What is this course all about?

• **Introduction to digital integrated circuits.**

• **What will you learn?**
  – Understanding, designing, and optimizing digital circuits with respect to different quality metrics: cost, speed, power dissipation, and reliability
Digital Integrated Circuits

- Introduction: Issues in digital design
- The CMOS inverter
- Combinational logic structures
- Sequential logic gates
- Design methodologies
- Interconnect: R, L and C
- Timing
- Arithmetic building blocks
- Memories and array structures

Why does it matter?

**FIG 1.1** Size of worldwide semiconductor market

Source: Semiconductor Industry Association.
A Brief History

• 1947: First Transistor at Bell Lab [John Bardeen and Walter Brattain]
• 1958: First Integrated circuit at Texas Instruments [Jack Kilby]
• 1965: Moore’s Law, Intel [Gordon Moore]
• 1994: Integrated circuits became $100B/year business
• 2003: Industry manufactured $10^{18}$ (one quintillion) transistors (200M per human being)

The First Computer

The Babbage Difference Engine (1832)
25,000 parts
cost: £17,470
ENIAC - The first electronic computer (1946)

- Vacuum tube based digital computer
- “The Giant Brain” as labeled by the press
- ENIAC facts
  - Occupied 1,800 sq. feet
  - Weighted 30 tons
  - 18000 vacuum tubes
- Application: calculate firing tables for World War II artillery guns

The Transistor Revolution

First transistor
Bell Labs, 1948
The First Integrated Circuits

Bipolar logic
1960’s

ECL 3-input Gate
Motorola 1966

IC Evolution

- SSI – Small Scale Integration (early 1970s)
  - contained 1 – 10 logic gates
- MSI – Medium Scale Integration
  - logic functions, counters
- LSI – Large Scale Integration
  - first microprocessors on the chip
- VLSI – Very Large Scale Integration
  - now offers 64-bit microprocessors,
    complete with cache memory (L1 and often L2),
    floating-point arithmetic unit(s), etc.
IC Evolution

- Bipolar technology
  - TTL (transistor-transistor logic), 1962; higher integration density
  - ECL (emitter-coupled logic), 1974; high-performance

- MOS (Metal-oxide-silicon)
  - although invented before bipolar transistor (1925, 1935), was initially difficult to manufacture
  - nMOS (n-channel MOS) technology developed in late 1970s required fewer masking steps, was denser, and consumed less power than equivalent bipolar ICs => an MOS IC was cheaper than a bipolar IC and led to investment and growth of the MOS IC market.
  - aluminum gates for replaced by polysilicon by early 1980
  - CMOS (Complementary MOS): n-channel and p-channel MOS transistors => lower power consumption, simplified fabrication process

Intel 4004

- Introduction date: November 15, 1971
- Clock speed: 108 KHz
- Number of transistors: 2,300 (10 microns)
- Bus width: 4 bits
- Addressable memory: 640 bytes
- Typical use: calculator, first microcomputer chip, arithmetic manipulation
### Pentium 4

- **0.18-micron process technology**
  - Introduction date: August 27, 2001 (2.0 GHz); August 20, 2000 (1.5, 1.4 GHz)
  - Level Two cache: 256 KB Advanced Transfer Cache (Integrated)
  - System Bus Speed: 400 MHz
  - SSE2 SIMD Extensions
  - Transistors: 42 Million
  - Typical Use: Desktops and entry-level workstations

- **0.13-micron process technology**
  - Introduction date: January 7, 2002
  - Level Two cache: 512 KB Advanced
  - Transistors: 55 Million

### Intel’s McKinley

- Introduction date: Mid 2002
- Caches: 32KB L1, 256 KB L2, 3MB L3 (on-chip)
- Clock: 1GHz
- Transistors: 221 Million
- Area: 464mm²
- Typical Use: High-end servers
- Future versions: 5GHz, 0.13-micron technology
Moore’s Law

- In 1965, Gordon Moore noted that the number of transistors on a chip doubled every 18 to 24 months.
- He made a prediction that semiconductor technology will double its effectiveness every 18 months.

