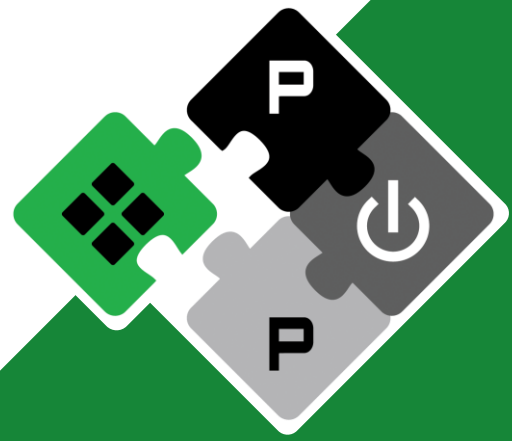
ETH zürich ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

CachePool: Many-core cluster of customizable, lightweight scalar-vector PEs for irregular L2 data-plane workloads

Integrated Systems Laboratory (ETH Zürich)

Zexin Fu, Diyou Shen zexifu, dishen@iis.ee.ethz.ch

Alessandro Vanelli-Coralli avanelli@iis.ee.ethz.ch
Luca Benini lbenini@iis.ee.ethz.ch



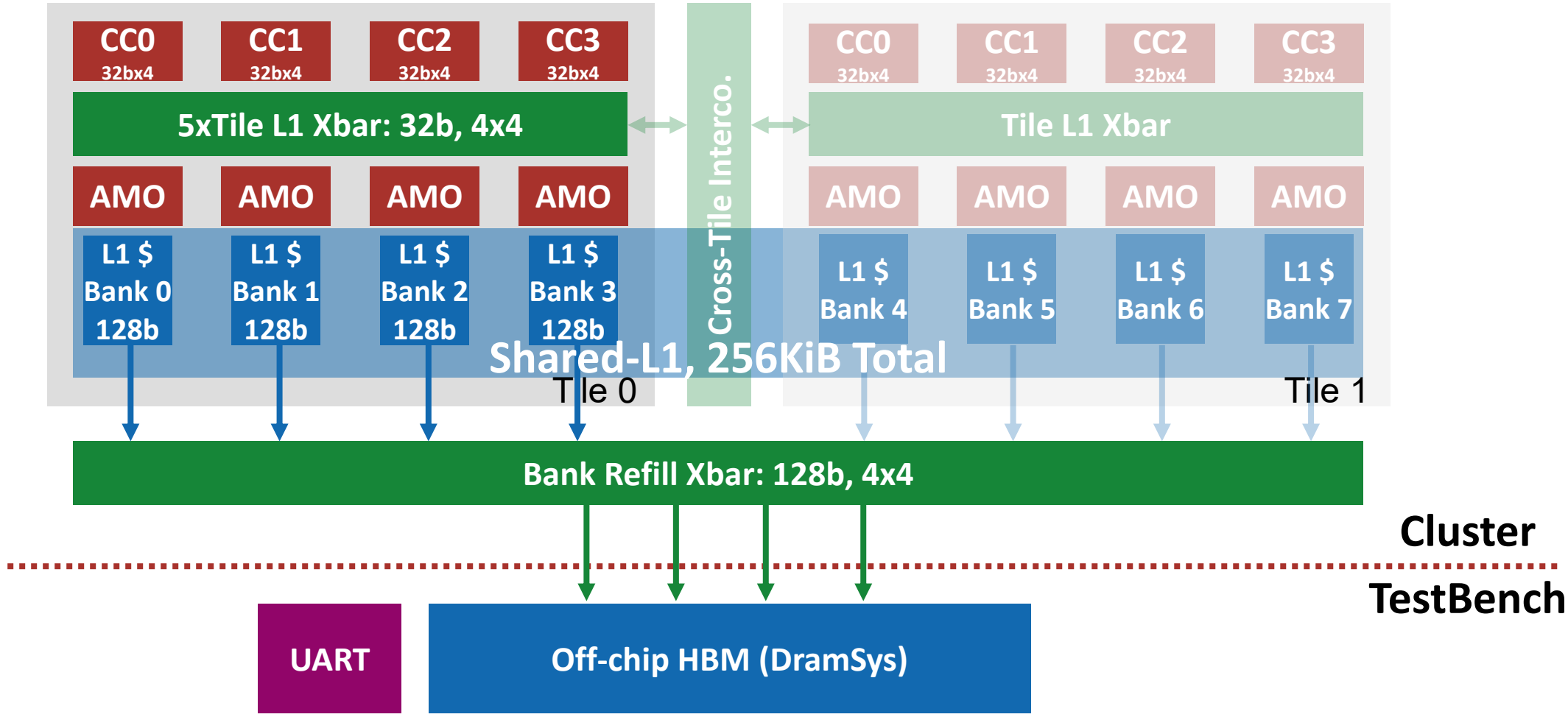
@pulp_platform

pulp-platform.org

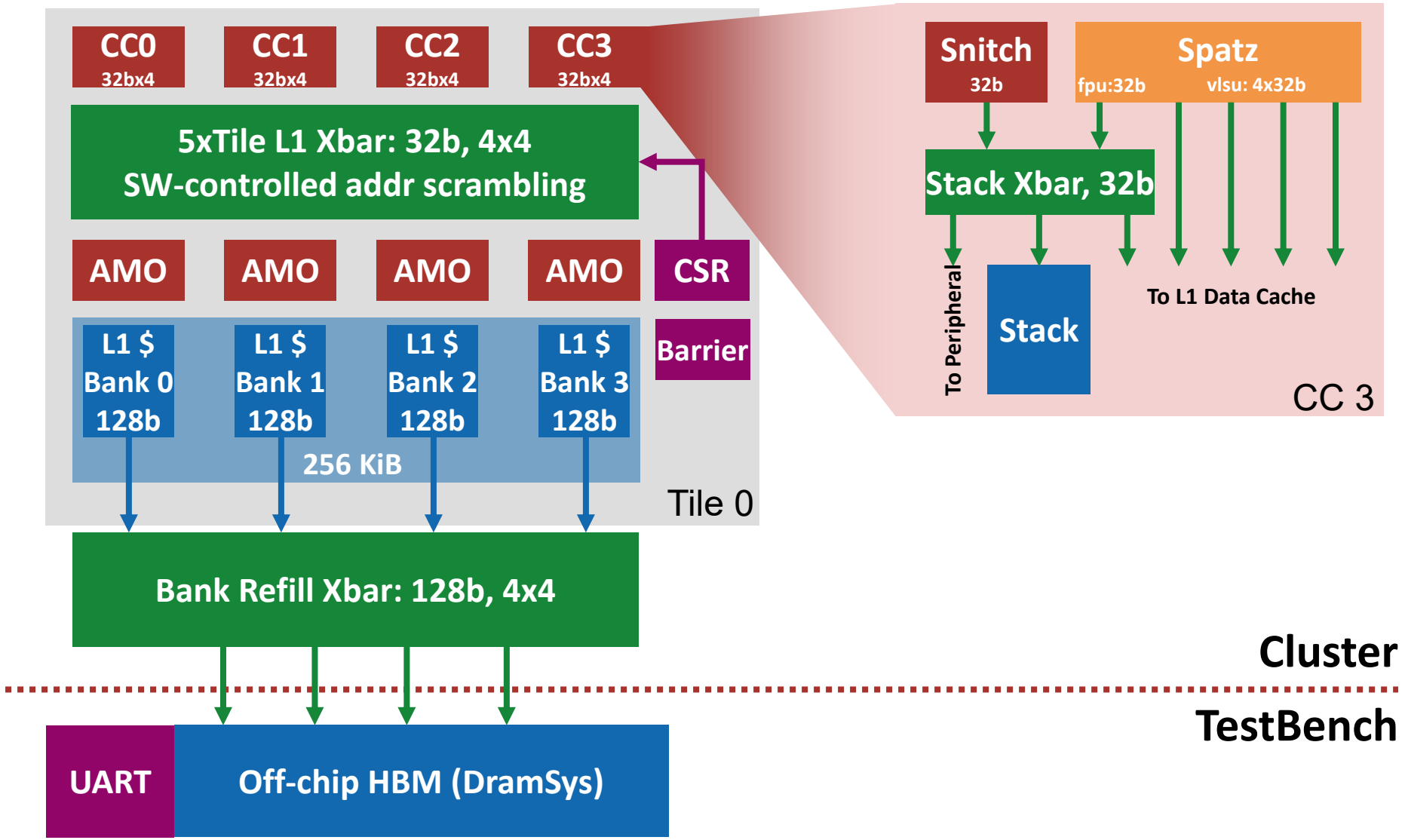
youtube.com/pulp_platform

PULP Platform
Open Source Hardware, the way it should be!

