February 21, 2024, TomsHardware: Intel announces new roadmap at IFS Direct Connect 2024: New 14A node, Clearwater Forest taped-in, five nodes in four years remains on track.

Intel unveiled a new roadmap that includes a new 14A node, the industry’s first to use High-NA EUV, here at its Intel Foundry Services (IFS) Direct Connect 2024 event. IFS also added new extensions to its existing lineup of process nodes and announced that it is now producing wafers of its Clearwater Forest processors on its 18A node.

Announcements

• Lab 4b slightly delayed
• Homework 2 due next week
MOS Transistor and Gate Delay Models
Modeling Goals

• Models that traverse design hierarchy
• Start with transistor models
• Gate delay models
• Use models to time the design
• Modeling variability

• Based on 251A, approach
  • Start simple
  • Increase accuracy, when needed
Device Models

• Transistor models
  • I-V characteristics
  • C-V characteristics

• Interconnect models
  • R, C, L
  • Covered in EE240A
Transistor Modeling

• Different levels:
  • Hand analysis
  • Computer-aided analysis (e.g. Matlab, Python, Excel,…)
  • Switch-level simulation (some flavors of ‘fast Spice’)
  • Circuit simulation (Hspice)

• These levels have different requirements in complexity, accuracy and speed of computation

• We are primarily interested in delay and energy modeling, rather than current modeling

• But we have to start from the currents…
Transistor Modeling

• **DC**
  - Accurate I-V equations
  - Well behaved conductance for convergence (not necessarily accurate)

• **Transient**
  - Accurate I-V and Q-V equations
  - Accurate first derivatives for convergence
  - Conductance, as in DC

• **Physical vs. empirical**

---

from BSIM group
Transistor I-V Modeling

• BSIM
  • Superthreshold and subthreshold models
  • Need smoothening between two regions

• EKV/PSP
  • One continuous model based on channel surface potential
Long-Channel MOS On-Current
Start with the basics:

\[ I_{DS} = W C_{ox} (V_{GS} - V_{Th} - V_C(x)) \mu E(x) \]

\( Q'(x) \): Charge density in channel at \( x \)

\( v \): Velocity at \( x \)
MOS I-V (BSIM)

Start with the basics:

\[ I_{DS} = WC_{ox}(V_{GS} - V_{Th} - V_C(x)) \mu E(x) \]

\[ I_{DS} = WC_{ox}(V_{GS} - V_{Th} - V_C(x)) \mu (dV_C(x)/dx) \]

• When integrated over the channel:

\[ I_{DS} = \frac{W}{L} \mu C_{ox} \left( V_{GS} - V_{Th} - \frac{V_{DS}}{2} \right) V_{DS} \]

Transistor saturates when \( V_{GD} = V_{Th} \) - the channel pinches off at drain’s side.

\[ I_{DS} = \frac{W}{2L} \mu C_{ox} (V_{GS} - V_{Th})^2 \]
MOS Currents (32nm CMOS with $L \gg 1 \mu m$)

Currents according to the quadratic model
Correct for long channel devices ($L \sim \mu m$)
Simulated 32nm Transistor

$L = 32\text{nm}$
Simulation vs. Model

Major discrepancies:
- shape
- saturation points
- output resistances
Velocity Saturation
Velocity Saturation

\[ E_C = 1.5 \]

\[ E_C = 2 \frac{v_{sat}}{\mu_{eff}} \]

\[ v_{sat} = 10^5 \text{ m/s} \]

Constant velocity

Constant mobility (slope = \( \mu \))
Modeling Velocity Saturation

• Fit the velocity-dependence curve

\[ v = \frac{\mu_{\text{eff}} E}{\left(1 + \left(\frac{E}{E_C}\right)^n\right)^{1/n}} \]

- NMOS: \( n = 2 \)
- PMOS: \( n = 1 \)
Modeling Velocity Saturation

- A few approximations: (a) $n \to \infty$, (b) $n = 1$, (c) piecewise

\[ E_{c}/2 = \frac{\nu_n}{\mu_{\text{eff}}} \]

\[ \nu_n = \frac{E}{\mu_{\text{eff}}/2} \]

Piecewise
Short-Channel MOS On-Current
Approximation $n \to \infty$

1) $v = \mu_{\text{eff}} E, \quad E < E_C$

$$I_{DS} = \mu_{\text{eff}} \frac{W}{L} \left( (V_{GS} - V_{Th})V_{DS} - \frac{V_{DS}^2}{2} \right)$$

2) $v = v_{\text{sat}}, \quad E > E_C$

$$I_{Dsat} = \mu_{\text{eff}} \frac{W}{L} \left( (V_{GS} - V_{Th})V_{Dsat} - \frac{V_{Dsat}^2}{2} \right)$$

