TRISHUL: A Single-pass Optimal Two-level Inclusive Data Cache Hierarchy Selection Process for Real-time MPSoCs

Md. S. Haque, Akash Kumar, Yajun Ha, Qiang Wu and Shaobo Luo
Department of Electrical and Computer Engineering
National University of Singapore

Research Question

- Can we still use the existing Single-pass cache simulation techniques as a fast, accurate and resource generous method to find the most appropriate multi-level cache hierarchy for a Multi-processor system on chip (MPSoC)?

Presentation Overview

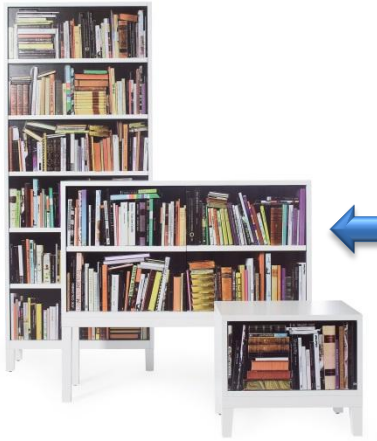
- Cache memories and single-pass cache simulation techniques.
- Finding the appropriate cache hierarchy in MPSoCs using single-pass simulation.
- Performance of our proposed single-pass simulation technique “TRISHUL”.

