

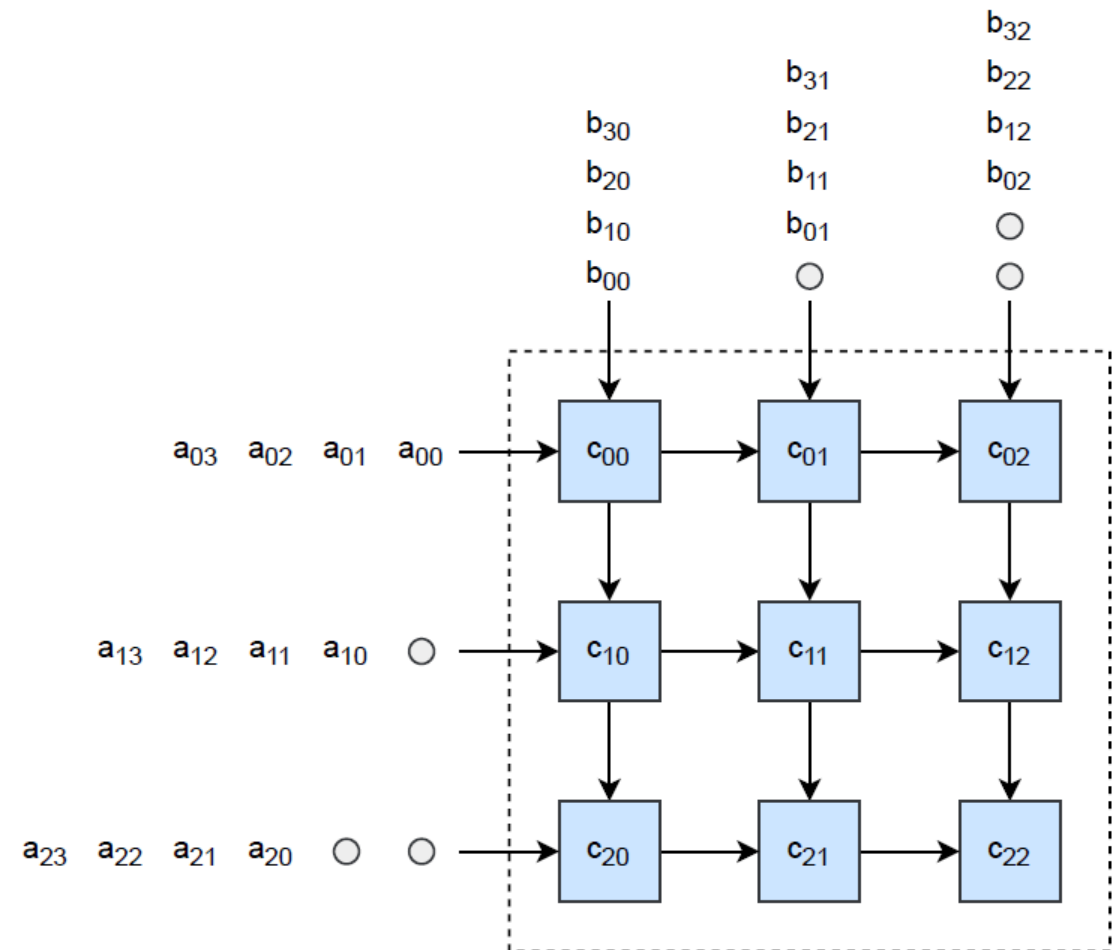
# MemPool meets Systolic

Samuel Riedel  
Matheus Cavalcante  
Prof. Luca Benini



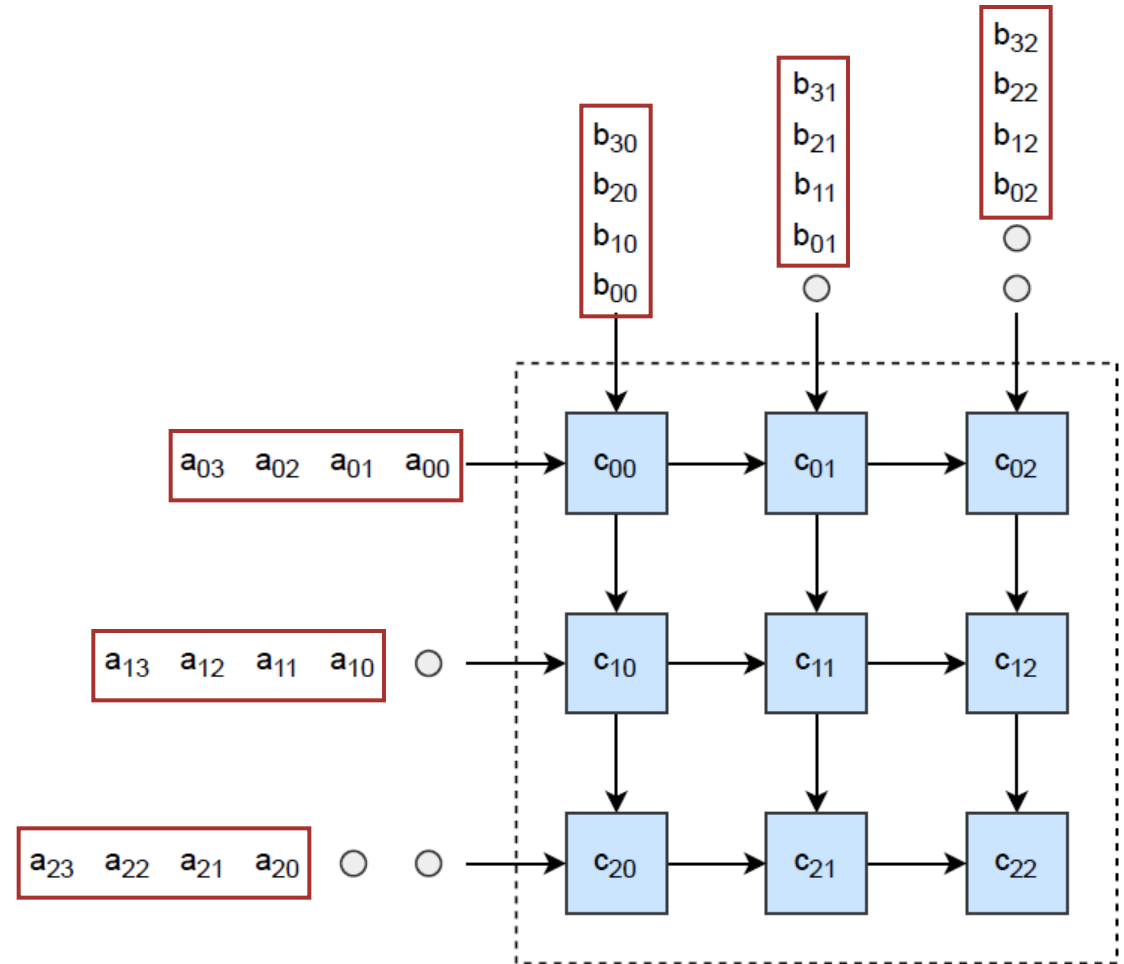
# Baseline systolic matrix multiplication

