EE 476 VLSI I

Introduction, Design Metrics

What will you learn?

- Understanding, designing, and optimizing digital circuits. How?

- with respect to different quality metrics:

IC Evolution

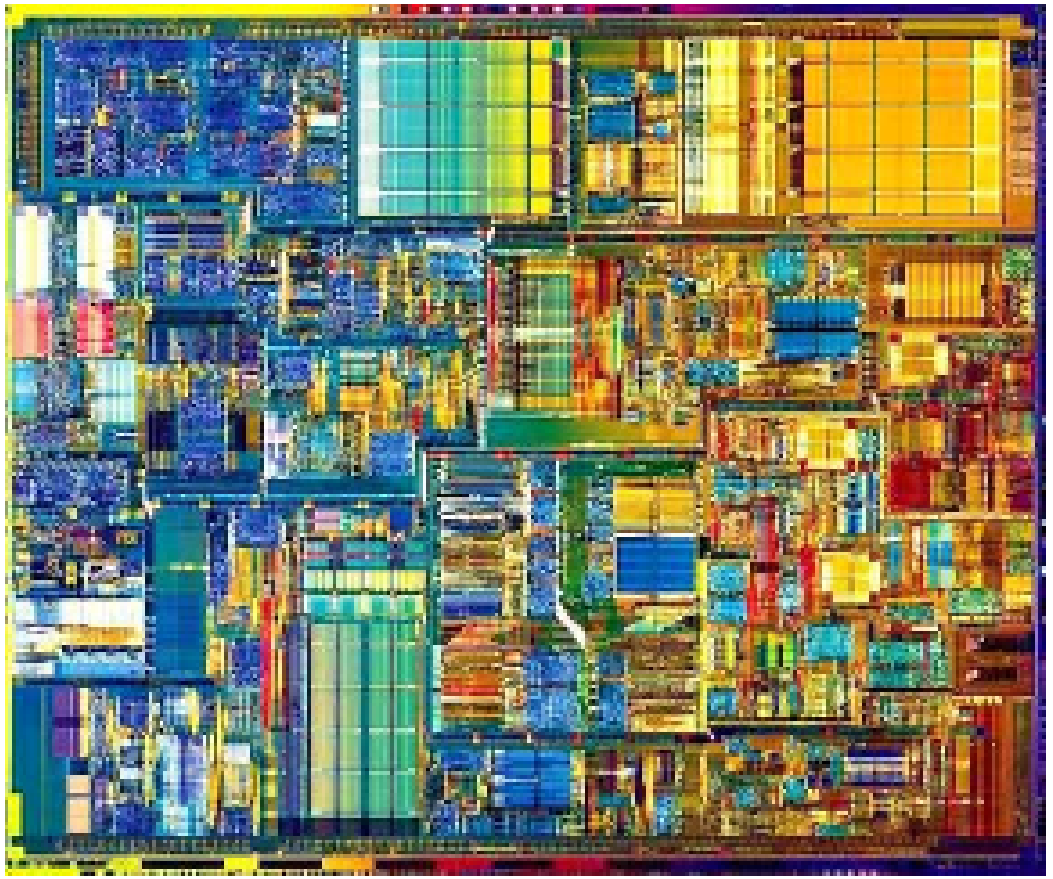
- SSI – Small Scale Integration (early 1970s)
  - contained 1 – 10 logic gates
- MSI – Medium Scale Integration
  - logic functions, counters
- LSI – Large Scale Integration
  - first microprocessors on the chip
- VLSI – Very Large Scale Integration
  - now offers 64-bit microprocessors, complete with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

IC Evolution

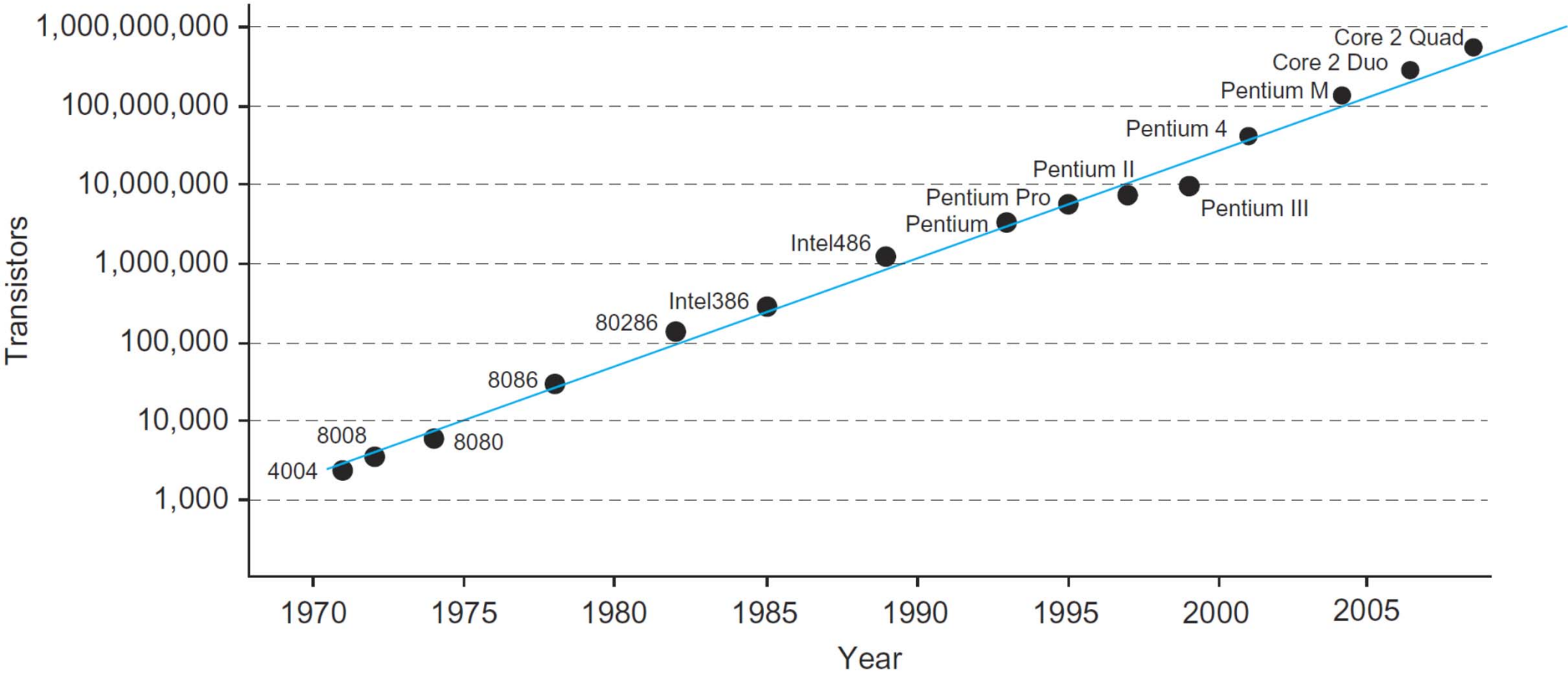
- Bipolar technology
  - TTL (transistor-transistor logic), 1962;
  - ECL (emitter-coupled logic), 1974;
- MOS (Metal-oxide-silicon)
  - although invented before bipolar transistor (1925, 1935), was initially difficult to manufacture
  - nMOS (n-channel MOS) technology developed in late 1970s required 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

Pentium 4

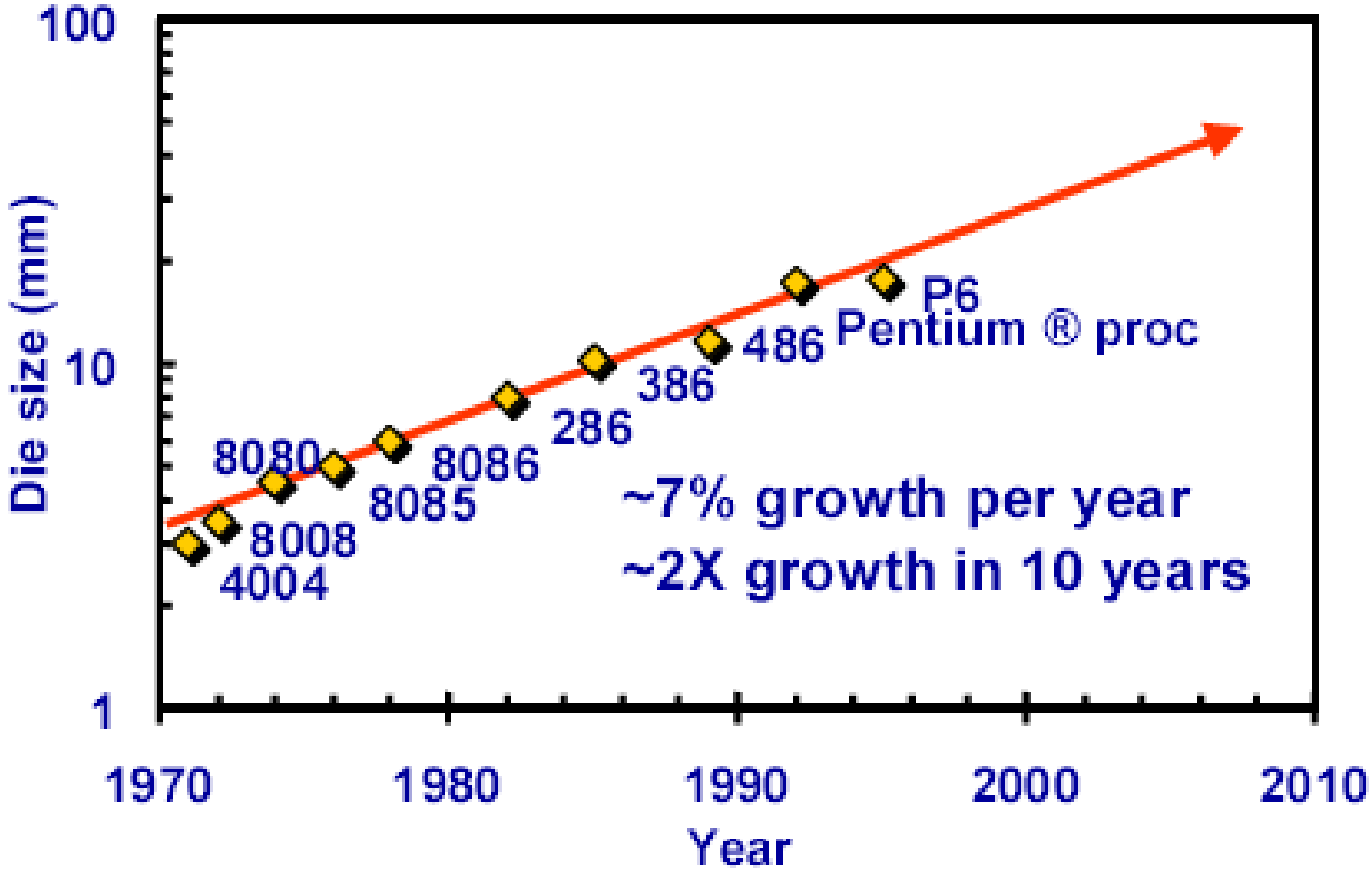
- **0.18-micron process technology (2, 1.9, 1.8, 1.7, 1.6, 1.5, and 1.4 GHz)**
  - Introduction date: August 27, 2001 (2, 1.9 GHz); …; November 20, 2000 (1.5, 1.4 GHz)
  - Level Two cache: 256 KB Advanced Transfer Cache (Integrated)
  - System Bus Speed: 400 MHz
  - SSE2 SIMD Extensions
  - Transistors: 42 Million
  - Typical Use: Desktops and entry-level workstations
- **0.13-micron process technology (2.53, 2.2, 2 GHz)**
  - Introduction date: January 7, 2002
  - Level Two cache: 512 KB Advanced
  - Transistors: 55 Million



Transistors in Intel microprocessors [Intel]



