EECS 151/251A Problem Set 1

Due Monday, Jan 29th, 2024

Problem 1: Moore's Law [10 pts]

Consider state-of-the-art processor chips from the 1970's, 1980's, 1990's, 2000's, and after 2010. Choose a processor from each period. (You may choose which every processor you like, but make sure they are spaced out by around 10 years. Use of Wikipedia or WikiChip is acceptable.)

1. For each look up the approximate number of transistor per chip. Plot the number of transistors per chip over time, with a log scale on the y-axis [5 pts].
2. On another set of axis, plot the clock frequency of each over time (on a log plot) [5 pts].

Solution:

1. The graphs will vary by which processors were used. Some example processors are listed below.

Some example Processors

Chip	Year	Transistors	Clock Freq. Low (MHz)	Clock Freq. High (MHz)	Key Advance
Intel 8080	1974	6K	2	3.125	One of the first widespread microprocessors owing in part to its 40-pin package, faster transistors (NMOS), compatibility with TTL, and and increased memory bus size (with ability to access 64 KB of memory)
Intel 80386	1985	275K	12	40	First Implementation of 32-bit extension to 80286. A common ancestor of modern x86 processors and is also known by the name i386.
Intel Pentium P55C	1997	4.5M	120	233	First Implementation of MMX (SIMD) extension in Intel Processors. Predecessor to SSE and AVX vector extensions
Intel Prescott	2004	125M	2400	3067	First of the Pentium line to introduce SSE3. Major overhaul that hit the heat barrier
Intel Ivy Bridge-E	2013	1.86B	3400	3600	Die shrinking of Sandy Bridge-E to 22 nm (finfet)

Problem 2: Pareto Optimality [10 pts]

Frequency (f)	Chip Capacitance (C)	Cost
2 GHz	1 nF	$10
2 GHz	3 nF	$25
4 GHz	10 nF	$25
4 GHz	5 nF	$25
4 GHz	6 nF	$50
1 GHz	15 nF	$50
5 GHz	3 nF	$100
5 GHz	2 nF	$135

After performing a design exploration, John has found the follow options for processors for his robotics project. He wants to maximize clock frequency while minimizing cost. His robot is only able to efficiently cool 30W. Assume that each processor has $V_{dd} = 1.5V$. Pick which points are on the frequency-cost Pareto Optimal frontier and meet his power constraints. (You can use the following equation for chip power: $P = \frac{1}{2} \cdot C \cdot V_{dd}^2 \cdot f$)

Solution:

Frequency	Chip Capacitance (nF)	Cost	Power
2 GHz	1 nF	\$10	2.25 W
2 GHz	3 nF	\$25	6.75 W
4 GHz	10 nF	\$25	45 W
4 GHz	5 nF	\$25	22.5 W
4 GHz	6 nF	\$50	27 W
1 GHz	15 nF	\$50	16.9 W
5 GHz	3 nF	\$100	16.9 W
5 GHz	2 nF	\$135	11.25 W

We see that we can remove the 3rd row as it violates the power constraints. Now, we can select the Pareto-optimal frontier for Cost and Frequency.

Frequency (GHz)	Chip Capacitance (nF)	Cost
2 GHz	1 nF	\$10
4 GHz	5 nF	\$25
5 GHz	3 nF	\$100

We can remove the 2nd row since it has the same frequency but higher cost. Similarly, we can remove the 5th and 8th row. On the other hand, the 7th row is more expensive but offers a higher frequency so it is on the frontier. Note that even though rows may have different usage powers, these values do not factor into the frontier since our power is given as a constraint and not something that is needed to be optimized.

Problem 3 - Dennard Scaling [21 pts]

Let us assume that we live in a world with perfect Dennard Scaling. I currently have access to a chip built with transistors with a width of 5nm. The chip has operational frequency of 2.5 GHz, V_{dd} of 1.5V, and a performance of 20TFLOPs/sec (FLOPs is Floating Point Operations). The total chip consumes 20W and has 10 billion transistors.

1. If I switch to the exact same chip but upgrade to a 4nm width transistor, what will the following values be in order to continue Dennard Scaling [15 pts]:
   (a) Voltage
   (b) Frequency
   (c) Power per transistor
   (d) Total Power of the Chip
   (e) Performance (FLOPs/sec) (assume that the FLOPs per cycle scales proportionally to the the number of transistors)
2. How small should the transistor be in order for me to run my 50 TFLOP workload in less than one second? [6 pts]

Solution:

1. The scaling factor (κ) is 1.25 (5/4)
   (a) Voltage scales by $\frac{1}{\kappa}$. $\frac{1}{1.25} \cdot 1.5 = 1.2$
   (b) Frequency scales by κ. $2.5 \cdot 1.25 = 3.125$ GHz
   (c) Power per transistor scales by $\frac{1}{\kappa^2}$. Current Power per transistor = $\frac{20}{10e9} = 2$nW. New Power per transistor is $2nW \cdot \frac{1}{1.25^2} = 1.28$nW
   (d) Total Power of the Chip stays the same, 20W
   (e) Performance scales by κ^3 as the number of transistors scales by κ^2 and the frequency scales by κ. Therefore the new performance is $20 \cdot 1.25^3 = 39.06$ T FLOPs/sec
2. We can first solve for κ. Our target performance if 50 TFLOPs/sec, and we know performance scales by κ^3. So, $\kappa^3 = 50/20 = 1.36$. We know the width of the transistor scales by $\frac{1}{\kappa} = 5 \cdot \frac{1}{1.36} = 3.67$nm.

In EdStem, we also state that we accept that the new chips has a smaller area, but keep the number of transistors constant. Therefore, we will also accept the following:

1. The scaling factor (κ) is 1.25 (5/4)
   (a) Voltage scales by $\frac{1}{\kappa}$. $\frac{1}{1.25} \cdot 1.5 = 1.2$
   (b) Frequency scales by κ. $2.5 \cdot 1.25 = 3.125$ GHz
   (c) Power per transistor scales by $\frac{1}{\kappa^2}$. Current Power per transistor = $\frac{20}{10e9} = 2$nW. New Power per transistor is $2nW \cdot \frac{1}{1.25^2} = 1.28$nW
   (d) Total Power of the Chip decreases by $\frac{1}{\kappa^2}$ as the area scales by $\frac{1}{\kappa^2}$ and power density is constant. Therefore, the new Power of the chip is $20W \cdot \frac{1}{1.25^2} = 12.8W$
   (e) Performance scales by κ as the number of transistors is constant and the frequency scales by κ. Therefore the new performance is $20 \cdot 1.25 = 25$ T FLOPs/sec
2. We can first solve for κ. Our target performance if 50 TFLOPs/sec, and we know performance scales by κ. So, $\kappa = 50/20 = 2.5$. We know the width of the transistor scales by $\frac{1}{\kappa} = 5 \cdot \frac{1}{2.5} = 2$nm.

Problem 4 - Technology Survey [5 pts]

Take a look at your laptop/desktop computer. What processor is it using? Which technology node was it designed in (14 nm, 12 nm, 7 nm, etc.)? How many cores does it have? Does it have dedicated hardware accelerator blocks?

Solution:

Solutions vary by device.