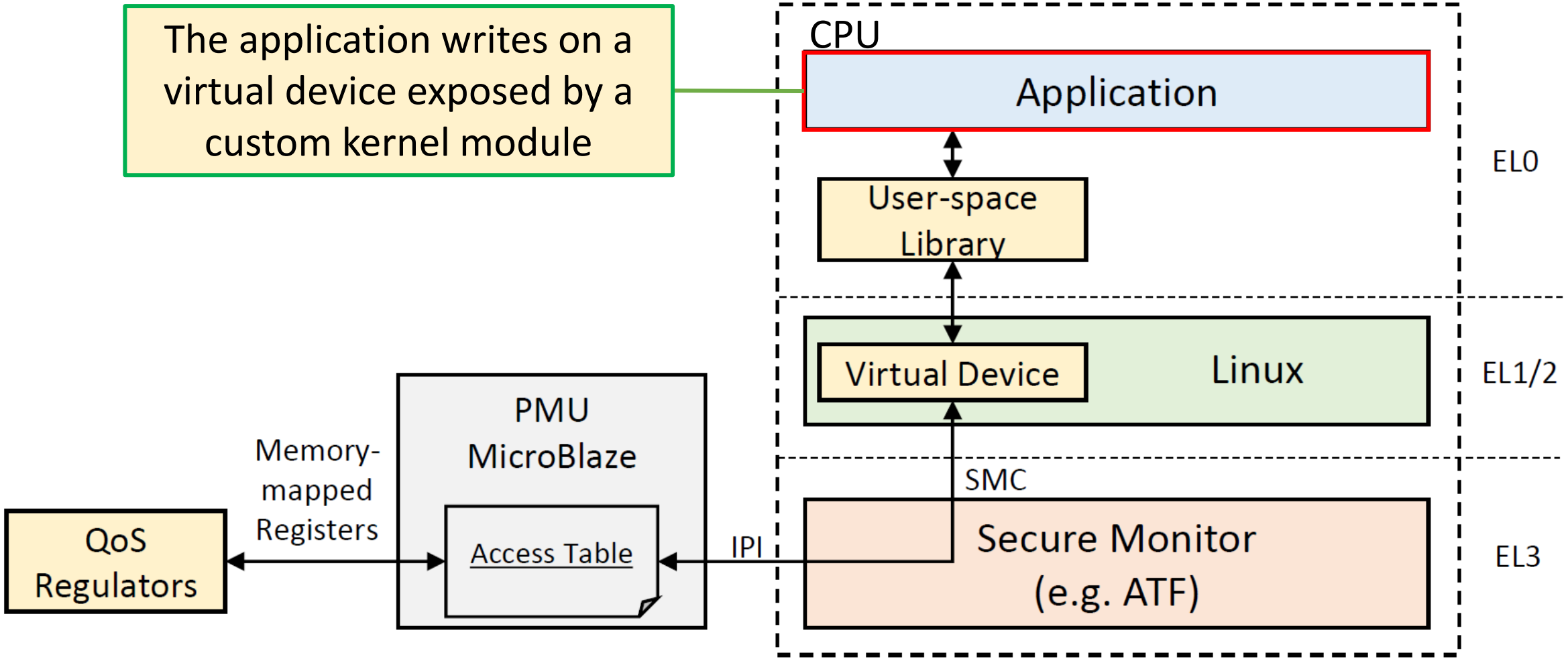
QoS-400 configuration



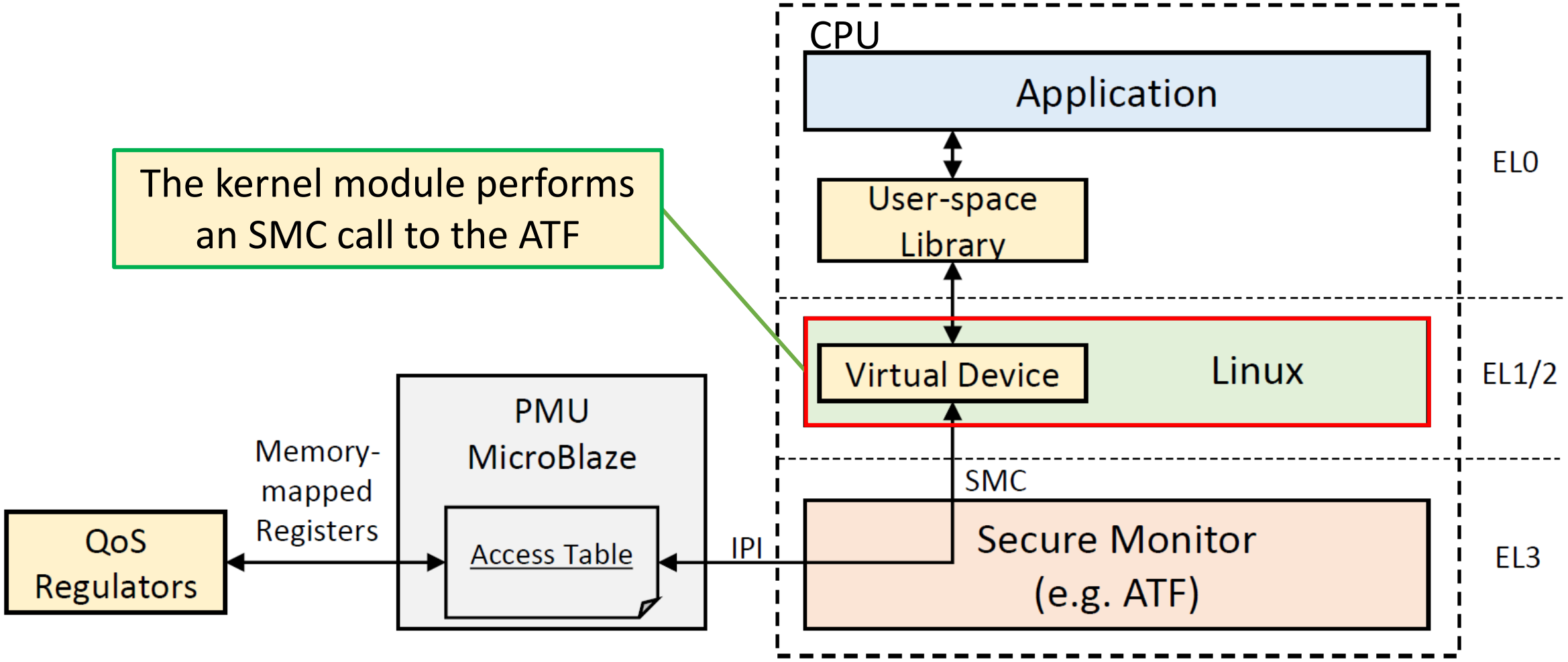
QoS-400 configuration



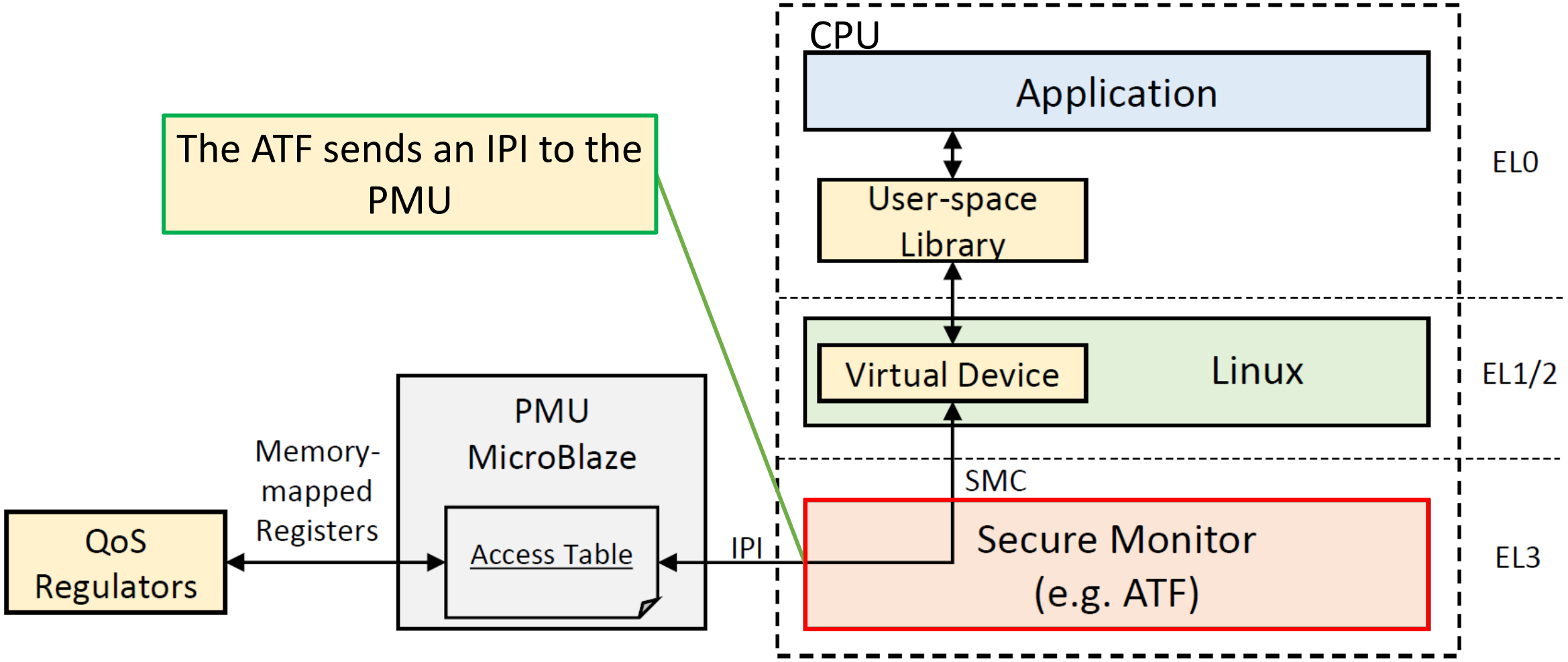
QoS-400 configuration



