Announcement

Next class: Thursday, 13. March
Room 308 CCE
Optional Topics

Please, choose 2 out of the following 4 topics
Final date: 14th of March

1. Future trends in VLSI design
2. Basics of Hspice-Simulations
3. Effects in nanometer CMOS circuits
4. Reliability problems in current and future designs
Goals

- Where do we find Integrated Circuits?
- History and Trends
- CMOS: basic ideas
- Logic gates
- Delay estimation
- Sizing
Where do we find chips?

“Computer are the workhorses of the semiconductors industry.”

Motivation
- Performance
- Flexibility
- Mobility

- ~ 2 % are processors
- ~ 6.5 Billion processors per year
- ~ 40 % of all parts are used in the PC area

Source: WSTS ‘02
Scenarios

- Obviously tasks
- High performance demands
- Fast execution
Scenarios cont’d

- Hidden helper
- Low performance demands
History

- 1906 – Semiconductors used to detect radio signals
- 1925 – FET concept patent by J. Lilienfeld
- 1941 – Z3 by Konrad Zuse – first computer
- 1946 – ENIAC – first electronic computer
- 1947 – Transistor “Invented”
  - AT&T ignores Lilienfeld
  - Bardeen, Brattain and Schockley, AT&T, Nobel Prize in 1956
- 1958 – Integrated Circuit
  - Kilby & Noyce (died 1990)
  - Kilby - Noble Prize in 2000
- 1960 - MOSFET manufactured and patented
- 1963 - CMOS logic invented
  - Resistors replaced by transistors
History cont’d

Zuse Z3 – First computer* (1941)

- First working programmable, fully automatic computing machine
- 2,000 Relays
- Clock frequency of ~5 - 10 Hz
- Word length of 22 bits
- Programmed by punched film stock
- Addition, Multiplication, Division, Square root

History cont’d

ENIAC – First electronic computer (1946)

- Electronic Numerical Integrator And Computer
- At Moore School of Electrical Engineering, University of Pennsylvania
- 17,468 vacuum tubes, 7,200 diodes (+ ca. 80k resistors & capacitors)
- 5 Million hand-soldered joints
History cont’d

Vacuum Tubes in ENIAC
History cont’d

(a) First transistor (1947, Bardeen & Brattain, Bell labs)
(b) First integrated circuit (1958, Kilby, AT&T)

Source: Weste, “CMOS VLSI design”, 2003
Moore’s Law

- Prediction by Gordon Moore in 1965
- Semiconductor technology will double its effectiveness every 18 months

Source: Moore, 1665
Moore’s Law cont’d

Source: Moore, ISSCC 2003
Trend: Cost per function

Source: VLSI Research

Price of a transistor


Price DRAM

Price All Semi
Trend: Performance

Source: Moore, ISSCC 2003
Trend: Power

Source: Moore, ISSCC 2003
Trend: Power Density

Source: Moore, ISSCC 2003
Dimensions

Source: „Spektrum der Wissenschaften“

„65 nm“-Transistor
Source: Intel
The CMOS Technology

- CMOS = Complementary Metal Oxide Semiconductor

- Currently most applied logic family

- Main advantages:
  - Low Power (compared to other technologies)
  - Very good scalability
  - High Speed
  - High packaging density
The CMOS Technique cont’d

- Main Idea:
  - Combination of two complementary switches
  - Switches are metal-oxide-semiconductor field-effect transistors (MOSFET)
  - Realization of logic gates (AND, NAND, …)

- “Metal–Oxide–Semiconductor“:
  - Physical structure of MOSFETs (metal gate electrode, oxide insulator, semiconductor material)
  - Today: polysilicon instead of metal
What is a transistor?

A MOS Transistor \[ \rightarrow \] A Switch!

