

Berti: An Accurate Local-Delta Data Prefetcher



Agustín Navarro-Torres¹, Biswabandan Panda²,
Jesús Alastruey-Benedé¹, Pablo Ibáñez¹, Víctor Viñals-Yúfera¹,
Alberto Ros³

¹Universidad de Zaragoza
{**agusnt**, jalastru, imarin, victor}@unizar.es

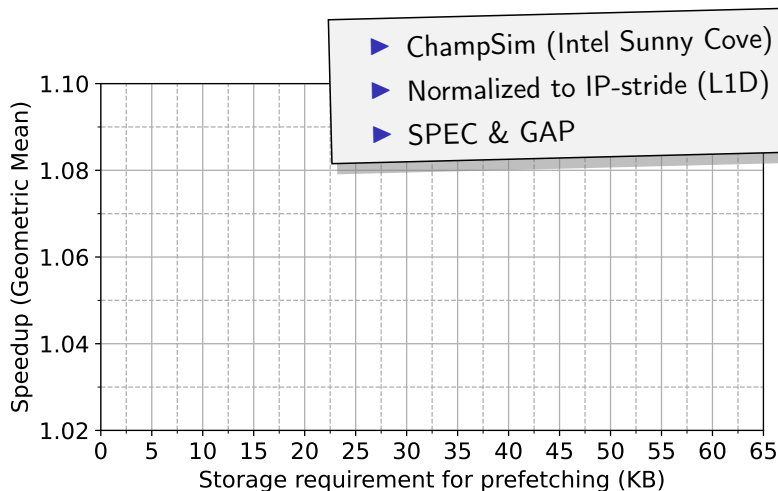
²Indian Institute of Technology Bombay
biswa@cse.iitb.ac.in

³University of Murcia
aros@ditec.um.es

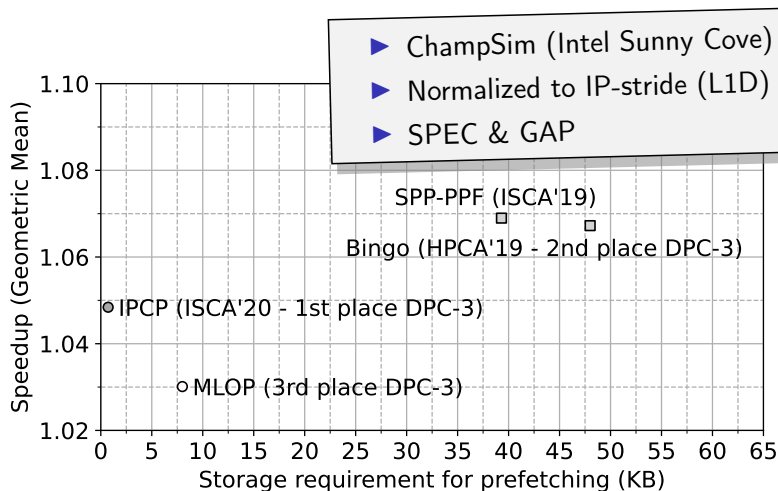
MICRO'55, Session Microarchitecture II, October 4th, 2022

- ▶ Accurate and timely local delta L1D prefetcher
- ▶ Outperforms state-of-the-art prefetchers
- ▶ Only 2.55 KB of storage

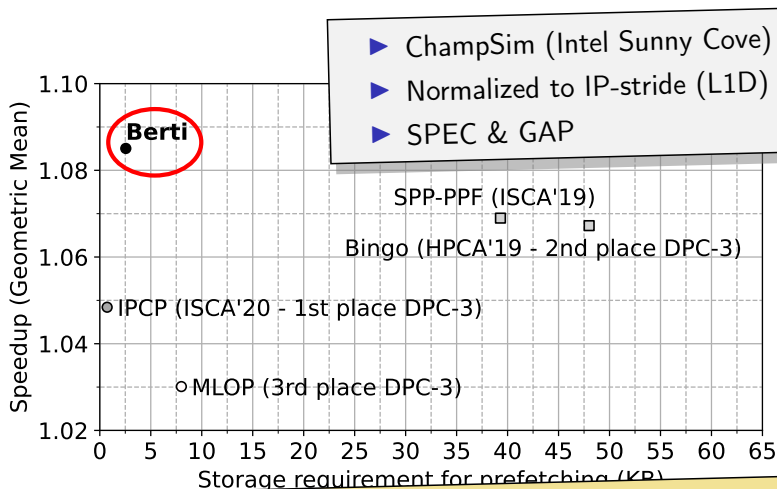
- ▶ Accurate and timely local delta L1D prefetcher
- ▶ **Outperforms state-of-the-art prefetchers**
- ▶ **Only 2.55 KB of storage**