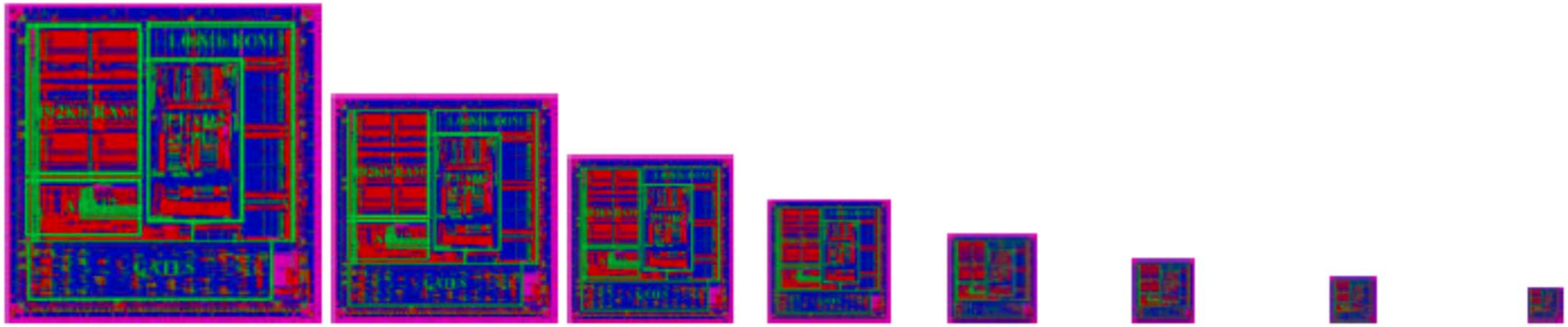
Die Size Growth



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

Courtesy, Intel

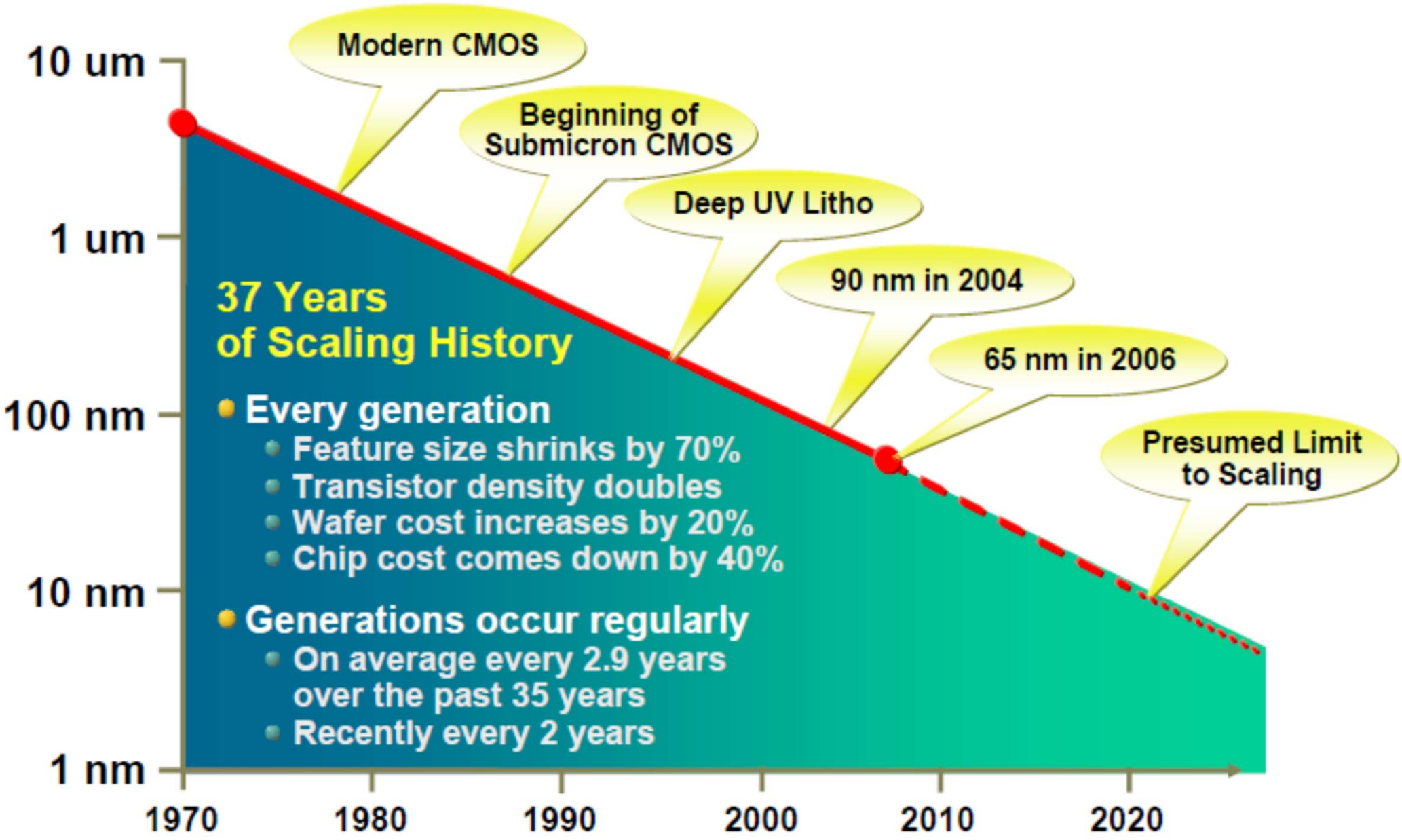
GSM Digital Baseband Evolution



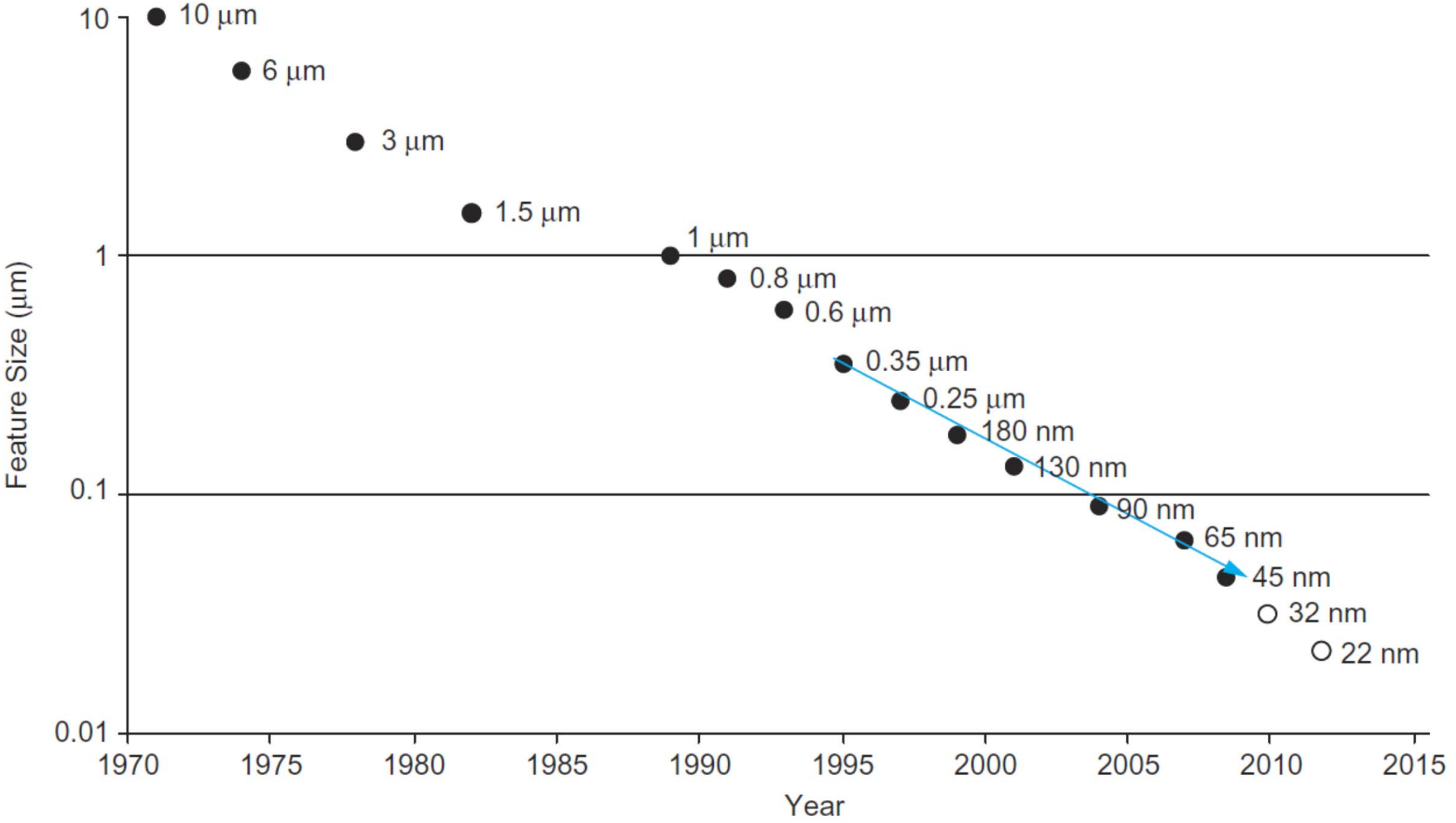
Year	1994	1997	1999	2000	2002	2004	2006	2008
Nano-meter	500nm	350nm	250nm	180nm	130nm	90nm	65nm	45nm
Wafer size	6"	8"	8"	8"	12"	12"	12"	12"
Die size (mm^2)	80.7	46.6	19.2	10.7	6.7	4.2	2.4	1.4
Dies per wafer	310	950	2550	4700	12,200	18,700	26,500	46,500

150X increase in die per wafer

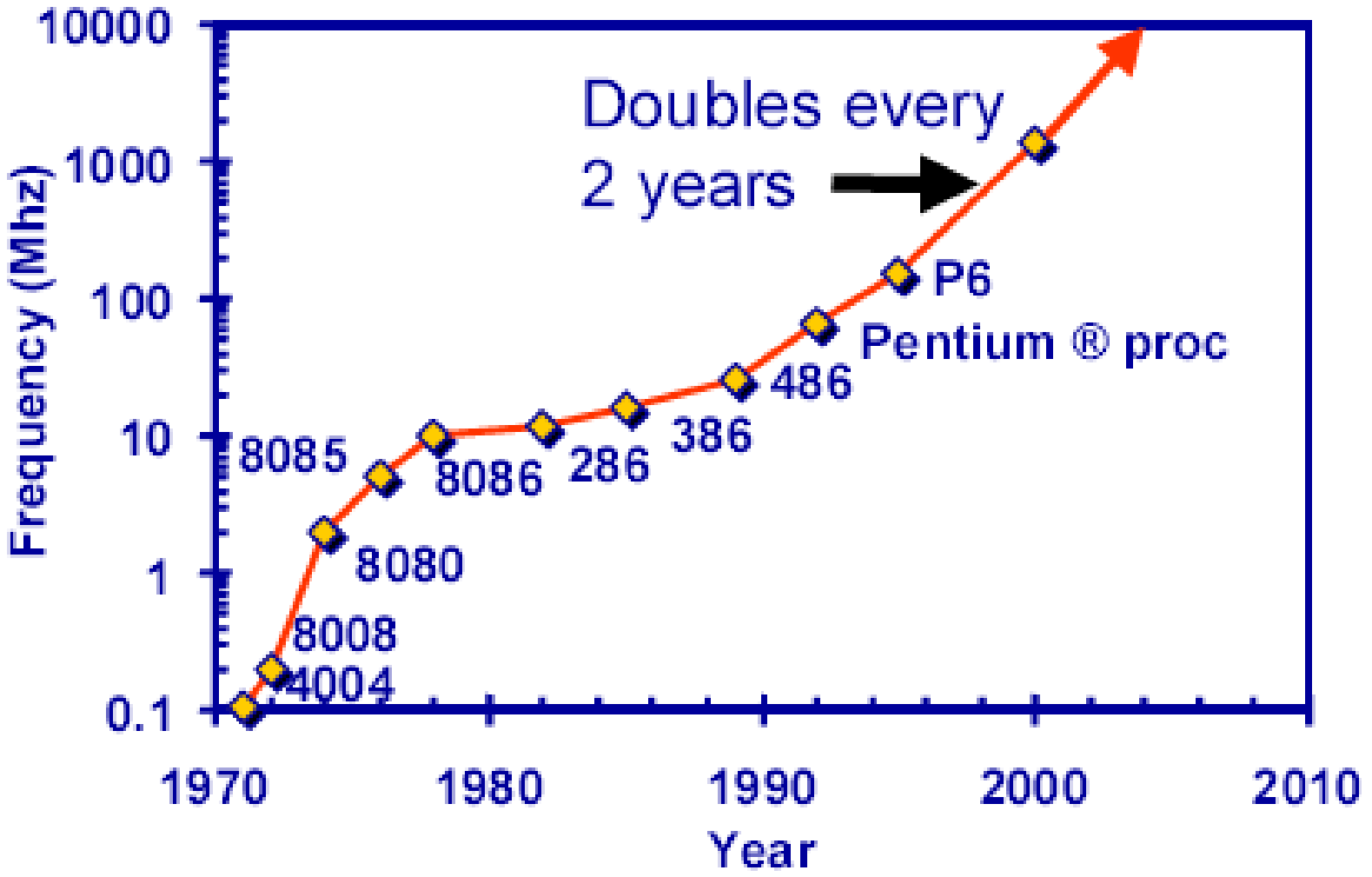
Semi-Conductor Scaling



Process generations. Future predictions from [SIA2007]