PMOS and NMOS

@NMOS: Body is (commonly) tied to ground (0 V)
@PMOS: Body is (commonly) tied to VDD

NMOS-Transistor (2)

Cross section of NMOS and PMOS

Source: Weste, “CMOS VLSI design”, 2003
Layout Mask Set

- Transistors and wires are defined by *masks*
- Cross-section taken along dashed line
I-V Curves of NMOS

Source: Weste, "CMOS VLSI design", 2003
Threshold Voltage $V_{th}$

- Transistor characteristic
- If: „Gate-Source“-Voltage $V_{gs}$ **higher** than $V_{th}$
  - Channel under Gate
  - Current between Drain and Source
- If: $V_{gs}$ **lower** than $V_{th}$
  - No current
Logic Gates

- Task (e.g. calculation)
- Transfer into Logic Gates (Synthesis)
- Gate characteristics:
  - Delay
  - Power dissipation
  - more ...
- Gates realized by transistors
  - Transistors determine gate characteristics

\[ Y = A + B \]
Example: Half-adder

- How do you add the two bits $A_0$ and $B_0$ in binary logic?

- So called Half-adder:

<table>
<thead>
<tr>
<th>$A_0$</th>
<th>$B_0$</th>
<th>Result</th>
<th>Carry</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>01</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>01</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- $\text{In1 (}A_0\text{)}$ | $\text{In2 (}B_0\text{)}$ | AND | XOR |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
CMOS Scheme

VDD (supply voltage)

PUN – Pull-up Network

PDN – Pull-down Network

IN1

... INx

OUT

GND (ground)
CMOS Inverter

<table>
<thead>
<tr>
<th>IN1</th>
<th>OUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (GND)</td>
<td>1 (VDD)</td>
</tr>
<tr>
<td>1 (VDD)</td>
<td>0 (GND)</td>
</tr>
</tbody>
</table>
Transistor as Water-tap
Transistor as Water-tap cont’d

Voltage (Volt, V) ↔ Water pressure (bar)
Current (Ampere, A) ↔ Water quantity (liter)

Source: Timmernann, 2007
NAND Gate

<table>
<thead>
<tr>
<th>In1</th>
<th>In2</th>
<th>PUN</th>
<th>PDN</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>OFF</td>
<td>ON</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>ON</td>
<td>OFF</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>ON</td>
<td>OFF</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>ON</td>
<td>ON</td>
<td>1</td>
</tr>
</tbody>
</table>
NOR Gate

<table>
<thead>
<tr>
<th>In1</th>
<th>In2</th>
<th>PUN</th>
<th>PDN</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>OFF</td>
<td>ON</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>OFF</td>
<td>ON</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>OFF</td>
<td>ON</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>ON</td>
<td>OFF</td>
<td>1</td>
</tr>
</tbody>
</table>
AND and OR Gate

<table>
<thead>
<tr>
<th>In1</th>
<th>In2</th>
<th>Out\text{_AND}</th>
<th>Out\text{_NAND}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In1</th>
<th>In2</th>
<th>Out\text{_OR}</th>
<th>Out\text{_NOR}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Delay Definitions

- **Input waveform**:
  - $V_{in}$
  - $V_{in}$ waveform 50% level

- **Output waveform**:
  - $V_{out}$
  - $V_{out}$ waveform 50% level

- **Signal slopes**:
  - $t_{r}$
  - $t_{pHL}$
  - $t_{pLH}$
  - 90% and 10% levels

- **Propagation delay** $t_p$
RC-Delay Model

- Simple but effective delay model
- Use equivalent circuits for MOS transistors
  - Ideal switch
  - Transistor capacitances
  - ON resistance (when transistor is conducting (=ON) ➔ channel between Drain to Source acts as resistor)
- Delay $t \sim R \cdot C$
MOSFET capacitances

- Any two conductors separated by an insulator create a capacitor
- MOS capacitances have three origins:
  - The basic MOS structure
  - The channel charge
  - The pn-junctions depletion regions

![Diagram of MOSFET capacitances](image)
RC-Delay Model: Inverter

Rising Slope

\[ C_{P,\text{gate}} \quad R_{P,\text{DS}} \quad C_{N,\text{gate}} \quad C_{\text{out}} \]
RC-Delay Model: Inverter

Falling Slope

\[ C_{P,\text{gate}} \]
\[ R_{N,DS} \]
\[ C_{N,\text{gate}} \]
\[ C_{\text{out}} \]
Where does \( C_{\text{out}} \) come from?

