Microprocessor Technology & Case Studies

NBA 6120

January 29, 2018

Donald P. Greenberg

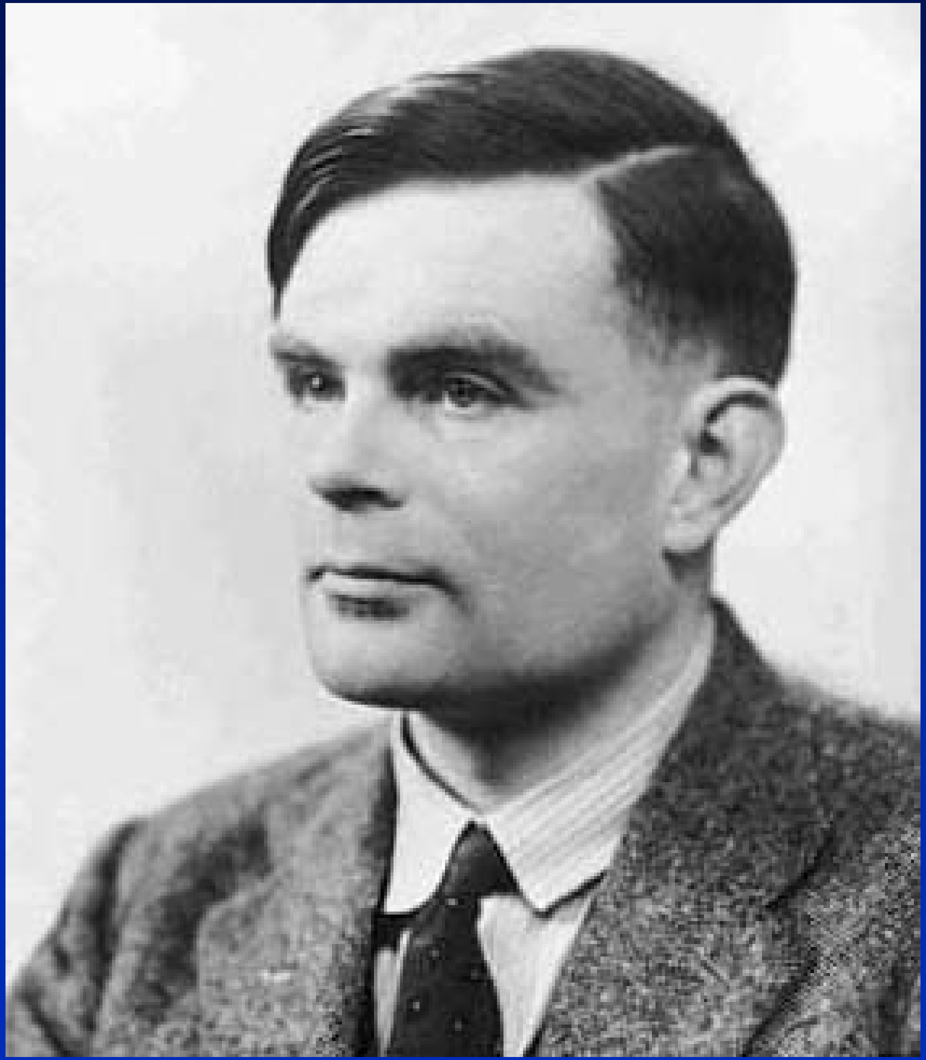
Lecture 2

- ***Required Reading:***
  - Craig R. Barrett. From Sand to Silicon: Manufacturing an Integrated Circuit, Scientific American, Special Issue, The Solid-State Century, January 1998, pp. 55-61. (Search: e-Journals/ Scientific American Archive Online/article (full text) http://www.library.cornell.edu/johnson/library/general/emba.html
  - Peter J. Denning and Ted G. Lewis. "Exponential Laws of Computing Growth." Communications of the ACM. January 2017. ACM.org.
- ***Optional Reading:***
  - Mack, Chris. "The Multiple Lives of Moore's Law." *IEEE Spectrum* Apr. 2015: 30-37. *Cornell University Library*. Web. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7065415

