

Aufgabe 1: Cache-Adressierung

Ihr Cachingsystem soll 16 GiB an Speicher auf Datenwortebene adressieren können. Die Datenwortlänge beträgt 4 Byte, der Cache umfasst 512 KiB und besteht aus 128 Byte großen Blöcken.

- a) Betrachten Sie für den Cache die Varianten *Direct Mapped Cache*, *8-Way Set Associative Cache* und *Fully Associative Cache*. Berechnen Sie für jede dieser Varianten die Längen von *Tag*, *Index*, *Offset* und die Adresslänge.

Direct Mapped Cache:
Adresslänge 32 bit
Index 12 bit
Offset 7 bit
Tag 13 bit

8-Way Set Associative Cache:
Adresslänge 32 bit
Index 9 bit
Offset 7 bit
Tag 16 bit

Fully Associative Cache:
Adresslänge 32 bit
Index 0 bit
Offset 7 bit
Tag 25 bit

- b) Berechnen Sie zu den Adressen 0xBADCAB1E und 0x23454321 die Werte für *Tag*, *Index* und *Offset* zu jedem der drei Caches aus Teilaufgabe a). Geben Sie die Werte jeweils binär und hexadezimal an, wobei führende Nullen weggelassen werden können.

Direct Mapped Cache:

0xBADCAB1E
Tag 1011101011011 (0x 175B)
Index 100101010110 (0x 956)
Offset 0011110 (0x1E)

8-Way Set Associative Cache:

0xBADCAB1E
Tag 1011101011011100 (0x BADC)
Index 101010110 (0x156)
Offset 0011110 (0x1E)

Fully Associative Cache:

0xBADCAB1E
Tag 1011101011011100101010110 (0x175B956)
Index
Offset 0011110 (0x1E)

0x23454321
Tag 0010001101000 (0x468)
Index 101010000110 (0xA86)
Offset 0100001 (0x21)

0x23454321
Tag 0010001101000101 (0x2345)
Index 010000110 (0x86)
Offset 0100001 (0x21)

0x23454321
Tag 0010001101000101010000110 (0x468A86)
Index
Offset 0100001 (0x21)

- c) Angenommen, zur Realisierung der Ersetzungsstrategie (z.B. LRU, LFU, etc.) werden beim *8-Way Set Associative Cache* 4 Bit pro Block und beim *Fully Associative Cache* 7 Bit pro Block benötigt. Bei schreibendem Zugriff soll darüber hinaus bei einem *Hit* das Verfahren *Copy-Back* angewendet werden, anderenfalls *Fetch-on-Write*.

Wie viel Speicher muß für die Verwaltung der drei Caches (alle gespeicherten Informationen außer den Nutzdaten) jeweils berücksichtigt werden?

4096 Blöcke
1 dirty-bit

Direct Mapped Cache:
4 KiB

8-Way Set Associative Cache:
20 KiB

Fully Associative Cache:
32 KiB

Aufgabe 2: Multilevel Caching

Gehen Sie von einem Prozessor mit folgenden Eckdaten aus:

- Prozessortakt: 8 GHz
- Zugriffsdauer auf L1-Cache: 4 Taktzyklen
- Zugriffsdauer auf L2-Cache: 40 Taktzyklen
- Zugriffsdauer auf Hauptspeicher: 100 Taktzyklen

Hinweis: In der Zugriffsdauer auf den L2-Cache ist die Zugriffsdauer für den *Miss* auf den L1-Cache bereits inkludiert. Selbiges gilt für die Zugriffsdauer auf den Hauptspeicher.

- a) Angenommen, die CPU führt 10.000 Adressanfragen aus. Die *Hit-Rate* (h) für den L1-Cache betrage 90% und für den L2-Cache 80%. Wie viele *Misses* treten beim L1-Cache bzw. beim L2-Cache auf?

L1-Misses: $10000 \cdot 10\% = 1000$

L2-Misses: $1000 \cdot 20\% = 200$

- b) Geben Sie die Zugriffsdauer auf den L1-Cache (t_{L1}), auf den L2-Cache (t_{L2}) und auf den Hauptspeicher (t_{main}) in Nanosekunden an.

$$\begin{aligned} f &= 8 \cdot 1000^3 &= 8 \times 10^9 \\ l_1 &= \frac{1}{f} \cdot 4 \cdot 1000^3 &= 0.5 \text{ ns} \\ l_2 &= \frac{1}{f} \cdot 40 \cdot 1000^3 &= 5 \text{ ns} \end{aligned}$$

- c) Um welchen Faktor verbessert sich die effektive Speicherzugriffszeit durch Verwendung des L2-Caches? Berechnen Sie dazu $t_{\text{eff},L1}$ für das Szenario ohne L2-Cache und $t_{\text{eff},L2}$ für das Szenario mit L2-Cache. Der Beschleunigungsfaktor ergibt sich dann als Quotient $t_{\text{eff},L1}/t_{\text{eff},L2}$.

