Predicting Coherence Communication by Tracking Synchronization Points at Run Time

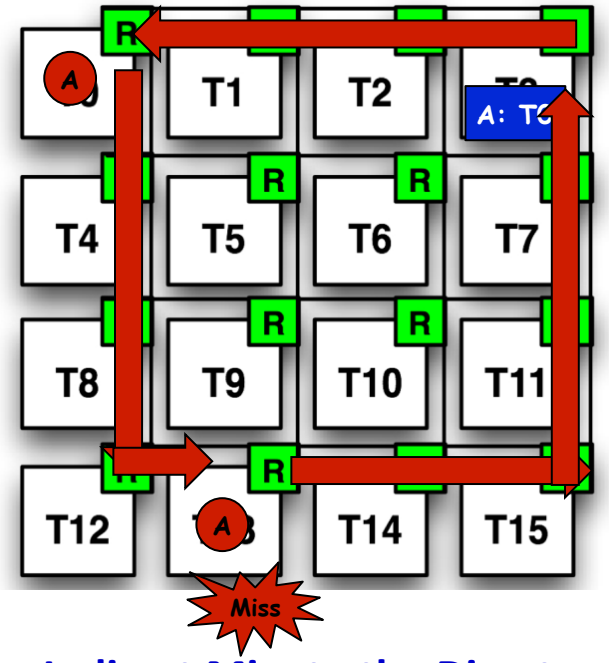
Socrates Demetriades
and Sangyeun Cho

University of Pittsburgh

45th International Symposium in Microarchitecture, December 2012

Communication Overheads

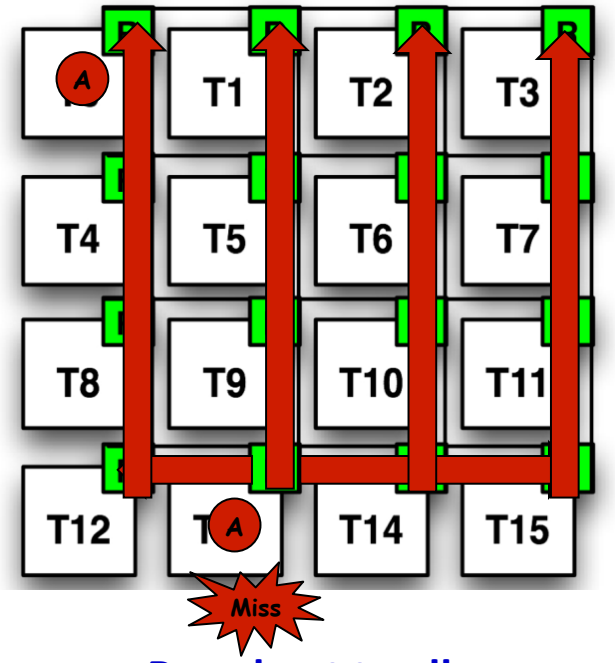
Directory-based Coherence Protocol



Indirect Miss to the Directory

=> Increase Miss Latency

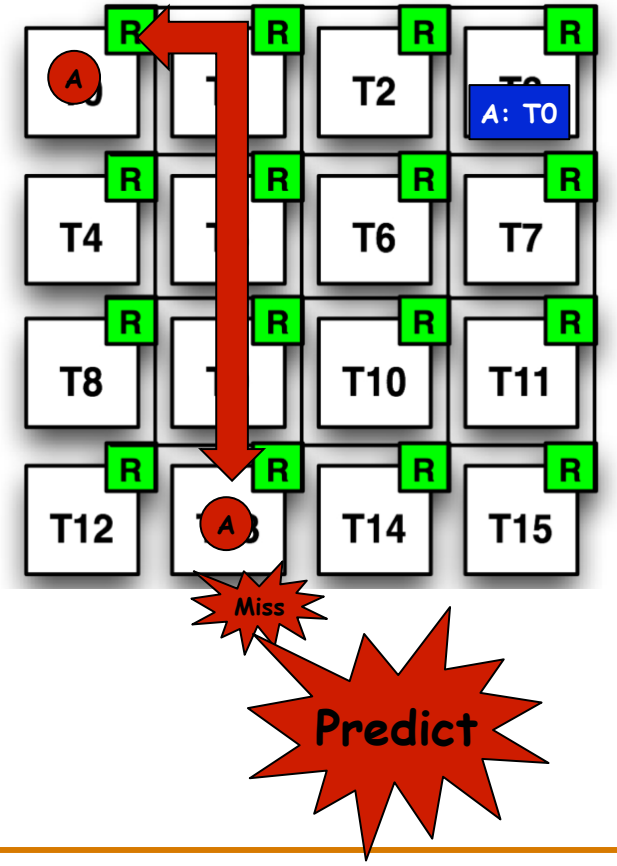
Snoop-based Coherence Protocol

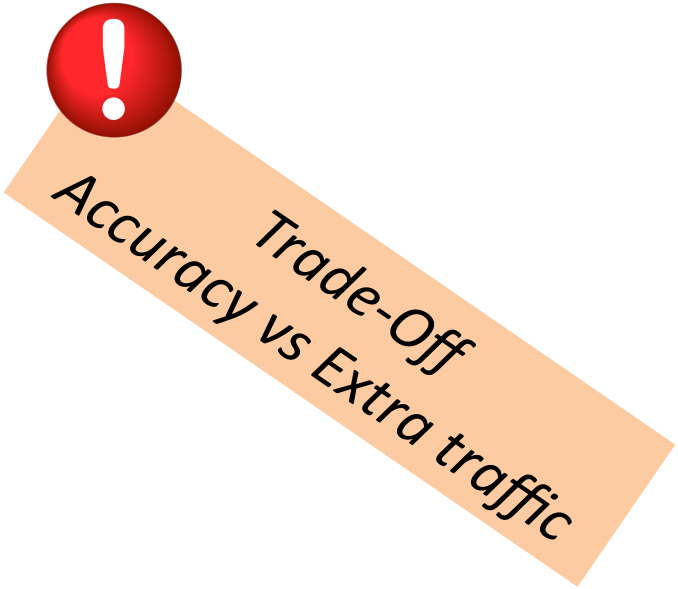


Broadcast to all

