arm

DeLorean: Directed Statistical Warming through Time Traveling

Best Paper Nominee

Nikos Nikoleris
Arm Research

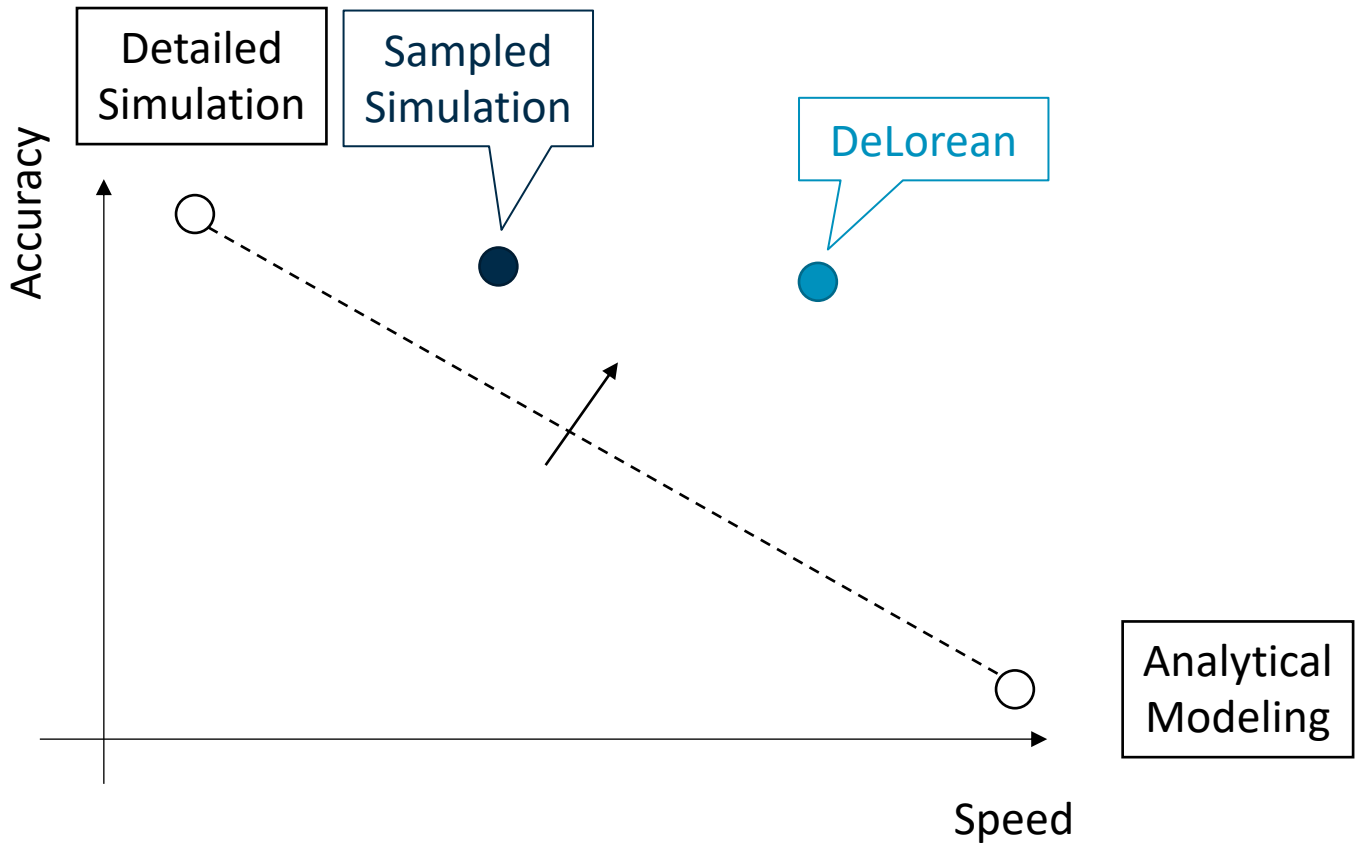
Lieven Eeckhout
Ghent University

Erik Hagersten
Uppsala University

Trevor E. Carlson
National University of Singapore

