



Zentralübung Rechnerstrukturen: Cache-Kohärenz und -Konsistenz

Musterlösung zu Aufgabenblatt 6

1 Cache allgemein

1.1 Cacheleistung

$$\begin{aligned} \text{a) } t_a &= t_{L1} * r_{H1} + (1 - r_{H1}) * (t_{L2} * r_{H2} + t_{HS} * (1 - r_{H2})) \\ t_a &= 1ns * 75\% + (1 - 75\%) * (8ns * 60\% + 128ns * (1 - 60\%)) \\ t_a &= 0.75ns + 0.25 * (3.6ns + 51,2ns) = 14,45ns \end{aligned}$$

$$\begin{aligned} \text{b) } t_{L1} &= 1ns, r_{H1} = 75\%, t_{L2} = 6ns, r_{H2} = 60\% \\ t_{L3} &= 18ns, r_{H3} = 60\%, t_{HS} = 128ns \end{aligned}$$

Antwort Zugriffszeit:

$$\begin{aligned} t_a &= t_{L1} * r_{H1} + (1 - r_{H1}) * (t_{L2} * r_{H2} + (1 - r_{H2}) * (t_{L3} * r_{H3} + t_{HS} * (1 - r_{H3}))) \\ t_a &= 1ns * 75\% + (1 - 75\%) * (6ns * 60\% + (1 - 60\%) * (18ns * 60\% + 128ns * \\ & (1 - 60\%))) \\ t_a &= 0,75ns + 0,25 * (3,6ns + 0,4 * (10,8ns + 51,2ns)) \\ t_a &= 0.75ns + 0,25 * (28,4ns) \\ t_a &= 0.75ns + 7,1ns = \underline{7,85ns} \end{aligned}$$

Speed-Up:

$$\begin{aligned} S &= \frac{t_a(L2)}{t_a(L3)} \\ S &= 14,45ns / 7,85ns \sim \underline{1,84} \end{aligned}$$

1.2 Beweise

- a) Gegeben Sei ein 2-fach assoziativer Cache mit 2 Blöcken, sowie ein 4-fach assoziativer Cache mit 1 Block. Beide Caches haben also eine Größe von 4 Einträgen.

Weiterhin seien sechs Datenwörter mit den folgenden Adressen gegeben: A, B, C, D, E, F, wobei A, B, E, F auf den ersten Block des 2-fach assoziativen Cache, C und D auf den zweiten Block abgebildet werden. Alle 6 Adressen werden auf den Block des 4-fach assoziativen Cache abgebildet.

Folgende Zugriffsfolge erzeugt bei 4-fach assoziativen Cache mehr Misses, als beim 2-fach assoziativen Cache: $ABCDABEFCD(ABEFCD)^n$

- b) Gegeben Sei ein Direkt-Mapped-Cache, sowie ein vollassoziativer Cache, beide haben eine Kapazität von 4 Cachezeilen. Weiterhin seien 5 Datenwörter mit den folgenden Adressen: A, B, C, D und E, wobei A auf die erste Cachezeile, B auf die zweite Cachezeile, C auf die dritte Cachezeile und D und E auf die vierte Cachezeile abgebildet werden.

Folgende Zugriffsfolge erzeugt bei vollassoziativen Cache mehr Misses, als beim direkt abgebildeten Cache: $ABCDE(ABCDE)^n$

1.3 Verständnisfragen

- Räumliche und Zeitliche Lokalität (90 / 10-Regel)
- Conflict-, Capacity-, Compulsory-Misses
- SRAM-Zellen benötigen viel Chipfläche, deswegen ist der Aufbau des Hauptspeichers aus SRAM-Zellen entweder zu teuer oder man könnte nur geringe Kapazitäten anbieten.

2 Cache-Kohärenzprotokolle

2.1 MESI

Prozessor	Aktion	Prozessor 1		Prozessor 2		Prozessor 3	
		Line 1	Line 2	Line 1	Line 2	Line 1	Line 2
	init	-	-	-	-	-	-
1	rd 6	6/E					
2	rd 2			2/E			
1	rd 4		4/E				
3	rd 4		4/S			4/S	
2	rd 3				3/E		
3	wr 7						7/M
1	wr 4		4/M			4/I	
2	rd 7			7/S			7/S
3	wr 5					5/M	
1	rd 3	3/S			3/S		
3	wr 3	3/I			3/I		3/M
2	wr 7			7/M			

2.2 MOESI

Prozessor	Aktion	Prozessor 1			Prozessor 2		
		Line 1	Line 2	Line 3	Line 1	Line 2	Line 3
	init	-	-	-	-	-	-
1	rd 1	1/E					
2	rd 2				2/E		
1	rd 2		2/S		2/S		
1	rd 4			4/E			
2	wr 6					6/M	
1	rd 6	6/S				6/O	
2	rd 2		2/S		2/S		
1	wr 6	6/M				6/I	
1	wr 4			4/M			
2	rd 3					3/E	
1	rd 5		5/E				
2	rd 6	6/O					6/S
2	wr 5		5/I		5/M		

2.3 Verständnisfragen

a)

- Zustandsbits zur Speicherung des MESI-Zustands
- Snooping-Logik zur Überwachung des Speicherbusses
- Signale zu den weiteren Caches

b) In DSM-Systemen existiert kein gemeinsamer Speicherbus, den die Snooping-Logik überwachen könnte.

c) In DSM-Systemen kommen verzeichnisbasierte Cache-Kohärenzprotokolle zum Einsatz.