

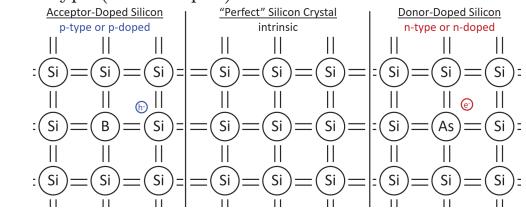
CG2027 Cheatsheet AY21/22 Sem 1

by richwill28

The Device

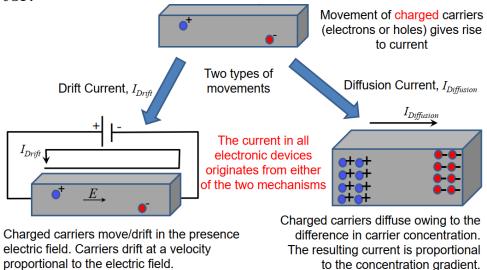
Semiconductor materials and doping

- Common semiconductor materials: Germanium, Silicon, and Gallium Arsenide.
- A charge carrier is a particle that is free to move, carrying an electric charge, especially the particles that carry electric charges in electrical conductors. Examples are electrons, ions and holes.
- Charge carrier density, also known as carrier concentration, denotes the number of charge carriers per volume.
- The carrier concentration of a semiconductor can be adjusted with doping.
- Doping is the intentional introduction of impurities into an intrinsic (pure) semiconductor. There are two types of doping, namely p-type (acceptor-doped) and n-type (donor-doped).



Current flow in a semiconductor

- There are two types of current flow in a semiconductor:



- Calculating current flow in a semiconductor:

Carrier Concentration	Drift Current Density	Diffusion Current Density
n	$qn\mu_n E$	$-qD_n \Delta n$
Total Current Density	Sum of <u>drift and diffusion</u> current densities for <u>electrons and holes</u>	

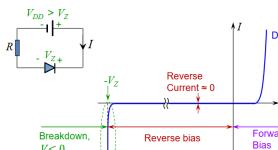
$$J_{TOTAL} = q_e n \mu_n E + q_h \mu_p E - q_e D_n \Delta n - q_h D_p \Delta p$$

Drift Diffusion

electrons holes electrons holes

- Current flow in circuit elements:

Devices	Movement Mechanism in on-state	Type of Carriers
Resistor	Drift	• Electrons (Metal) • Electrons and holes (Semiconductor)
Diode	Diffusion	• Electrons and holes
Bipolar Junction Transistor	Diffusion	• Electrons and holes
MOSFET	Drift	• Electrons (NMOS) • Holes (PMOS)



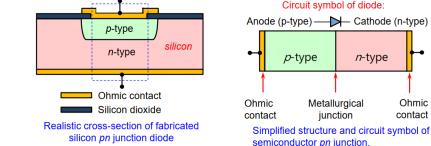
With an external voltage supply that reverse biases the diode, $V_{DD} > V_Z$ (breakdown voltage), the reverse current is no longer ≈ 0 but increases rapidly with practically no increase in the voltage across the diode. This condition is called **breakdown**.

Under breakdown condition, the voltage across the pn junction diode stays practically constant at $-V_Z$. Minus sign highlights that breakdown is a reverse biased condition.

- Operation in the breakdown region **does not** destroy the diode, provided the current through it is kept below a certain level, such that the power dissipation $V \cdot I$ is below what the diode can handle.
- Current, while operating in the breakdown region, can be limited by connecting a resistor, R , of suitable value in series with the pn-junction diode, such that $I = (V_{DD} - V_Z)/R$.

pn-junction diode

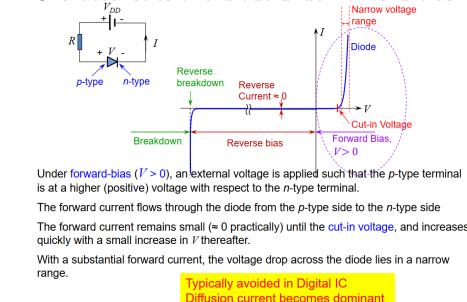
- Diode (semiconductor pn-junction) is the simplest (2-terminal) and most fundamental nonlinear circuit element.