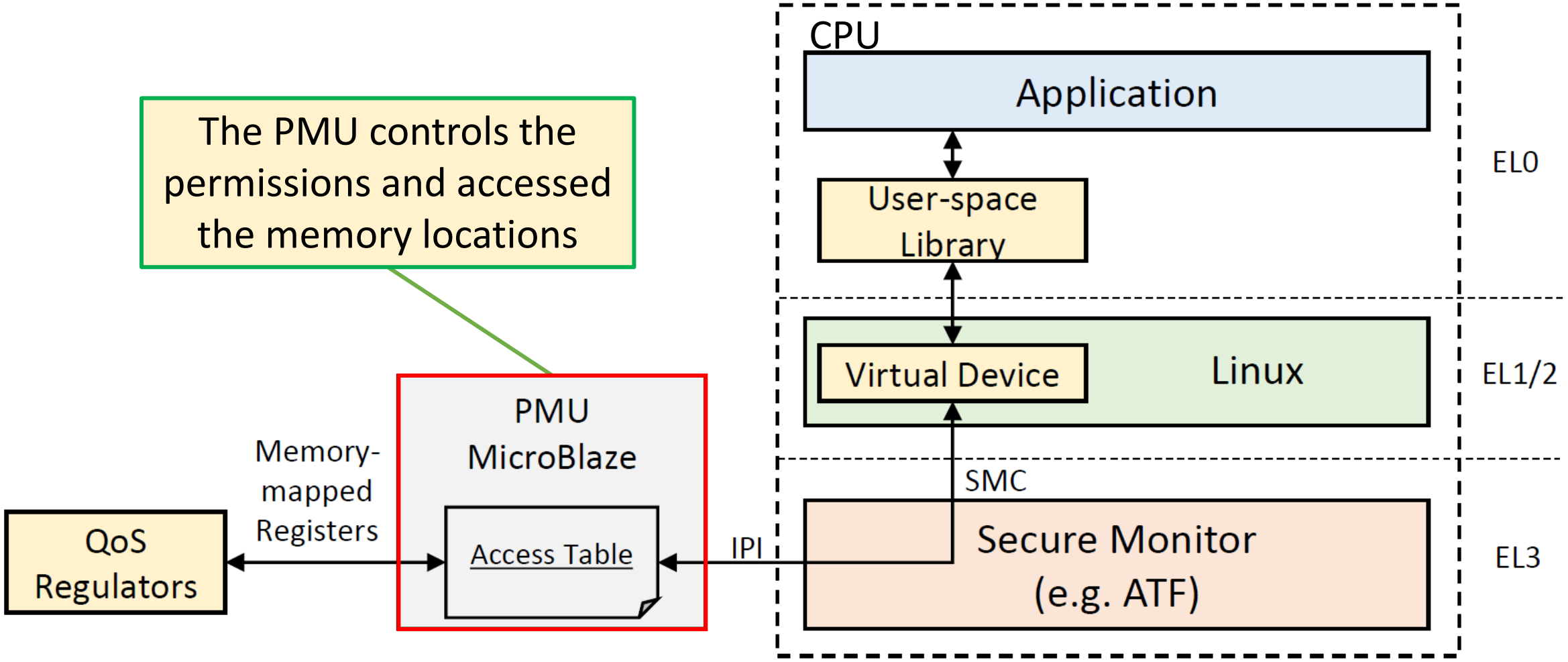
QoS-400 configuration



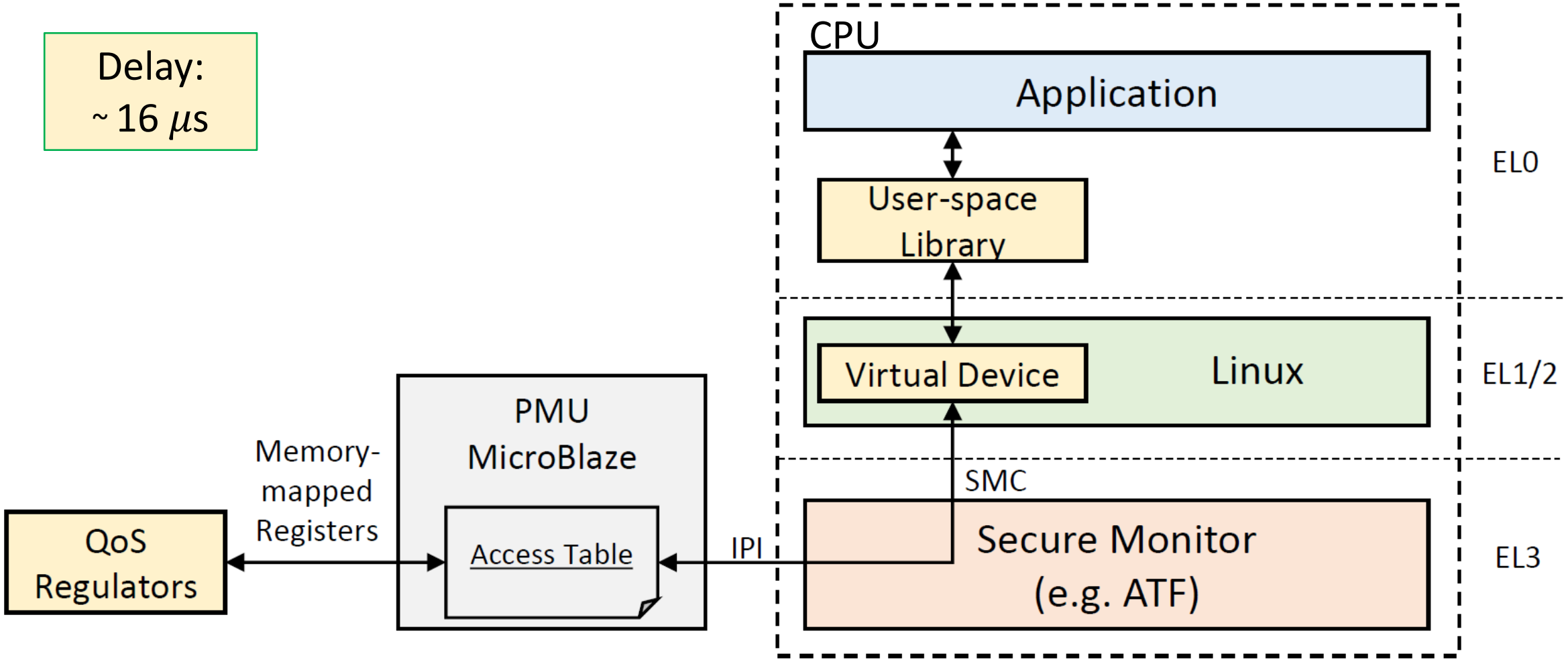
QoS-400 configuration



QoS-400 configuration

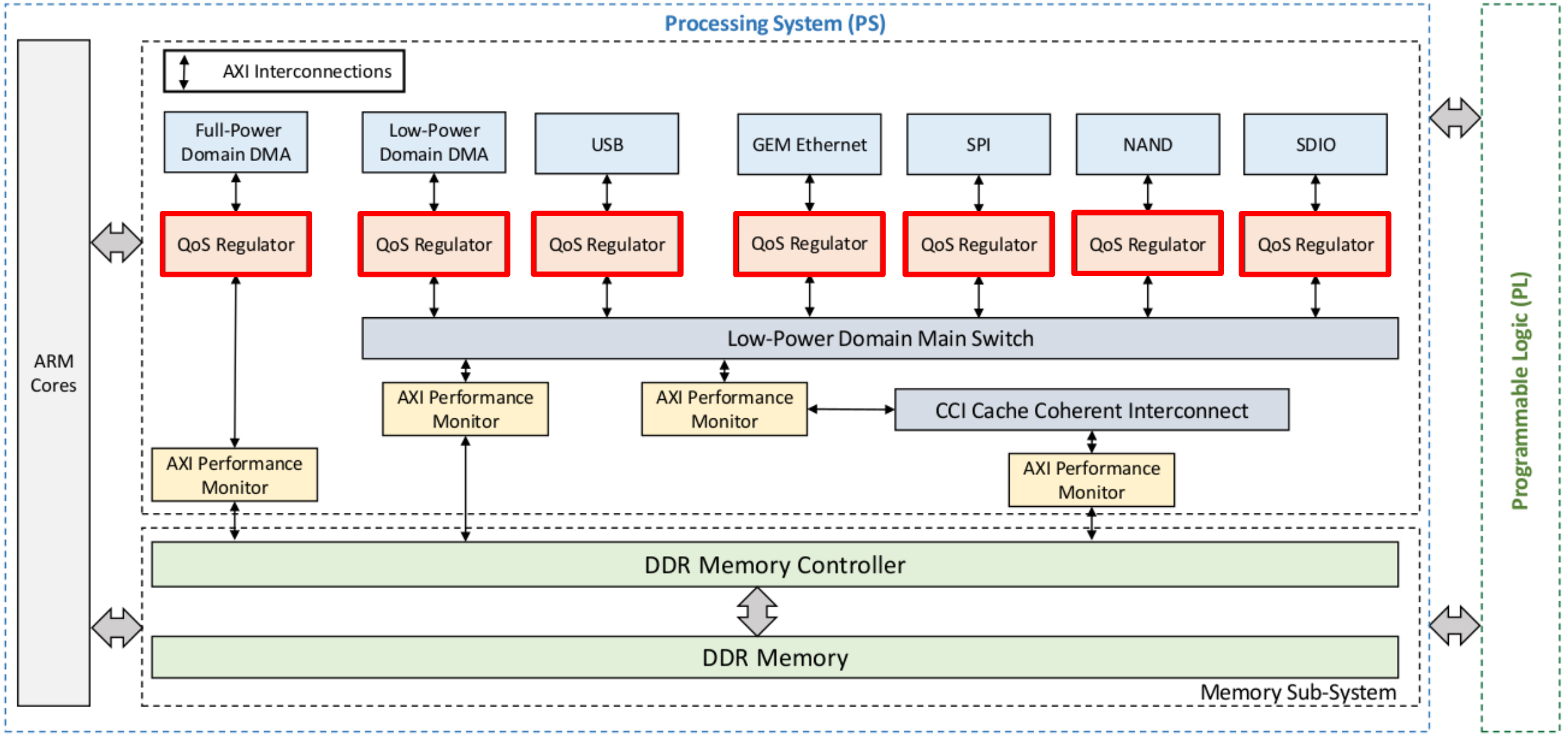


