15. Deep Submicron Issues

J. A. Abraham
Department of Electrical and Computer Engineering
The University of Texas at Austin
EE 360R – Computer-Aided IC Design
Fall 2011
October 19, 2011

Ideal Transistor I-V

Shockley first-order transistor models

\[ I_{ds} = \begin{cases} 
0 & V_{gs} < V_t \\
\beta \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{dsat} \\
\frac{\beta}{2} (V_{gs} - V_t)^2 & V_{ds} > V_{dsat}
\end{cases} \]

cutoff
linear
saturation
Ideal nMOS I-V Plot

180 nm TSMC process

- Ideal Models
  - $\beta = 155(W/L) \ \mu A/V^2$
  - $V_t = 0.4 \ V$
  - $V_{DD} = 1.8 \ V$

Simulated nMOS I-V Plot

180 nm TSMC process
BSIM3 3V3 SPICE models

- What differs?
  - Less ON current
  - No square law
  - Current increases in saturation
**Velocity Saturation**

- We assumed carrier velocity $\propto$ E-field
  - $\nu = \mu E_{lat} = \mu V_{ds}/L$
- Carriers scatter off atoms
- Velocity reaches $\nu_{sat}$
  - Electrons: $6 - 10 \times 10^6$ cm/s
  - Holes: $4 - 8 \times 10^6$ cm/s
- Better model
  - $\nu = \frac{\mu E_{lat}}{1 + \frac{E_{lat}}{E_{sat}}} \Rightarrow \nu_{sat} = \mu E_{sat}$

**Velocity Saturation I-V Effect**

- Ideal transistor ON current increases with $V_{DD}^2$
  - $I_{ds} = \mu C_{ox} \frac{W}{L} \frac{(V_{gs} - V_t)^2}{2} = \frac{\beta}{2} (V_{gs} - V_t)^2$
- Velocity-saturated ON current increases with $V_{DD}$
  - $I_{ds} = C_{ox} W (V_{gs} - V_t) \nu_{max}$
- Real transistors are partially velocity saturated
  - Approximate with $\alpha$-power law model
  - $I_{ds} \propto V_{DD}^{\alpha}$
  - $1 < \alpha < 2$ determined empirically
**α-Power Model**

\[
I_{ds} = \begin{cases} 
0 & V_{gs} < V_{t} \\
I_{dsat} \frac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \\
I_{dsat} & V_{ds} > V_{dsat}
\end{cases}
\]

- \(I_{dsat} = P_c \beta \frac{V_{gs} - V_{t}}{2} \alpha\)
- \(V_{dsat} = P_\nu (V_{gs} - V_{t})^{\alpha/2}\)

α, β, \(P_c\), and \(P_\nu\) are parameters determined empirically from a curve-fit of I-V characteristics.

---

**Channel Length Modulation**

- Reverse-biased p-n junctions form a **depletion region**
  - Region between n and p with no carriers
  - Width of depletion \(L_d\) region grows with reverse bias
  - \(L_{eff} = L - L_d\)
- Shorter \(L_{eff}\) = more current
  - \(I_{ds}\) increases with \(V_{ds}\)
  - Even in saturation

- [Diagram of depletion region and channel length modulation](image)
Channel Length Modulation I-V

\[ I_{ds} = \frac{\beta}{2} (V_{gs} - V_t)^2 (1 + \lambda V_{ds}) \]

\[ \lambda = \text{channel length modulation coefficient} \]
- Not feature size
- Empirically fit to I-V characteristics

Body Effect

- \( V_t \): gate voltage necessary to invert channel
- Increases if source voltage increases because source is connected to the channel
- Increase in \( V_t \) with \( V_s \) is called the body effect

Body Effect Model

\[ V_t = V_{t0} = \gamma \left( \sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right) \]

\( \phi_s \) = surface potential at threshold
- Depends on doping level \( N_A \)
- As well as intrinsic carrier concentration \( n_i \)

\( \gamma \) = body effect coefficient

\[ \gamma = \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2q\varepsilon Si N_A} \]

\[ \frac{\sqrt{2q\varepsilon Si N_A}}{C_{ox}} \]
OFF Transistor Behavior

- What about current in cutoff?
- Simulated results don’t match measurements
- What differs?
  - Current doesn’t go to 0 in cutoff