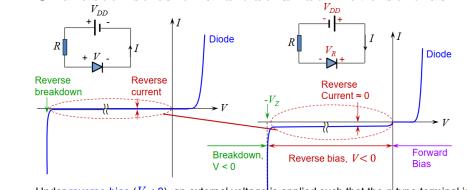
- Relationship between current and voltage:

$$I_D = I_{D0} \cdot (e^{\frac{V_D}{V_T}} - 1), \text{ where } V_T = \text{threshold voltage and } \eta = \text{non-ideality factor.}$$

- Characteristics of a diode under forward bias:



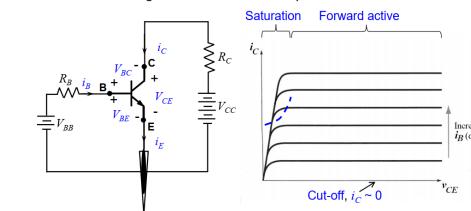
- Characteristics of a diode under reverse bias:



- Characteristics of a diode under breakdown region:

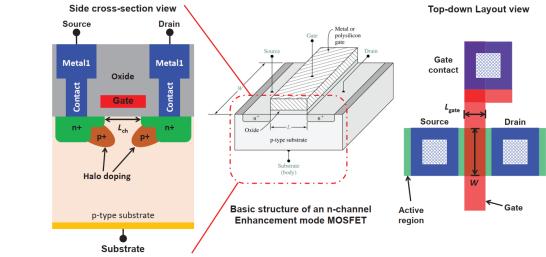
The regions corresponding to the forward active, saturation and cut-off modes of operation are as indicated in the plot below.

IV characteristics show the relationships between the collector current, i_C , and the collector-emitter voltage, V_{CE} , for different base currents, i_B , (or equivalently different V_{BE}). The current gain i_C/i_B is high around 100 in the forward active region which is used in amplifiers.

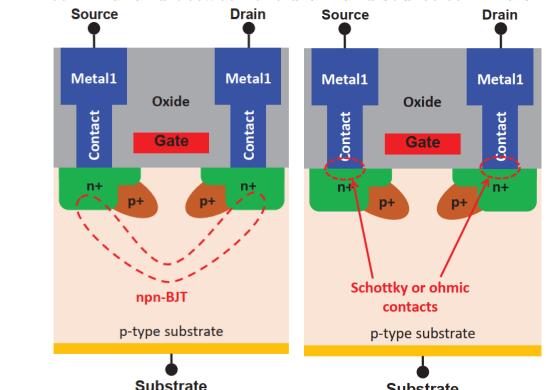


MOSFET

- MOSFET has 4 terminals, namely the gate, drain, source and substrate.



- The basic structure of the MOSFET is the region of the semiconductor just next to the oxide below the gate terminal and between the drain and source terminals.

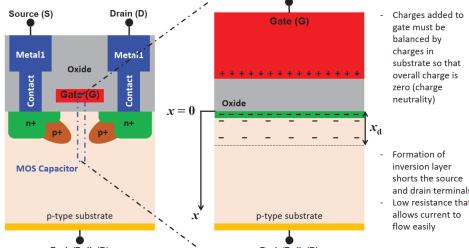


- Naively making the metal contacts to the source and drain terminals can form Schottky contacts, which have rectifying characteristics like pn-junction.
- The metal contacts should be engineered to be ohmic contacts instead, which behave more like resistor.

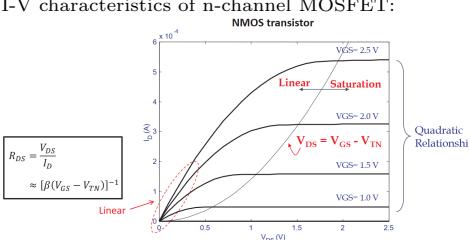
n-channel enhancement mode MOSFET

- The electrical characteristics n-channel MOSFET:

Mode of Operation	Emitter-Base Junction	Collector-Base Junction	Applications
Cut-off	Reverse biased ($V_{BE} < 0$ for npn)	Reverse biased ($V_{BC} < 0$ for npn)	Logic - OFF State
Forward Active	Forward biased ($V_{BE} > 0$ for npn)	Reverse biased ($V_{BC} < 0$ for npn)	Amplifier
Saturation	Forward biased ($V_{BE} > 0$ for npn)	Forward biased ($V_{BC} > 0$ for npn)	Logic - ON State
Reverse Active	Reverse Biased ($V_{BE} < 0$ for npn)	Forward Biased ($V_{BC} > 0$ for npn)	Not used