Cache Memories

Processor

Cache Memory

Main Memory



Student

Bookshelf

Library

Cache Memories

Processor

Cache Memory

Main Memory

Chef

Refrigerator

Supermarket

Deciding The Right Cache Memories

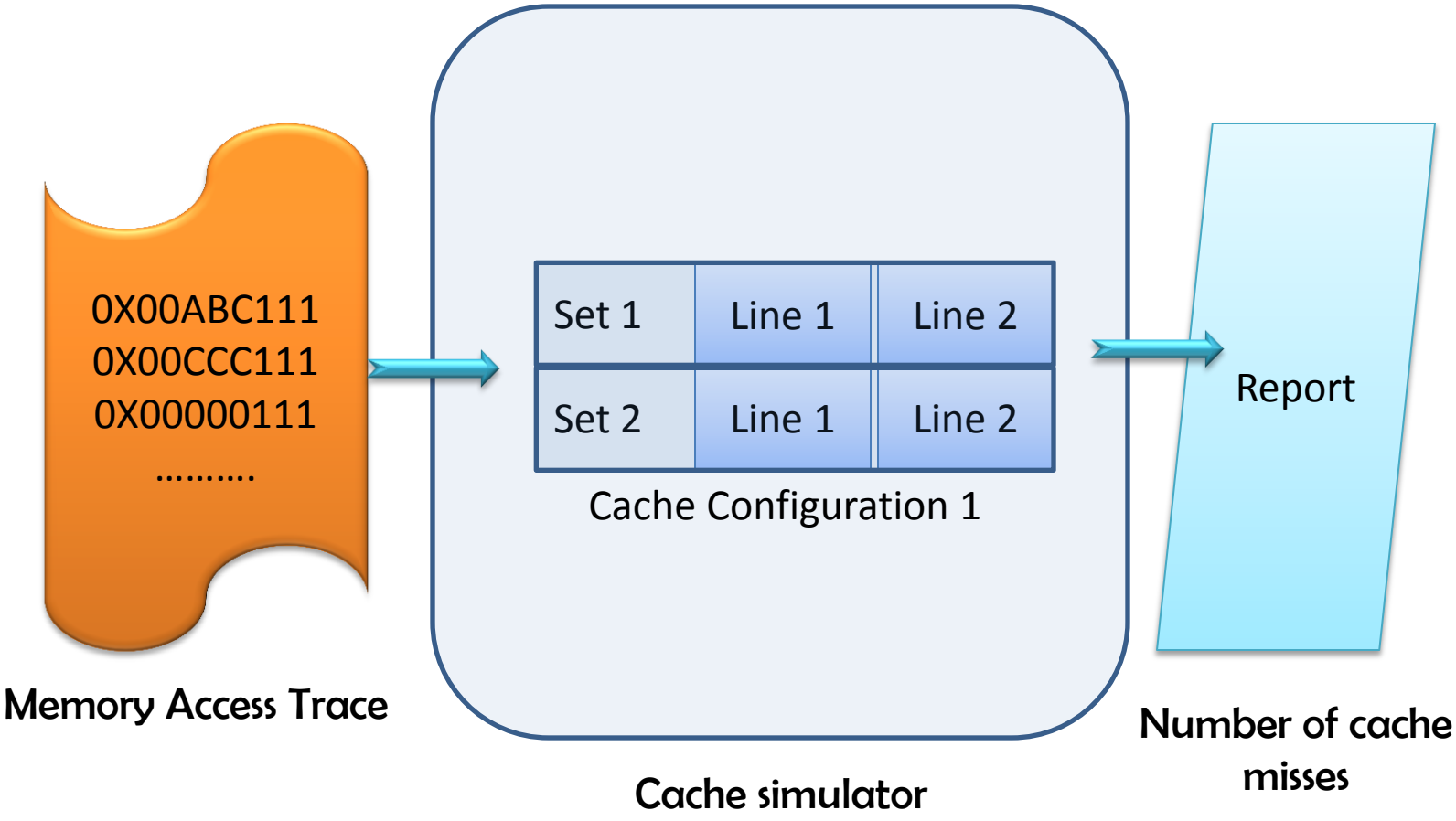
Refrigerator

Time
Wastage

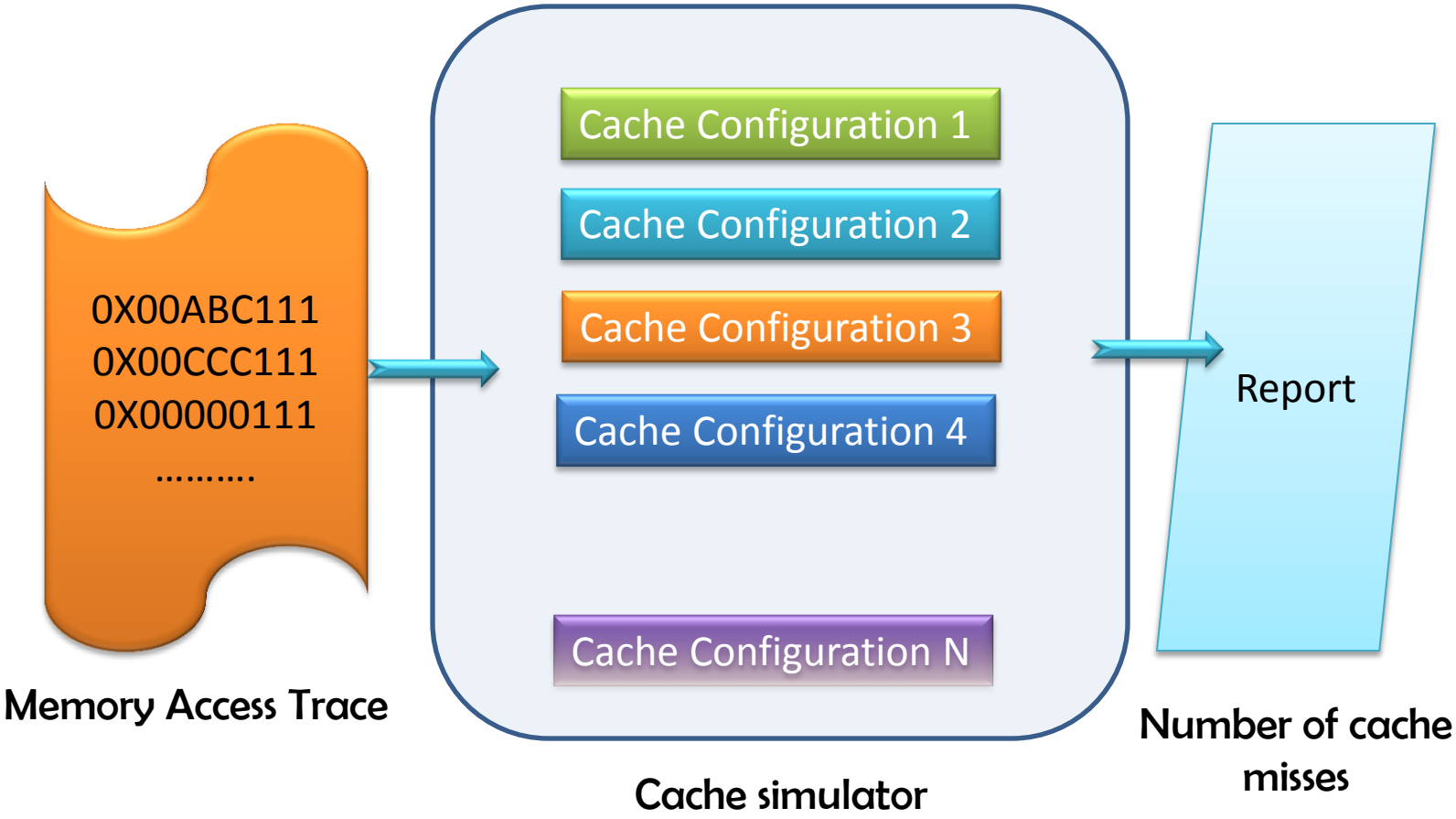
Energy
Wastage

How to Decide The Right Cache Memory?

Trace Driven Cache Simulation



Single-pass Cache Simulation



Challenges in Deciding the Optimal Two-level Inclusive Cache Hierarchy in Real-time MPSoCs

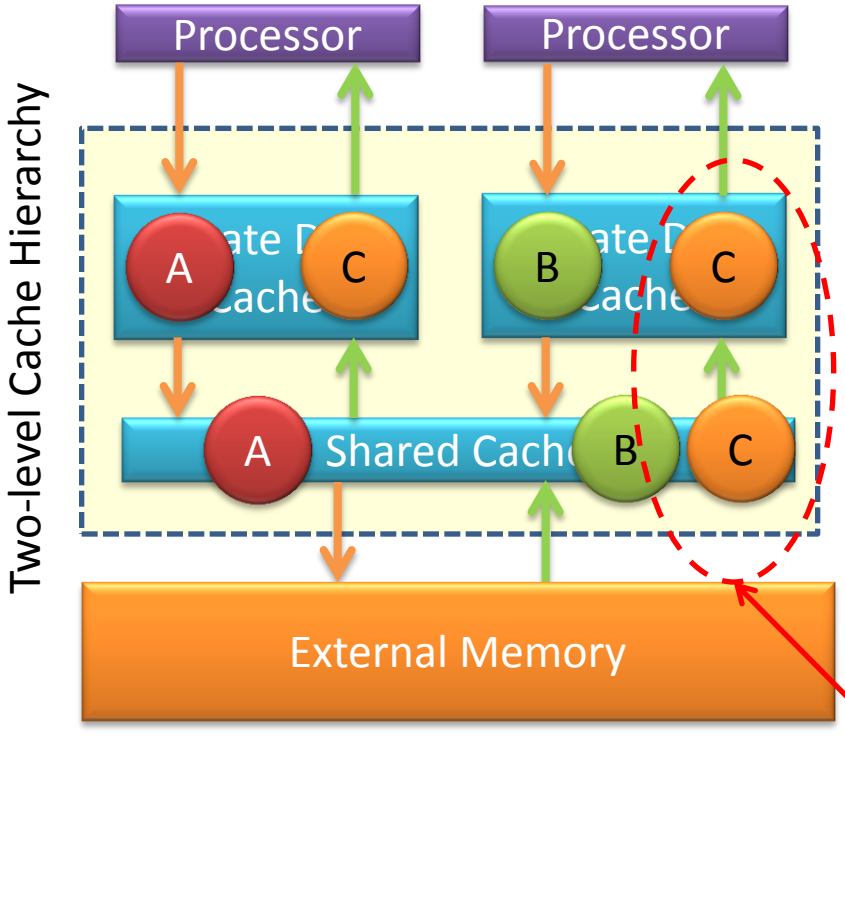
For a cache memory:

- $(Cache\ Capacity) \propto \left(\frac{1}{Cache\ misses}\right)$
- $(Cache\ Capacity) \propto Area\ Consumption.$
- $(Cache\ Capacity) \propto Price$
- $(Cache\ Capacity) \propto Energy\ Consumption.$

Application Specific Uniprocessor Systems vs. Application Specific Real-time Systems

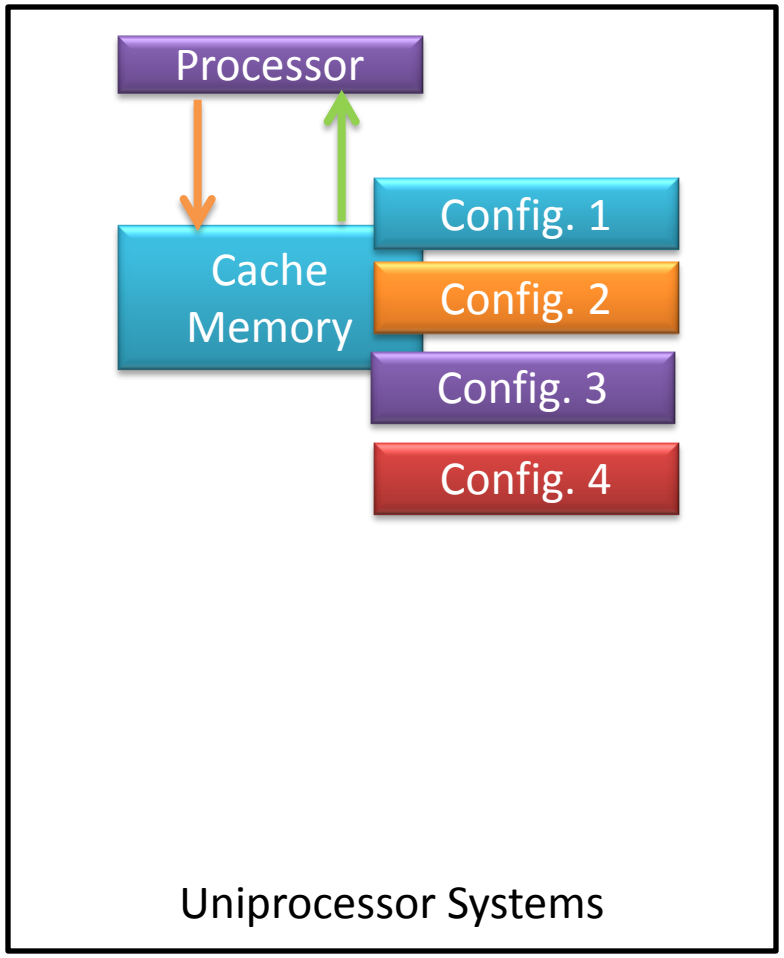
- In Real-time systems, applications need to be completed with a given time limit to avoid catastrophic consequences.
- Example of real-time systems can be life saving devices in the hospital, auto-pilot in the aircraft, image compressor in the digital camera, etc.

