UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Small is Beautiful - Temporal Accelerators for Embedded FPGAs

Workshop on Collaborative Technologies and Data Science in Smart City Applications 2020

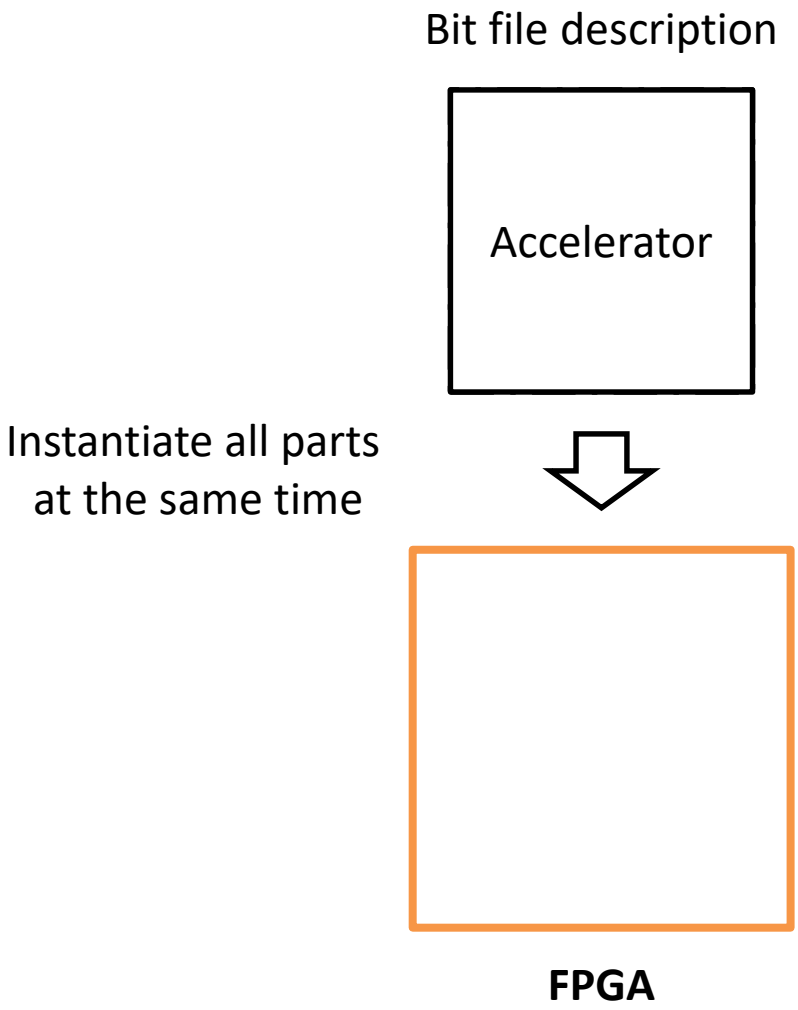
Christopher Cichiwskyj, Gregor Schiele ■ Uni Duisburg-Essen

Motivation

- Field Programmable Gate Arrays (FPGAs) are a promising type of hardware
- Can instantiate powerful accelerators
  - Perform complex calculations in hardware
  - e.g. for AI and Machine learning algorithms

