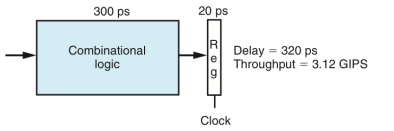
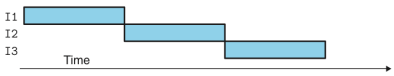
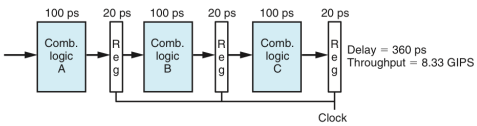


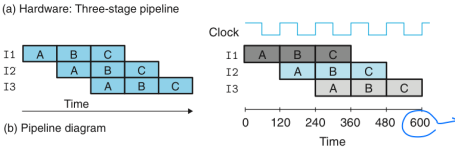
(a) Hardware: Unpipelined



(b) Pipeline diagram



(a) Hardware: Three-stage pipeline



(b) Pipeline diagram

Measure circuit delays in units of picoseconds (ps) → 10^{-12} seconds

Throughput : maximum rate at which we could operate the system ← limited by the slowest stage

$$Throughput = \frac{1 \text{ instruction}}{(20 + 300) \text{ picoseconds}} \cdot \frac{1{,}000 \text{ picoseconds}}{1 \text{ nanosecond}} \approx 3.12 \text{ GIPS}$$

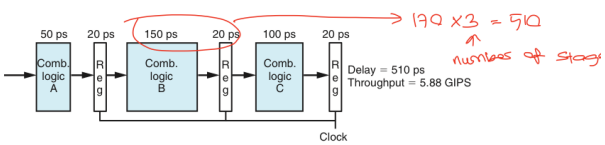
↳ giga-instructions per sec..

Latency: The total time required to perform a single instruction from beginning to end.

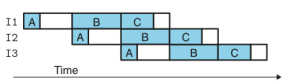
Latency = 3 x 120 = 360 ps ← processing a single instruction requires 3 clock-cycle.

$$\text{Throughput} = \frac{3 \text{ instructions}}{360 \text{ ps}} \times \frac{1{,}000 \text{ ps}}{1 \text{ ns}} \approx 8.33 \text{ GIPS}$$

Non-uniform partitioning



(a) Hardware: Three-stage pipeline, nonuniform stage delays



(b) Pipeline diagram

Sum of the delays through all of the stages remains 300 ps. BUT, the rate at which we can operate the clock is limited by the delay of the slowest stage.

clock cycle : 150 + 20 = 170 ps

$$\text{throughput} = \frac{1}{170} \times \frac{1000 \text{ ps}}{1 \text{ ns}} = 5.88 \text{ GIPS}$$

$$\text{latency} = 3 \times 170 = 510 \text{ ps}$$