Course Overview

• Targeted for graduate students who have already taken basic VLSI design classes
• Real world challenges and solutions in designing high-performance and low-power circuits
• Relations to VLSI Design
  – Recent developments in digital IC design
  – Project oriented
  – Student participation: class presentation
Prerequisite

• MOS VLSI Design or equivalent
  – MOS transistor
  – Static, dynamic logic
  – Adder

• Familiarity with VLSI CAD tools
  – Magic or Cadence: LVS, DRC
  – HSPICE

• Basic knowledge on solid-state physics
Class Materials

• Lecture notes: primary reference


Class Organization

• One exam (40% of overall grade)
• Term-long project (60%)
  – Proposal (5%)
  – Midterm presentation (15%) – background material and proposed work
  – Final presentation (20%)
  – Final report (20%)
CAD Tools

- Cadence
  - Schematic editor, layout editor, DRC, LVS
- HSPICE, awaves
- Technology files
  - TSMC 0.18\(\mu\)m, BPTM 70nm, …
- Synopsys design compiler, library compiler
- Taurus-device, Taurus-medici

Everyone should have some experience with these tools
Term Project

• Single person project

• Proposal (~week 3)
  – 2 pages
  – Topic, problem statement, research plan, references

• Midterm presentation (~week 7)
  – 15 mins
  – Literature survey
  – Off campus students can give presentations over the phone

• Final presentation (early December)
  – 20 mins
  – Background, final results, contributions

• Final report (Dec. 10)
  – Publishable quality
Project Topic

- Students pick the research topic they want to work on
- After the literature survey, choose a paper that you would like to evaluate yourself
- Has to be on digital VLSI circuit DESIGN
  - Op-amp design alone is not acceptable
  - Op-amp design for digital applications is acceptable
- Show the paper’s claim using your own simulations
- Your contribution must be clearly shown at the end
  - Improve previous design
  - New circuit, modeling technique
  - Show limitation of previous techniques
- Talk to the instructor in case you need help
How to Find a Project Topic?

• Conferences
  – International Solid-State Circuits Conference (ISSCC, top conference!): slides posted on IEEExplore
  – Symposium on VLSI Circuits (VLSIC), DAC, ICCAD
  – Custom Integrated Circuits Conference (CICC)

• Journal
  – IEEE TVLSI, IEEE TCAD, IEEE TED
  – IEEE Journal of Solid-State Circuits (JSSC)
  – Intel Technology Journal
  – IBM Journal on R & D
How to Find a Project Topic?

• Funding agencies
  – Research needs document (www.src.org)

• Presentation
  – University of Michigan VLSI seminar series (www.eecs.umich.edu/vlsi_seminar/)
  – Design automation conference (www.dac.com)

• Pick a recent issue in VLSI design (< 5 years)
• I suggest you start doing the literature survey ASAP (deadline coming up in 3 weeks)
Acknowledgements

• Prof. Chris Kim
• Intel circuit research labs (S. Borkar and many others)
• IBM
• Copy right 2002 J. Rabaey et al.
Academic Misconduct

• Students caught engaging in an academically dishonest practice will receive a failing grade for the course.

• University policy on academic dishonesty will be followed strictly.
Course Topics

- Scaling issues
- High performance design
  - High performance logic family, clocking strategies, interconnects
- Low power design
  - Low voltage designs, leakage control techniques, circuit/device/technology issues, low power SRAM
- Variation tolerant design
  - Process compensating techniques
- Power delivery, interconnect, reliability
- Bulk and SOI
A physical system as a computing medium

• We need to create a bit first. Information processing always requires physical carrier, which are material particles.

• **First** requirement to physical realization of a bit implies creating *distinguishable* states within a system of such material particles.

• **The second** requirement is *conditional* change of state.

• The properties of *distinguishability* and *conditional change of state* are two fundamental properties of a material subsystem to represent information. **These properties can be obtained by creating energy barriers** in a material system.
Particle Location is an Indicator of State

1 1 0 0 1 0
Two-well bit
Barrier engineering in semiconductors

By doping, it is possible to create a built-in field and energy barriers of controllable height and length within semiconductor. It allows one to achieve conditional complex electron transport between different energy states inside semiconductors that is needed in the physical realization of devices for information processing.
Kroemer’s Lemma of Proven Ignorance

• If in discussing a semiconductor problem, you cannot draw an Energy-Band-Diagram, this shows that you don’t know what are you talking about

• If you can draw one, but don’t, then your audience won’t know what are you talking about
Intel founder and chairman Gordon Moore predicted in 1965 that the number of transistors on a chip will double every 18-24 months.
Transistor Scaling

- Constant E-field scaling: voltage and dimensions (both horizontal and vertical) are scaled by the same factor $k$, ($\sim 1.4$), such that the electrical field remains unchanged.
Technology Scaling

**Dimensions** $\xrightarrow{\text{scale}} 0.7$, $V_{dd} \xrightarrow{\text{scales}} \beta$, $V_t \xrightarrow{\text{scales}} \beta$

\[
I = \frac{kW}{T_{ox}} (V_{dd} - V_t) \xrightarrow{\text{scales}} \frac{0.7}{0.7} \times \beta = \beta
\]

\[
D = \frac{CV_{dd}}{I} \xrightarrow{\text{scales}} \frac{0.7 \times \beta}{\beta} = 0.7 \quad (30\% \text{ delay reduction})
\]

\[
E = CV_{dd}^2 \xrightarrow{\text{scales}} 0.7 \beta^2
\]
IC Frequency & Power Trends

- Clock frequency improves 50%
- Gate delay improves ~30%
- Power increases 50%
- Power = $C_L V^2 f$

Active switched capacitance “$C_L$” is increasing.

$\text{Frequency (MHz)}$ vs $\text{Technology Generation (\(\mu\text{m}\))}$

- Chip Power (W)
- Power = $C_L V^2 f$
- Frequency = $\frac{1}{2\pi f}$

- 486DX CPU
- 386 CPU
- Pentium Processor
- Pentium II Processor
Vdd vs. Vt scaling

- Recently: constant e-field scaling, aka voltage scaling
  - $V_{cc} \propto 1V$
  - $V_{cc}$ & modest $V_T$ scaling
  - Loss in gate overdrive ($V_{cc}-V_T$)

Voltage scaling is good for controlling IC’s active power, but it requires aggressive $V_T$ scaling for high performance.
Delay

\[ \tau_d = \frac{C_L V_{DD}}{I_D} \]

Long Channel MOSFET

\[ \tau_d = \frac{C_L}{(W/2L)\mu C_{ox} V_{DD}(1 - \frac{V_T}{V_{DD}})^2} \]

Short Channel MOSFET

\[ \tau_d = \frac{C_L}{WC_{ox} V_{SAT}(1 - \frac{V_T}{V_{DD}})} \]

\[
\tau = \frac{C_L 0.5 T_{ox} 0.5}{V_{DD}^{0.3}(0.9 - \frac{V_T}{V_{DD}})^{1.3}} \left( \frac{1}{W_n} + \frac{2.2}{W_p} \right) \]

[1]


Performance significantly degrades when \( V_{DD} \) approaches \( 3V_T \).
**VT Scaling: VT and IOFF Trade-off**

Performance vs Leakage:

\[ \text{VT} \downarrow \text{IOFF} \uparrow \text{ID(SAT)} \uparrow \]

\[ I_{OFF} \propto I_{subth} \propto \frac{W_{eff}}{L_{eff}} K_1 e^{(V_{GS}-V_T)} \]

\[ I_D(SAT) \propto \frac{W_{eff}}{L_{eff}} K_2 (V_{GS} - V_T)^2 \]

\[ I_D(SAT) \propto K_3 W_{eff} C_{ox} v_{SAT} (V_{GS} - V_T) \]

\[ \downarrow \text{As VT decreases, sub-threshold leakage increases} \]

\[ \downarrow \text{Leakage is a barrier to voltage scaling} \]
## Constant Field Scaling

<table>
<thead>
<tr>
<th>Scaling assumptions</th>
<th>Device and circuit parameters</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Device dimensions ($t_{ox}$, $L$, $W$, $X_j$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td></td>
<td>Doping concentration ($N_a$, $N_d$)</td>
<td>$k$</td>
</tr>
<tr>
<td></td>
<td>Voltage ($V$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td>Device parameters</td>
<td>Electric field ($E$)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Capacitance ($C=\varepsilon A/t$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td></td>
<td>Current ($I$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td></td>
<td>Channel resistance ($R_{ch}$)</td>
<td>1</td>
</tr>
<tr>
<td>Circuit parameters</td>
<td>Delay ($CV/I$)</td>
<td>$1/k$</td>
</tr>
<tr>
<td></td>
<td>Power ($VI$)</td>
<td>$1/k^2$</td>
</tr>
<tr>
<td></td>
<td>Switching energy ($CV^2$)</td>
<td>$1/k^3$</td>
</tr>
<tr>
<td></td>
<td>Circuit density ($1/A$)</td>
<td>$k^2$</td>
</tr>
<tr>
<td></td>
<td>Power density ($P/A$)</td>
<td>1</td>
</tr>
</tbody>
</table>
Scaling in the Vertical Dimension

- Transistor $V_t$ rolls off as the channel length is reduced
- Shallow junction depth reduces $V_t$ roll-off
- However, sheet resistance increases
Scaling in the Vertical Dimension