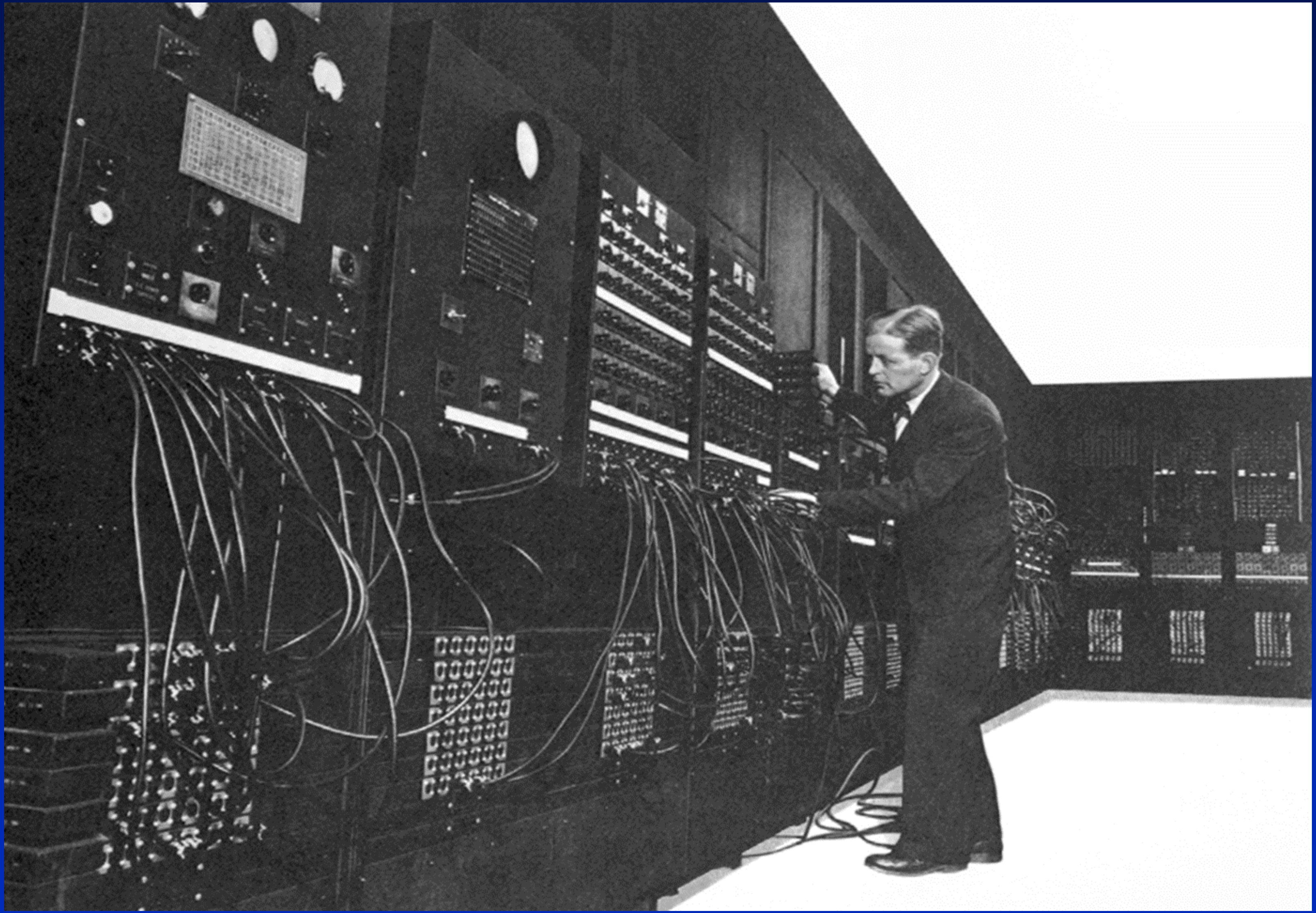
Bletchley Park

1940s

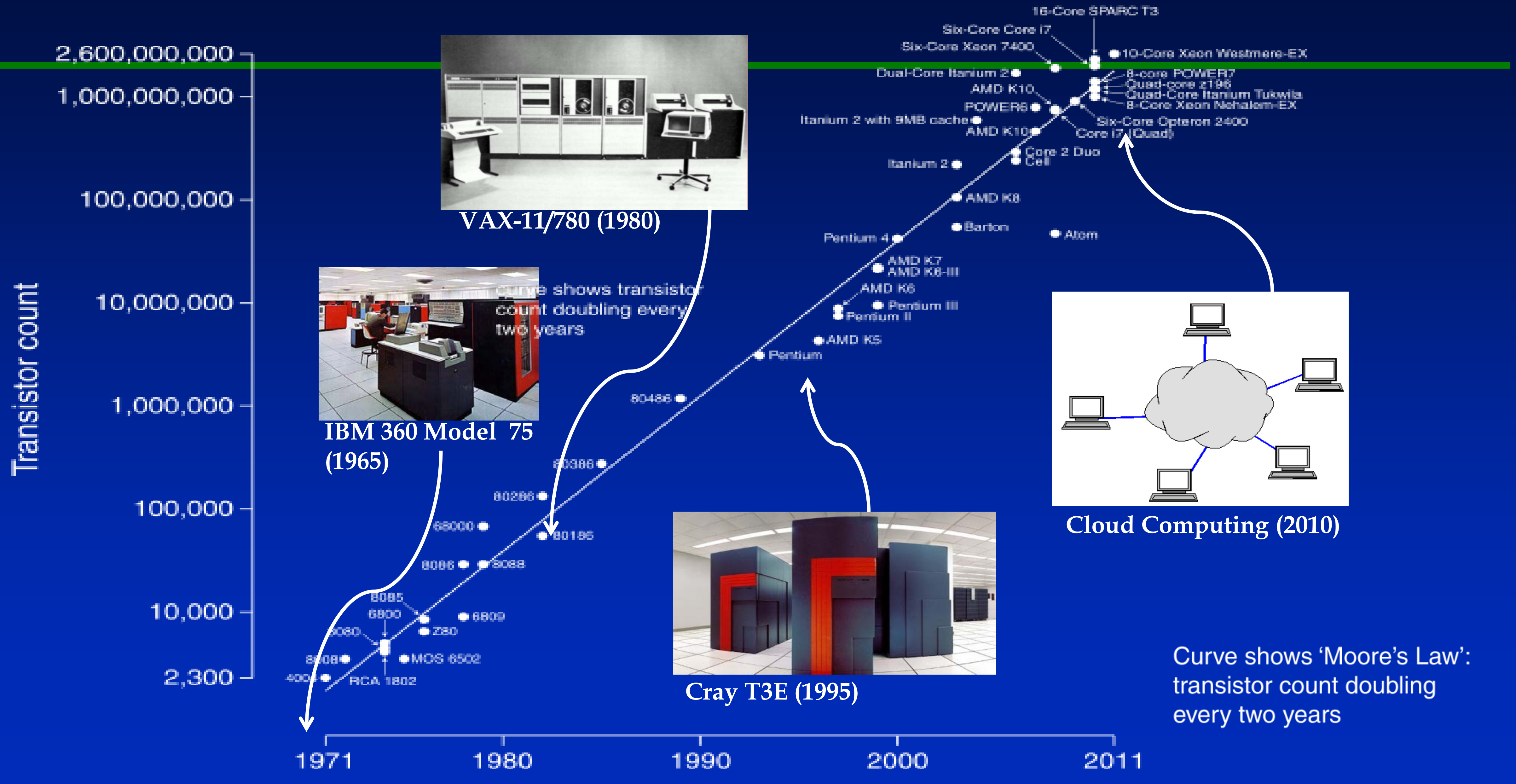
Alan Turing



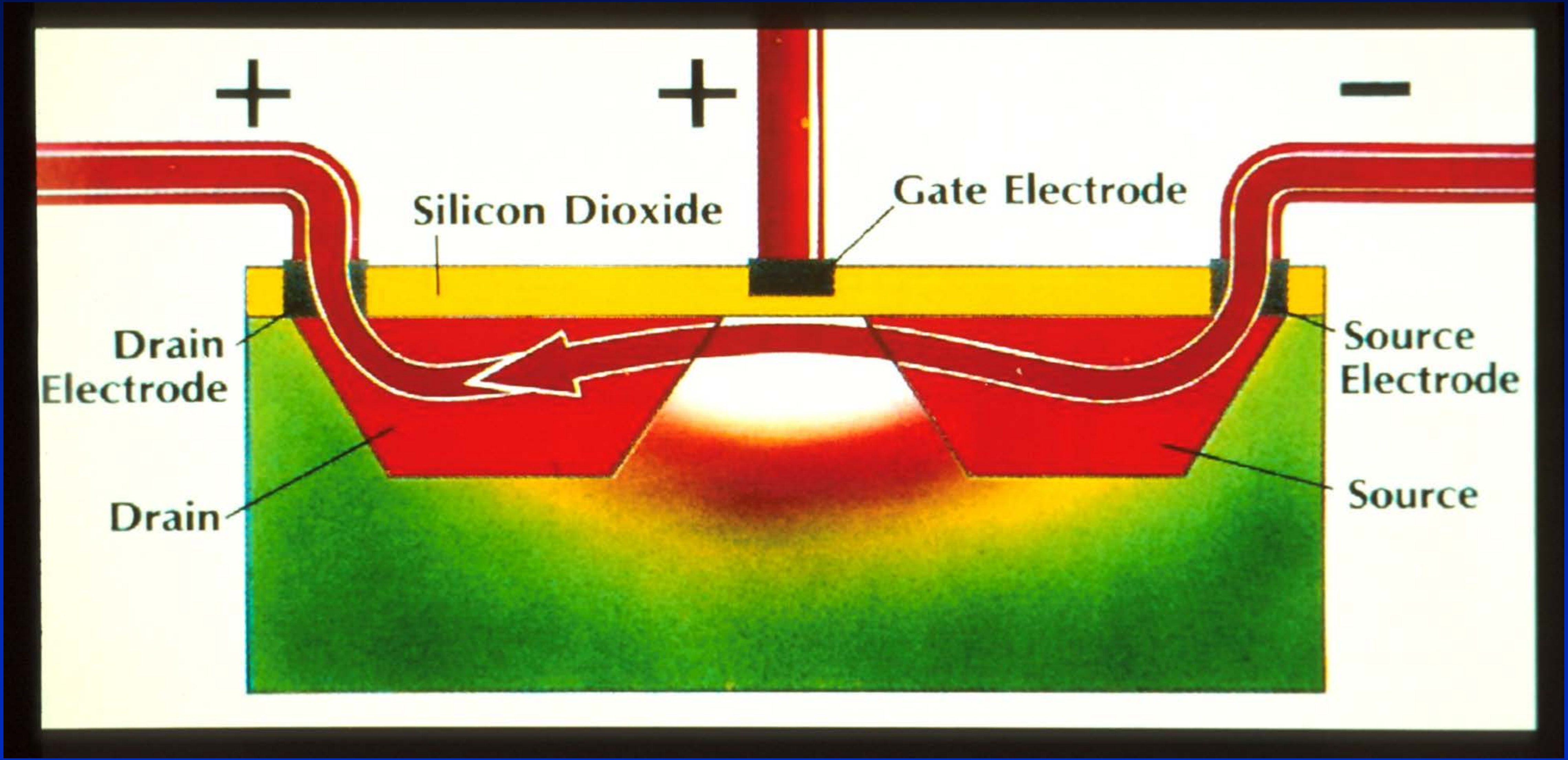
Eniac 1946



Microprocessor Transistor Counts 1971-2011 & Moore's Law

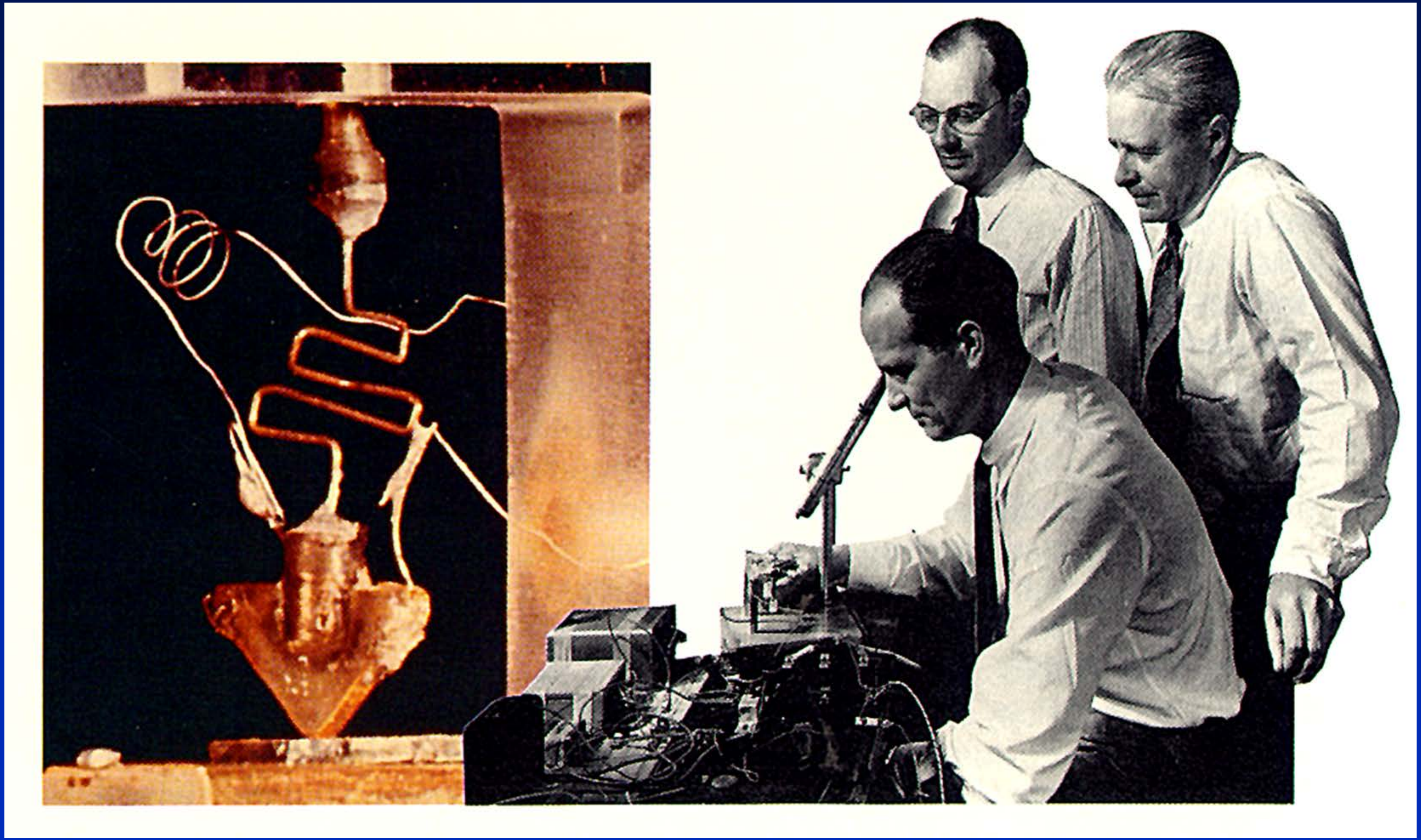


Transistor



+
+
−
Silicon Dioxide
Gate Electrode
Drain
Electrode
Source
Electrode
Drain
Source

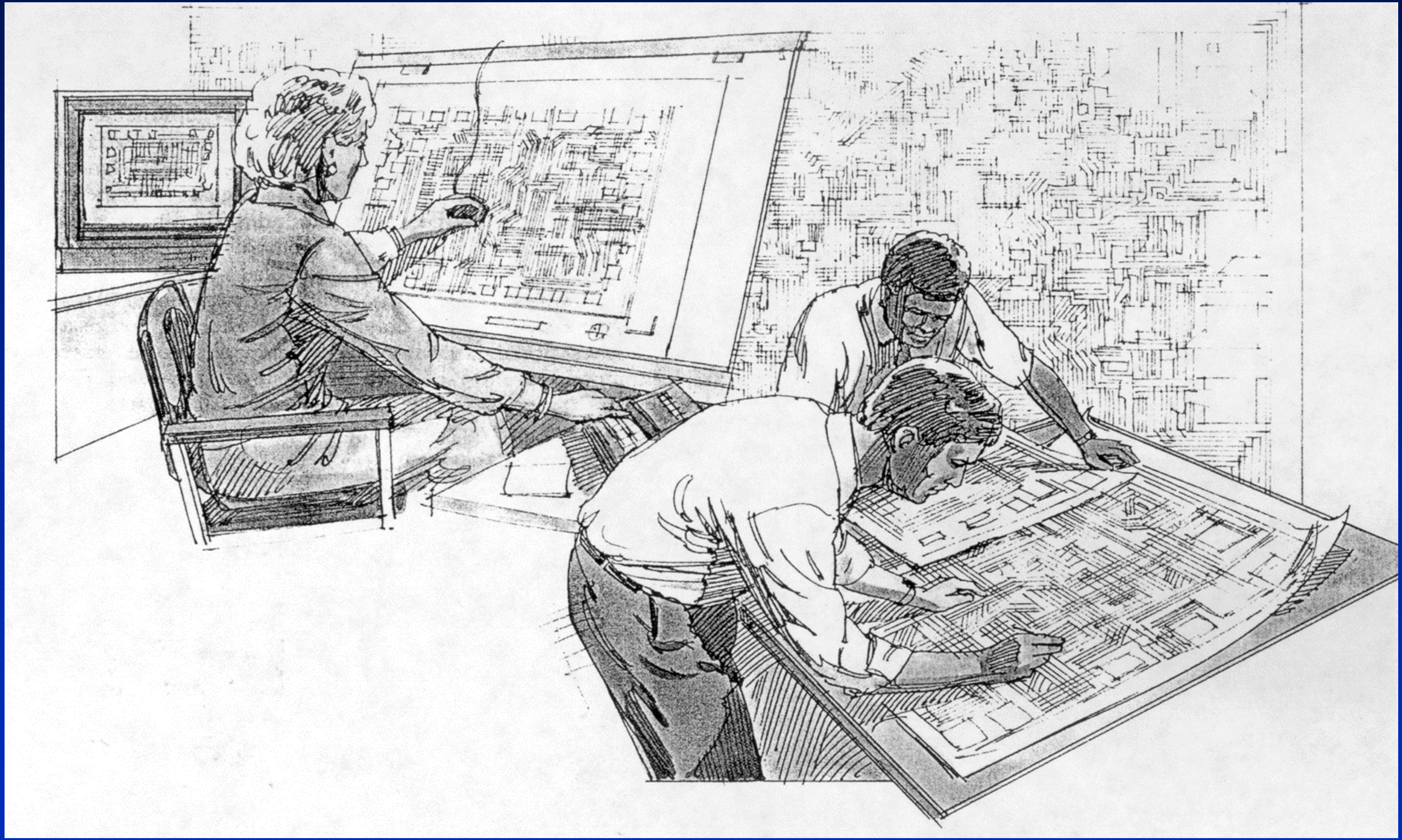
Shockley, Bardeen & Brattain 1947



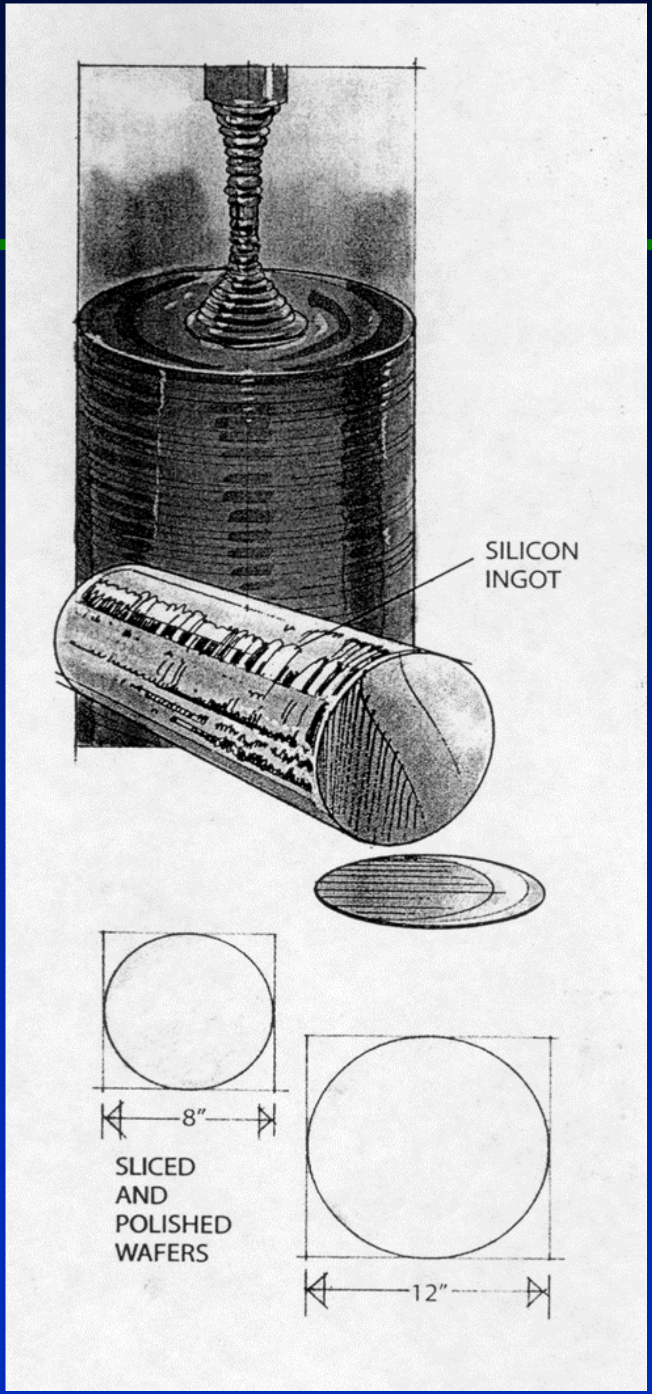
From Sand to Silicon – Manufacturing an Integrated Circuit