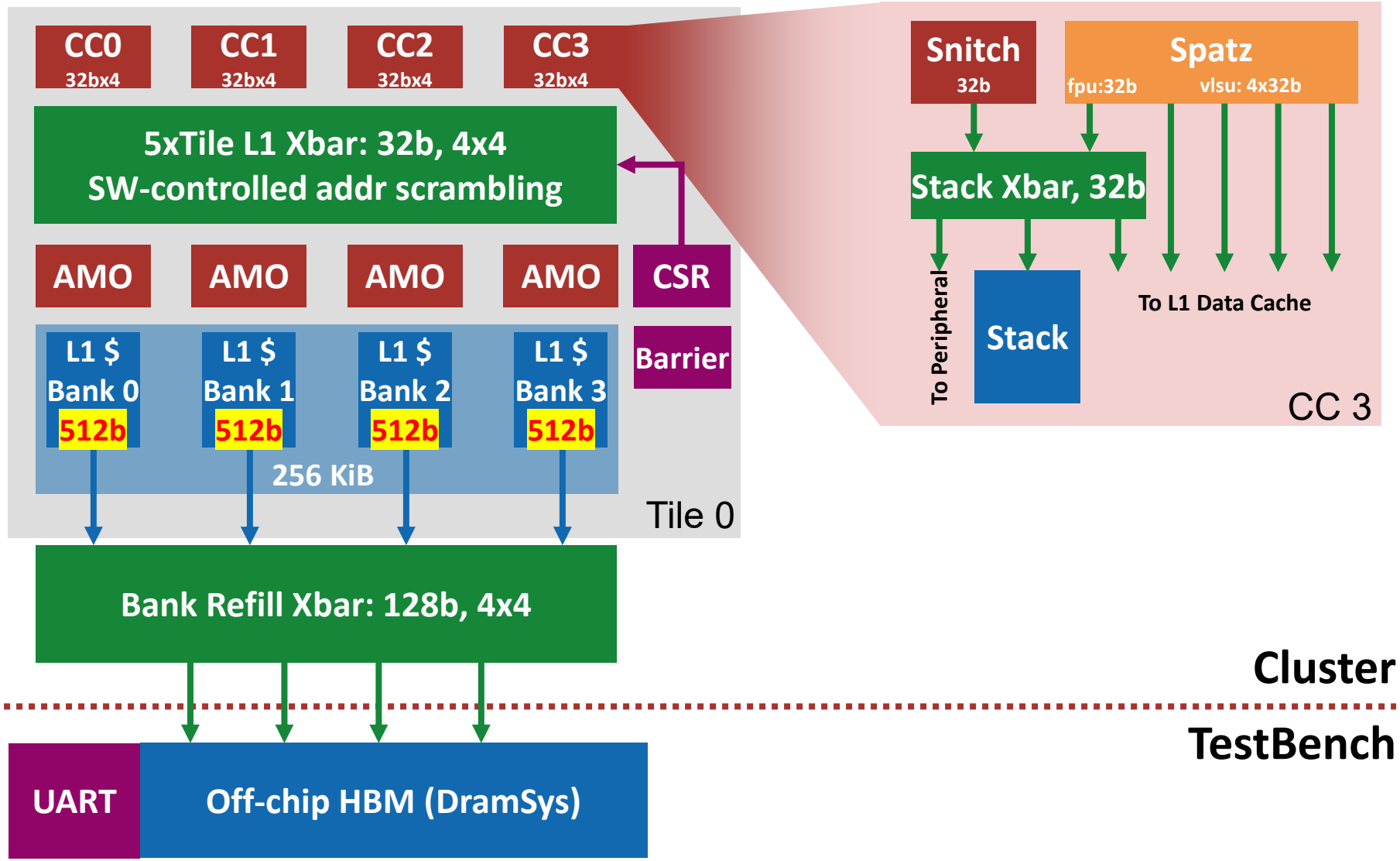
Hardware Development



Hardware Development



Hardware Development

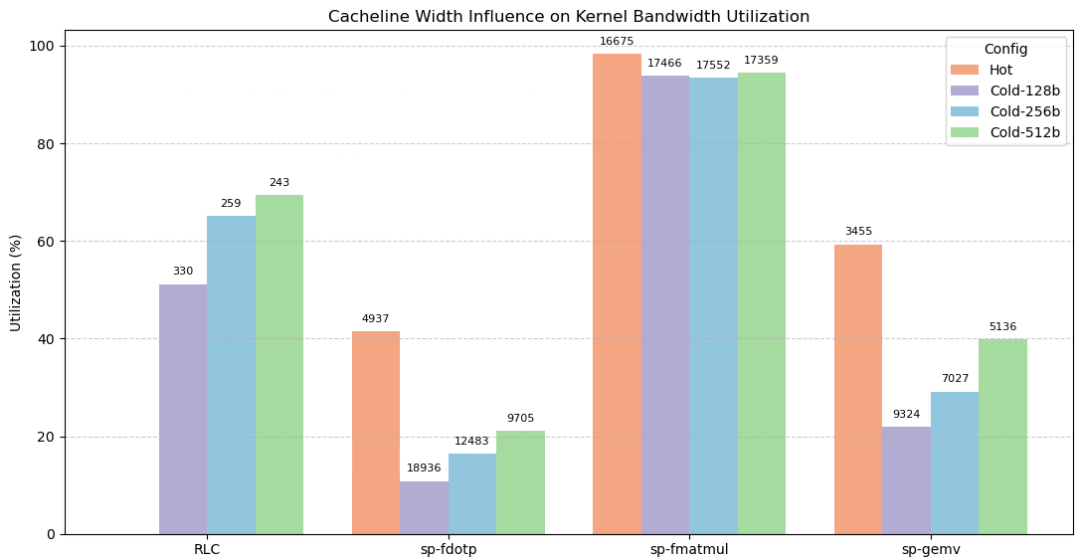


Hardware Development

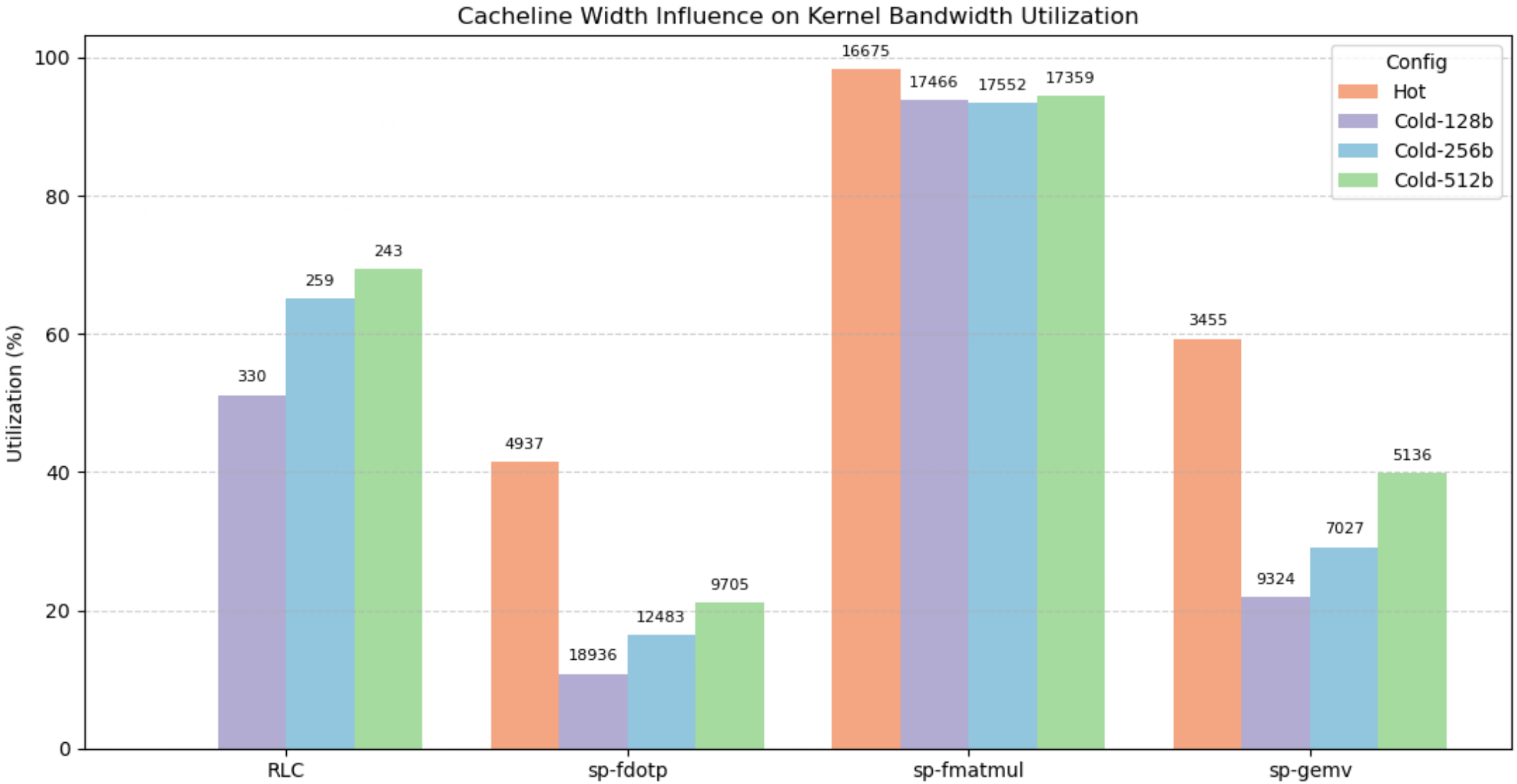
- **Add configurable cacheline width: 128b, 256b, 512b**
  - Interconnection width kept to 128b
  - Burst transaction for refilling, evicting and flushing
  - Use FSM to control the response reassemble in the controller wrapper
    - No changes inside the core of controller
    - Out-of-order support without extra buffers
- **Add a hardware configuration without FPU**
  - Will start BE this week
  - With area analysis on cacheline explorations

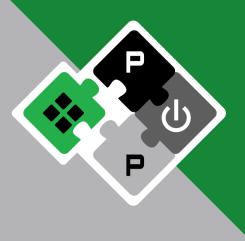
Software Analysis – Different Cacheline Width

- **Longer cacheline gives better cold-cache performance**
  - Hot cache analysis still undergoing
    - Have some tiny bugs to fix
    - Current hot cache data from 512b cacheline
  - Use FPU/LSU utilization in the plot to uniform different kernels
    - RLC: LSU utilization (ideal cycles: data to move / LSU bandwidth)
    - Matmul, dotp, gemv: FPU utilization



Software Analysis – Different Cacheline Width





Thank you!