Hinweis: Die effektive Speicherzugriffszeit wird bei Verwendung eines Caches grundsätzlich nach folgender Formel berechnet:

$$t_{\text{eff}} = h \cdot t_{\text{cache}} + (1 - h) \cdot t_{\text{main}}$$

$$\begin{aligned} t_{\text{eff},L1} &= 0.9 \cdot t_1 + 0.1 \cdot t_h &= 1.7 \text{ ns} \\ t_{\text{eff},L2} &= 0.9 \cdot t_1 + 0.08 \cdot t_2 + 0.02 \cdot t_h &= 1.1 \text{ ns} \\ \frac{t_{\text{eff},L1}}{t_{\text{eff},L2}} & &= 1.545454545 \end{aligned}$$

- d) Nehmen Sie an, dass die Kapazität des L2 Caches beträchtlich größer ist als die des L1 Caches. Gibt es, unter der weiteren Annahme, dass es sich um einen korrekten Hardwareentwurf handelt, eine Erklärung dafür, dass die Hitrate des L1 Caches höher ist als die des L2 Caches? Wenn ja, geben Sie eine mögliche Erklärung an. Wenn nein, erklären Sie, wieso das nicht möglich ist.

Beispiel: Die meisten Daten, die das Programm braucht liegen in der Nähe und somit im L1-Cache. Wenn das Programm doch mal andere Daten braucht, liegen diese irgendwo im Speicher und sind somit auch im L2-Cache unwahrscheinlich zu finden.

Aufgabe 3: Virtuelle Adressierung

Sie arbeiten auf einem Betriebssystem mit 32 Bit virtuellem Adressraum und 32 KiB großen Pages. Die Adressierung erfolgt byteweise. Nehmen Sie an, dass nur 1/8 (= 12.5%) des adressierbaren Speichers im System tatsächlich physisch verbaut ist.

- a) Wie lang ist der Offset?

$$32 \text{ KiB} = 2^{15} \Rightarrow 15 \text{ bits}$$

- b) Aus wie vielen Bits besteht die Page-Nummer?

$$\text{VPN} = 32 - 15 = 17$$

$$\text{PPN} = 14 \text{ (1/8 der Adressen)}$$

- c) Wie lang sind die physikalischen Adressen?

$$\text{PPN} + \text{Offset} = 29$$

- d) Wie viel Speicher ist im System physisch verbaut? Geben Sie den Wert in MiB an!

$$\frac{2^{29}}{1024^2} = 512 \text{ MiB}$$

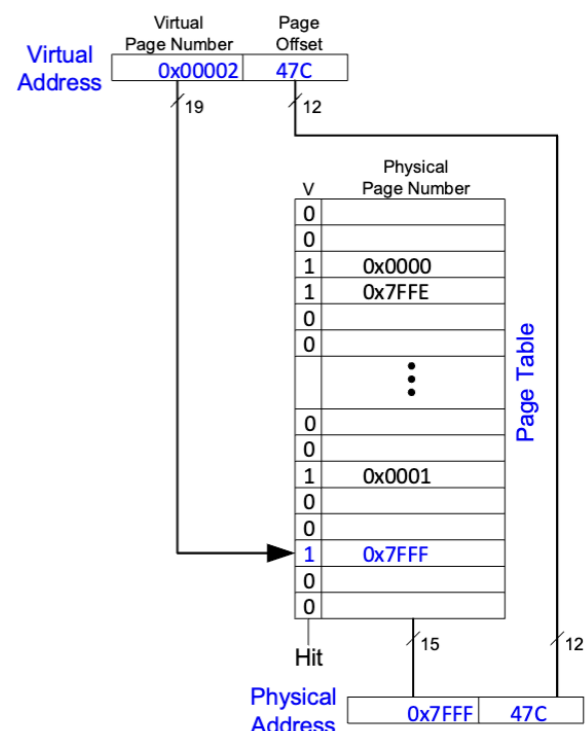
- e) Wie groß ist die Page-Table maximal, wenn Page-Nummern explizit gespeichert werden und jeder Eintrag zusätzlich 8 *Permission Bits* enthält? Geben Sie die Größe der Page-Table in KiB an!