Scientific American: The Solid-State Century, Special Issue 1998

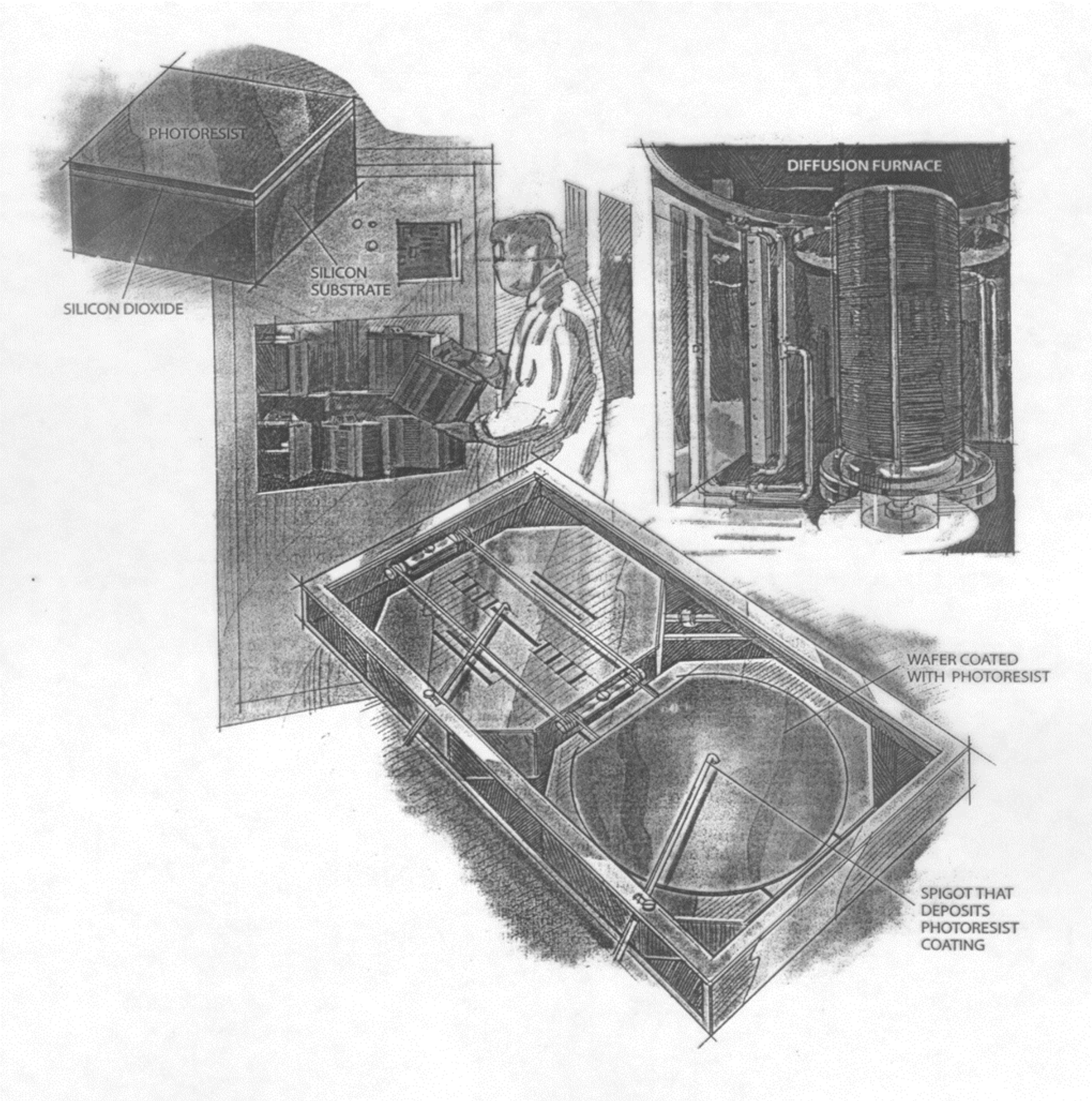
Chip Design



Silicon Crystal

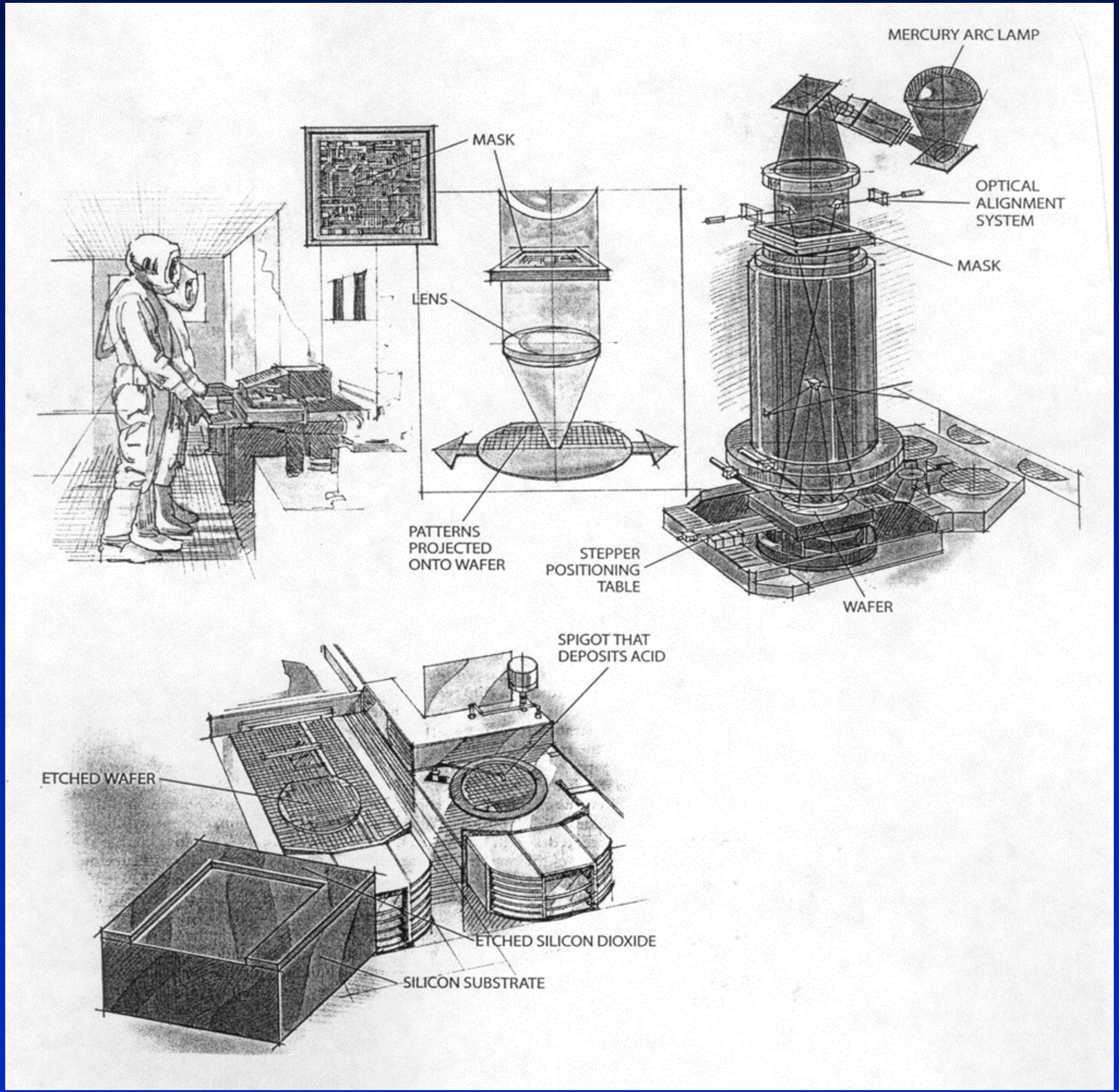


Layering

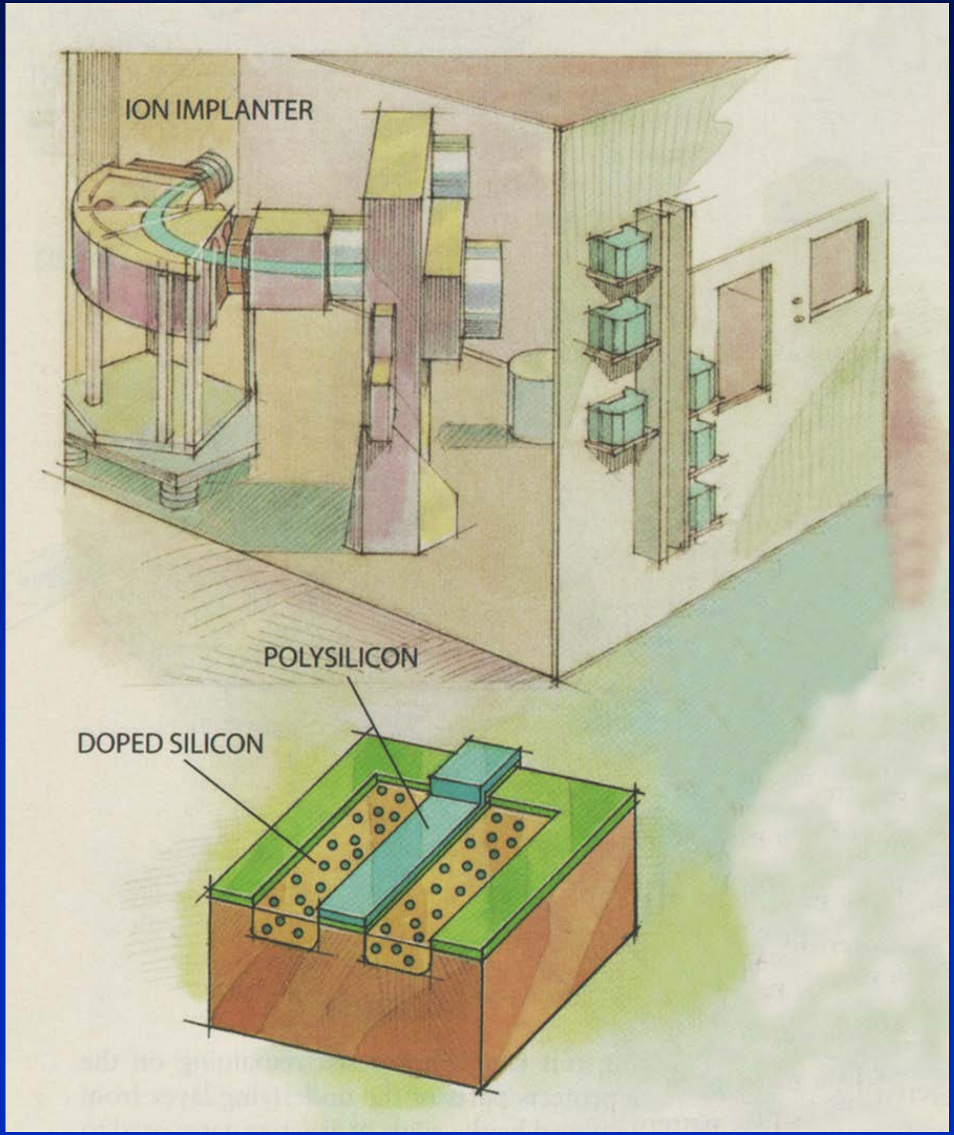


PHOTORESIST
SILICON SUBSTRATE
SILICON DIOXIDE
DIFFUSION FURNACE
WAFER COATED WITH PHOTORESIST
SPIGOT THAT DEPOSITS PHOTORESIST COATING

