

SAFARI

DSPATCH: DUAL SPATIAL PATTERN PREFETCHER

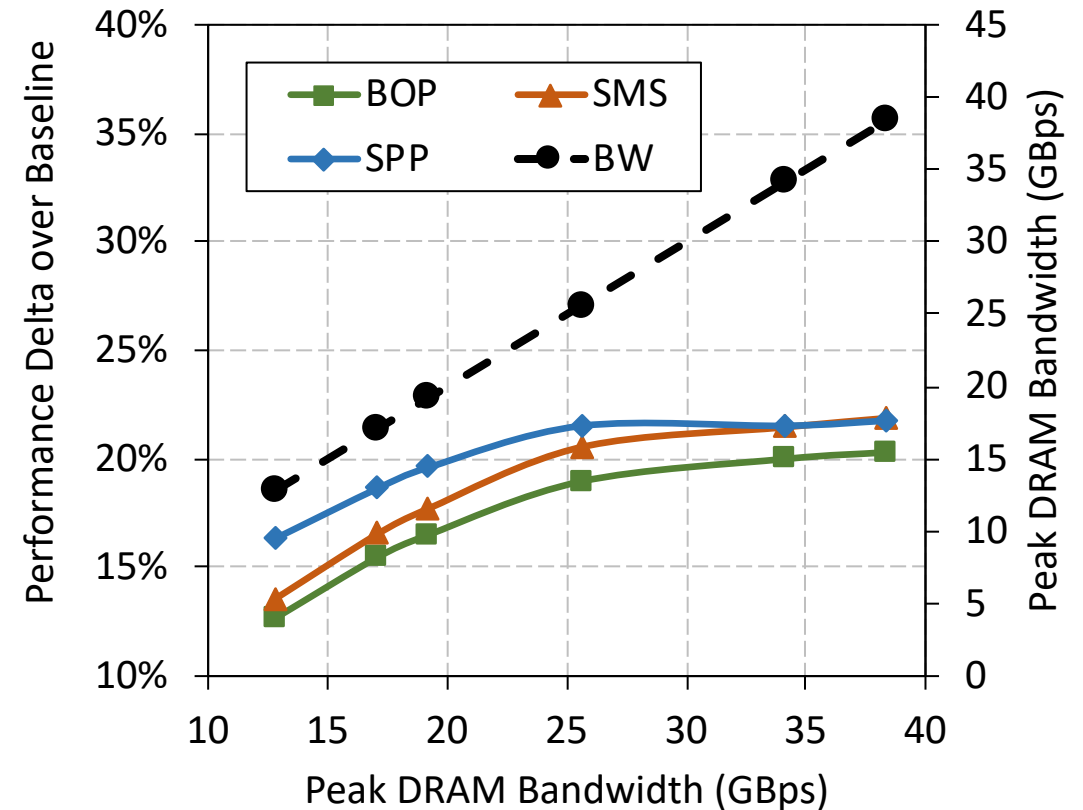
Rahul Bera¹, Anant V. Nori¹, Onur Mutlu², Sreenivas Subramoney¹