DPC-3 = 3rd Data Prefetching Championship



DPC-3 = 3rd Data Prefetching Championship

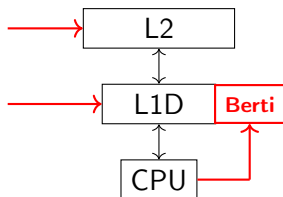


Berti outperforms state-of-the-art prefetchers with less storage

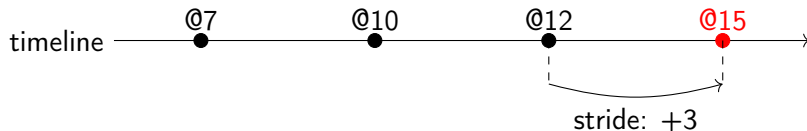
DPC-3 = 3rd Data Prefetching Championship

BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER

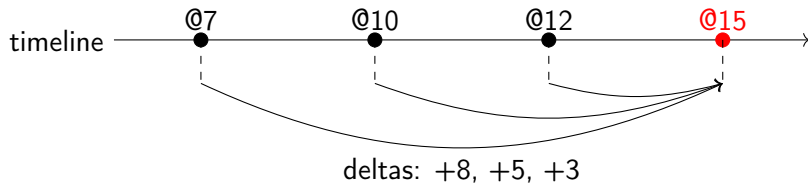
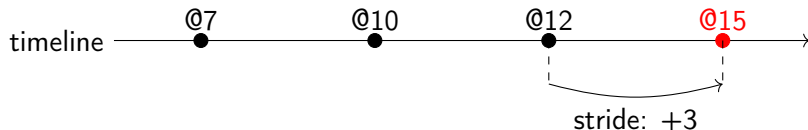
- ▶ Orchestrates prefetch requests across L1D/L2
- ▶ Instruction Pointer (IP) and unfiltered memory references
- ▶ Virtual addresses: cross-page prefetching



Definition of delta

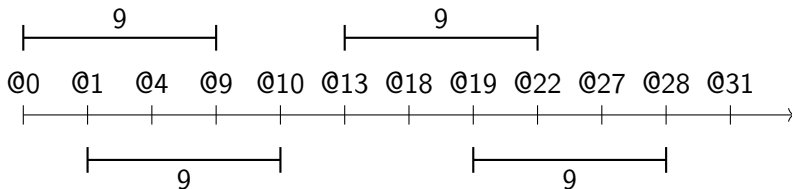


Definition of delta



BERTI: ACCURATE AND TIMELY LOCAL **DELTA** L1D PREFETCHER

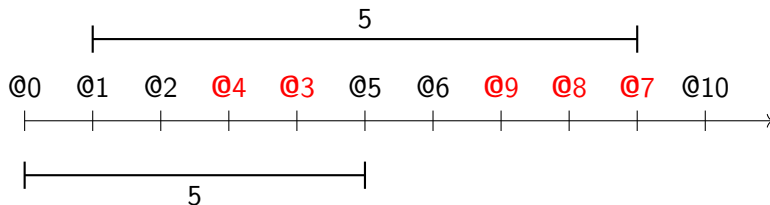
Stride: **+1, +3, +5**



With which delta should I prefetch?
 $\text{delta} = 1 + 3 + 5 = 9 \rightarrow$ **always hit**