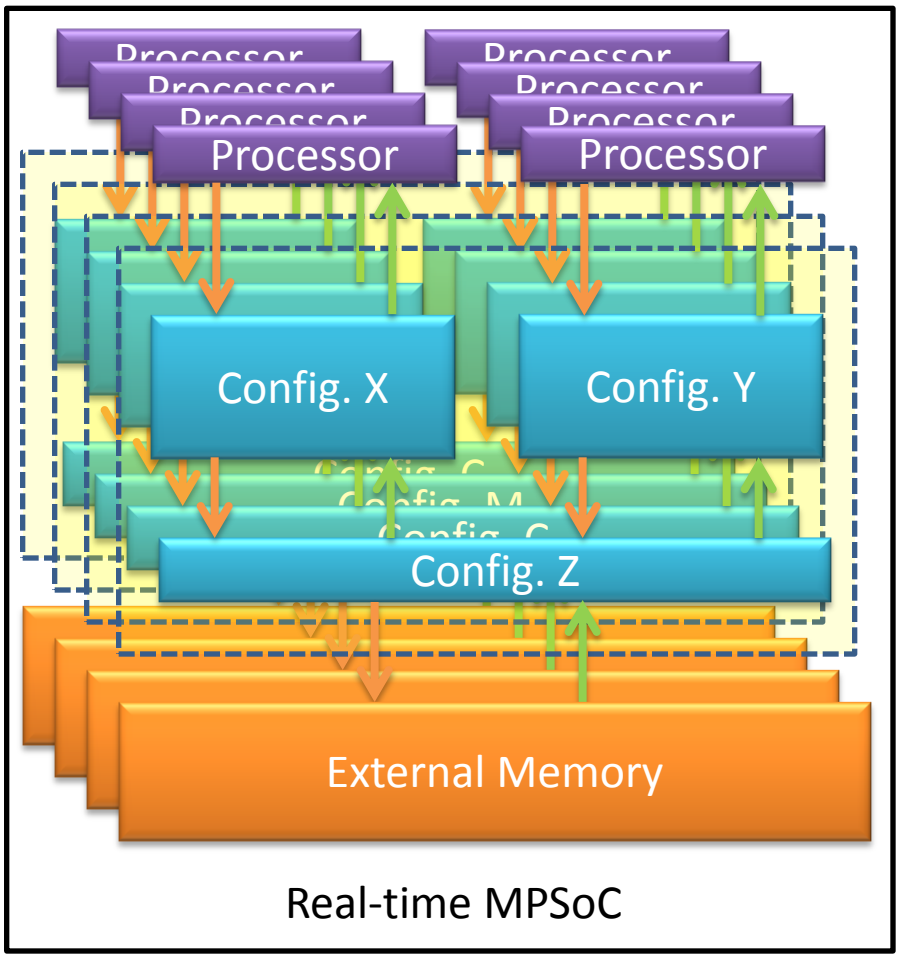
Real-time MPSoC with Two-level Inclusive Data Cache Hierarchy



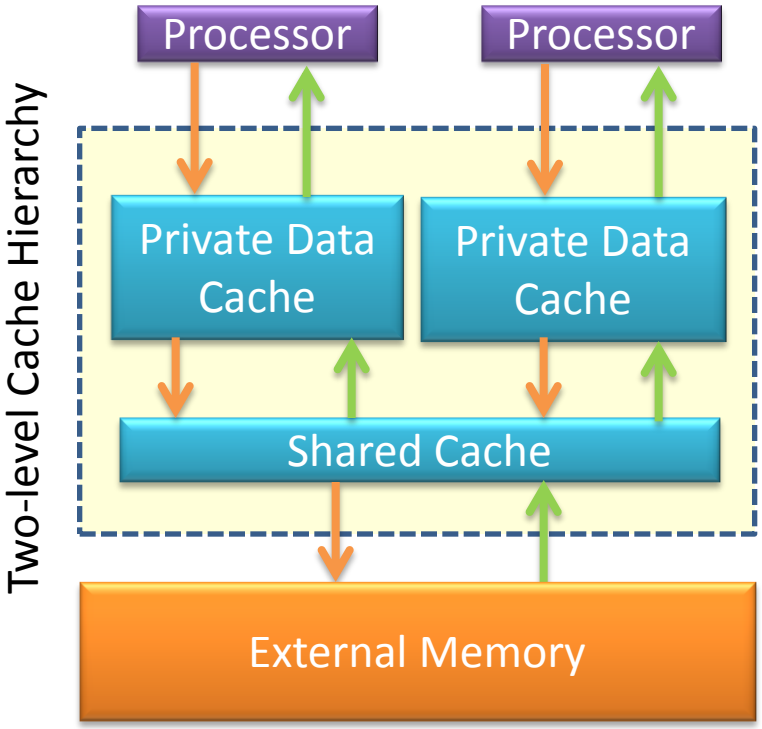
- Every processor has own private data cache memory.
- Shared cache holds the superset of the private cache contents.
- Private caches can have copy of the same data.
- All the cache memories must be aware of a data updat

Challenges in Deciding the Optimal Two-level Inclusive Cache Hierarchy in Real-time MPSoCs



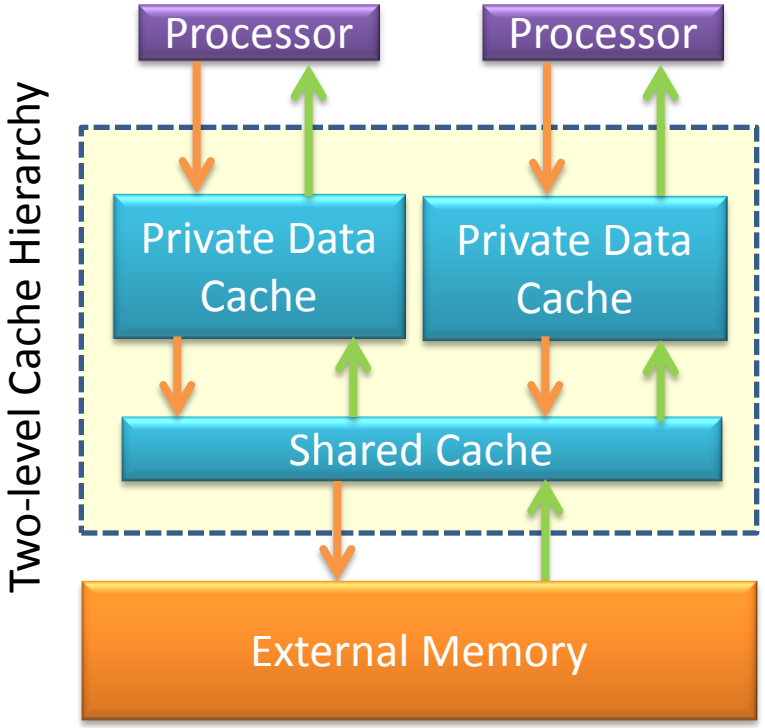


Challenges in Deciding the Optimal Two-level Inclusive Cache Hierarchy in Real-time MPSoCs