- Vertical dimension scales less than horizontal
- Aggravates short channel effect ($V_t$ roll-off)
## Constant Voltage Scaling

<table>
<thead>
<tr>
<th>Scaling assumptions</th>
<th>Device and circuit parameters</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device dimensions ($t_{ox}$, $L$, $W$, $X_j$)</td>
<td>$1/k$</td>
<td></td>
</tr>
<tr>
<td>Doping concentration ($N_a$, $N_d$)</td>
<td>$k$</td>
<td></td>
</tr>
<tr>
<td>Voltage ($V$)</td>
<td>$1$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Device parameters</th>
<th>Electric field ($E$)</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacitance ($C=\varepsilon A/t$)</td>
<td>$1/k$</td>
<td></td>
</tr>
<tr>
<td>Current ($I$)</td>
<td>$k$</td>
<td></td>
</tr>
<tr>
<td>Channel resistance ($R_{ch}$)</td>
<td>$1/k$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Circuit parameters</th>
<th>Delay ($CV/I$)</th>
<th>$1/k^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power ($VI$)</td>
<td>$k$</td>
<td></td>
</tr>
<tr>
<td>Switching energy ($CV^2$)</td>
<td>$1/k$</td>
<td></td>
</tr>
<tr>
<td>Circuit density ($1/A$)</td>
<td>$k^2$</td>
<td></td>
</tr>
<tr>
<td>Power density ($P/A$)</td>
<td>$k^3$</td>
<td></td>
</tr>
</tbody>
</table>
Constant Voltage Scaling

- More aggressive scaling than constant field
- Limitations
  - Reliability problems due to high field
  - Power density increases too fast

- Both constant field and constant voltage scaling have been followed in practice
- Field and power density has gone up as a byproduct of high performance, but till now designers are able to handle the problems
## ITRS Roadmap

<table>
<thead>
<tr>
<th>Year</th>
<th>2001</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2010</th>
<th>2013</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM ½ pitch [nm]</td>
<td>130</td>
<td>100</td>
<td>80</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>MPU transistors/chip</td>
<td>97M</td>
<td>153M</td>
<td>243M</td>
<td>386M</td>
<td>773M</td>
<td>1.55G</td>
<td>3.09G</td>
</tr>
<tr>
<td>Wiring levels</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>High-perf. phys. gate [nm]</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>25</td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>High-perf. VDD [V]</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Local clock [GHz]</td>
<td>1.7</td>
<td>3.1</td>
<td>5.2</td>
<td>6.7</td>
<td>11.5</td>
<td>19.3</td>
<td>28.8</td>
</tr>
<tr>
<td>High-perf. power [W]</td>
<td>130</td>
<td>150</td>
<td>170</td>
<td>190</td>
<td>218</td>
<td>251</td>
<td>288</td>
</tr>
</tbody>
</table>

Transistor Scaling

- 90nm is in production, 65nm in research phase
- New technology generation introduced every 2-3 years
Cost per Transistor

- You can buy 10M transistors for a buck
- They even throw in the interconnect and package for free
Transistors Shipped Per Year

- Today, there are about 100 transistors for every ant
  - Gordon Moore, ISSCC ‘04
Transistors per Chip

- 1.7B transistors in Montecito (next generation Itanium)
- Most of the devices used for on-die cache memory
Moore’s Wrong Prediction
Chip Frequency

- 30% higher frequency every new generation
Die Size

- ~15% larger die every new generation
- This means more than 2X increase in transistors per chip

Graph shows:
- ~7% growth per year
- ~2X growth in 10 years
- Key models: 8080, 8008, 8085, 286, 386, 486, Pentium® proc
Supply Voltage Scaling

- Supply voltage is reduced for active power control

\[ P_{\text{active}} \propto CV_{dd}^2 f \]
4 Decades of Transistor Scaling: Itanium 2 Processor

- 130nm process
- 410M transistors
- 374mm² die size
- 6MB on-die L3 cache
- 1.5GHz at 1.3V
- 6.4GB/s 400MT/s 4-way bus interface
- System compatible with existing Itanium 2 platforms
- Extensive RAS, DFT and DFM features
Power Density

- High-end microprocessors: Packaging, cooling
- Mobile/handheld applications: Short battery life
Active and Leakage Power

\[ \text{delay} \propto \frac{C_L}{V_t} \frac{1}{1 - \frac{V_t}{V_{dd}}} \]

\[ I_{\text{leak}} \propto \exp\left(\frac{-V_t}{mkT/q}\right) \]

- Transistors are becoming dimmers
Leakage Power Crawling Up in Itanium 2

- Same thermal design envelope as the 180nm Itanium® 2 processor
  - 50% frequency increase
  - 2X larger L3 cache
  - Leakage increased 3.5X