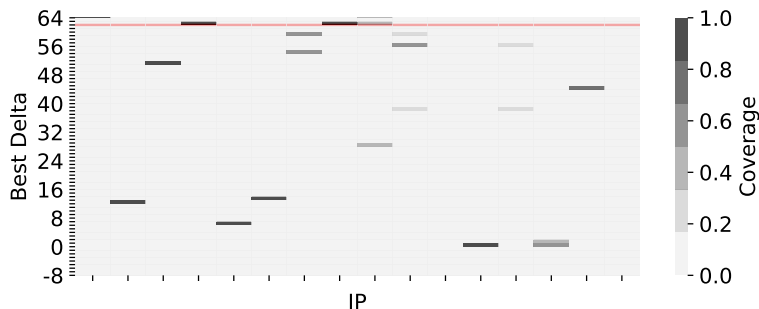
BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER

Addresses reordered by **out-of-order** processor



Stride prefetch requires specific order
We can prefetch with $\text{delta} = 5$, for example

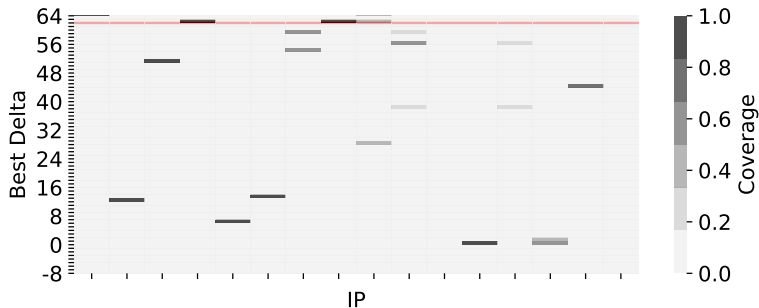
mcf-1554B



- ▶ Red line: best delta by BOP¹, coverage: 2%
- ▶ Black lines: per-IP local deltas, coverage: 10%

¹Winner of 2nd Data Prefetching Championship (DPC2)

mcf-1554B



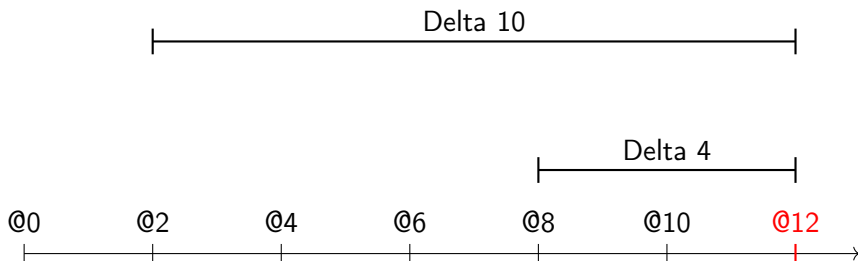
- ▶ Red line: best delta by [unclear]
- ▶ Black lines: per-IP local

Every IP can have its own deltas!

¹Winner of 2nd Data Prefetching Championship (DPC2)

BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER

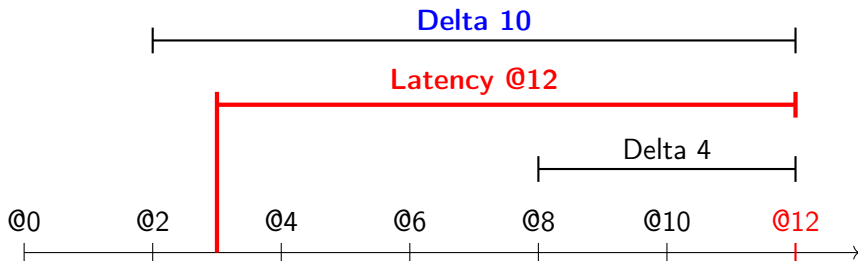
Stride **+2**



How far in advance should I prefetch address 12?

BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER

Stride +2



How far in advance should I prefetch address 12?
Depends on its latency

TRAINING

1. Measure fetch latency
2. Learn timely and accurate deltas
3. Compute coverage of deltas

History table		
IP	@	Time
A	2	0
A	5	30
B	10	50

Table of deltas			
IP	Delta	Coverage	Destination

TRAINING

1. Measure fetch latency
2. Learn timely and accurate deltas
3. Compute coverage of deltas

History table			Table of deltas			
IP	@	Time	IP	Delta	Coverage	Destination
A	2	0				
A	5	30				
B	10	50				
A	12	70				

