ECE 3030: Physical Foundations of Computer Engineering

Spring 2024

Final Exam

May 1, 2024

Time: 2 hour 50 min

Instructor: Asif Khan

Instructions:

1. There are 15 pages in this test. Count the number of pages and notify the proctor if you are missing a page.
2. Read all the problems carefully and thoroughly before you begin working.
3. A list of constants and equations is provided on pages 14, 15.
4. You are required to answer all 5 questions. There are 100 total points. Observe the point value of each problem and allocate your time accordingly.

Q1	10 pts
Q2	25 pts
Q3	25 pts
Q4	20 pts
Q5	20 pts
Q6 (HW11)	25 pts
Q7 (HW12)	25 pts
Total	150 pts

5. Show all your work and circle/underline your final answer. For numerical answers, write the units. Write legibly. If I cannot read it, it will be considered a wrong answer. Do all work on the space provided; use scratch paper when necessary. Turn in all scratch paper, even if it did not lead to an answer.
6. Report any and all ethics violations to the instructor/proctor.

Sign your name on ONE of the two following cases:

I DID NOT observe any ethical violations during this exam:

I observed an ethical violation during this exam:

[Q1.1] **Physics of resistors:** The following table lists different properties of different interconnect metals. Rank them in terms of your preference for their use of as interconnect metals in a chip. Provide a brief explanation of your answer. [10 pts]

Material	Electron density (m^{-3})	Electron mobility (m^2/Vs)
Al	1.98×10^{29}	1.2×10^{-3}
Cu	8.5×10^{28}	4.32×10^{-3}
W	6×10^{28}	1.8×10^{-3}
AB	5×10^{28}	3×10^{-3}

Q2 **MOSFETs and Delay and Power in Inverter:** If decrease the doping density N_A in a MOSFET with all the parameters unchanged, how will the following quantities change? [Total 25 pts]

[Q2.1] The MOSFET threshold voltage, V_t. [5 pts]

[Q2.2] The on-state current, I_{ON} of the MOSFET. [5 pts]

[Q2.3] The off-state leakage current, I_{OFF} of the MOSFET. [5 pts]

[Q2.4] The corresponding inverter delay. [5 pts]

[Q2.5] The active power in the corresponding inverter. The clock frequency did not change. [5 pts]

Q3 **Scaling:** Consider that all three physical dimensions of MOSFETs (W, L, t_{ox}) are downscaled by factor of x and the power supply voltage, V_{DD} and the threshold voltage, V_t are decreased by a factor of y in every subsequent generation. In addition, the total area of the chip, A_c, also increases by 5% in every subsequent generation. [Total 25 pts]

[Q3.1] If, in every subsequent generation, the total number of transistors doubles, what is the nominal value of x? [5 pts]

[Q3.2] Consider the clock frequency and the total active/dynamic chip power of the n-th generation are f_n and $P_{chip,n}$. Find an expression of the ratio of total chip powers of two subsequent generation, $P_{chip,n+1}/P_{chip,n}$ in terms of x, y, f_n and f_{n+1}. [5 pts]

[Space for Q3.2]

[Q3.3] What is the relation between f_{n+1} and f_n if you wanted the total active/dynamic chip power to remain the same across generations. [5 pts]

[Q3.4] In recent years, why is it not possible to reduce the threshold voltage of the MOSFETs significantly in subsequent generations? Use necessary figures. [Total 5 pts]

[Q3.5] In the face of the inability to downscale the threshold voltage significantly in successive generation, what should be the value of threshold voltage scaling factor y if you wanted to physical scaling to continue at the same rate x as you calculated in Q3.1 while keeping the total active/dynamic chip power and the clock frequency constant in each subsequent generation. [5 pts]

Q4 **Memory technologies.** [Total 20 pts]

[Q4.1] **SRAM Array:** Consider the SRAM array shown in figure 1. You want to read all the cells in row 2. What is the sequence of operation you will need to perform? Make sure that, after your prescribed operations, you keep the data in the cells you read intact. [Total 10 pts]