Electronics, April 19, 1965.
Evolution in Complexity

Transistor Counts
Moore’s law in Microprocessors

Transistors on Lead Microprocessors double every 2 years

Die Size Growth

Die size grows by 14% to satisfy Moore’s Law

Courtesy, Intel
Frequency

Doubles every 2 years

Power Dissipation

Lead Microprocessors power continues to increase

Courtesy, Intel
Power will be a major problem

Power delivery and dissipation will be prohibitive

Power density too high to keep junctions at low temp

Courtesy, Intel
### Technology Directions: SIA Roadmap

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature size (nm)</td>
<td>180</td>
<td>130</td>
<td>100</td>
<td>70</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Logic trans/cm²</td>
<td>6.2M</td>
<td>18M</td>
<td>39M</td>
<td>84M</td>
<td>180M</td>
<td>390M</td>
</tr>
<tr>
<td>Cost/trans (mc)</td>
<td>1.735</td>
<td>.580</td>
<td>.255</td>
<td>.110</td>
<td>.049</td>
<td>.022</td>
</tr>
<tr>
<td>#pads/chip</td>
<td>1867</td>
<td>2553</td>
<td>3492</td>
<td>4776</td>
<td>6532</td>
<td>8935</td>
</tr>
<tr>
<td>Clock (MHz)</td>
<td>1250</td>
<td>2100</td>
<td>3500</td>
<td>6000</td>
<td>10000</td>
<td>16900</td>
</tr>
<tr>
<td>Chip size (mm²)</td>
<td>340</td>
<td>430</td>
<td>520</td>
<td>620</td>
<td>750</td>
<td>900</td>
</tr>
<tr>
<td>Wiring levels</td>
<td>6-7</td>
<td>7</td>
<td>7-8</td>
<td>8-9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Power supply (V)</td>
<td>1.8</td>
<td>1.5</td>
<td>1.2</td>
<td>0.9</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>High-perf pow (W)</td>
<td>90</td>
<td>130</td>
<td>160</td>
<td>170</td>
<td>175</td>
<td>183</td>
</tr>
</tbody>
</table>

### Not Only Microprocessors

**Cell Phone**

**Digital Cellular Market (Phones Shipped)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Units</td>
<td>48M</td>
<td>86M</td>
<td>162M</td>
<td>260M</td>
<td>435M</td>
</tr>
</tbody>
</table>

(data from Texas Instruments)
Why Scaling?

- Technology shrinks by 0.7/generation
- With every generation can integrate 2x more functions per chip; chip cost does not increase significantly
- Cost of a function decreases by 2x
- But …
  - How to design chips with more and more functions?
  - Design engineering population does not double every two years…
- Hence, a need for more efficient design methods
  - Exploit different levels of abstraction

Design Abstraction Levels
Major Design Challenges

- **Microscopic issues**
  - ultra-high speeds
  - power dissipation and supply rail drop
  - growing importance of interconnect
  - noise, crosstalk
  - reliability, manufacturability
  - clock distribution

- **Macroscopic issues**
  - time-to-market
  - design complexity (millions of gates)
  - high levels of abstractions
  - design for test
  - reuse and IP, portability
  - systems on a chip (SoC)
  - tool interoperability

<table>
<thead>
<tr>
<th>Year</th>
<th>Tech.</th>
<th>Complexity</th>
<th>Frequency</th>
<th>3 Yr. Design Staff Size</th>
<th>Staff Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>0.35</td>
<td>13 M Tr.</td>
<td>400 MHz</td>
<td>210</td>
<td>$90 M</td>
</tr>
<tr>
<td>1998</td>
<td>0.25</td>
<td>20 M Tr.</td>
<td>500 MHz</td>
<td>270</td>
<td>$120 M</td>
</tr>
<tr>
<td>1999</td>
<td>0.18</td>
<td>32 M Tr.</td>
<td>600 MHz</td>
<td>360</td>
<td>$160 M</td>
</tr>
<tr>
<td>2002</td>
<td>0.13</td>
<td>130 M Tr.</td>
<td>800 MHz</td>
<td>800</td>
<td>$360 M</td>
</tr>
</tbody>
</table>