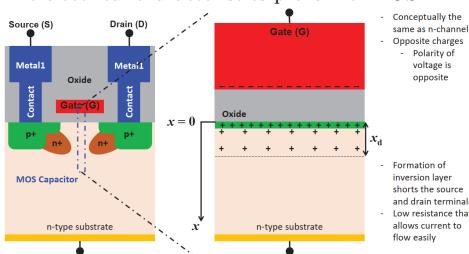
- Current must always flow from **drain to source**.
- The MOSFET input is the **gate voltage**, which controls the **resistance/conductance** between the **drain and source** terminals.
- For the n-channel MOSFET to work properly, the pn-junctions between the source and body terminals and between the drain and body terminals must both be reverse-biased. Hence, the body terminal must be connected to the **highest voltage** the source or drain terminal can have.
- The square law model:
Cut-off: $V_{GS} < V_{TH} \Rightarrow I_D \approx 0$
Linear/triode: $V_{GS} \geq V_{TH} \& V_{DS} < V_{DS,SAT} \Rightarrow I_D = \mu_n C_{ox} \frac{W}{L_{ch}} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \cdot (1 + \lambda_n V_{DS})$
Saturation: $V_{GS} \geq V_{TH} \& V_{DS} \geq V_{DS,SAT} \Rightarrow I_D = \mu_n C_{ox} \frac{W}{L_{ch}} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \cdot (1 + \lambda_n V_{DS})$
- In this module, we assume $V_{DS,SAT} = V_{GS} - V_{TH}$
- If voltages at all terminals are fixed, increasing W or W/L_{ch} will increase I_D .
- I-V characteristics of n-channel MOSFET:



- In the linear region of operation, the drain current varies **linearly** with V_{DS} .
- In the saturation region of operation, the drain current is controlled by V_{GS} and thus, the MOSFET behaves like a **voltage-controlled current source**. As we increase V_{GS} , I_D increases **quadratically**.

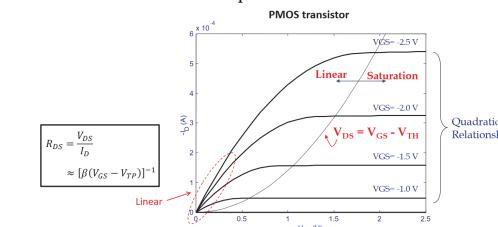
p-channel enhancement mode MOSFET

- The electrical characteristics p-channel MOSFET:



- Current must always flow from **source to drain**.

- The MOSFET input is the **gate voltage**, which controls the **resistance/conductance** between the **source and drain** terminals.
- For the p-channel MOSFET to work properly, the pn-junctions between the source and body terminals and between the drain and body terminals must both be reverse-biased. Hence, the body terminal must be connected to the **lowest voltage** the source or drain terminal can have.
- The square law model:
Cut-off: $V_{GS} > V_{TH} \Rightarrow I_D \approx 0$
Linear/triode: $V_{GS} \leq V_{TH} \& V_{DS} > V_{DS,SAT} \Rightarrow I_D = \mu_n C_{ox} \frac{W}{L_{ch}} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \cdot (1 + \lambda_n V_{DS})$
Saturation: $V_{GS} \leq V_{TH} \& V_{DS} \leq V_{DS,SAT} \Rightarrow I_D = \mu_n C_{ox} \frac{W}{L_{ch}} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \cdot (1 + \lambda_n V_{DS})$
- In this module, we assume $V_{DS,SAT} = V_{GS} - V_{TH}$
- If voltages at all terminals are fixed, increasing W or W/L_{ch} will increase I_D .
- I-V characteristics of p-channel MOSFET:



- In the linear region of operation, the drain current varies **linearly** with V_{DS} .
- In the saturation region of operation, the drain current is controlled by V_{GS} and thus, the MOSFET behaves like a **voltage-controlled current source**. As we decrease V_{GS} , I_D increases **quadratically**.

Design flow and layout rules

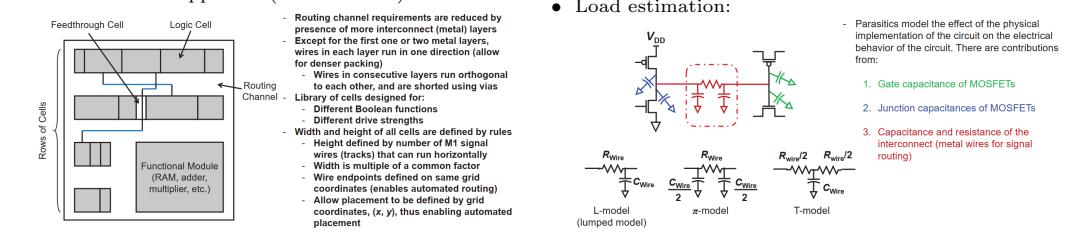
- Common design flow for integrated circuits:



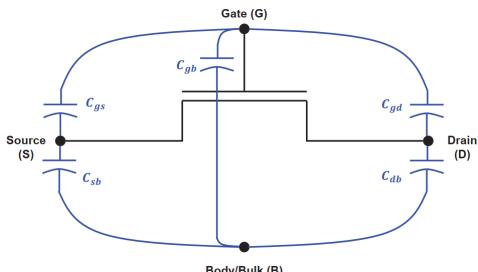
- **Array-based approach:** Larger circuits can be built by smaller circuits by configuring the connections or functions between them.
- **Cell-based approach:**
 - In the standard cell design flow, simple logic gates are designed and form a library from which more complex designs can be quickly synthesized. Therefore, it is considered cost effective because the cost of designing the library is amortized across design projects and the turn around time of implementing a design is fairly quick.
 - In the macro cell design flow, circuit macros (which are functional circuit blocks that perform specialized functions, much like a software macro) are included along side the standard cells in the library.
- Design rules in IC:
 1. Minimum spacing between metal lines
 2. Line width
 3. Transistor channel length
- The followings are IC design (layout) rules:
 - Interface between designer and process engineer
 - Guidelines for constructing process masks
 - Unit dimension: Minimum line width
 - Scalable design rules: lambda parameters
 - Absolute design rules: micron rules
 - The number of metal layers available
 - The spacing between shapes in the same layer
 - The width/length of a shape in a layer
 - The overlap between shapes in different layers
- The shapes and their geometry used to define a MOSFET only specifies the channel width and lengths, and well as the location of the gate terminal, contacts to the source/drain/gate/body terminals, and the circuit connections.
- FinFET and GAA-FET has quantized widths due to fixed height. This has significant constraints on analog designs, but not as restrictive on digital design.

CMOS Inverter The standard cell methodology

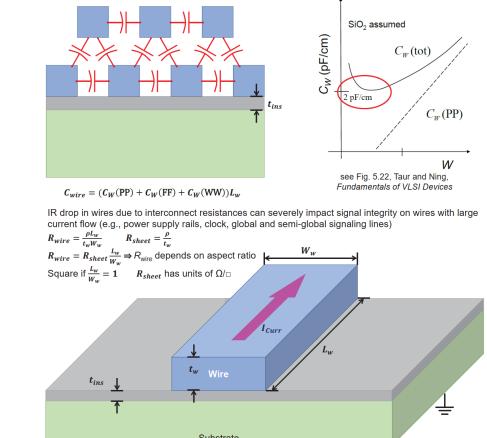
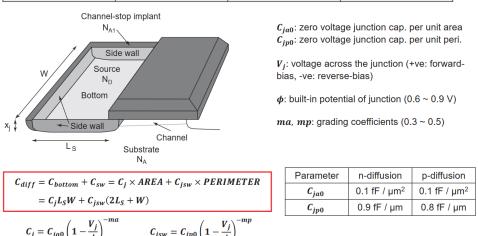
- A standard cell is a group of transistor and interconnect structures that provide a boolean logic function (e.g. AND, OR, XOR, XNOR, inverters) or a storage function (flipflop or latch). The simplest cells are direct representations of the elemental NAND, NOR, XOR boolean function, although cells of much greater complexity are commonly used (such as a 2-bit full-adder, or muxed D-input flipflop).
- A standard cell library is a collection of low-level electronic logic functions such as AND, OR, INVERT, flip-flops, latches, and buffers. These cells are realized as fixed-height, variable-width full-custom cells. The key aspect with these libraries is that they are of a fixed height, which enables them to be placed in rows, easing the process of automated digital layout.
- The cell-based approach (standard cell):
 - Routing channel requirements are reduced by presence of more interconnect (metal) layers
 - Except for the first one or two metal layers, wires in each layer run in one direction (allow for denser packing)
 - Width in successive layers runs orthogonal to each other, and are shorted using vias
 - Library of cells designed for:
 - Different Boolean functions
 - Different drive strengths
 - Width and height of all cells are defined by rules
 - Height defined by number of M1 signal wires (number of vertical metal layers)
 - Width is multiple of a common factor
 - Wire endpoints defined on same grid coordinates (automated routing)
 - Allow placement to be defined by grid coordinates, (x, y), thus enabling automated placement



- There are **5 parasitic capacitances** in a MOSFET, namely three intrinsic capacitances and two junction capacitances (one between source and body terminal and the other between drain and body terminal).