+10



TRAINING

1. Measure fetch latency
2. Learn timely and accurate deltas
3. Compute coverage of deltas

History table		
IP	@	Time
A	2	0
A	5	30
B	10	50
A	12	70

Table of deltas			
IP	Delta	Coverage	Destination
A	+10	1/1 (100%)	

TRAINING

1. Measure fetch latency
2. Learn timely and accurate deltas
3. Compute coverage of deltas

History table			Table of deltas			
IP	@	Time	IP	Delta	Coverage	Destination
A	2	0	A	+10	2/2 (100%)	
A	5	30	A	+13	1/2 (50%)	
B	10	50				
A	12	70				
A	15	140				

+10, +13



ISSUING PREFETCH REQUESTS

1. Select deltas
2. Orchestration

History table		
IP	@	Time
A	2	0
A	5	30
B	10	50
A	12	70
A	15	140

Table of deltas			
IP	Delta	Coverage	Destination
A	+10	2/2 (100%)	
A	+13	1/2 (50%)	

ISSUING PREFETCH REQUESTS

1. Select deltas
2. Orchestration

History table		
IP	@	Time
A	2	0
A	5	30
B	10	50
A	12	70
A	15	140

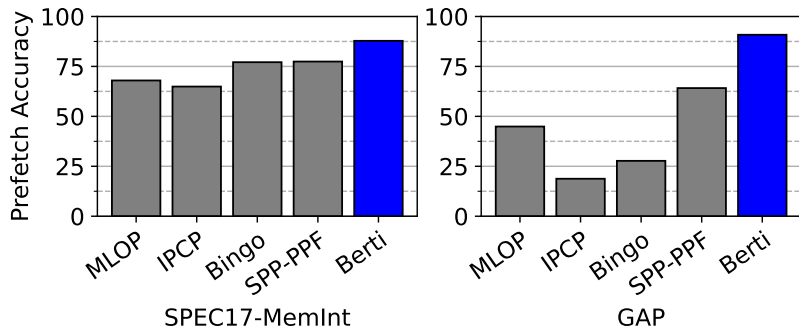
Table of deltas			
IP	Delta	Coverage	Destination
A	+10	2/2 (100%)	L1D
A	+13	1/2 (50%)	L2

Coverage

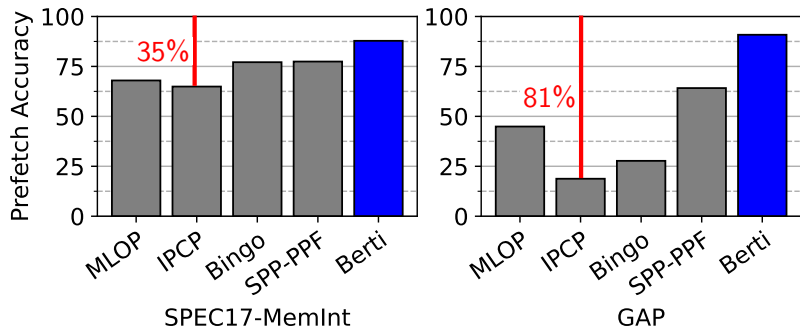
> **65%** → **L1D**

> **35%** → **L2**

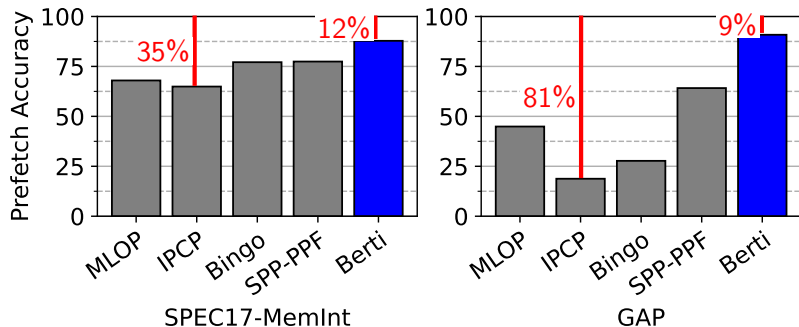
BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER



BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER



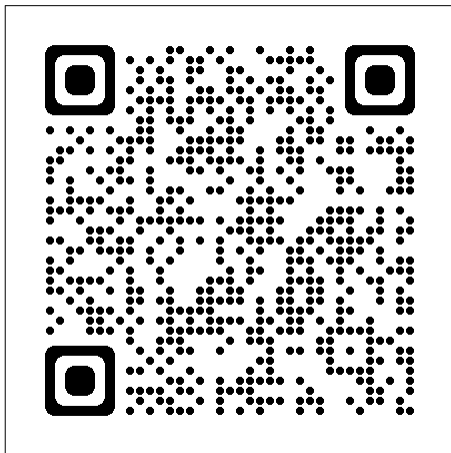
BERTI: ACCURATE AND TIMELY LOCAL DELTA L1D PREFETCHER



Improved accuracy reduces
BW overhead and L1D pollution

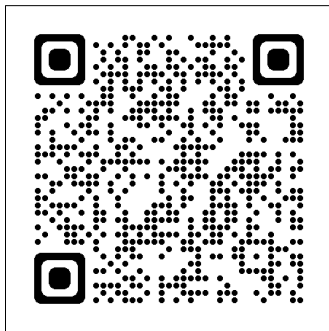
BERTI: CONCLUSIONS

- ▶ **Accurate** and **timely local delta L1D** prefetcher
- ▶ Outperforms state-of-the-art prefetchers
- ▶ Learns the best **deltas** to prefetch
- ▶ **Artifact available** ([Webpage QR](#))

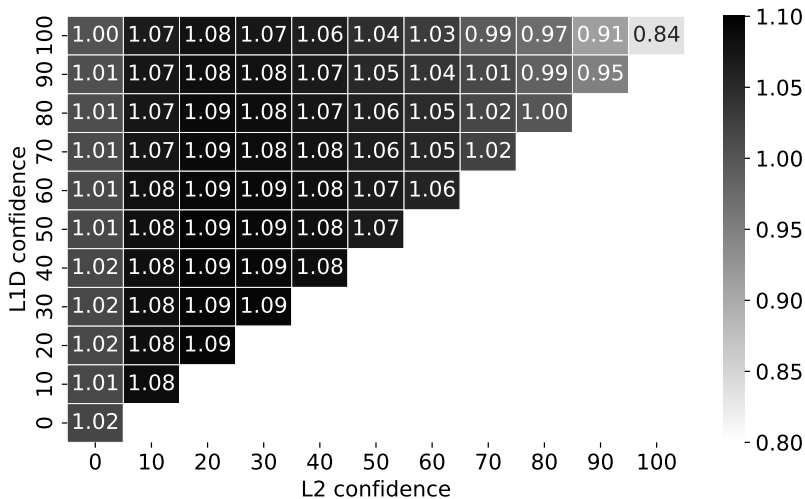


Thanks!
Questions?