Productivity Trends

- Complexity outpaces design productivity

Source: Sematech

Courtesy, ITRS Roadmap
**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

---

**Cost of Integrated Circuits**

- NRE (non-recurring engineering) costs
  - Fixed cost to produce the design
    - design effort
    - design verification effort
    - mask generation
    - Influenced by the design complexity and designer productivity
    - More pronounced for small volume products
- Recurring costs – proportional to product volume
  - silicon processing
    - also proportional to chip area
  - assembly (packaging)
  - test

\[
\text{Cost per IC} = \text{Variable cost per IC} + \frac{\text{Fixed cost}}{\text{Volume}}
\]
NRE Cost is Increasing

“The club of people who can afford an extreme sub-micron ASIC or COTS design is getting pretty exclusive.”

Ron Wilson, EE Times (May 2000)

Cost per Transistor

Fabrication capital cost per transistor (Moore’s law)
Silicon Wafer

From http://www.amd.com
Going up to 12" (30cm)

Recurring Costs

Variable cost = \( \frac{\text{Die cost} + \text{Testing cost} + \text{Packaging cost}}{\text{Final test yield}} \)

Cost of die = \( \frac{\text{Cost of wafer}}{\text{Dies per wafer} \times \text{Die yield}} \)
Dies per Wafer

\[
\text{Dies per wafer} = \frac{\pi \times (\text{Wafer diameter}/2)^2}{\text{Die area}} - \frac{\pi \times \text{Wafer diameter}}{\sqrt{2} \times \text{Die area}}
\]

Yield

\[
\text{Die yield} = \text{Wafer yield} \times \left(1 + \frac{\text{Defects per unit area} \times \text{Die area}}{\alpha}\right)^{-\alpha}
\]

\(\alpha\) is approximately 3

\[
\text{die cost} = f(\text{die area})^4
\]
### Examples of Cost Metrics (1994)

<table>
<thead>
<tr>
<th>Chip</th>
<th>Metal layers</th>
<th>Line width</th>
<th>Wafer cost</th>
<th>Defects/cm²</th>
<th>Area (mm²)</th>
<th>Dies/wafer</th>
<th>Yield</th>
<th>Die cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>386DX</td>
<td>2</td>
<td>0.90</td>
<td>$900</td>
<td>1.0</td>
<td>43</td>
<td>360</td>
<td>71%</td>
<td>$4</td>
</tr>
<tr>
<td>486DX2</td>
<td>3</td>
<td>0.80</td>
<td>$1200</td>
<td>1.0</td>
<td>81</td>
<td>181</td>
<td>54%</td>
<td>$12</td>
</tr>
<tr>
<td>PowerPC 601</td>
<td>4</td>
<td>0.80</td>
<td>$1700</td>
<td>1.3</td>
<td>121</td>
<td>115</td>
<td>28%</td>
<td>$53</td>
</tr>
<tr>
<td>HP PA 7100</td>
<td>3</td>
<td>0.80</td>
<td>$1300</td>
<td>1.0</td>
<td>196</td>
<td>66</td>
<td>27%</td>
<td>$73</td>
</tr>
<tr>
<td>DEC Alpha</td>
<td>3</td>
<td>0.70</td>
<td>$1500</td>
<td>1.2</td>
<td>234</td>
<td>53</td>
<td>19%</td>
<td>$149</td>
</tr>
<tr>
<td>Super SPARC</td>
<td>3</td>
<td>0.70</td>
<td>$1700</td>
<td>1.6</td>
<td>256</td>
<td>48</td>
<td>13%</td>
<td>$272</td>
</tr>
<tr>
<td>Pentium</td>
<td>3</td>
<td>0.80</td>
<td>$1500</td>
<td>1.5</td>
<td>296</td>
<td>40</td>
<td>9%</td>
<td>$417</td>
</tr>
</tbody>
</table>

### Yield Example

- **Example #1:**
  - 20-cm wafer for a die that is 1.5 cm on a side.
  - Solution: Die area = 1.5x1.5 = 2.25cm².
    Dies per wafer = 3.14x(20/2)²/2.25 = 3.14x20/(2x2.5)² = 110.

