INSTITUTO FEDERAL Brasília	**Instituto Federal de Brasília** **Campus Taguatinga** **Superior em Computação** **Exercício de Fixação 3** **Arquitetura de Computadores II - 2023/1**

1) Considere que temos um computador onde os ciclos de clock por instrução (CPI) seja de 1,0 quando todos os acessos à memória têm acertos na cache. Os únicos acessos a dados são loads e stores, e estes totalizam 50% das instruções. Se a penalidade por falta for 50 ciclos de clock e a taxa de falta for 1%, quão mais rápido o computador será se todas as instruções forem acertos de cache?

```
Desempenho para computador que sempre acerta:

Tempo de execução = (Ciclos de clock da CPU + Ciclos de Stall da memória) *
 Ciclo de Clock

=((IC * CPI)+0) * Ciclo de clock
=IC * 1,0 * Ciclo de Clock

Para o computador com a cache real, primeiro calculo o ciclo de stall de me
mória

Ciclos de Stall = IC * ((Acesso à memória / instrução) * Taxa de falta) * p
enalidade de falta

= IC * (1 + 0,5) * 0,01 * 50 = IC * 0,75

(1+0,5) indica 1 acesso à instrução e 0,5 acesso à dados por instrução(50%)

tempo de execução com a cache:

Tempo de execução da CPU(cache) = (IC * 1,0 + IC * 0,75) * Ciclo de Clock
=1,75 * IC * Ciclo de clock

O desempenho é o inverso dos tempos de execução

Tempo de execução CPU(cache) / Tempo de execução CPU = 1,75 * IC * Ciclo de
 Clock / 1,0 * IC * Ciclo de Clock = 1,75

Computador sem faltas é 1,75 vezes mais rápido.
```

2) Considere uma cache write-back totalmente associativa com muitas entradas de cache que começam vazias. A seguir apresentamos uma sequência de cinco operações de memória (o endereço está entre colchetes):

Read Mem[300];
Write Mem[100];
Write Mem[100];
Read Mem[200];
Write Mem[200];
Read Mem[100];
Write Mem[100].
Write Mem[100];

Quais são os números de acertos e faltas quando se usam a no-write allocate e a write allocate?

```
No write allocate                        writallocate
  miss              Read Mem[300];        miss
  miss              Write Mem[100];       hit
  miss              Write Mem[100];       hit
  miss              Read Mem[200];        miss
  hit               Write Mem[200];       hit
  miss              Read Mem[100];        hit
  hit               Write Mem[100];       hit
  hit              Write Mem[100];       hit

5 miss                                    2 miss
3 hit                                     6 hit
```

3) Qual tem a menor taxa de falta: uma cache de instruções de 16 KiB com uma cache de dados de 16 KiB ou uma cache unificada de 32 KiB? Use as taxas de falta da Figura abaixo para ajudar a calcular a resposta correta, considerando que 36% das instruções são instruções de transferência de dados. Considere que um acerto utiliza um ciclo de clock e a penalidade por falta é de 100 ciclos de clock. Um acerto no load ou store utiliza um ciclo de clock extra em uma cache unificada se houver apenas uma porta de cache para atender a duas solicitações simultâneas. Qual é o tempo médio de acesso à memória em cada caso? Considere caches write-through com um buffer de escrita e ignore os stalls devidos ao buffer de escrita.

Tamanho (KiB)	Cache de instruções	Cache de dados	Cache unificada
8	8,16	44,0	63,0
16	3,82	40,9	51,0
32	1,36	38,4	43,3
64	0,61	36,9	39,4
128	0,30	35,3	36,2
256	0,02	32,6	32,9

Figura: Taxa de Falta por 1000 instruções

```
36% são instruções de transferência de dados
penalidade por falta - 100

tempo médio de acesso a memória = tempo de acerto + taxa de falta + penalid
ade da falta

taxa de falta = ((faltas/1000 instruções)/100)/ (acesso a memória / instruç
ão)

taxa de falta 16KB(instrução) = (3,82/1000) / 1 = 0,004

36% são transferências de dados, logo:
taxa de falta 16KB(dados) = (40,9/100) / 0,36 = 0,114

a taxa de falta unificada leva em consideração os acessos a dados e instruç
ões

taxa de falta unificada 32KB = (43,3/1000) / (1 + 0,36) = 0,318

a cada 100 acessos à cache de instruções, 36 acessos são a cache de dados:
(100/136) * 100 ~= 74% a memória de instrução

(36/136) * 100 ~= 36% a memória de dados

taxa de falta geral para as caches divididas

(0,74 * 0,004) + (0,26 * 0,114) = 0,0326

uma cache unificada de 32KB tem taxa de falta efetiva ligeaimente meor que
de duas caches de 16KB

Tempo médio de acesso a memória = %instrução * (taxa de acerto + taxa de fa
lta de instrução + penalidade de falta) +
                                   %dados * (taxa de acerto + taxa de falta
de instrução + penalidade de falta)

cache dividida = 0,74 * (1 + 0,004 * 200) + 0,26 * (1 + 0,114 * 200) = 7,52

cache unificada = 0,74 * (1 + 0,0318 * 200) + 0,26 * (1+ 0,0318 * 200) = 7,
62
```

Bons estudos!