¹Processor Architecture Research Lab, Intel Labs

²ETH Zürich

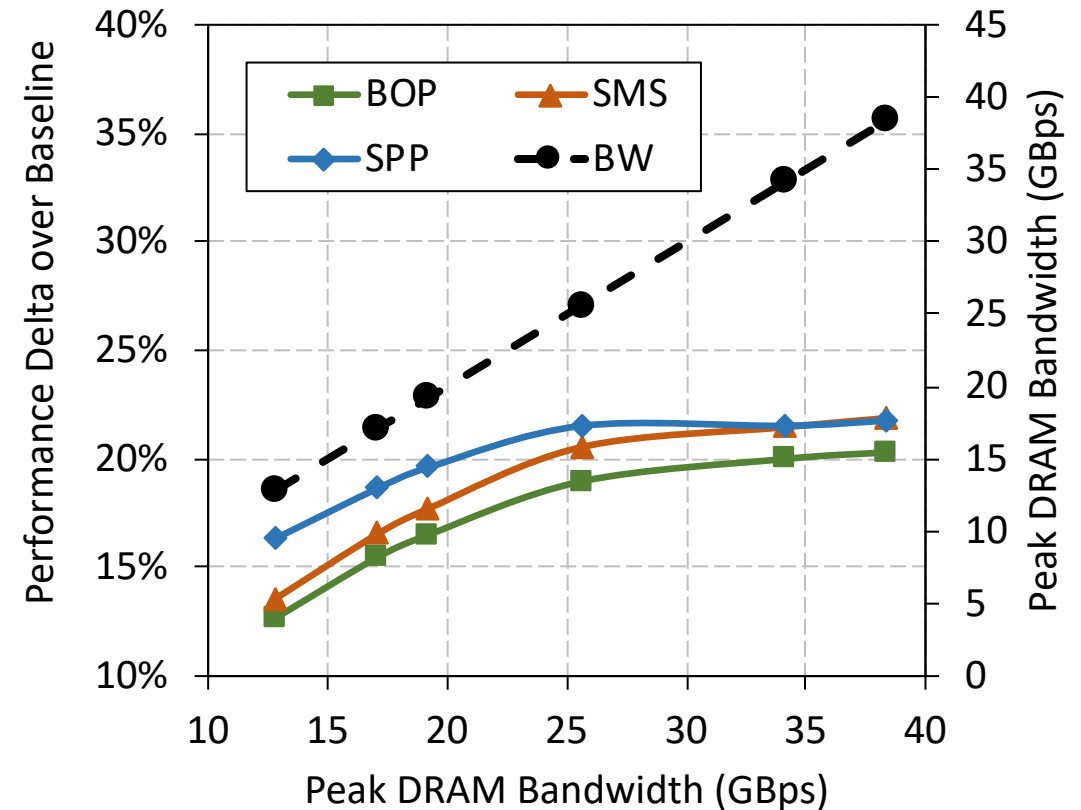
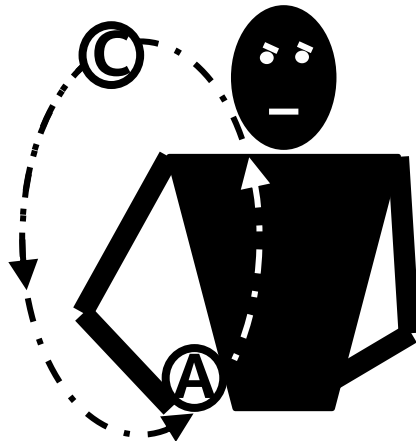
Motivation and Challenge

- DRAM bandwidth increases every generation
- Prefetchers need to adapt in speculation and coverage to utilize this valuable resource