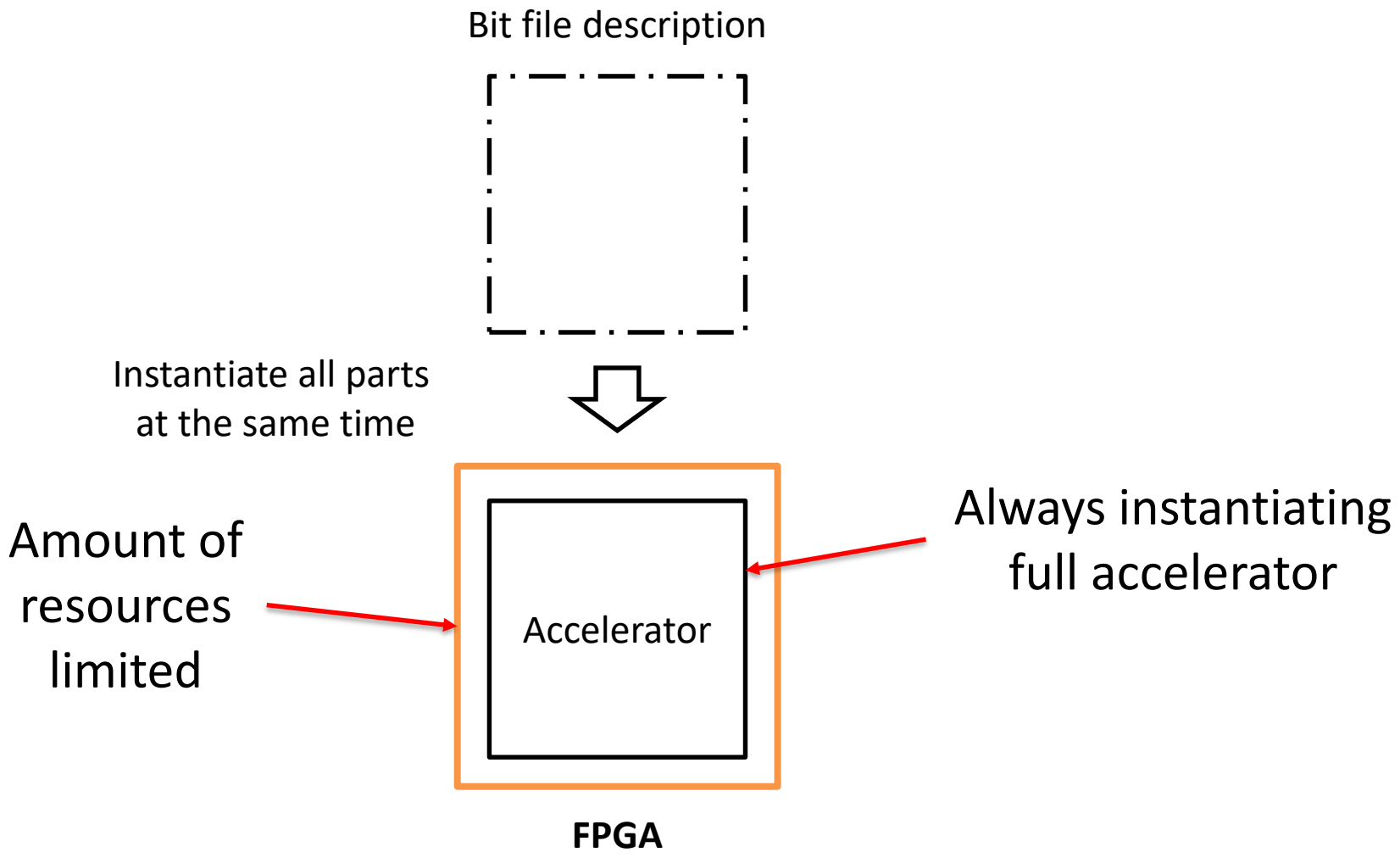
- Using FPGAs can be expensive

Conventional Accelerators



- Design stored in bit file
- FPGA reconfigures

Conventional Accelerators



Problem

→ Complexity defines FPGA size

App. requirements define optimization goal

- Smaller accelerator
  - cheaper device
  - lower power consumption
  - worse performance
- Larger accelerator
  - better performance
  - higher power consumption
  - more expensive device