QoS-400 configuration



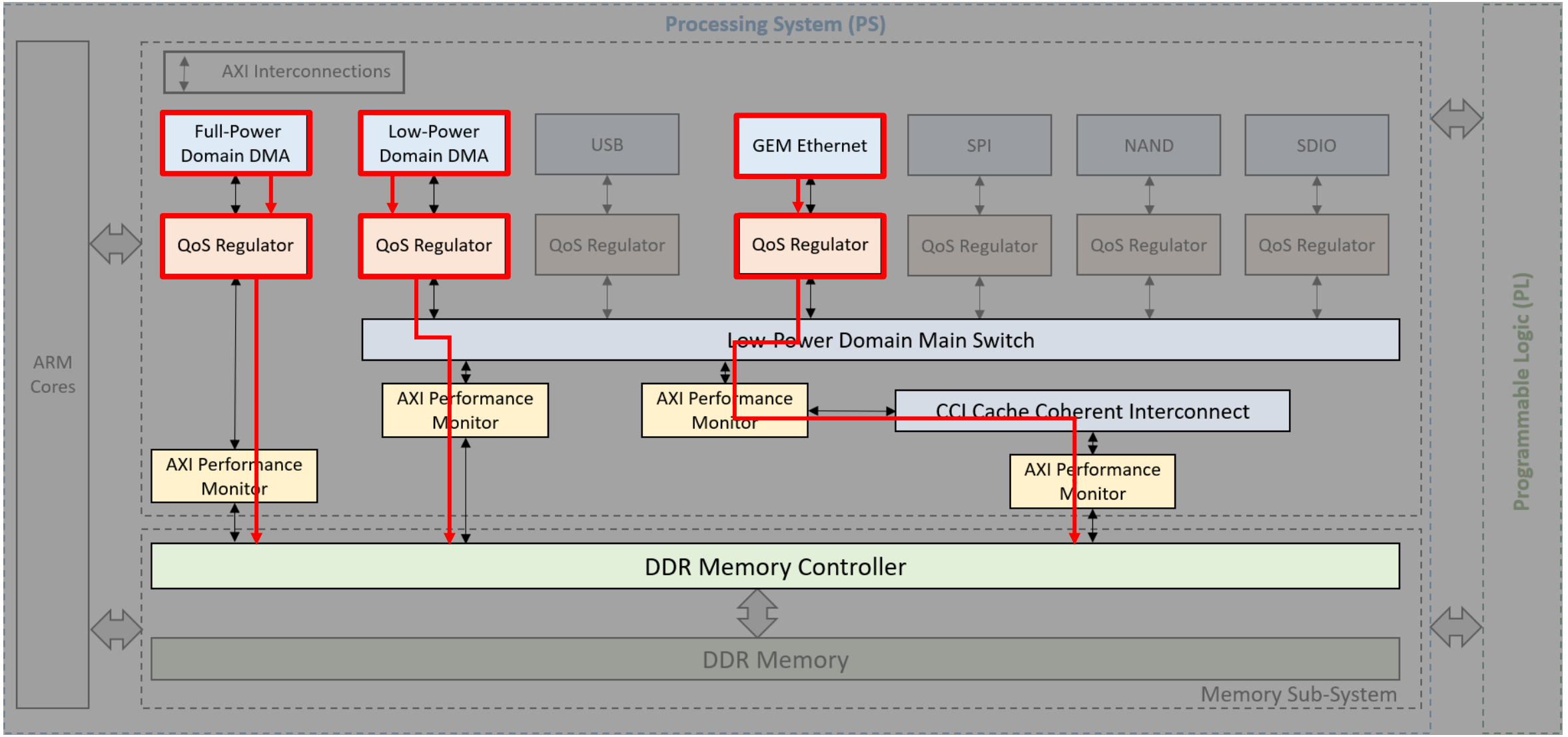
Retis
Real-Time Systems Laboratory

Ultrascale+ architecture



Experimental results

Ultrascale+ architecture