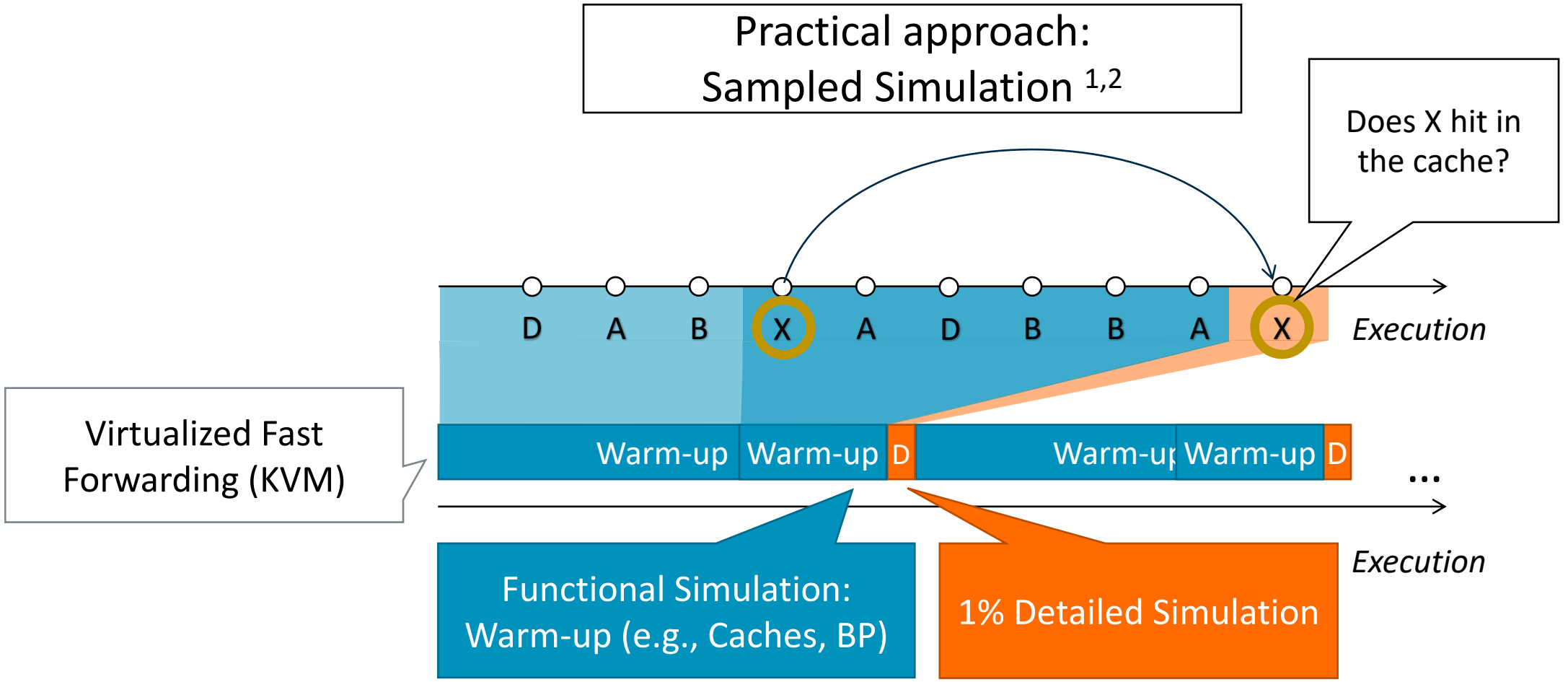
MICRO-52

Accuracy/Speed Trade-off



arm

Sampled Simulation

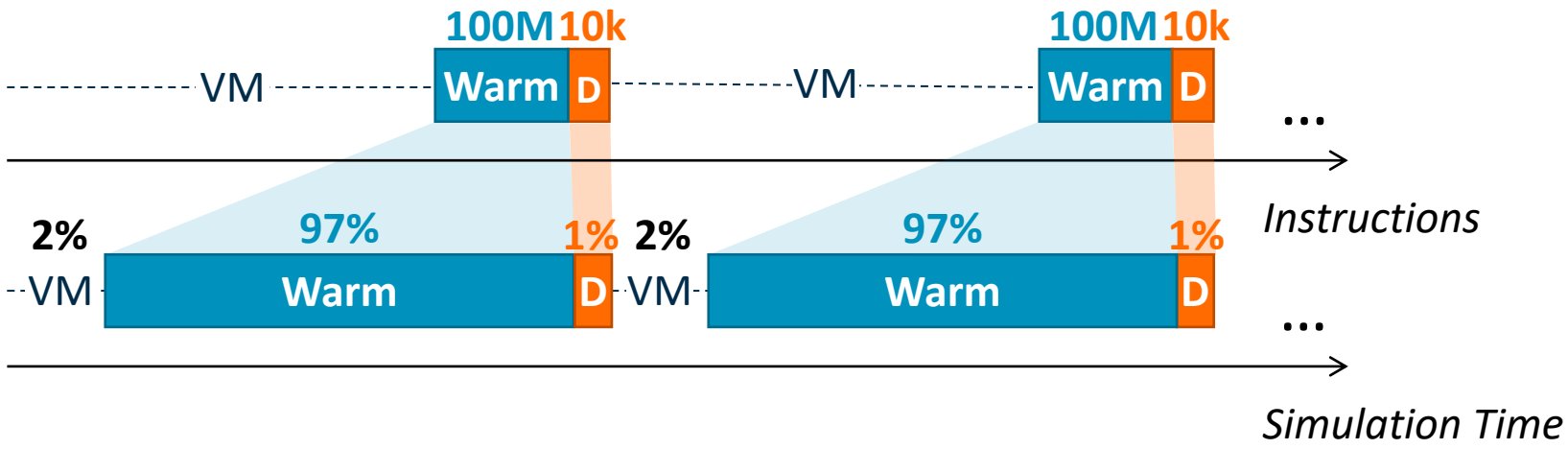


[1] SMARTS: Accelerating Microarchitecture Simulation via Rigorous Statistical Sampling. Wunderlich et al. ISCA 2003

[2] Full Speed Ahead: Detailed Architectural Simulation at Near-Native Speed. Sandberg et al. IISWC 2015

arm

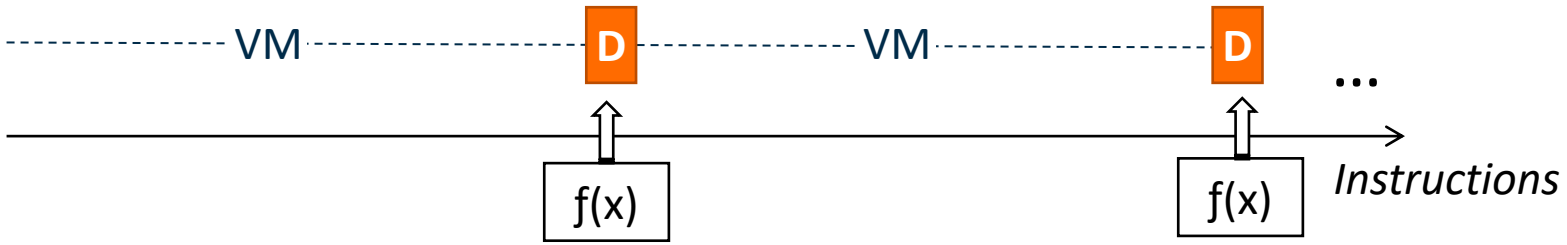
Sampled Simulation Bottleneck



Cache warmup
97% of simulation time

arm

Goal: Eliminate Traditional Cache Warm-up



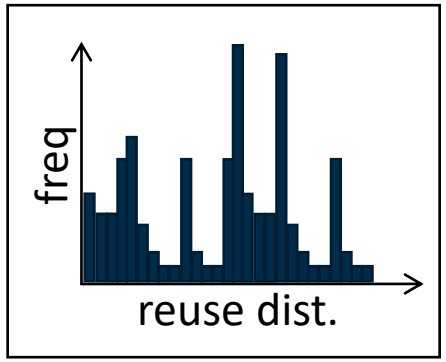
Feasible simulation for systems w/ large caches

arm

But how?

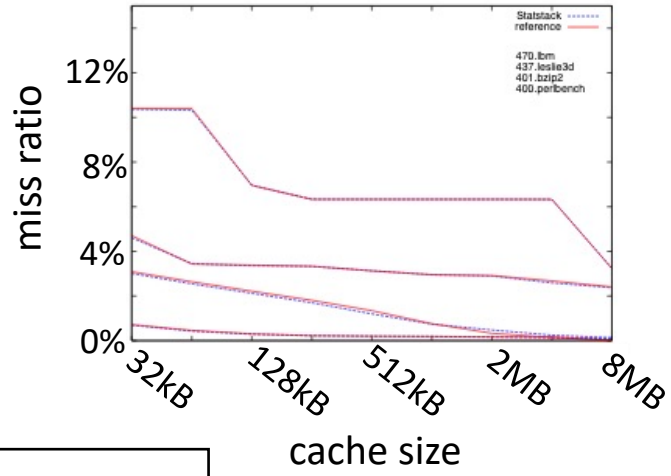
Statistical Cache Modeling

Input:
memory reuse



Cache Model

Output:
miss ratio=ƒ(cache size)



StatStack[1], StatCache[2], etc:

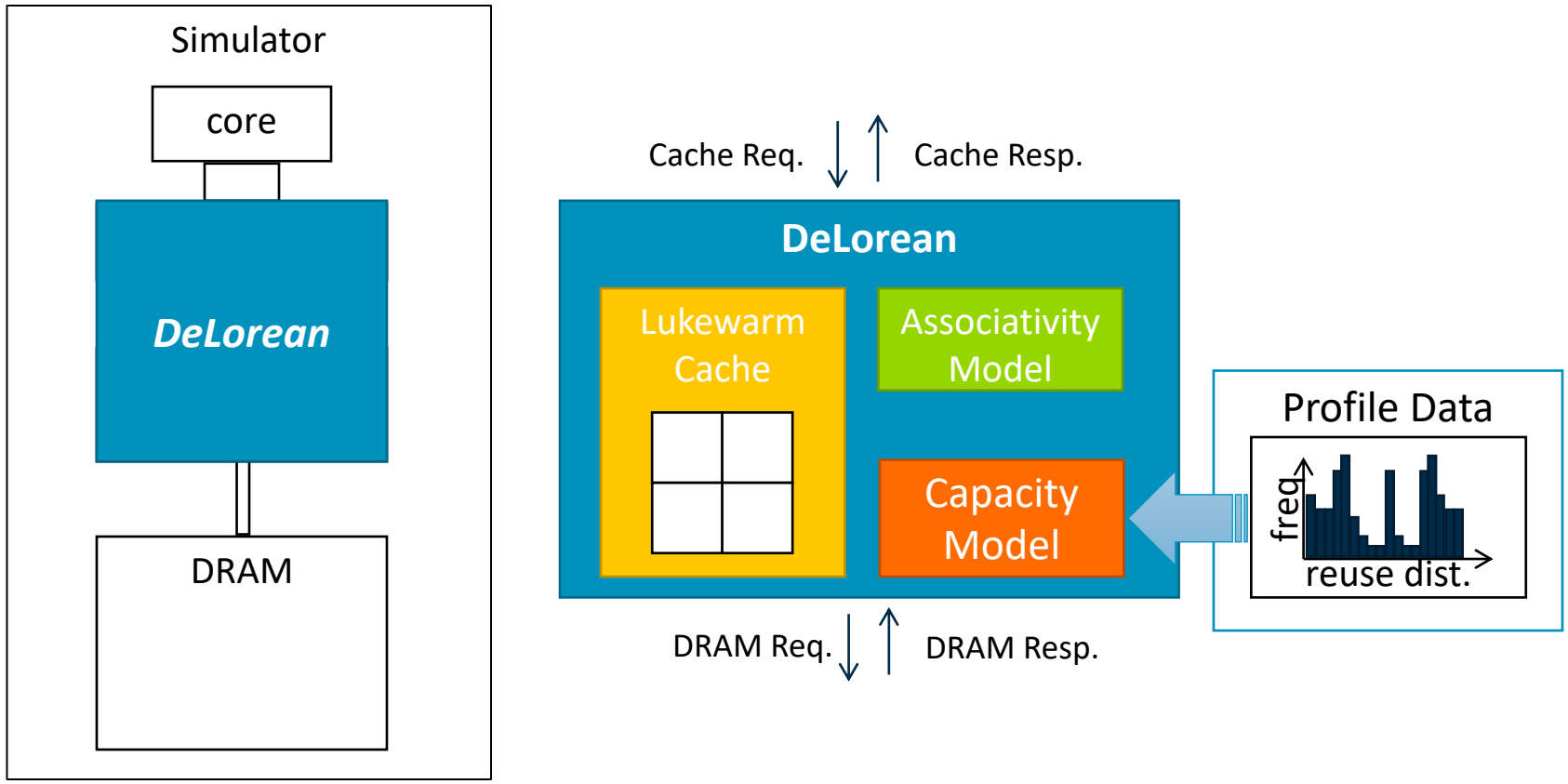
- **Sparse** input
- 20% overhead on native execution