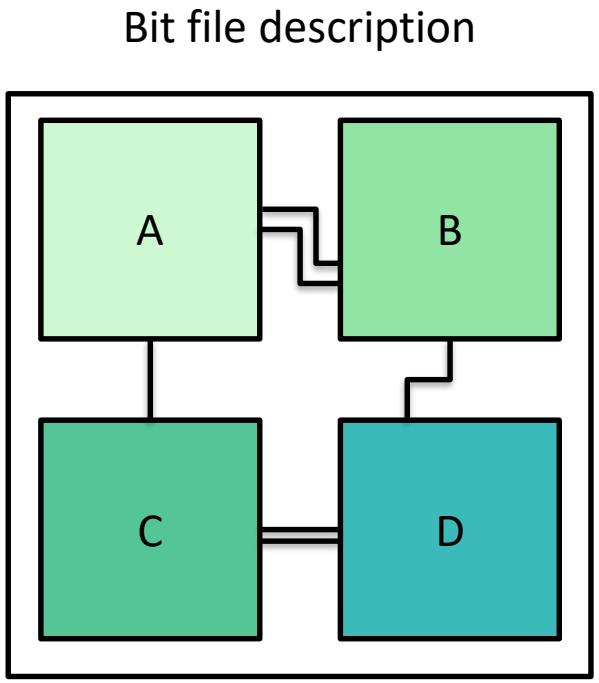
Hard to find an optimal solution

Temporal Accelerator

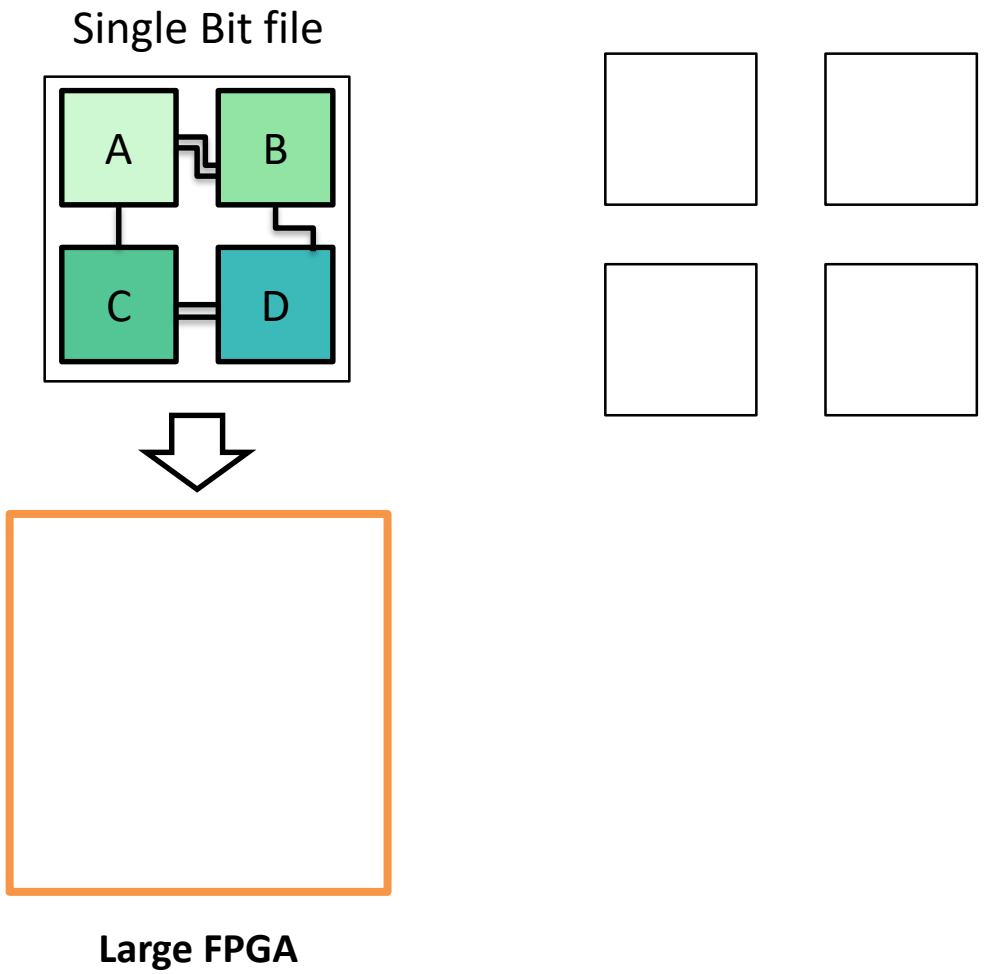
- Main Idea
  - Use smaller (cheaper) FPGA
    - Can’t instantiate full accelerator
  - Reconfigurations to increase available resources ***over time***