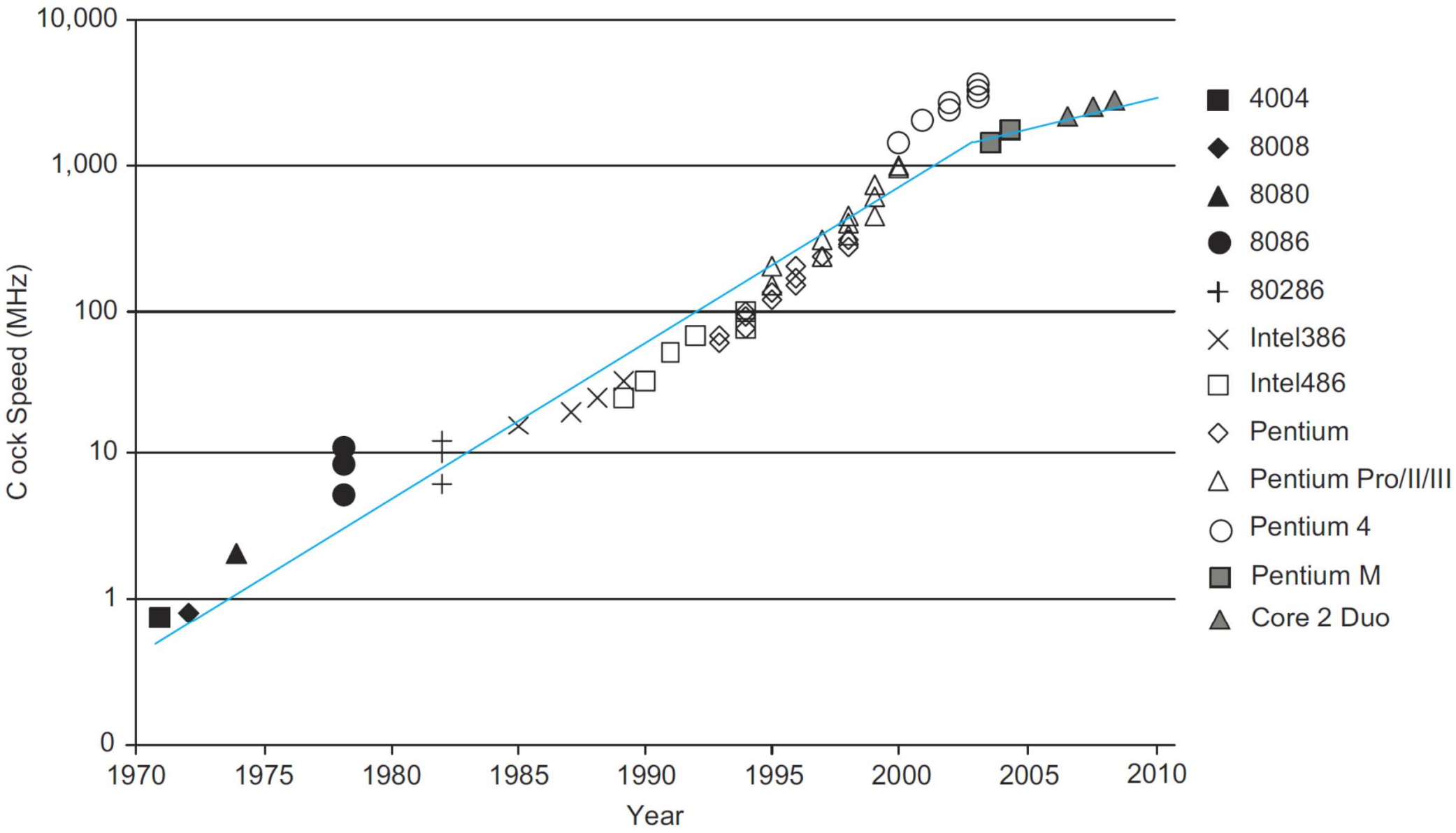
Frequency



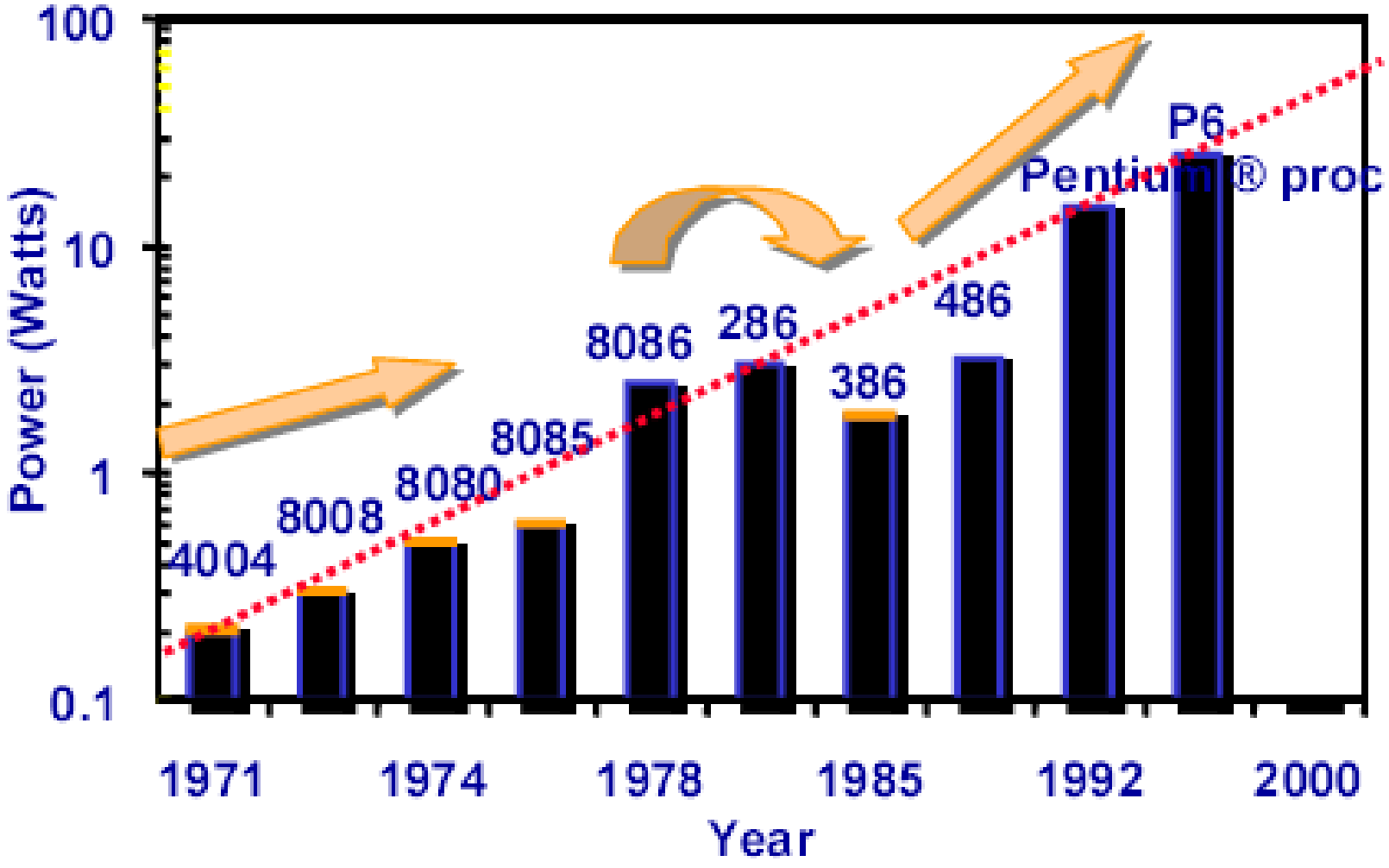
Lead Microprocessors frequency doubles every 2 years

