ECE 3030: Physical Foundations of Computer Engineering

Fall 2021

Final exam

Time: 2 hr 50 min

Instructor: Asif Khan

Instructions:

1. There are 14 pages in this test. A list of constants and equations is provided on pages 13, and 14.
2. Read all the problems carefully and thoroughly before you begin working.
3. You are allowed to look at your notes, the video lectures, textbooks and any material and can have access to the Internet to search for answers to the questions.
4. There are a total of 6 problems that you are required to answer and 1 bonus problem that you may or may not choose to answer. If you answer the 6 mandatory problems correctly and do not answer the bonus problem, your score will be 100. If you answer the 6 mandatory and the bonus problem correctly, your score will be 110 (out of 100).

Q1	15 pts
Q2	5 pts
Q3	5 pts
Q4	15 pts
Q5	30 pts
Q6	30 pts
Bonus	10 pts
Total	100 pts

5. Show all your work and circle/underline your final answer. For numerical answers, write the units. Write legibly. If l 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. Answer the following questions. Circle the correct answer when appropriate. [Total 15 pts]

[Q1.1] What are three classes of materials used in an integrated circuit? [3 pts]
Solution: Semiconductor, metal, oxide.

[Q1.2] Which one of the following devices uses all these three classes materials? **(A capacitor/An inductor/A p-n junction diode/✓A MOSFET)** [2 pts]

[Q1.3] Which one of the following devices uses only one of the three classes of material? **(A capacitor/An inductor/✓A p-n junction diode/A MOSFET)** [2 pts]

[Q1.4] "When the input to an inverter is a logical 1, the n-MOSFET is in the on-state."–is this statement true? **✓Yes/No** [1 pts]

[Q1.5] "In general, the delay in an inverter increases if the power supply voltage increases."–is this statement true? **Yes/✓No** [1 pts]

[Q1.6] "In general, the delay in an interconnect increases when the power supply voltage increases.."–is this statement true? **Yes/✓No** [1 pts]

[Q1.7] When the device dimensions, power supply voltage and the clock frequency are all scaled as per Dennard's scaling law, does the active power dissipated in a MOSFET decrease? **✓Yes/No** [1 pts]

[Q1.8] The total length of wires/interconnects in a microprocessor chip (such as the one in your laptop computer) is of the order of **5m/50 m/500 m/5 km/✓50 km**. [1 pts]

[Q1.10] "To make the voltage transfer curve of an inverter symmetric, the p-MOSFET needs to be wider than the n-MOSFET."–is this statement true? **✓Yes/No** [1 pts]

[Q1.11] "The current (I_D) vs. gate voltage (V_G) curve of a n-MOSFET at a given drain voltage (V_D) looks like the curve shown in figure 1."–is this statement true? **Yes/✓No** [2 pts]

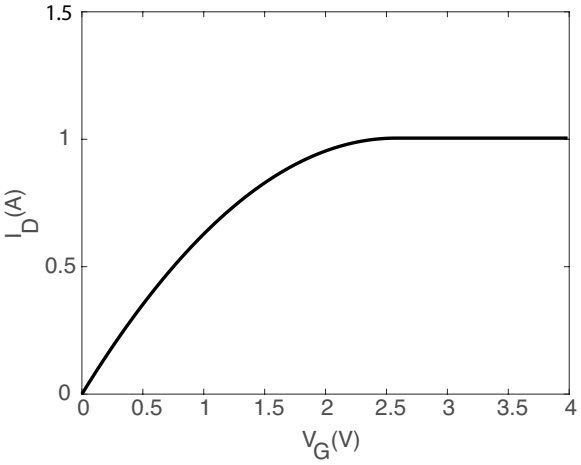
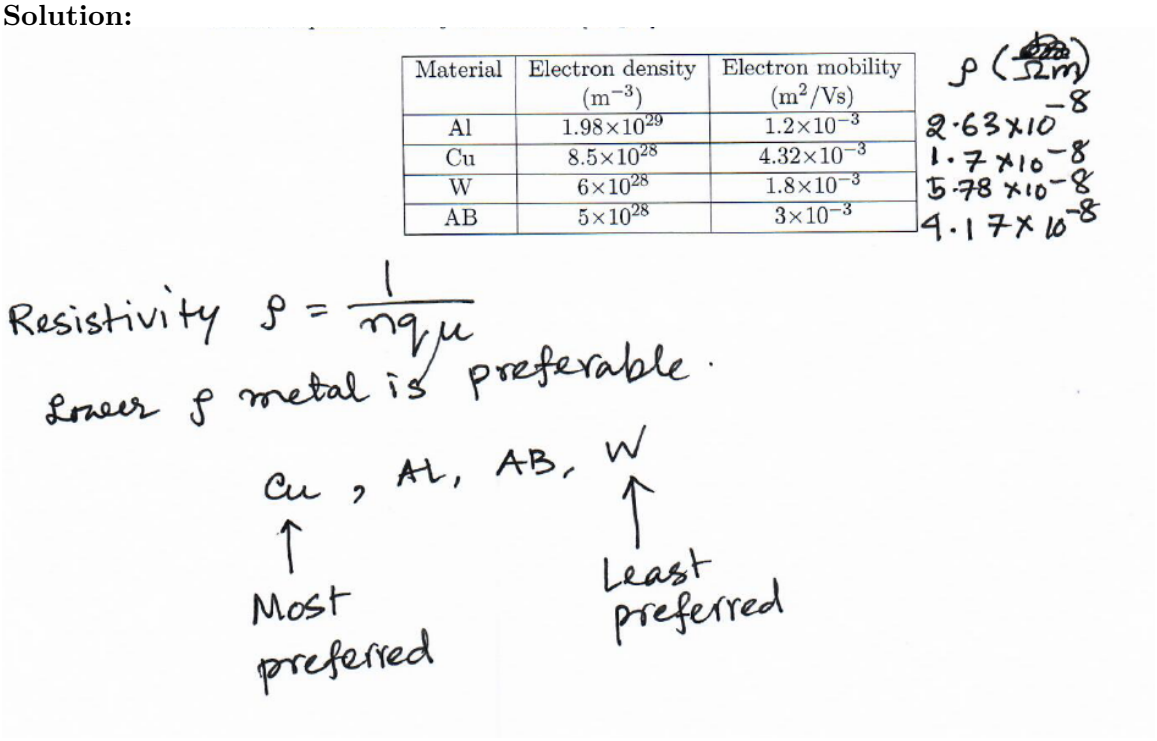