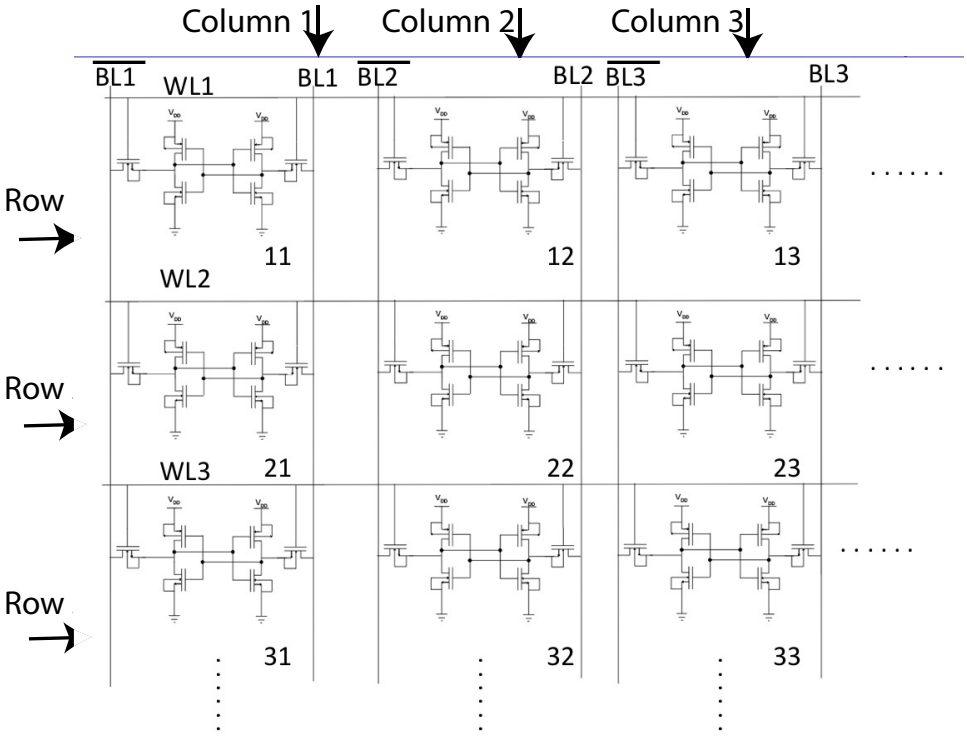


Figure 1: An SRAM array.

[Q4.2] **Magnetic Hard Drives:** Briefly explain why the magnetic hard drive technology is not classified as a random access memory technology by clearly stating what random memory access means. [5 pts]

[Q4.3] Say you have deployed a bunch of IoT sensors to measure the distribution of temperature at different location in a forest which is then sent to a central server through internet telemetry. These sensors do not have any reliable source of power–*i.e.*, they do not have batteries; they harvest power from vibrations caused by the wind breeze, and the sensors may loose power at any moment. To store the temperature data, what kind of memory will you use: SRAM, DRAM, FLASH or magnetic hard drive? Explain your answer. [Total 5 pts]

Q5 **Floating-gate transistor for storage:** Consider the following floating-gate transistor.

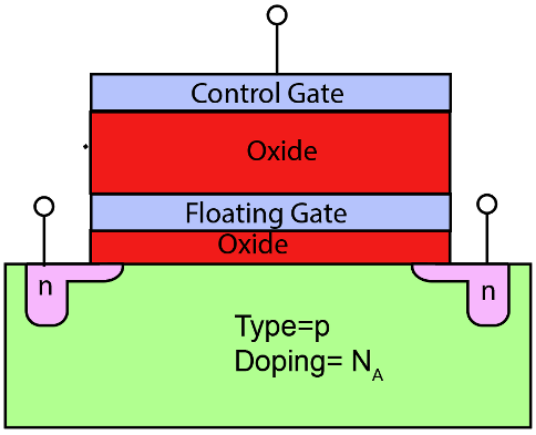


Figure 2: A floating gate transistor.

[Q5.1] How is data stored in this device. [Total 5 pts]

[Q5.2] How would you read the stored data in the device by reading the drain current? [Total 5 pts]

[Q5.3] Say you want to design the device for two different application: (1) Portable device, for example, a smart phone which a consumer will use only for an average of 3 years, (2) a long term storage, for example, an external hard-drive, which a consumer will use only for an average of 3 years. What will you change in the device and what will be the trade-off between data retention and voltage required for writing data into the device. [Total 10 pts]

Q6 **DRAM Array:** Consider the DRAM array shown in figure 3.

