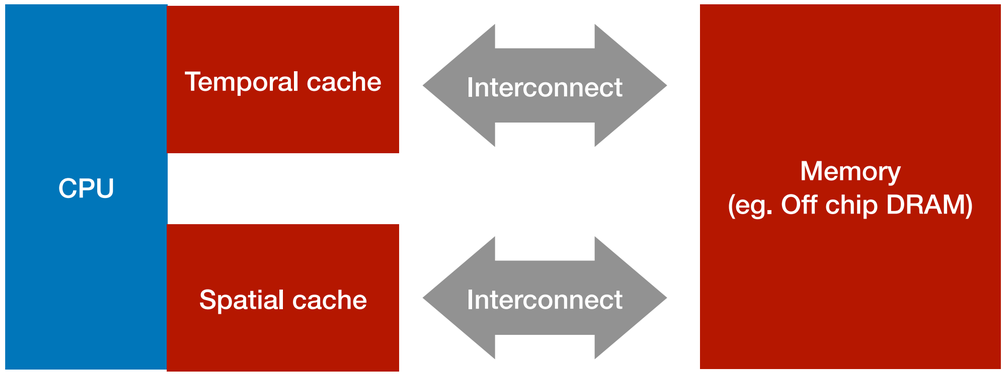
A single case

Then use a temporal/spatial-spilt cache



- The cache setup (decided by PaLM algorithm) is :
- Temporal cache: 4 words cache line size, 2KB in total
  - Variable with strong temporal locality get cached here
- Spatial cache: 8 words cache line size, 6KB in total
  - Variable with strong spatial locality get cached here
- Final memory bandwidth needed: 0.26B/cycle
- Maintained the same hit rate: 95%

We do achieve a much SMALLER bandwidth with NO degeneration on performance,

how this contribute to Power Saving then?

A single case

Power comparison - p1

- For off chip **memory access**
- dram_power = 2W 8bit(1Byte/cycle), keeping access dram @100MHz
- bus_power = 0.6W 8bit(1Byte) bus, being active every cycle @100MHz
- mem_power = bandwidth * (dram_power + bus_power)

Cpu is not visiting memory all the time

Hence its unit is Byte/cycle, it can be understood as **the percentage utilization** of the memory port

Less use means less power consumption

- bandwidth_tradition(0.32) > bandwidth_spilt(0.26)
- Hence: **mem_power_tradtion > mem_power_split**