Operating Region	C_{gb}	C_{gs}	C_{gd}
Cutoff	$C_{Ox}WL_{EFF}$	0	0
Linear/Triode	0	$C_{Ox}WL_{EFF}$	$C_{Ox}WL_{EFF}$
Saturation	0	$\frac{2}{3}C_{Ox}WL_{EFF}$	0



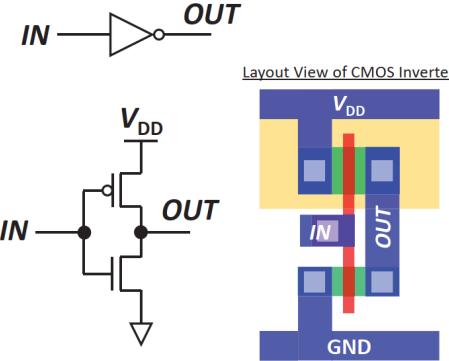
- Sheet resistance is a measure of resistance of thin films that are nominally uniform in thickness.
- The dimension of Ω/\square is actually equivalent to Ω . This is an advantage, because sheet resistance of could be taken out of context and misinterpreted as bulk resistance of 1 ohm, whereas sheet resistance of cannot thus be misinterpreted.
- The junction capacitance of the MOSFET is independent of the channel length.
- All parasitic capacitances of the MOSFET are directly proportional to the width of the MOSFET.

Static CMOS logic

- Static CMOS logic gate consists of a pull-up network and a pull-down network.
- Every input signal is given to the gates of at least one NMOS and one PMOS.
- Output voltage is very close to ideal (0 V for '0', V_{DD} for '1')
- The output responds immediately to changes to input signals.
- Works so long as sufficient power is supplied to the logic gate.
- No short circuit power dissipation at steady-state (i.e., after input and output voltages have stopped changing).
- Fan-in:** number of inputs the logic gate has.
- Fan-out:** number of inputs receiving the output of the logic gate.

Static CMOS inverter

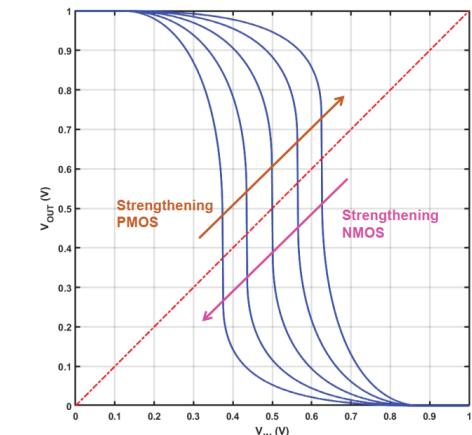
- Schematic:



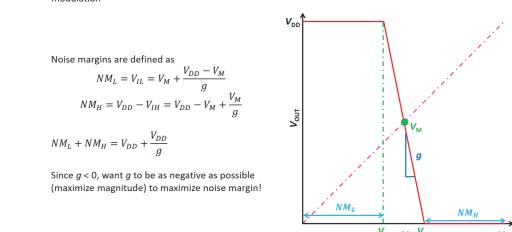
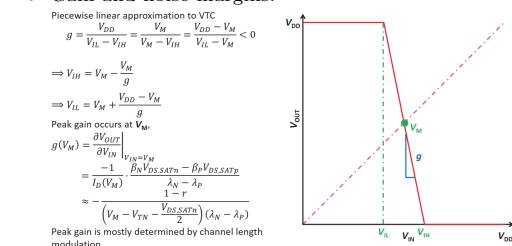
• Relationship between V_M and β -ratio:

$$V_M = \frac{\left(V_{TN} + \frac{V_{DS,SATn}}{2}\right) + r\left(V_{BD} + V_{TP} + \frac{V_{DS,SATp}}{2}\right)}{1+r}; r = \frac{-\beta_p V_{DS,SATp}}{\beta_n V_{DS,SATn}} = \frac{-\mu_p (W_p/L_p)V_{DS,SATp}}{\mu_n (W_n/L_n)V_{DS,SATn}}$$

$$\beta\text{-ratio} = \frac{\beta_p}{\beta_n} = r \text{ for when } \frac{|V_{DS,SATp}|}{|V_{DS,SATn}|} = 1$$

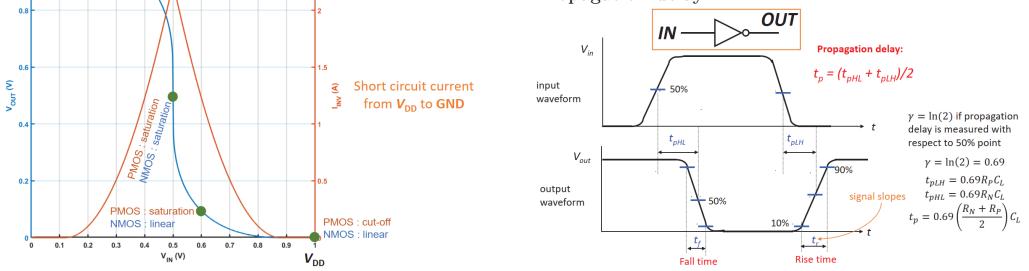


- Gain and noise margins:

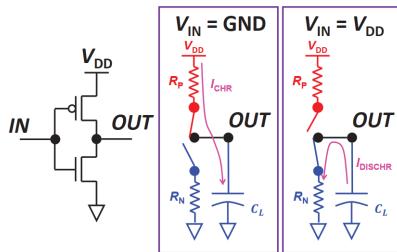


Dynamic response of static CMOS inverter

- Propagation delay:



- Dynamic response:



- During $0 \rightarrow 1$ transition:

$$V_{out} = V_{DD} (e^{-\frac{t}{R_P C_L}})$$

- During $1 \rightarrow 0$ transition:

$$V_{out} = V_{DD} (e^{1 - \frac{t}{R_N C_L}})$$

- Common design requirements:

- Balanced propagation delay
 - Design for $t_{PHL} = t_{PLH}$ (i.e., $0.69R_N C_L = 0.69R_P C_L$), achieved by setting $(R_N = R_P)$
 - Since $\frac{R_P}{R_N} \propto \frac{(W_L/L_D)}{(W_N/L_N)}$, we achieve this by adjusting the aspect ratio (W/L ratio) of the MOSFETs
- Fastest speed (i.e., smallest delay)
 - Reduce C_L
 - internal diffusion capacitance of the gate itself
 - keep the drain diffusion as small as possible
 - interconnect capacitance
 - fanout
 - Increase W/L ratio of the transistors
 - the most powerful and effective performance optimization tool in the hands of the designer
 - watch out for self-loading – when the intrinsic capacitance dominates the extrinsic load
 - may come increase input capacitance which loads the gate that generates the input signal
 - Increase V_{DD}
 - trade-off energy for performance
 - increasing V_{DD} above a certain level yields only very minimal improvements
 - reliability concerns enforce a firm upper bound on V_{DD}

Power consumption of static CMOS inverter

- The key components of power dissipation are:
- 1. **Dynamic/active power consumption**, due to charging capacitances.

- 2. **Short circuit current paths**: short circuit current between power rails during switching.
- 3. **Leakage**: Leaking diodes and MOSFETs.

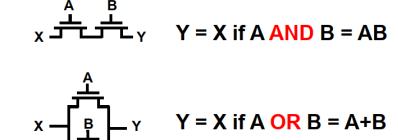
- Dynamic power dissipation:

- During $0 \rightarrow 1$ transition (output node charged to V_{DD})
 - Total supplied = $C_L V_{DD}^2$
 - Stored in load capacitor = $0.5 C_L V_{DD}^2$
 - Energy loss = $0.5 C_L V_{DD}^2$
 - During $1 \rightarrow 0$ transition (output node discharged to GND)
 - Total supplied = 0
 - Stored in load capacitor = 0
 - Energy loss = $0.5 C_L V_{DD}^2$
 - Energy is consumed by circuit (i.e. taken from power supply) only when there is $0 \rightarrow 1$ transition
 - Dynamic/active Power = rate of energy consumption = $C_L V_{DD}^2 f_{dk}$
- f_{dk} : operating frequency $\alpha_{0 \rightarrow 1}$: activity factor (probability of $0 \rightarrow 1$ transition)

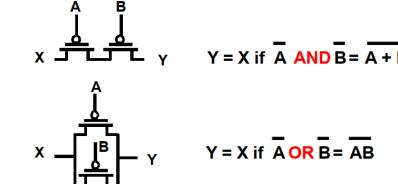
CMOS Logic Static CMOS logic gate

- To design a static CMOS logic gate, one needs to:
 - Design the topology of the logic gate (i.e., how the MOSFETs are connected in the circuit)
 - Design the sizes of the MOSFETs (i.e., the channel widths)
 - Draw the layout of the logic gate

- NMOS Transistors pass a “strong” 0, but a “weak” 1.



- PMOS Transistors pass a “strong” 1, but a “weak” 0.

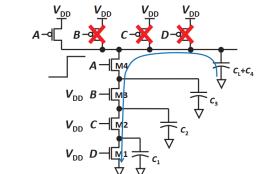


Fan-in and fan-out

- Fan-in affects number of transistors in the logic gate, which determines internal parasitic capacitances.