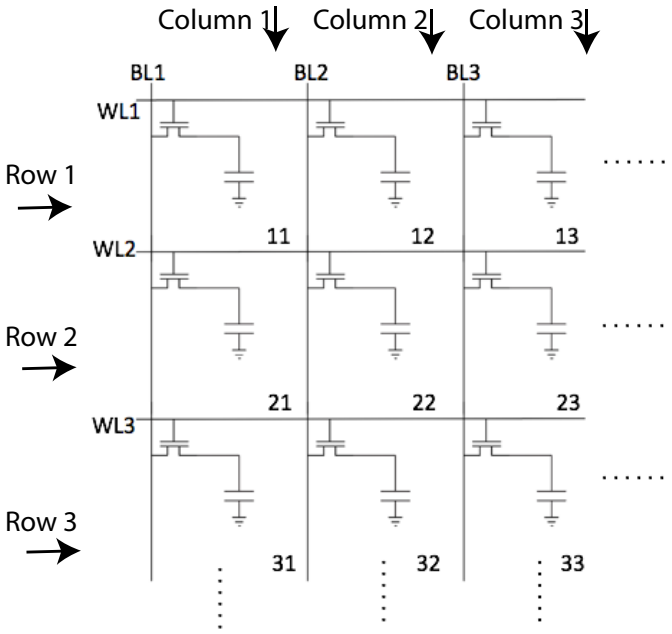


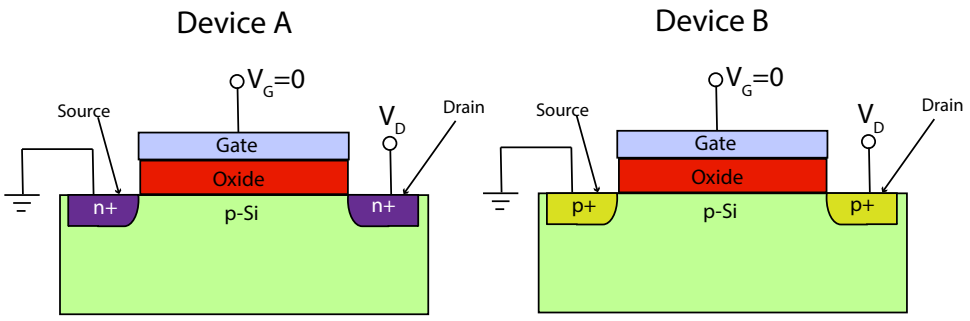
Figure 3: A DRAM array.

[6.1] Say you want to read all the cells in row 2. What is the sequence of operation you will need to perform? Make sure that, after your prescribed operations, you keep the data in the cells you read intact. [Total 5 pts]

[6.2] Explain in short why it necessary to have cache memories in today's microprocessor technology. [Total 10 pts]

[6.3] Compare and contrast the different memory technologies (SRAM, DRAM, Flash and magnetic disk) used in a memory hierarchy with respect to density, speed, and volatility [Total 10 pts]

Q7 **MOSFETS:** Consider the two devices shown in the following figure.



[7.1] Device A is the MOSFET structure that we discussed in the class and in which the source and drain are n^+-type (heavily doped n-type). On the other hand, in device B source and drain are p^+-type (heavily doped p-type). Based on what we discussed in class, do you think device B will behave like a switch–*i.e.* will device B be able to block current from flowing between the drain and the source terminal when the gate voltage V_G is zero (or less than the threshold voltage)? Provide justification for your answer. [Total 10 pts]

[7.2] **Dynamic Voltage and Frequency Scaling (DVFS):** Consider the logic blocks shown in figure 4. Logic block 2 and 3 receives input from logic block 1 at the same time. Logic block 5 needs inputs from blocks 3 and 4 for generate the final output. Logic block 3 receives input from logic block 2. Logic block 2, 3 and 4 requires 50, 10, and 30 cycles, respectively, to generate the respective output.