=> Increase traffic

Communication Prediction





Contribution

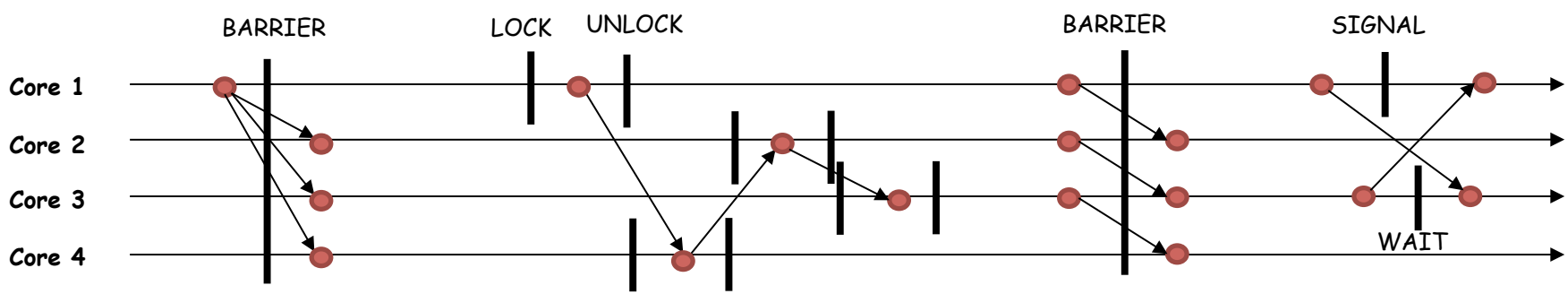
Synchronization Point based Prediction (*SP-prediction*)

Inter-thread communication caused by coherence transactions is tightly related with the synchronization points in parallel execution

- Main Idea: Associate the communication behavior with synchronization points and utilize this association to predict the destination of misses.

- Main Advantage: Has very low storage cost, yet delivers relatively high performance.

Why Synchronization Points?