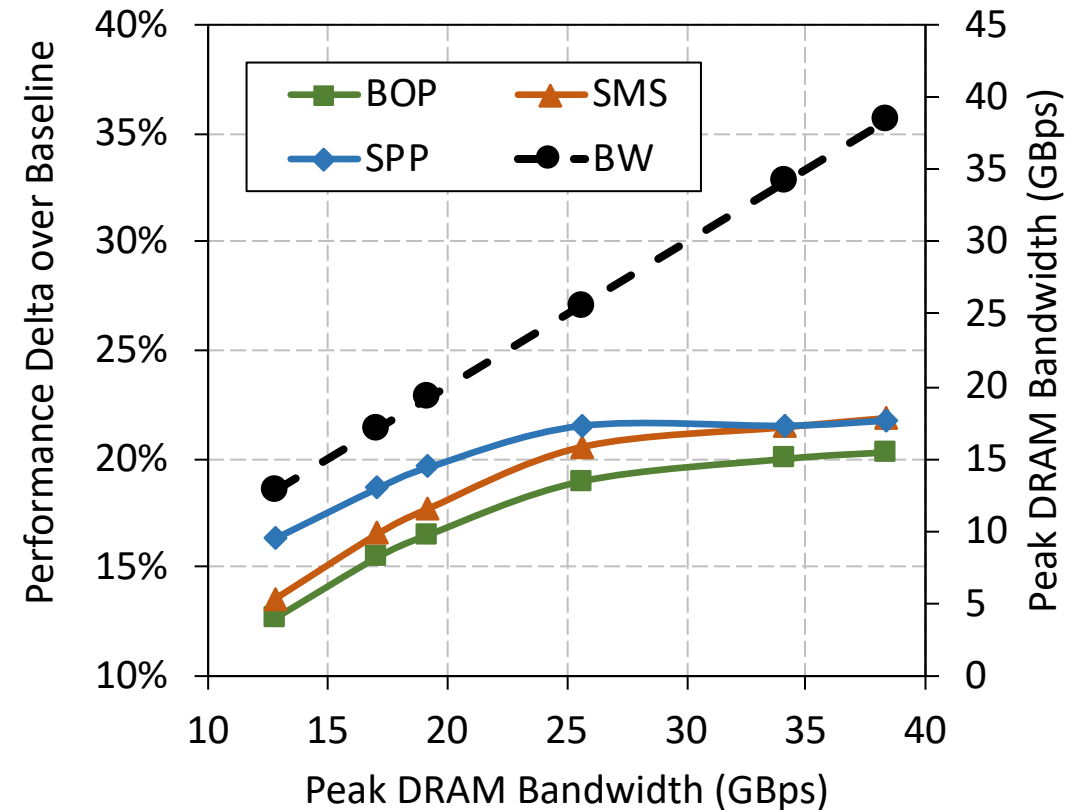
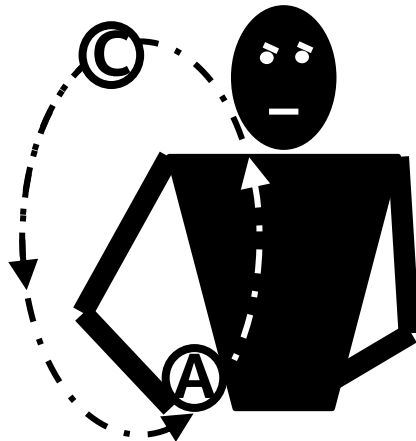
Motivation and Challenge

- DRAM bandwidth increases every generation
- Prefetchers need to adapt in speculation and coverage to utilize this valuable resource
- Traditionally, prefetchers have juggled between coverage and accuracy



Motivation and Challenge

- DRAM bandwidth increases every generation
- Prefetchers need to adapt in speculation and coverage to utilize this valuable resource
- Traditionally, prefetchers have juggled between coverage and accuracy



Need to *boost Coverage* while *simultaneously optimizing* for Accuracy

Dual Spatial Pattern Prefetcher

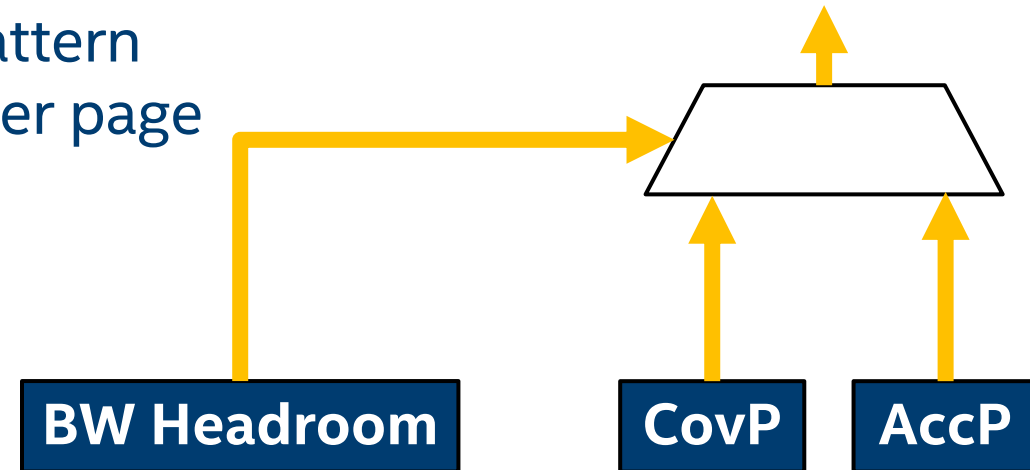
- Simultaneously learn **two** spatial bit-pattern representations of program accesses per page
 - Coverage-Biased
 - Accuracy-Biased

