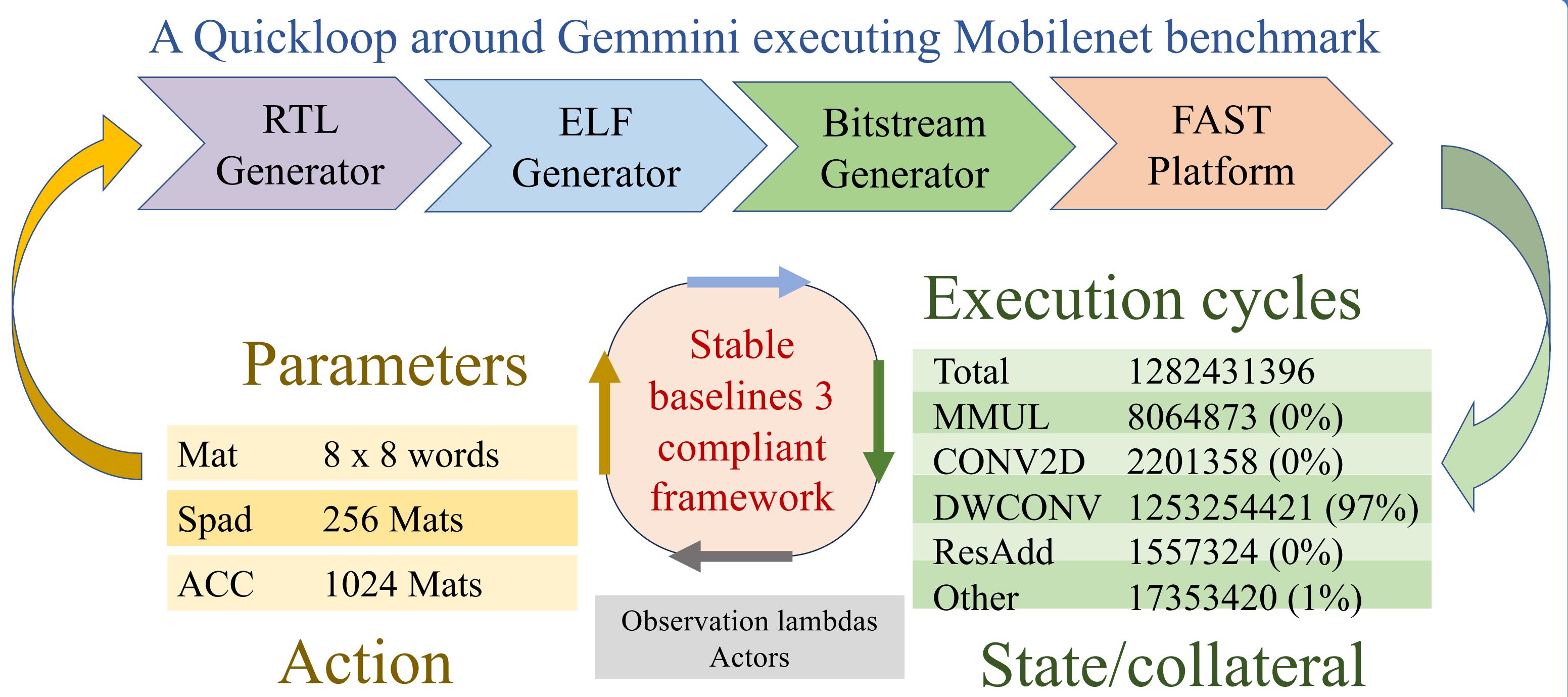


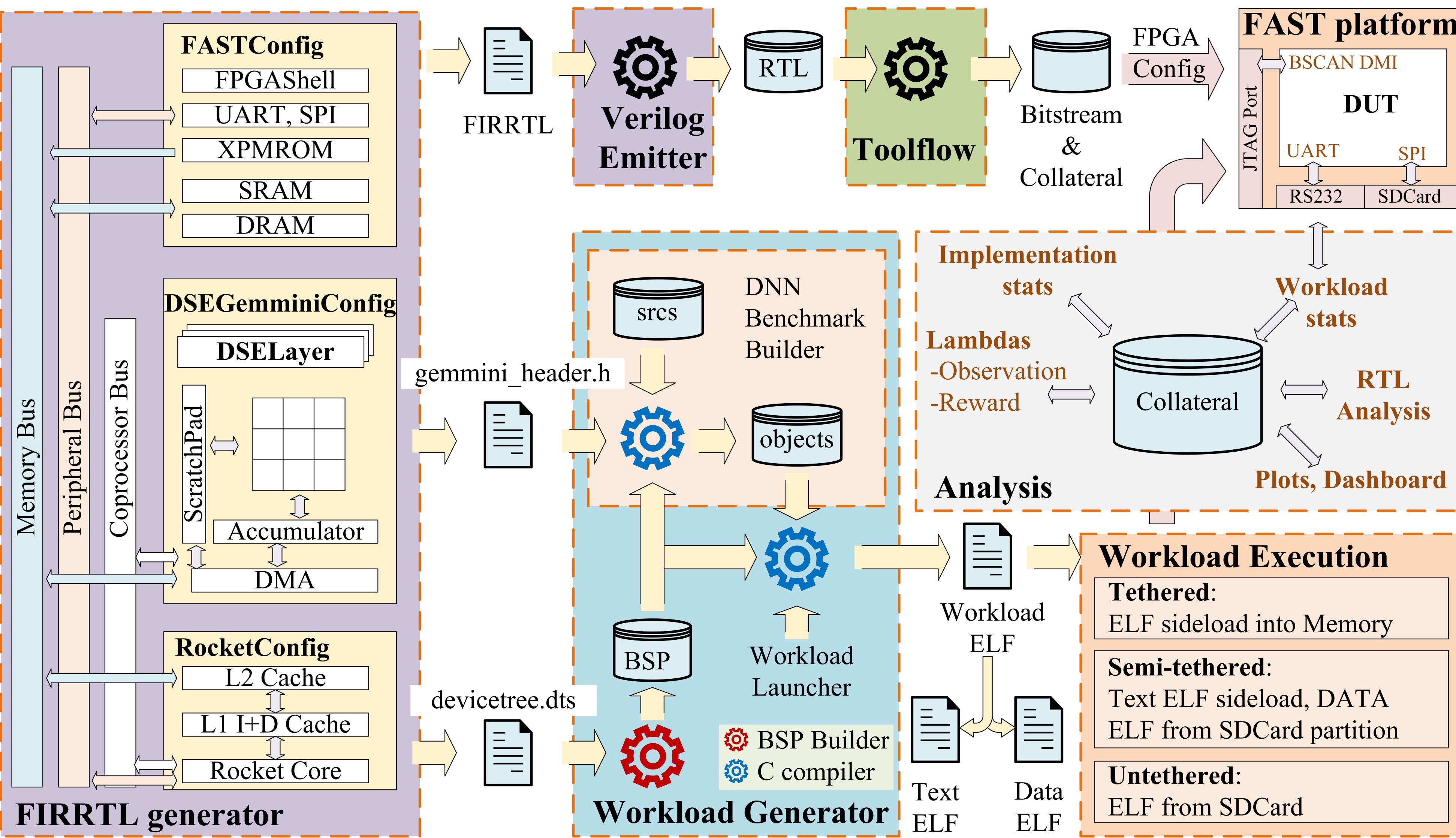
Tayyeb Mahmood (tayyeb.mahmood@gmail.com), Jaeyong Chung (jychung@inu.ac.kr)

1- Introduction

Quickloop enables the workload-driven Design-Space Exploration of parameterized SoC generators and RTL templates on FPGA. Quickloop wraps RTL generator, software stack, and FPGA tool flow into Quicksteps, which are isolated environments optimized for cascadability, scalability and replay. Quicksteps are compliant with Stable-baselines3 of OpenAI Gym and therefore support drop-in interface with conventional and reinforcement learning algorithms. In this WIP, we demonstrate a Quickloop around UCBAR's Gemmini DNN Accelerator on FPGA. We also optimize FPGA turn-around time with DSE-aware tool flow strategies.