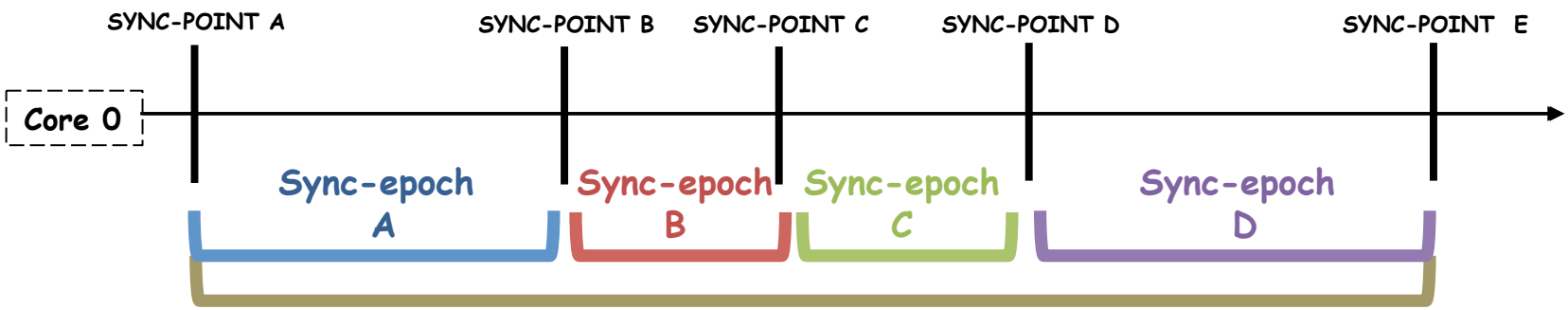


shared data

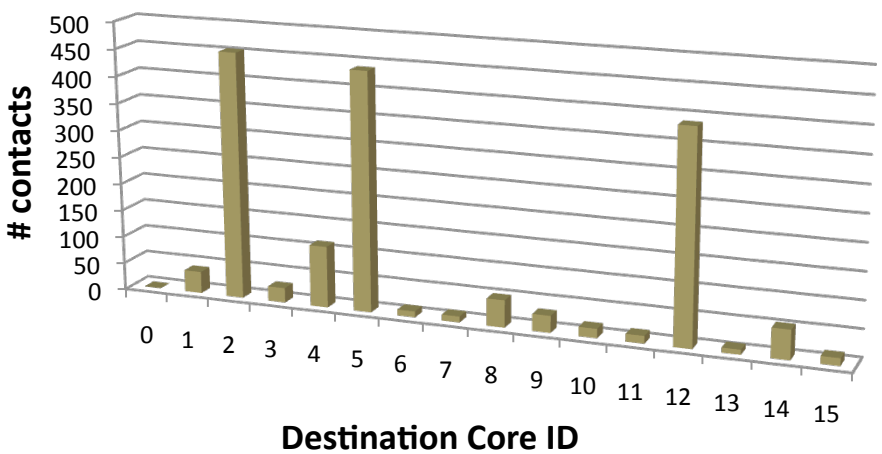
communication direction

[Pthread notation]

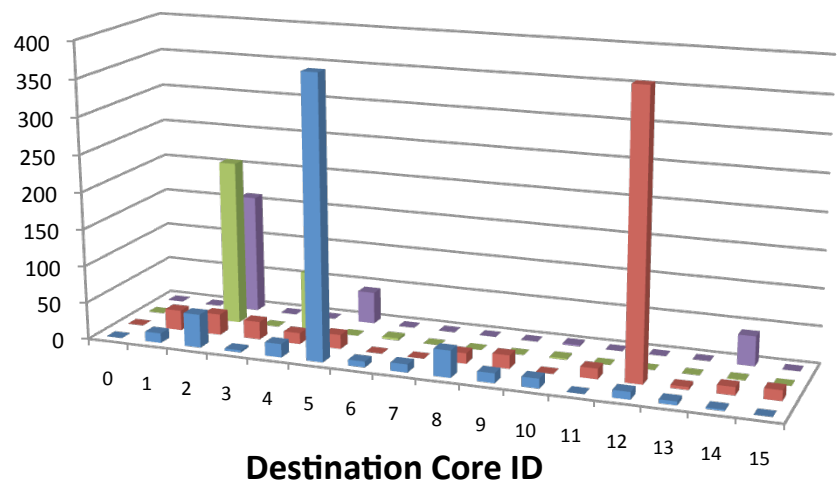
Synchronization Epochs



Communication Distribution of Core 0 (full interval)

