

Metric Prefixes

peta	P	10^{15}	1 000 000 000 000 000
tera	T	10^{12}	1 000 000 000 000
giga	G	10^9	1 000 000 000
mega	M	10^6	1 000 000
kilo	k	10^3	1 000
hecto	h	10^2	100
deca	da	10^1	10
one		10^0	1
deci	d	10^{-1}	0.1
centi	c	10^{-2}	0.01
milli	m	10^{-3}	0.001
micro	μ	10^{-6}	0.000 001
nano	n	10^{-9}	0.000 000 001
pico	p	10^{-12}	0.000 000 000 001
femto	f	10^{-15}	0.000 000 000 000 001

De Morgan's Laws

- $\overline{AB} = \overline{A} + \overline{B}$
- $\overline{A + B} = (\overline{A})(\overline{B})$

Silicon

- Si
- P-type:
 - * doped with material to remove electrons (add electron holes), usually Boron (B), Aluminum (Al), or Gallium (Ga)
- N-type:
 - * doped with material to add electrons, usually Antimony (Sb), Arsenic (As), or Phosphorous (P)
- Silicon dioxide: SiO_2
- Polysilicon is just silicon without the crystal structure

Transistors

- pMOS:
 - * has the bubble
 - * on when input is 0, off when input is 1
- nMOS
 - * no bubble
 - * on when input is on, off when input is off
- CMOS: when you combine a nMOS and pMOS network together to make a gate, where one is the compliment of the other
- V_t : Threshold voltage. Nominal voltage below which the transistor is off
 - * below as in closer to 0, not less
 - * $V_t > 0$ for nMOS, $V_t < 0$ for pMOS
 - * this is compared to the gate to source voltage, V_{gs}
- regions: (for nMOS)
 - * accumulation: gate is negatively charged, attracts positive voids in p-substrate, which block flow in the channel
 - * depletion: small positive charge on gate repels positive voids from channel, forming a depletion below the gate
 - * inversion: higher positive charge ($> V_t$) is applied to gate, attracting electrons to the channel and allowing flow

D Flip Flop vs Latch

- latch is level triggered
- flip flop is edge triggered

Fabrication

- n-well: use diffusion or ion implantation
- positive lithography: expose to UV where you want to remove material
- negative lithography: expose to UV where you want to keep material

Stick Diagram vs Boolean Function**• TODO**

- there is more than one way of making a stick diagram for an expression
 - * stuff in series could be in different order, for instance
- if you do it manually, you need to check it with tools:
 - * LVS: Layout Vs Schematic
 - * DRC: Design Re-Check
 - * These aren't really needed for designs automatically generated from verilog code, because of course that's correct

Lithography

- the process of printing onto a chip at nanometer scale
- generally uses UV light, wavelength around 150nm
 - * must use fancy tricks to make 10nm features with 150nm light
 - * would be nice to use even lower wavelength X-rays, but those are hard to focus
- negative lithography: use the lithography mask to cover what you want to keep.
- positive lithography: mask what you want to remove
- a lens is used to focus the light
 - * ideally, want a point source for the light, but that is not practical
 - * Optical Proximity Effect: what happens when your focus from the lens is not just right
 - * results in rounded corners, inaccurate critical dimensions, and shorter wire ends
 - * can use Optical Proximity Correction to fix: basically over-emphasize all the features, and/or add extra lines at outset

MOS transitive I - V Characteristics and Parasitics

- I - V : current-voltage relationship
- Transistors are not really ideal switches, they have 3 zones of operation: cutoff, linear, saturation
- definitions:
 - V_{gs} : voltage gate to source
 - V_{gd} : voltage gate to drain
 - V_{ds} : voltage source to drain (across the channel)
 - V_t : critical voltage at which transistor is saturated
 - channel: space between the source and drain, where the electrons flow
- remember that the gate is insulated from the area under it by a thin layer of Silicon Dioxide (SiO_2)
- by convention, the source is the terminal at lower voltage
- **cutoff**:
 - * when $V_{gs} < 0$
 - * electrons on the gate attract positive voids in the silicon below, and inhibit current flow. Therefore, the transistor is closed.
 - * $I_{ds} = 0$
- **linear**:
 - * $V_{gs} > V_t, V_{gd} = V_{gs}, V_{ds} = 0$ or $V_{gs} > V_t, V_{gs} > V_{gd} > V_t, 0 < V_{ds} < V_{gs} - V_t$
 - * I_{ds} linearly proportional to V_{ds}
 - * channel of electrons forms, allowing current to flow
- **saturated**:
 - * $V_{gs} > V_t, V_{gd} < V_t, V_{ds} > V_{gs} - V_t$
 - * channel pinches off due to electrons attracting to source
 - * I_{ds} is independent of V_{ds}

capacitor effect

- gate and channel can have a parallel plate capacitor effect, with the thin layer of SiO_2 acting as the insulator
- $C = \frac{Q}{V} = \epsilon_{\text{SiO}_2} w l / t_{\text{SiO}_2}, V = V + g c - V_t = (v_{gs} - V_{ds}/2) - V_t$