- **Input capacitance** (= gate capacitances) of following gate
- **Diffusion capacitances** (Drain-Bulk) of PMOS- and NMOS transistors
Gate width $W$ can be changed by Designer ($L$, $T_{ox}$, $V_{DD}$... are fixed)

- Capacitance proportional to width: $C \sim W$
- Resistance inversely proportional to width: $R \sim 1/W$
- Resistance of NMOS approx. two times smaller than PMOS with same width:

\[
\begin{align*}
W_P &= W_N \\
R_P &= 2R_N \\
W_P &= 2W_N \\
R_P &= R_N \Rightarrow C_P = 2C_N!
\end{align*}
\]
RC-Delay Model: Fanout

\[ \text{fanout} : f = \frac{C_{load}}{C_{in}} \]
RC-Delay Model: Rising Slope

\[ R_{N,DS} = \frac{R}{W_N}, \quad R_{P,DS} = \frac{2R}{W_P} \]

\[ C_{N,DB} = C \cdot W_N, \quad C_{N,gate} = C \cdot W_N \]

\[ C_{P,DB} = C \cdot W_P, \quad C_{P,gate} = C \cdot W_P \]

\[ t = RC = R_{P,DS} \cdot \left( C_{N,DB} + C_{P,DB} + C_{load} \right) \]

\[ = \frac{2R}{W_P} \cdot \left( C \cdot W_N + C \cdot W_P + f \cdot C_{in} \right) \]

\[ = \frac{2R}{2n} \left( nC + 2nC + 3nfC \right) \]

\[ = 3(1 + f) \cdot R \cdot C \]
RC-Delay Model: Falling Slope

\[ R_{N,DS} = \frac{R}{W_N}, \quad R_{P,DS} = \frac{2R}{W_P} \]

\[ C_{N,DB} = C_{N} \cdot W_N, \quad C_{N,gate} = C_{N} \cdot W_N \]

\[ C_{P,DB} = C_{P} \cdot W_P, \quad C_{P,gate} = C_{P} \cdot W_P \]

\[ t = RC = R_{N,DS} \cdot \left( C_{P,DB} + C_{N,DB} + C_{\text{load}} \right) \]

\[ = \frac{R}{W_N} \cdot \left( C_{N} \cdot W_P + C_{P} \cdot W_N + f \cdot C_{in} \right) \]

\[ = \frac{R}{n} \left( 2nC_{N} + nC_{P} + 3nfC_{in} \right) \]

\[ = 3(1 + f) \cdot R \cdot C_{\text{in}} \]
RC-Delay Model: Examples

- Delay of an Inverter with a fanout of 64:

\[
\begin{align*}
  t &= 3(1 + f) \cdot R \cdot C \\
  &= 3(1 + 64) \cdot R \cdot C \\
  &= 195 \cdot R \cdot C
\end{align*}
\]
RC-Delay Model: Examples cont’d

- Chain of Inverters with $C_{load} = 192 \, C_\square$ and $C_{in} = 3 \, C_\square$

\[
f_{chain} = \frac{C_{load,chain}}{C_{in,chain}} = 64 = f_{INV_3} \cdot f_{INV_2} \cdot f_{INV_1}
\]

\[
= \frac{C_{load,INV_3}}{C_{in,INV_3}} \cdot \frac{C_{load,INV_2}}{C_{in,INV_2}} \cdot \frac{C_{load,INV_1}}{C_{in,INV_1}} = \frac{C_{load,chain}}{C_{in,chain}} \cdot \frac{C_{in,INV_3}}{C_{in,INV_2}} \cdot \frac{C_{in,INV_2}}{C_{in,chain}}
\]

\[
t_{chain} = t_{INV_1} + t_{INV_2} + t_{INV_3}
\]

\[
= 3 \cdot R_{\square} \cdot C_{\square} \left[ (1 + f_{INV_1}) + (1 + f_{INV_2}) + (1 + f_{INV_3}) \right]
\]
RC-Delay Model: Examples cont’d