Courtesy, Intel

Clock frequencies of Intel microprocessors



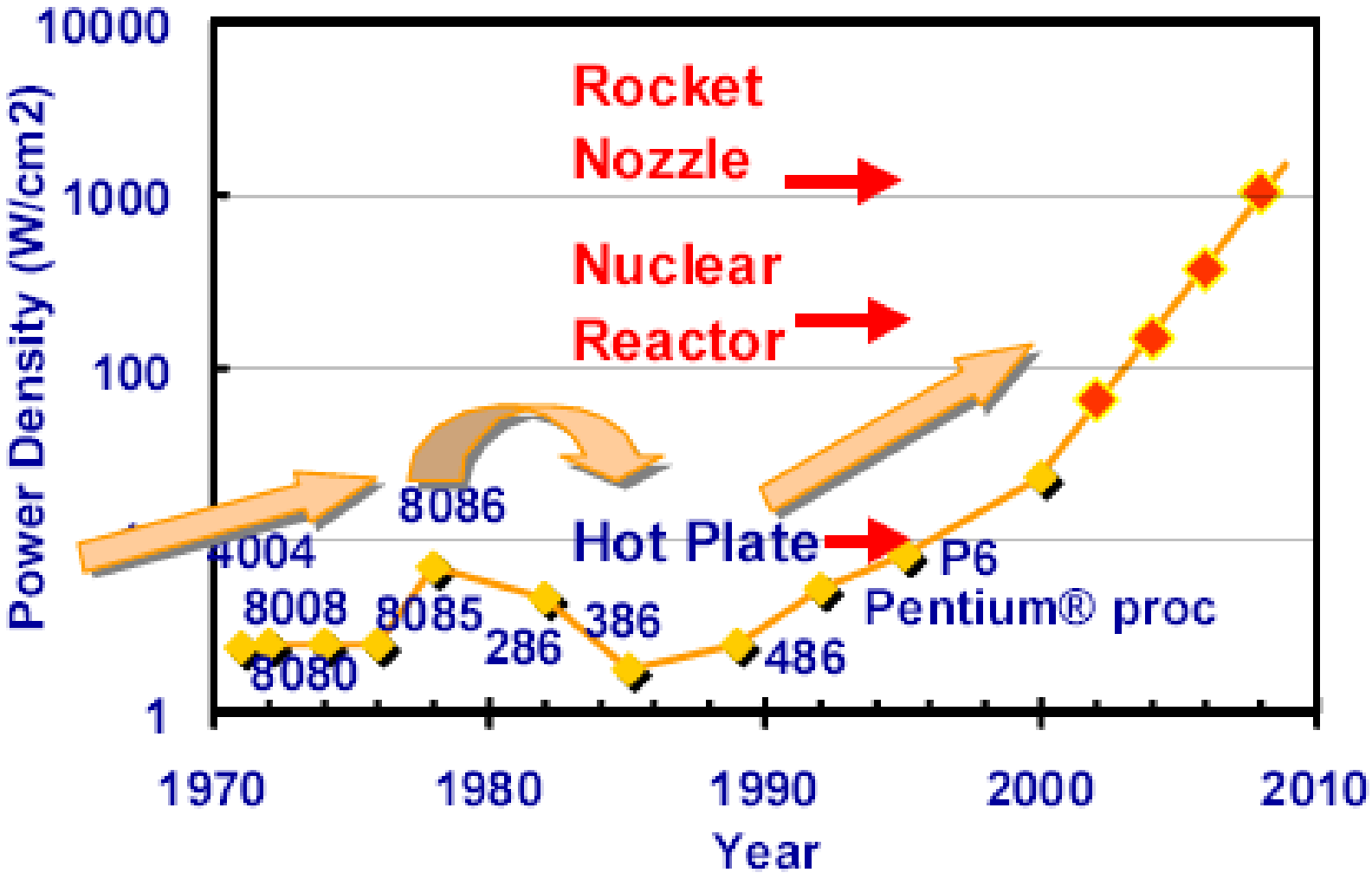
Power Dissipation



Lead Microprocessors power continues to increase

Courtesy, Intel

Power density



Power density too high to keep junctions at low temp

Courtesy, Intel

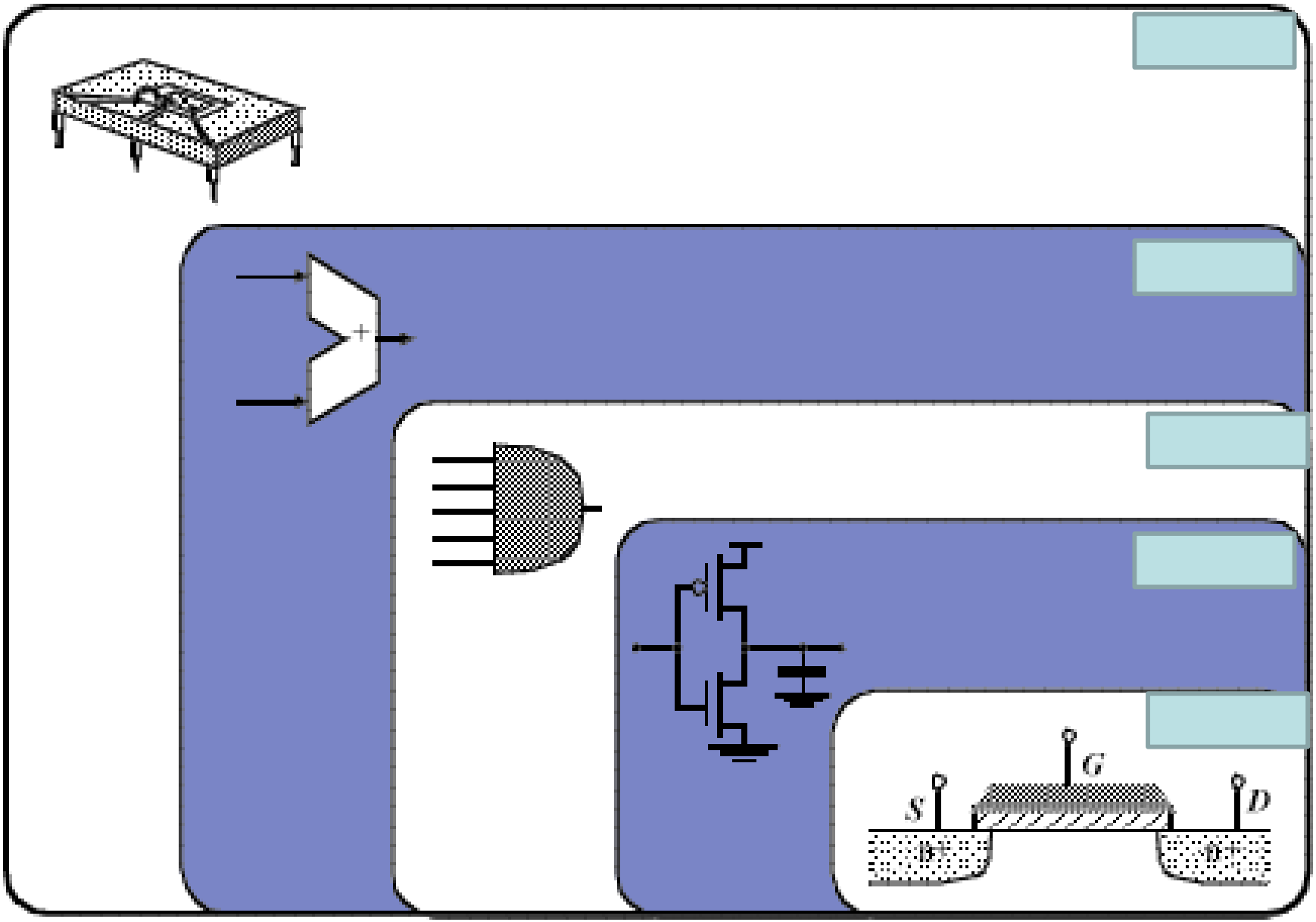
Technology Directions: SIA Roadmap

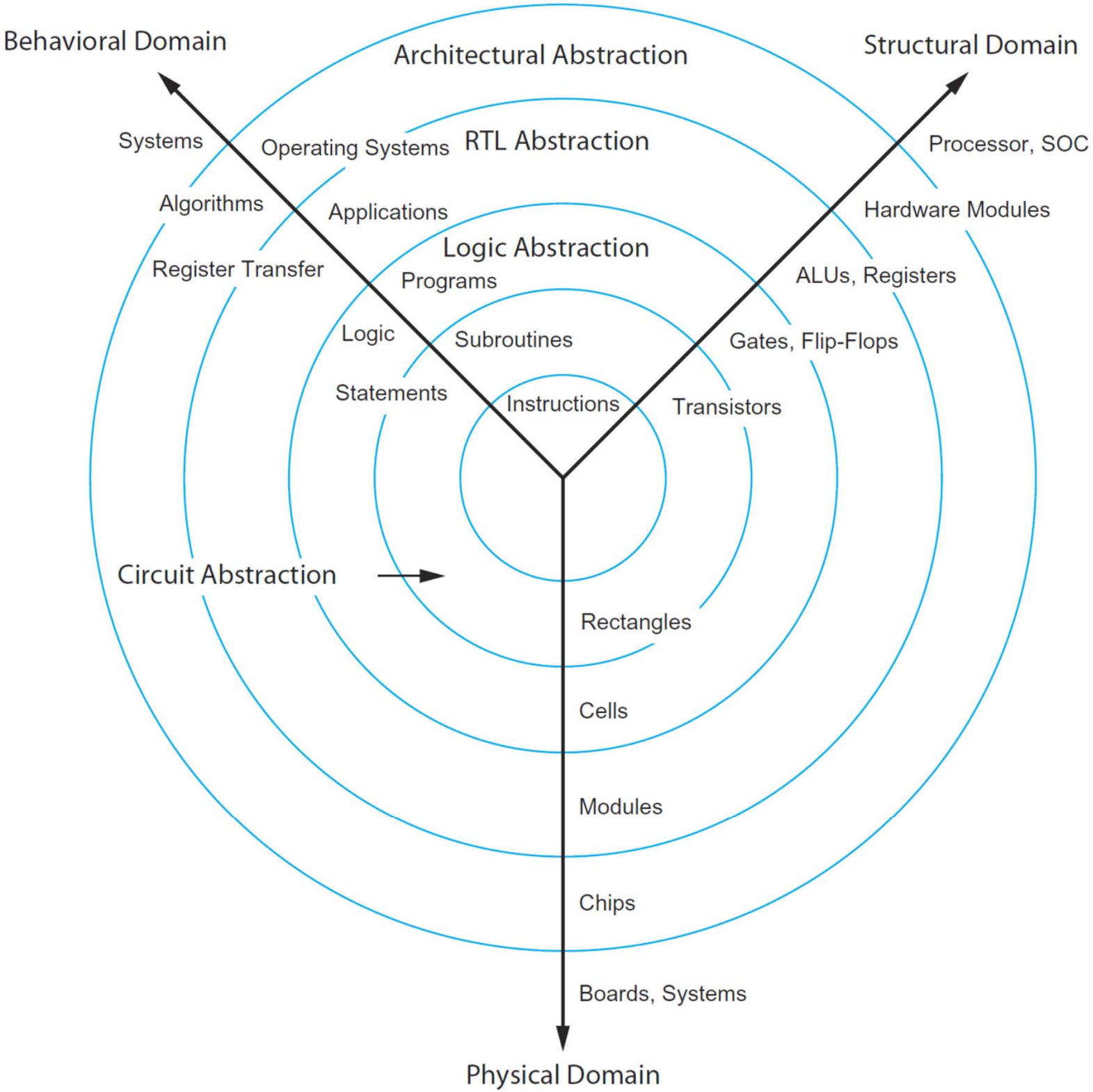
Year	1999	2002	2005	2008	2011	2014
Feature size (nm)	180	130	100	70	50	35
Logic trans/cm^2	6.2M	18M	39M	84M	180M	390M
Cost/trans (mc)	1.735	.580	.255	.110	.049	.022
#pads/chip	1867	2553	3492	4776	6532	8935
Clock (MHz)	1250	2100	3500	6000	10000	16900
Chip size (mm^2)	340	430	520	620	750	900
Wiring levels	6-7	7	7-8	8-9	9	10
Power supply (V)	1.8	1.5	1.2	0.9	0.6	0.5
High-perf pow (W)	90	130	160	170	175	183

Why Scaling?

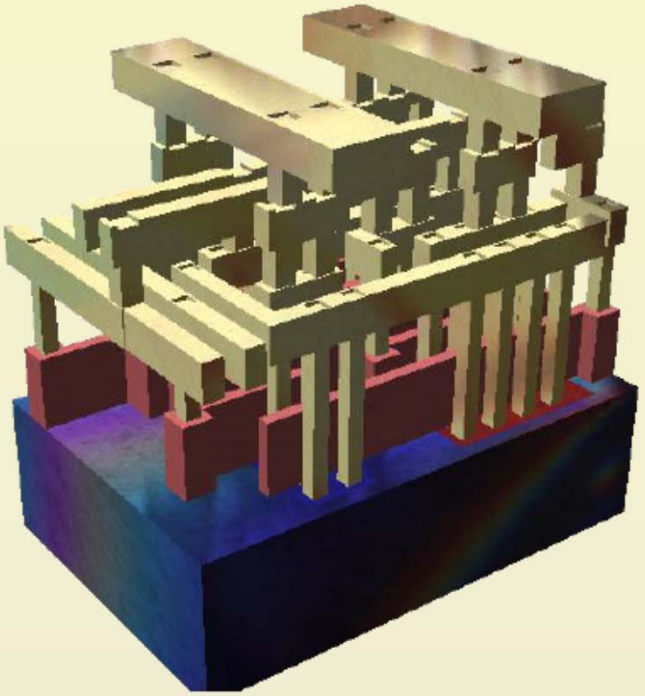
- Technology shrinks by 0.7/generation
- With every generation can integrate 2x more functions per chip; chip cost does not increase significantly
- Cost of a function decreases by 2x
- But …
  - How to design chips with more and more functions?
  - Design engineering population does not double every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

Design Abstraction Levels




Behavioral Domain
Architectural Abstraction
Structural Domain
Systems
Operating Systems
RTL Abstraction
Processor, SOC
Algorithms
Applications
Hardware Modules
Logic Abstraction
Register Transfer
Programs
ALUs, Registers
Logic
Subroutines
Gates, Flip-Flops
Statements
Instructions
Transistors
Circuit Abstraction
Rectangles
Cells
Modules
Chips
Boards, Systems
Physical Domain

General overview of the design hierarchy



System Specification

↓

Architectural Design

↓

Logic Design

↓

Circuit Design

↓

Physical Design

↓

Physical Verification and Signoff

↓

Fabrication

↓

Packaging and Testing

↓

CHIP

Chip planning

↓

Placement

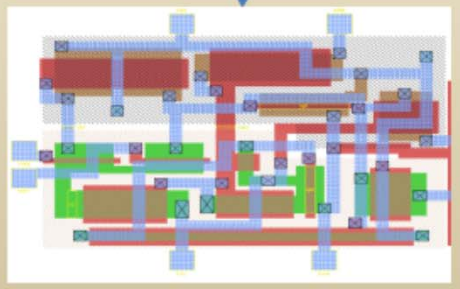
↓

Signal Routing

↓

Verification

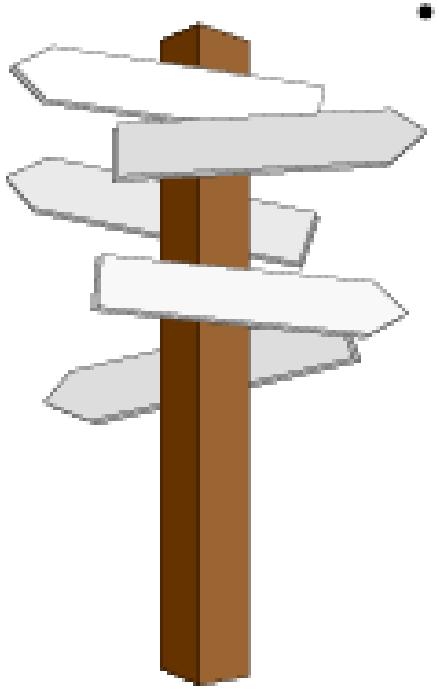
↓



The LAYOUT of an IC defines the geometry of the masks used in fabrication.

Major Design Challenges

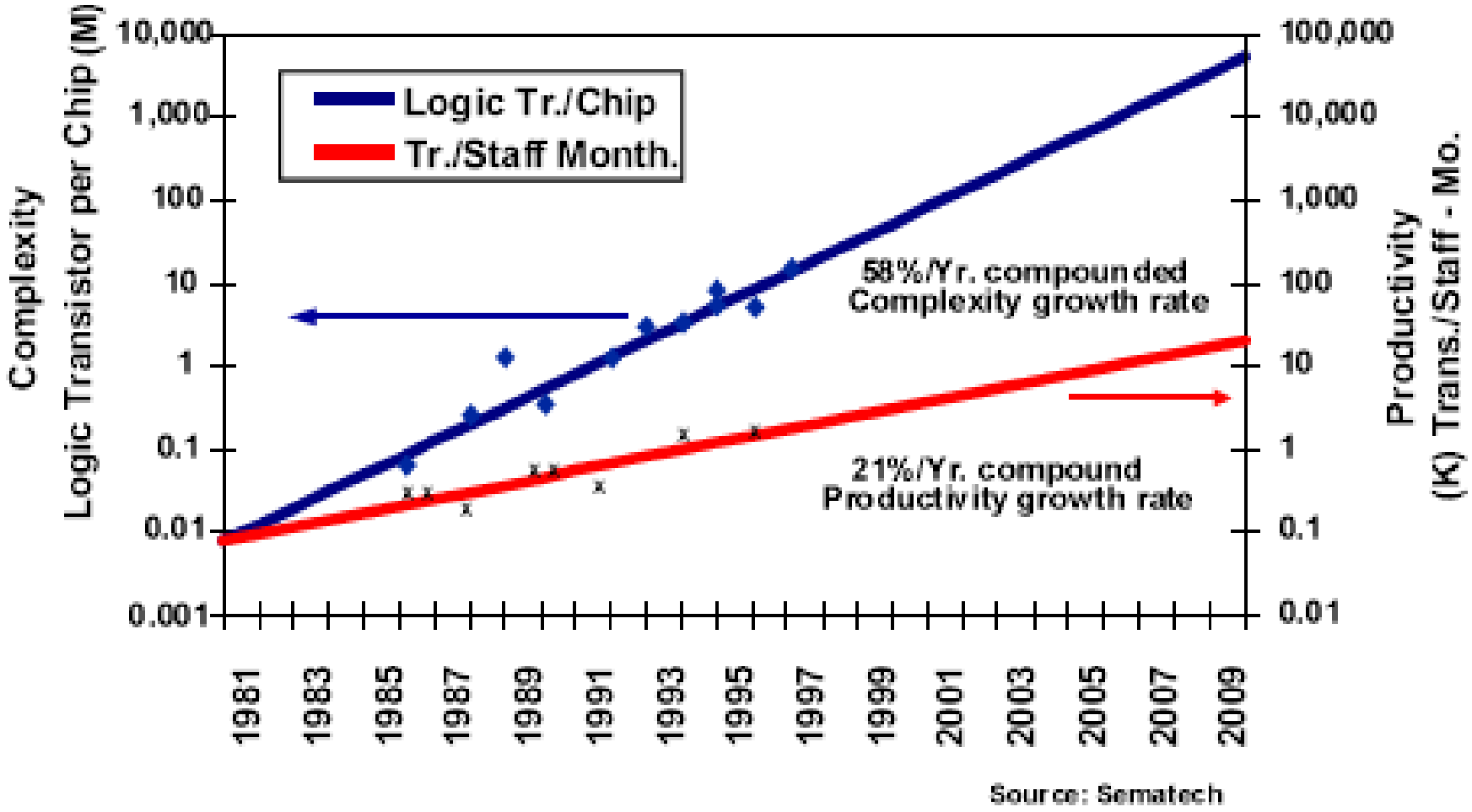
- Microscopic issues
  - ultra-high speeds
  - power dissipation and supply rail drop
  - growing importance of interconnect
  - noise, crosstalk
  - reliability, manufacturability
  - clock distribution