- * l, w : length, width of section of gate above channel
- * ϵ_{SiO_2} : permittivity of SiO_2 layer
- * t_{SiO_2} : thickness of SiO_2 layer
- general capacitance per unit area: $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$
- * ϵ_{ox} : permittivity of oxidation layer
- * t_{ox} : thickness of oxidation layer
- carrier velocity: velocity of the electrons?
- * proportional to the electric field running horizontally between source and drain
- * $v = \mu E, E = V_{ds}/L, t = L/v = L/(\mu E) = L/(\mu \frac{V_{ds}}{L})$
- μ : mobility. electrons move about twice as fast as positive voids
- L : length of channel
- actual velocity of electrons is the speed of light, but they don't travel in a straight line, they travel atom-to-atom
- * this slowdown is called the **scattering** effect

Shockley model of transistor

- $V_{dsat} = V_{gs} - V_t$
- 1st order model: $\beta = \mu C_{SiO_2} \frac{w}{l}$
- $V_{gs} < V_t$ $I_{ds} = 0$ cutoff
- $V_{ds} < V_{dsat}$ $I_{ds} = \beta(V_{gs} - V_t - \frac{V_{ds}}{2})V_{ds}$ linear
- $V_{ds} > V_{dsat}$ $I_{ds} = \frac{\beta}{2}(V_{gs} - V_t)^2$ saturation
- Must be able to derive this model on exam!

Non-Ideal I - V Effects

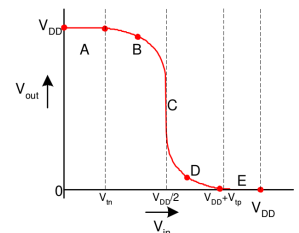
- velocity saturation (due to scattering)
- * also called short channel effect
- * this is hard to make into a mathematical formula
- sub-threshold leakage, junction leakage, gate tunneling
- **Body Effect**
- * affected by V_{sb} : voltage of the p-substrate (which should be ground)
- * $V_t = V_{t0} + \lambda(\sqrt{|-2\phi_F + V_{sb}|} - \sqrt{|-2\phi_F|})$
- V_{t0} : threshold without body bias
- ϕ_F : Fermi potential
- negative for nMOS, positive for pMOS
- λ : body effect coefficient
- positive for nMOS, negative for pMOS (reversed)
- * generally fixed/caused by biasing
- * Forward Body Bias (FBB):
- $V_{sb} < 0, V_t < V_{s0}$
- gates switch faster, but leak more current
- * Reverse Body Bias (RBB):
- $V_{sb} > 0, V_t > V_{s0}$
- gates switch slower, but consume less power (because less leakage)
- **Temperature**
- * higher temperature means higher electron mobility, more leakage, and threshold decreases
- **Diffusion Capacitance**
- * capacitance on the source/drain: C_{sb}, C_{db}
- source/drain are diffusion nodes
- C is comparable to C_g (gate) for connected nodes, $\frac{1}{2}C_g$ for unconnected
- (depends on process)
- * this is the capacitance we care most about (because we must fill it every time the gate switches?)
- * if two capacitors share a source/drain, that reduces diffusion capacitance (reducing this is a good thing)
- * also, you can remove unconnected diffusion spots

DC Response

- example: inverter
- must settle to $I_{dsn} = |I_{dsp}|$
- V_{tn}, V_{tp} : threshold voltages for nMOS and pMOS (defined in Non-Ideal I - V Effects)

- $V_{in} = V_{DD} \rightarrow V_{out} = 0, V_{in} = 0 \rightarrow V_{out} = V_{DD}$
- nMOS
- * $V_{gsn} = V_{in}, V_{dsn} = V_{out}$
- * cutoff: $V_{in} < V_{tn}$
- * linear: $V_{in} > V_{tn}, V_{out} < V_{in} - V_{tn}$
- * saturated: $V_{in} > V_{tn}, V_{out} > V_{in} - V_{tn}$
- pMOS
- * $V_{gsp} = V_{in} - V_{DD}, V_{dsp} = V_{out} - V_{DD}, T_{tp} < 0$
- * cutoff: $V_{in} > V_{DD} + V_{tp}$
- * linear: $V_{in} < V_{DD} + V_{tp}, V_{out} > V_{in} - V_{tp}$
- * saturated: $V_{in} < V_{DD} + V_{tp}, V_{out} < V_{in} - V_{tp}$
- to calculate actual output voltage, balance $I_{dsn} = I_{dsp}$ (easiest to do graphically)
- * end up with a graph of V_{out} as function of V_{in}

Region	nMOS	pMOS
A	Cutoff	Linear
B	Saturation	Linear
C	Saturation	Saturation
D	Linear	Saturation
E	Linear	Cutoff