Divide and Conquer

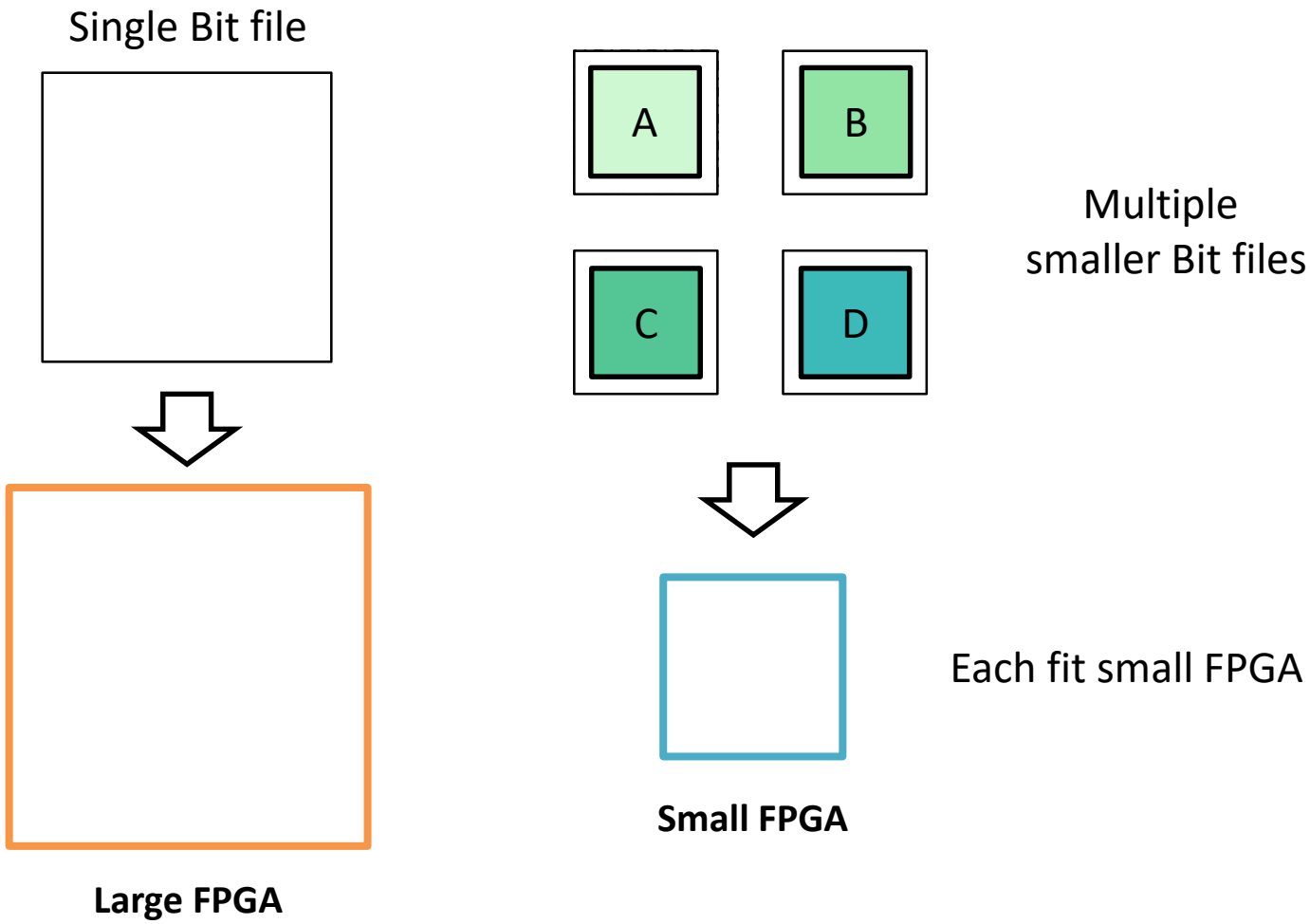
- Accelerator often made of subcomponents
  - Often (fairly) isolated
  - Exchange data