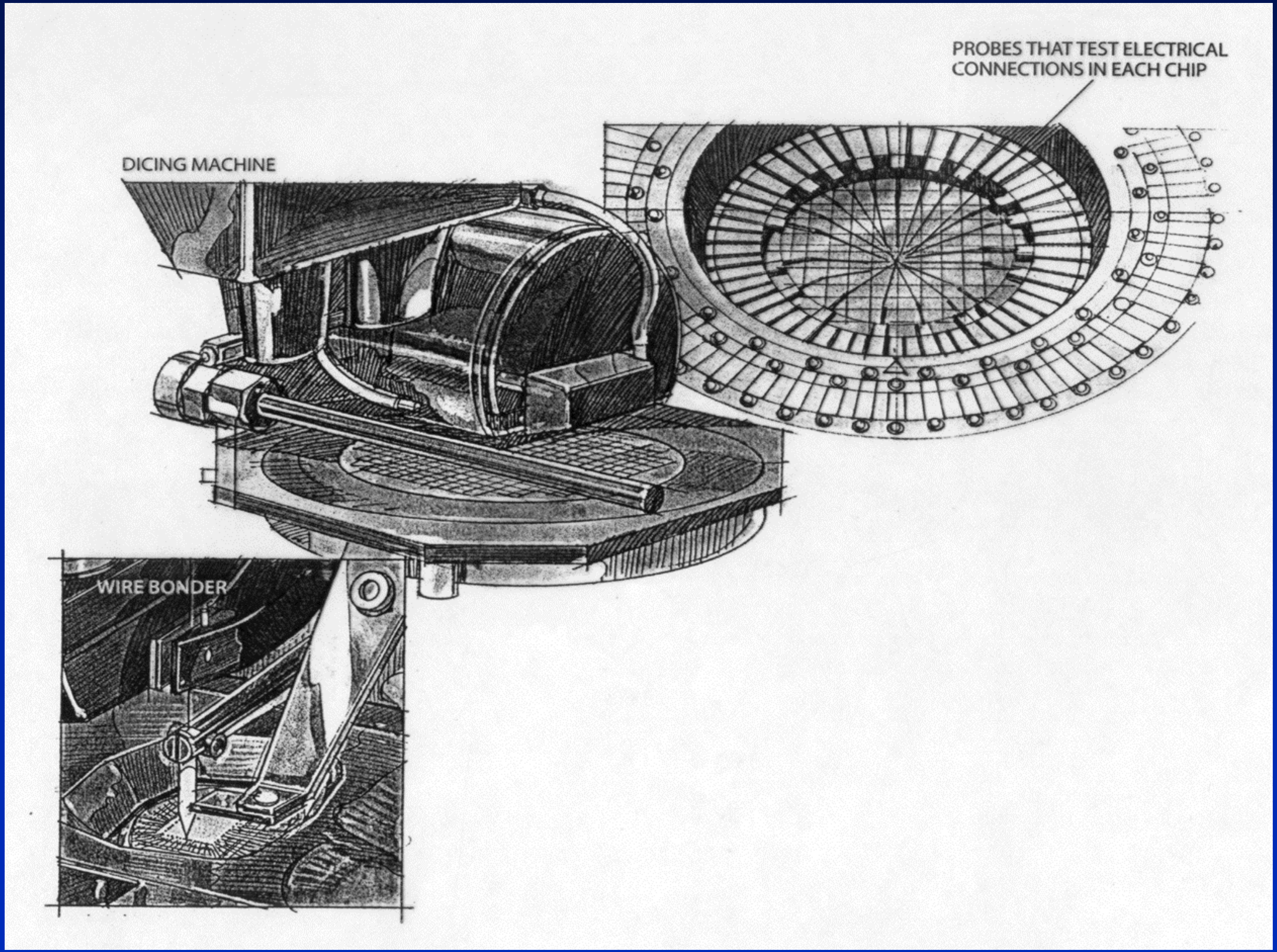
Masking & Etching



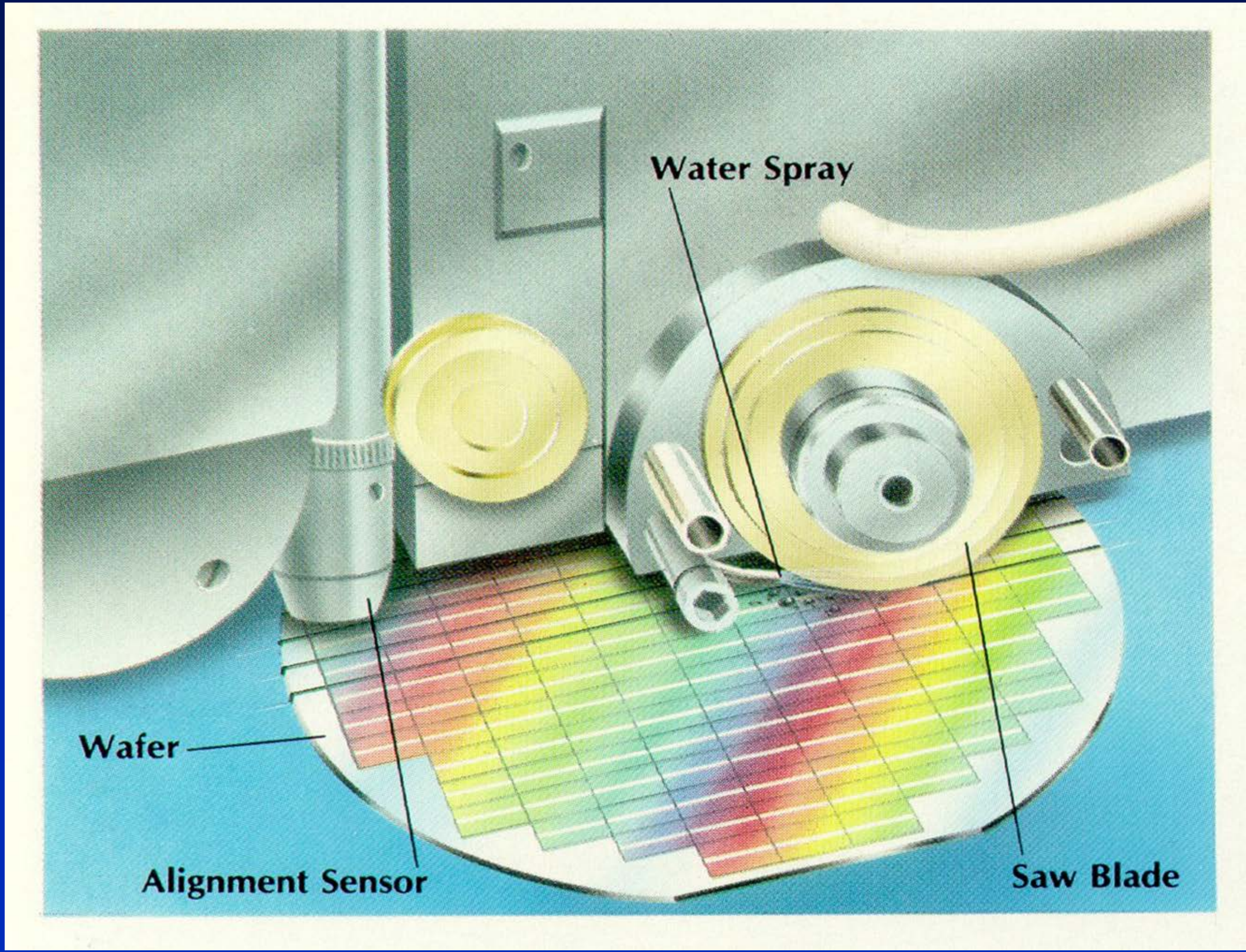
Doping



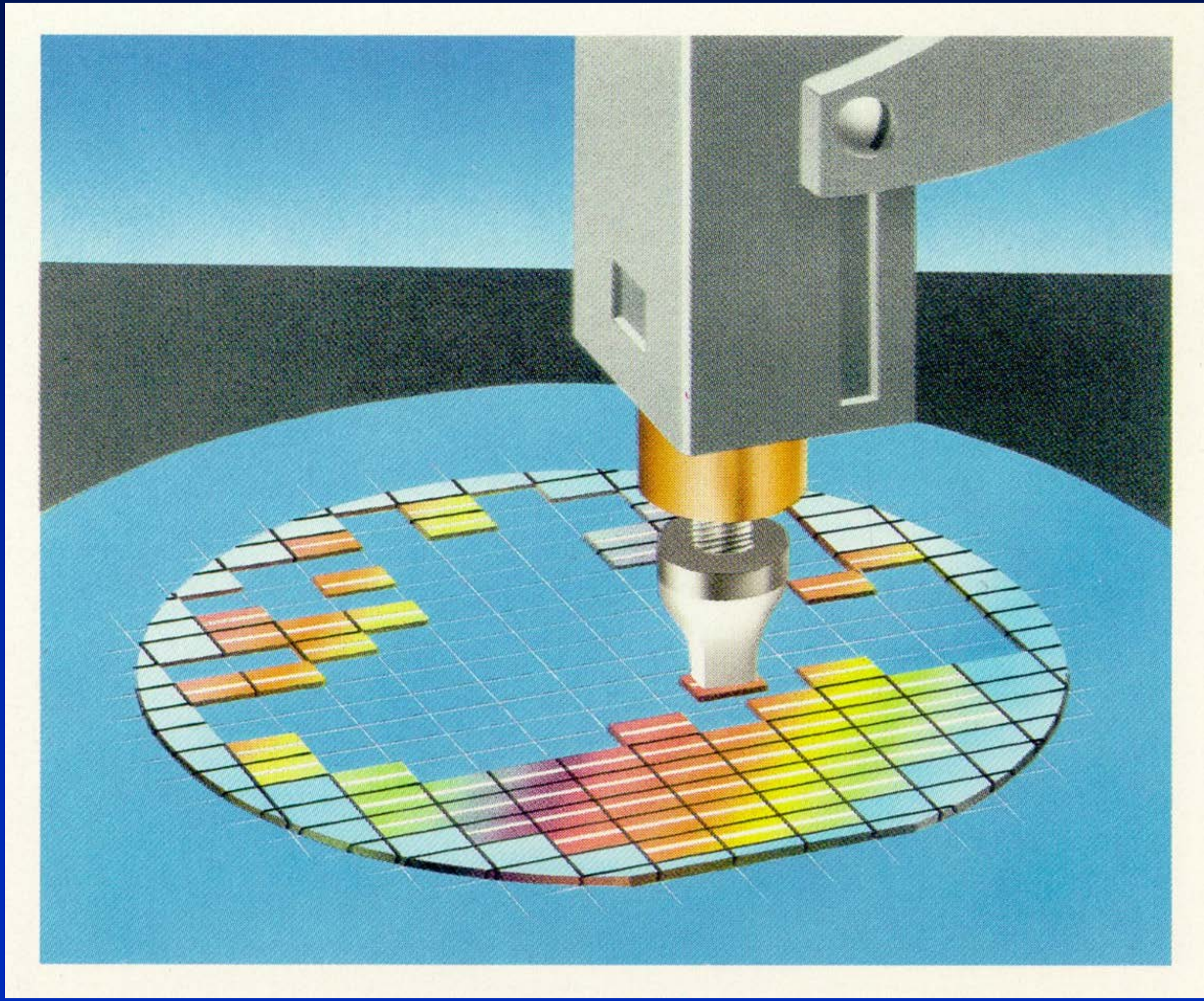
Interconnections & Dicing



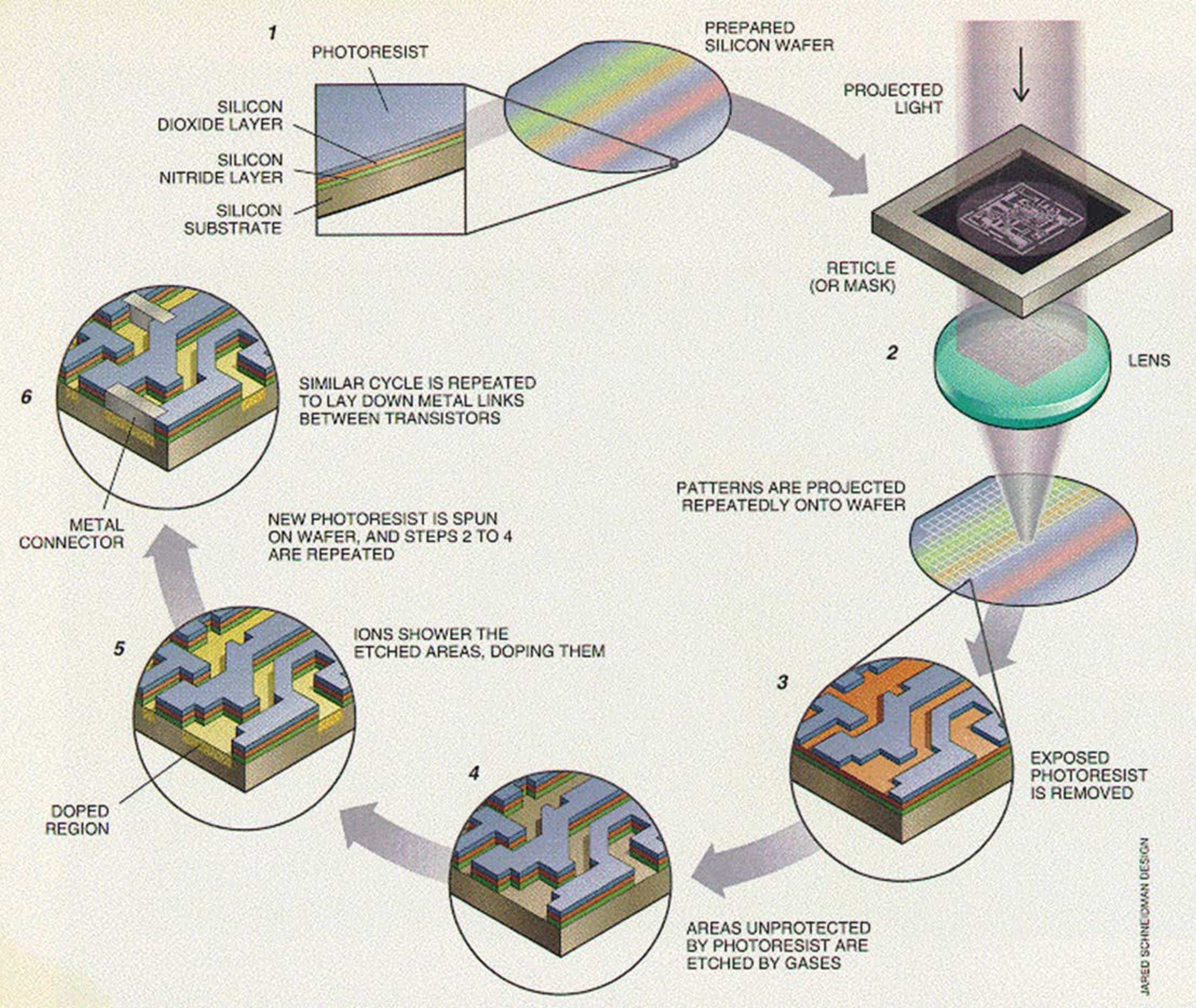
Dicing



Chip Selection



Chip Fabrication



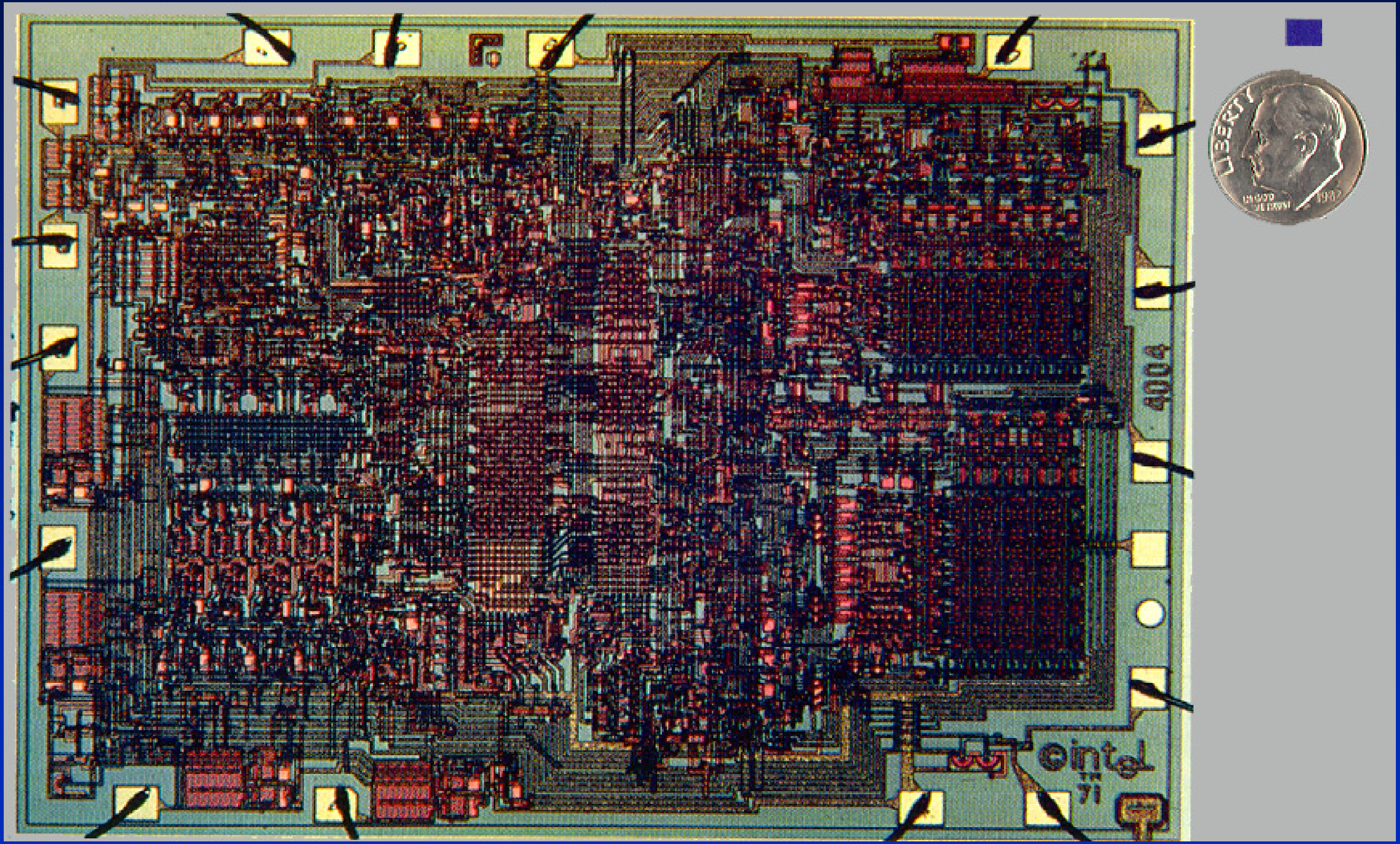
International Technology Roadmap for Semiconductors

	2001	2004	2007	2010	2013	2016
Technology (nanometers)	**130nm**	**90nm**	**65nm**	**45nm**	**32nm**	**22nm**
Functions per Chip (millions)	97	193	386	1546	3092	6184
Clock Speed (Ghz)	2.5Ghz	4.1Ghz	9.3Ghz	15Ghz	23Ghz	40Ghz
Wafer Size (millimeters)	200mm	300mm	300mm	300mm	450mm	450mm
Chip Size (mm^2)	140 mm^2	140 mm^2	140 mm^2	140 mm^2	140 mm^2	140 mm^2

Roughly 0.5 shrink every 3 years.

Intel released 22 nm chips in 2013

Intel 4004 November 1971



Moore's Law – CPU Transistor Counts

Processor	Transistor count	Date of introduction	Manufacturer	Process	Area
Core 2 Duo	291,000,000	2006	Intel	65 nm	
AMD K10	463,000,000	2007	AMD	65 nm	
AMD K10	758,000,000	2008	AMD	45 nm	
Itanium 2 with 9MB cache	592,000,000	2004	Intel	130 nm	
Core i7 (Quad)	731,000,000	2008	Intel	45 nm	263 mm²
POWER6	789,000,000	2007	IBM	65 nm	341 mm²
Six-Core Opteron 2400	904,000,000	2009	AMD	45 nm	
Six-Core Core i7	1,170,000,000	2010	Intel	32 nm	
Dual-Core Itanium 2	1,700,000,000	2006	Intel	90 nm	596 mm²
Six-Core Xeon 7400	1,900,000,000	2008	Intel	45 nm	
Quad-Core Itanium Tukwila	2,000,000,000	2010	Intel	65 nm	
Six-Core Core i7 (Sandy Bridge-E)	2,270,000,000	2011	Intel	32 nm	434 mm²
8-Core Xeon Nehalem-EX	2,300,000,000	2010	Intel	45 nm	684 mm²
10-Core Xeon Westmere-EX	2,600,000,000	2011	Intel	32 nm	512 mm²
Six-core zEC12	2,750,000,000	2012	IBM	32 nm	597 mm²
8-Core Itanium Poulson	3,100,000,000	2012	Intel	32 nm	544 mm²
15-Core Xeon Ivy Bridge-EX	4,310,000,000	2014	Intel	22nm	541 mm²
62-Core Xeon Phi	5,000,000,000	2012	Intel	22 nm	350 mm²
Xbox One Main SoC	5,000,000,000	2013	Microsoft	28 nm	363 mm²
18-core Xeon Haswell-E5	5,560,000,000	2014	Intel	22 nm	661mm²
IBM z14 Storage Controller	9,700,000,000	2017	IBM	14 nm	696mm²
32-core SPARC M7	10,000,000,000	2015	Oracle	20 nm	
Centriq 2400	18,000,000,000	2017	Qualcomm	10 nm	398 mm²
32-core AMD Epyc	19,200,000,000	2017	AMD	14 nm	4× 192 mm2

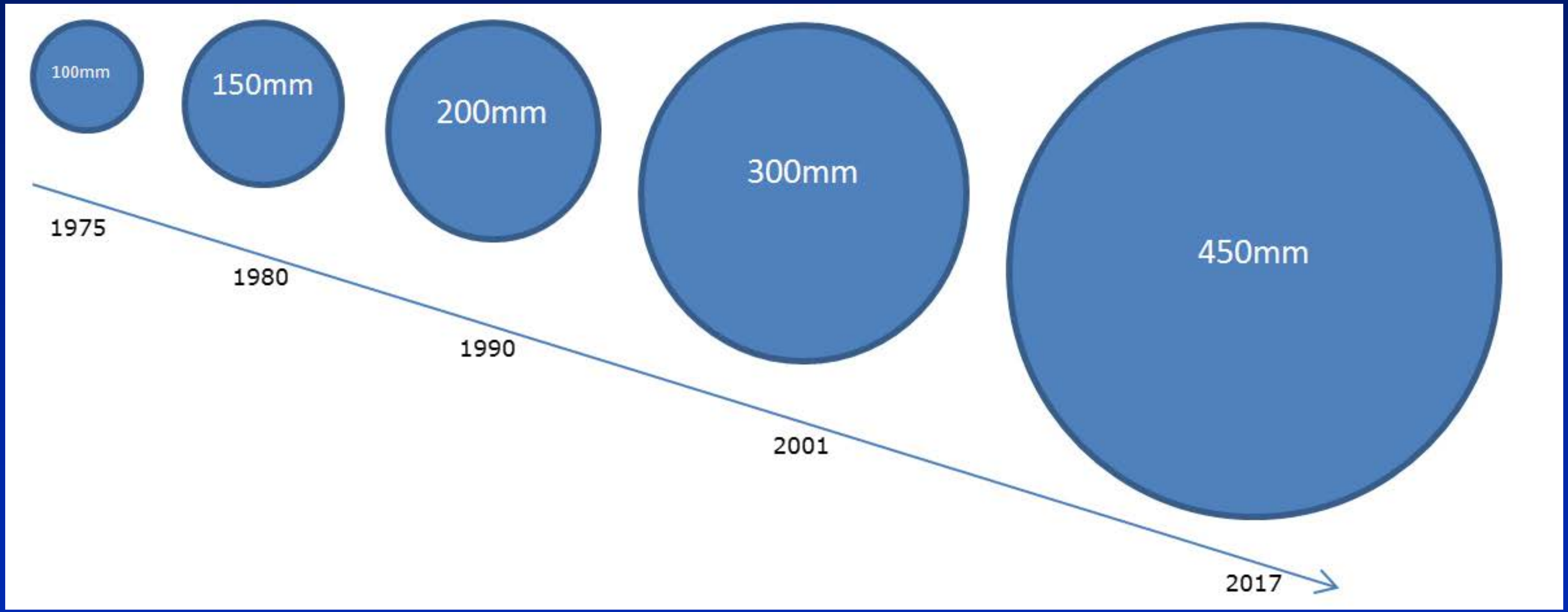
http://en.wikipedia.org/wiki/Transistor_count

Moore’s Law – GPU Transistor Counts

Processor	Transistor count	Date of introduction	Manufacturer	Process	Area
R520	321,000,000	2005	AMD	90 nm	288 mm^2
R580	384,000,000	2006	AMD	90 nm	352 mm^2
G80	681,000,000	2006	NVIDIA	90 nm	480 mm^2
R600 Pele	700,000,000	2007	AMD	80 nm	420 mm^2
G92	754,000,000	2007	NVIDIA	65 nm	324 mm^2
RV790XT Spartan	959,000,000	2008	AMD	55 nm	282 mm^2
GT200 Tesla	1,400,000,000	2008	NVIDIA	65 nm	576 mm^2
Cypress RV870	2,154,000,000	2009	AMD	40 nm	334 mm^2
Cayman RV970	2,640,000,000	2010	AMD	40 nm	389 mm^2
GF100 Fermi	3,200,000,000	Mar 2010	NVIDIA	40 nm	526 mm^2
GF110 Fermi	3,000,000,000	Nov 2010	NVIDIA	40 nm	520 mm^2
GK104 Kepler	3,540,000,000	2012	NVIDIA	28 nm	294 mm^2
Tahiti RV1070	4,312,711,873	2011	AMD	28 nm	365 mm^2
GK110 Kepler	7,080,000,000	2012	NVIDIA	28 nm	561 mm^2
RV1090 Hawaii	6,300,000,000	2013	AMD	28 nm	438 mm^2
GM204 Maxwell	5,200,000,000	2014	NVIDIA	28 nm	398 mm^2
GM200 Maxwell	8,100,000,000	2015	NVIDIA	28 nm	601 mm^2
Fiji	8,900,000,000	2015	AMD	28 nm	596 mm^2
GP102 Pascal	12,000,000,000	2016	Nvidia	12 nm	471 mm^2
Vega 10	12,500,000,000	2017	AMD	14 nm	484 mm^2
GP100 Pascal	15,300,000,000	2016	Nvidia	16 nm	610 mm^2
GV100 Volta	21,100,000,000	2017	Nvidia	12 nm	815 mm^2

http://en.wikipedia.org/wiki/Transistor_count

Paul S. Otellini – Intel Corporation CEO 2007



Why are we continuing to strive for smaller and smaller technology?

