

EECS 151/251A Homework 7

Due Monday, April 5th, 2021

For this Assignment

Please include a short (1-2 sentence) explanation of your answer with each response unless otherwise directed to by the problem.

Problem 1: Technology Survey (No short explanation needed)

Take a look at your phone. What is its battery capacity in kWh and in J. How often do you charge your phone, and from what battery level do you start charging from each day? Estimate your average power consumption from those values. Repeat this exercise for your laptop.

Problem 2: Parallelization & Pipelining

One of your co-workers has designed an accumulator that takes a running sum of 8 integers. The integers arrive as an 8-wide integer array and the accumulator cycles through each element of the array and returns the final sum. They aggressively designed their accumulator to meet a throughput target of 64×10^7 sums/sec by pushing their accumulator to be as fast as possible. However, in doing so, their design has a dynamic power consumption of 75 W, which is unacceptably high for a simple accumulator. At your next weekly meeting, they sketch their design on the board and bring up their dilemma and you immediately recall learning in EECS151 about two ways to achieve the same throughput while using less power.

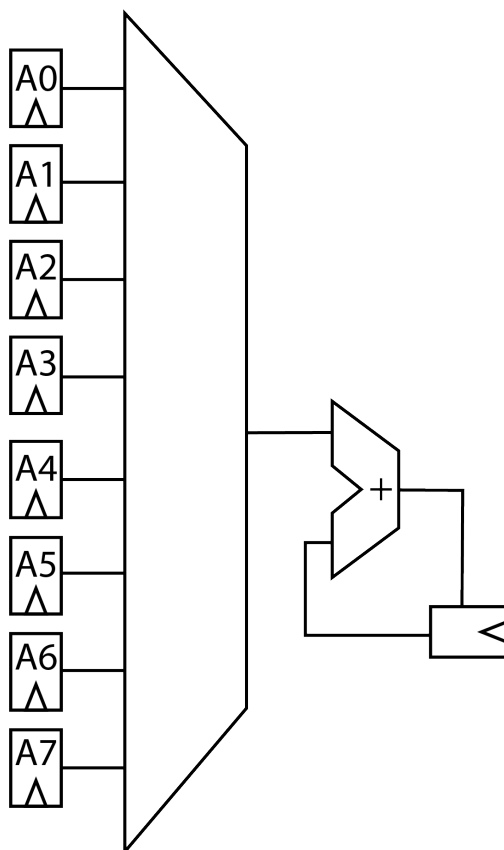


Figure 1: Sketch of original design

- (a) The first idea that comes to your mind is just to parallelize the design. How would you modify the design to parallelize the hardware as much as possible while maintaining the same throughput? How and why does this improve the dynamic power consumption? (*Hint: Think of a binary tree*)
- (b) Now as you consider your parallelized approach, it dawns on you that you can also leverage the other technique you learned and pipeline the resulting design. How would you pipeline the parallelized design that you came up with in the last part? How and why does this further improve the dynamic power consumption?

Problem 3: Race to Halt

One scheme for potentially improving energy efficiency if static power is a significant proportion of the total power consumption is a technique known as "race to halt". Basically, we run the circuits at maximum speed to finish the computation as quickly as possible, then cut off the power so that we don't suffer the static power loss.

Suppose we have a CPU that takes 10 seconds to run a particular application, consuming 12 W, where some proportion δ of the total power is consumed by dynamic power, with the remaining $\sigma = 1 - \delta$ the proportion lost to static power consumption. Assuming there are no other applications

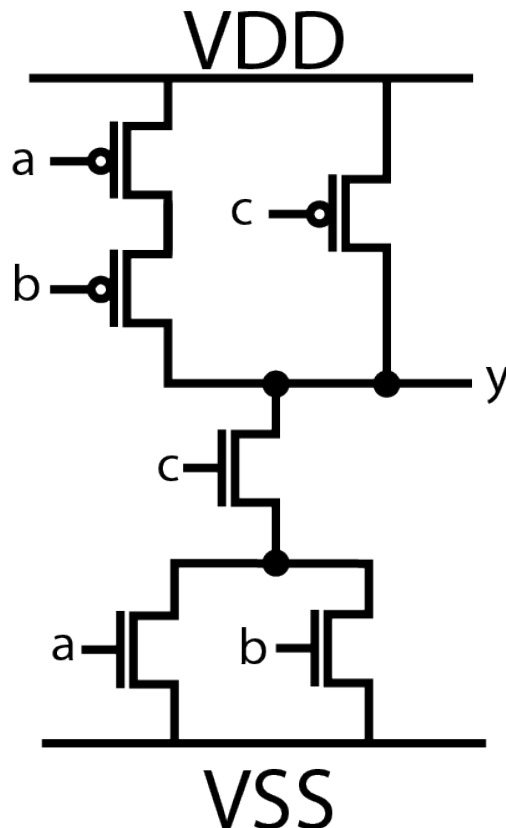
running on the CPU and that these two proportions cover the entire power budget of the CPU, we would like to find a scheme that would minimize power consumption.

As an alternative to the race to halt method, we could also consider a more traditional frequency/voltage scaling method for reducing power consumption. The technology we are working with can tolerate a V_{DD} reduction by at most 25%, and we can assume that the voltage to delay scaling is linear (e.g. a 2x reduction in supply voltage will need a 2x decrease in clock frequency).