- Fan-out affects C_L and thus, affects delay and active power dissipation.

Generalization: N-input NAND Gate



- Compare the delay for when ABCD changes to 1110:

i. 1111

ii. 0111

Initial conditions:

i. C_1, C_2, C_3 , and C_4 are charged to V_{DD}

ii. C_1, C_2, C_3 , and C_4 are charged to GND

Final conditions:

- C_1, C_2, C_3 , and C_4 are all discharged to GND

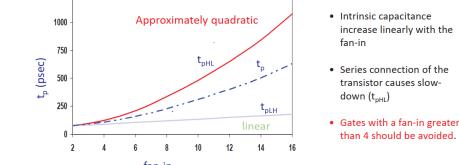
in both cases

- Signature based on Elmore delay model:

i. $t_{PLH} = R_{ON}(C_1 + C_2 + R_{OFF}) + C_1(V_{DD} + R_{ON} + R_{OFF})$

ii. $t_{PHL} = R_{ON}(C_1 + C_2 + R_{OFF}) + C_2(V_{DD} + R_{ON} + R_{OFF})$

NMOS: 0.5μm, PMOS: 1.5μm



If R_{Mj} are the same and all C_i are the same,

$$t_{d,worst} = 0.69 \left[NRC_L + RC \sum_{i=1}^N i \right] = 0.69R \left[C \left(\frac{N(N+1)}{2} \right) + NC_L \right]$$

Design techniques to reduce delay for large fan-in:

- Transistor sizing

- Only effective when fan-out dominates the delay

$$t_{d,worst} = 0.69R \left[C \left(\frac{N(N+1)}{2} \right) + NC_L \right] \approx 0.69NRC_L$$

Sizing to reduce R and delay

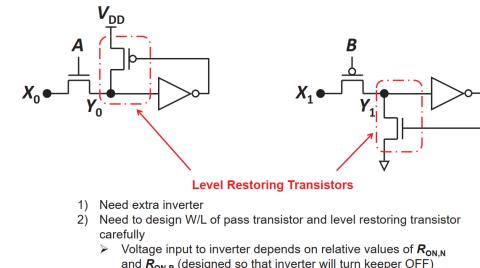
- Progressive transistor sizing

- $M_1 > M_2 > M_3 > \dots > M_N$ (distributed delay line)
- Easy to do in schematic but may be difficult in layout due to design rules

- Input reordering

- Logic restructuring

- Level restoring transistor (keeper transistor):



- 1) Need extra inverter
- 2) Need to design W/L of pass transistor and level restoring transistor carefully
 - Voltage input to inverter depends on relative values of $R_{ON,P}$ and $R_{ON,N}$ (designed so that inverter will turn keeper OFF)

CMOS Arithmetic Logic Unit Digital Processor

- The digital processor is partitioned into:
 - Control unit: Sends command to other units
 - Arithmetic Logic Unit (ALU): Perform computation
 - Memory unit: Stores data
 - Input/output: Interacts with external devices

Arithmetic Logic Unit (ALU)

- The Arithmetic Logic Unit (ALU) is the part of the CPU performs various arithmetic and logic operation:
 - Additions and subtractions
 - Multiplication and division
 - Boolean logic such as AND, OR, XOR, etc.
 - Bit-shifts
 - Bit-rotations
 - Evaluate conditions (if block)
 - Calculate addresses for program counter (but does not keep track of it)
 - Math functions (e.g. sine, exp, etc.)

Half and full adders

- Half adder:

Half-adder		
A	B	Carry (C_0)
0	0	0
0	1	0
1	0	0
1	1	1

C_0 : carry-out
Result of $A+B$: $C_0 S$

A	B	Sum (S)
0	0	0
0	1	1
1	0	1
1	1	0

$C_0 = A \cdot B$
 $S = A \oplus B$

- Full adder:

A	B	Carry-in (C_i)	Carry (C_o)	Sum (S)
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

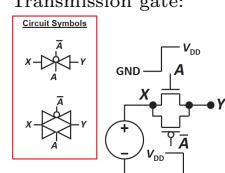
Pass transistor and transmission gate

- NMOS: Strong ‘0’, but weak ‘1’.

- PMOS: Strong ‘1’, but weak ‘0’.

- Pass transistor: use only one type of transistor, either NMOS or PMOS. Cascading in a staircase-like shape causes functionality issue, because $V_{out} = V_{DD} - N \cdot V_{TN}$. Cascading horizontally (in series) causes propagation delay, which increases quadratically with N .