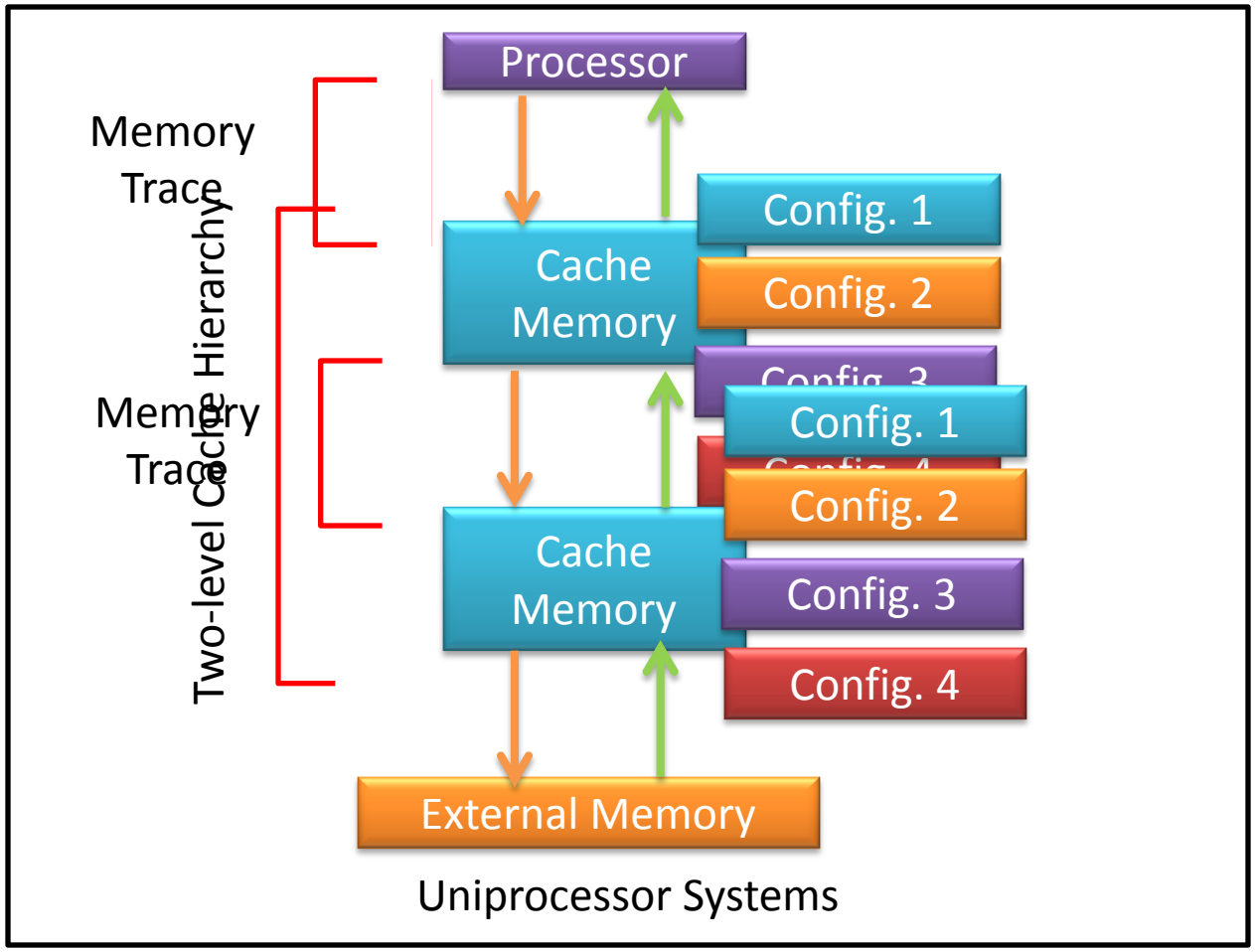
- If every cache memory has 500 configurations,
- If one cache hierarchy configuration needs 5 minutes to simulate,
- If each processor in a quad core machine simulates a cache hierarchy.

Challenges in Deciding the Optimal Two-level Inclusive Cache Hierarchy in Real-time MPSoCs

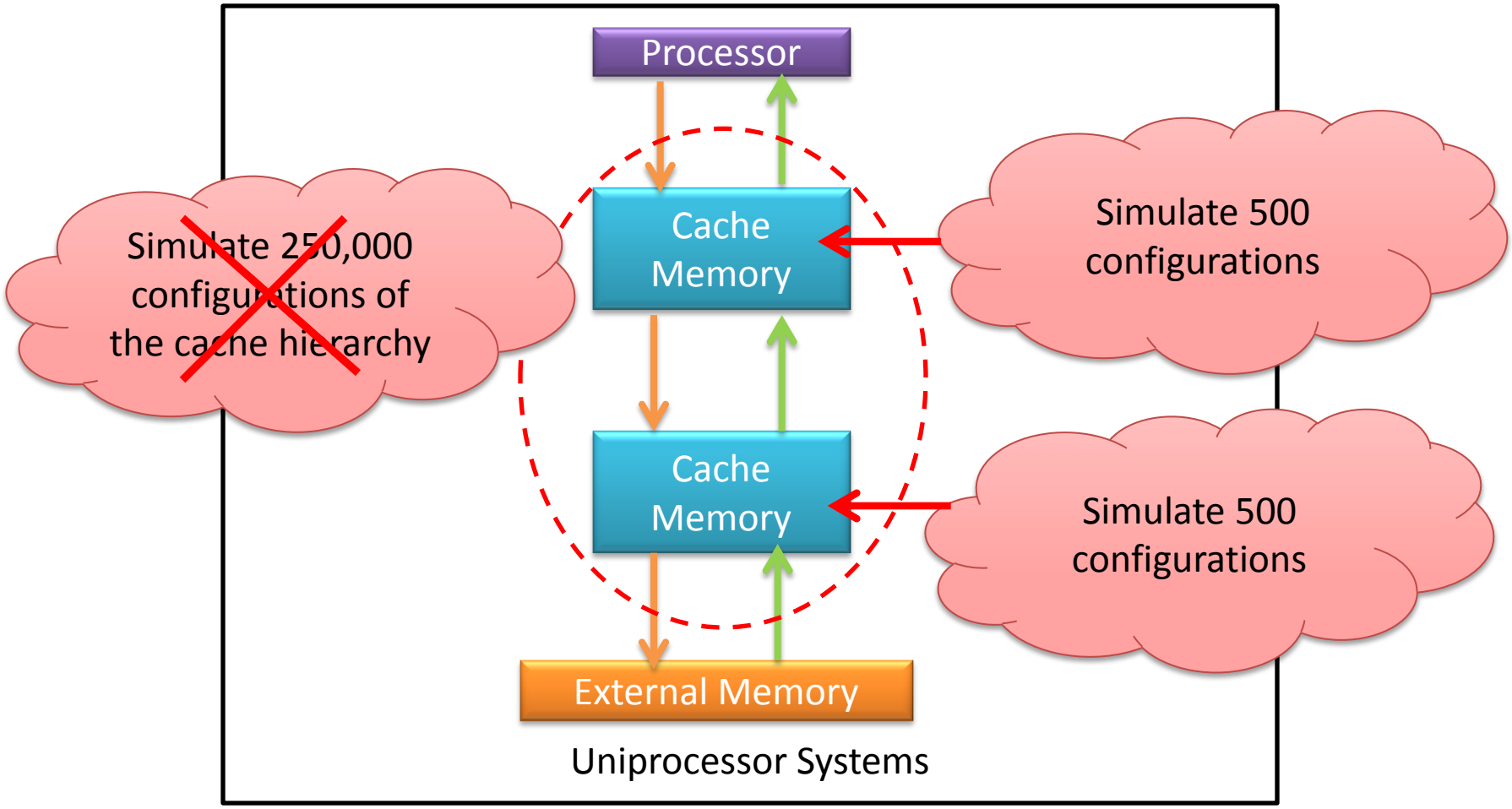


- It will take 297 years to complete simulation of the 125,000,000 cache hierarchy configurations.

