CHAIR FOR COMPILER CONSTRUCTION

cfaed CENTER FOR ADVANCING ELECTRONICS DRESDEN

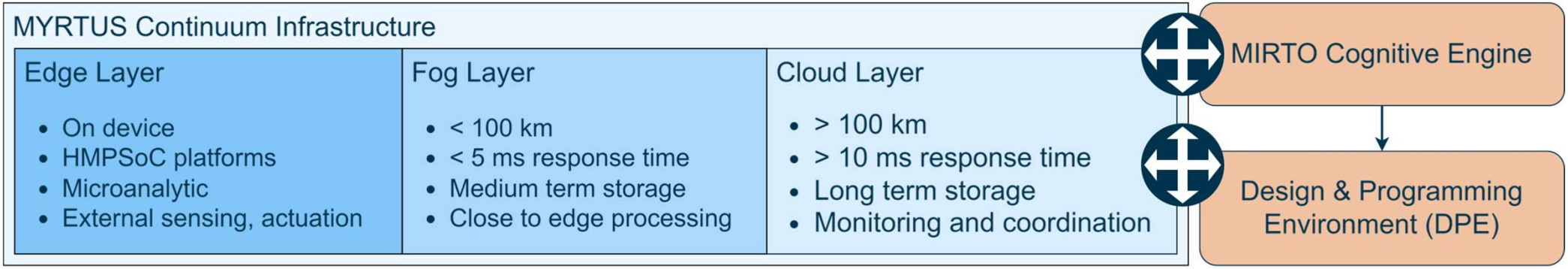
Leveraging the MLIR Infrastructure for Computing Continuum

Jiahong Bi

CPS Summer School Workshop

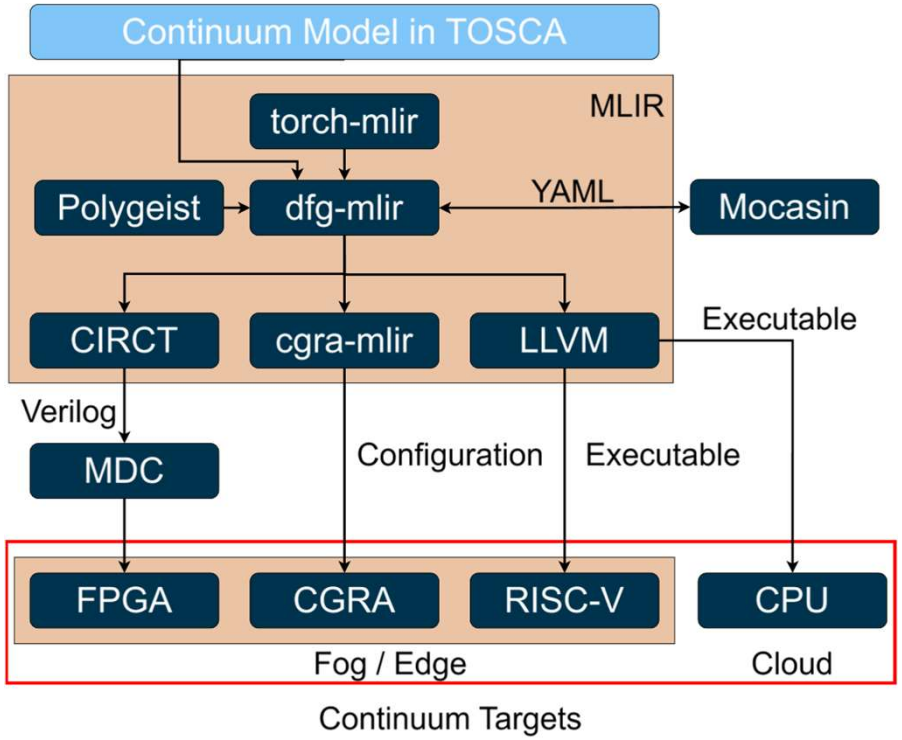
September 16th , 2024

MYRTUS[1] Continuum Insfrastucture



- Computing scenarios are getting more complex
- MYRTUS is built as a layered cloud-fog-edge continuum
- MIRTO engine manages resources, connections and workloads across the continuum
- Design & Programming Environment (DPE) supports definition, implementation and deployment