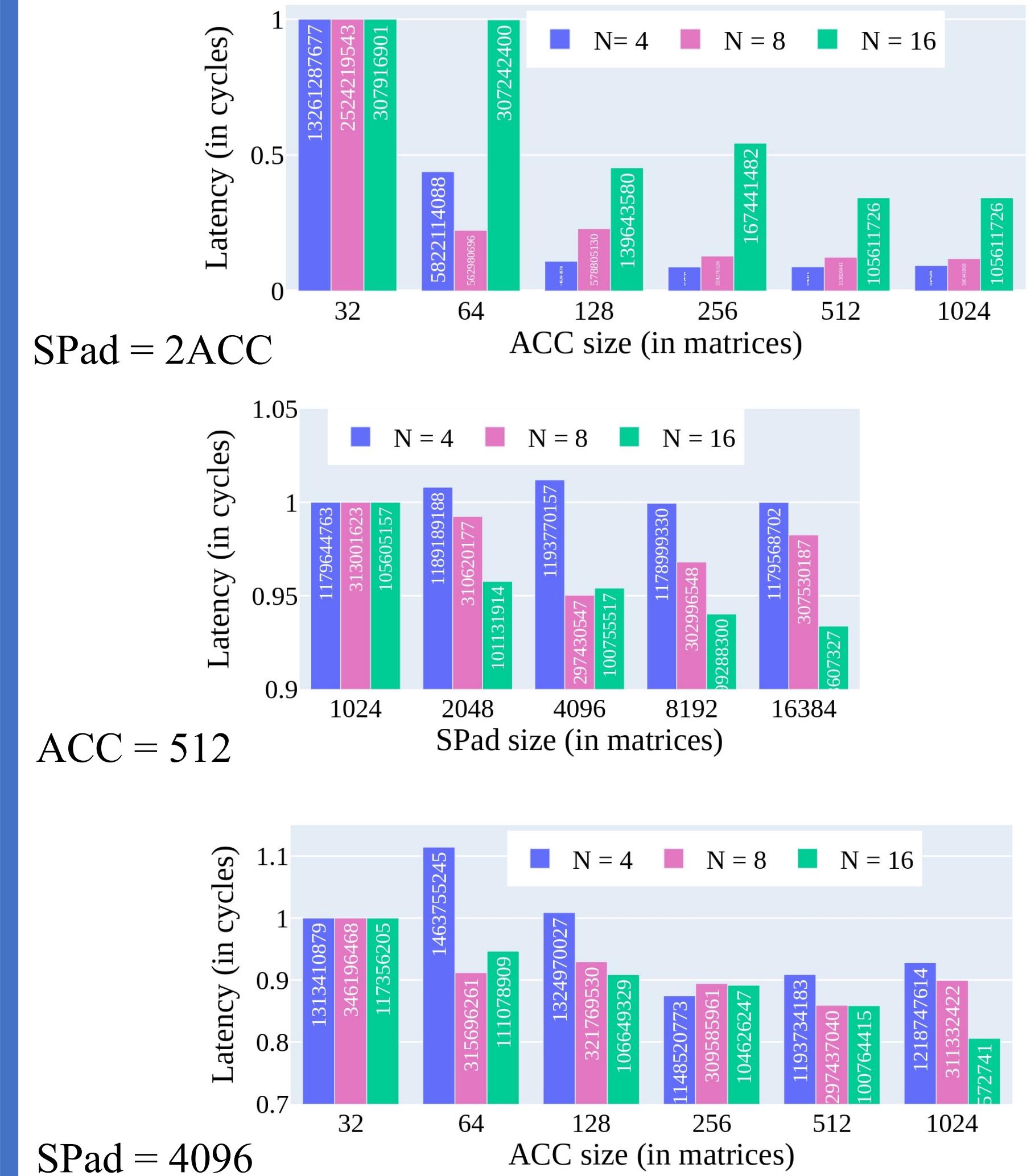
2- Quickloop framework Architecture



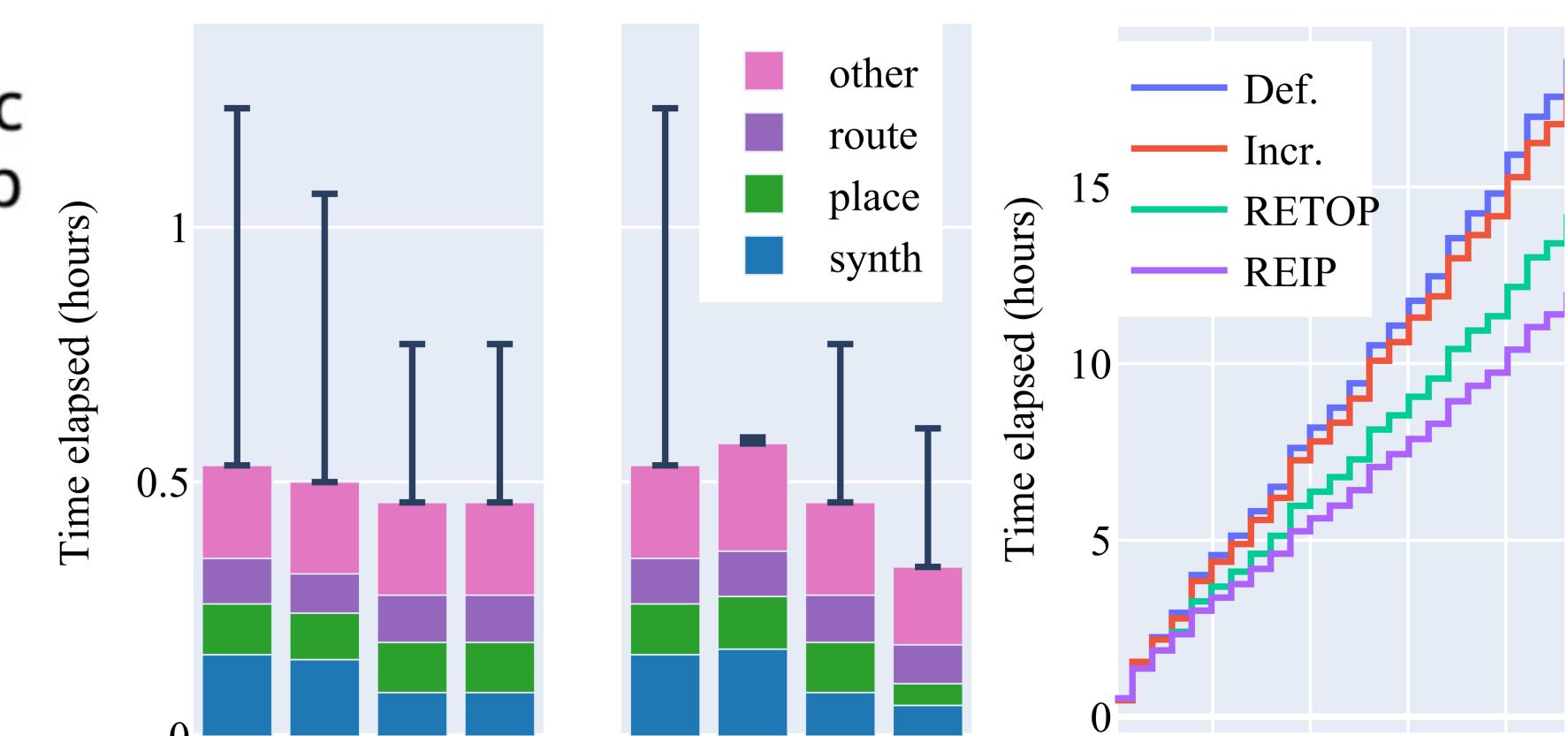
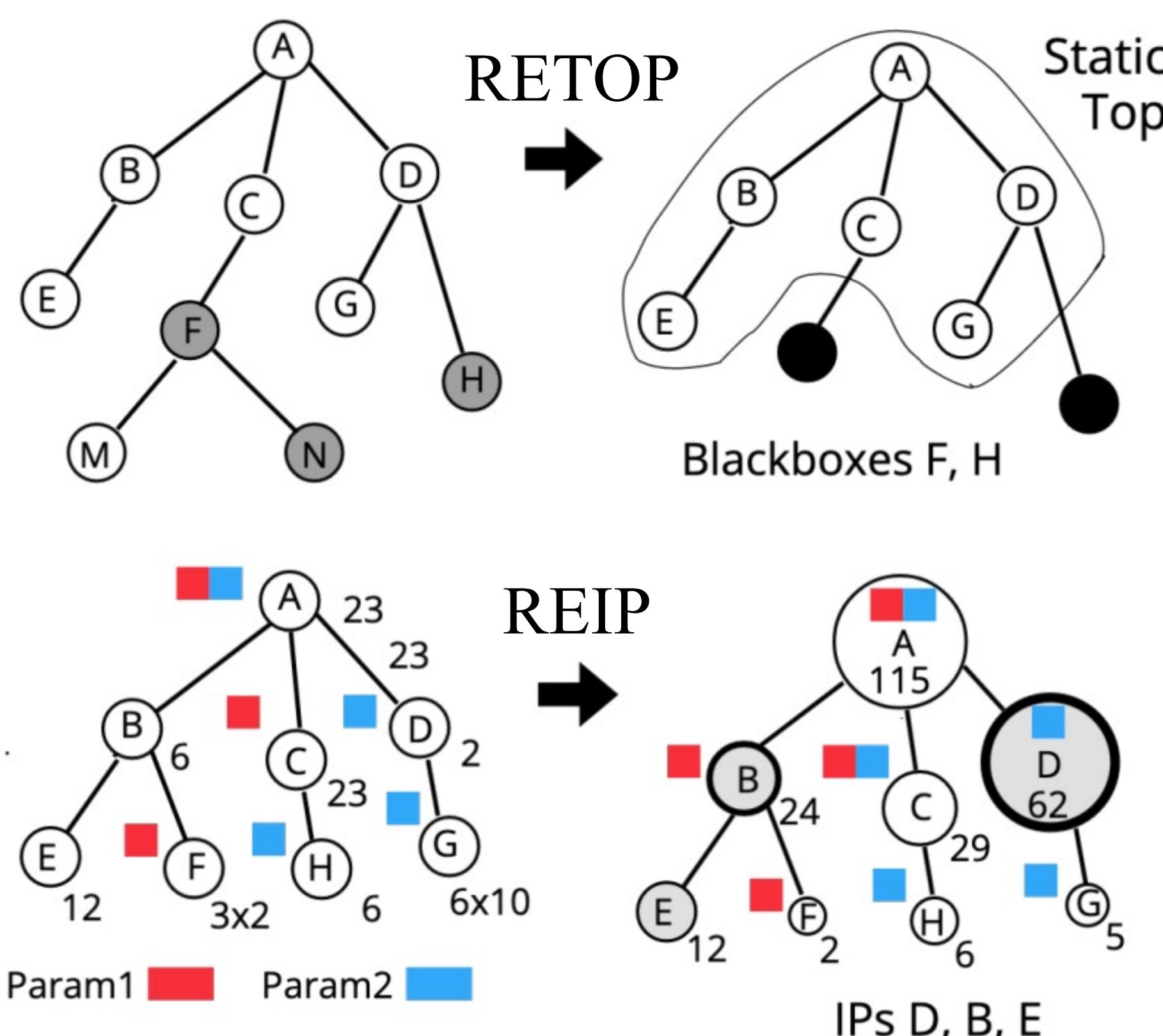
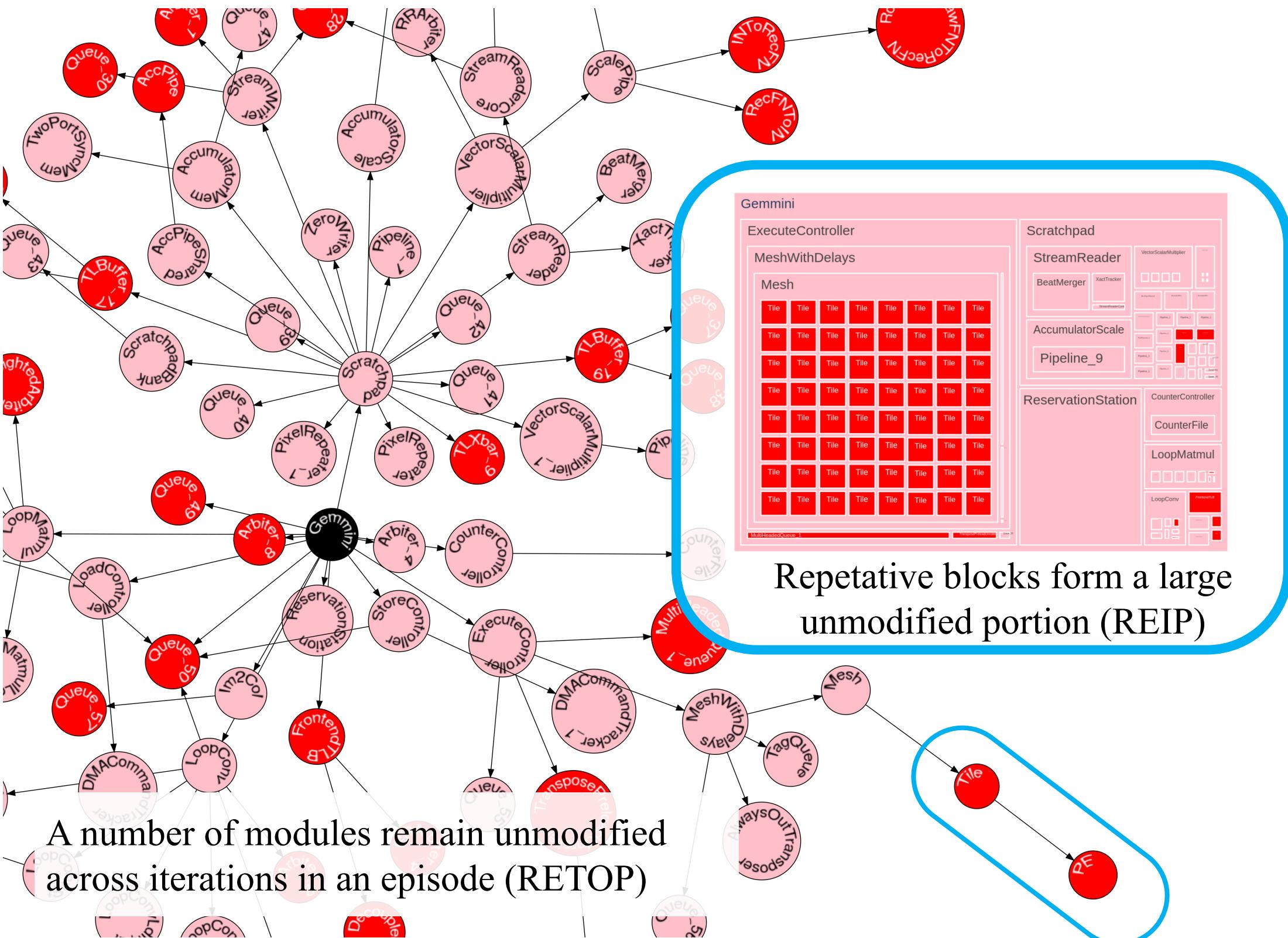
* FAST, FPGA-assisted Scalable Test Platform consists of Xilinx Virtex-7 T2000 FPGA platform with 8GB DDR3 SODIMM. FPGA wrapper is configured with FASTConfig, based on Sifive's fpga-shells infrastructure. Workload generator layers gemmini-rock-tests on freedom-e-sdk. Tethered execution with OpenOCD JTAG BSCAN and UART readback constitutes HIL.

3- DSE Case study

We change systolic array (matrix multiplier) size N, Scratch pad size (SPad) and Accumulator size (ACC) to observe execution pattern of Resnet50. Insets show execution cycles. Large parametric variations are observed.



4- Turnaround Time optimization



Minimum and maximum TAT for various tool flow strategies with synthesis, placement, routing breakups. a) with episode A, b) with episode B , c) Episodal TAT with 24 iterations (randomized parameter sweep)



5- Conclusion

- A realistic workload-based, cycle-accurate simulation is crucial for the optimization of domain-specific architectures like Gemmini.
- A data-driven approach is presented to minimize FPGA Turnaround time (TAT), a key bottleneck in FPGA-accelerated simulation during DSE.
- In future, we aim to
 - Extend our framework to a wider vendor-base of FPGA platforms, from in-expensive development kits to cloud-based scalable infrastructures.
 - Include other RTL generators and benchmarks.