- Per element, each core does:
  - 2 pop
  - 1 MAC
  - 2 push
- More queue operations than MACs
- → Software queue could handle multiple matrix elements per queue element



# Improved systolic matrix multiplication

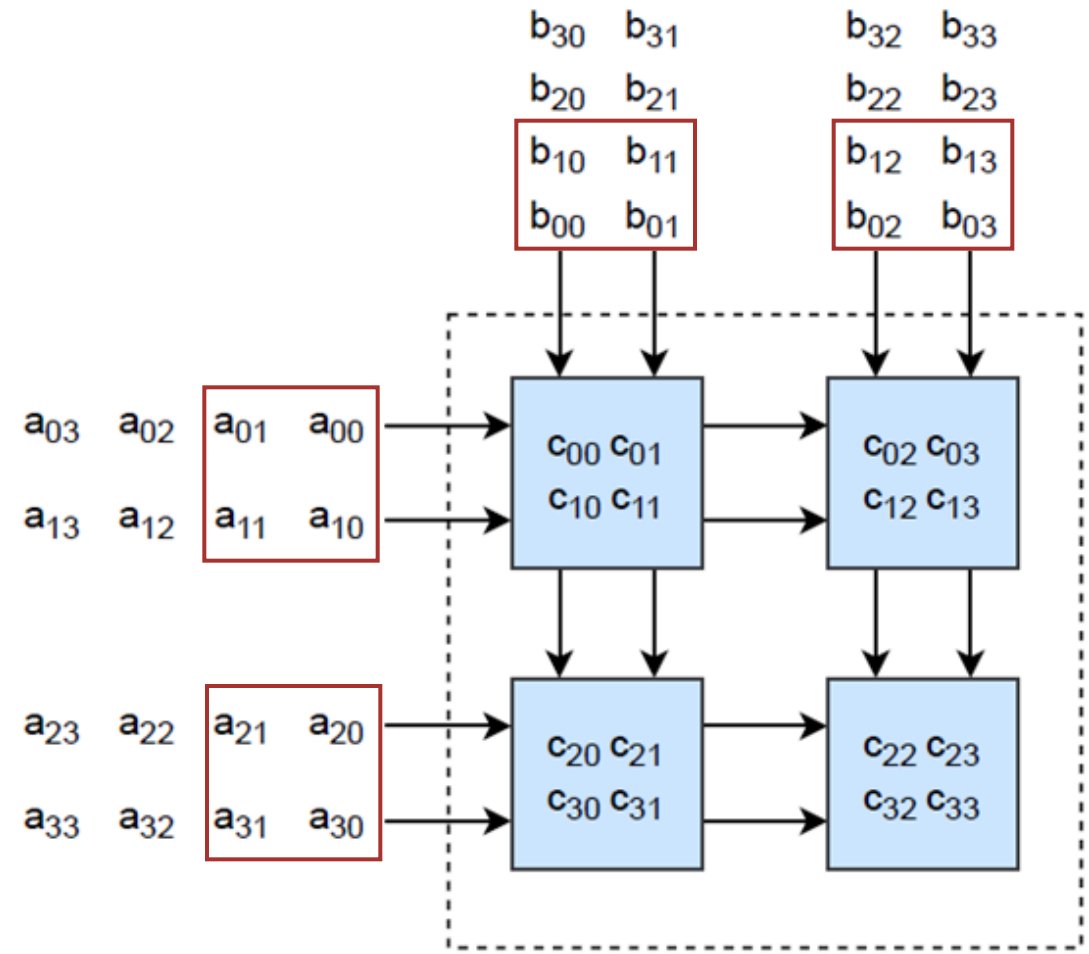
- Pop/push multiple elements at once
- Per four elements, each core does:
  - 2 pop
  - 4 MAC
  - 2 push
- Just as many MACs as queue operations
- ➔ 2.78 times faster than previously





# Improved computational intensity

- Pop/push multiple elements at once
- Per four element, each core does:
  - 2 pop
  - 8 MAC
  - 2 push
- Half the number of queue operations
- ➔ 1.86 times faster than previously



# Outlook

- Clean up and merge
- Add second benchmark
- Start on hardware extensions