- * don't want to switch nMOS and pMOS because then nMOS would saturate at far end of graph (and vice versa for pMOS)
- * horizontal position of graph can be varied by tuning β_p/β_n
- called beta ratio, or skewed gate
- $\beta_p/\beta_n > 1 \rightarrow$ right, $\beta_p/\beta_n \rightarrow$ left
- * unity gain slope: part of response graph where the slope is -1
- want to tune β_p/β_n to put logic levels at these regions to maximize noise margins
- * Noise Margins:
- $NM_H = |V_{OH} - V_{IH}|, NM_L = |V_{OL} - V_{IL}|$
- that's just the higher/lower of the two axes

Transient Analysis

- for instance, find step response of gate to determine rise time.
- rise/fall delay: time from when V_{in} crosses $\frac{V_{DD}}{2}$ to when V_{out} crosses it
- rise/fall time: (of V_{in} or V_{out}): time for that signal to go from $0.1V_{DD}$ to $0.9V_{DD}$ (or reverse)
- TODO many equations and such for inverter step response (from slides)
- TODO pass transistors (from slides)

Pass Transistors

- for nMOS trying to pass V_{DD} or pMOS trying to pass 0
- nMOS can pull no higher than $V_{DD} - V_{tn}$ if $V_g = V_{DD}$
- * more generally, $V_g - V_{tn}$
- * called degraded 1
- pMOS can pull no lower than $|V_{tp}|$

Delay

- generally estimated with RC models
- * for nMOS with width k :
- * resistance of R/k
- * caps of kC on all terminals
- * pMOS same except resistance is $2R/k$
- depends on effective R and C of transistors
- * exactly what parasitic caps depends on exact layout (which stuff is shared between transistors)
- width:
- * C proportional to width (approx $2 \text{ fF}/\mu\text{m}$)
- * R inversely proportional to width (approx $6 \text{ k}\Omega \cdot \mu\text{m}$)
- * TODO unity transistors
- find widths necessary for rise and fall resistance to be same as standard inverter

- * pMOS is about half as conductive as nMOS, so the inverter has nMOS=1, pMOS=2
- * larger width \rightarrow smaller R
- * transistors in series: delay adds, so double the width
- * transistors in parallel: same as a single transistor (because we assume worst case of only one being active)
- use effective resistance in RC model: $I_{ds} = V_{ds}/R$ (just good enough for a RC model, not for current at arbitrary time)
- find delay of circuit
 - * decompose to RC model (take into account widths of each transistor)
 - replace transistors with resistors
 - add parasitic caps on either side of every transistor (cap value = width of transistor)
 - caps with both pins short to ground don't count, are never charged
 - caps from V_{DD} to ground don't count, always charged
 - * add together all R and C's to get delay
 - don't count R of nMOS and pMOS at the same time, because they're never on at the same time
- Elmore Delay:
 - * ON transistors look like resistors, so pullup/pulldown network is modeled as RC ladder
 - * $t = \sum_{i \in \text{nodes}} R_{i \text{--} \text{to--} \text{source}} C_i$
 $= R_1 C_1 + (R_1 + R_2) C_2 + \dots + (R_1 + \dots + R_n) C_n$

Static Timing Analysis

- worst case at each step
- in form arrival time / required arrival time / slack
- arrival time: input to output, take max
- required arrival time: output, take min?
 - * or work backward?
- slack: required arrival time – arrival time
- Contamination delay: just best case delay (smallest delay)
 - * in this case, you'll count parallel transistors as parallel resistors
- not sure about this stuff, Peter just said to not worry about this question on the homework

Power Estimation

- Dynamic Power
 - * power required to charge the load capacitor
 - * only counted when transistor switches
 - therefore this power usage is data dependent
 - therefore, you have to count the falling transitions in the output per time period
 - * $P_{dynamic} = \frac{1}{T} \int_0^T i_{DD}(t) V_{DD} dt$
 - * for any gate: $P_{dynamic} = C V_{DD}^2 f_{sw}$
- Activity Factor α
 - * α : how often this gate switches in terms of the base clock frequency
 - * $\alpha = 1$: clock; $\alpha = 0.5$: every other cycle, etc...
 - * for system clock at frequency f , $P_{dynamic} = \alpha C V_{DD}^2 f$
- Short Circuit Current
 - * nMOS and pMOS may both be on for a short instant during switching, leading to a short instant of short circuit current (from V_{DD} to ground)
 - * $P_s \propto (V_{DD} - 2V_t)^3 t_r f_p$
 assume $t_r = t_f$ for input
 f_p : frequency of input
 - * this is less than 10% of dynamic power if the rise/fall times are comparable
- Static Power
 - * leakage when gate is off

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_t}{n v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}} \right)$$

$$V_t = V_{t0} - \eta V_{ds} + \gamma \left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right)$$