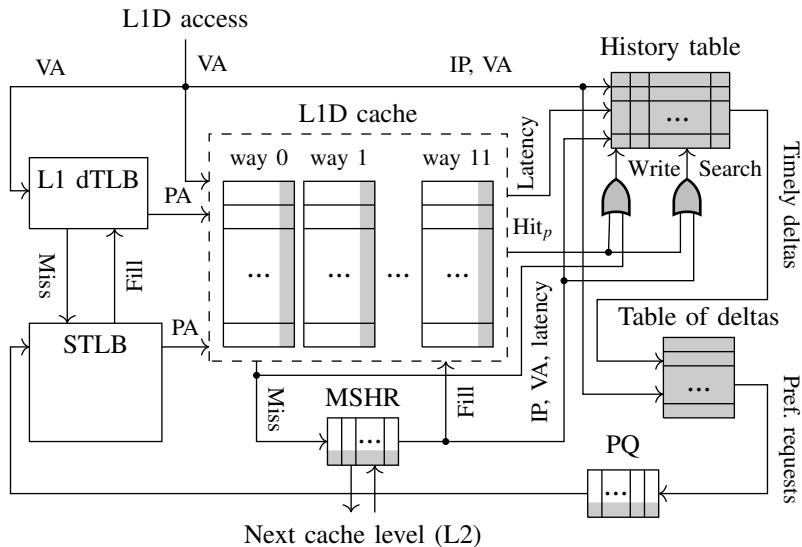
agusnt@unizar.es



WATERMARK SENSITIVITY



HARDWARE IMPLEMENTATION



HISTORY TABLE

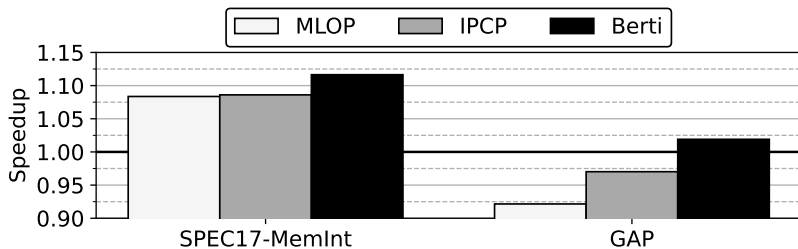
	IP tag	line address	timestamp
History table	7	24	16

DELTA TABLE

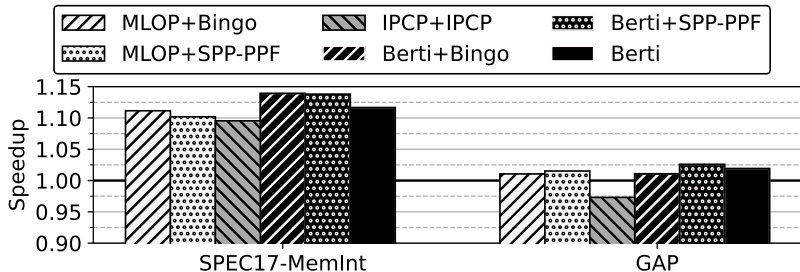
Table of deltas

IP tag	counter	delta	coverage status			delta	coverage status	
10	4	13	4	2	...	13	4	2

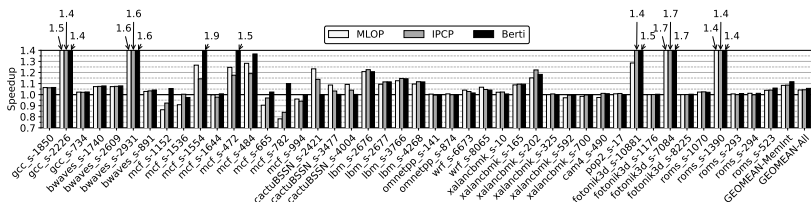
L1D PREFETCHERS PERFORMANCE



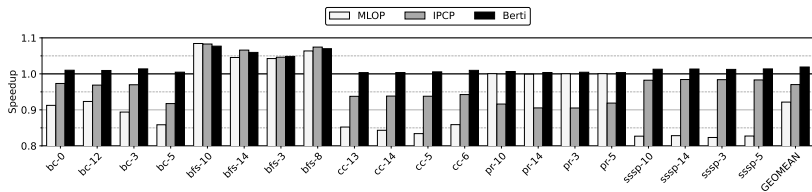
MULTI-LEVEL PREFETCHERS PERFORMANCE



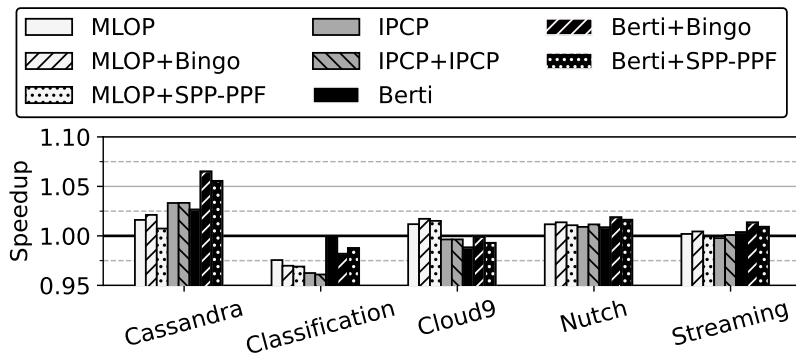
L1D PREFETCHERS SPEC PERFORMANCE



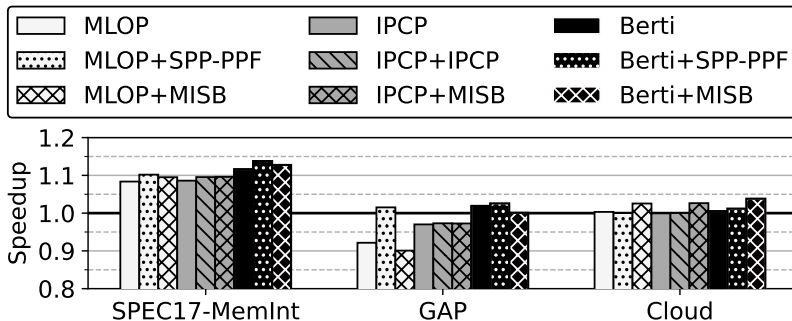
L1D PREFETCHERS GAP PERFORMANCE



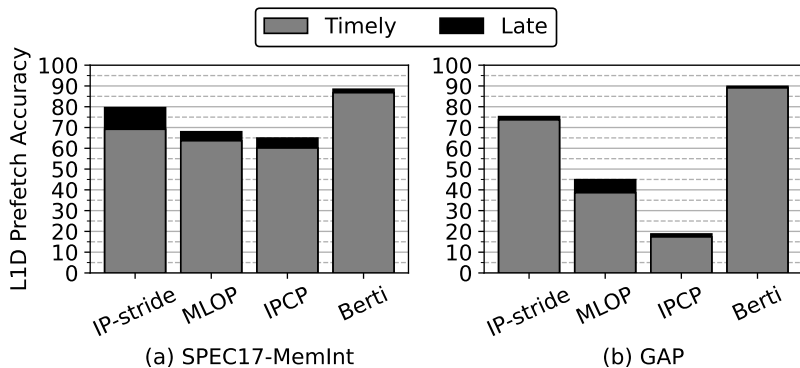
CLOUDSUITE PERFORMANCE



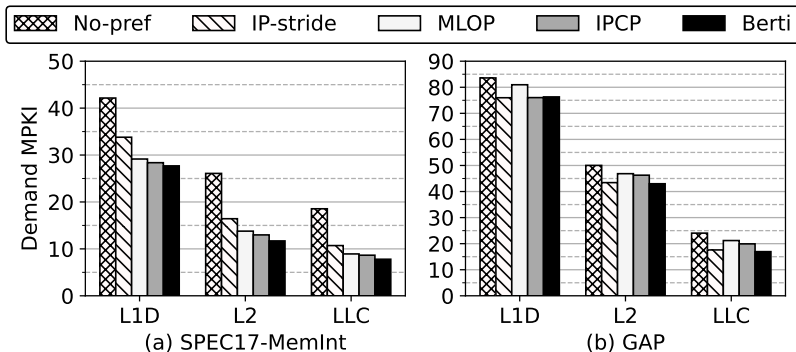
ISB PERFORMANCE



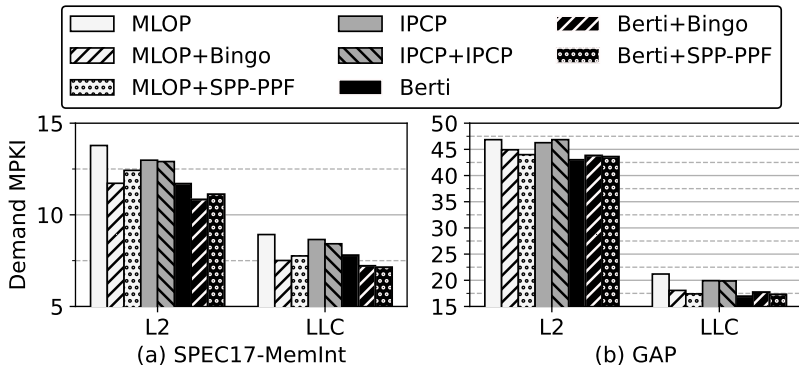
