Essentials of Computer Systems - Exercises #5

1 Performance Equations - continued

Exercise 1.1 *CPI and Memory Hierarchy*

When running a given benchmark with 30% of load-store insns in the insn mix, we found the following:

- *Assuming perfect memory, the average CPI is 2.*
- *The I$ miss rate is 0% (always a hit).*
- *the D$ miss rate is 4%, and the D$ miss penalty is 100 cycles*

For the given benchmark, compare the execution time assuming perfect[1] *memory with execution time considering the real system.*

Solution:
Recall execution time equation $T = IC \times CPI \times t_{\text{CLK}} = \frac{\#insn}{program} \cdot \frac{\#cycles}{insn} \cdot \frac{\#seconds}{cycle}$.

We extend the execution time T *as follows:* $T_{\text{real mem}} = T_{\text{CPU}} + T_{\text{wait}}$, *the time spent executing on the CPU and the time spent waiting*[2] *for memory hierarchy to react. Assuming perfect memory, the memory access is always a hit, thus no waiting time (*$T_{\text{wait}} = 0$*), that is* $T_{\text{perfect mem}} = T_{\text{CPU}}$. *Our task is to compute the following:*

$$\text{Speedup} = \frac{T_{\text{real mem}}}{T_{\text{perfect mem}}} = \frac{T_{\text{CPU}} + T_{\text{wait}}}{T_{\text{CPU}}}$$

1. *The* T_{CPU} *calculation is straightforward:* $T_{\text{CPU}} = IC \times CPI \times t_{\text{CLK}} = IC \cdot \frac{\#cycles}{insn} \cdot t_{\text{CLK}} = IC \cdot 2 \cdot t_{\text{CLK}}$.

2. *For the* T_{wait} *calculation, we need to consider:*
   - *the fraction of instructions that result in a wait:* IC_{wait}
   - *#penalty cycles, i.e., wait cycles: in this example we have* $\frac{\#\text{penalty cycles}}{insn} = 100$

   Let us think about IC_{wait} *further:*
   - *The waiting is caused by D$ access: only the load and store insn, that is only 30% of all insn,* **can cause** *this wait.*
   - **can cause** *means that we only consider the miss rate: only 4% of all load-store insn cause the wait.*
   - *hence* $IC_{\text{wait}} = IC \times 0.3 \times 0.04$

   The T_{wait} *calculation is as follows:* $T_{\text{wait}} = (IC \times 0.3 \times 0.04) \cdot 100 \cdot t_{\text{CLK}}$

3. *Finally, we assemble the equation and see that* IC *and* t_{CLK} *cancel out:*

$$\text{Speedup} = \frac{T_{\text{real mem}}}{T_{\text{perfect mem}}} = \frac{T_{\text{CPU}} + T_{\text{wait}}}{T_{\text{CPU}}} = \frac{IC \cdot 2 \cdot t_{\text{CLK}} + (IC \times 0.3 \times 0.04) \cdot 100 \cdot t_{\text{CLK}}}{IC \cdot 2 \cdot t_{\text{CLK}}} = \frac{2 + 0.3 \times 0.04 \times 100}{2} = 1.6$$

In conclusion: a perfect memory would yield a speedup factor 1.6.

---

[1] always a hit, no need to wait for memory hierarchy to react
[2] the CPU is stalling while waiting for memory

Exercise 1.2 *CPI and Memory Hierarchy*

When running a given benchmark with 30% of load-store insns in the insn mix, we found the following:

- *Assuming perfect memory, the average CPI is 2.*
- *The I$ miss rate is 2%, and the I$ miss penalty is 80 cycles*
- *the D$ miss rate is 4%, and the D$ miss penalty is 100 cycles*

For the given benchmark, compare the execution time assuming perfect memory with execution time considering the real system.

Solution:
Recall execution time equation $T = IC \times CPI \times t_{\text{CLK}} = \frac{\#insn}{program} \cdot \frac{\#cycles}{insn} \cdot \frac{\#seconds}{cycle}$.

We extend the execution time T *as follows:* $T_{\text{real mem}} = T_{\text{CPU}} + T_{\text{wait}}$*, the time spent executing on the CPU and the time spent waiting for memory hierarchy to react. Assuming perfect memory, the memory access is always a hit, thus no waiting time (*$T_{\text{wait}} = 0$*), that is* $T_{\text{perfect mem}} = T_{\text{CPU}}$*. Our task is to compute the following:*

$$\text{Speedup} = \frac{T_{\text{real mem}}}{T_{\text{perfect mem}}} = \frac{T_{\text{CPU}} + T_{\text{wait}}}{T_{\text{CPU}}}$$

1. *The* T_{CPU} *calculation is straightforward:* $T_{\text{CPU}} = IC \times CPI \times t_{\text{CLK}} = IC \cdot \frac{\#cycles}{insn} \cdot t_{\text{CLK}} = IC \cdot 2 \cdot t_{\text{CLK}}$.

2. *For the* T_{wait} *calculation, we need to consider two causes:*

   - *the D$ misses with penalty 100 (see previous exercise:* $IC_{\text{D\$ wait}} = IC \times 0.3 \times 0.04$*)*
   - *and the I$ misses with penalty 80, whereby:*
     - *The waiting is caused by I$ access: all insn, that is 100% of all insn,* **can cause** *this wait.*
     - **can cause** *means that we only consider the miss rate: only 2% of all insn cause the wait.*
     - *hence* $IC_{\text{I\$ wait}} = IC \times 1 \times 0.02$

   The T_{wait} *calculation is as follows:* $T_{\text{wait}} = (IC \times 0.3 \times 0.04) \cdot 100 \cdot t_{\text{CLK}} + (IC \times 1 \times 0.02) \cdot 80 \cdot t_{\text{CLK}}$

3. *Finally, we assemble the equation and see that* IC *and* t_{CLK} *cancel out:*

$$\text{Speedup} = \frac{T_{\text{real mem}}}{T_{\text{perfect mem}}} = \frac{T_{\text{CPU}} + T_{\text{wait}}}{T_{\text{CPU}}} = \frac{IC \cdot 2 \cdot t_{\text{CLK}} + (IC \times 0.3 \times 0.04) \cdot 100 \cdot t_{\text{CLK}} + (IC \times 1 \times 0.02) \cdot 80 \cdot t_{\text{CLK}}}{IC \cdot 2 \cdot t_{\text{CLK}}} =$$

$$= \frac{2 + 0.3 \times 0.04 \times 100 + 1 \times 0.02 \times 80}{2} = 2.4$$

In conclusion: a perfect memory would yield a speedup factor 2.4.