- Chain of Inverters with $C_{\text{load}} = 192 \ C \square$ and $C_{\text{in}} = 3 \ C \square$

\[ f_{\text{INV}_1} = 1, \ f_{\text{INV}_2} = 1, \ f_{\text{INV}_3} = 64 \]
\[ t_{\text{chain}1,1,64} = 207 \cdot R \ C \square \]

\[ f_{\text{INV}_1} = 4, \ f_{\text{INV}_2} = 4, \ f_{\text{INV}_3} = 4 \]
\[ t_{\text{chain}4,4,4} = 45 \cdot R \ C \square \]

Chain of Inverters: Optimum result (for speed) at equal fanout!
# Chains of Inverters

<table>
<thead>
<tr>
<th>Circuit</th>
<th># of stages</th>
<th>f</th>
<th>delay/stage</th>
<th>total delay</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /> 64</td>
<td>1</td>
<td>64</td>
<td>195</td>
<td>195</td>
</tr>
<tr>
<td><img src="image2" alt="Diagram" /> 64</td>
<td>2</td>
<td>8</td>
<td>27</td>
<td>54</td>
</tr>
<tr>
<td><img src="image3" alt="Diagram" /> 64</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td><img src="image4" alt="Diagram" /> 64</td>
<td>4</td>
<td>2.8</td>
<td>11.4</td>
<td>45.9</td>
</tr>
</tbody>
</table>
Sizing

- Increasing Width
  - Resistance get down
  - Increasing current
  - Decreasing delay

- BUT
  - Capacitance increase too
  - Internal capacitances increase
  - Output load of previous gates increases
Sizing for Performance

Any sizing factor that is sufficiently larger than $\alpha$ will give only minimal performance gains at substantial area costs.

Source: Irwan, PSU, 2001
Alpha Power Law Model

\[ t_{\text{rise}} = \frac{k' \cdot C_L \cdot V_{DD}}{(W_{PMOS} / L) \cdot (V_{DD} - V_{TH,PMOS})^\alpha} \]

\[ t_{\text{fall}} = \frac{k' \cdot C_L \cdot V_{DD}}{(W_{NMOS} / L) \cdot (V_{DD} - V_{TH,NMOS})^\alpha} \]
**Logical Effort**

**DEF:** Logical effort is the ratio of the input capacitance of a gate to the input capacitance of an inverter delivering the same output current.

- Measured from delay vs. fanout plots of simulated or measured gates
- Or estimated, counting capacitance in units of transistor width:

![Logic Circuit Diagrams](Image)

**Inverter:**
- $C_{in} = 3$
- $g = 1$ (def)

**NAND2:**
- $C_{in} = 4$
- $g = 4/3$

**NOR2:**
- $C_{in} = 5$
- $g = 5/3$

Source: Harris ‘05
Logical Effort (LE) cont’d

\[
gain = \frac{C_{out}}{C_{in}} \times LE = f \times LE
\]

LE of the whole circuit:

\[
LE_{sum} = \prod_{i} \text{LE}_i
\]

fanout of the whole circuit:

\[
f_{sum} = \frac{C_{out}}{C_{in,firstGate}}
\]

gain of the whole circuit:

\[
gain_{sum} = LE_{sum} \times f_{sum}
\]
Logical Effort (LE) cont’d

\[ \text{gain}_{\text{sum}} = \sqrt{\text{gain}_{\text{sum}}} \]

⇒ for every gate (starting at the last gate):

\[ C_{\text{out},\text{gate}} = \frac{\text{gain}_{\text{gate}} \cdot C_{\text{in},\text{gate}-1}}{\text{LE}_{\text{gate}}} \]