L1D PREFETCHERS ACCURACY

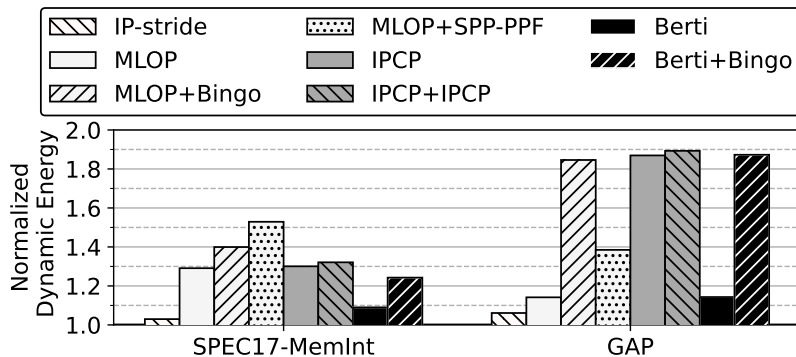


L1D PREFETCHERS DEMAND MPKI

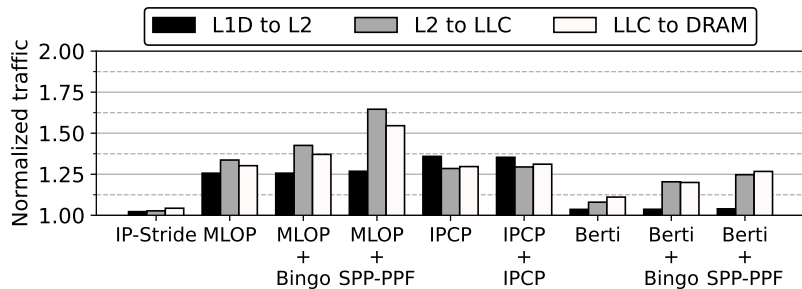


MULTI-LEVEL PREFETCHER DEMAND MPKI

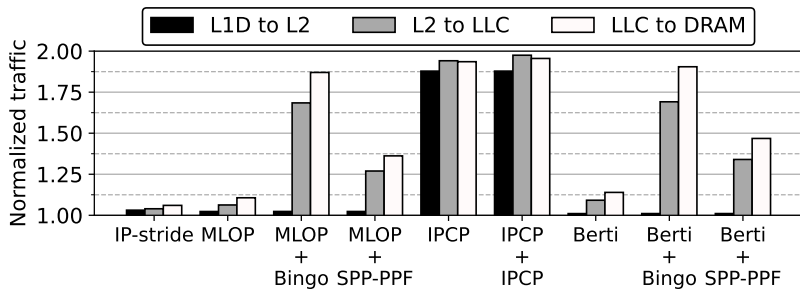




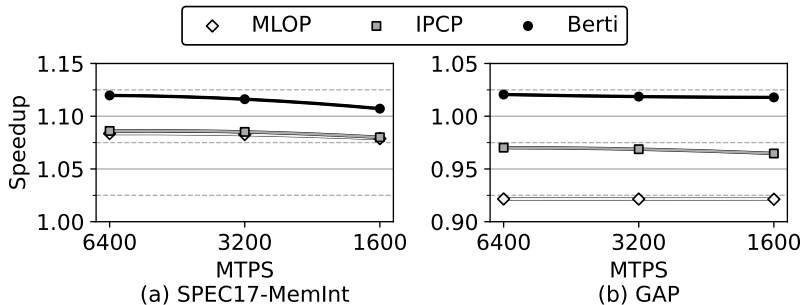
SPEC NORMALIZED TRAFFIC



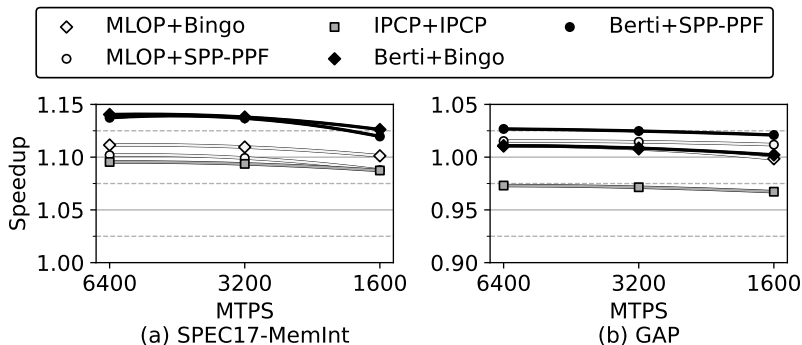
GAP NORMALIZED TRAFFIC



L1D PREFETCHER LOW BANDWIDTH



MULTI-LEVEL PREFETCHER LOW BANDWIDTH



PERFORMANCE VS. SIZE