Figure 1: Q1.10

Q2 **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. [5 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}

Solution:



Q3 **Physics of semiconductors:** Draw the band diagram (the relative positions of conduction band edge E_C, valence band edge E_v, Fermi level E_F) for Si wafer with acceptor doping N_A=5×10^{23} m^{-3} and no donor doping. Clearly indicate the values of $E_C - E_F$, $E_F - E_V$, $E_i - E_F$, $E_G = E_C - E_V$. E_i is the intrinsic Fermi level. Assume N_C=N_V=10^{25} m^{-3}, E_G=1.1 eV, n_i=1.5×10^{16} m^{-3}, kT=0.026 eV. [Total 5 pts]

Solution:

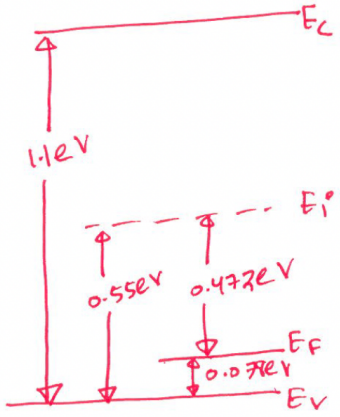
$P = N_A = 5 \times 10^{23}\ m^{-3}$

$\Rightarrow N_A = N_V\, e^{\frac{E_V - E_F}{K_B T}}$

$\Rightarrow E_V - E_F = K_B T \ln \frac{N_A}{N_V}$

$= 0.026 \ln \frac{5 \times 10^{23}}{10^{25}}$

$= -0.078\ eV$



Q4 **Current-voltage characteristics of a MOSFET:** The following figure shows how the drain current I_D in a long channel MOSFET changes as a function of the gate voltage V_{GS} at given value of the drain voltage V_{DS}. The range of the V_{GS} range for different modes of operation (cut-off, saturation and linear) also indicated. [Total 10 pts]