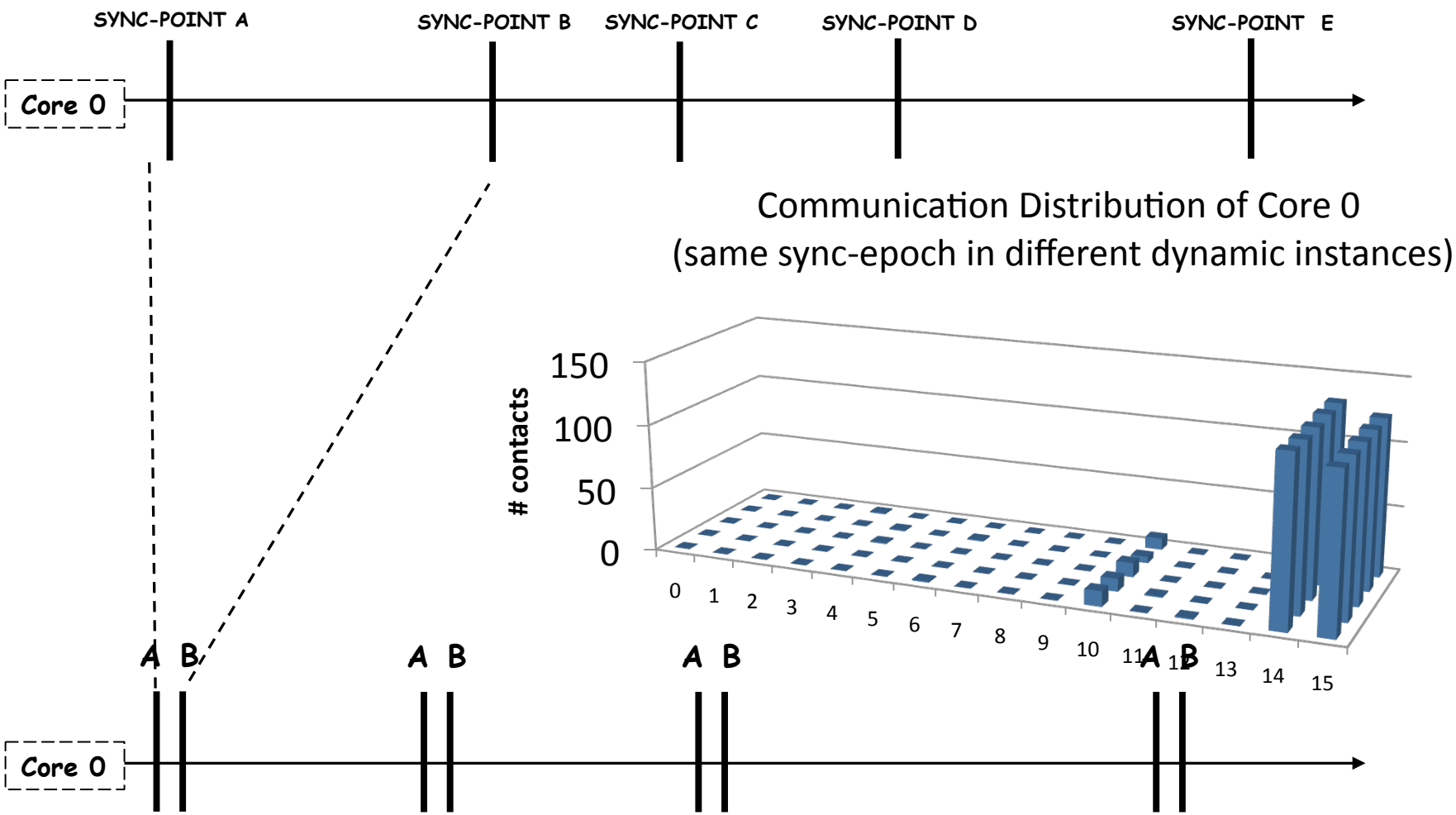


Communication Distribution of Core 0 (different sync-epochs)



