

Multithread accelerators on FPGAs: a Dataflow-based Approach



Francesco Ratto¹, Luigi Raffo¹, Francesca Palumbo²
¹University of Cagliari (IT), ²University of Sassari (IT)
 francesco.ratto@unica.it

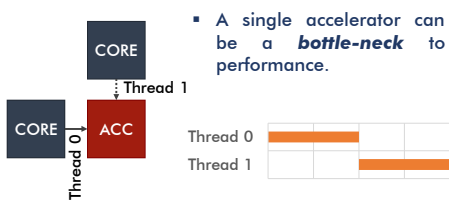


Abstract

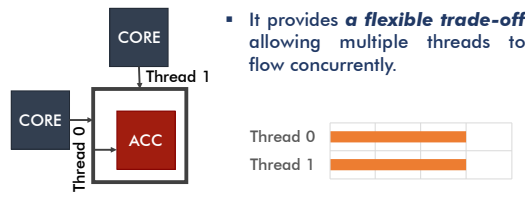
Computing elements of cyber-physical systems should ensure **flexibility** (to guarantee interoperability) and **adaptivity** (to cope with evolving requirements, such as internal changes in battery level and or newly requested critical tasks). This contribution describes a **design methodology** that leverages dataflow models with tagged tokens for designing **specialized multithread hardware accelerators**. Targeting FPGA chips, the resulting hardware allows the concurrent hardware execution of multiple threads and resource sharing among them. Results demonstrate that the proposed solution guarantees a valuable tradeoff between a set of parallel single-thread accelerators and one time-multiplexed single-thread accelerator.

Motivation and goal

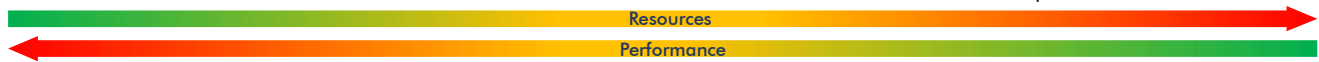
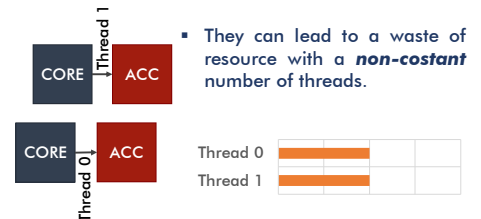
a) Single-thread accelerator



b) Multithread hardware accelerator

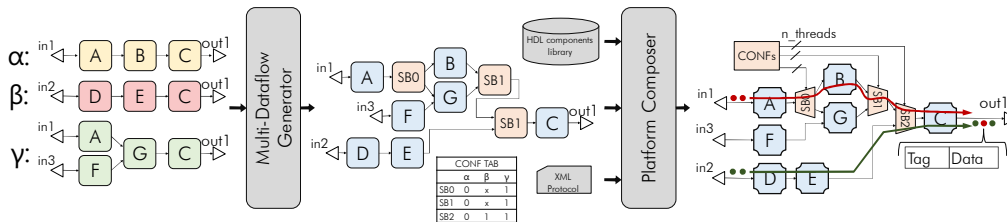


c) Replicated accelerators



Design

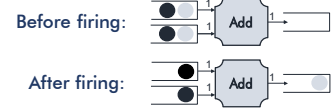
The proposed **model-based** approach implements **hardware-level multi-threading** support leveraging on **tagged dataflow** models. It make possible to automate the design process leveraging on **datapath merging** techniques. Actors process tokens of different threads in a shared computation logic, while memory resources are extended to store the status of each thread.



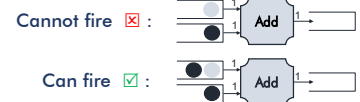
Overview of the **MDC** [3] functionality. The three input dataflows (α , β , γ) are merged by the **Multi-Dataflow Generator** in a unique dataflow that shares common actors (A, C) and addresses tokens depending on the configuration using switching boxes (SB). On the right, the output of the **Platform Composer** is depicted. Two threads, in red and green, are concurrently processed according to the different supported configurations. The sboxes address tokens considering both the configuration and their tag.

FIFOs and actors functional requirements

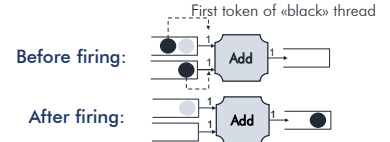
1. A firing actor must **tag the output tokens** with the same tag of the input ones



2. The firing rules must be adjusted so that only **matching tokens** can fire the execution.

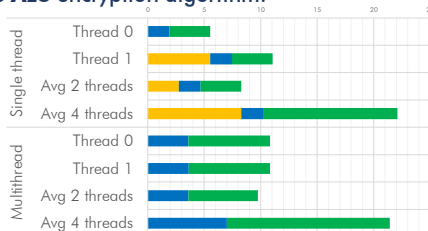


3. **FIFOs must provide semi-out-of-order read**, letting the reading actors choose among the first token of each flow of execution.

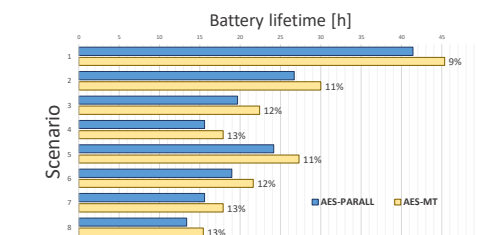


Results

An assessment of the design approach has been carried out on an **HEVC** interpolation filter and on the **AES** encryption algorithm.



Sketch of the **execution of multiple HEVC threads** on single thread and multithread accelerators. For each one the time to wait to start running (yellow), wait for first token produce (blue) and completion of execution (green) are reported. The multithread accelerator is able to provide speed-up with limited resource overhead [1].



Energy consumption estimation carried out on the **AES** accelerators under different testbenches shows that proposed solution delivers energy saving with respect to a set parallel accelerators [2].

Ongoing and future work

A complete host **processor – accelerator environment** with OS support and an API for thread instantiation is **under development**.

Integration of the proposed solution with the MLIR ecosystem within the coming **MYRTUS** [4] project.

- Ratto, F., et al.: Multithread accelerators on fpgas: A dataflow-based approach. doi.org/10.4230/OASiCS.PARMA-DITAM.2022.6
- Ratto, F., et al.: A multithread aes accelerator for cyber-physical systems . doi.org/10.1145/3587135.3592819
- Sau, C., et al.: The multi-dataflow composer tool: An open-source tool suite for optimized coarse-grain reconfigurable hardware accelerators and platform design . doi.org/10.1016/j.micpro.2020.103326
- Myrtus will receive funding by the Horizon Europe program (topic HORIZON-CL4-2023-DATA-01-04) under grant agreement 101135183. Tentative starting date January 2024.