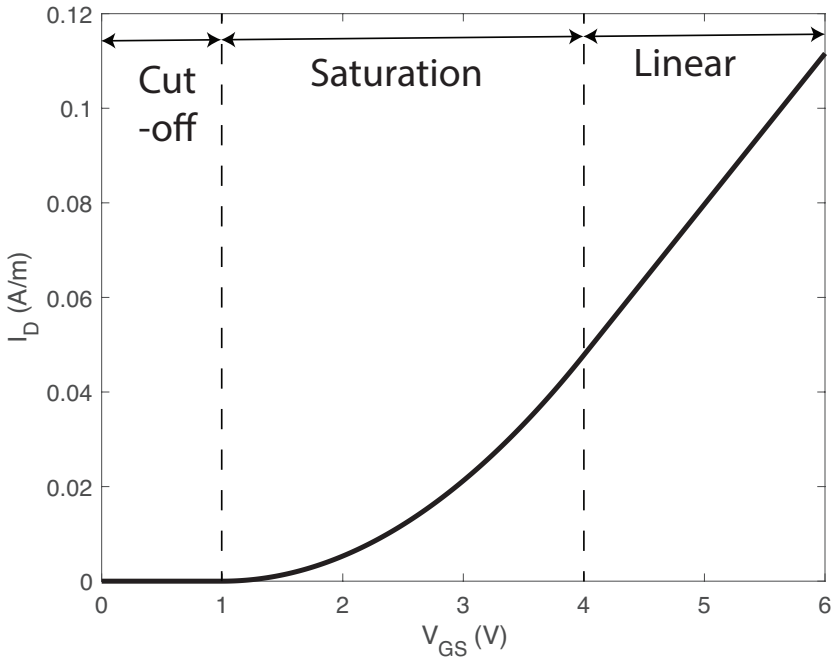


Figure 2: MOSFET $I_D - V_{GS}$ characteristics.

[Q4.1] What is the threshold voltage V_t of the MOSFET? [5 pts]

Solution:
Cut-off region: V_{GS}: 0—1 V. Hence, threshold voltage V_t=1 V.

[Q4.2] What is the value of the drain voltage V_{DS}? [10 pts]

Solution:

In the linear region, $V_{GS} - V_t \geq V_{DS}$

At the boundary between linear and saturation region, $V_{GS} - V_t = V_{DS}$.

Hence, $V_{DS} = V_{GS} - V_t = 4 - 1 = 3V$.

Q5 **Memory Technologies:** [Total 30 pts]

[Q5.1] In DRAMs, why is it necessary to perform a restore operation after every time a cell is read. [5 pts]

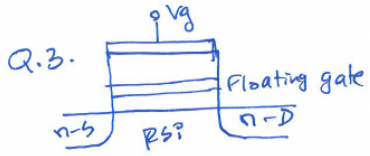
Solution: The read operation in a DRAM is destructive–*i.e.*, after a DRAM cell is read, the data is lost. As such, the data read needs to be re-written after a cell has been read.

[Q5.2] **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]

Solution: A random access memory device allows data items to be read or written in almost the same amount of time irrespective of the physical location of data inside the memory. In a magnetic hard drive, the read head first needs to move to the target track on the magnetic disk, then the disk needs to rotate such the target section comes under the head and finally the entire sector is read by the head. As such, the total amount of time to retrieve the data depends in the initial position of the head and the physical location of the data–*i.e.*, the access time depends on the physical location of the data. Hence, the magnetic hard drive technology is not classified as a random access memory technology.

[Q5.3] How is a floating gate MOSFET is programmed and erased. Explain with necessary diagrams. [10 pts]

Solution:



Programming: In order to program the Mosfet & put data = 0, we need to put electrons in the floating gate. To do that, we should ground the Source, put a large Voltage at ~~V_G~~ Gate and Drain. Then due to tunneling current or hot electron effect, charge will be stored in the floating gate & the cell stores 0.

Erasing: In order to erase ie. put data = 1, we need to ~~put~~ remove electrons from the floating gate. To do that, we should keep the ground open-circuited, put a large Voltage at Drain & put 0 voltage at Gate. Then, the tunneling will drain the charge from the floating gate to the source & the cell will thus store 1.

[Q5.4] Rank the different memory technologies (SRAM, DRAM, Floating Gate, Magnetic) with respect to their speed. [3 pts]
Solution: Speed: SRAM (fastest)>DRAM>Floating Gate>Magnetic (slowest)

[Q5.5] Rank these memory technologies with respect to their maximum achievable densities. [3 pts]
Solution: Density: SRAM (least dense)<DRAM<Floating Gate<Magnetic (densest)

[Q5.6] Which of these technologies are volatile and which are non-volatile. [4 pts]
Solution: Volatile: SRAM, DRAM; non-volatile: Floating Gate, Magnetic.

Q6 **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 30 pts]

[Q6.1] If, in every subsequent generation, the total number of transistors doubles, what is the nominal value of x? [10 pts]

Solution to Q6.1:

is the nominal value of x? [2 pts]

Total #transistors : $N_n = \frac{A_{c,n}}{W_n L_n}$ $N_{n+1} = \frac{1.05 A_{c,n}}{W_{n/x} L_{n/x}} = 1.05 x^2 N_n$

$\frac{N_{n+1}}{N_n} = 1.05 x^2 = 2 \Rightarrow x = \sqrt{\frac{2}{1.05}} = 1.38$

[Q6.2] 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}. [10 pts]

Solution to Q6.2:

Total gate cap of a transistor: $C_{L,n} = C_{ox,n} W_n L_n = \frac{\varepsilon_0 \varepsilon_{ox}}{t_{ox,n}} W_n L_n$