- More transistors/chip → increased functionality and performance
- Higher speeds → partially depends on how close together the components are placed
- Cheaper – more chips/wafer, greater yields

Yield Ratio

$$yield = \frac{n_w}{n_t}$$

$$n_w = yield \bullet n_t$$

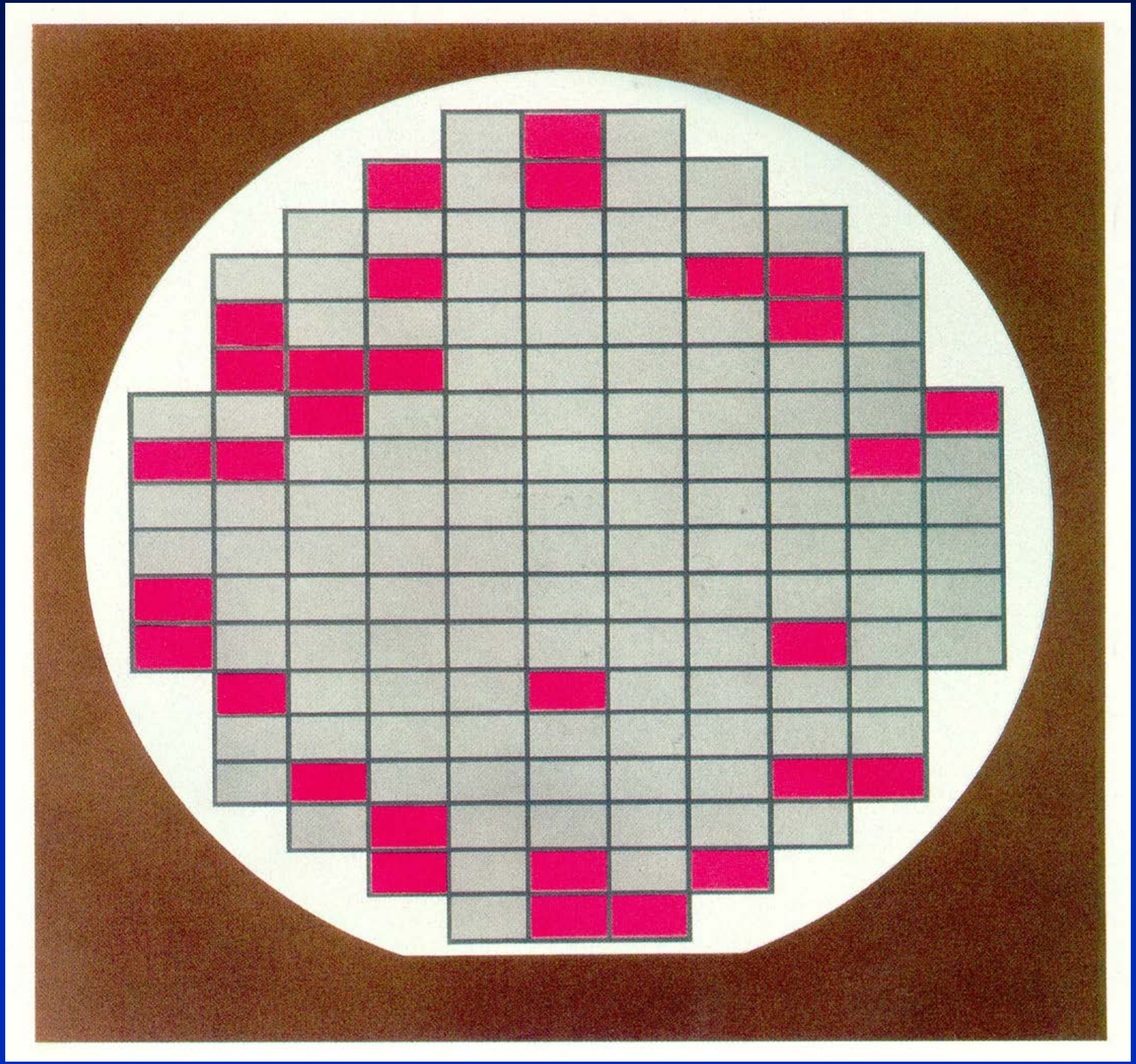
$$n_w = number\ of\ working\ chips/wafer$$

$$n_t = total\ number\ of\ chips/wafer$$

$$Old\ fab\ lines, yield \rightarrow > 90\%$$

$$New\ fab\ lines, yield \rightarrow < 40\%$$

Yield per Wafer



Yield Ratio

Number of defects/unit area depends on the process

$$\therefore Yield \approx \frac{1}{Chip\ Area}$$

Total chips (n_t) for a given wafer size is also inversely proportional to the chip area

Why does the shrinking technology make the cost of manufacturing cheaper per component?

Example:

For a 10% shrink in feature size :

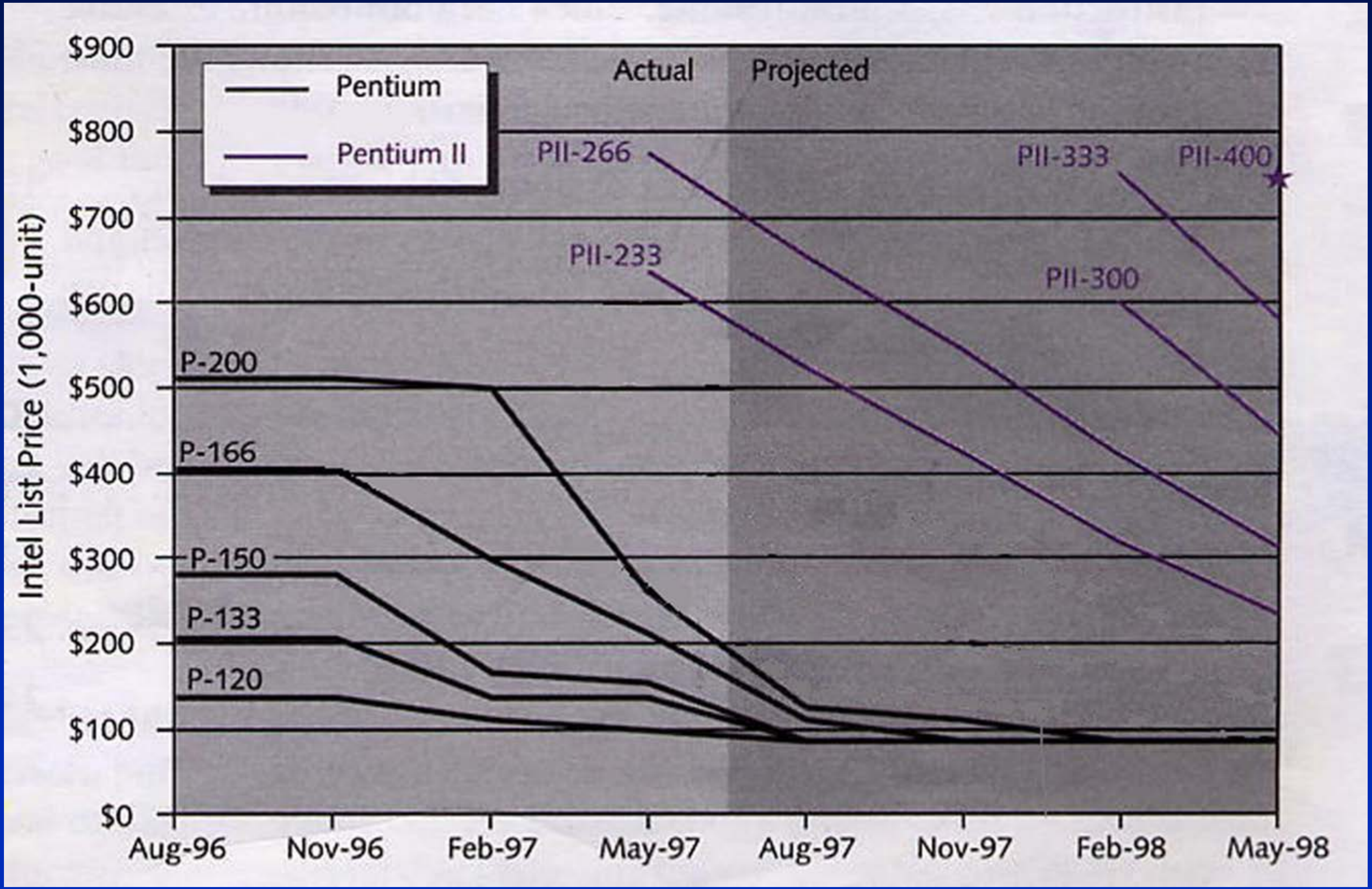
$$n_{w_{new}} = n_{w_{old}} \left(\frac{1}{.9}\right)^2 \left(\frac{1}{.9}\right)^2$$