Hinweis: Maximal bedeutet hier, dass zu jeder Page-Nummer ein Eintrag (ggf. mit lauter Nullen initialisiert) in der Page-Table angelegt wird.

$$2^{17} \cdot \frac{(1 + 14 + 8)}{1024} = 2944 \text{ KiB}$$

- f) Wie viele Frames muss sich das Betriebssystem reservieren, um die Page-Table aus Teilaufgabe e) im physikalischen Speicher zu halten?

$$2944/32 = 92$$



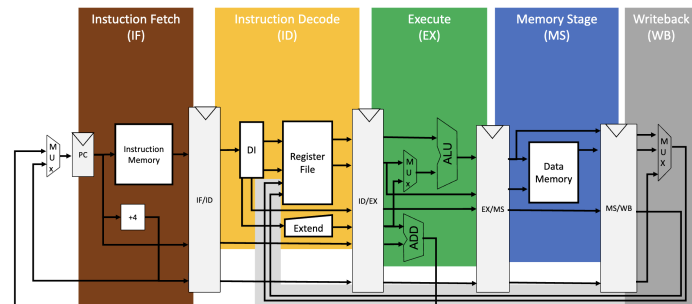
Aufgabe 4: Basic RISC-V Pipeline: Hazards und Branch Target Buffer

Gegeben ist der folgende RISC-V-Programmcode. Nehmen Sie an, dass dieser auf der aus der Vorlesung bekannten 5-stufigen Pipeline (in-order, skalar) ausgeführt wird, Instruction sowie Data Memory innerhalb des gleichen Zyklus Daten liefern (keine Verzögerung), Forwarding zunächst nicht implementiert ist, Sprünge in der dritten Pipelinestufe verarbeitet werden und zunächst **always not taken** als statische Sprungvorhersage genutzt wird.

```

1 0x00: function1: addi t0,zero,0
2 0x04:             addi t1,t0,0    DH t0 (1-2)
3 0x08:             addi t5,zero,0
4 0x0C:             addi t6,zero,0
5 0x10: oloop:     addi t2,zero,0
6 0x14: iloop:     t0,zero,0
7 0x18:             lw t3,0(a0)    KH (6)
8 0x1C:             add t5,t5,t3   DH t3 (7-8) KH (6)
9 0x20:
10 0x24: else:      lw t4,0(a0)    KH (9)
11 0x28:             add t6,t6,t4   DH t4 (10-11) KH (9)
12 0x2C: join:      xori t0,t0,1
13 0x30:             addi a0,a0,4
14 0x34:             addi t2,t2,1
15 0x38:             t2,a2,        DH t2 (14-15)
16 0x3C:             addi t1,t1,1   KH (15)
17 0x40:             t1,a1,        DH t1 (16-17) KH (15)
18 0x44:             sub a0,t6,t5   KH (17)
19 0x48:             ret            KH (17)