Leakage Sources

- Subthreshold conduction
  - Transistors can’t abruptly turn ON or OFF
- Junction leakage
  - Reverse-biased PN junction diode current
- Gate leakage
  - Tunneling through ultrathin gate dielectric
- Subthreshold leakage is the biggest source of leakage in modern transistors

Subthreshold Leakage

- Subthreshold leakage is exponential with $V_{gs}$
  
  $$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_t}{nV_T}} \left( 1 - e^{-\frac{V_{ds}}{V_T}} \right), \quad I_{ds0} = \beta V_T^2 e^{1.8}$$

- $n$ is process dependent, typically 1.4 – 1.5
Other Leakage Sources

Drain-Induced Barrier Lowering (DIBL)
- Drain Voltage also affects $V_t$ ($V'_t = V_t - \eta V_{ds}$)
- High drain voltage causes subthreshold leakage to increase

Junction Leakage
- Reverse-biased p-n junctions have some leakage
  $$I_D = I_S \left( \frac{V_D}{e^{VT} - 1} \right)$$
- $I_S$ depends on doping levels
  - As well as area and perimeter of diffusion regions
  - Typically $< 1 \text{ fA/\mu m}^2$

Gate Leakage
- Carriers may tunnel through very thin gate oxides
- Negligible for older processes
- Becoming critically important for nanoscale transistors
- However, use of metal gates and rare-earth dielectrics (Hf) may reduce this significantly

Predicted tunneling current (from Song, 2001)
Temperature Sensitivity

- Increasing temperature
  - Reduces mobility
  - Reduces $V_t$
- $I_{ON}$ decreases with temperature
- $I_{OFF}$ increases with temperature

So What?

- So what if transistors are not ideal?
  - They still behave like switches, and isn’t that enough for digital logic?
- But these effects matter for . . .
  - Supply voltage choice
  - Logical effort
  - Quiescent power consumption
  - Pass transistors
  - Temperature of operation
Parameter Variations

- Transistors have uncertainty in parameters
  - Process: $L_{eff}$, $V_t$, $t_{ox}$ of nMOS and pMOS
  - Vary around typical (T) values

- Fast (F)
  - $L_{eff}$: short
  - $V_t$: low
  - $t_{ox}$: thin

- Slow (S): opposite

- Not all parameters are independent for nMOS and pMOS

Environmental Variation

- $V_{DD}$ and $T$ also vary in time and space
- Fast:
  - $V_{DD}$: high
  - $T$: low

<table>
<thead>
<tr>
<th>Corner</th>
<th>Voltage</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1.98</td>
<td>0°C</td>
</tr>
<tr>
<td>T</td>
<td>1.8</td>
<td>70°C</td>
</tr>
<tr>
<td>S</td>
<td>1.62</td>
<td>125°C</td>
</tr>
</tbody>
</table>
Process Corners

- Process corners describe worst case variations
  - If a design works in all corners, it will probably work for any variation
- Describe corner with four letters (T, F, S)
  - nMOS speed
  - pMOS speed
  - Voltage
  - Temperature

### Important Corners

Some critical simulation corners include

<table>
<thead>
<tr>
<th>Purpose</th>
<th>nMOS</th>
<th>pMOS</th>
<th>$V_{DD}$</th>
<th>Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle time</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Power</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Subthreshold leakage</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>Pseudo-nMOS</td>
<td>S</td>
<td>F</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Causes of Variations?

**Minimum Feature Size**

<table>
<thead>
<tr>
<th>Feature Size (microns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.01</td>
</tr>
<tr>
<td>0.01</td>
</tr>
</tbody>
</table>

- Human hair, 100 μm
- Amoeba, 15 μm
- Red blood cell, 7 μm
- AIDS virus, 0.1 μm
- Buckyball, 0.001 μm
Features Smaller than Wavelengths

What is drawn is not what is printed on silicon

Source: Raul Camposano, Synopsys

Random Dopant Fluctuations
Dynamic Temperature Variations

Thermal Map – 1.5 GHz Itanium Chip

Dynamic Voltage and Power Variations

Voltage variations

Power variations