$V_{Dsat} = ?$

Can be reduced to Rabaey DIC model by making $V_{Dsat} = \text{const}$

Is this physically justified?
MOS Model from DIC, 2nd ed.

\[ I_D = 0 \text{ for } V_{GT} \leq 0 \]

\[ I_D = k' \frac{W}{L} \left( V_{GT} V_{\text{min}} - \frac{V_{\text{min}}^2}{2} \right) (1 + \lambda V_{DS}) \text{ for } V_{GT} \geq 0 \]

with \( V_{\text{min}} = \min(V_{GT}, V_{DS}, V_{DSAT}) \),

\[ V_{GT} = V_{GS} - V_T \]

and \( V_T = V_{T0} + \gamma (\sqrt{-2F} + V_{SB}) - \sqrt{-2F} \)

\( \gamma \) - body effect parameter

From Rabaey, 2nd ed.
Unified MOS Model

• Model presented is compact and suitable for hand analysis.

• Still have to keep in mind the main approximation: that $V_{DSat}$ is constant. When is it going to cause largest errors?
  • When does $E$ scale? – Transistor stacks.

• But the model still works fairly well.
  • Except for stacks
Approximation $n = 1$, piecewise

• $n = 1$ is solvable, piecewise closely approximates

\[
v = \begin{cases} 
\frac{\mu_{\text{eff}} E}{1 + E/E_{C}}, & E < E_{C} = \frac{2v_{\text{sat}}}{\mu_{\text{eff}}} \\
 v_{\text{sat}}, & E > E_{C}
\end{cases}
\]

Velocity, $v$:

Sodini, Ko, Moll, TED’84
Toh, Ko, Meyer, JSSC’88
BSIM model
Drain Current

• We can find the drain current by integrating $I_{DS}$

$$I_{DS} = WC_{ox}(V_{GS} - V_{Th} - V_C(x)) \, V$$

Linear:

$$I_{DS} = \frac{\mu C_{ox} W}{1 + (V_{DS}/E_C L) L} \left( (V_{GS} - V_{Th})V_{DS} - \frac{V_{DS}^2}{2} \right)$$

In saturation:

$$I_{DSat} = C_{ox} W v_{sat} (V_{GS} - V_{Th} - V_{Dsat})$$

$$I_{Dsat} = \frac{\mu C_{ox} W}{1 + (V_{DSat}/E_C L) L} \left( (V_{GS} - V_{Th})V_{Dsat} - \frac{V_{Dsat}^2}{2} \right)$$
Drain Current in Velocity Saturation

• Solving for $V_{Dsat}$

$$V_{DSat} = \frac{(V_{GS} - V_{Th})E_CL}{(V_{GS} - V_{Th}) + E_CL}$$

⇒ And saturation current

$$I_{DSat} = \frac{W \mu_{eff} C_{ox} E_CL}{L} \frac{(V_{GS} - V_{Th})^2}{(V_{GS} - V_{Th}) + E_CL}$$
Velocity Saturation

![Velocity Saturation Graph](image)
Velocity Saturation

- $E_{CL}$ is $V_{GS}$-dependent
- Can calculate $V_{DSat}$ \[(V_{Th} \sim 0.4\text{V in 28nm})\]

<table>
<thead>
<tr>
<th>$V_{GS}$ [V]</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{DSat}$ [V]</td>
<td>0</td>
<td>0.05</td>
<td>0.11</td>
<td>0.18</td>
<td>0.25</td>
<td>0.33</td>
</tr>
</tbody>
</table>

- For $V_{GS} - V_{Th} << E_{CL}$, $V_{DSat}$ is close to $V_{GS} - V_{Th}$
- For large $V_{GS}$, $V_{DSat}$ bends upwards toward $E_{CL}$
- Therefore, $E_{CL}$ can be sometimes approximated with a constant term
  - But also need to understand the limitation of the approximation
Application of I-V Models
Application of Models: NAND Gate

• 2-input NAND gate

Sizing for equal transitions:
• P/N ratio (β-ratio): 1 for L < 20nm, 1.6 for 20nm < L < ~65nm, 2 for L > 90nm
• Upsizing stacks by a factor proportional to the stack height
Transistor Stacks

• With transistor stacks, $V_{DS}$, $V_{GS}$ reduce.
• Unified model assumes $V_{DSat} = \text{const.}$
• For a stack of two, appears that both have exactly double $R_{ekv}$ of an inverter with the same width
• Therefore, doubling the size of each, should make the pull-down $R$ equivalent to an inverter
Velocity Saturation