- **Example #2**
  - wafer size of 12 inches, die size of 2.5 cm², 1 defects/cm²,
    \( \mu = 3 \) (measure of manufacturing process complexity)
  - 252 dies/wafer (remember, wafers round & dies square)
  - die yield of 16%
  - 252 x 16% = only 40 dies/wafer die yield !

- **Die cost is strong function of die area**
  - proportional to the third or fourth power of the die area
Functionality and Robustness

- Prime requirement – IC performs the function it is designed for
- Normal behavior deviates due to:
  - variations in the manufacturing process (dimensions and device parameters vary between runs and even on a single wafer or die)
  - presence of disturbing on- or off-chip noise sources
- Noise: Unwanted variation of voltages or currents at the logic nodes

Reliability
Noise in Digital Integrated Circuits

- from two wires placed side by side
  - inductive coupling
    - current change on one wire can influence signal on the neighboring wire
  - capacitive coupling
    - voltage change on one wire can influence signal on the neighboring wire
    - cross talk
- from noise on the power and ground supply rails
  - can influence signal levels in the gate
Example of Capacitive Coupling

- Signal wire glitches as large as 80% of the supply voltage will be common due to crosstalk between neighboring wires as feature sizes continue to scale. 

Crosstalk vs. Technology

0.16m CMOS
0.12m CMOS
0.35m CMOS
0.25m CMOS

Pulsed Signal
Black line quiet
Red lines pulsed
Glitches strength vs technology

From Dunlop, Lucent, 2000

Static Gate Behavior

- Steady-state parameters of a gate – static behavior – tell how robust a circuit is with respect to both variations in the manufacturing process and to noise disturbances.

- Digital circuits perform operations on Boolean variables $x \in \{0, 1\}$

- A logical variable is associated with a nominal voltage level for each logic state

$$V_{OH} = \neg V_{OL}$$
$$V_{OL} = \neg V_{OH}$$

- Difference between $V_{OH}$ and $V_{OL}$ is the logic or signal swing $V_{SW}$
DC Operation
Voltage Transfer Characteristic

\[ V_{OH} = f(V_{IL}) \]
\[ V_{OL} = f(V_{IH}) \]

\[ V_{OH} = f(V_{OL}) \]
\[ V_{OL} = f(V_{OH}) \]
\[ V_M = f(V_M) \]

Mapping between analog and digital signals

- The regions of acceptable high and low voltages are delimited by \( V_{IH} \) and \( V_{IL} \) that represent the points on the VTC curve where the gain = -1 (dVout/dVin)
**Definition of Noise Margins**

- For robust circuits, want the “0” and “1” intervals to be as large as possible.

```
\[ NM_H = V_{OH} - V_{IH} \]
```

- Large noise margins are desirable, but not sufficient …

**The Regenerative Property**

- A gate with regenerative property ensures that a disturbed signal converges back to a nominal voltage level.
Conditions for Regeneration

\[ v_1 = f(v_0) \Rightarrow v_1 = f^{-1}(v_2) \]

Regenerative Gate

Nonregenerative Gate

- To be regenerative, the VTC must have a transient region with a gain greater than 1 (in absolute value) bordered by two valid zones where the gain is smaller than 1. Such a gate has two stable operating points.