$\uparrow$ New yield

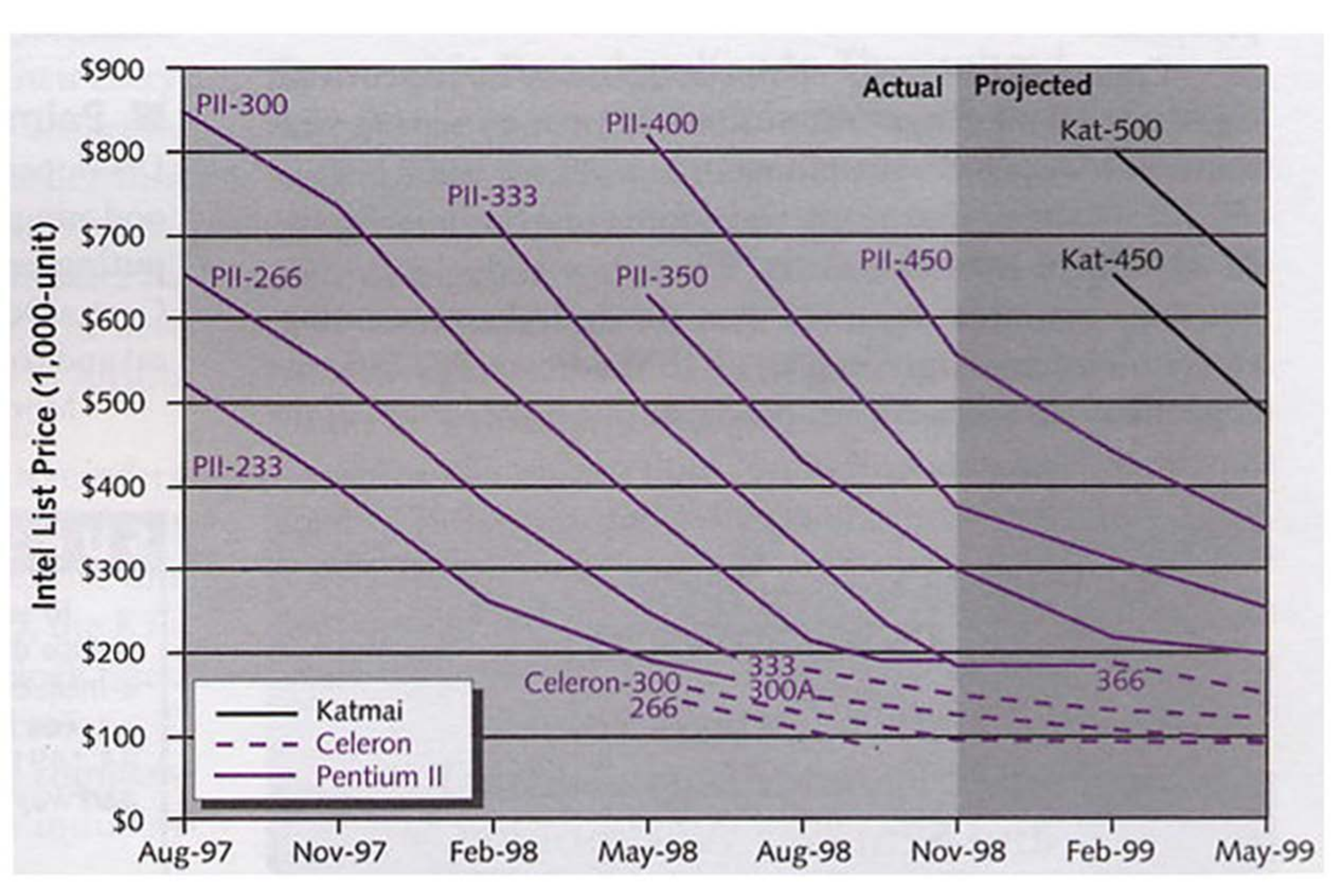
$\uparrow$ New n_t

$$n_{w_{new}} = 1.52 n_{w_{old}}$$

Projected Chip Pricing



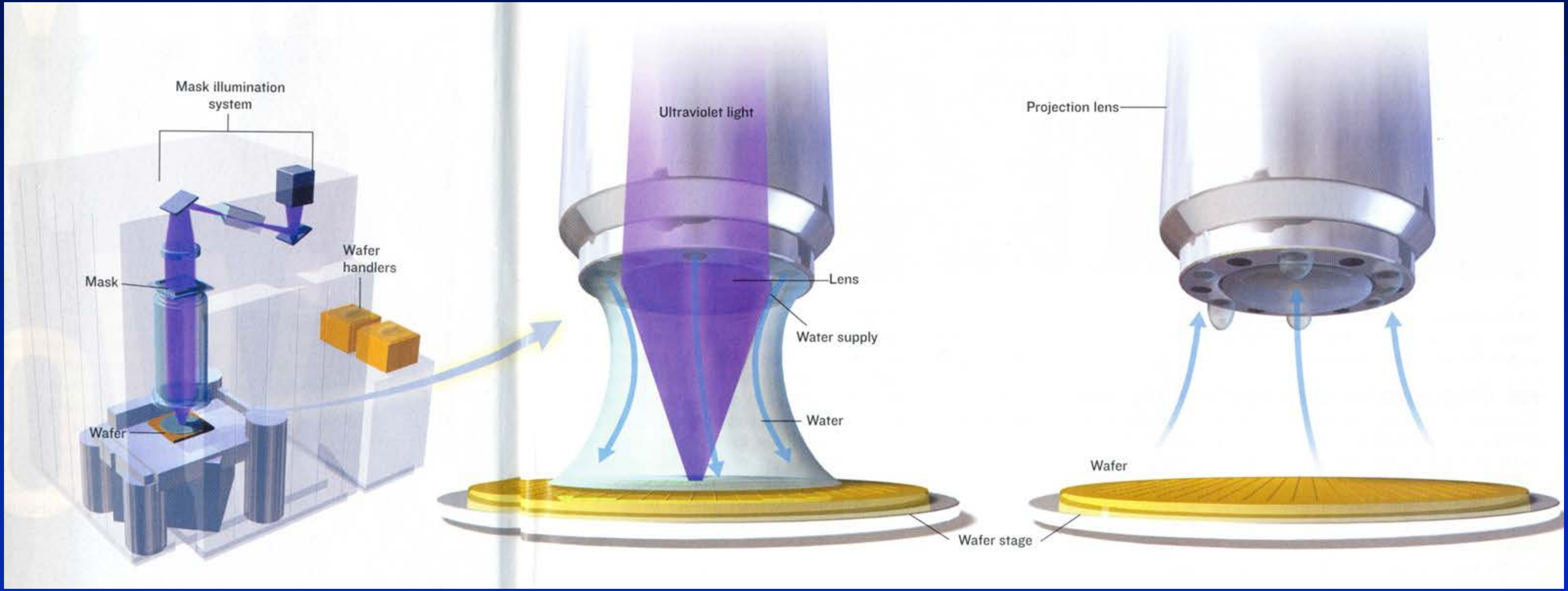
Projected Chip Pricing



Keeping Up with Moore’s Law

Remarkably, Moore's Law-the number of transistors that can fit on a microchip will double every 18-24 months-has held true for many years. Keeping up with this famous prediction by Intel founder Gordon Moore is getting harder.

Getting Wafers Wet



By adding a thin layer of water between the projection lens and the wafer, the immersion system can create features 30 percent smaller.

Photolithography

- Defining the smallest components requires short wavelengths of light.
- Currently, most fabrication processors use extreme ultra-violet light at 193nm.
- Can pass the light through water. The water slows the light (less velocity) shrinking its wavelength. It is estimated that this technique will meet demands for 7 more years.
- On February 20, 2006 IBM Almaden & JSR Micro demonstrated a system using an “unidentified” light slowing liquid yielding patterns 29.9nm wide.

Science News, March 2, 2006

10 Nanometer Technology

- Nov. 15, 2012, Samsung unveiled a 64 gigabyte (GB) multimedia card (eMMC) based on 10 nm technology.
- April 11, 2013, Samsung announced it was mass-producing High-Performance 128-gigabit NAND Flash Memory with 10 nm and 20 nm technology.
- April 2015, TSMC announced that 10 nm production would begin at the end of 2016.
- May 23rd 2015, Samsung Electronics showed off a wafer of 10nm FinFET chips.

Factors Contributing to Advancing Microprocessor Performance

- Shrinking Component Size
- Increasing Speed
- Reducing Circuit Resistance
- New Materials

Factors Contributing to Advancing Microprocessor Performance

- RISC vs. CISC
- VLIW
- Multi-level Cache
- Parallelism & Pipelining

Factors Contributing to Advancing Microprocessor Performance

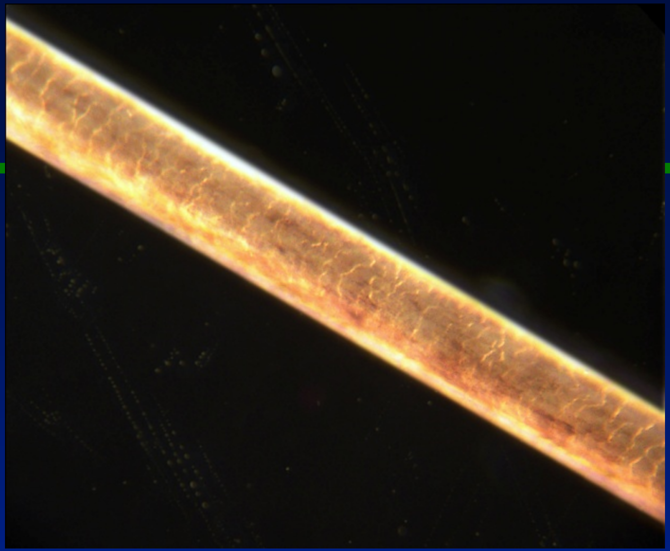
- RISC vs. CISC
- VLIW
- Multi-level Cache
- Parallelism & Pipelining
- Multi-core Technology

Multicore Craze

- For years, the trend was to make chips faster

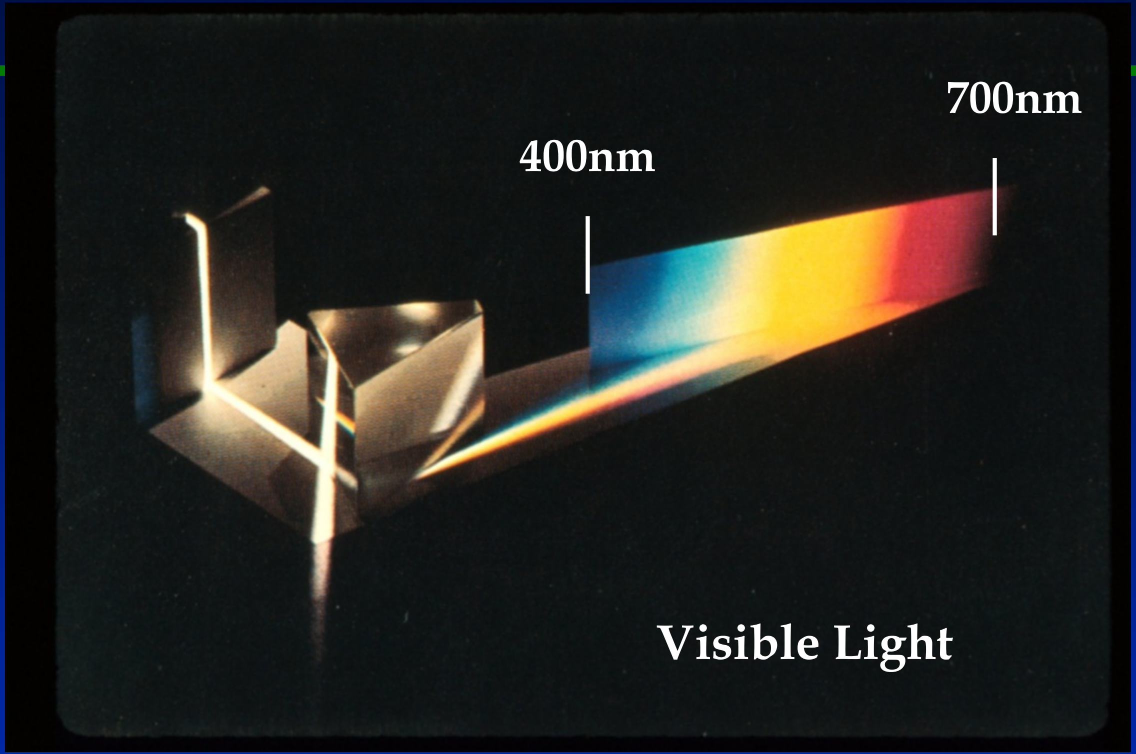
  Today → 3 Ghz

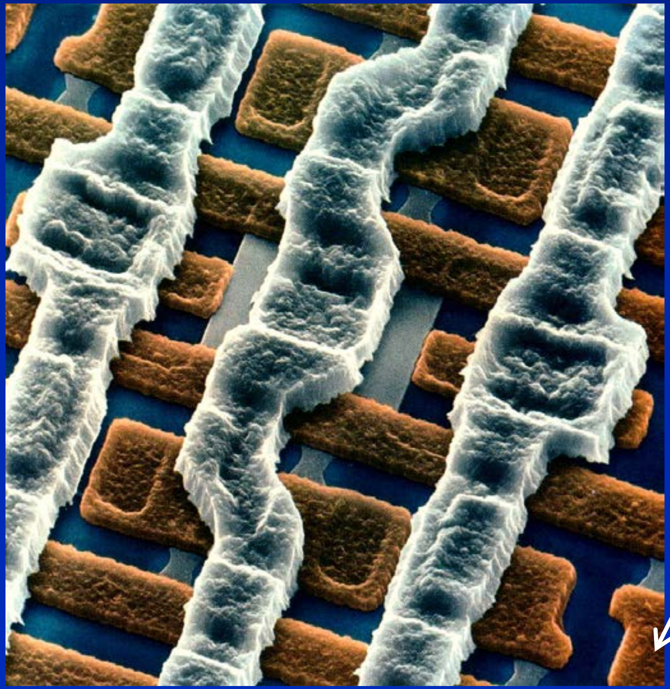
- But the power required (Watts) and the heat generated is proportional to the frequency squared.
- Therefore, put more computers on the chip but run at slower speeds.



Human Hair

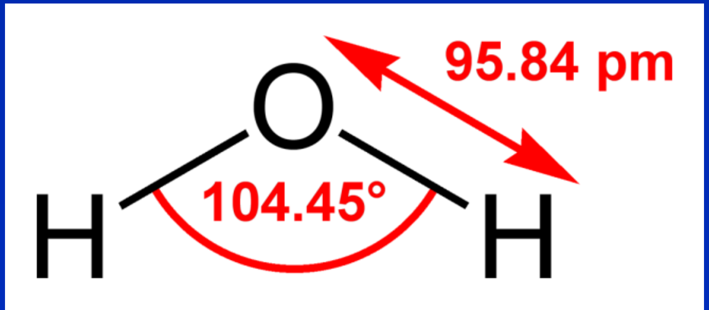
50μm





Integrated Chip

32nm



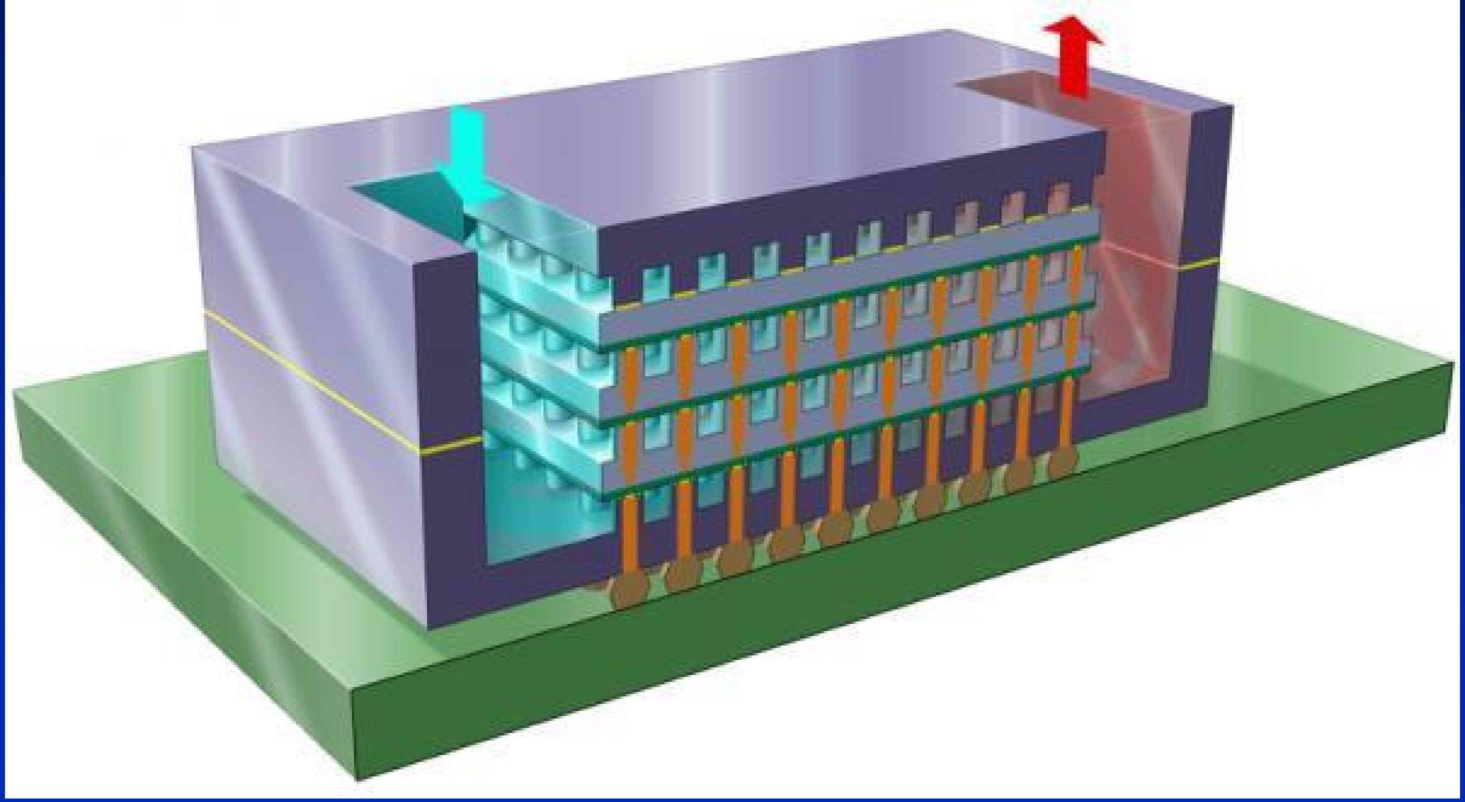
- How long can Moore’s Law continue?
- What are the limits to this integrated circuit technology?