[1] StatStack: Efficient modeling of LRU caches. Eklöv and Hagersten. ISPASS 2010.

[2] StatCache: A Probabilistic Approach to Efficient and Accurate Data Locality Analysis. Berg and Hagersten. ISPASS 2004.

arm

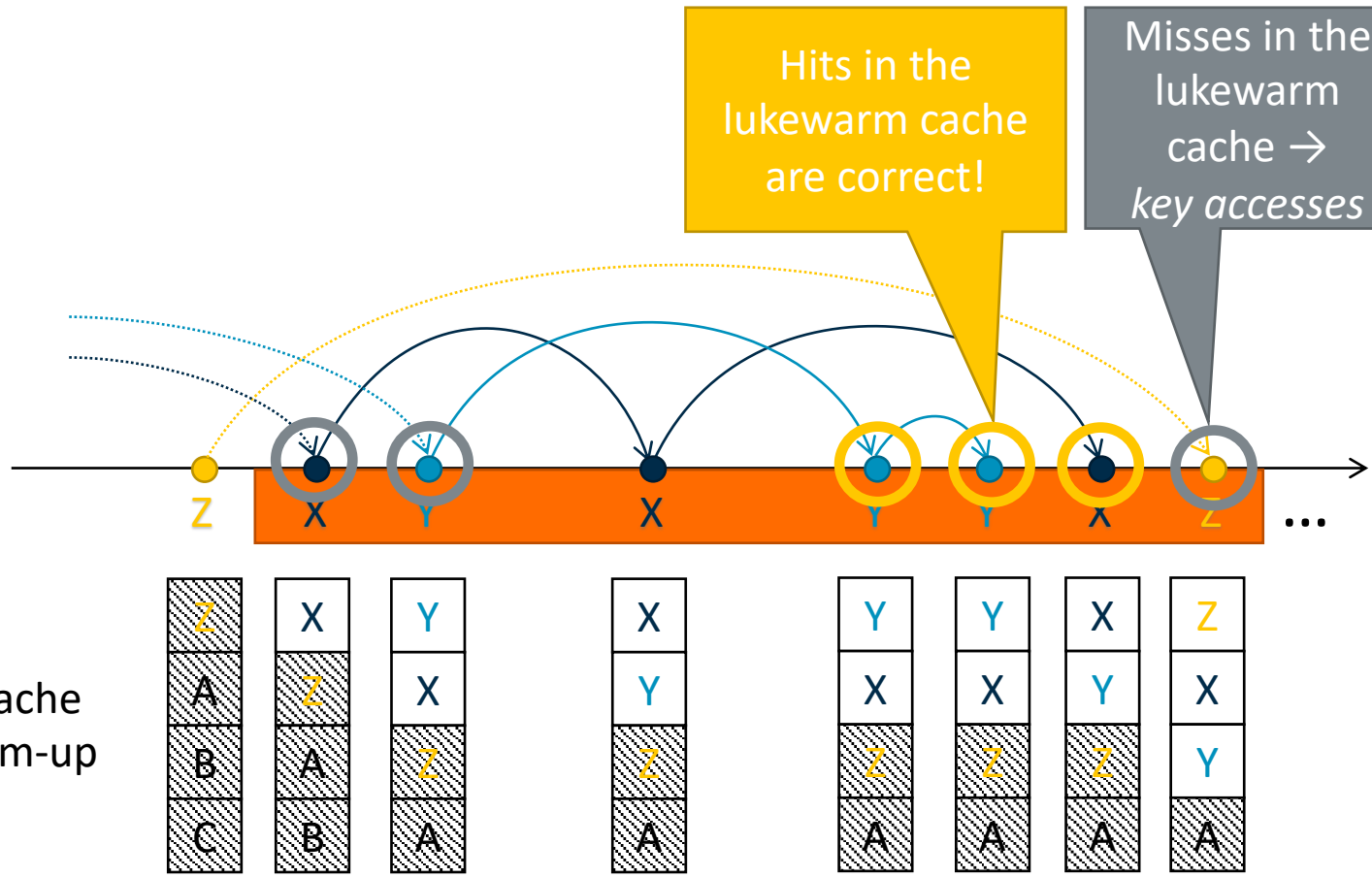
DeLorean Overview



arm

The Lukewarm Cache

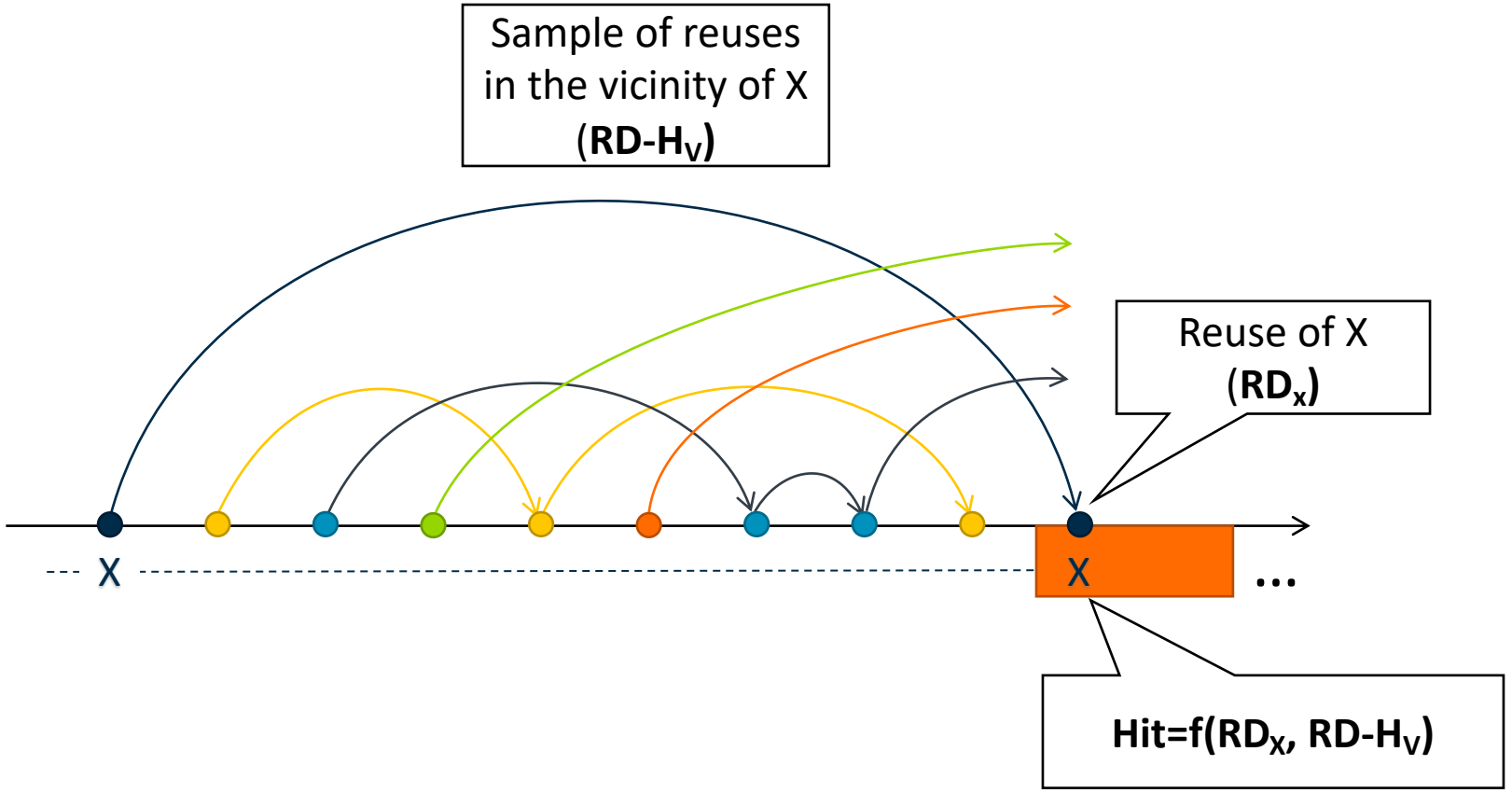
Do we need to statistically model every memory instruction?