- 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

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

Productivity Trends



Complexity outpaces design productivity

Courtesy, ITRS Roadmap

Fundamental Design Metrics

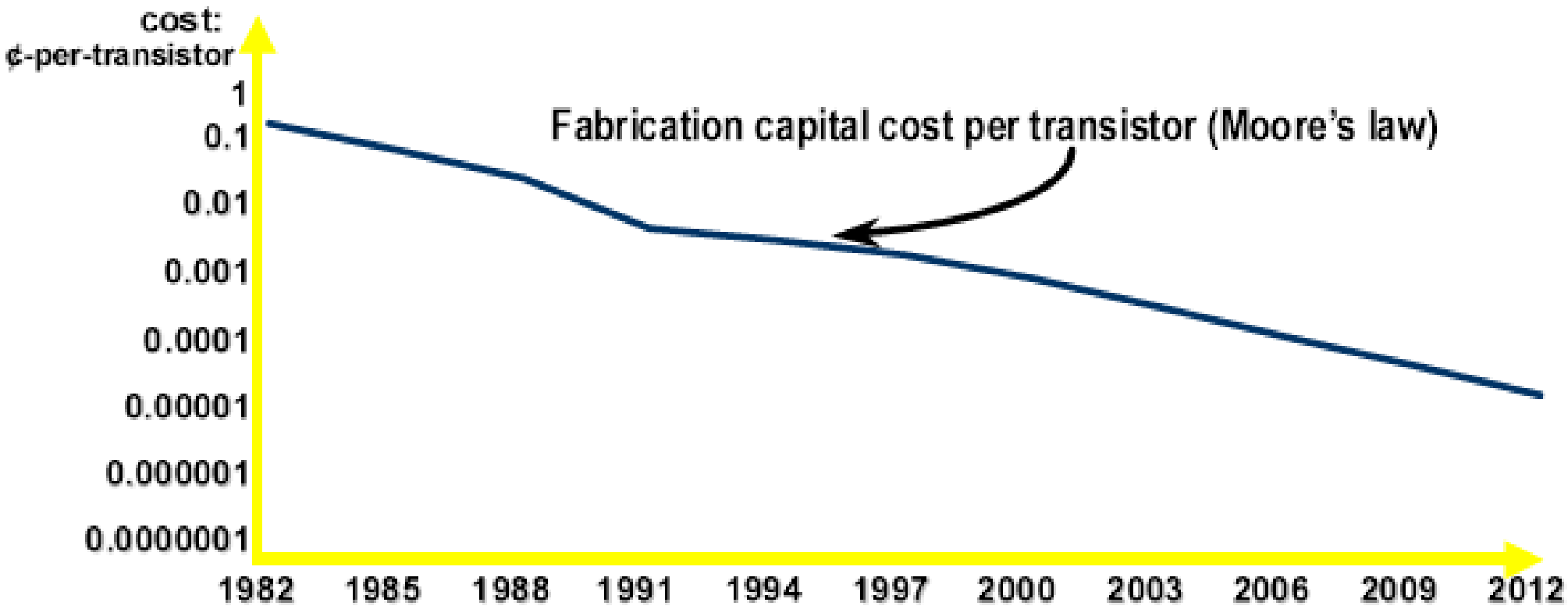
- Functionality
- Cost
  - NRE (fixed) costs - design effort
  - RE (variable) costs - cost of parts, assembly, test
- Reliability, robustness
  -
  -
- Performance
  -
  -
- Time-to-market

Cost of Integrated Circuits

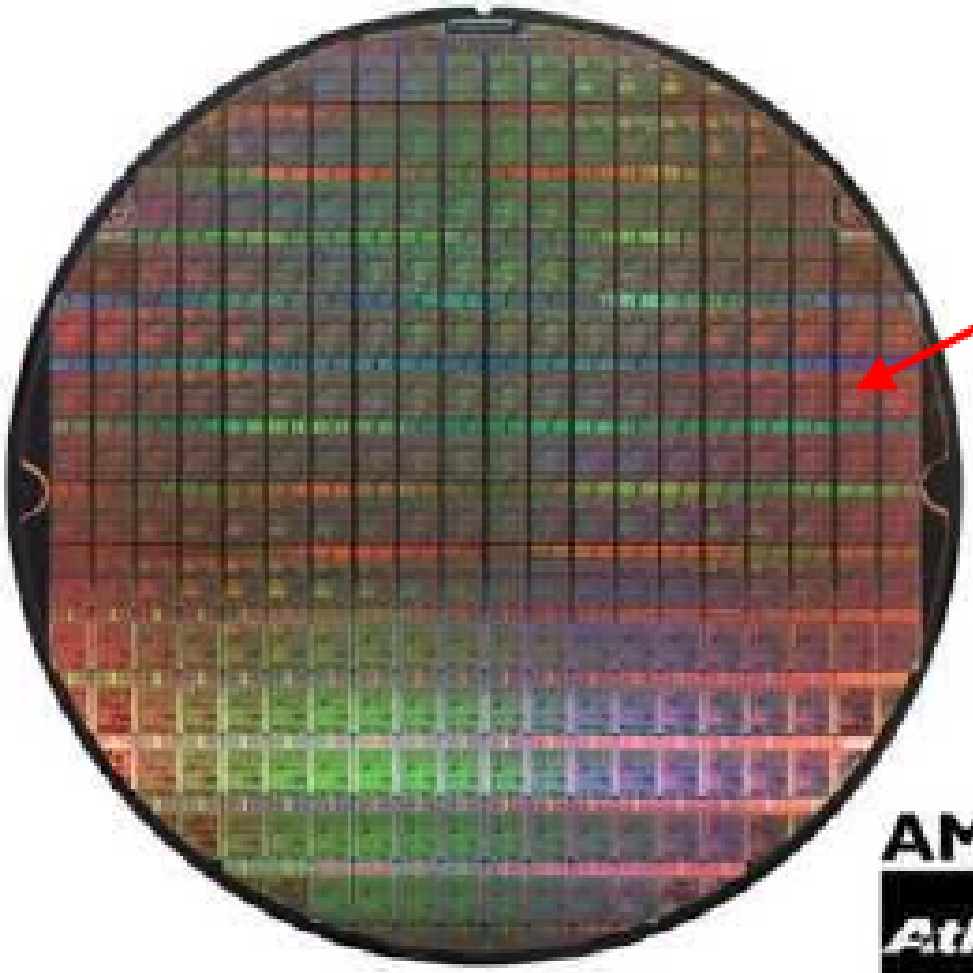
- NRE (non-recurring engineering) costs
  - Fixed cost to produce the design
    - design effort
    - design verification effort
    - mask generation
  - Influenced by the design complexity and designer productivity
  - More pronounced for small volume products
- Recurring costs – proportional to product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

$$\text{Cost per IC} = \text{Variable cost per IC} + \frac{\text{Fixed cost}}{\text{Volume}}$$

Cost per Transistor



Silicon Wafer



Single die

Wafer

AMD Athlon

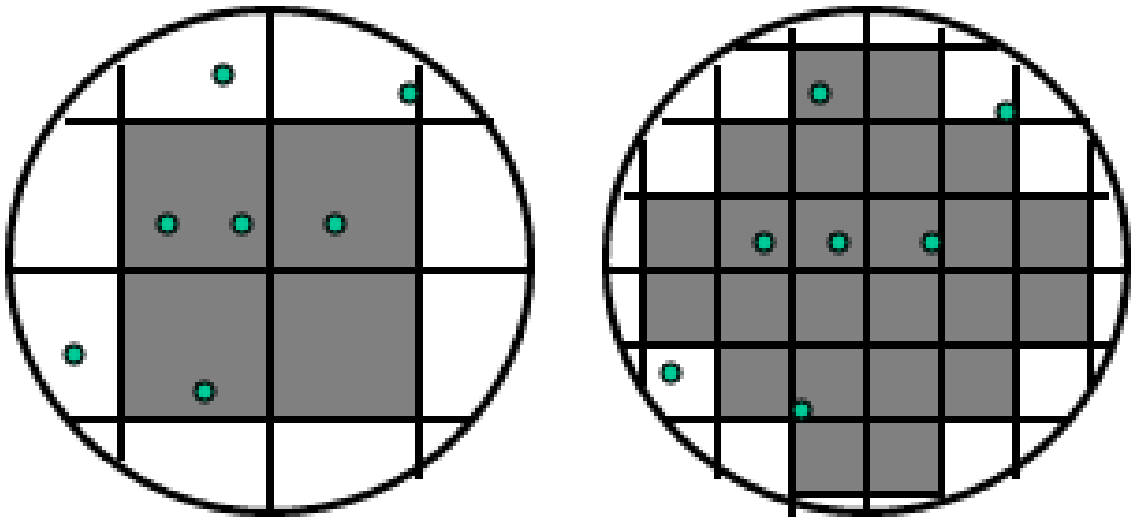
From http://www.amd.com

Going up to 12″ (30cm)

Recurring Costs

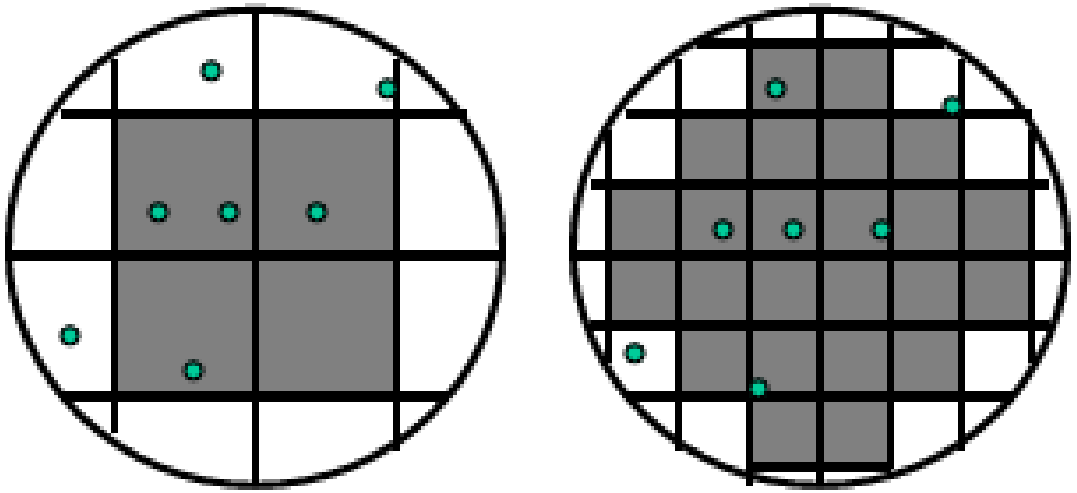
$$\text{Variable cost} = \frac{\text{Die cost} + \text{Testing cost} + \text{Packaging cost}}{\text{Final test yield}}$$

$$\text{Cost of die} = \frac{\text{Cost of wafer}}{\text{Dies per wafer} \times \text{Die yield}}$$

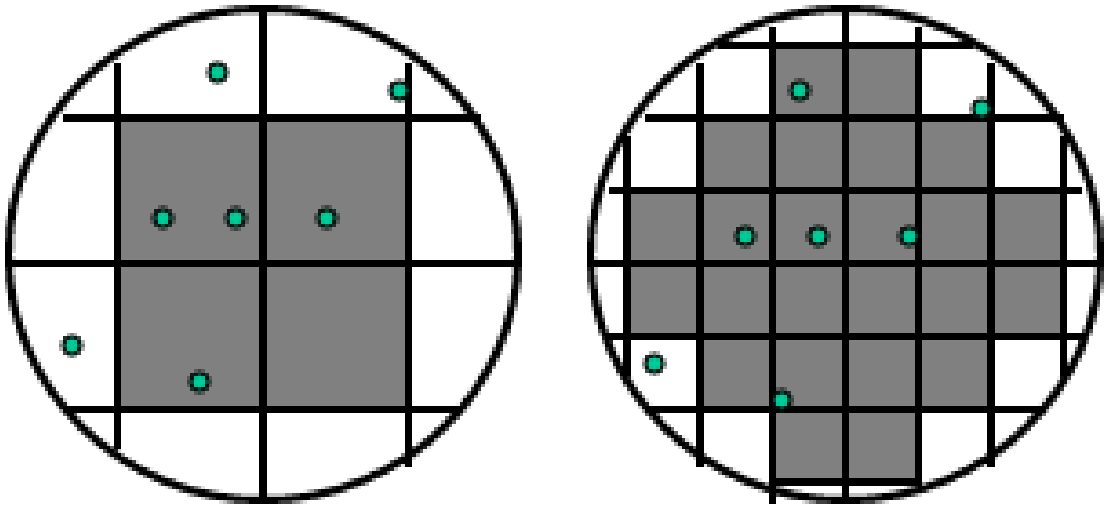


Dies per Wafer

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



Yield



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

α is approximately 3

$$\text{die cost} = f(\text{die area})^4$$

α is a parameter → complexity of the manufacturing process, ~ the number of masks

Example: Die Yield

Assume a wafer size of 12 inch, a die size of 2.5cm^2, 1 defects/cm^2, and $\alpha = 3$. Determine the die yield of this CMOS process run.

$$\text{dies per wafer} = \frac{\pi \times (\text{wafer diameter}/2)^2}{\text{die area}} - \frac{\pi \times \text{wafer diameter}}{\sqrt{2 \times \text{die area}}}$$

$$= 296 - 44 = 252$$

$$\text{die yield} = \left(1 + \frac{\text{defects per unit area} \times \text{die area}}{\alpha}\right)^{-\alpha} = 16\%$$

