### Review: Fundamental Design Metrics

- **Functionality**
- **Cost**
  - NRE (fixed) costs - design effort
  - RE (variable) costs - cost of parts, assembly, test
- **Reliability, robustness**
  - Noise margins
  - Noise immunity
- **Performance**
  - Speed (delay)
  - Power consumption; energy
- **Time-to-market**

### Design Abstraction Levels

```
SYSTEM
  └── MODULE
      └── CIRCUIT
          └── DEVICE
```

### The MOS Transistor

In the NMOS transistor cross section:
- **n** areas have been doped with donor ions (arsenic) of concentration $N_D$, electrons are the majority carriers.
- **p** areas have been doped with acceptor ions (boron) of concentration $N_A$, holes are the majority carriers.
### Switch Model of NMOS Transistor

- **Gate**: Switches on (‘1’) or off (‘0’)
- **Source** (of carriers)
- **Drain** (of carriers)
- **$|V_{GS}| < |V_T|$**: Open (off) (Gate = ‘0’)
- **$|V_{GS}| > |V_T|$**: Closed (on) (Gate = ‘1’)
- **$R_{on}$**

### Switch Model of PMOS Transistor

- **Gate**: Switches on (‘0’) or off (‘1’)
- **Source** (of carriers)
- **Drain** (of carriers)
- **$|V_{GS}| > |V_{DD} - |V_T||$**: Open (off) (Gate = ‘1’)
- **$|V_{GS}| < |V_{DD} - |V_T||$**: Closed (on) (Gate = ‘0’)
- **$R_{on}$**

### CMOS Inverter: A First Look

#### Steady State Response

- $V_{DD}$
- $V_{out} = 0$
- $V_{in} = V_{DD}$
- $R_{p}$
- $V_{out} = 1$
- $V_{in} = 0$
- $V_{OL} = 0$
- $V_{OH} = V_{DD}$
- $V_{M} = f(R_{n}, R_{p})$

### CMOS Fabrication

- CMOS transistors are fabricated on silicon wafer
- Lithography process similar to printing press
- On each step, different materials are deposited or etched
- Easiest to understand by viewing both top and cross-section of wafer in a simplified manufacturing process

### Inverter Cross-section

- Typically use p-type substrate for nMOS transistors
- Requires n-well for body of pMOS transistors
**Well and Substrate Taps**

- Substrate must be tied to GND and n-well to $V_{DD}$
- Metal to lightly-doped semiconductor forms poor connection called Shottky Diode
- Use heavily doped well and substrate contacts / taps

**Inverter Mask Set**

- Transistors and wires are defined by masks
- Cross-section taken along dashed line

---

**Detailed Mask Views**

- Six masks
  - n-well
  - Polysilicon
  - n+ diffusion
  - p+ diffusion
  - Contact
  - Metal

**Fabrication Steps**

- Start with blank wafer
- Build inverter from the bottom up
- First step will be to form the n-well
  - Cover wafer with protective layer of SiO$_2$ (oxide)
  - Remove layer where n-well should be built
  - Implant or diffuse n dopants into exposed wafer
  - Strip off SiO$_2$

**Oxidation**

- Grow SiO$_2$ on top of Si wafer
  - 900 – 1200 C with $H_2O$ or $O_2$ in oxidation furnace

**Photoresist**

- Spin on photoresist
  - Photoresist is a light-sensitive organic polymer
  - Softens where exposed to light
Lithography
• Expose photoresist through n-well mask
• Strip off exposed photoresist

Etch
• Etch oxide with hydrofluoric acid (HF)
  – Seeps through skin and eats bone, nasty stuff!!!
• Only attacks oxide where resist has been exposed

Strip Photoresist
• Strip off remaining photoresist
  – Use mixture of acids called piranah etch
• Necessary so resist doesn’t melt in next step

n-well
• n-well is formed with diffusion or ion implantation
  • Diffusion
    – Place wafer in furnace with arsenic gas
    – Heat until As atoms diffuse into exposed Si
  • Ion Implantation
    – Blast wafer with beam of As ions
    – Ions blocked by SiO₂, only enter exposed Si

Strip Oxide
• Strip off the remaining oxide using HF
• Back to bare wafer with n-well
• Subsequent steps involve similar series of steps

Polysilicon
• Deposit very thin layer of gate oxide
  – < 20 Å (6-7 atomic layers)
• Chemical Vapor Deposition (CVD) of silicon layer
  – Place wafer in furnace with Silane gas (SiH₄)
  – Forms many small crystals called polysilicon
  – Heavily doped to be good conductor
Polysilicon Patterning
- Use same lithography process to pattern polysilicon

Self-Aligned Process
- Use oxide and masking to expose where n+ dopants should be diffused or implanted
- N-diffusion forms nMOS source, drain, and n-well contact

N-diffusion
- Pattern oxide and form n+ regions
- Self-aligned process where gate blocks diffusion
- Polysilicon is better than metal for self-aligned gates because it doesn’t melt during later processing

N-diffusion cont.
- Historically dopants were diffused
- Usually ion implantation today
- But regions are still called diffusion

N-diffusion cont.
- Strip off oxide to complete patterning step

P-Diffusion
- Similar set of steps form p+ diffusion regions for pMOS source and drain and substrate contact
Contacts

- Now we need to wire together the devices
- Cover chip with thick field oxide
- Etch oxide where contact cuts are needed

Metalization

- Sputter on aluminum over whole wafer
- Pattern to remove excess metal, leaving wires

Layout

- Chips are specified with set of masks
- Minimum dimensions of masks determine transistor size (and hence speed, cost, and power)
- Feature size $f$ = distance between source and drain
  - Set by minimum width of polysilicon
- Feature size improves 30% every 3 years or so
- Normalize for feature size when describing design rules
- Express rules in terms of $\lambda = \frac{f}{2}$
  - E.g. $\lambda = 0.3 \, \mu m$ in 0.6 $\mu m$ process

Simplified Design Rules

- Conservative rules to get you started

Inverter Layout

- Transistor dimensions specified as Width / Length
  - Minimum size is $4\lambda / 2\lambda$, sometimes called 1 unit
  - In $f = 0.6 \, \mu m$ process, this is 1.2 $\mu m$ wide, 0.6 $\mu m$ long

Summary

- MOS Transistors are stack of gate, oxide, silicon
- Can be viewed as electrically controlled switches
- Build logic gates out of switches
- Draw masks to specify layout of transistors
- Now you know everything necessary to start designing schematics and layout for a simple chip!
Design Rules
- Interface between the circuit designer and process engineer
- Guidelines for constructing process masks
- Unit dimension: minimum line width
  - scalable design rules: lambda parameter
  - absolute dimensions: micron rules
- Rules constructed to ensure that design works even when small fab errors (within some tolerance) occur
- A complete set includes
  - set of layers
  - intra-layer: relations between objects in the same layer
  - inter-layer: relations between objects on different layers

Why Have Design Rules?
To be able to tolerate some level of fabrication errors such as
1. Mask misalignment
2. Dust
3. Process parameters (e.g., lateral diffusion)
4. Rough surfaces

Intra-Layer Design Rule Origins
- Minimum dimensions (e.g., widths) of objects on each layer to maintain that object after fab
  - minimum line width is set by the resolution of the patterning process (photolithography)
- Minimum spaces between objects (that are not related) on the same layer to ensure they will not short after fab

Intra-Layer Design Rules
- Same Potential
- Different Potential

Inter-Layer Design Rule Origins
1. Transistor rules – transistor formed by overlap of active and poly layers

Inter-Layer Design Rules
- Transistors
- Unrelated Poly & Diffusion

Transistor Layout
Select Layer

Inter-Layer Design Rule Origins, Con't

2. Contact and via rules

- M1 contact to p-diffusion
- M1 contact to n-diffusion
- M1 contact to poly
- Mx contact to My

Contact Mask

Via Masks

Both materials

Contact: 0.44 x 0.44

mask misaligned

CMOS Process Layers

<table>
<thead>
<tr>
<th>Mask/Layer name</th>
<th>Generation from drawn layers</th>
<th>Alternative names for mask/layer</th>
<th>MOSIS mask label</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-well</td>
<td>nwell</td>
<td>bulk, substrate, tub, n-tub, n+</td>
<td>CWN</td>
</tr>
<tr>
<td>p-well</td>
<td>pwell</td>
<td>bulk, substrate, tub, p-tub, p+</td>
<td>CPW</td>
</tr>
<tr>
<td>active</td>
<td>active</td>
<td>active, tub, trench, island, gate oxide</td>
<td>GOD</td>
</tr>
<tr>
<td>poly/diffusion</td>
<td>poly</td>
<td>poly, gate</td>
<td>CPG</td>
</tr>
<tr>
<td>p-diffusion implant</td>
<td>pimplant</td>
<td>n-select, p+</td>
<td>CWN</td>
</tr>
<tr>
<td>n-diffusion implant</td>
<td>nimplant</td>
<td>p-select, n+</td>
<td>CSN</td>
</tr>
<tr>
<td>contact</td>
<td>contact</td>
<td>contact, poly contact, diffusion contact</td>
<td>CXP and CCA</td>
</tr>
<tr>
<td>metal1</td>
<td>m1</td>
<td>first-level metal</td>
<td>CMP</td>
</tr>
<tr>
<td>metal2</td>
<td>m2</td>
<td>second-level metal</td>
<td>CMS</td>
</tr>
<tr>
<td>via2</td>
<td>via2</td>
<td>metal2/metal2 via</td>
<td>CVS</td>
</tr>
<tr>
<td>metal3</td>
<td>m3</td>
<td>third-level metal</td>
<td>GBF</td>
</tr>
<tr>
<td>glass</td>
<td>glass</td>
<td>passivation, overglass, pass</td>
<td>COG</td>
</tr>
</tbody>
</table>

Vias and Contacts

CMOS Transistor Theory

Outline

- Introduction
- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- Gate and Diffusion Capacitance
- Pass Transistors
- RC Delay Models
Introduction

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - \[ I = C \left( \frac{\Delta V}{\Delta t} \right) \] -> \[ \Delta t = \frac{C}{I} \Delta V \]
  - Capacitance and current determine speed
- Also explore what a "degraded level" really means

MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
  - Accumulation
  - Depletion
  - Inversion

Terminal Voltages

- Mode of operation depends on \( V_{gs}, V_{gd}, V_{ds} \)
  - \( V_{gs} = V_g - V_s \)
  - \( V_{gd} = V_g - V_d \)
  - \( V_{ds} = V_d - V_s = V_{gs} - V_{gd} \)
- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence \( V_{ds} \geq 0 \)
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - Cutoff
  - Linear
  - Saturation

nMOS Cutoff

- No channel
- \( I_{ds} = 0 \)

nMOS Linear

- Channel forms
- Current flows from d to s
  - \( e^- \) from s to d
- \( I_{ds} \) increases with \( V_{ds} \)
- Similar to linear resistor

nMOS Saturation

- Channel pinches off
- \( I_{ds} \) independent of \( V_{ds} \)
- We say current saturates
- Similar to current source
I-V Characteristics

In Linear region, \( I_{ds} \) depends on
- How much charge is in the channel?
- How fast is the charge moving?

Channel Charge

MOS structure looks like parallel plate capacitor while operating in inversion
- Gate – oxide – channel
- \( Q_{channel} = CV \)

\[ C = C_g = \frac{\varepsilon_{ox} WL}{t_{ox}} = C_{ox}WL \]
\[ C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \]
\[ V = V_{go} - V_t = (V_{gs} - V_{ds}/2) - V_t \]

Carrier velocity

Charge is carried by e-
- Carrier velocity \( v \) proportional to lateral E-field between source and drain
- \( v = \)
Carrier velocity

- Charge is carried by e-
- Carrier velocity $v$ proportional to lateral E-field between source and drain
  - $v = \mu E$
  - $E = \frac{V_{ds}}{L}$

- Time for carrier to cross channel:
  - $t = \frac{L}{v}$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{\text{channel}}$ is in the channel
  - How much time $t$ each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$= \mu C_m \frac{W}{L} \left( V_{gS} - V_t - \frac{V_{ds}}{2} \right) V_{ds}$$

$$= \beta \left( V_{gS} - V_t - \frac{V_{ds}}{2} \right) V_{ds}$$

$$\beta = \mu C_m \frac{W}{L}$$
nMOS Saturation I-V

- If $V_{gs} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta \left( V_{gs} - V_t - \frac{V_{dsat}}{2} \right) V_{dsat}$$

nMOS I-V Summary

- Shockley 1st order transistor models

$$I_{ds} = \begin{cases} 
0 & V_{gs} < V_t \quad \text{cutoff} \\
\beta \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} & V_{gs} < V_{dsat} \quad \text{linear} \\
\frac{\beta}{2} (V_{gs} - V_t)^2 & V_{ds} > V_{dsat} \quad \text{saturation}
\end{cases}$$

Example

- For a 0.6 $\mu$m process
  - From AMI Semiconductor
    - $t_{ox} = 100$ Å
    - $\mu = 350$ cm$^2$/V·s
    - $V_t = 0.7$ V
- Plot $I_{ds}$ vs. $V_{ds}$
  - $V_{gs}$ = 0, 1, 2, 3, 4, 5
  - Use W/L = 4/2

$$\beta = \mu C_V \frac{W}{L} = \left( 350 \text{ cm}^2/\text{V·s} \right) \left( \frac{W}{L} \text{ cm} \right) = 120\beta \mu C_V$$

pMOS I-V

- All dopings and voltages are inverted for pMOS
- Mobility $\mu_p$ is determined by holes
  - Typically 2-3x lower than that of electrons $\mu_n$
  - 120 cm$^2$/V·s in AMI 0.6 $\mu$m process
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$
  - *** plot I-V here
Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
  - Creates channel charge necessary for operation
- Source and drain have capacitance to body
  - Across reverse-biased diodes
  - Called diffusion capacitance because it is associated with source/drain diffusion

Gate Capacitance

- Approximate channel as connected to source
- \( C_{gs} = \frac{\varepsilon_{ox}WL}{t_{ox}} = C_{ox}WL = C_{permicro}W \)
- \( C_{permicro} \) is typically about 2 fF/\( \mu m \)

Diffusion Capacitance

- \( C_{sb}, C_{db} \)
- Undesirable, called parasitic capacitance
- Capacitance depends on area and perimeter
  - Use small diffusion nodes
  - Comparable to \( C_g \)
    - for contacted diff
  - \( \frac{1}{2} C_g \) for uncontacted
  - Varies with process