```



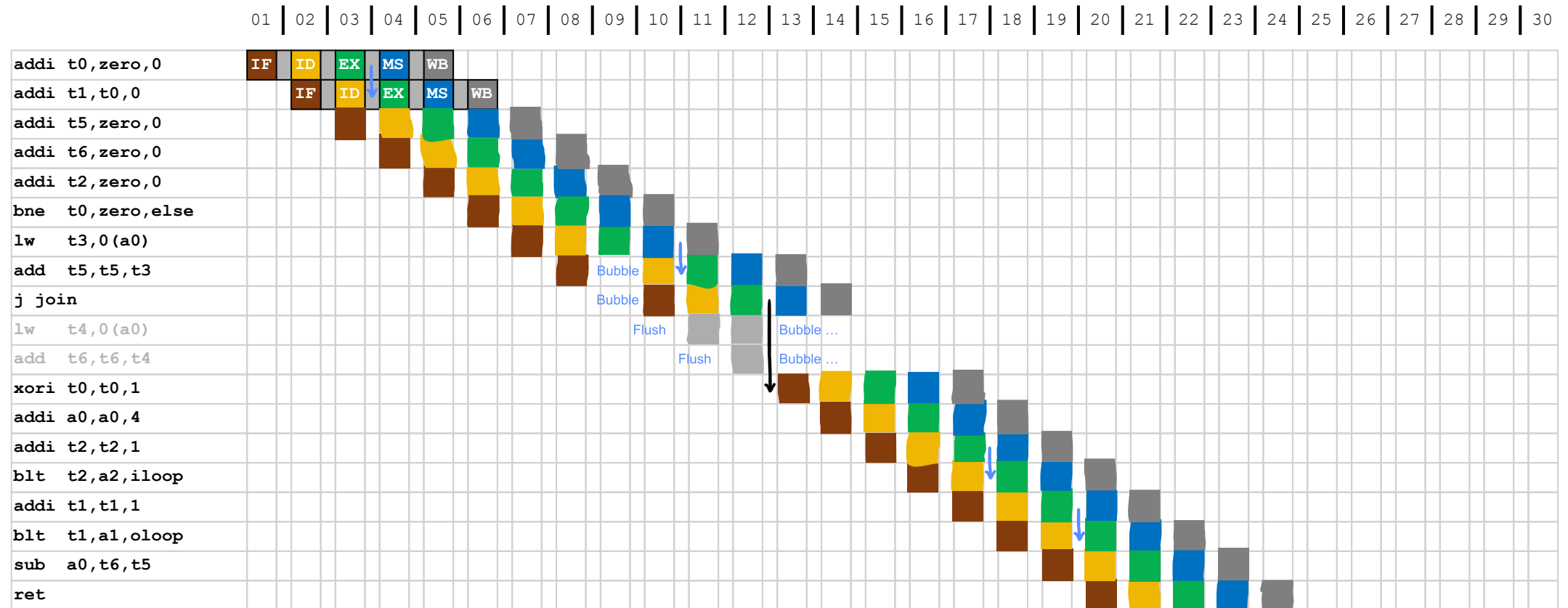
- a) Ermitteln Sie alle Codestellen, bei denen Datenhazards (DH) und Kontrollhazards (KH) bei Ausführung des gegebenen Codes auftreten könnten. Nutzen Sie dafür folgende Notation: $DH_{reg}(ZQI - ZZI)$ bzw. $KH(ZQI)$, wobei reg =Registernummer, ZQI =Zeile der Quellinstruktion und ZZI =Zeile der Zielinstruktion.

Beispiel: $DH_{t0}(1 - 2)$, d.h. ein Datenhazard, der das Register `t0` betrifft und durch das Schreiben von `t0` in Zeile 1 und das nachfolgende Lesen von `t0` in Zeile 2 ausgelöst wird.

Siehe oben

- b) Erstellen Sie ein Pipelinediagramm für den gegebenen Code, das die Ausführung auf der oben beschriebenen Pipeline zeigt, wobei diese jedoch um Forwarding erweitert wurde. Sonstige Kontroll- und Datenhazards sollen automatisch in der Pipeline durch Einfügen von Stalls behandelt werden. Kennzeichnen Sie in Ihrem Pipelinediagramm die auftretenden Stalls. Nehmen Sie an, dass $a1 = 1$ und $a2 = 1$.

Nächste Seite



- c) Die gegebene Pipeline mit Forwarding wird nun um eine Einheit zur dynamischen Sprungvorhersage erweitert, die sich in der ersten Pipelinestufe befindet. In dieser Einheit kommt ein Branch Target Buffer (BTB) zum Einsatz, der für den Fall, dass ein Sprung zu **taken** ausgewertet wird, die Adresse des Sprungs selbst (BIA) sowie die dazugehörige Sprungzieladresse (BTA) speichert. Gehen Sie davon aus, dass der BTB zu Beginn leer ist und über ausreichend Kapazität verfügt, um alle in diesem Codeausschnitt ausgeführten Sprünge speichern zu können (nehmen Sie an, dass die erste Instruktion an Adresse 0x0 liegt). Gehen Sie weiterhin davon aus, dass die Richtung des Sprunges (taken/not taken) durch einen Prädiktor stets korrekt vorhergesagt wird, bei einem BTB Miss jedoch die Vorhersage auf **not taken** geändert wird. Zeigen Sie den Verlauf der Änderungen im BTB, d.h. Zugriffe und mögliche Änderungen der Einträge im BTB, die zugreifende Instruktion (nehmen Sie an, dass nur **branch** und **j/jal** Instruktionen auf den BTB zugreifen und letztere hier im Hinblick auf den BTB wie eine branch Instruktion behandelt werden sollen) und ob der Zugriff ein Hit oder Miss war.