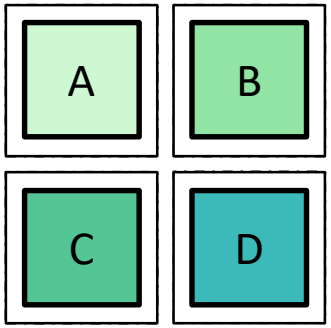
Temporal Accelerator



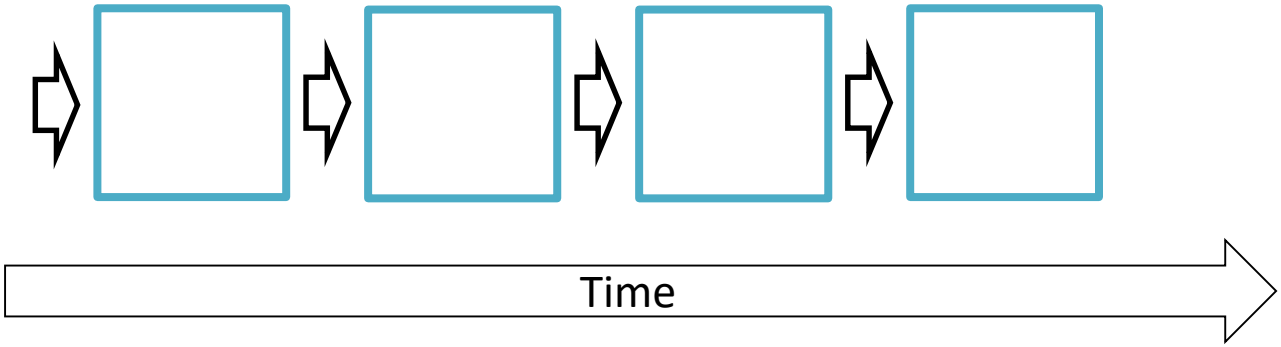
Temporal Accelerator



Temporal Accelerator

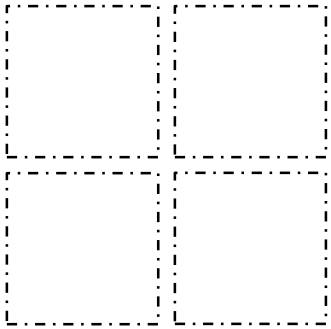


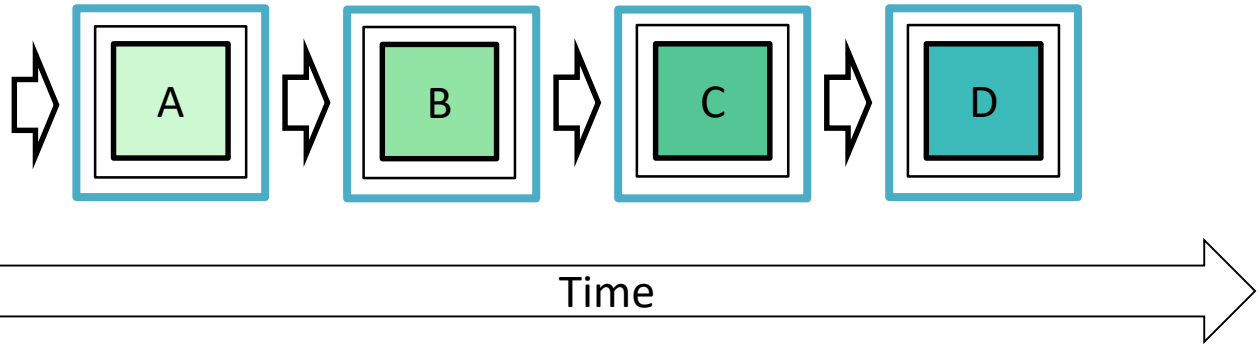
Small FPGA

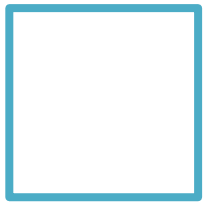


- Execute parts sequentially
  - Not all parts present at all times

Temporal Accelerator







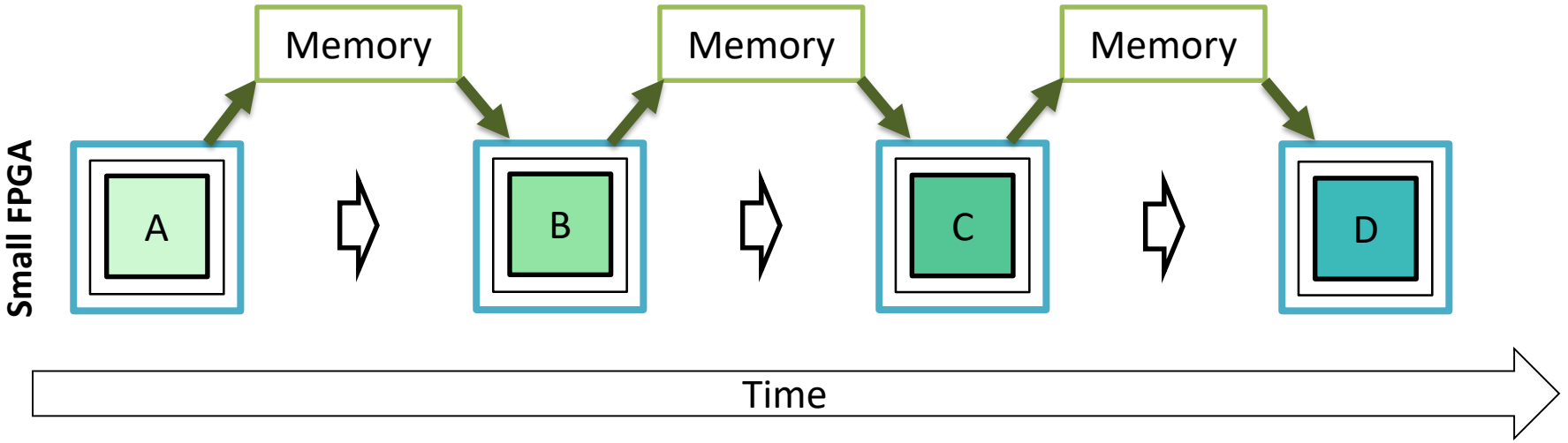
Small FPGA

- Execute parts sequentially
  - Not all parts present at all times

Additional Management

- Can't buffer results during reconfiguration
  → Offload of intermediate results