• As \((V_{GS} - V_{Th})/E_CL\) changes, the depth of saturation changes

\[
I_{DSat} = \frac{W \mu_{eff} C_{ox} E_CL}{L} \frac{(V_{GS} - V_{Th})^2}{2 (V_{GS} - V_{Th}) + E_CL}
\]

• For \(V_{GS}, V_{DS} = 1.0V, E_CL\) is \(\sim 0.75V\)
• With double length, \(E_CL\) is 1.5V (in this model in 28nm)
• Stacked transistors are less saturated
• \(V_{GS} - V_{Th} = 0.6V, I_{DSat} \sim 2/3\) of inverter \(I_{DSat}\) (64%)
• Therefore NAND2 should have pull-down sized 1.5X
• Check any library NAND2’s
• Current halved in a stack of 3
• PMOS less impacted by velocity saturation since lower mobility
Note about FinFETs

- Widths are quantized
Example: Logical Effort

- **Older CMOS (130nm)**

- **Planar CMOS (~28nm, bulk, FDSOI)**

- **FinFET (7nm)**

\[
\begin{align*}
g_{\text{inv}} &= 1 \\
g_{\text{NAND2}} &= \\
g_{\text{NOR2}} &= \\
\end{align*}
\]
Example: Logical Effort

- **Older CMOS (130nm)**

  \[ g_{\text{inv}} = 1 \]

  \[ g_{\text{NAND2}} = \frac{4}{3} = 1.33 \]

- **Planar CMOS (~28nm, bulk, FDSOI)**

  \[ g_{\text{inv}} = 1 \]

  \[ g_{\text{NAND2}} = \frac{3.1}{2.6} = 1.2 \]

- **FinFET (7nm)**

  \[ g_{\text{inv}} = 1 \]

  \[ g_{\text{NAND2}} = \frac{2.5}{2} = 1.25 \]
Example: Logical Effort

- Older CMOS (130nm)

\[ g_{\text{inv}} = 1 \]

\[ g_{\text{NAND2}} = \frac{4}{3} = 1.33 \]

\[ g_{\text{NOR2}} = \frac{5}{3} = 1.66 \]

- Planar CMOS (~28nm, bulk, FDSOI)

\[ g_{\text{inv}} = 1 \]

\[ g_{\text{NAND2}} = \frac{3.1}{2.6} = 1.2 \]

\[ g_{\text{NOR2}} = \frac{3.9}{2.6} = 1.5 \]

- FinFET (7nm)

\[ g_{\text{inv}} = 1 \]

\[ g_{\text{NAND2}} = \frac{2.5}{2} = 1.25 \]

\[ g_{\text{NOR2}} = \frac{2.5}{2} = 1.25 \]
Other Velocity Saturation Models
Other Models: Alpha Power Law Model

- Simple model, sometimes useful for hand analysis

\[ I_{DS} = \frac{W}{2L} \mu C_{ox} (V_{GS} - V_{Th})^\alpha \]

Parameter \( \alpha \) is between 1 and 2.

Sakurai, Newton, JSSC 4/90
Alpha Power Law Model

• This is not a physical model

• Simply empirical:
  • Can fit (in minimum mean squares sense) to variety of $\alpha$’s, $V_{Th}$
  • Need to find one with minimum square error – fitted $V_{Th}$ can be different from physical
  • Can also fit to $\alpha = 1$
    • What is $V_{Th}$?
$K(V_{GS} - V_{THZ})$ Model ($\alpha = 1$)

Drain current vs. gate-source voltage
## Saturation Current Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_{DS} = K \frac{W}{L} (V_{GS} - V_{THZ}) )</td>
<td>Delay estimates with ( V_{DD} \gg V_{TH} )</td>
</tr>
<tr>
<td>( I_{DS} = \frac{W}{L} \frac{\mu C_{ox}}{2} (V_{GS} - V_{TH})^2 )</td>
<td>Long channel devices (rare in digital)</td>
</tr>
<tr>
<td>( I_{DS} = \frac{W}{L} \frac{\mu C_{ox}}{2} (V_{GS} - V_{TH})^\alpha )</td>
<td>Delay estimates in a wider range of ( V_{DD} )'s</td>
</tr>
<tr>
<td>( I_{DS} = \frac{W}{L} \frac{\mu C_{ox}}{2} \left( (V_{GS} - V_{TH}) V_{Dsat} - \frac{V_{Dsat}^2}{2} \right) )</td>
<td>Easy to remember, does not handle stacks correctly</td>
</tr>
<tr>
<td>( I_{DS} = \frac{W}{L} \frac{\mu C_{ox}}{2} \frac{E_{CL} (V_{GS} - V_{TH})^2}{(V_{GS} - V_{TH}) + E_{CL}} )</td>
<td>Handles stacks correctly, sizing</td>
</tr>
</tbody>
</table>
Next Lecture

- Gate models