CovP

AccP

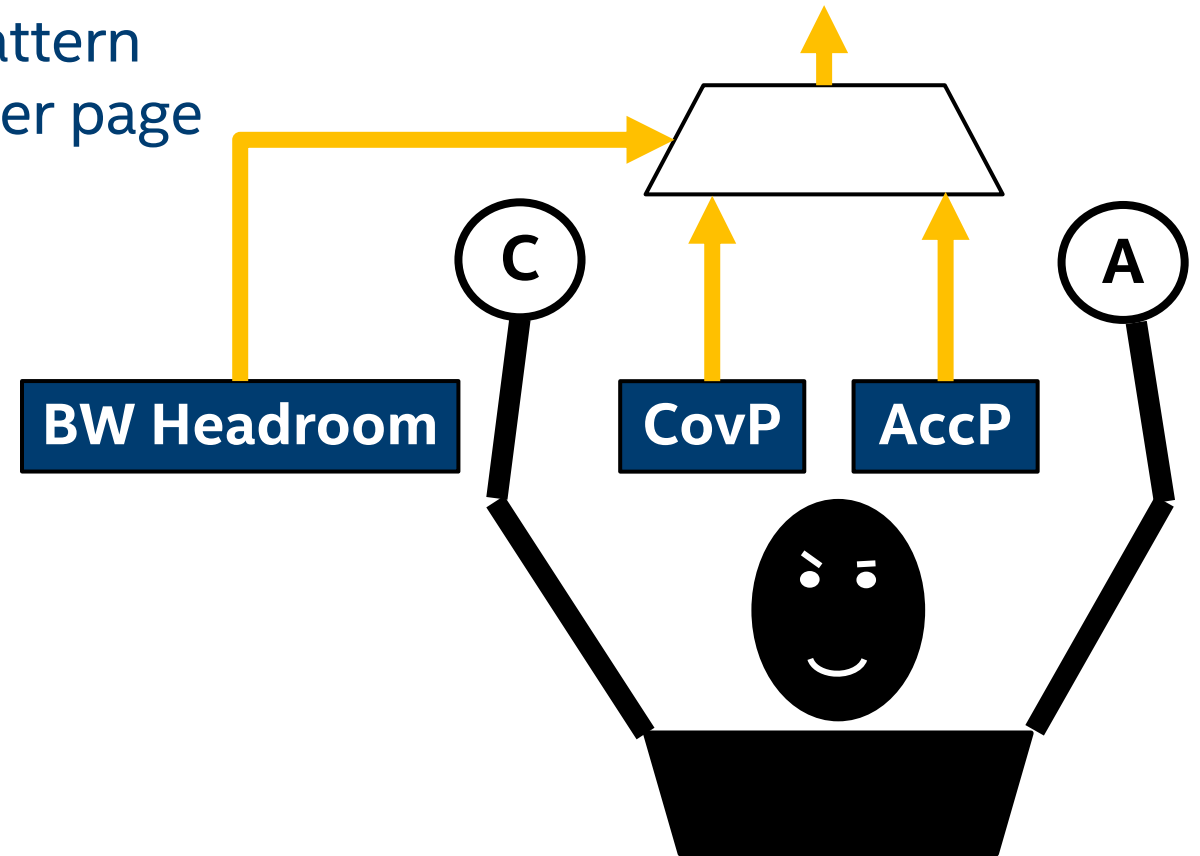
Dual Spatial Pattern Prefetcher

- Simultaneously learn **two** spatial bit-pattern representations of program accesses per page
- Coverage-Biased
- Accuracy-Biased
- Predict **one** of the patterns based on current DRAM bandwidth headroom