[Benchmark: Bodytrack / 16-threads]

Sync-Epoch Dynamic Instances



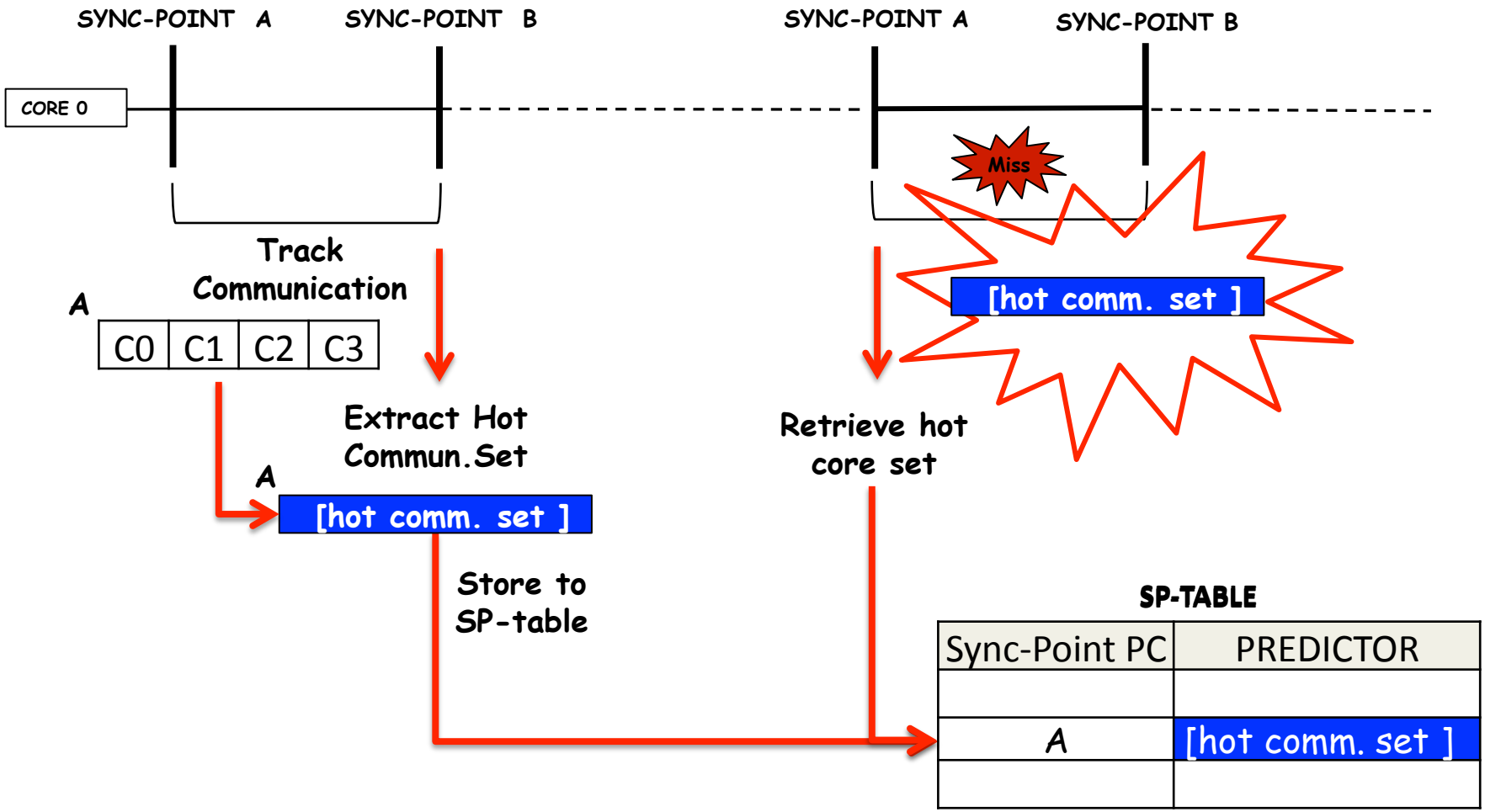
[Benchmark: Bodytrack / 16-threads]