- Transistor leakage is perhaps the biggest problem
Leakage power is problematic in active mode for high performance microprocessors.
Thermal Runaway

- Destructive positive feedback mechanism
- Leakage increases exponentially with temperature
- May destroy the test socket → thermal sensors required
Gate Oxide Thickness

- Electrical $t_{ox} >$ Physical $t_{ox}$
- Due to gate depletion and carrier quantization in the channel
Gate Tunneling Leakage

- Placing a few SiO$_N$ species uniformly in billions of devices

- MOSFET no longer have infinite input resistance
- Impacts both power and functionality of circuits
Process Variation in Microprocessors

- Fast chips burn too much power
- Slow chips cannot meet the frequency requirement
Process Variation in Transistors

- More than 2X variation in $I_{on}$, 100X variation $I_{off}$
- Within-dies, die-to-die, lot-to-lot
Sources of Process Variation

- Intrinsic parameter variation (static)
  - Channel length, random dopant fluctuation
- Environmental variation (dynamic)
  - Temperature, supply variations
Sub-wavelength Lithography
Line Edge/Width Roughness

- $I_{off}$ and $I_{dsat}$ impacted by LER and LWR
Random Dopant Fluctuation

- $V_t$ variation caused by non-uniform channel dopant distribution
Supply Voltage Integrity

- IR noise due to large current consumption
- Ldi/dt noise due to new power reduction techniques (clock gating, power gating, body biasing) with power down mode
Supply Voltage Integrity

• Degrades circuit performance
• Supply voltage overshoot causes reliability issues
• Power wasted by parasitic resistance causes self-heating
• $V_{dd}$ fluctuation should be less than 10%

Courtesy IBM
Productivity Gap

- Design complexity surpasses manpower
- Effective CAD tools, memory dominated chips
What will end Moore’s law, economics or physics?
Interconnect Scaling

- Global interconnects get longer due to larger die size
- Wire scaling increases R, L and C
- Example: local vs. global interconnect delay
Interconnect Delay Problem

1997 SIA technology roadmap

- Local interconnect has sped up (shorter wires)
- Global interconnect has slowed down (RC doesn’t scale)
Interconnect Metal Layers

- Local wires have high density to accommodate the increasing number of devices
- Global wires have low RC (tall, wide, thick, scarce wires)
Interconnect distribution scaling trends

- RC/μm scaling trend is only one side of the story...

Source: Intel
Power Delivery & Distribution Challenges

- High-end microprocessors approaching > 10 GHz
  - How to deliver and distribute ~100A at < 1V for < $20!
  - On-die power density >>> hot-plate power density
    - crossover happened back in 0.6\(\mu\)m technology!
    - di/dt noise only worsening with scaling: drivers are one of the sources.
Example multi-layer system

M7
2.00µm

M6
2.00µm

M5
1.6µm

M4
1.6µm

M3
0.64µm

M2
0.64µm

M1
0.50µm

Poly
Substrate

Al
1.00 x 1.80µm

0.80 x 1.6µm

0.80 x 1.6µm

0.32 x 0.64µm

0.32 x 0.64µm

0.25 x 0.48µm

K. Soumyanath et al. [2]
Cross Talk Noise

- As wires are brought closer with scaling, capacitive coupling becomes significant.
- Adjacent wires on same layer have stronger coupling.
Cross Talk Noise

- Multiple aggressors multiple victims possible
- Cross talk noise can cause logic faults in dynamic circuits

Noise $\propto Cc \ dV/dt$
Cross Talk and Delay

- Capacitive cross talk can affect delay
- If aggressor(s) switch in opposite direction, effective coupling capacitance is doubled
- On the other hand, if aggressor(s) switch in the same direction, $C_c$ is eliminated
- Significant difference in RC delay depending on adjacent switching activity
Soft Error In Storage Nodes

• Soft errors are caused by
  – Alpha particles from package materials
  – Cosmic rays from outer space
Soft Error In Storage Nodes

- Error correction code
- Shielding
- SOI
- Radiation-hardened cell
More Roadblocks

● Memory stability
● Long term reliability
● Mixed signal design issues
● Mask cost
● Testing multi-GHz processors
● Skeptics: Do we need a faster computer?
● …

● Eventually, it all boils down to economics
Summary

- Digital IC Business is Unique
  - Things Get Better Every Few Years
  - Companies Have to Stay on Moore’s Law Curve to Survive

- Benefits of Transistor Scaling
  - Higher Frequencies of Operation
  - Massive Functional Units, Increasing On-Die Memory
  - Cost/MIPS Going Down

- Downside of Transistor Scaling
  - Power (Dynamic and Static)
  - Process Variation
  - Design/Manufacturing Cost
  - ....