$C_{L,n+1} = C_{ox,n+1} W_{n+1} L_{n+1}$

$= \frac{\varepsilon_0 \varepsilon_{ox}}{t_{ox,n/x}} \frac{W_n}{x} \frac{L_n}{x} = C_{ox,n} \times \frac{1}{x^2}$

Power per transistor : $P_n = C_L V_{DD}^2 f_n$ $P_{n+1} = C_{L,n+1} \frac{V_{DD}^2}{y^2} f_{n+1}$

$= C_{L,n} \frac{1}{x^2} \frac{V_{DD}^2}{y^2} f_{n+1}$

$= P_n \frac{1}{x^2 y^2} \frac{f_{n+1}}{f_n}$

[Space for Q6.2]

Total chip power : $P_{chip,n} = P_n N_n = P_n \frac{A C_n}{V_n f_n}$

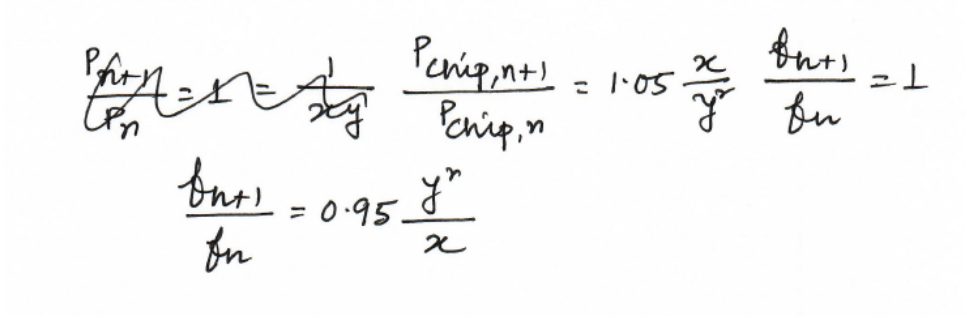
$$P_{chip,n+1} = P_{n+1} N_{n+1}$$

$$= P_n \frac{1}{xy^2} \frac{f_{n+1}}{f_n} 1.05 x^2 N_n = P_n N_n \times \frac{x}{y^2} 1.05 \times \frac{f_n}{f_n}$$

$$\frac{P_{chip,n+1}}{P_{chip,n}} = 1.05 \frac{x}{y^2} \frac{f_{n+1}}{f_n}$$

[Q6.3] 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. [10 pts]

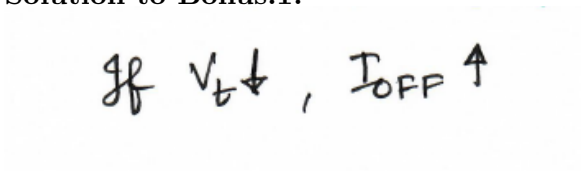
Solution to Q6.3:



[Bonus] The bonus problem is a continuation of Q6.

[Bonus.1] 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]

Solution to Bonus.1:



[Bonus.2] 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 Q7.1 while keeping the total active/dynamic chip power and the clock frequency constant in each subsequent generation. [5 pts]

Solution to Bonus.2:

$$P_{chip,n} = P_{chip,n+1}$$

$$f_{n+1} = f_n$$

$$\frac{P_{chip,n+1}}{P_{chip,n}} = 1.05 \frac{x}{y^2} \frac{f_{n+1}}{f_n}$$

$$\Rightarrow 1 = 1.05 \frac{1.38}{y^2}$$

$$\Rightarrow y = \sqrt{1.05 \times 1.38} = 1.2$$

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:
Conductivity of a metal $\sigma = nq\mu$
Resistivity of a metal $\rho = 1/\sigma$
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 $V_t = \frac{\sqrt{4\epsilon_\circ \epsilon_{Si} qN_A \psi_B}}{C_{ox}} + 2\psi_B$; ψ_B=$\frac{|E_i - E_F|}{q}$

MOSFET:
Long channel MOSFET (square law model):

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

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

$$\frac{I_D}{W} = \begin{cases} I_{sub-V_t} e^{\frac{q(V_G-V_t)}{mkT}}(1-e^{\frac{-qV_D}{kT}}); \text{ when } V_G < V_t \text{ (sub-threshold)} \\ I_{sub-V_t}(1-e^{\frac{-qV_D}{kT}}) + \mu C_{ox}\frac{1}{L}((V_G-V_t)V_D - \frac{1}{2}V_D^2); \text{ when } V_G - V_t > V_D \text{ (linear)} \\ I_{sub-V_t}(1-e^{\frac{-qV_D}{kT}}) + \mu C_{ox}\frac{1}{2L}(V_G-V_t)^2; \text{ when } V_G - V_t < V_D \text{ (saturation)} \end{cases}$$

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.
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.