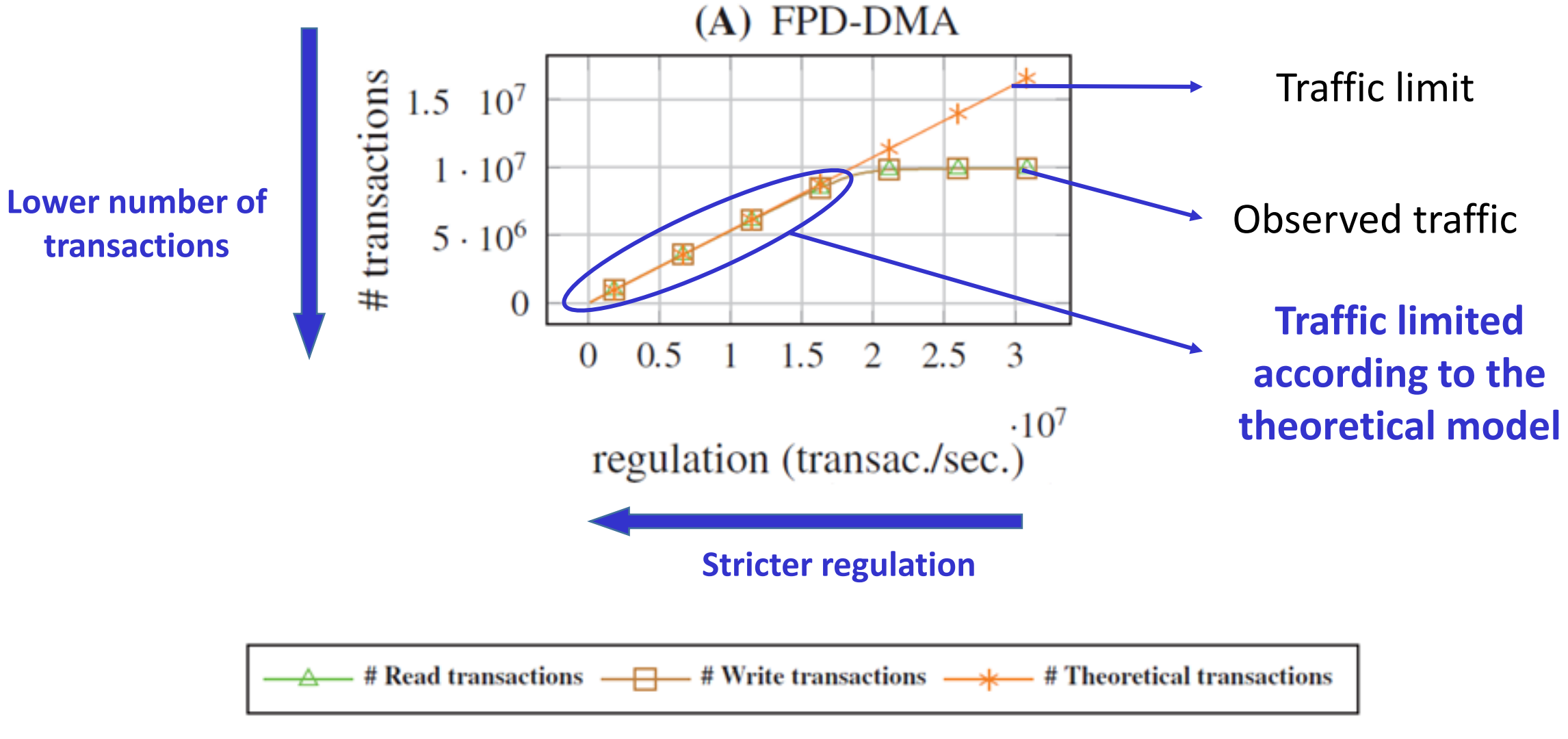


Retis
Real-Time Systems Laboratory

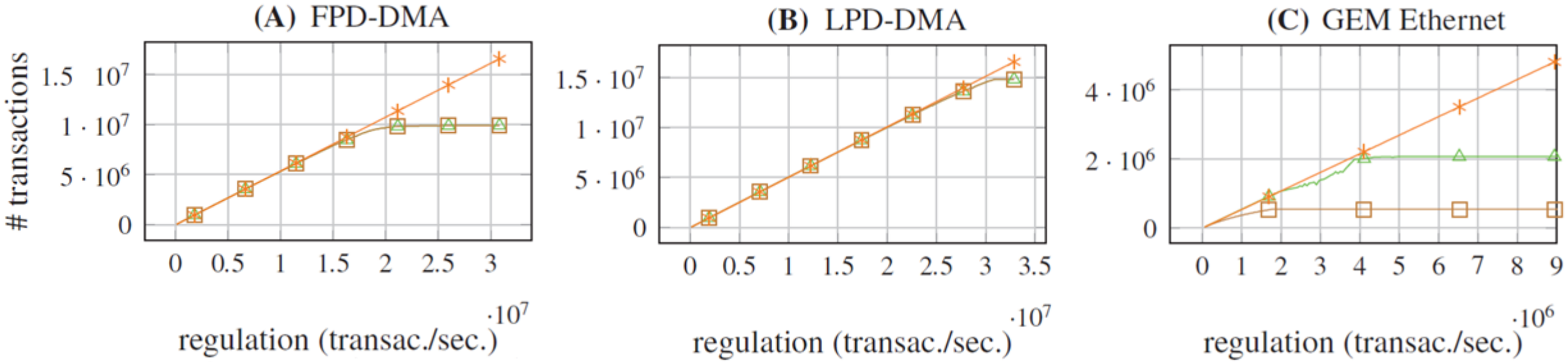
Benchmarks

- Validation of theoretical model