Based on what you have learned in class, how will you apply DVFS in these system? For which logic block(s), will you increase or decrease the power supply voltage and by how much? How much energy will you save in your prescribed process? Explain your answer. Make necessary and simplest possible reasonable assumptions. [Total 10 pts]

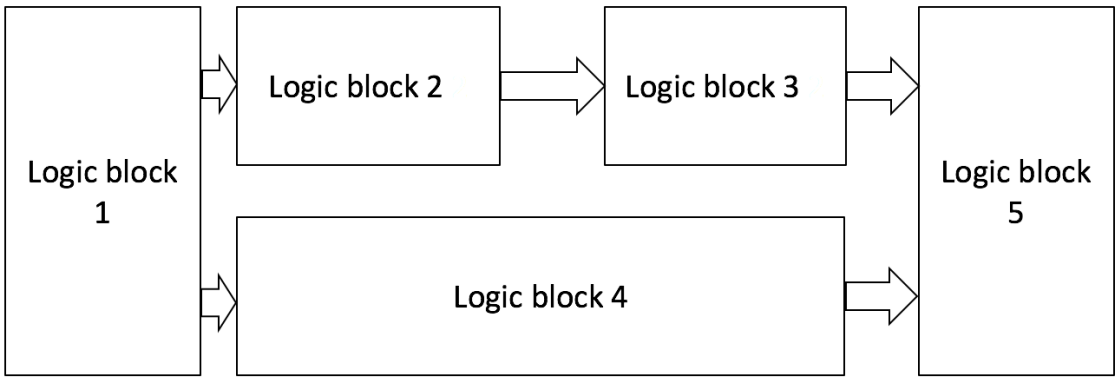


Figure 4: Dynamic Voltage and Frequency Scaling (DVFS).

[7.3] **Energy, Power, and Heat** As we know in Moore's law, the transistor density of devices increases at an exponential rate. However, the power supply voltage, V_{DD}, did not decrease until after 1999. Provide a few reasons why modern semiconductors face these limitations. [Total 5 pts]

Constants:
Electron charge q=1.6×10^{-19} C
Vacuum permittivity $\epsilon_\circ$=8.854×10^{-12} F/m Intrinsic carrier density of Si $n_i = 1.5 \times 10^{16}$ m^{-3} Relative dielectric constant of Si ϵ_{Si}=12
Relative dielectric constant of SiO_2 ϵ_{ox}=4
$N_C \approx N_V$=10^{25} m^{-3}
Bandgap of Si E_g=1.1 eV
Boltzmann constant k_B= 1.38 ×10^{-23} m^{2}kg s^{-2} K^{-1}
k_BT/q=26 mV (T=room temperature).

Equations:
Intrinsic carrier density $n_i = \sqrt{N_C N_V} e^{-\frac{E_g}{2k_BT}}$; Electron density $n = N_C e^{\frac{E_F - E_C}{k_BT}}$; Hole density $p = N_V e^{\frac{E_V - E_F}{k_BT}}$
In a p-type semiconductor, hole density p=N_A and $n = \frac{n_i^2}{N_A}$.
In a n-type semiconductor, hole density $p = \frac{n_i^2}{N_D}$ and $n = N_D$.

p-n Junctions:
Built in potential $V_{bi} = \frac{k_BT}{q} \ln \frac{N_A N_D}{n_i^2}$
Depletion width $W = \sqrt{\frac{2\epsilon_\circ \epsilon_{Si}}{q}(\frac{1}{N_A} + \frac{1}{N_D})(V_{bi} - V)}$; V is the voltage applied across the pn junction. V is positive and negative when the pn junction is forward and reverse biased, respectively.
Depletion width in p-side $W_p = \frac{N_D}{N_A + N_D} W$
Depletion width in n-side $W_p = \frac{N_A}{N_A + N_D} W$
Maximum electric field $E_{max} = \frac{qN_A W_p}{\epsilon_\circ \epsilon_{Si}} = \frac{qN_D W_n}{\epsilon_\circ \epsilon_{Si}}$

MOS Capacitor:
Gate voltage $V_G = V_{ox} + \psi_s$; V_{ox} is the voltage drop across the oxide and ψ_S is the surface potential (electrostatic potential at the oxide-semiconductor interface).
Oxide capacitance $C_{ox} = \frac{\epsilon_\circ \epsilon_{ox}}{t_{ox}}$
Depletion width $W = \sqrt{\frac{2\epsilon_\circ \epsilon_{Si}}{qN_A}\psi_s}$
Maximum depletion width $W_{max} = \sqrt{\frac{2\epsilon_\circ \epsilon_{Si}}{qN_A} 2\psi_B}$
Threshold voltage $V_t = \frac{\sqrt{4\epsilon_\circ \epsilon_{Si} q N_A \psi_B}}{C_{ox}} + 2\psi_B$; $\psi_B = \frac{|E_i - E_F|}{q}$