“There are two constraints:

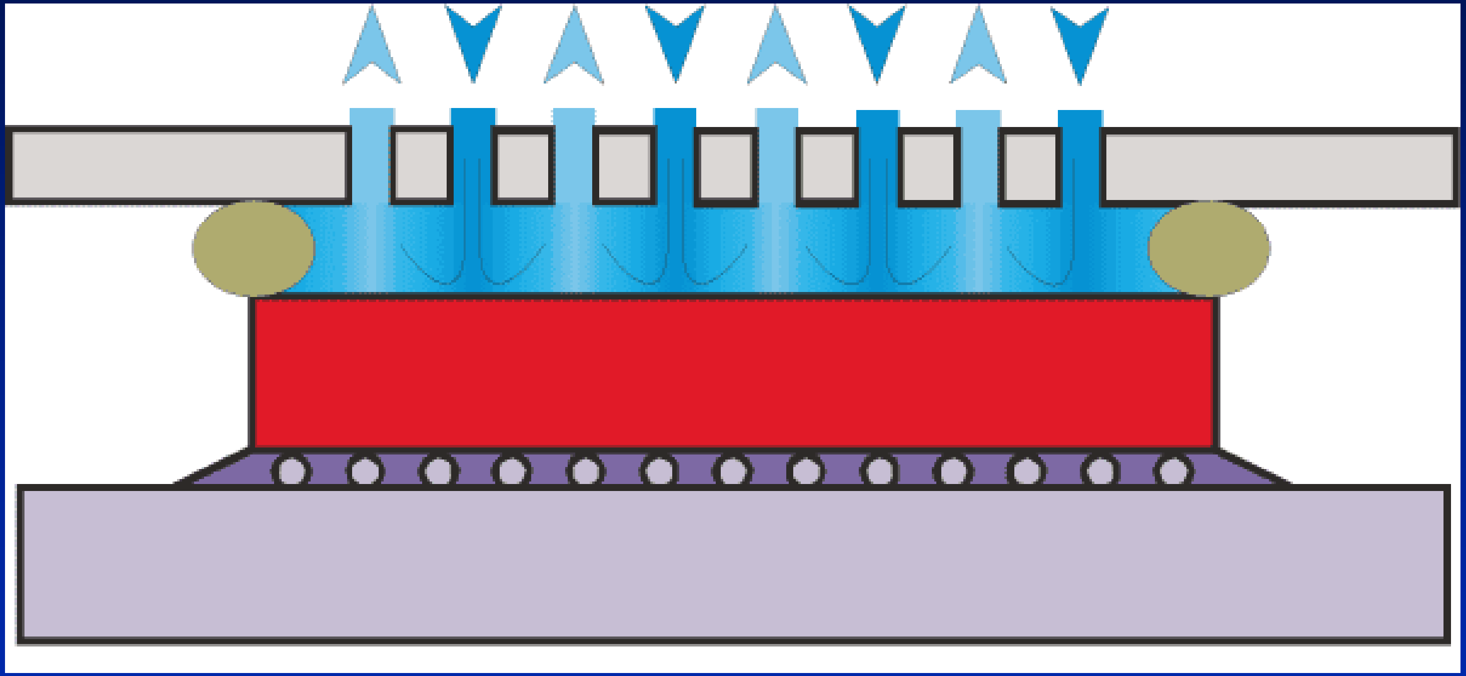
- *The finite velocity of light*
- *The atomic nature of materials”*

- Stephen Hawking

IBM’s Chip Stacking Technology



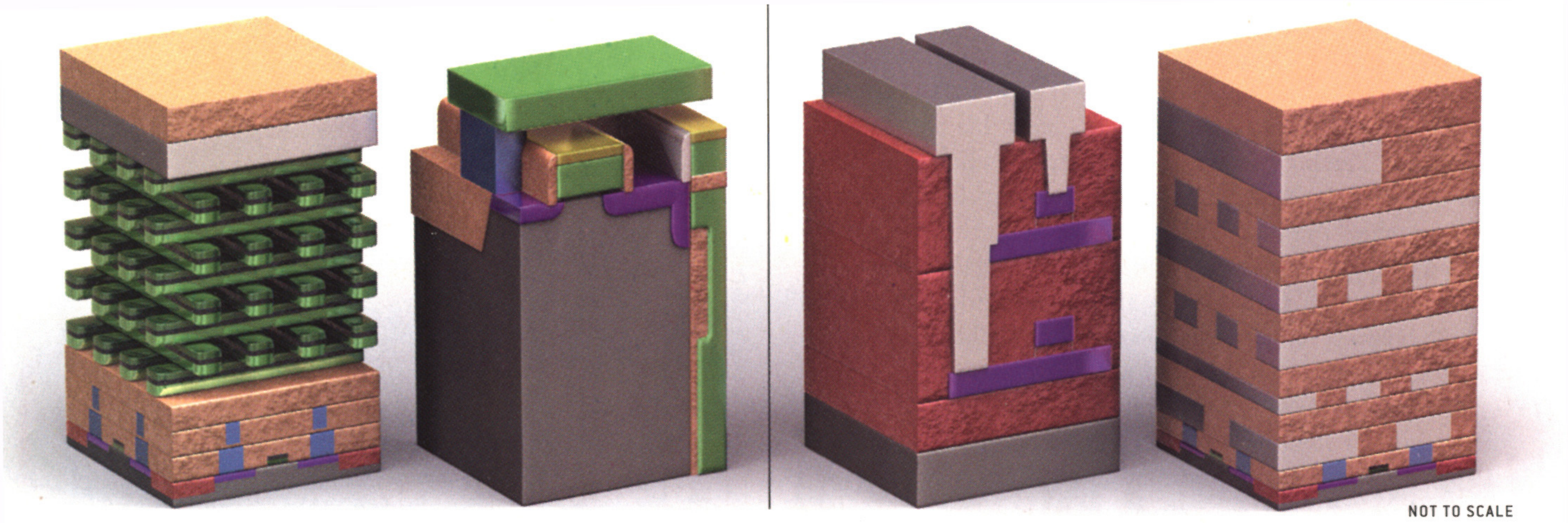
Single-phase, miniaturized convective cooling



Distributed return architecture with cross section showing inlet jets with neighboring drainage holes.

Interior Structure of 3-D chips

3-D Volatile Memory [Matrix Semiconductor]	2-D Random-Access Memory [IBM 256-Megabit]	3-D Logic Circuit [Lab Prototype]	2-D Microprocessor [Advanced Micro Devices Athlon]



- Monosilicon substrate
- Insulators
- Aluminum wires
- Polysilicon
- Tungsten plugs
- Ion-doped silicon
- Isolation oxides
- Silicide

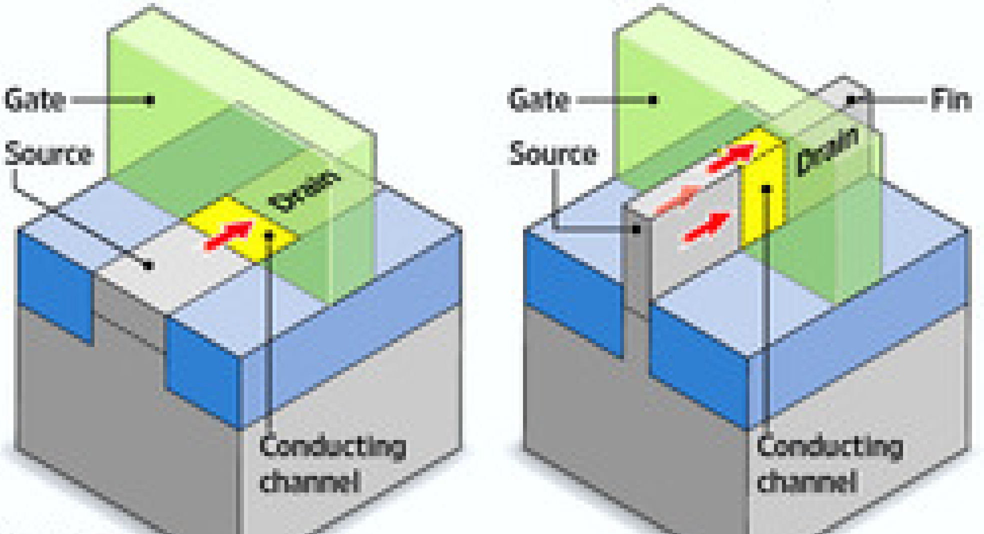
Intel’s 3D Transistor 2011

Intel’s Move Into 3-D

The chip maker breaks from conventional approaches to make transistors.

Conventional transistor: Electrons flow between components called a **source** and a **drain**, forming a two-dimensional **conducting channel**. A component called a **gate** starts and stops the flow, switching a transistor on or off.

Intel’s new transistor: A **fin-like structure** rises above the surface of the transistor with the **gate** wrapped around it, forming **conducting channels** on three sides. The design takes less space on a chip, and improves speed and reduces power consumption.



Source: Intel

“Every economic era is based on a key abundance and a key scarcity.”

George Gilder,
Forbes ASAP, 1992

What are the key scarcities?

Computer Processing
Case Studies

NBA 6120

January 29, 2018

Donald P. Greenberg

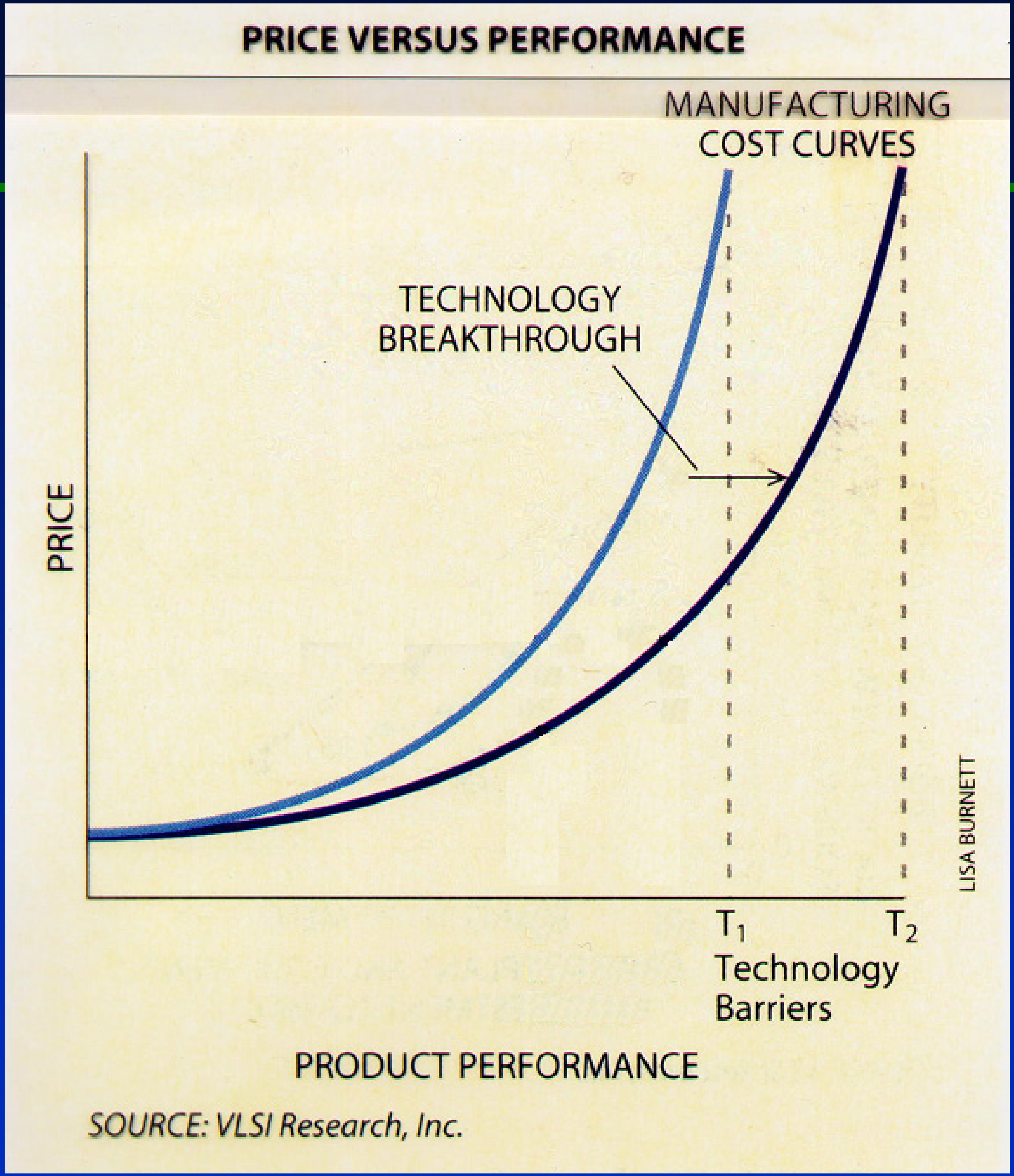
Lecture 2

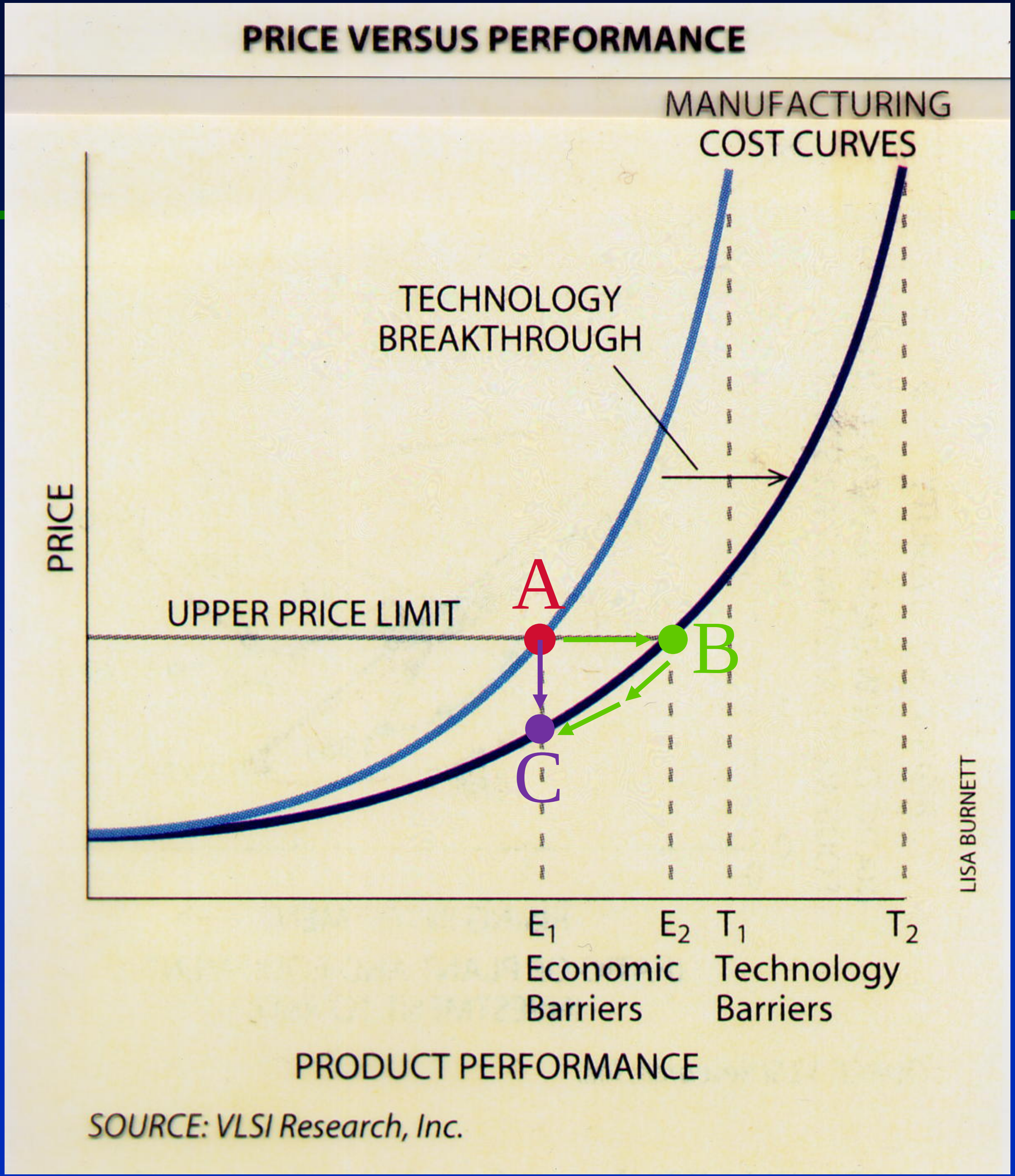
Required Reading

- G. Dan Hutcheson and Jerry D. Hutcheson. Technology & Economics in the Semiconductor Industry, Scientific American, January 1996.

Economics of the Semiconductor Industry

- G. Dan Hutcheson and Jerry D. Hutcheson. Technology & Economics in the Semiconductor Industry, Scientific American, January 1996.