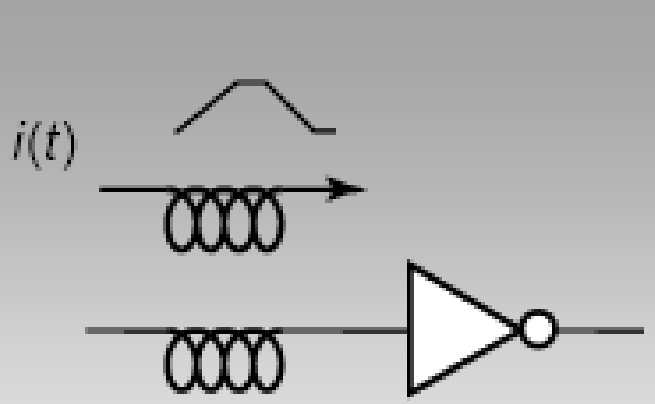
Examples of Cost Metrics (1994)

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

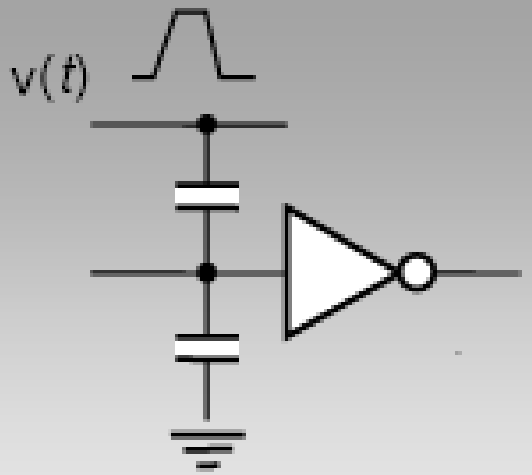
Functionality and Robustness

- 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

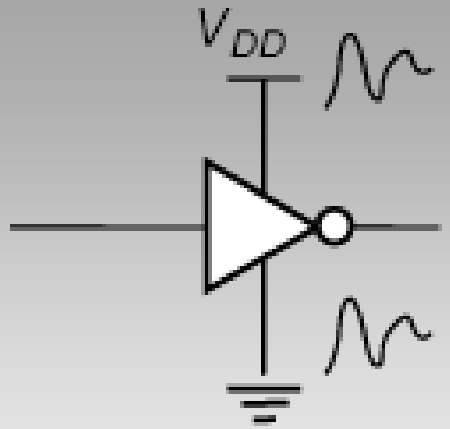
Reliability
Noise in Digital Integrated Circuits



Inductive coupling



Capacitive coupling



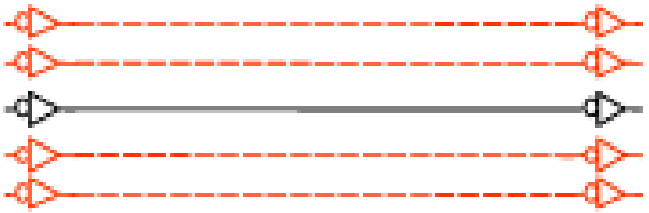
Power and ground noise

- from two wires placed side by side
  - inductive coupling
    - [redacted]
  - capacitive coupling
    - [redacted]
    - 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