SP-prediction – Overview

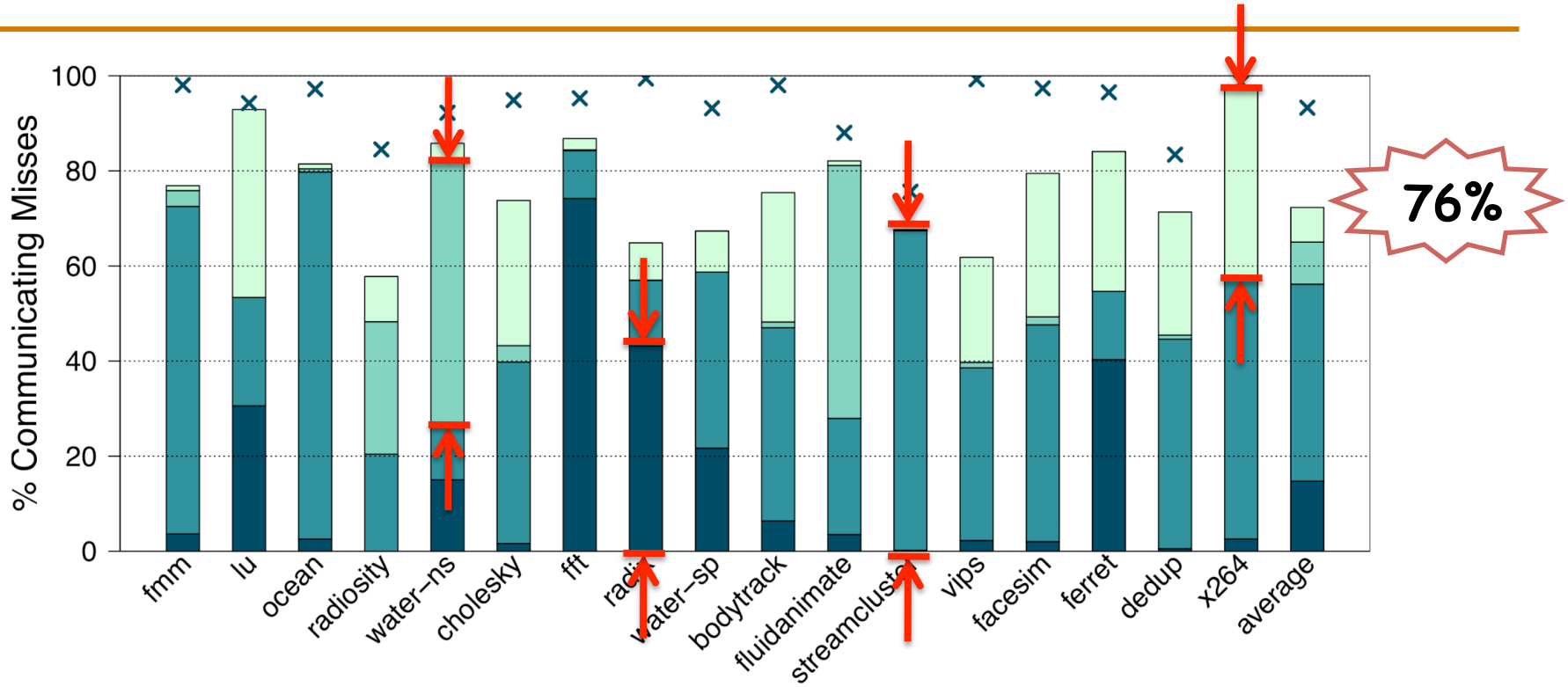
- Monitor destinations of each miss on each core.
- Extract communication signatures for each sync-epoch.
- Store and later reuse those signatures to predict misses in future sync-epoch instances.
- When initial predictions do not exist or are inaccurate, reconstruct the signatures within the sync-epochs.

- Sync Points must be exposed to the hardware so it can sense the beginning and end of sync-epochs.
  - A dedicated instruction must be inserted at the calling location of the synchronization point.
  - PC, lock variable and type must be extracted and pass to a history table.