Nehmen Sie hier an, dass $a1 = 1$ und $a2 = 2$. Ein entsprechendes Protokoll der ausgeführten Instruktionen finden Sie im Folgenden:

```

1 0x00: addi t0,zero,0
2 0x04: addi t1,t0,0
3 0x08: addi t5,zero,0
4 0x0C: addi t6,zero,0
5 0x10: addi t2,zero,0
6 0x14: bne t0,zero,else Predict not taken, correct guess
7 0x18: lw t3,0(a0)
8 0x1C: add t5,t5,t3
9 0x20: j join First occurrence, miss, save to btb
10 0x2C: xori t0,t0,1
11 0x30: addi a0,a0,4
12 0x34: addi t2,t2,1
13 0x38: blt t2,a2,loop Branch taken, miss, save to btb
14 0x14: bne t0,zero,else Branch taken, miss, save to btb
15 0x24: lw t4,0(a0)
16 0x28: add t6,t6,t4
17 0x2C: xori t0,t0,1
18 0x30: addi a0,a0,4
19 0x34: addi t2,t2,1
20 0x38: blt t2,a2,loop Predict not taken
21 0x3C: addi t1,t1,1
22 0x40: blt t1,a1,loop Predict not taken
23 0x44: sub a0,t6,t5
24 0x48: ret

```

Füllen Sie folgende Tabellen aus:

Initial state:

Valid	BIA	BTA
0		
0		
0		
0		
0		
0		

Valid	BIA	BTA
1	0x20	0x2C

Accessed BIA: 0x20
Hit (H)/Miss (M): M

Valid	BIA	BTA
1	0x20	0x2C
1	0x38	0x14

Accessed BIA: 0x38
Hit (H)/Miss (M): M

Valid	BIA	BTA
1	0x20	0x2C
1	0x38	0x14
1	0x14	0x24

Accessed BIA: 0x38
Hit (H)/Miss (M): M

Valid	BIA	BTA

Accessed BIA:
Hit (H)/Miss (M):

Valid	BIA	BTA

Accessed BIA:
Hit (H)/Miss (M):

Valid	BIA	BTA

Accessed BIA:
Hit (H)/Miss (M):

Valid	BIA	BTA

Accessed BIA:
Hit (H)/Miss (M):

- b) Es wird ein lokaler 2-bit Prädiktor verwendet, d.h. der Prädiktor wird für alle relevanten Sprünge getrennt genutzt. Nehmen Sie an, dass die initiale Vorhersage jeweils **predict weakly not taken** (PWNT) lautet.

	01	02	03	04	05	06	07	08	09
Sprung (Zeile)	Z06	Z15	Z06	Z15	Z06	Z15	Z06	Z15	Z17
Vorhersage (Z06)	PWNT		PSNT		PWNT		PSNT		
Vorhersage (Z15)	-	PWNT		PWT		PST		PST	
Vorhersage (Z17)	-								PWNT
Sprungrichtung	NT	T	T	T	NT	T	T	NT	T
Korrekt?	●	●	●	●	●	●	●	●	●

	10	11	12	13	14	15	16	17	18
Sprung (Zeile)	Z06	Z15	Z06	Z15	Z06	Z15	Z06	Z15	Z17
Vorhersage (Z06)	PWNT		PSNT		PWNT		PSNT		
Vorhersage (Z15)		PWT		PST		PST		PST	
Vorhersage (Z17)									PWT
Sprungrichtung	NT	T	T	T	NT	T	T	NT	NT
Korrekt?	●	●	●	●	●	●	●	●	●

- c) Welche Fehlvorhersagequote (**misprediction rate**) kann für die beiden oben genannten Konfigurationen der Sprungvorhersage in der **function1**-Funktion beobachtet werden?