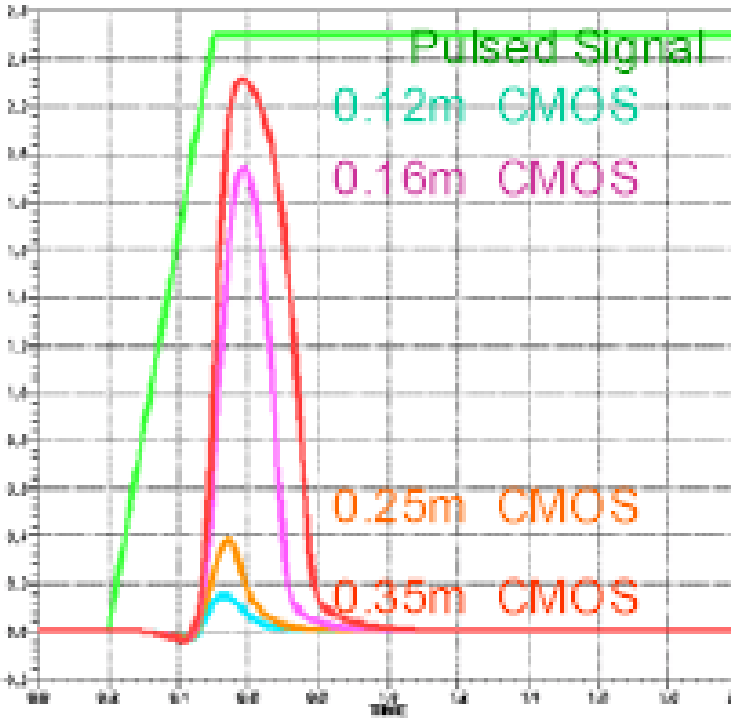
Crosstalk vs. Technology



Black line quiet

Red lines pulsed

Glitches strength vs technology



From Dunlop, Lucent, 2000

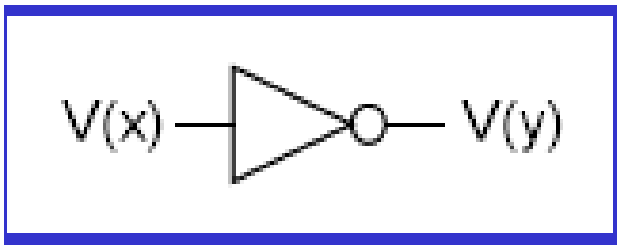
Static Gate Behavior

- 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

$$1 \Leftrightarrow V_{OH} \text{ and } 0 \Leftrightarrow V_{OL}$$

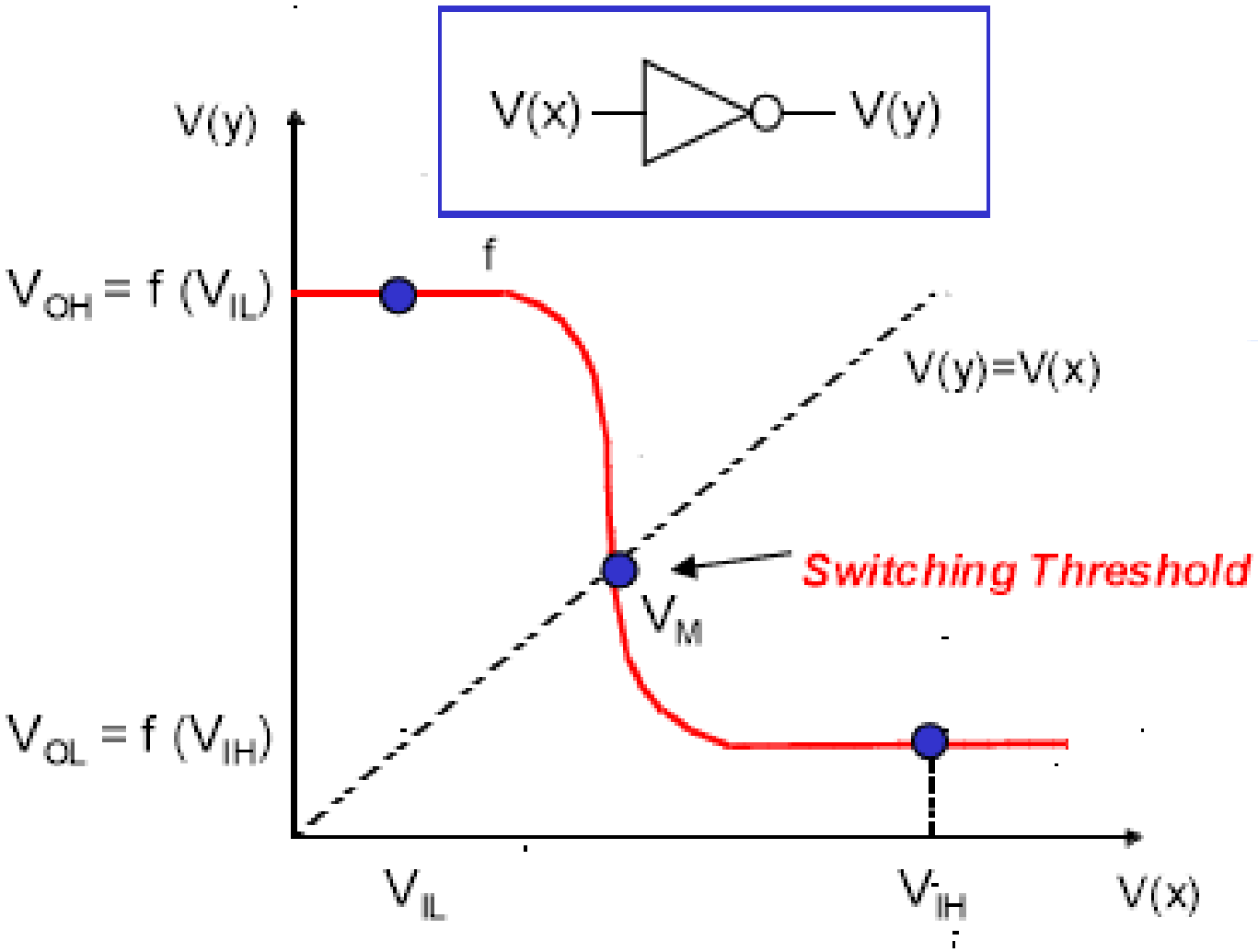


$$V_{OH} = !\,(V_{OL})$$
$$V_{OL} = !\,(V_{OH})$$

- Difference between V_{OH} and V_{OL} is the logic or *signal swing* V_{sw}

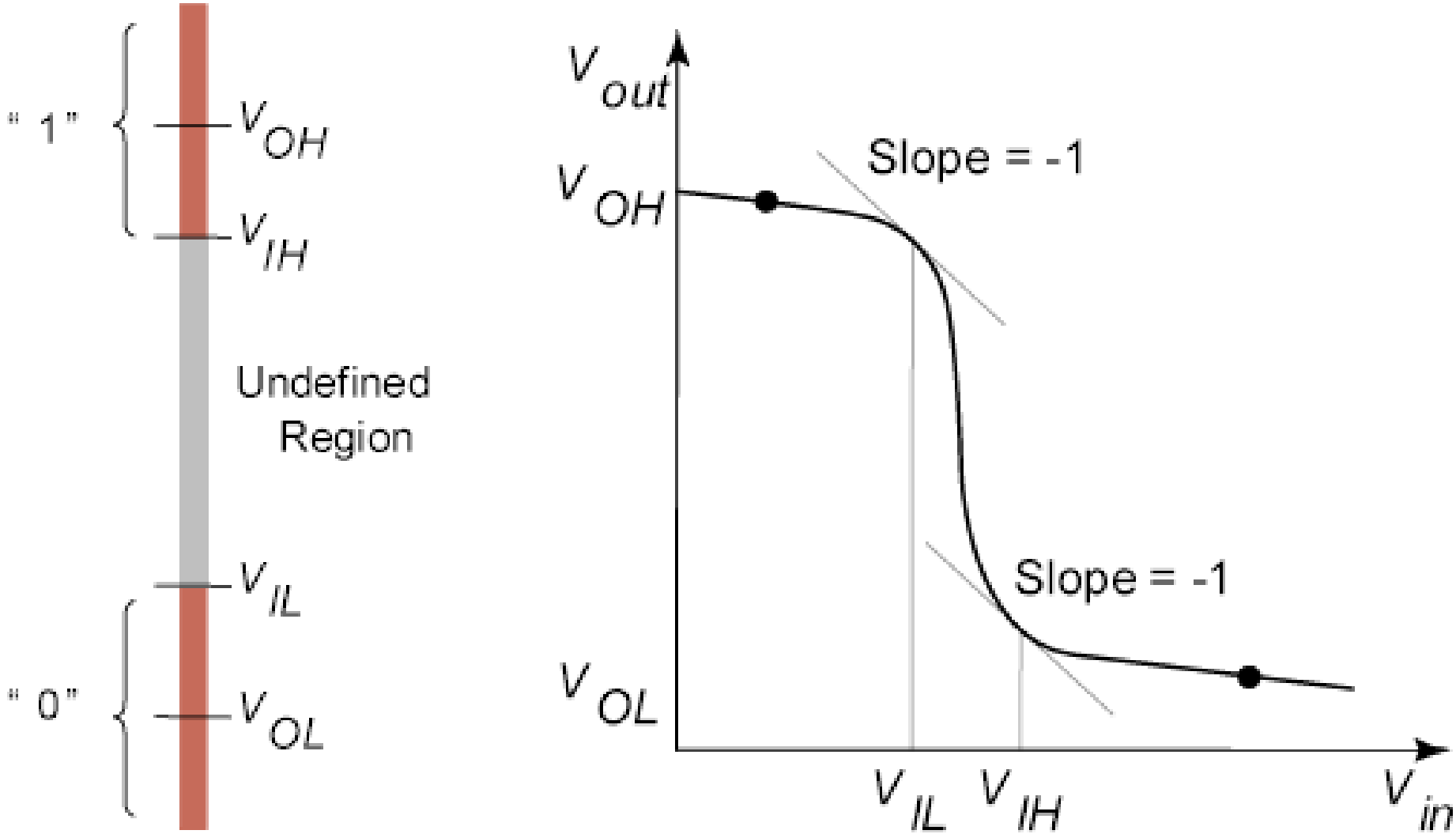
DC Operation

Voltage Transfer Characteristic



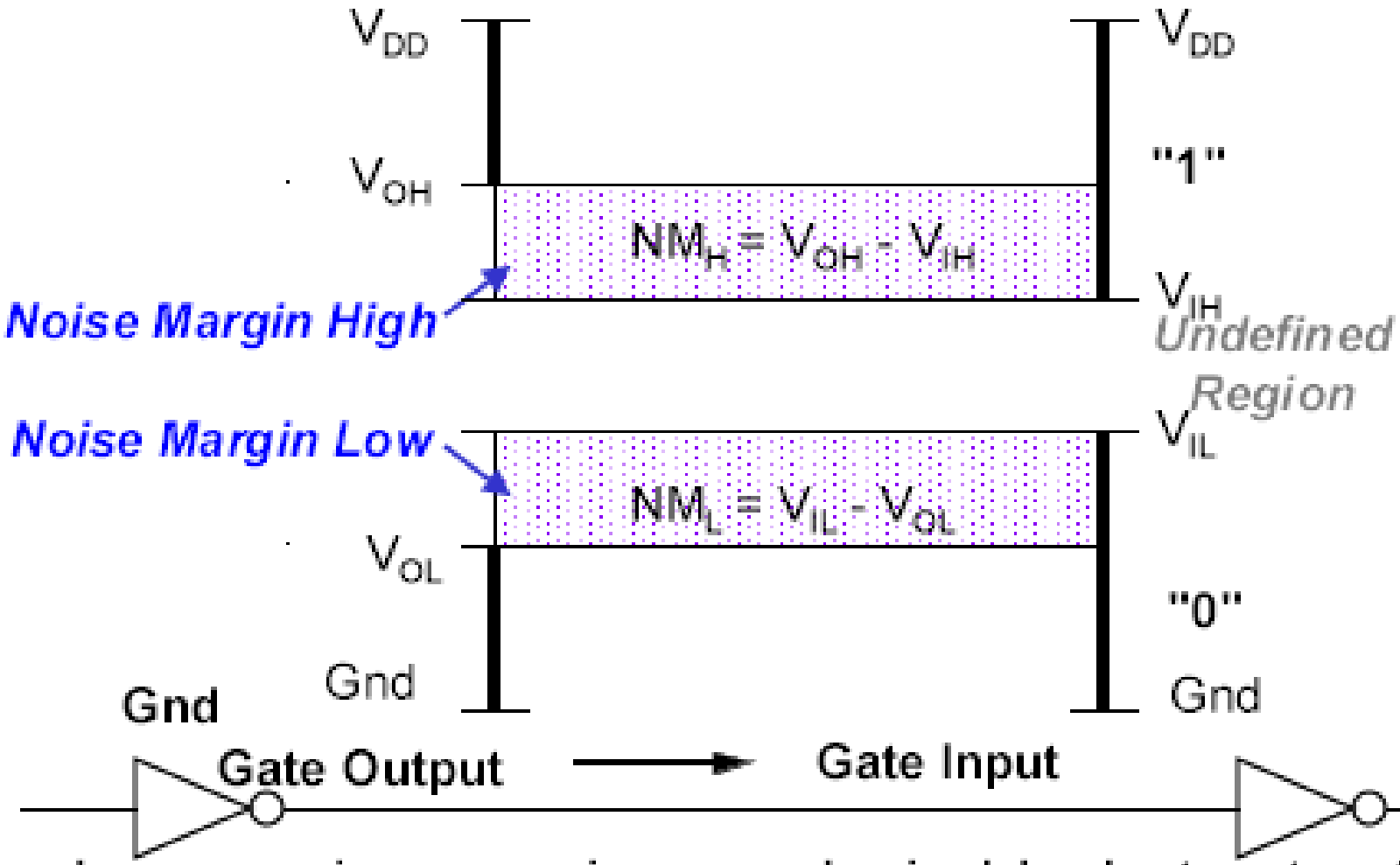
Mapping between analog and digital signals

- The regions of acceptable high and low voltages are delimited by VIH and VIL that represent the points on the VTC curve where the gain = -1 (dVout/dVin)



Definition of Noise Margins

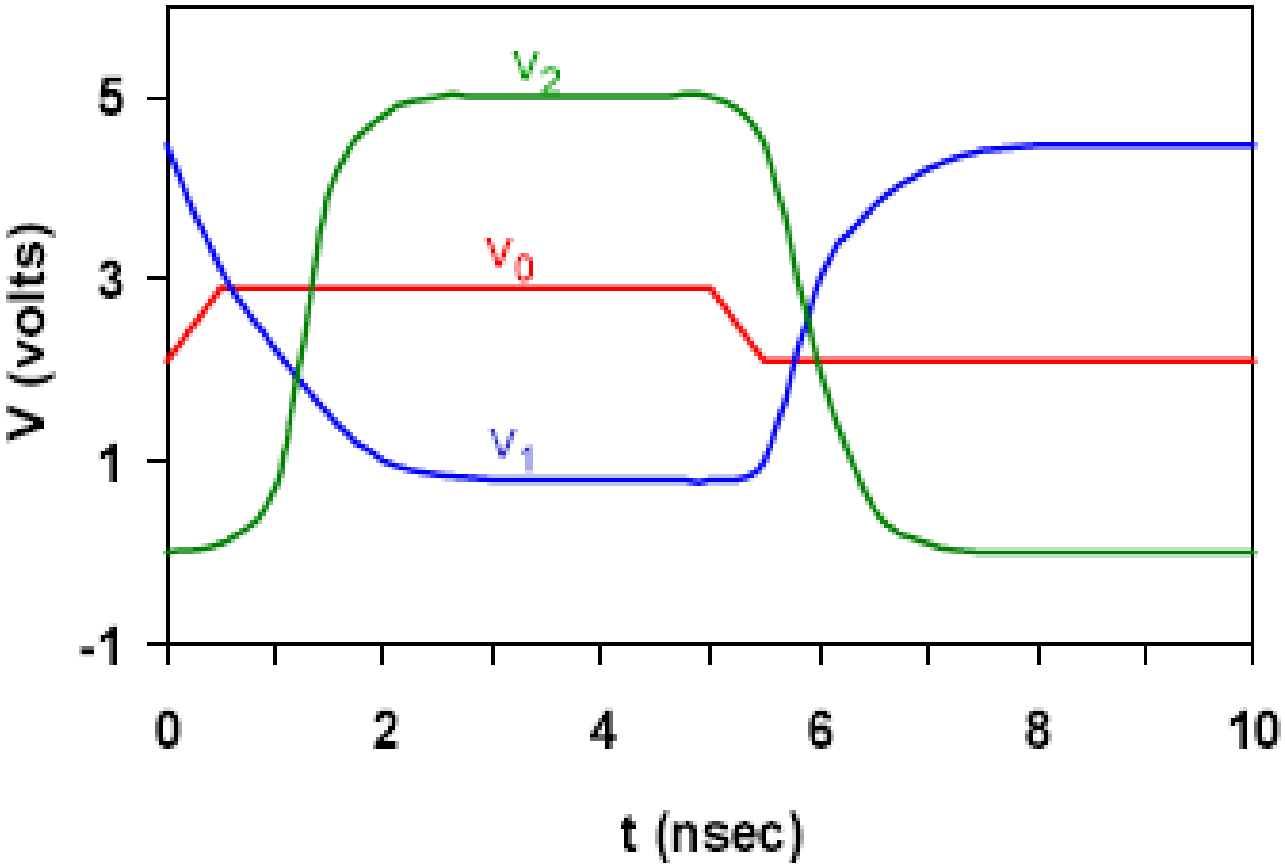
- For robust circuits, want the "0" and "1" intervals to be as large as possible



- Large noise margins are desirable, but not sufficient …

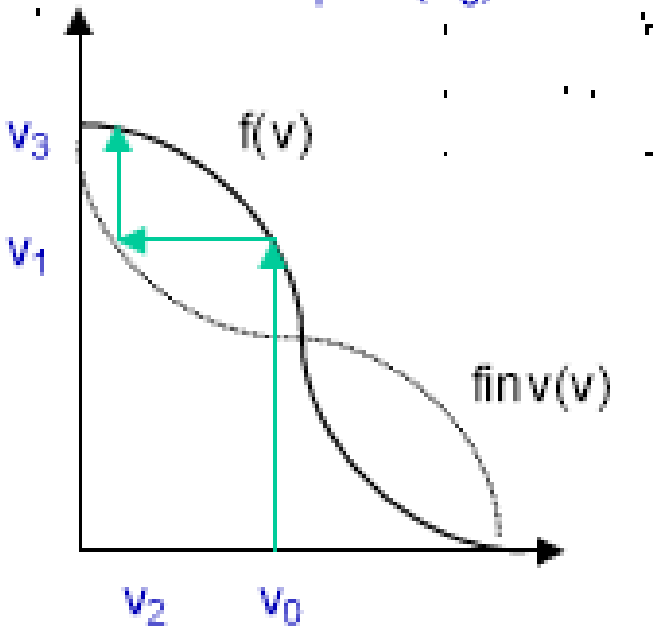
The Regenerative Property

A gate with regenerative property ensure that a disturbed signal converges back to a nominal voltage level

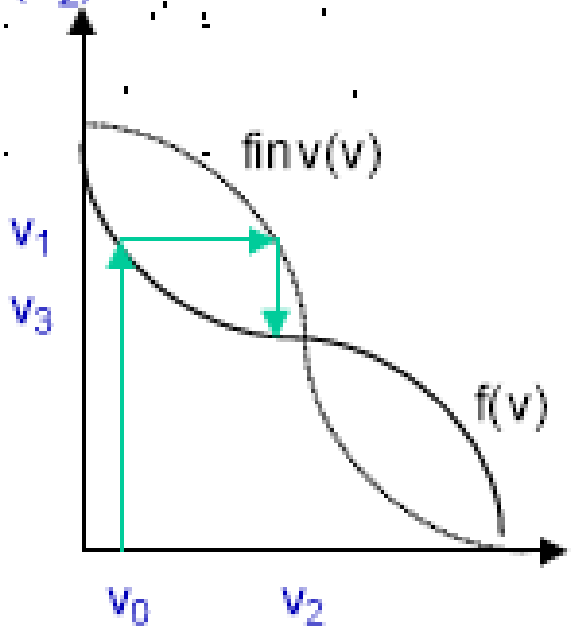


Conditions for Regeneration

$V_1 = f(V_0) \Rightarrow V_1 = finv(V_2)$



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

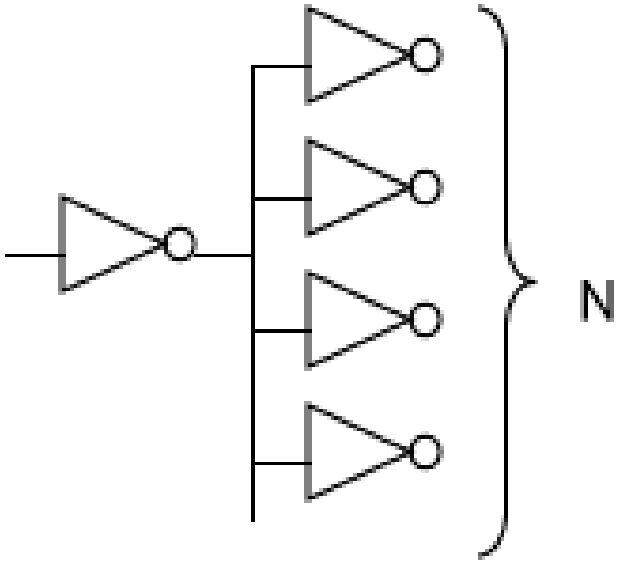
Directivity

- A gate must be unidirectional: 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)

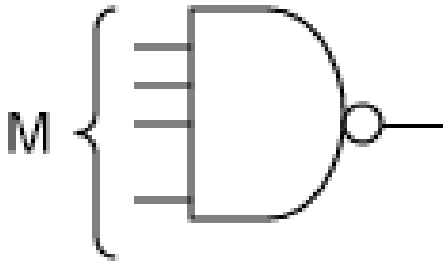
- Key metrics: *output impedance* of the driver and *input impedance* of the receiver
  - ideally, the output impedance of the driver should be zero
  - input impedance of the receiver should be infinity

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

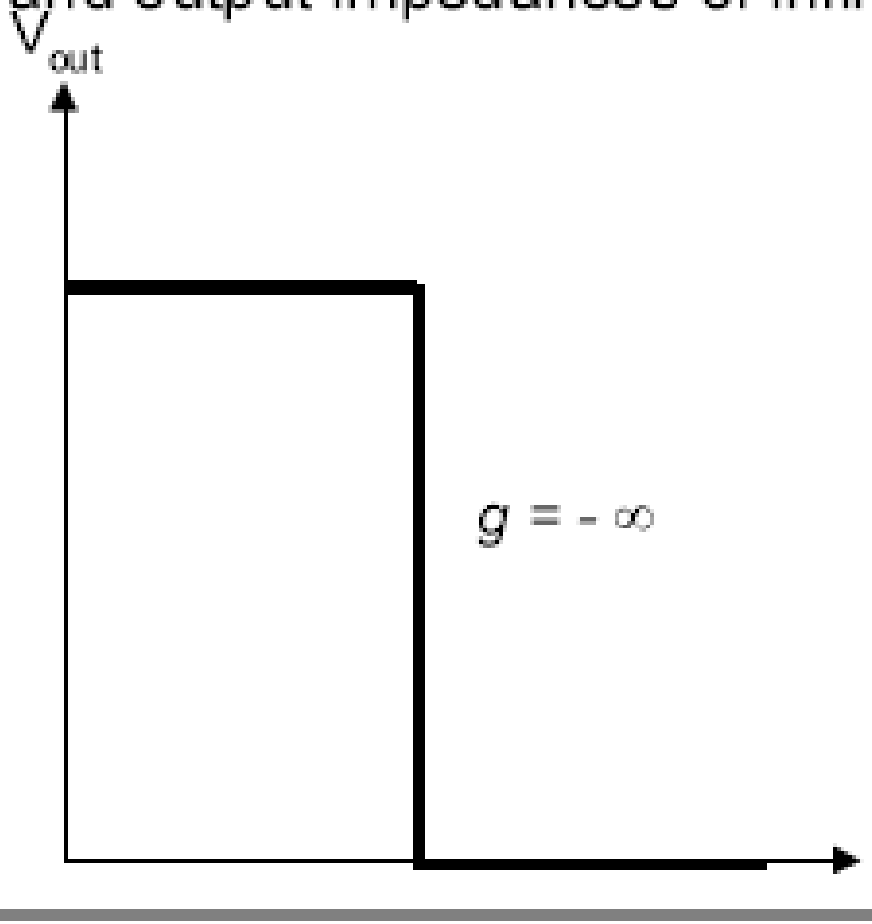


- Fan-in – the number of inputs to the gate
  - gates with large fan-in are bigger and slower



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.