- Transmission gate:



- A transmission gate is formed using a p-MOSFET and an n-MOSFET connected as shown:

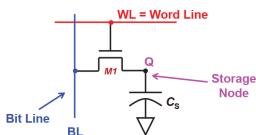
- The gates are connected to complementary signals so that both of them are either ON or OFF at the same time.
- The source terminal of the n-MOSFET is connected to the drain terminal of the p-MOSFET.
- The drain terminal of the n-MOSFET is connected to the source terminal of the p-MOSFET.
- The n-MOSFET helps to pass strong ‘0’.
- The p-MOSFET helps to pass strong ‘1’.

Need complementary signals AND twice the number of MOSFETs to pass strong ‘0’ and strong ‘1’!
Does not solve problem of delays in long chains of transmission gates!

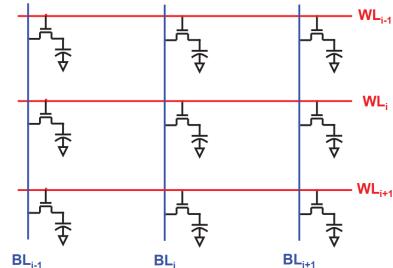
- 3. Wait until steady state
- 4. Sense $V_{BL} - V_{\overline{BL}}$
- 5. Turn off WL
- something

1-Transistor DRAM

- Schematic:



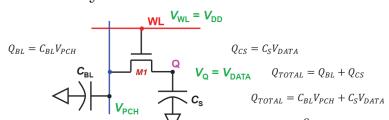
- Unlike SRAM
 - The 1T DRAM cell has only one transistor and one capacitor
 - Single-ended
- Simple structure and compact in size (good for density)
- Data is stored as the amount of charge in the storage capacitor
- Node Q is in high-Z state when $V_{WL} = GND$
 - Stored charge will leak due to transistor leakage
 - Requires dynamic refresh of charge stored to retain data (hence, dynamic RAM)



- Data is stored as a charge on the capacitor.
- Operations that can be performed on the 1T DRAM:
 - Read
 - Write
 - Hold
- Reducing the leakage currents helps to increase the time that data can be stored in the DRAM cell (also called the retention time).
- The DRAM cell is not sensitive to noise because the storage node is in the high-Z state when it is not being accessed.

DRAM read and write operations

- Write operation:
 1. Initially, $V_{WL} = GND$
 2. Charge BL
 - Logic 0: $V_{BL} = GND$
 - Logic 1: $V_{BL} = V_{DD}$
 3. Turn on WL
 4. Wait for the capacitor to be charged or discharged.
 5. Turn off WL , $V_{WL} = GND$
 6. Disconnect voltage source from BL
- Read operation:
 1. Precharge BL , so that C_{BL} is charged to V_{DD}
 2. Turn on WL
 3. Wait until steady state



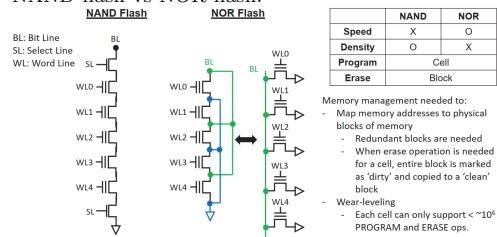
- M1 electrically shorts C_{BL} and C_s together
 - Charge sharing between C_{BL} and C_s
 - ΔV_Q and ΔV_{BL} depends on size of C_{BL} relative to C_s , and whether M1 gets turned off or not

- 4. Sense V_{BL}
- 5. Turn off WL

- The read operation in the 1T DRAM is destructive, which requires the data to be written back into the cell after every read operation.
- The reason that read operations in the 1T DRAM are destructive is that charge sharing between the bit-line capacitance and storage capacitance occurs.

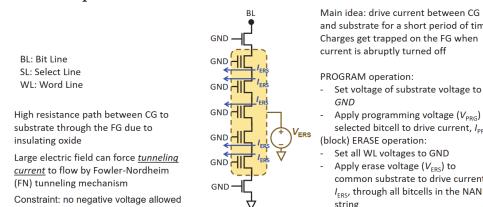
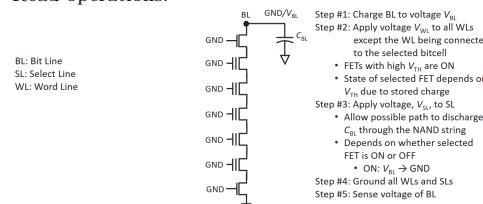
Flash memory

- NAND flash vs NOR flash:



NAND flash read and write operations

- Read operations:



- Read operations:

