1.

a.

$$I_A = 10^5 Inst$$
$$I_B = 2 \times 10^5 Inst$$
$$I_C = 5 \times 10^5 Inst$$
$$I_D = 2 \times 10^5 Inst$$

$$ET_1 = \frac{\left(CPI_{A_1} \times I_A + CPI_{B_1} \times I_B + CPI_{C_1} \times I_C + CPI_{D_1} \times I_D\right)}{F_{CLK1}} = \frac{10^5 + 4 \times 10^5 + 10^6 + 2 \times 10^5}{2 \times 10^9}$$
$$= 0.58ms$$

$$ET_2 = \frac{\left(CPI_{A_2} \times I_A + CPI_{B_2} \times I_B + CPI_{C_2} \times I_C + CPI_{D_2} \times I_D\right)}{F_{CLK2}}$$
$$= \frac{2 \times 10^5 + 6 \times 10^5 + 2 \times 10^6 + 8 \times 10^5}{4 \times 10^9} = 0.9ms$$
$$ET_1 < ET_2$$

So P_1 is faster.

b.

$$avg\ CPI_1 = 10\% CPI_{A_1} + 20\% CPI_{B_1} + 50\% CPI_{C_1} + 20\% CPI_{D_1} = 0.1 + 0.4 + 1 + 0.2 = 1.7$$
$$avg\ CPI_2 = 10\% CPI_{A_2} + 20\% CPI_{B_2} + 50 CPI_{C_2} + 20\% CPI_{D_2} = 0.2 + 0.6 + 2 + 0.8 = 3.6$$

c.

$$Clock\ cycles_1 = avgCPI_1 \times IC = 1.7 \times 10^6$$
$$Clock\ cycles_2 = avg \times IC = 3.6 \times 10^6$$

d.

$$IPS_1 = \frac{F_{CLK1}}{avgCPI_1} = \frac{2GHz}{1.7} = 1.18 \times 10^9 Inst/Sec$$
$$IPS_2 = \frac{F_{CLK2}}{avgCPI_2} = \frac{4GHz}{3.6} = 1.11 \times 10^9 Inst/Sec$$
$$IPS_1 > IPS_2$$

Thus, P_1 has the highest throughout performance.

e.

P_1 is more energy efficient because it has lower CPI and P_2 needs more clock cycles in each instruction so it consumes more power to the same work.

2.

a.

50% of the total energy.

b.

$$\frac{E_{new}}{E_{old}} = \frac{\frac{1}{2} \times C \times V_{new}^2}{\frac{1}{2} \times C \times V_{old}^2} = \frac{1}{4}$$

So, 3/4 of the total energy can be saved.

3.

a.

$$P_{new_1} = 90\% \times (1-60\%) \times P_{max} = 36\% P_{max}$$
$$P_{old} = 90\% P_{max}$$
$$\frac{P_{old} - P_{new_1}}{P_{old}} = \frac{90\% - 36\%}{90\%} = 60\%$$

b.

$$P_{new_2} = 90\% \times P_{max} \times (1-60\%) + 20\% \times P_{max} \times 60\% = 48\% P_{max}$$
$$\frac{P_{old} - P_{new_2}}{P_{old}} = \frac{90\% - 48\%}{90\%} = 46.7\%$$

c.

$$1 - \frac{P_{new_3}}{P_{old}} = 1 - \frac{C_{old} \times (0.8V_{old})^2 \times F_{old} \times 0.6}{C_{old} \times V_{old}^2 \times F_{old}} = 1 - 38.4\% = 61.6\%$$

d.

$$1 - \frac{P_{new_4}}{P_{old}} = 1 - \frac{90\% \times P_{max} \times (1-60\%) + 20\% \times P_{max} \times 30\%}{90\%} = 1 - \frac{42\%}{90\%} = 53.3\%$$

4.

a.

MTTF for the system=35/3=11.7=12days

b.

If the MTTF is doubled, the running time for computers will increase, so we can save cost and time.

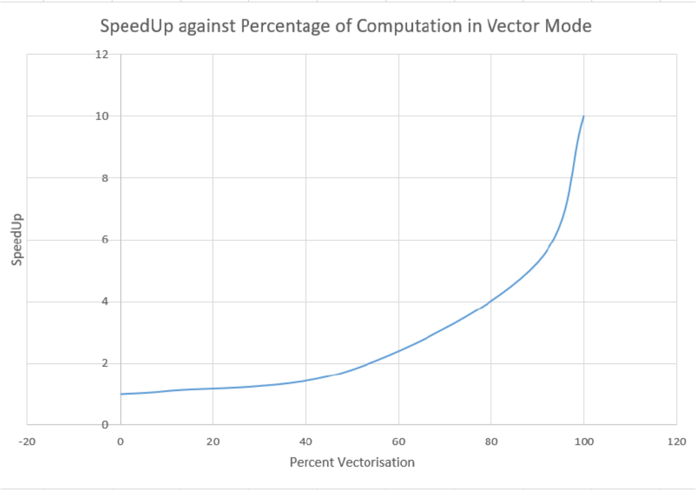
5.

a.

x=0, y=1

x=1, y=10

y=10/(10-9x)



b.

$$\frac{10}{10-9x}=2$$

Then,

$$x=\frac{10}{18}=55.56\%$$

c.

Assume the old execution time = 1.

$$Percentage=\frac{vector\ mode\ time}{new\ execution\ time}=\frac{55.56\%\times 0.1}{\frac{1}{2}}=11.112\%$$

d.

$$y=\frac{10}{2}=5=\frac{10}{10-9x}$$

$$x=\frac{40}{45}=88.89\%$$

e.

$$x=70\%$$

$$y=\frac{10}{10-9x}=2.70$$

$$y_{goal}=2.70\times 2=5.40=\frac{10}{10-9x}$$

$$x_{goal}=\frac{44}{48.6}=90.53\%$$

6.

a.

$$ET=\frac{I\times(CPI_1+CPI_2+CPI_3)}{F_{CLK}}=\frac{150\times 10^6\times(0.5\times 3+0.3\times 4+0.2\times 5)}{2.7\times 10^9}=205.6ms$$

b.

$$CPI_{new}=0.5\times 3+0.3\times 2+0.2\times 2=2.5$$

$$ET_{new}=\frac{I\times CPI_{new}}{F_{CLKnew}}=\frac{150\times 10^6\times 2.5}{2.1\times 10^9}=178.6ms$$

$$Improvement=\frac{ET}{ET_{new}}-1=\frac{205.6}{178.6}=1.15-1=0.15$$

7.

a.

$$\begin{aligned}ET_{old}&=(CPI_1\times I_1+CPI_2\times I_2+CPI_3\times I_3)\times Clock\ Cycle\ Time\\&=(1\times 500+10\times 300+3\times 100)\times 10^6\times Clock\ Cycle\ Time\\&=3.8\times 10^9\times Clock\ Cycle\ Time\end{aligned}$$

$$\begin{aligned}ET_{new}&=(1\times 500\times 75\%+10\times 300+3\times 100)\times 10^6\times Clock\ Cycle\ Time\times 110\%\\&=4.0426\times 10^9\times Clock\ Cycle\ Time\end{aligned}$$

Since $ET_{old}<ET_{new}$, it's not a good design choice.

b.

If the performance is doubled, $CPI_1 = 0.5$.

$$ET_{new} = (0.5 \times 500 + 10 \times 300 + 3 \times 100) \times 10^6 \times Clock\ Cycle\ Time$$
$$= 3.35 \times 10^9 \times Clock\ Cycle\ Time$$
$$\frac{ET_{new}}{ET_{old}} = \frac{3.55}{3.8} = 0.9342$$
$$Speedup = \frac{1}{0.9342} = 1.07$$

If the performance is improved by 10 times, then $CPI_1 = 0.1$

$$ET_{new} = (0.1 \times 500 + 10 \times 300 + 3 \times 100) \times 10^6 \times Clock\ Cycle\ Time$$
$$= 3.35 \times 10^9 \times Clock\ Cycle\ Time$$
$$\frac{ET_{new}}{ET_{old}} = \frac{3.35}{3.8} = 0.8815$$
$$Speedup = \frac{1}{0.8815} = 1.13$$

8.

$$ET_{old} = \frac{CPI \times I}{F_{CLK}}$$
$$ET_{new} = \frac{CPI \times \left(1 - 30\% \times \left(1 - \frac{2}{3}\right)\right) I}{0.95 F_{CLK}}$$
$$\frac{ET_{old}}{ET_{new}} = \frac{0.95}{0.9} = 1.056$$

Since $ET_{old} > ET_{new}$, the optimized version is faster.