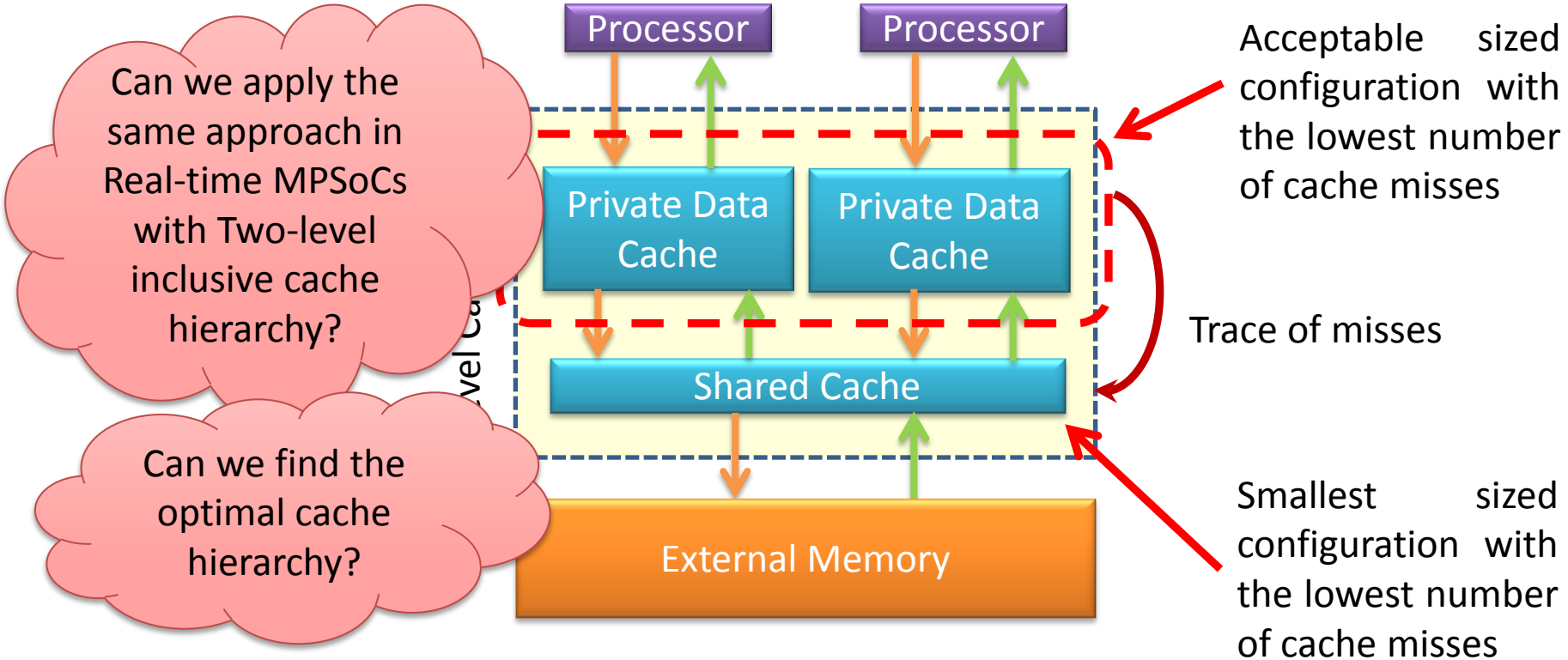
Using Single-pass Simulation in Deciding the Optimal Two-level Inclusive Data Cache Hierarchy in Uniprocessor Systems



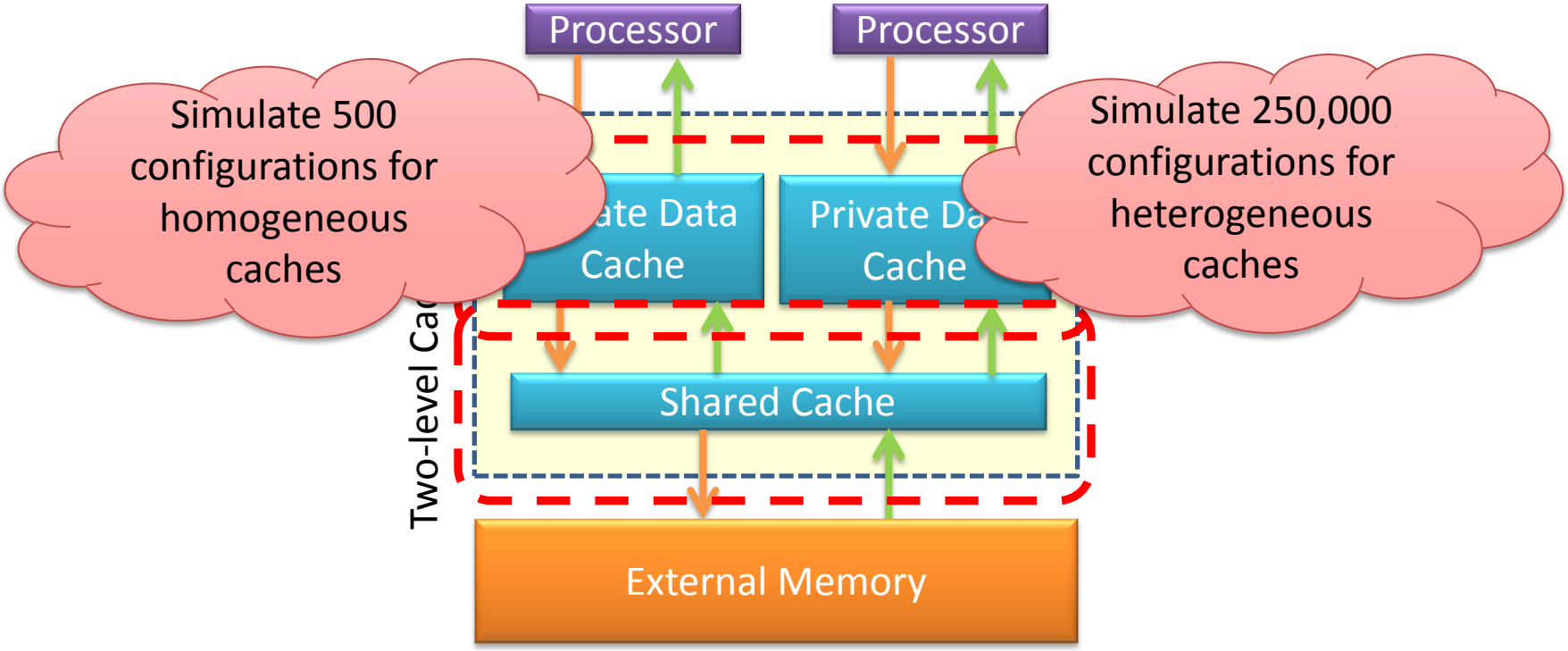
Using Single-pass Simulation in Deciding the Optimal Two-level Inclusive Data Cache Hierarchy in Uniprocessor Systems