- *Does not have to be on a memory chips*



Additional Management

- FPGA can (theoretically) perform offloading & scheduling management
- Increases resource consumption overhead
  - Reduces available area for accelerator

→ Combine FPGA with a conventional CPU/MCU chip

Overhead?

- Intuition suggests:

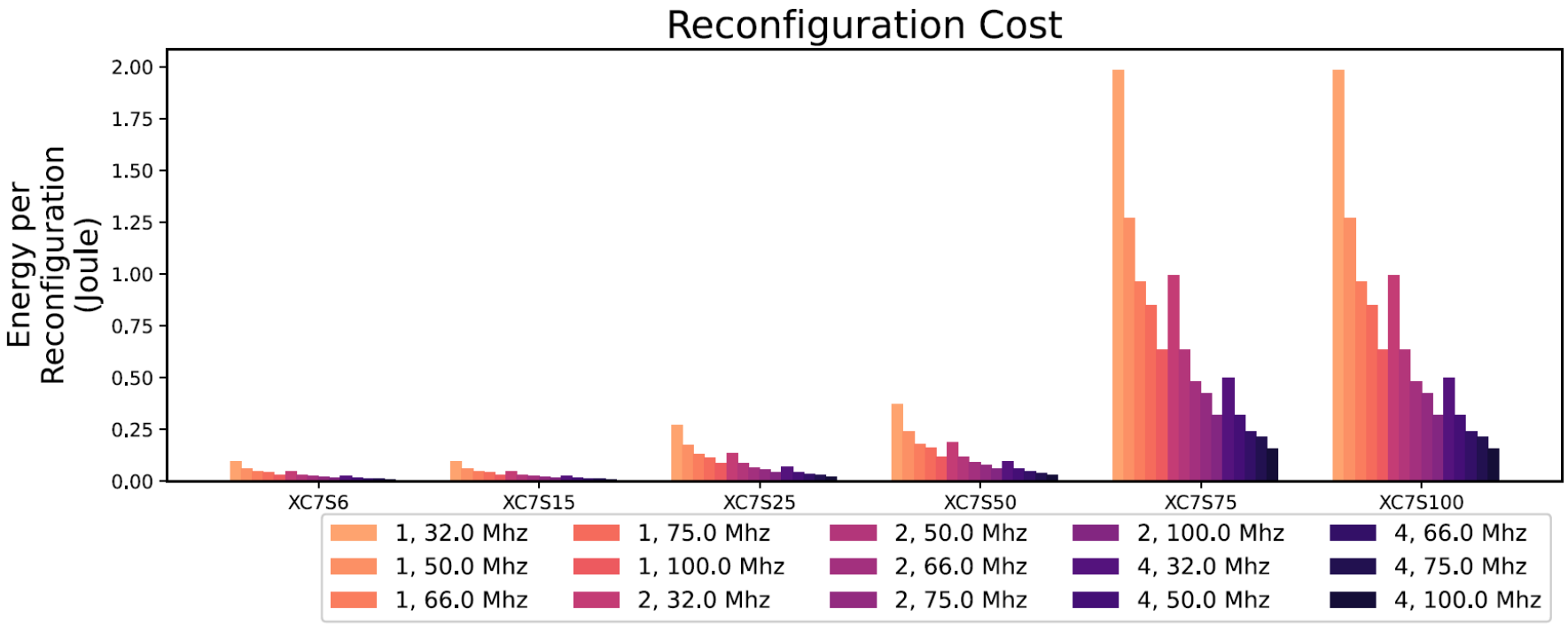
*„So many reconfigurations make it **prohibitively** expensive“*

Overhead?

- Intuition, however, forgot:

Smaller FPGAs have** (usually) **a shorter reconfiguration time

Less Overhead than expected



Conclusion

- Temporal Accelerators are viable
- Can achieve equivalent performance on smaller & cheaper devices