

Memorias Cache

Maximiliano A. Eschoyez
Programación Eficiente — UBP

2010

Mediciones

- *Hit Rate*: Tasa de cantidad de aciertos en la memoria cache, es decir, cuando se va a buscar un dato que está allí. Como este valor suele ser muy alto, se utiliza *Miss Rate*.
- *Average Memory–Access Time*: Tiempo promedio para acceder a un dato:

$$= \text{Hit Time} + \text{Miss Rate} \times \text{Miss Penalty} \quad [\eta s \text{ o clocks}]$$

- *Miss Penalty*: Tiempo para reemplazar un bloque de memoria desde un nivel inferior, incluyendo el reemplazo en la CPU:
 - *Access Time*: Retardo desde el nivel inferior
 - *Transfer Time*: Tiempo de transferencia del bloque

Terminología

- *Hit*: Los datos se encuentran en algún bloque de niveles superiores:
 - *Hit Rate*: Fracción de accesos a memoria encontrados en los niveles superiores
 - *Hit Time*: Tiempo para acceder a los niveles superiores, que consiste en tiempo de acceso a la RAM + tiempo para determinar *Hit/Miss*
- *Miss*: Se necesita traer los datos desde niveles inferiores:
 - *Miss Rate* = 1 - (*Hit Rate*)
 - *Miss Penalty*: el tiempo para reemplazar un bloque en un nivel superior + el tiempo para entregar el bloque al procesador
- *Hit Time* << *Miss Penalty*

Performance

El tiempo de ejecución de un programa está dado por

$$T = I_{\text{count}} \times \text{CPI} \times T_{\text{cycle}} \quad (1)$$

$$I_{\text{count}} = I_{\text{ALU}} + I_{\text{MEM}} \quad (2)$$

$$\text{CPI} = \frac{I_{\text{ALU}}}{I_{\text{count}}} \times \text{CPI}_{\text{ALU}} + \frac{I_{\text{MEM}}}{I_{\text{count}}} \times \text{CPI}_{\text{MEM}} \quad (3)$$

Desarrollando las instrucciones que deben realizarse

$$M_{\text{ALU}} = \frac{I_{\text{ALU}}}{I_{\text{count}}} \quad (4)$$

$$M_{\text{MEM}} = \frac{I_{\text{MEM}}}{I_{\text{count}}} \quad (5)$$

$$1 = M_{\text{ALU}} + M_{\text{MEM}} \quad (6)$$

$$\text{CPI} = M_{\text{ALU}} \times \text{CPI}_{\text{ALU}} + M_{\text{MEM}} \times \text{CPI}_{\text{MEM}} \quad (7)$$

$$T = I_{\text{count}} \times (M_{\text{ALU}} \times \text{CPI}_{\text{ALU}} + M_{\text{MEM}} \times \text{CPI}_{\text{MEM}}) \times T_{\text{cycle}} \quad (8)$$

$$\text{CPI}_{\text{MEM}} = \text{CPI}_{\text{MEM-HIT}} + r_{\text{MISS}} \times \text{CPI}_{\text{MEM-MISS}} \quad (9)$$

Quedando el tiempo total como

$$T = I_{\text{count}} \times [M_{\text{ALU}} \times \text{CPI}_{\text{ALU}} + M_{\text{MEM}} \times (\text{CPI}_{\text{MEM-HIT}} + r_{\text{MISS}} \times \text{CPI}_{\text{MEM-MISS}})] \times T_{\text{cycle}} \quad (10)$$

Donde:

T = Tiempo total de ejecución

T_{cycle} = Tiempo de un ciclo de procesador

I_{count} = Número total de instrucciones

I_{ALU} = Número de instrucciones de ALU (registro-registro)

I_{MEM} = Número de instrucciones de memoria (load/store)

CPI = Promedio de ciclos por instrucción

CPI_{ALU} = Promedio de ciclos por instrucción de ALU

CPI_{MEM} = Promedio de ciclos por instrucción de memoria

r_{MISS} = Tasa de *cache miss*

r_{HIT} = Tasa de *cache hit*

$\text{CPI}_{\text{MEM-MISS}}$ = Promedio de ciclos por *cache miss*

$\text{CPI}_{\text{MEM-HIT}}$ = Promedio de ciclos por *cache hit*

M_{ALU} = Mezcla de instrucciones de ALU

M_{MEM} = Mezcla de instrucciones de memoria

Ejemplos

T ? = Tiempo total de ejecución

$T_{\text{cycle}} = 0,5 \text{ ns}$

$I_{\text{count}} = 10^{11}$

$I_{\text{MEM}} = 2 \times 10^{10}$

$\text{CPI}_{\text{ALU}} = 1$

$\text{CPI}_{\text{MEM-MISS}} = 100$

$\text{CPI}_{\text{MEM-HIT}} = 1$

$$I_{\text{ALU}} = I_{\text{count}} - I_{\text{MEM}} = 8 \times 10^{10}$$

$$M_{\text{ALU}} = \frac{I_{\text{ALU}}}{I_{\text{count}}} = \frac{8 \times 10^{10}}{10^{11}} = \frac{8}{10} = 0,8$$

$$M_{\text{MEM}} = \frac{I_{\text{MEM}}}{I_{\text{count}}} = \frac{2 \times 10^{10}}{10^{11}} = \frac{2}{10} = 0,2$$

Para $r_{\text{HIT}} = 0,9$

$$\begin{aligned}
r_{MISS} &= 1 - r_{HIT} = 0,1 \\
CPI_{MEM} &= CPI_{MEM-HIT} + r_{MISS} \times CPI_{MEM-MISS} \\
&= 1 + (1 - 0,9) \times 100 = 1 + 10 = 11 \\
T &= 10^{11} \times [(0,8 \times 1) + (0,2 \times 11)] \times 5 \times 10^{-10} \\
&= 150 \text{ s}
\end{aligned}$$

Para $r_{HIT} = 0,7$

$$\begin{aligned}
r_{MISS} &= 1 - r_{HIT} = 0,3 \\
CPI_{MEM} &= CPI_{MEM-HIT} + r_{MISS} \times CPI_{MEM-MISS} \\
&= 1 + (1 - 0,7) \times 100 = 1 + 30 = 31 \\
T &= 10^{11} \times [(0,8 \times 1) + (0,2 \times 31)] \times 5 \times 10^{-10} \\
&= 205 \text{ s}
\end{aligned}$$

Para $r_{HIT} = 0,5$

$$\begin{aligned}
r_{MISS} &= 1 - r_{HIT} = 0,5 \\
CPI_{MEM} &= CPI_{MEM-HIT} + r_{MISS} \times CPI_{MEM-MISS} \\
&= 1 + (1 - 0,5) \times 100 = 1 + 50 = 51 \\
T &= 10^{11} \times [(0,8 \times 1) + (0,2 \times 51)] \times 5 \times 10^{-10} \\
&= 550 \text{ s}
\end{aligned}$$