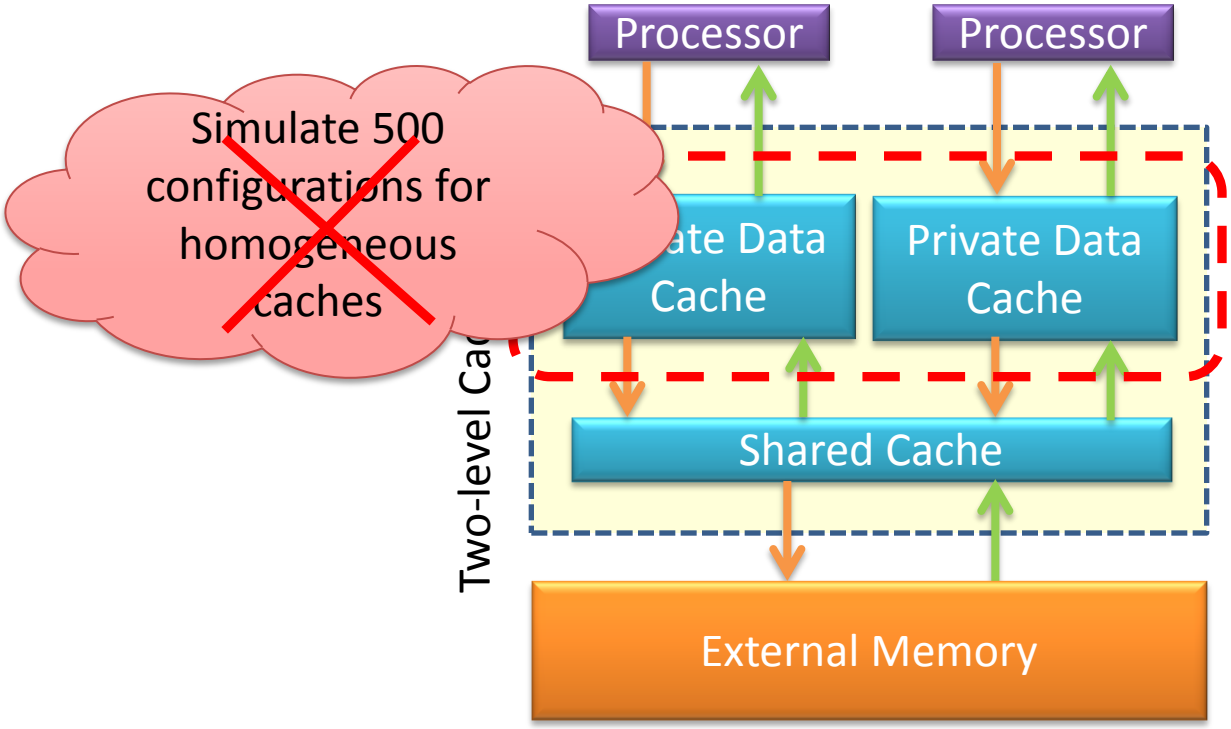
Using Single-pass Simulation in Deciding the Optimal Two-level Inclusive Data Cache Hierarchy in MPSoCs



Using Single-pass Simulation in Deciding the Optimal Two-level Inclusive Data Cache Hierarchy in MPSoCs



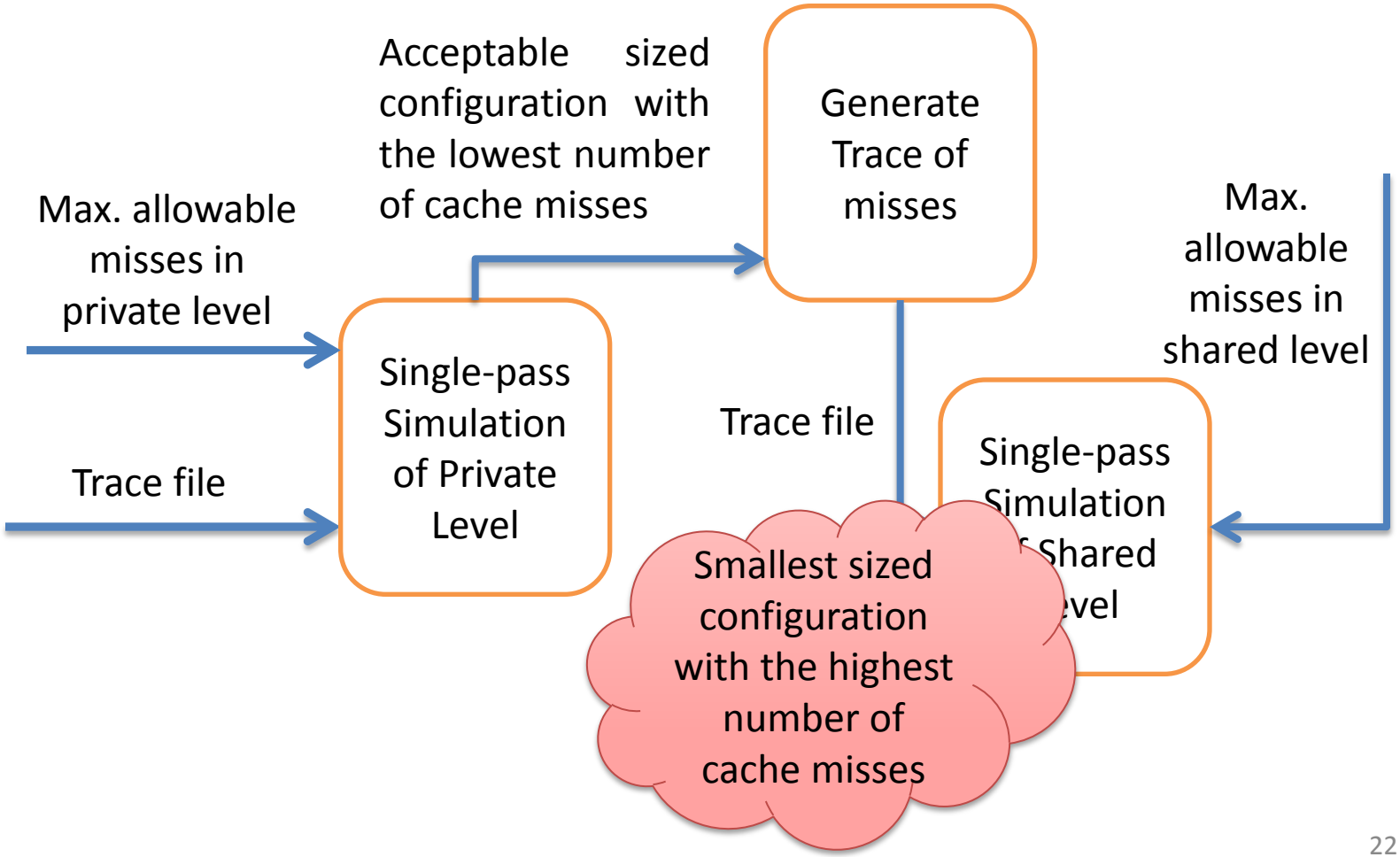
Using Single-pass Simulation in Deciding the Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs



Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass

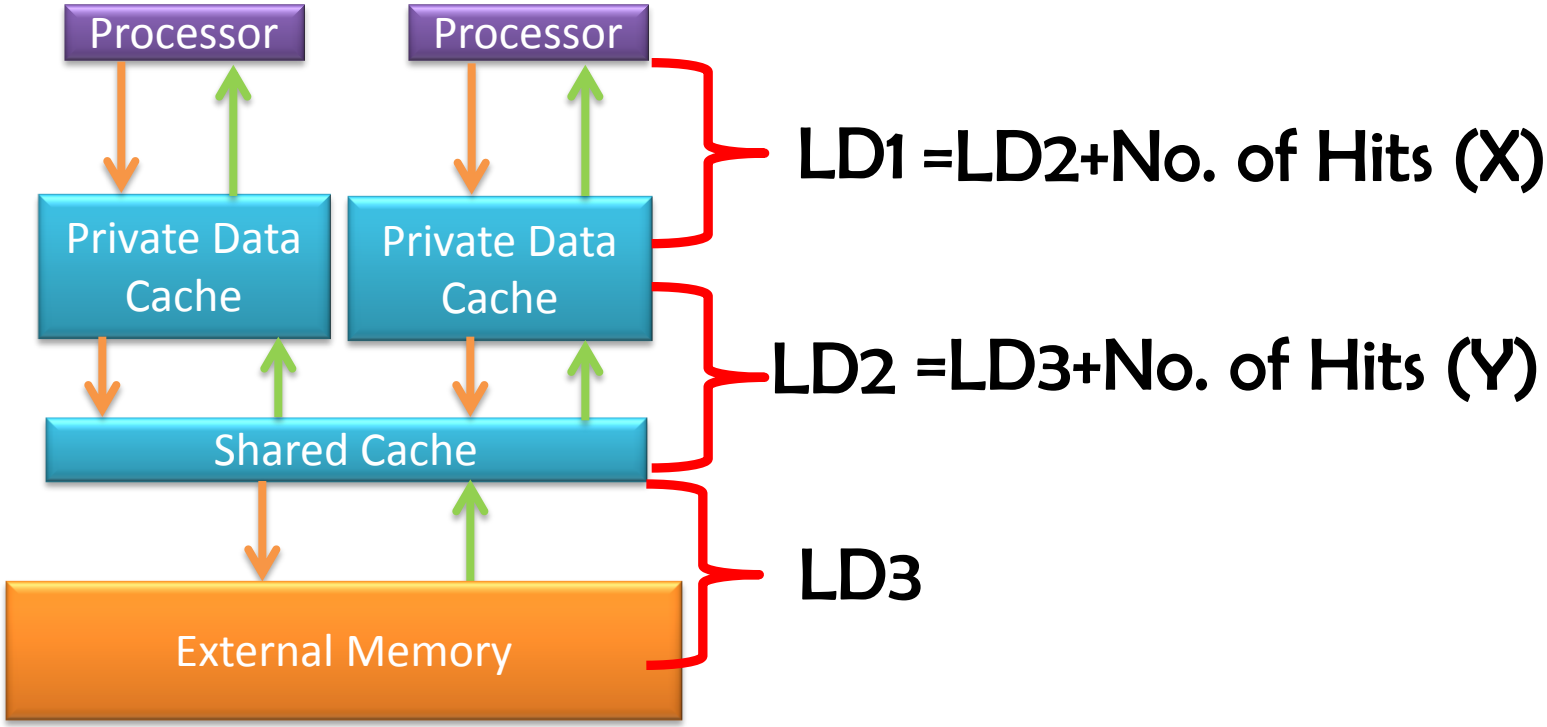
- It is possible to determine the Worst Case Data Memory Operation Time (WCDMOT) from the Execution time limit in Real-time MPSoC.

Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass



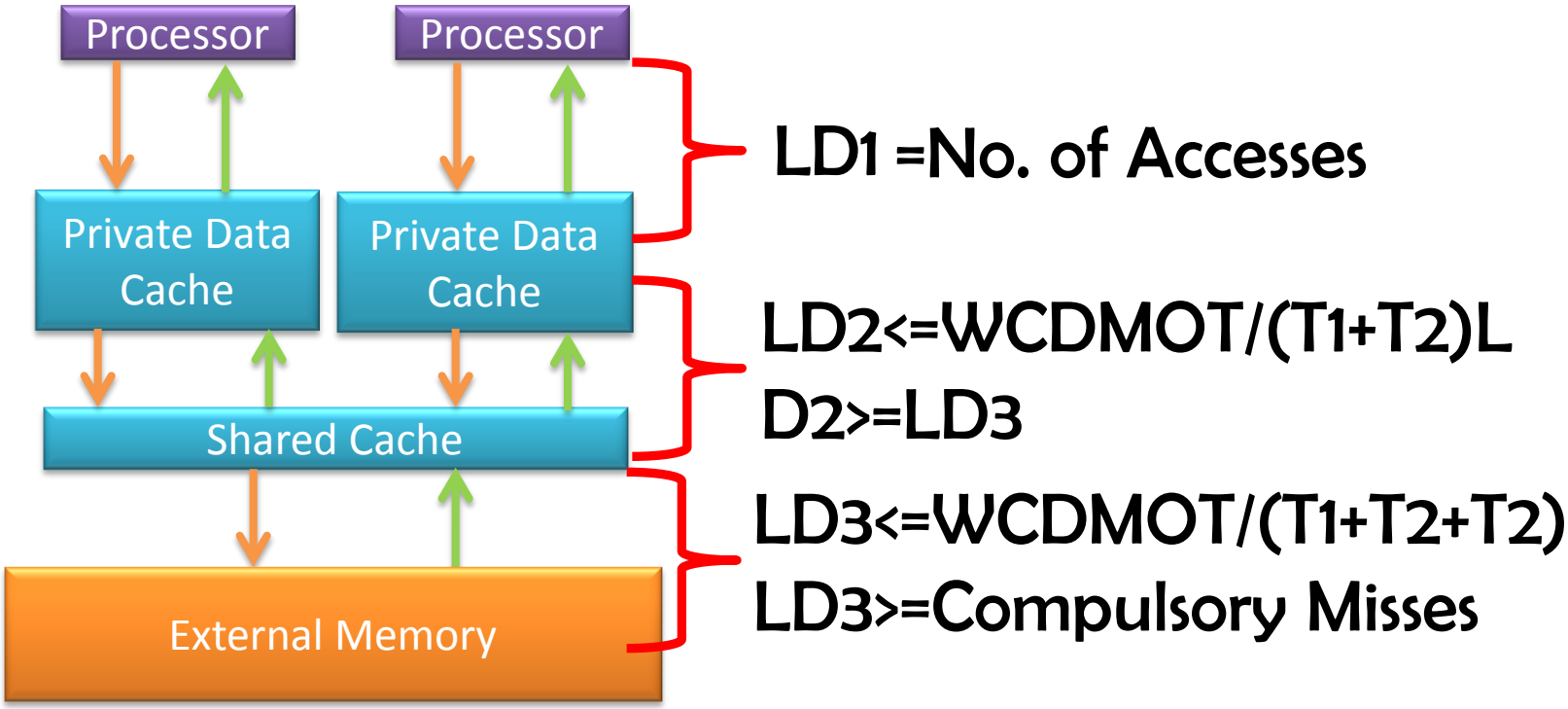
Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass

$$WCDMOT = (LD1 \times T1) + (LD2 \times T2) + (LD3 \times T3)$$



Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass

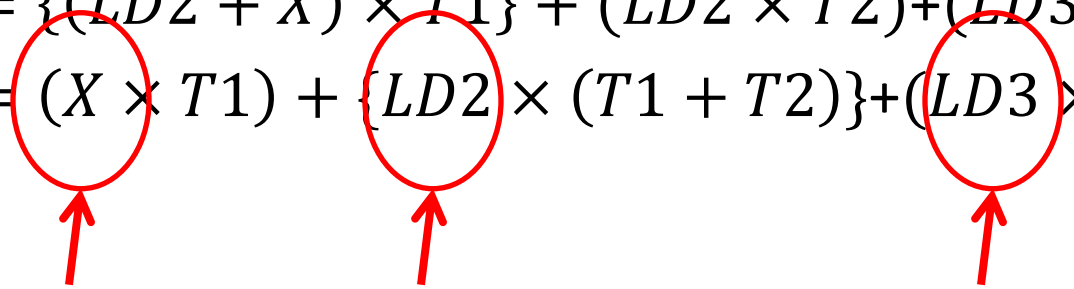
$$WCDMOT = (LD1 \times T1) + (LD2 \times T2) + (LD3 \times T3)$$



Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass

$$WCDMOT = \{(LD2 + X) \times T1\} + (LD2 \times T2) + (LD3 \times T3)$$

$$WCDMOT = (X \times T1) + \{LD2 \times (T1 + T2)\} + (LD3 \times T3)$$



Minimum Value Maximum Value Maximum Value

Finding Optimal Two-level Inclusive Data Cache Hierarchy in Real-time MPSoCs using Single-pass

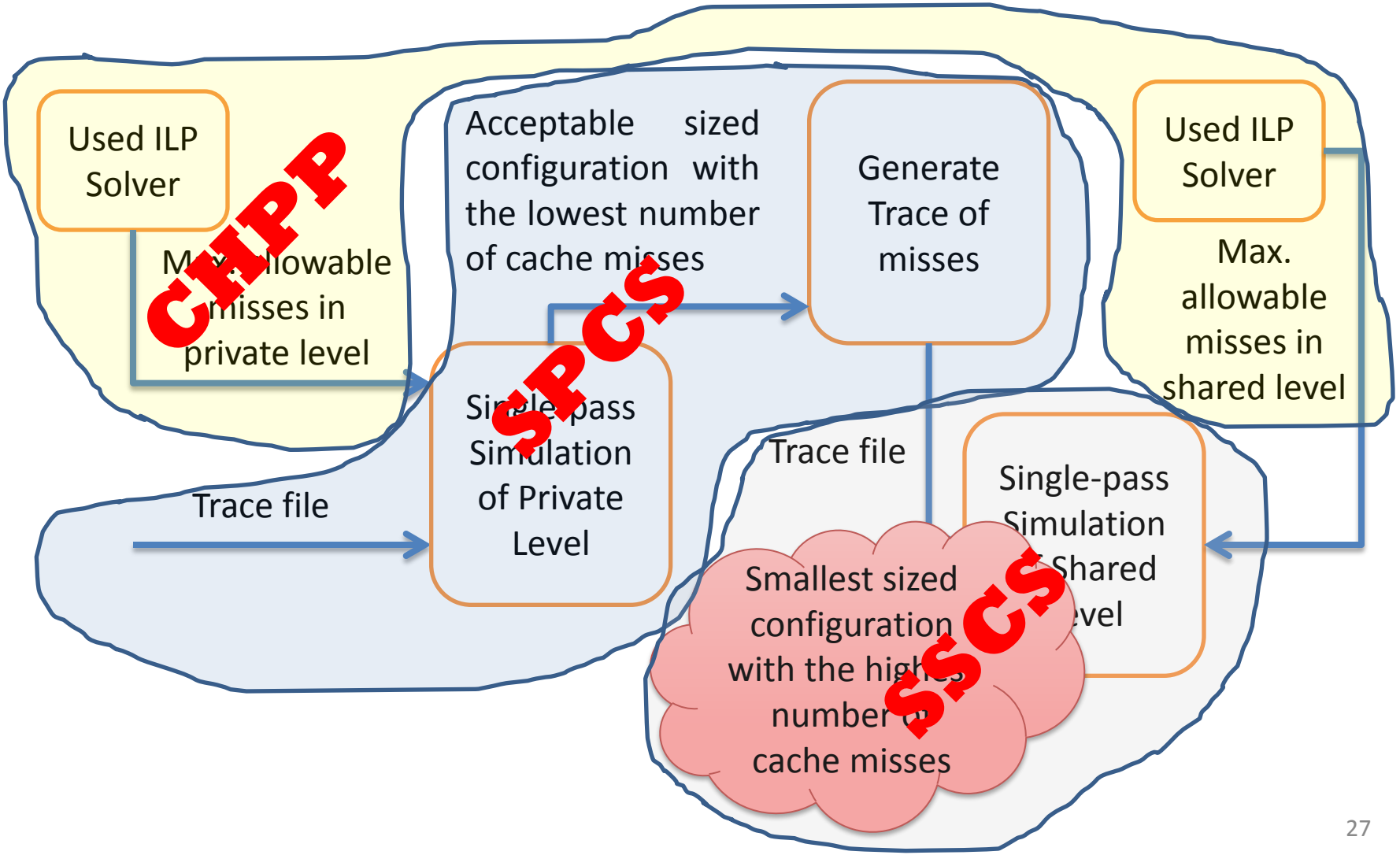
$$WCDMOT = (LD1 \times T1) + \{(LD3 + Y) \times T2\} + (LD3 \times T3)$$

$$WCDMOT = (LD1 \times T1) + (Y \times T2) + \{LD3 \times (\mathrm{T2} + \mathrm{T3})\}$$

Minimum Value

Maximum Value

TRISHUL Multi-level Cache Hierarchy Simulator for Real-Time MPSoCs



Simulation Results

Trace	WCDMOT (Sec)	TRISHUL (Optimal) Private Config.	TRISHUL (Optimal) Shared Config.	DIMSim Shared Config.
barbara	1	(8X2)	(1X2)	(8X16)
	0.4	(4X16)	(1X2)	N/A
	0.15	(16X16)	(64X16)	N/A
graph	1	(8X2)		
	0.4	(4X16)		
	0.15	(16X16)	(128X8)	N/A
lena	1	(8X2)	(1X2)	(8X16)
	0.4	(4X16)	(1X2)	N/A
	0.15	(16X16)	(64X16)	N/A

Entire Cache hierarchy capacity is 320Bytes if cache line size is 4Bytes

Shared cache capacity is 512Bytes if cache line size is 4Bytes

Simulation Results

Trace	WCDMOT (Sec)	TRISHUL AMT (Sec)	TRISHUL Decision (Sec)	DIMSim Decision (Sec)
barbara	1	0.96	1700	1832
	0.4	0.4	361	N/A
	0.15	0.15	281	N/A
criss	1	0.96	1699	1758
	0.4	0.4	345	N/A
	0.15	0.15	280	N/A
graph	1	0.96	1792	1752
	0.4	0.4	354	N/A
	0.15	0.15	283	N/A
lena	1	0.96	1761	1735
	0.4	0.4	344	N/A
	0.15	0.15	281	N/A

An Additional Advantage of TRISHUL

- When WCDMOT slightly reduces due to system overhead and the system is with reconfigurable cache hierarchy, without performing any simulation, we can adjust the cache hierarchy to the optimal setup.

Conclusion

- In this article, we have shown with evidences that single-pass simulation is still an effective method to decide the optimal cache hierarchy in Real-time MPSoCs.
- By using strategically, Single-pass simulation doesn't loose any benefit of Uniprocessor Systems in MPSoCs with Two level inclusive data cache hierarchy.
- For additional optimization, users can use our simulator TRISHUL.

Thank You Very Much