EECS151/251A Discussion 8

Christopher Yarp

Mar. 22, 2019

Plan for Today

- Practice Problem
- Your Questions

Quick Reminder about Dynamic Power

- $P_{SW} = \frac{1}{2}\ \alpha\ C\ V_{DD}^2\ F$
- P_{SW}: Switching Power
- α: Activity Factor (Avg. Percentage of Nodes to Switch)
- C: Total Chip Capacitance
- V_{DD}: VDD of Chip
- F: Clock Frequency

- From the Chip-Level “Dynamic” Power Slide of Lecture 12

Problem 1: Overclocking

- Let us take a hypothetical CPU which nominally operates with a 3 GHz clock at a VDD of 0.9 V.
- Let us assume that the VDD required to support a particular clock frequency can be approximated as the linear function VDD = 0.0002f + 0.3 where f is the clock frequency in MHz and VDD is in V.

a) Approximately what VDD is required to run the CPU at 4 GHz?

b) What is the ratio of active power dissipation between an overclocked CPU running at 4 GHz compared to the same CPU running at 3 GHz?

Problem 1(a): Finding New VDD

- $V_{DD} = 0.0002f + 0.3$
- $V_{DD} = 0.0002(4000) + 0.3$
- $V_{DD} = 1.1V$

Problem 1(b): Power Ratio

- 3 GHz Operating Point
  - $P_{SW3} = \frac{1}{2}\ \alpha\ C\ V_{DD}^2\ F$
  - $P_{SW3} = \frac{1}{2}\ \alpha C (0.9V)^2 (3\ GHz)$
- 4 GHz Operating Point
  - $P_{SW4} = \frac{1}{2}\ \alpha\ C\ V_{DD}^2\ F$
  - $P_{SW4} = \frac{1}{2}\ \alpha C (1.1V)^2 (4\ GHz)$
- Capacitance is the same in both cases (chip did not change).
- Activity factor is assumed to be the same (running the same application)

- $\frac{P_{SW4}}{P_{SW3}} = (1.1V)^2(4\ GHz)/(1.1V)^2(4\ GHz)$
- $\frac{P_{SW4}}{P_{SW3}} = 1.998$

- Almost 2x the amount of power is required to run the system at a 1.33x faster clock rate!
- That's 2x the amount of heat that needs to be dissipated
- One of the reasons why overclocking requires a robust power supply and a good cooling solution
- One way to address this is to shut down cores when overclocking.

Problem 2: Parallelism for Lower Power

- A ASIC implementation of a video encoder (Version A) is being modified to be more power efficient (Version B). The new version of the encoder works on segments of a video in parallel. However, there is a price for this parallelism:
  - If τ_{orig} is the critical path delay in version A, the critical path in version B is increased by $0.1\ N\ \tau_{orig}$ where N is the number of parallel video encoders
  - If A_{orig} is the area of version A, in addition to the N replications of the encoder, $0.2\ \mathrm{N}\ A_{orig}$ in area overhead is also added to version B
- Version B includes N=4 transcoders

Problem 2: Parallelism for Lower Power

a) What is the new critical path length of version B?
b) If version B was run at the fastest rate possible and used all 4 of its encoders, how long would it take to transcode a video compared to version A? (*Assume the required cycles to encode a video can be split evenly across the decoders*)
c) Using all 4 transcoders, what clock rate is required to transcode a video in the same amount of time as version A?
d) What is the chip size of version B compared to version A?
e) How much power is consumed in version B as compared to version A when transcoding a video in the same amount of time (max rate of version A)?

Problem 2(a): New Critical Path Delay

- $\tau_{CPB} = \tau_{orig} + 0.1(4)\tau_{orig}$
- $\tau_{CPB} = 1.4\tau_{orig}$
- $\tau_{CPB} \approx 1.4(333.3\ \text{ps}) = 466.7$
- Fastest clk approx. 2.143 GHz

Problem 2(b): Transcode a Video at Full Speed

- Let L be the number of cycles required to transcode a given video with version A
  - $\tau_{videoA} = \tau_{CPA}\, L = \tau_{orig}\, L$
- Version B has 4 transcoders, each of which are capable of splitting the encoding work evenly between them
  - Number of required clock cycles to encode video in version B is L/4
- $\tau_{videoB} = \tau_{CPB}\ L/4 = 1.4\tau_{orig}\, L/4$
- $\tau_{videoB}/\tau_{videoA} = (1.4\tau_{orig}L/4)/(\tau_{orig}\, L\,) = 1.4/4$
- $\tau_{videoB}/\tau_{videoA} = 0.35$
- Not quite ¼ the time due to the overhead in version B

Problem 2(c): Slowing Clock Rate

- Want $\tau_{videoB} = \tau_{videoA}$
- Because of the parallel units, if version A requires L cycles to encode a video, version B requires L/4 cycles
- $\tau_{videoB} = \tau_{videoA}$
- $\tau_{clkB}\ L/4 = \tau_{orig}\ L$
- $\tau_{clkB} = 4\,\tau_{orig}$
- $\tau_{clkB} = 4\,(333.3\,ps) = 1.333\,ns$
- Required clock rate is 750 MHz
- This rate is achievable with version B as it is far below the maximum clock rate
- However, note that this is ¼ the max frequency of version A but is less than ¼ the max frequency of version B due to delay overhead

Problem 2(d): Chip Area

- There are 2 components to area growth in version B
  - Replicas of the video transcoder
  - Overhead
- $\mathrm{A_B} = \mathrm{N\ A_{orig}} + 0.2\ \mathrm{N}\ A_{orig}$
- $\mathrm{A_B} = 1.2\ \mathrm{N\ A_{orig}}$
- $\mathrm{A_B} = 1.2\ (4)\ \mathrm{A_{orig}}$
- $\mathrm{A_B} = 4.8\ \mathrm{A_{orig}}$

Problem 2(e): Dynamic Power Consumption

- Let us assume that the capacitance is proportional to the area of the chip.
  - $C_B = 4.8\ C_{orig}$
- The additional copies of the encoder as well as the overhead increase C for version B as compared to version A
- Activity factor is assumed to be the same (running the same application)
- Assume both chips were produced with the same frequency vs. VDD relationship
  - Version A operates at 0.9V at 3 GHz
  - Version B operates at 0.9V at 3 GHz
    - Find new Y intercept
    - $V_{DD} = 0.0002f + offset$
    - $0.9 = 0.0002(2.143) + offset$
    - $offset = 0.4714$

- $V_{DDB} = 0.0002(750) + 0.4714 = 0.6214V$
- Version A
  - $P_{SWA} = \frac{1}{2}\ \alpha\ C_{orig}\ V_{DD\_A}^2\ F$
  - $P_{SWA} = \frac{1}{2}\ \alpha\ C_{orig}(0.9V)^2(3\ GHz)$
- Version B
  - $P_{SWB} = \frac{1}{2}\ \alpha\ C_B\ V_{DD\_B}^2\ F$
  - $P_{SWB} = \frac{1}{2}\ \alpha\ 4.8\ C_{orig}(0.6214V)^2(0.75\ GHz)$
- $\frac{P_{SWB}}{P_{SWA}} = 4.8(0.6214V)^2(0.75\ GHz)/(0.9V)^2(3\ GHz)$
- $\frac{P_{SWB}}{P_{SWA}} = 0.5721$
- Note, if there was no area overhead, the change in C and the change in F would have canceled.