f(x)

arm

Capacity Model: StatStack



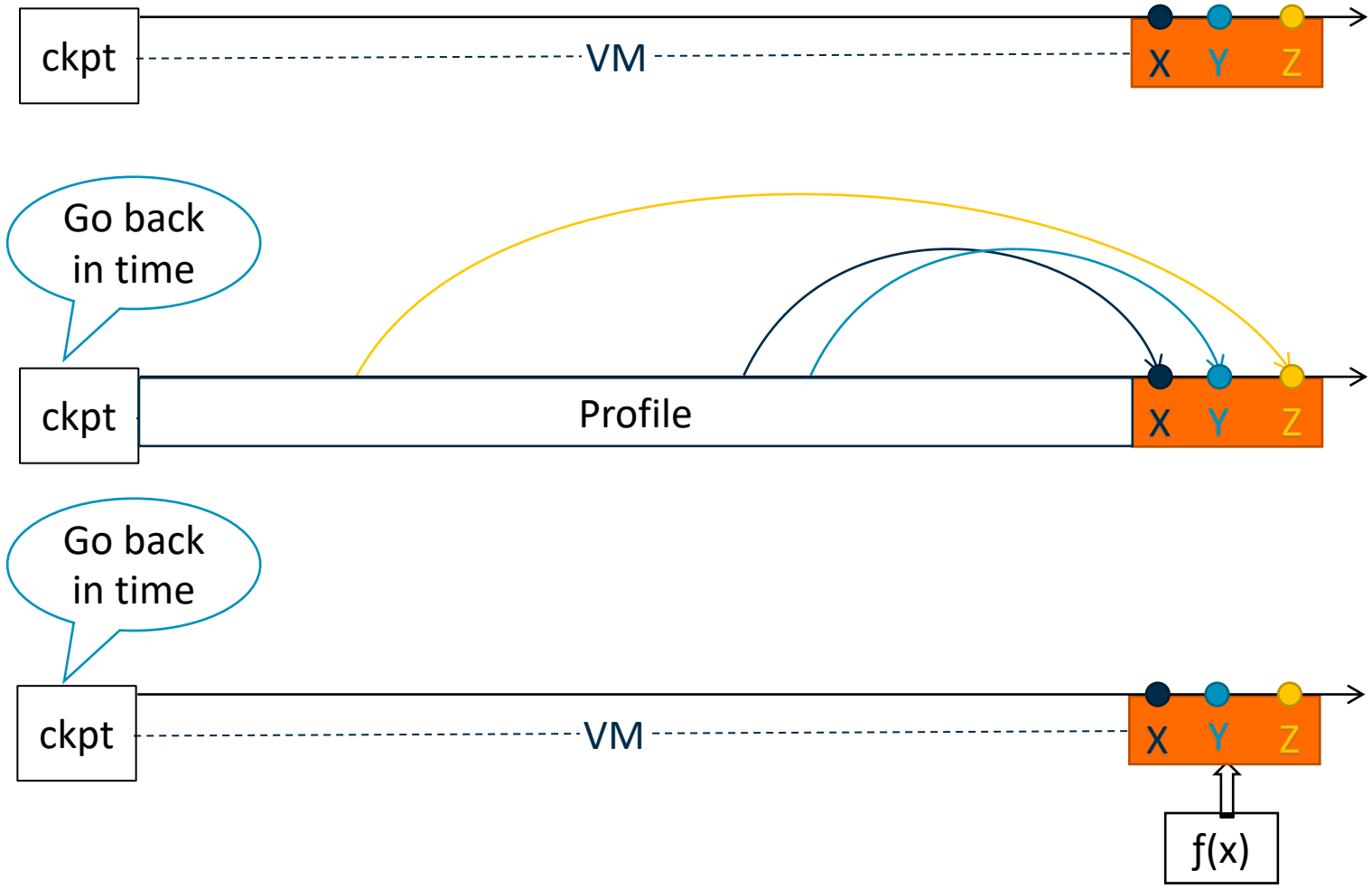
arm

Time Traveling

Record key accesses
Misses in the Lukewarm Cache

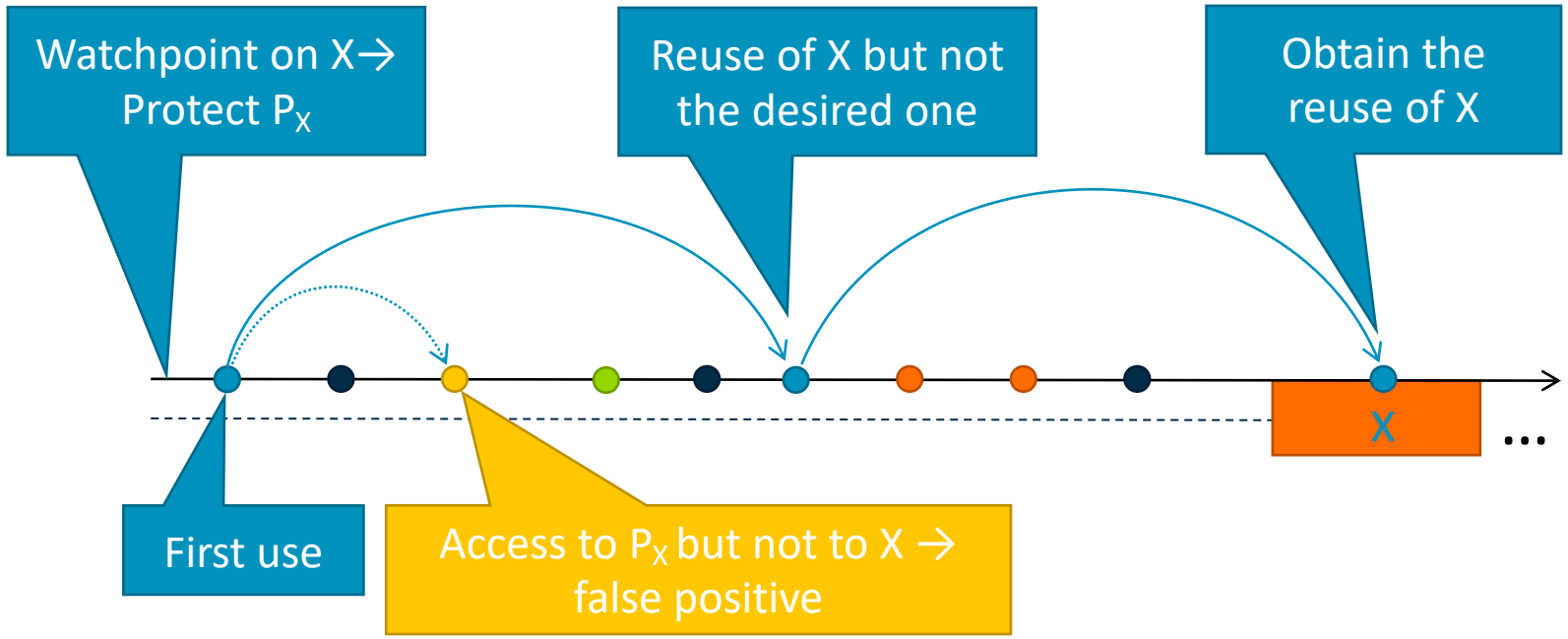
Obtain reuses for key accesses

Detailed simulation and directed cache warming



Directed Reuse Profiling

Obtain reuses for the key accesses while executing in KVM



Avoid false positives for efficient reuse collection

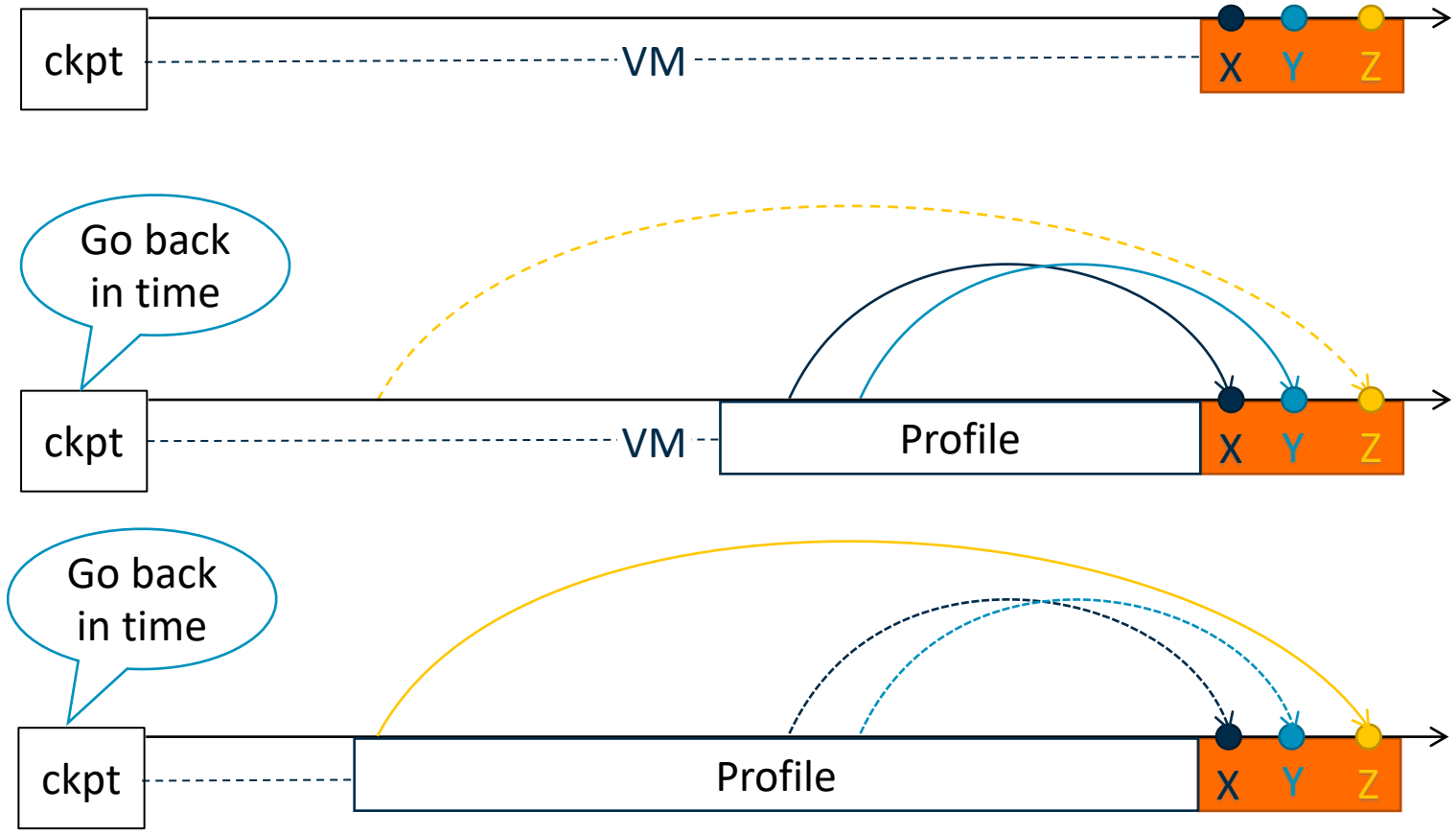
arm

Time Traveling

Record key accesses
Misses in the Lukewarm Cache

Obtain reuses key accesses