- Effects of regulation on tasks' execution time:
  - San Diego Vision Benchmarks Suite (SD-VBS)
- Effects of regulation on I/O devices' transfer time

Model validation

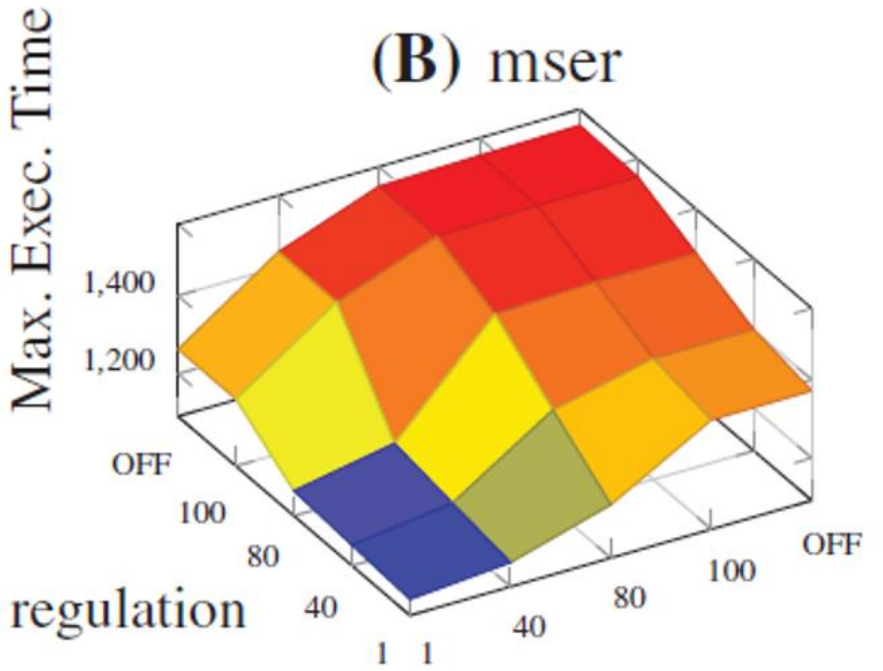


Model validation



Effect on tasks

- Maximum observed execution times (in milliseconds) as a function of the regulation control value of two devices reported on the axis of each chart.