SP-prediction: History-based Example

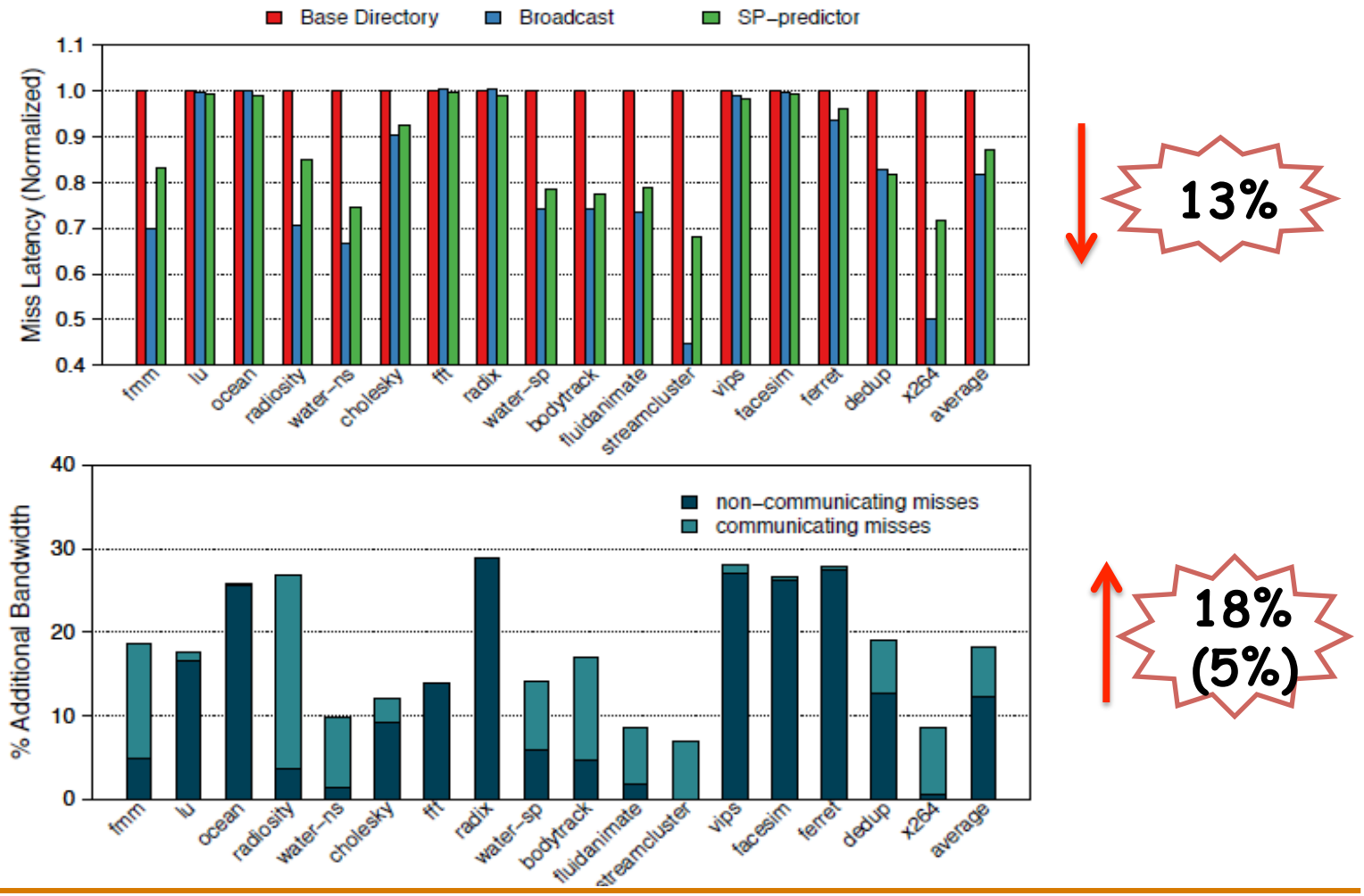


Results: Prediction Accuracy

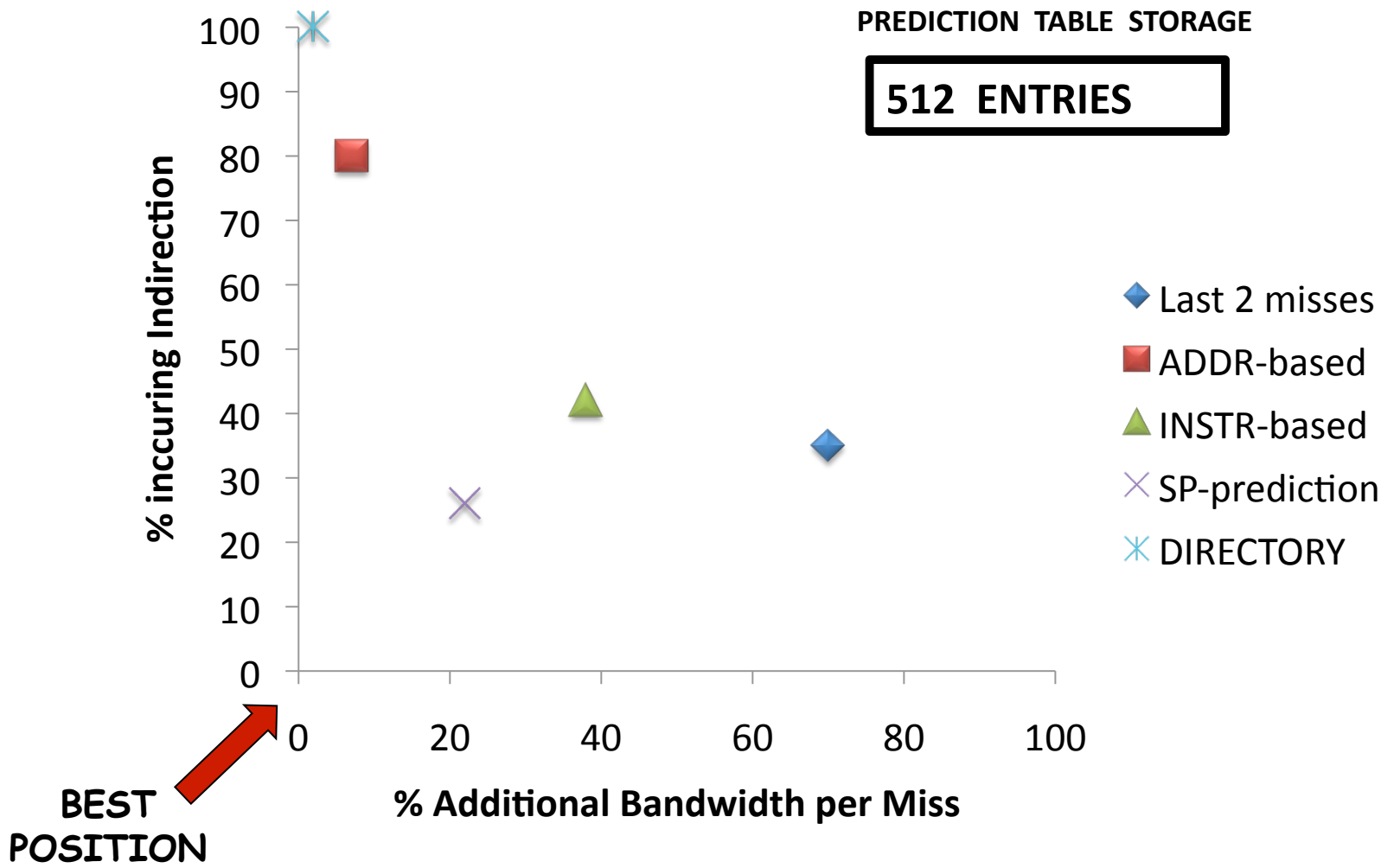


	Average
Destination Set Size (actual)	1.2
SP-prediction Set Size	2.6

Results: Latency & Bandwidth



Comparison with other Predictors



Conclusions

- SP-prediction is a new, run-time approach on coherence communication predictability.
- In contrast to traditional hardware based temporal predictors, exploits localities based on application-defined intervals.
- Promotes very low storage requirements, an important property for emerging CMP implementations.
- Scales independent of core count and cache sizes.
- Takes advantage of the existing shared memory programming paradigm.