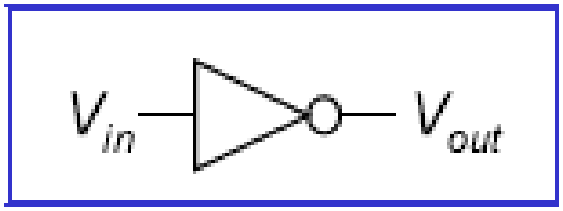
$R_i = \infty$

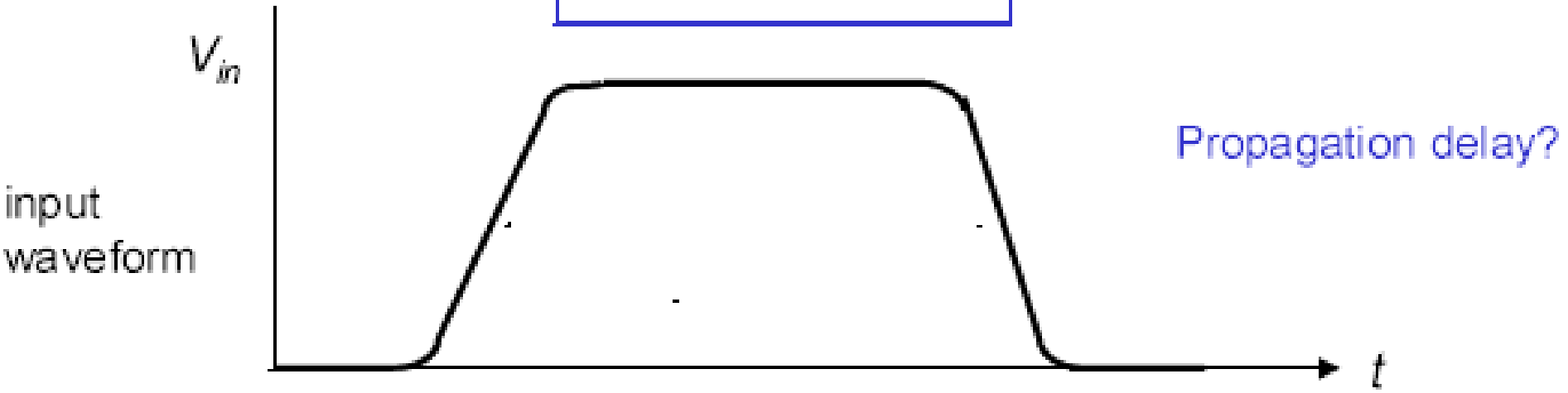
$R_o = 0$

Fanout = ∞

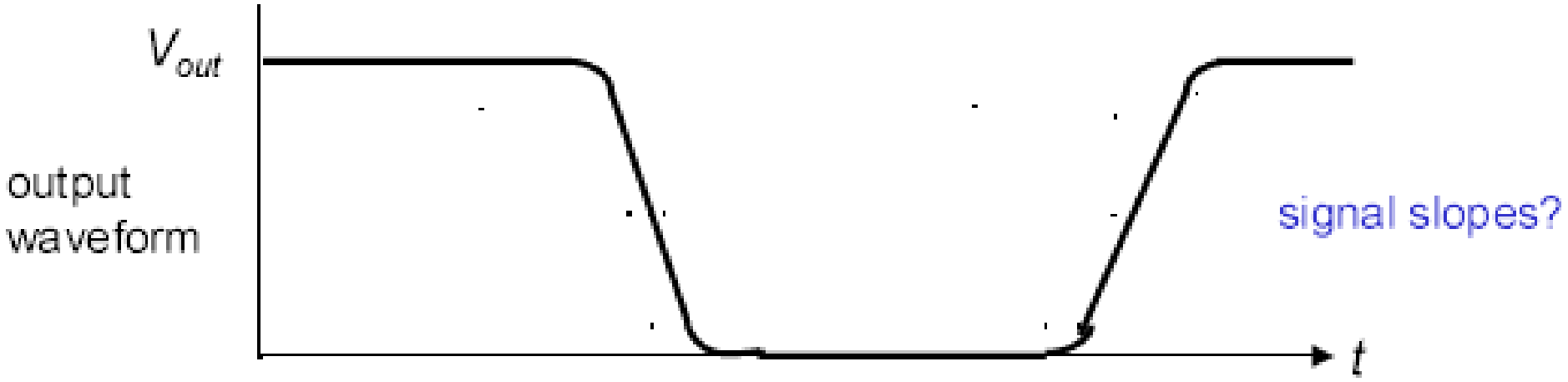
$NM_H = NM_L = VDD/2$

Delay Definitions



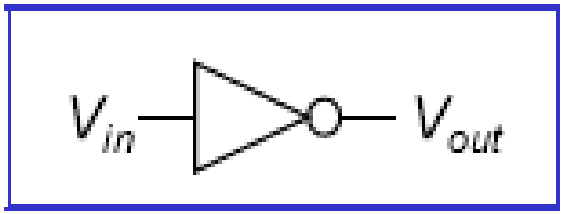


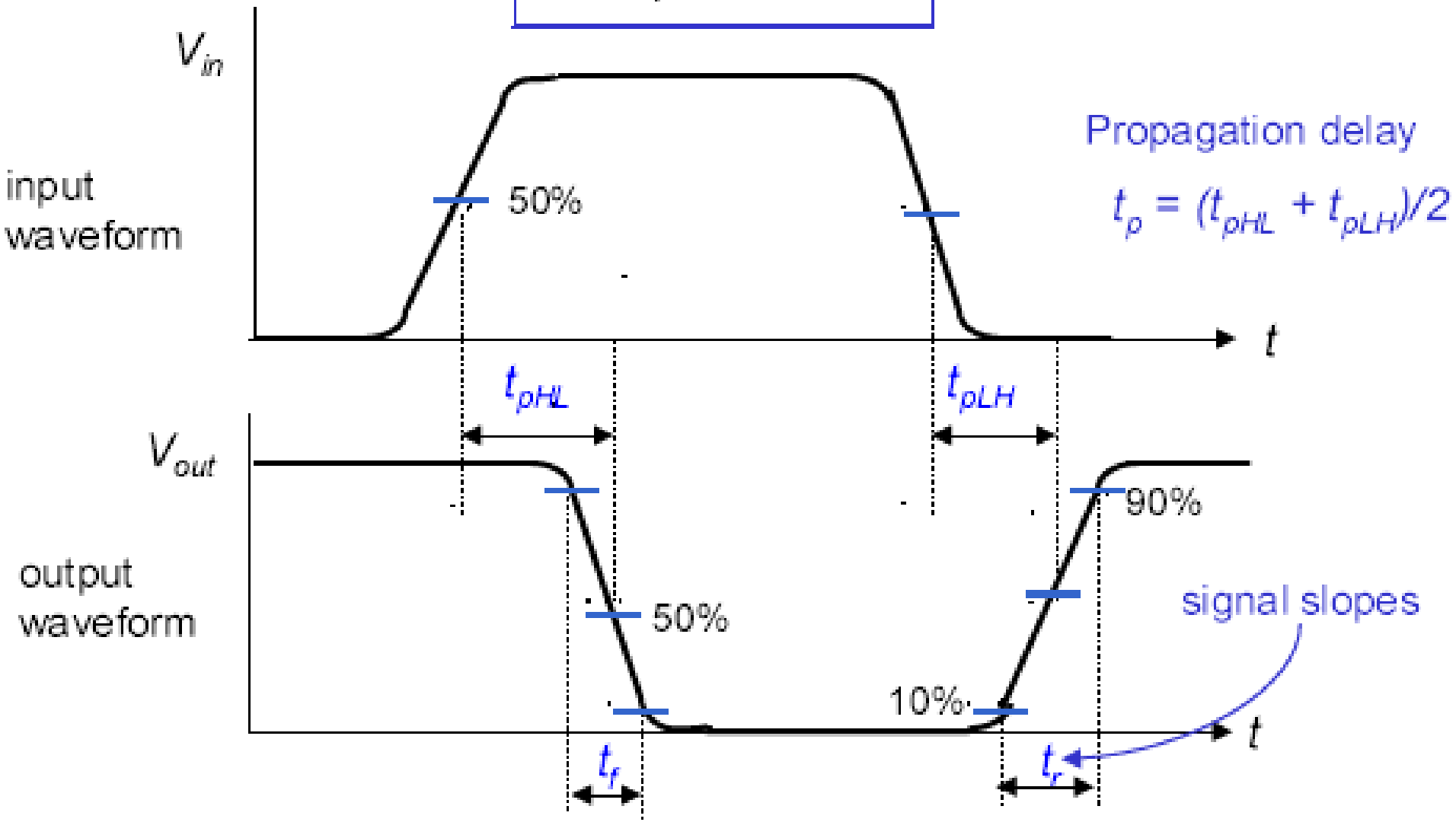
Propagation delay?



signal slopes?

Delay Definitions



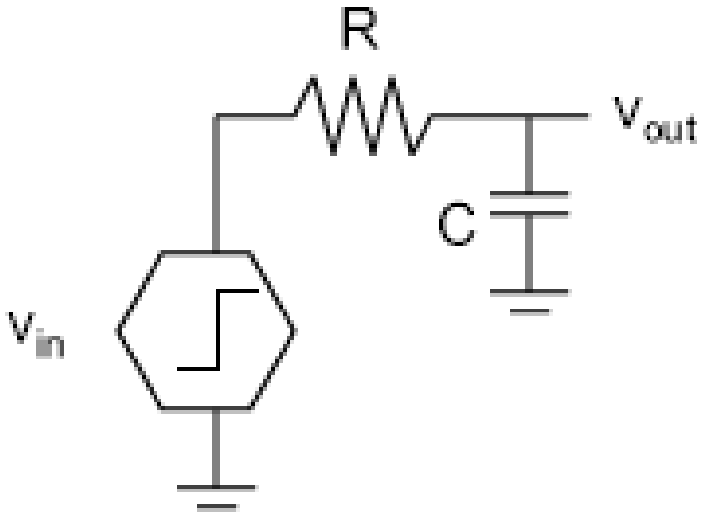


Propagation delay

$t_p = (t_{pHL} + t_{pLH})/2$

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
$t = \ln(10)\tau = 2.3\,\tau$

- 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)

$$p(t) = v(t)i(t) = V_{dd} i(t) \qquad P_{avg} = 1/T \int p(t)\,dt = V_{dd}/T \int i_{dd}(t)\,dt$$

  - packaging and cooling requirements
- Two important components: static and dynamic

$$E \text{ (joules)} = C_L V_{dd}^2 P_{0\to1} + t_{sc} V_{dd} I_{peak} P_{0\to1} + V_{dd} I_{leakage}$$

$$f_{0\to1} = P_{0\to1} * f_{clock}$$

$$P \text{ (watts)} = C_L V_{dd}^2 f_{0\to1} + t_{sc} V_{dd} I_{peak} f_{0\to1} + V_{dd} I_{leakage}$$

Power and Energy Dissipation (Cont.)

- 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}

Integrated Circuit Technologies

Why does CMOS dominate?

other technologies

- passive circuits
- III-V devices
- Silicon BJT

CMOS dominates because:

- **Silicon is cheaper** → preferred over other materials
- **physics** of CMOS is **easier** to understand
- CMOS is **easier to implement/fabricate**
- CMOS provides **lower power-delay product**
- CMOS is **lowest power**
- can get **more** CMOS transistors/functions **in same chip area**

- **BUT!** CMOS is not the fastest technology!
  - BJT and III-V devices are faster

CMOS Technology Trends

Variations over time

- # transistors / chip: increasing with time
- power / transistor: decreasing with time (constant power density)
- device channel length: decreasing with time
- power supply voltage: decreasing with time

low power/voltage is critical for future ICs

"Electronics" Building block(s)

- MOSFET Device-- 1950+ to 2020

- New elements in nano technologies are emerging include:
  - Memristor: memory resistor- see Dec IEEE Spectrum
  - Nano-tubes
  - Molecular devices
  - Quantum dots
  - Etc.

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