9.

a.

$$A_{operations} = 0.7 \times throughput_A = 9422.7$$
$$B_{operations} = 0.3 \times throughput_B = 10939.5$$
$$t_{TPU} = \frac{A_{operations}}{throughput_A} + \frac{B_{operations}}{throughput_B} = 0.0419 + 0.0391 = 0.081$$
$$Speedup = \frac{t_{GPU}}{t_{TPU}} = 12.35$$

b.

$$General: 70\% \times 42\% + 30\% \times 100\% = 59.4\%$$
$$GPU: 70\% \times 37\% + 30\% \times 100\% = 55.9\%$$
$$TPU: 70\% \times 80\% + 30\% \times 100\% = 86\%$$

c.

$$P_{GPU} = \frac{55.9\%}{991W - 357W}$$
$$P_{TPU} = \frac{86\%}{384W - 290W}$$

$$\frac{P_{TPU}}{P_{GPU}} = 10.3764$$

d.

General:

$$A = 40\% \times 5482 = 2192.8$$
$$B = 10\% \times 13194 = 1319.4$$
$$C = 50\% \times 12000 = 6000$$

$$t_{GPU} = \frac{2192.8}{13461} + \frac{1319.4}{36465} + \frac{6000}{15000} = 0.1629 + 0.0362 + 0.4 = 0.5991$$

$$Speedup_{GPU} = \frac{1}{0.5991} = 1.6692$$

$$t_{TPU} = \frac{2192.8}{225000} + \frac{1319.4}{280000} + \frac{6000}{2000} = 0.0097 + 0.0047 + 3 = 3.0144$$

$$Speedup_{TPU} = \frac{1}{3.0144} = 0.3317$$

e.

$$Haswell: \frac{14kW}{504W} = 27.78 = 27$$

$$NVIDIA: \frac{14kW}{1838W} = 7.62 = 7$$

$$TPU: \frac{14kW}{861W} = 16.26 = 16$$

f.

$$Haswell: \frac{200W \times 11}{504W} = 4 < 27$$

$$NVIDIA: \frac{200W \times 11}{1838W} = 1 < 7$$

$$TPU: \frac{200W \times 11}{861W} 2 < 16$$

Thus, only 1 cooling door is required for each.

10.

a.

The projected growth rate in 2015 was 3.5% performance bump per year. Therefore, by this estimate the processors in 2025 should be 1.03510 times better than the processors of 2015.

b.

The SUN 4/290 processor, which was released in 1987, was 10 times faster than the VAX-11/780. Since the projected growth rate at the time was 52% increase per year, the performance of a 2025 processor, by 1977 estimates, would be (10 ·1.5238) times faster than the VAX-11/780

c.

Moore's Law was built upon Denard Scaling, which proposed that as transistors decreased in size, their power density remained consistent. This meant that power consumption remained proportional to chip area, resulting in reductions in both voltage and current. However, as transistors continued to shrink, it became apparent that lowering voltage and current indefinitely would jeopardize the integrity of integrated circuits. Factors such as leakage current and threshold voltage became increasingly influential when determining power consumption levels.

11.

a.

Failures in time (FIT) is traditionally reported as failure per billion (1 × 10^9) hours of operation.

$MTTF \; = \frac{1}{\left(\frac{100}{10^9}\right)} = \frac{10^9}{100} = \; 10^7 hours$

b.

MTTR = 1 day = 24 hours

$$Availability\ of\ the\ system = \frac{MTTF}{(MTTF \; + \; MTTR)} = \frac{10^7}{10^7} + 24 \; = \; 0.999 \sim 1 \; = \; 99.99\%$$

c.

$$FIT_{1000} = Number\ of\ processors \; \times \; FIT\ per\ processor = 1000 \; \times \; 100 \; = \; 10^5$$

$$MTTF_{1000} = \frac{1}{\frac{10^5}{10^9}} = \frac{10^9}{10^5} = 10^4 \; hours$$