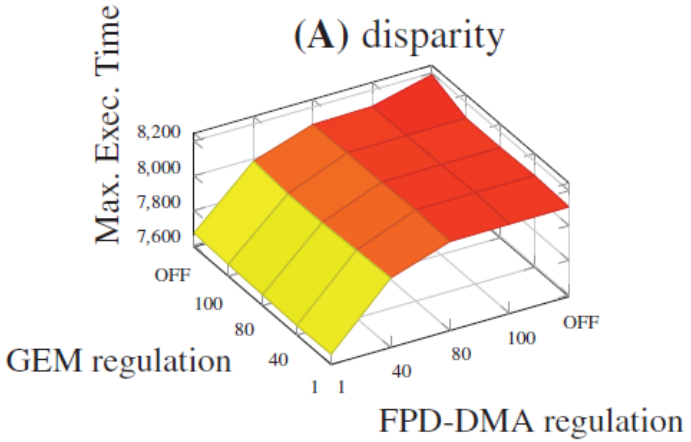
Execution time: - 26%

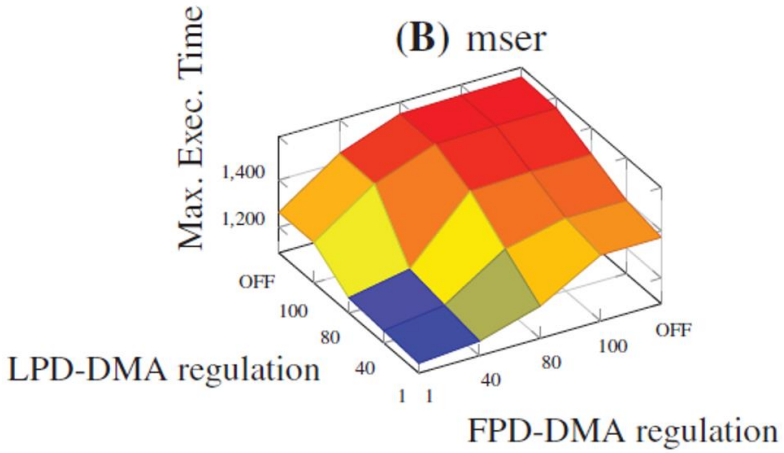
Stricter regulation

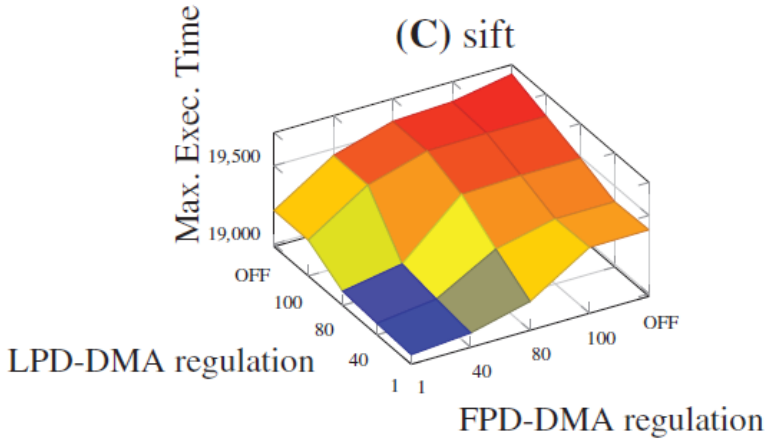
Stricter regulation

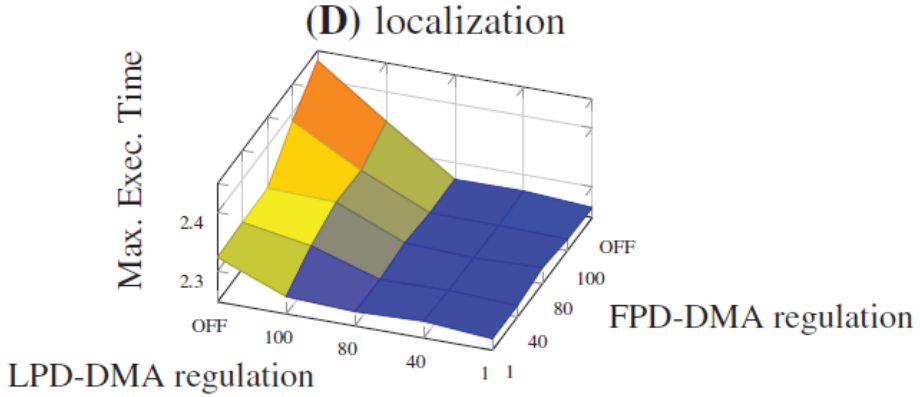
Effect on tasks

- Maximum observed execution times (in milliseconds) as a function of the regulation control value of two devices reported on the axis of each chart.