Explore race to halt versus frequency/voltage scaling. Assume that the voltage and frequency scaling will only affect dynamic power consumption, and does not affect the amount of static power consumption (which is not an unrealistic approximation). For what δ would race to halt be better than frequency/voltage scaling?

Problem 4: Logic Gate Dynamic Power

Consider the OR-AND-INV (OAI) gate shown below, operating at $V_{DD} = 1\text{V}$, with a load of $C_L = 20\text{fF}$, and $C_D = 0.5\text{fF/nm}$.



- (a) Size the devices in the OAI gate such that the input capacitance is equal to a unit-sized inverter, with NMOS width $W_N = 12\text{nm}$. Make sure the rise and fall time of the gate is balanced for the worst case.

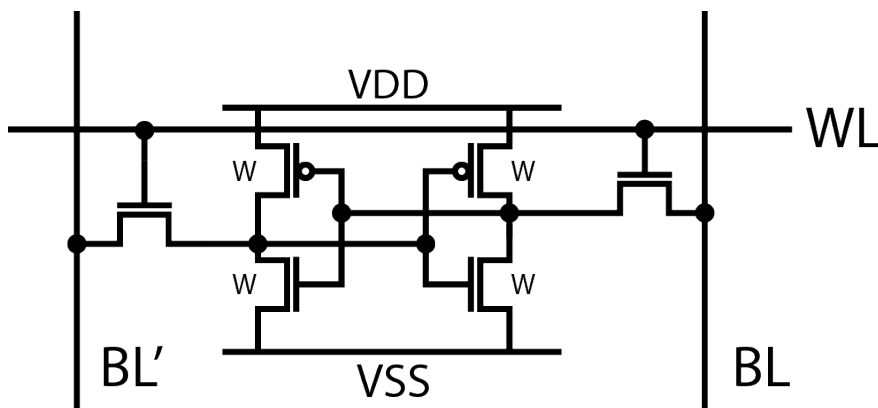
- (b) What is the dynamic power dissipation of the gate, assuming the design runs at a clock frequency of 4 GHz and the gate output has an activity factor of $\alpha = 0.4$?

Problem 5: Short Circuit Power

Set up a simulation in LTSpice to demonstrate short-circuit power. Build an inverter with $W_P = 2W_N = 0.5 \text{ nm}$ and $l_N = l_P = 16 \text{ nm}$ at a $V_{DD} = 0.9 \text{ V}$. Set up the schematic to have input transitions of 10 ps, 25 ps, and 50 ps. To measure the short-circuit current, consider this: In an ideal inverter with no short-circuit current, for a high-to-low transition the only device that should be conducting is the pull-down device. Any additional current through the pull-up device is unwanted current, and thus contributes to the short-circuit power that we would like to avoid. Therefore to measure the short circuit current, we just need to see how much current is conducting through the undesired device. Please turn in a screenshot of your schematic as well as a waveform of the three short-circuit currents.

Problem 6: 6T SRAM Cell

Consider the 6T SRAM Cell below, with the inverter device sizes labeled. In this technology, $R_P = R_N$. The cell operates at a supply voltage of 0.9 V



- (a) How would you size the access transistors to ensure a successful write? Assume for this technology that the voltage needs to be more than 20% V_{DD} away from the switching voltage ($V_{DD}/2$) for a successful write. You may assume the column drivers perfectly drive the bit lines, and the only significant resistances are from the SRAM cell and the access transistors.
- (b) A write-only SRAM cell is not very useful, so we will also need to read from this SRAM cell. However, you just sized the devices so that the SRAM cell is easy to write to, which means the typical read procedure has a high risk of corrupting the memory values stored in the SRAM cell. How would you resize the devices to reduce the chance of read corruption while still maintaining a successful write? You may now resize the devices in the SRAM cell itself if you wish to do so. Again, you may assume the only significant resistances are that of the SRAM cell and the access transistors. You can also assume that the bitlines have extremely high (effectively infinite) capacitance so the voltage on them will not change significantly during the read operation, but the sense amplifier will still be sensitive enough to register a read.

(c) **251A only** — *Optional Challenge Question for 151*

If you could redesign the SRAM cell, is there a way you could break the read vs. write sizing tradeoff? You may add or remove devices from the classic 6T structure to achieve this new design.

Problem 7: Building Bigger Blocks

As modern processors are able to handle more and more computations per second, the associated memories must also be able to hold more and more data to keep up. However, one major problem with making huge SRAM blocks is that the bit lines will become extremely long as the memory increases in capacity, which causes issues with readability and write speed. One way to get around this problem is to make the overall memory block out of smaller sub-blocks.

For this problem, you have access to an SRAM block that is 32 words deep with a 32-bit word length. The basic SRAM block has a single read port and a single write port.

- (a) Describe how you would make a 128KB, 32-bit wide dual-port memory (single read, single write) using these sub-cells. You will need to describe your periphery circuits. How would you assign the address bits for this design (row/column arrangement)?
- (b) Describe how you would implement a dual-read SRAM (1 write port, 2 read ports).

Problem 8: Address Decoding

Consider a 16-bit wide 2KB SRAM. How many rows and columns are in your design? Design an address decoder using the predecoder technique from the lecture. You may use only logic gates of no more than 4 inputs.