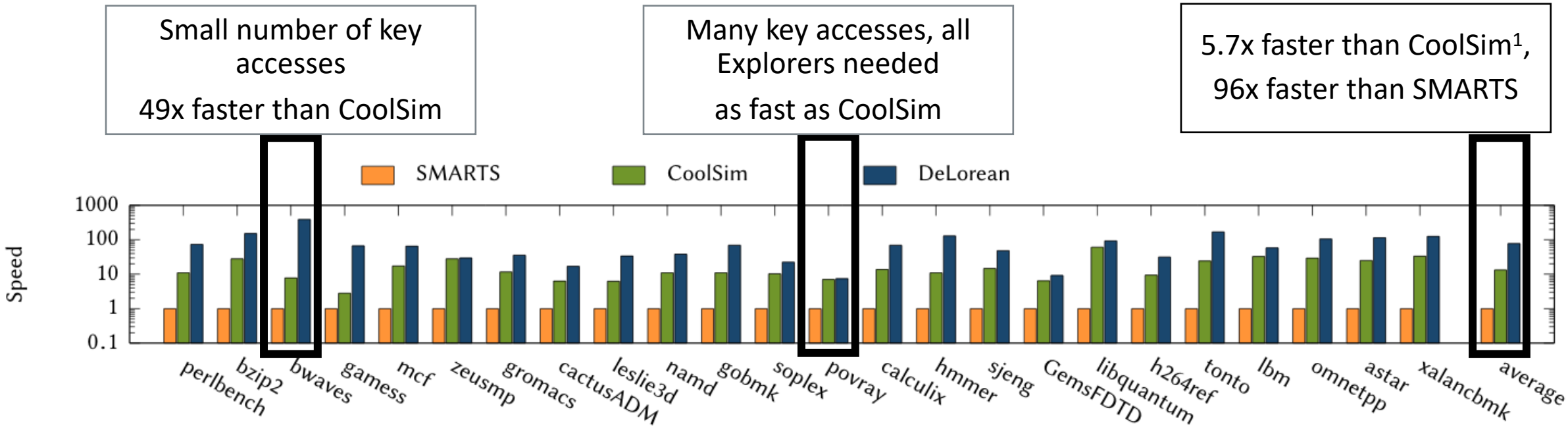
Detailed simulation and directed cache warming



Repeat until reuses for all key accesses are found.
Explorer-n+1 is not always needed!

arm

Results: Speed

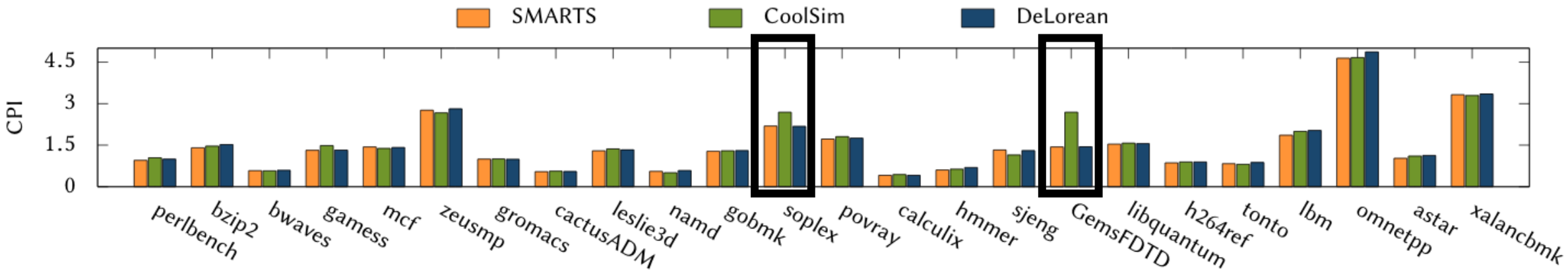


[1] CoolSim: Statistical Techniques to Replace Cache Warming with Efficient, Virtualized Profiling. Nikoleris et al. SAMOS 2016

arm

Results: Accuracy

Average error 3.5%



arm

Conclusion

- Multi-pass approach to identify and capture key memory reuses
  - Time Traveling: Exploits **near-native fast-forwarding** (**KVM**)
  - Targets memory reuse of **key accesses**
- Simulation speed
  - 96x faster than SMARTS
- Accuracy - 3% error on average
- Check the paper for more
  - Multicore simulation
  - Replacement policies
- Code publicly available at:
  - github.com/delorean-sim

arm

Thank You
Danke
Merci
谢谢
ありがとう
Gracias
Kiitos
감사합니다
धन्यवाद
شكرًا
תודה

arm

Directed Statistical Warming through Time Traveling

Nikos Nikoleris
Arm Research

Lieven Eeckhout
Ghent University

Erik Hagersten
Uppsala University

Trevor E. Carlson
National University of Singapore

MICRO-52