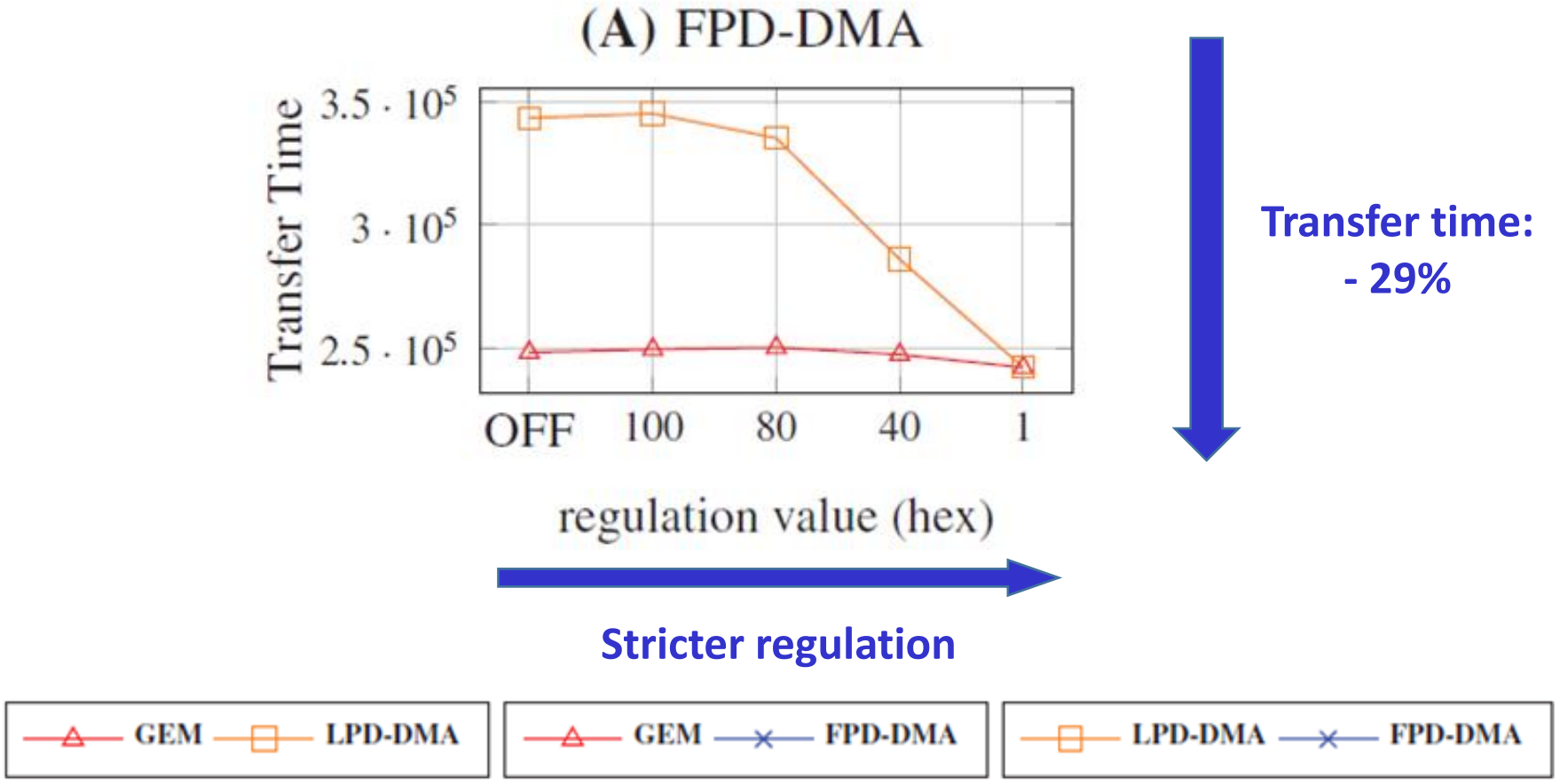






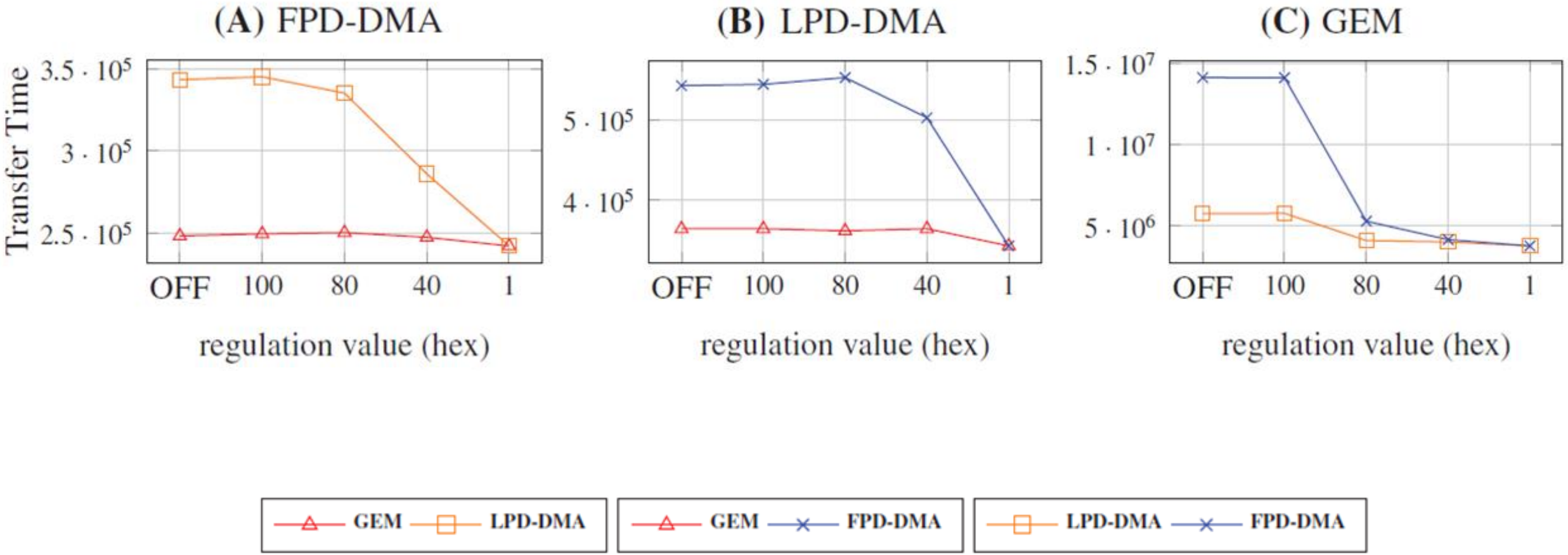
Effect on other I/O devices

- Effects of QoS regulation on the transfer time (in clock cycles) of other I/O devices



Effect on other I/O devices

- Effects of QoS regulation on the transfer time (in clock cycles) of other I/O devices



Conclusions and Future Work

- We analytically derived a model for the QoS-400 regulator
- We developed a strategy to access the configuration register at EL0 on the Ultrascale+ MPSoC
- We experimentally evaluated the effect of the memory interference generated by I/O devices and its regulation on the execution time of tasks and the transfer time of I/O devices.

Future work:

- Optimisation methods to simultaneously optimize the memory bandwidth of tasks and I/O deviced
- Study of the trade-off between tasks' response time and I/O bandwidth

Thank you!

Matteo Zini
matteo.zini@santannapisa.it