MOSFET:
Long channel MOSFET (square law model):

$$\frac{I_D}{W} = \begin{cases} 0; \text{ when } V_G < V_t \\ \mu C_{ox} \frac{1}{L}((V_G - V_t)V_D - \frac{1}{2}V_D^2); \text{ when } V_G - V_t > V_D \\ \mu C_{ox} \frac{1}{2L}(V_G - V_t)^2; \text{ when } V_G - V_t < V_D \end{cases}$$

Long channel MOSFET (square law model with correction for subthreshold current):

$$\frac{I_D}{W} = \begin{cases} I_{sub-V_t} e^{\frac{V_G - V_t}{m k_B T}} (1 - e^{\frac{-qV_D}{k_B T}}); \text{ when } V_G < V_t \\ I_{sub-V_t}(1 - e^{\frac{-qV_D}{k_B T}}) + \mu C_{ox} \frac{1}{L}((V_G - V_t)V_D - \frac{1}{2}V_D^2); \text{ when } V_G - V_t > V_D \\ I_{sub-V_t}(1 - e^{\frac{-qV_D}{k_B T}}) + \mu C_{ox} \frac{1}{2L}(V_G - V_t)^2; \text{ when } V_G - V_t < V_D \end{cases}$$

Body factor $m = 1 + \frac{C_D}{C_{ox}}$, where $C_D = \frac{\epsilon_\circ \epsilon_{Si}}{W}$ (depletion capacitance).

Velocity saturated MOSFET: $I_{D,sat} = \mu C_{ox} W v_{sat}(V_G - V_t)$; where v_{sat} is the saturation velocity of the carriers.

On-state current at a given power supply voltage V_{DD}, $I_{ON} = I_D(V_{GS} = V_{DS} = V_{DD})$
Off-state leakage current at a given power supply voltage V_{DD}, $I_{OFF} = I_D(V_{GS} = 0, V_{DS} = V_{DD})$

Inverter:

$$\text{Middle voltage } V_M = \frac{\sqrt{\frac{\beta_p}{\beta_n}}(V_{DD} - |V_{tp}|) + V_{tn}}{1 + \sqrt{\frac{\beta_p}{\beta_n}}}$$

where β_i=$\mu_i C_{ox}(\frac{W_i}{L_i})$, $i \equiv p, n$. μ_i=mobility, W_i=i-type MOSFET width; L_i=i-type MOSFET gate length. MOSFET effective resistance $R_t = \frac{10V_B + 3V_t}{6\beta V_B^2} = \frac{10V_{DD} - 7V_t}{6\beta(V_{DD} - V_t)^2}$
where $V_B = V_{DD} - V_t$; β=$\mu C_{ox}(\frac{W}{L})$. μ=mobility, W= MOSFET width; L= MOSFET gate length.

Inverter delay is proportional to $R_t C_L$. $C_L = C_{g,p} + C_{g,n}$ where $C_{g,i}$=gate capacitance of i-type MOSFET = $\epsilon_\circ \epsilon_{ox} W_i L_i / t_{ox}$, $i \equiv p, n$.

Power dissipation due to charging and discharging = $C_L V_{DD}^2 f$; f being the frequency.

Resistance of a wire $R = \rho \frac{\text{Length}}{\text{Area}}$; ρ being the resistivity.
Capacitance of a parallel plate capacitor $C = \frac{\epsilon_\circ \epsilon_{ox} \text{Area}}{\text{Thickness}}$.

Elmore delay $\delta_E = \sum_{1 \le i \le n} c_i \sum_{1 \le j \le i} r_j$, c_i and r_j being the capacitance and the resistance of i-th and j-th section respectively.
If all the segments have the same resistance and capacitance $\delta_E = n(n+1)rc/2$, n being the number of segments.

Chip junction temperature $T = T_A + P R_{thermal}$, T_A, P and $R_{thermal}$ being the ambient temperature, power and thermal resistance, respectively.
Thermal time constant $\tau_{thermal} = R_{thermal} C_{thermal}$, $C_{thermal}$ being the thermal capacitance.