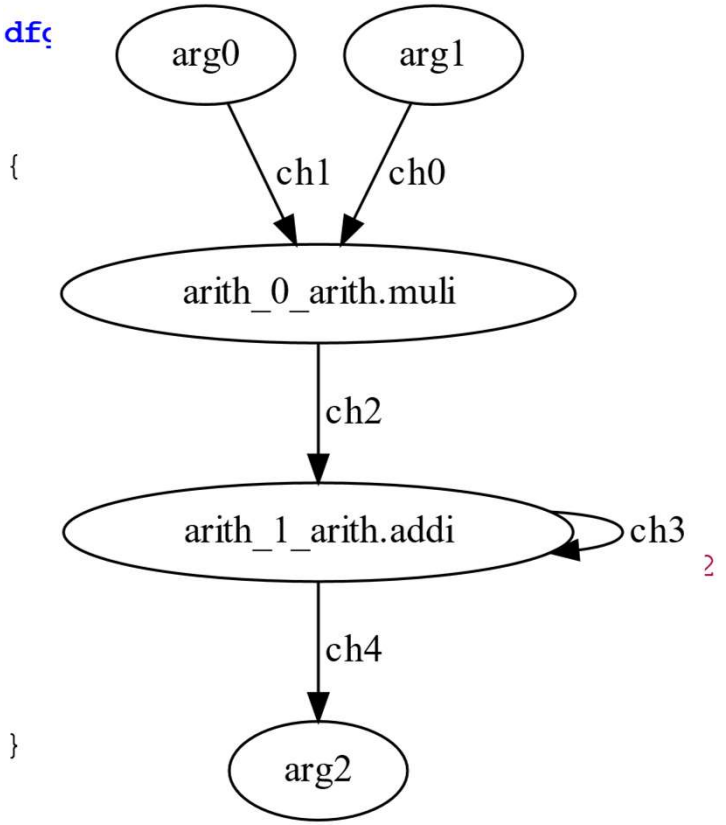
Node-Level Optimization and Deployment



- TOSCA[2] is utilized for orchestration in the continuum
- Input supports: PyTorch, C, …
- Dataflow Graph Generation
- Mocasin[3] runs Design Space Exploration for CGRA mappings
- RISC-V with FPGA/CGRA forms the HMPSoC acceleration platform

Dataflow Graph Dialect[4]

```
dfg.operator @mac
        inputs(%in0: i32, %in1: i32)
        outputs(%out: i32)
        iter_args(%sum: i32)
    initialize {
        %0 = arith.constant 0 : i32
        dfg.yield %0 : i32
    } {
    %0 = arith.muli %in0, %in1 : i32
    %1 = arith.addi %0, %sum : i32
    dfg.output %1 : i32
    dfg.yield %1 : i32
}
```



- OperatorOp is Synchronous Data Flow node
- OperatorOp can be converted to a ProcessOp
- Inner dataflow in OperatorOp can be extracted

Work in Progress and Future Work

- Middle-end and Backend
  - Mocasin for CGRA mapping DSE
  - Integration with MDC[5] for FPGA
  - New dialects for CGRA
- dfg-mlir atop CIRCT[6]
  - Similar semantics to existing dialects
  - Introduce higher abstraction level
- Time and Adaptivity
  - Adopt time semantics from Lingua Franca[7]
  - Hybrid mapping for higher energy efficiency[8]

References

1. Palumbo, Francesca, et al. "MYRTUS: Multi-layer 360 dYnamic orchestration and interopeRable design environmenT for compute-continUum Systems." Proceedings of the 21st ACM International Conference on Computing Frontiers: Workshops and Special Sessions. 2024.
2. OASIS TOSCA Standard. https://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.3/os/TOSCA-Simple-Profile-YAML-v1.3-os.html
3. Menard, Christian, et al. "Mocasin—rapid prototyping of rapid prototyping tools: A framework for exploring new approaches in mapping software to heterogeneous multi-cores." Proceedings of the 2021 Drone Systems Engineering and Rapid Simulation and Performance Evaluation: Methods and Tools Proceedings. 2021. 66-73.
4. Bi, Jiahong. "A Lowering for High-Level Data Flows to Reconfigurable Hardware." (2024).
5. Manca, Federico, Francesco Ratto, and Francesca Palumbo. "ONNX-to-Hardware Design Flow for Adaptive Neural-Network Inference on FPGAs." arXiv preprint arXiv:2406.09078 (2024).
6. CIRCT Project. https://circt.llvm.org/
7. Lohstroh, Marten, et al. "Toward a lingua franca for deterministic concurrent systems." ACM Transactions on Embedded Computing Systems (TECS) 20.4 (2021): 1-27.
8. Smejkal, Till, et al. "E-Mapper: Energy-Efficient Resource Allocation for Traditional Operating Systems on Heterogeneous Processors." arXiv preprint arXiv:2406.18980 (2024).

Thank you for listening!