Dual Spatial Pattern Prefetcher

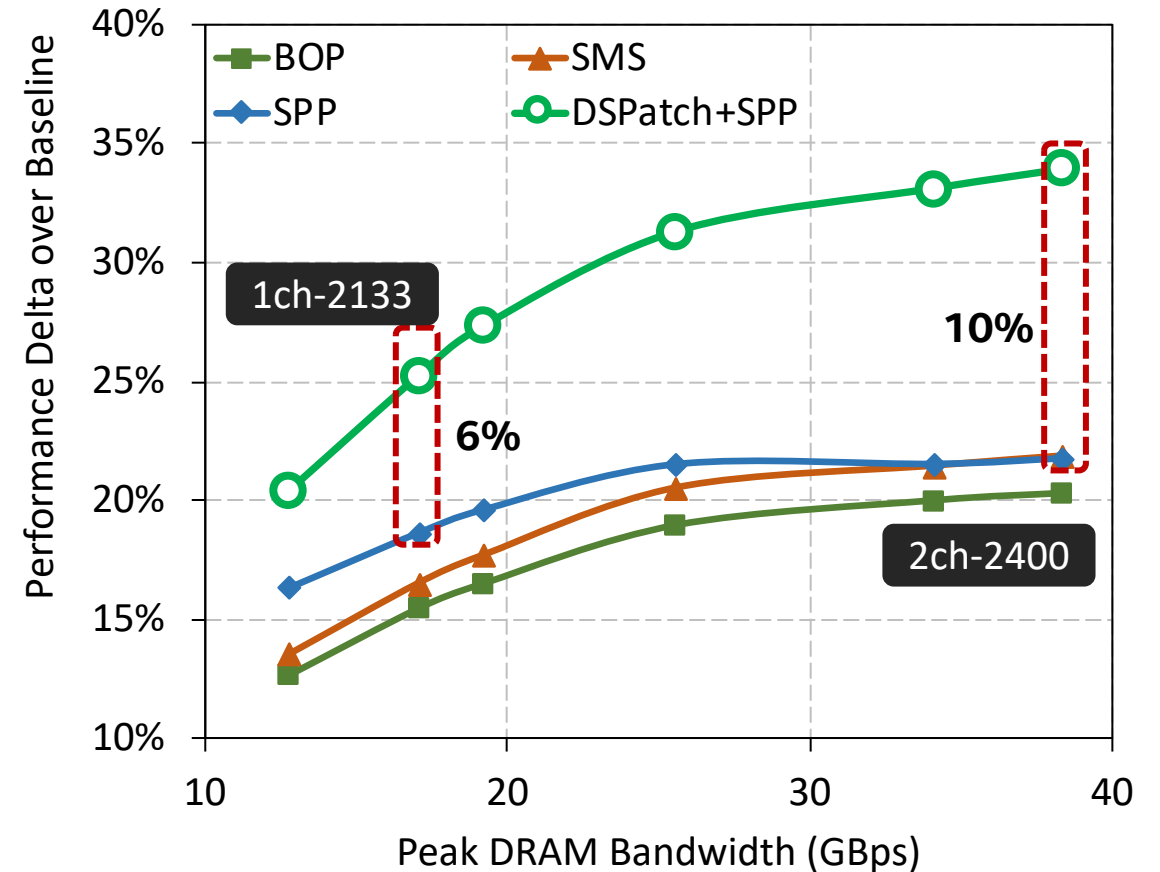
- Simultaneously learn **two** spatial bit-pattern representations of program accesses per page
- Coverage-Biased
- Accuracy-Biased
- Predict **one** of the patterns based on current DRAM bandwidth headroom

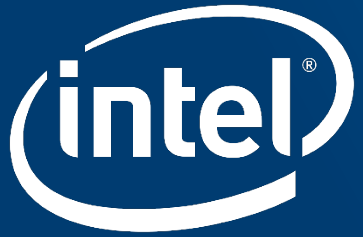


DSPatch can boost Coverage while *simultaneously* optimizing for Accuracy

Key Results Summary

- **6% average speedup** over baseline with PC-stride @ L1 and SPP @ L2
- **10% average speedup** if DRAM bandwidth is doubled
- **3.6 KB** of hardware storage





SAFARI

DSPATCH: DUAL SPATIAL PATTERN PREFETCHER

Rahul Bera¹, Anant V. Nori¹, Onur Mutlu², Sreenivas Subramoney¹

¹Processor Architecture Research Lab, Intel Labs

²ETH Zürich