

Exercise walk-through

6.6

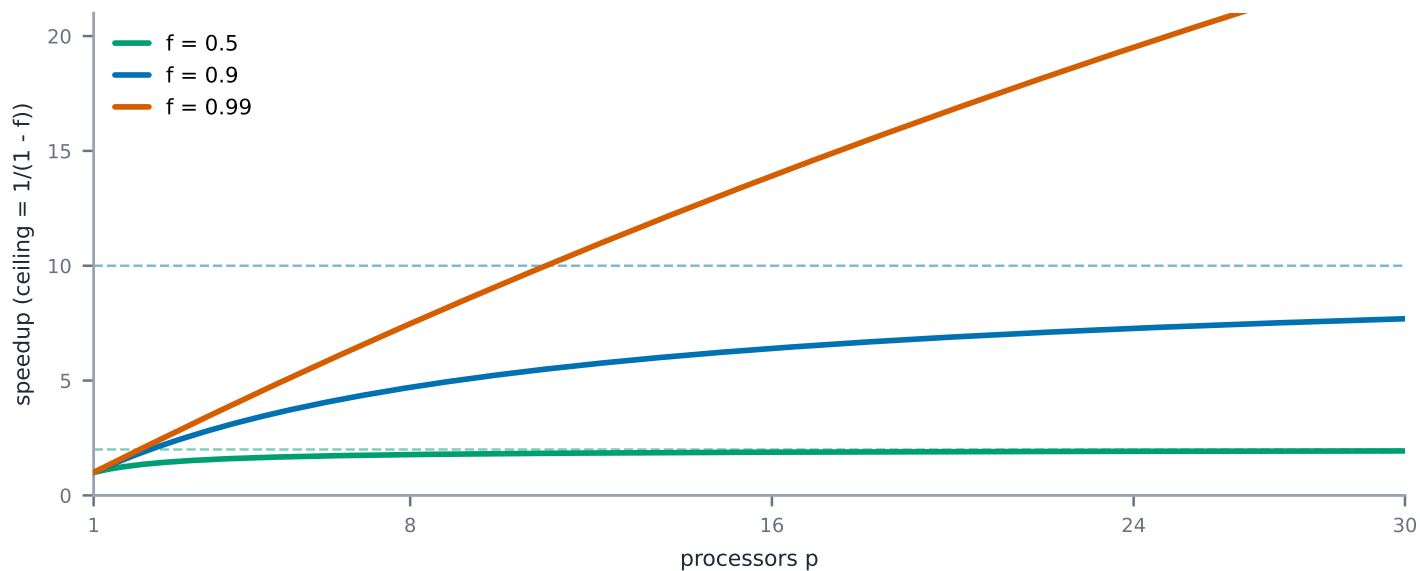
Parallelism and performance

University Computer Science

Parallelism and performance: read Amdahl

- 1 **Ceiling** $f = 0.9$ caps every machine at 10 times; $p = 8$ gives 4.71, $p = 16$ gives 6.40 - the second eight add 1.7.
- 2 **Price** 36 s serial + 64 s parallel: $f = 0.64$; on 4 processors 1.92 times and 52.0 s, on 8 it is 2.27 and 44.0 s.
- 3 **Exact** $f = 0.99$ reaches exactly 50 times at $p = 99$: diminishing returns are structural.

Amdahl speedup curves



The serial fraction caps every speedup curve.

Check yourself

- A 16-bit signed immediate tops out at...? (32 767.)
- Five instructions, five stages, total cycles...? (9.)
- 4 KiB cache, 16-byte lines: **index** bits...? (8.)
- $f = 0.9$: the speedup ceiling...? (10 times.)

Exercises 2.1-2.3

2.1 $f = 0.8$: state the ceiling. [3]

2.2 $f = 0.75$, $p = 3$: compute the speedup to two decimals. [3]

2.3 Name the three rungs of the parallelism ladder. [3]

Exercises 2.1-2.3

2.1 $f = 0.8$: state the ceiling. [3]

2.1

- $1/(1 - 0.8) = 5 \times \checkmark$.

2.2 $f = 0.75, p = 3$: compute the speedup to two decimals. [3]

2.2

- $1/(0.25 + 0.75/3) = 1/0.5 = 2.00 \times \checkmark$.

2.3 Name the three rungs of the parallelism ladder. [3]

2.3

- ILP within a core, threads per core, cores per chip $\checkmark$.

Exercise 3.1

3.1 A workload spends 12 s serial and 88 s parallelizable on one core.

Compute f . Compute the speedup and runtime on $p = 4$ and $p = 16$. Compute the smallest p that reaches a $4\times$ speedup (solve $1/((1-f) + f/p) \geq 4$ for p). State the ceiling and what it means for a planned 64-core purchase. [8]

Exercise 3.1

3.1 A workload spends 12 s serial and 88 s parallelizable on one core.

Compute f . Compute the speedup and runtime on $p = 4$ and $p = 16$. Compute the smallest p that reaches a $4\times$ speedup (solve $1/((1-f) + f/p) \geq 4$ for p). State the ceiling and what it means for a planned 64-core purchase. [8]

3.1

- $f = 88/100 = 0.88$ ✓.
- $p = 4$: $1/(0.12 + 0.22) = 2.94\times$, runtime 34.0 s ✓.
- $p = 16$: $1/(0.12 + 0.055) = 5.71\times$, runtime 17.5 s ✓.
- $4\times$ needs $0.12 + 0.88/p \leq 0.25 \Rightarrow 0.88/p \leq 0.13 \Rightarrow p \geq 6.77$ – smallest integer $p = 7$ ✓.
- Ceiling $= 1/0.12 \approx 8.3\times$: 64 cores cannot pass it – the serial 12 s owns the rest ✓.

Exercise 3.2

3.2 Two optimisations are proposed for a program that runs 50 s: A parallelises 90% of the time on 8 cores; B cuts the serial 10% remaining time in half.

Compute A's speedup and new runtime. Compute B's speedup and new runtime (no parallelism). Compute the combined effect of A then B, and state which single change a buyer should prefer at 8 cores, with the reason in one line. [7]

Exercise 3.2

3.2 Two optimisations are proposed for a program that runs 50 s: A parallelises 90% of the time on 8 cores; B cuts the serial 10% remaining time in half.

Compute A's speedup and new runtime. Compute B's speedup and new runtime (no parallelism). Compute the combined effect of A then B, and state which single change a buyer should prefer at 8 cores, with the reason in one line. [7]

3.2

- A: $f = 0.9$, $p = 8$: $1/(0.1 + 0.1125) = 4.71\times$, runtime $50/4.71 = 10.6$ s ✓.
- B: the serial 5 s halves to 2.5 s: new time = $2.5 + 45 = 47.5$ s; speedup $50/47.5 = 1.05\times$ ✓.
- Combined: serial $2.5 + 45/8 = 2.5 + 5.6 = 8.1$ s, speedup $50/8.125 = 6.15\times$ ✓.
- Prefer A: one line – cutting the parallel part multiplies; cutting the serial tail by half moved 2.5 s ✓.