Noise Immunity

- Noise margin expresses the ability of a circuit to overpower a noise source
  - noise sources: supply noise, cross talk, interference, offset
- Absolute noise margin values are deceptive
  - a floating node is more easily disturbed than a node driven by a low impedance (in terms of voltage)
- Noise immunity expresses the ability of the system to process and transmit information correctly in the presence of noise
- For good noise immunity, the signal swing (i.e., the difference between \( V_{OH} \) and \( V_{OL} \)) and the noise margin have to be large enough to overpower the impact of fixed sources of noise
Directivity

• A gate must be undirectional: changes in an output level should not appear at any unchanging input of the same circuit
  – In real circuits full directivity is an illusion (e.g., due to capacitive coupling between inputs and outputs)

• Key metrics: output impedance of the driver and input impedance of the receiver
  – ideally, the output impedance of the driver should be zero
  – input impedance of the receiver should be infinity

Fan-In and Fan-Out

- Fan-out – number of load gates connected to the output of the driving gate
  - gates with large fan-out are slower

- Fan-in – the number of inputs to the gate
  - gates with large fan-in are bigger and slower
The Ideal Inverter

- The ideal gate should have
  - infinite gain in the transition region
  - a gate threshold located in the middle of the logic swing
  - high and low noise margins equal to half the swing
  - input and output impedances of infinity and zero, resp.

\[ g = -\infty \]
\[ R_i = \infty \]
\[ R_o = 0 \]

Fanout = \( \infty \)

\[ \text{NM}_H = \text{NM}_L = VDD/2 \]

An Old-time Inverter
Delay Definitions

Propagation delay?

$V_{in} \rightarrow V_{out}$

$t_{p} = \frac{(t_{pHL} + t_{pLH})}{2}$

signal slopes?

$50\%$

$t_{pHL}$

$t_{pLH}$

$90\%$

$50\%$

$10\%$

$V_{out}$

input waveform

output waveform
Modeling Propagation Delay

- Model circuit as first-order RC network

\[ v_{out}(t) = (1 - e^{-t/\tau})V \]

where \( \tau = RC \)

Time to reach 50% point is
\[ t = \ln(2) \tau = 0.69 \tau \]

Time to reach 90% point is
\[ t = \ln(9) \tau = 2.2 \tau \]

- Matches the delay of an inverter gate

Power and Energy Dissipation

- Power consumption: how much energy is consumed per operation and how much heat the circuit dissipates
  - supply line sizing (determined by peak power)
    \[ P_{peak} = V_{dd} I_{peak} \]
  - battery lifetime (determined by average power dissipation)
    \[ P_{avg} = 1/T \int p(t) dt = V_{dd} I_{avg} \]

- Two important components: static and dynamic

\[ E \text{ (joules)} = C_L V_{dd}^2 f_{0 \rightarrow 1} + t_{sc} V_{dd} I_{peak} P_{0 \rightarrow 1} + V_{dd} I_{leakage} \]

\[ P \text{ (watts)} = C_L V_{dd}^2 f_{0 \rightarrow 1} + t_{sc} V_{dd} I_{peak} f_{0 \rightarrow 1} + V_{dd} I_{leakage} \]
Power and Energy Dissipation

- Propagation delay and the power consumption of a gate are related.
- Propagation delay is (mostly) determined by the speed at which a given amount of energy can be stored on the gate capacitors.
  - The faster the energy transfer (higher power dissipation) the faster the gate.
- For a given technology and gate topology, the product of the power consumption and the propagation delay is a constant.
  - Power-delay product (PDP) – energy consumed by the gate per switching event.
- An ideal gate is one that is fast and consumes little energy, so the ultimate quality metric is.
  - Energy-delay product (EDP) = power-delay \(^2\)

Summary

- Digital integrated circuits have come a long way and still have quite some potential left for the coming decades.
- Some interesting challenges ahead.
  - Getting a clear perspective on the challenges and potential solutions is the purpose of this course.
- Understanding the design metrics that govern digital design is crucial.
  - Cost, reliability, speed, power and energy dissipation.