1-bit Predictor: 10/18 = 55% misprediction rate

2-bit Predictor: 9/18 = 50% misprediction rate

Aufgabe 6: Advanced Pipelines: Dependencies

Gegeben ist der folgende RISC-V-Programmcode:

```

1 xori x10, x0, 7
2 lw x22, 0(x6)
3 lw x11, 4(x6)
4 sw x11, 0(x7)
5 sw x22, 4(x7)
6 mul x12, x10, x11
7 andi x13, x10, 15
8 div x14, x8, x13
9 mul x15, x11, x11
10 div x16, x8, x9
11 mul x13, x10, x16
12 addi x17, x14, 8
13 mul x18, x12, x12
14 mul x19, x13, x9
15 ori x19, x19, 31
16 sw x17, 8(x7)

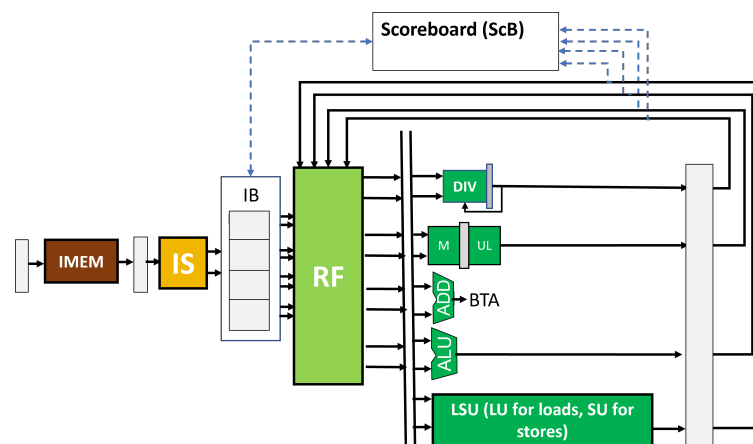
```

Ermitteln Sie alle echten Datenabhängigkeiten (true dependencies (TD)), Gegenabhängigkeiten (anti dependencies (AD)) und Ausgabeabhängigkeiten (output dependencies (OD)) im oben genannten Code. Nutzen Sie dafür folgende Notation: $\{T|A|O\}D_{reg}(ZQI - ZZI)$, wobei ZQI =Zeile der Quellinstruktion, ZZI =Zeile der Zielinstruktion und reg =Registernummer.

RAW/True Dependencies	WAR/Anti Dependencies	WAW/Output Dependencies
Beispiel: $TD_{x19}(14 - 15)$ TD x13 (11-14) TD x12 (6-13) TD x14 (8-12) TD x16 (10-11) TD x10 (1-11) TD x11 (4-9) TD x13 (7-8) TD x10 (1-7) TD x11 (4-6) TD x10 (1-6)	AD x13 (8-11)	OD x17 (12-16) OD x19 (14-15) OD x13 (7-11) OD x22 (2-5) OD x11 (3-4)

Aufgabe 7: Advanced Pipelines: Scoreboard

Gegeben ist die folgende Pipeline:



Diese hat die folgenden Eigenschaften:

- Issue Buffer-Größe (IB size) von 4
- 4 Ports um Instruktionen aus dem Issue Buffer zu emittieren (instruction issue)
- 4 Ports um Daten zurückzuschreiben (writeback)

Es sind folgende funktionale Einheiten vorhanden:

- ALU: 1 Zyklus Latenz; 1 Zyklus Initialisierungsintervall
- ADD: 1 Zyklus Latenz; 1 Zyklus Initialisierungsintervall
- MUL: 2 Zyklen Latenz; 1 Zyklus Initialisierungsintervall (pipelined)
- DIV: 4 Zyklen Latenz; 4 Zyklen Initialisierungsintervall (serial)
- LSU:
 - LU: 2 Zyklen Latenz; 1 Zyklus Initialisierungsintervall (non-blocking)
 - SU: 1 Zyklus Latenz, danach Store Buffer; 1 Zyklus Initialisierungsintervall

Gegeben ist zudem folgender RISC-V-Programmcode:

```
1 xori x10,x0,7
2 lw x22,0(x6)
3 lw x11,4(x6)
4 sw x11,0(x7)
5 sw x22,4(x7)
6 mul x12,x10,x11
7 andi x13,x10,15
8 div x14,x8,x13
9 mul x15,x11,x11
10 div x16,x8,x9
11 mul x13,x10,x16
12 addi x17,x14,8
13 mul x18,x12,x12
14 mul x19,x13,x9
15 ori x19,x19,31
16 sw x17,8(x7)
```

- a) Zeigen Sie für die genannte Pipeline die Ausführung des Codes. Erstellen Sie dazu ein Pipelinediagramm und zeigen Sie den Zustand des Scoreboards und den Status der Funktionseinheiten in jedem Zyklus. Nehmen Sie an, dass kein Register Renaming verwendet wird.

01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

[illegible]

Cycle: 02

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

Cycle: ----

Instr.	rd	rs1	rs2	Imm	RO	Complete

DIV	MUL	ALU	ADD	SU	LU

b) Bestimmen Sie den erreichten CPI-Wert des Codes bei der Ausführung auf dieser Pipeline.