PRICE VERSUS PERFORMANCE
MANUFACTURING
COST CURVES
TECHNOLOGY
BREAKTHROUGH
PRICE
LISA BURNETT
T1
T2
Technology
Barriers
PRODUCT PERFORMANCE
SOURCE: VLSI Research, Inc.


PRICE VERSUS PERFORMANCE
MANUFACTURING
COST CURVES
TECHNOLOGY
BREAKTHROUGH
PRICE
UPPER PRICE LIMIT
A
B
C
E1
E2
T1
T2
Economic
Barriers
Technology
Barriers
PRODUCT PERFORMANCE
LISA BURNETT
SOURCE: VLSI Research, Inc.

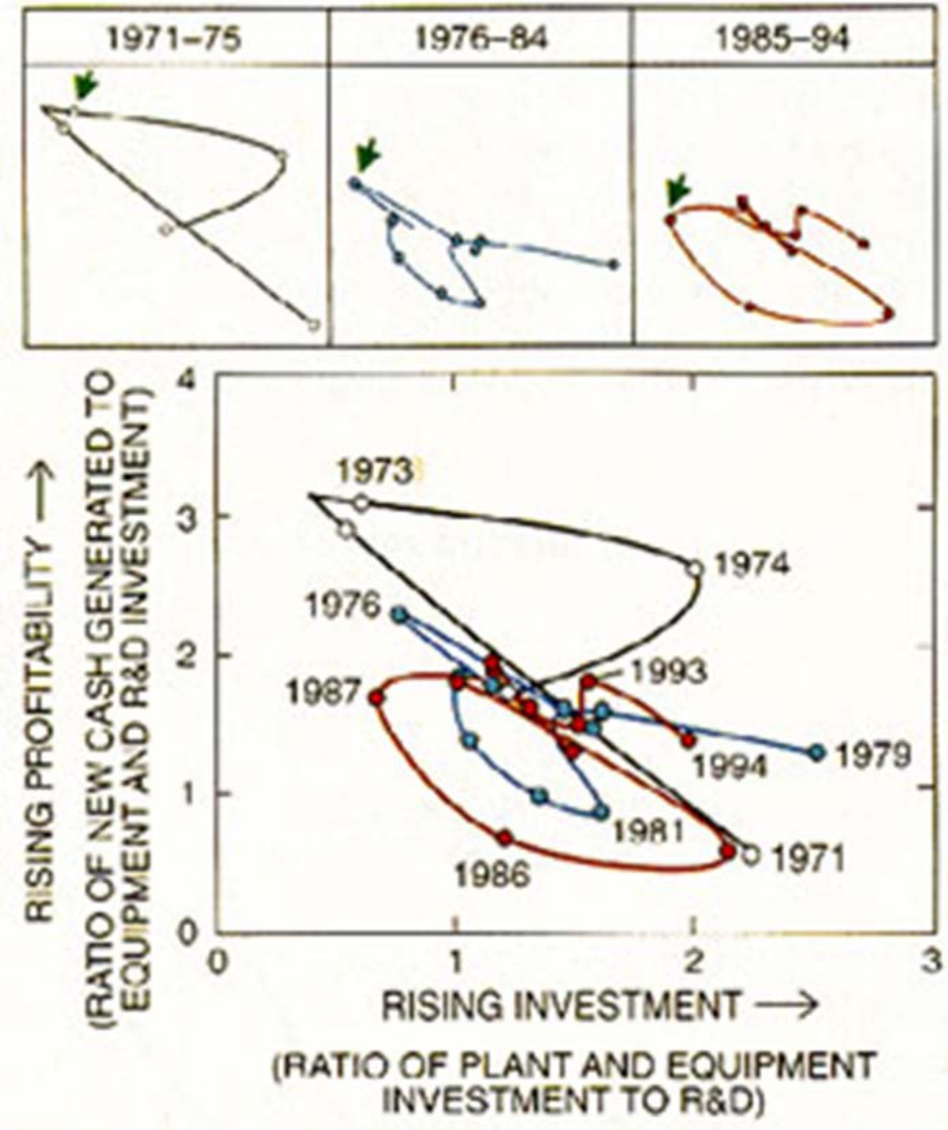
How do you predict what the technology, manufacturing cost, market demand, market supply, and competition will be five years in the future?

Return on Investment (ROI) Model does not work well

Difficulties:

- How long does the product last?
- What is the price (revenue)/unit?
- Exponential change
- Non-linear pricing behavior
- Competition (monopoly pricing)
- Prediction of demand
- Technical obstacles

Profitability vs. Investment in the Computer Industry



Profitability vs. Investment in the Computer Industry

Rising Profitability

$$\text{Measured by ratio} = \frac{\text{cash generated during year}}{\text{investments made in new technology previous year}}$$

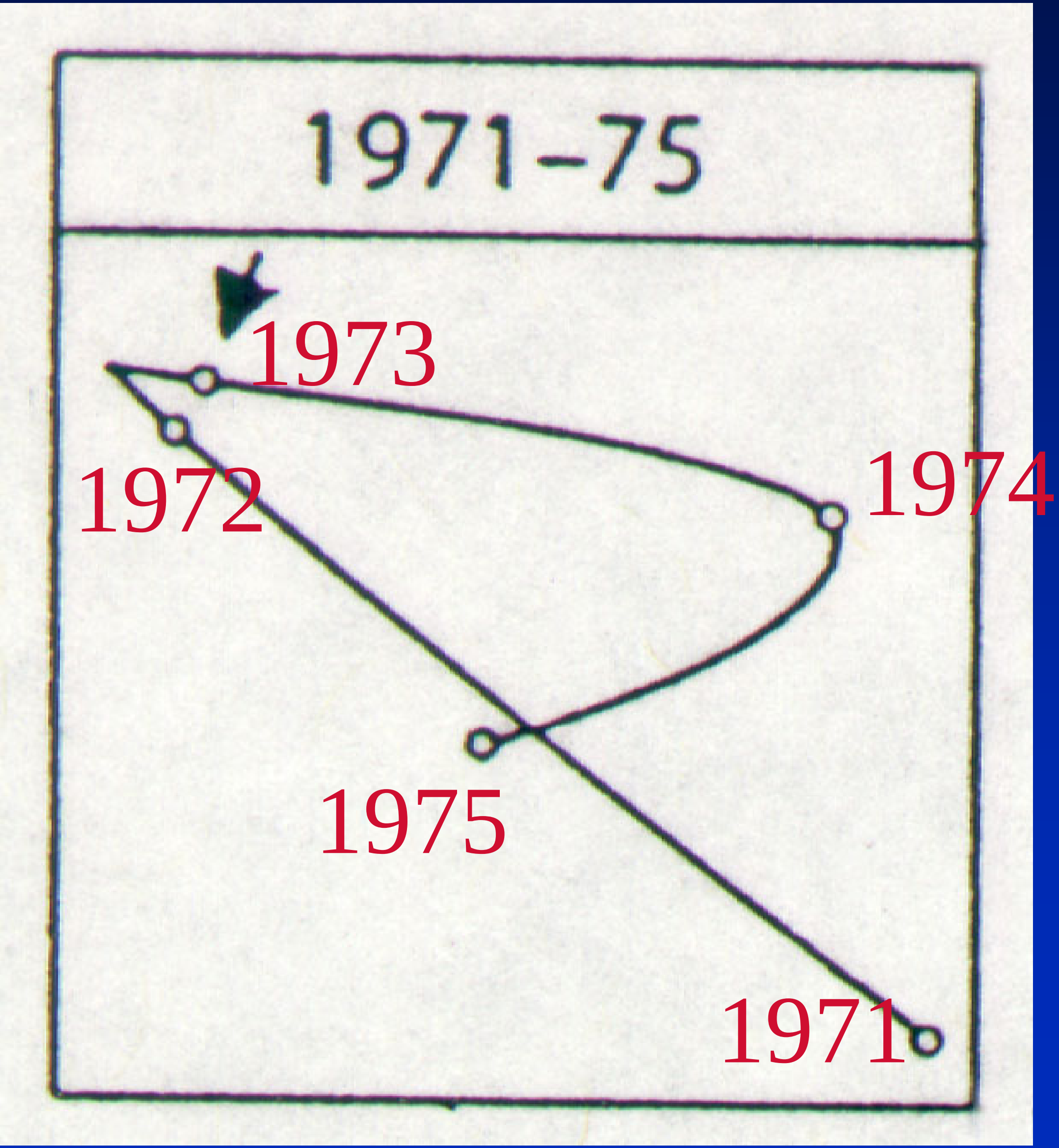
$$\text{where new technology} = \text{new equipment} + \text{R \& D}$$

$$\text{cash} = \text{gross profit (including R \& D)}$$

Rising Investment

$$\text{Measured by ratio} = \frac{\text{plant \& equipment investment}}{\text{R \& D}}$$

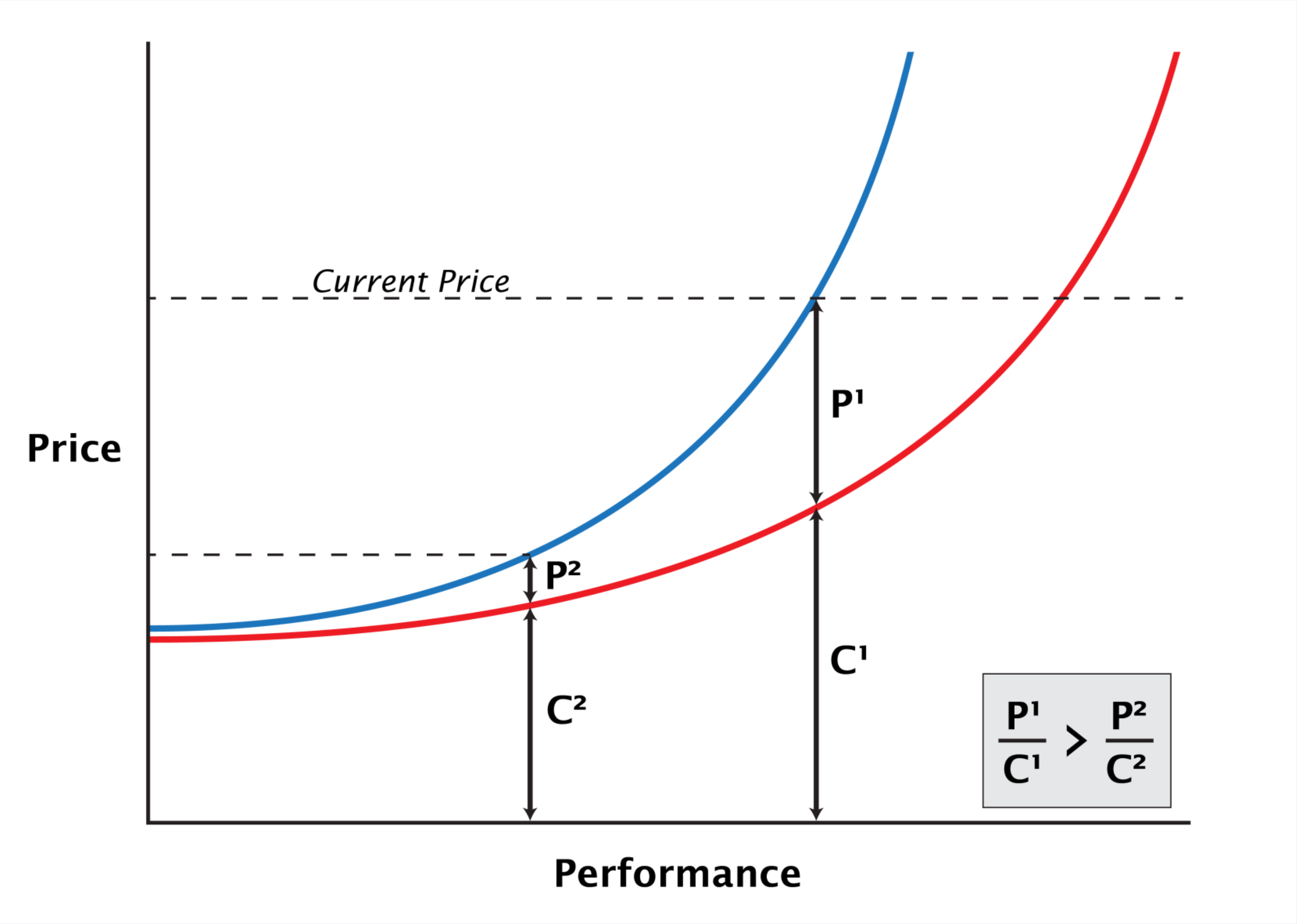
Profitability vs. Investment



Profitability vs. Investment in the Computer Industry

- It is obvious that with the shrinking technology, it is getting more expensive to move to the next generation process technology.
- It is also obvious that the manufacturing cost as well as the sales price of processing chips is decreasing rapidly.

Diminishing Profitability



Andrew S. Grove, Chief Executive and Chairman of Intel Corporation

From the New York Times, caption:
“Mr. Grove in 1991 with a silicon wafer, part of the process to make Intel’s 386 microprocessor.”

9/2/1936 - 3/21/2016

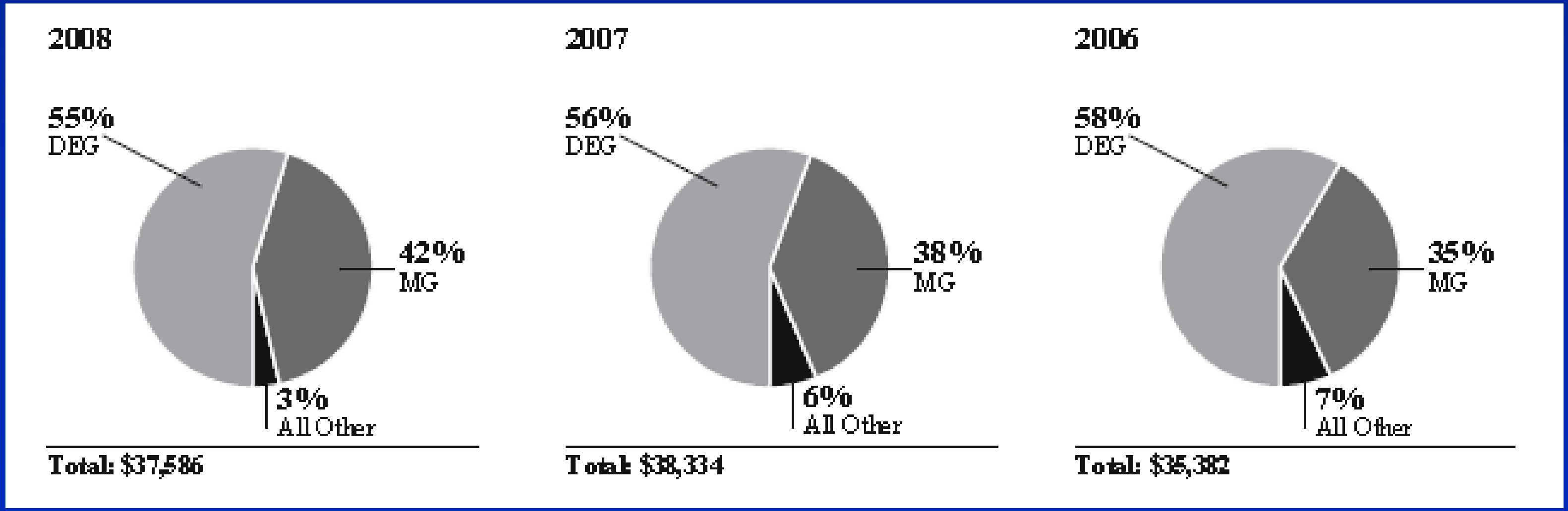
Impact of Changing Customer Demand

Intel 2007

- Mobile microprocessors ASP's are less than desktop microprocessor ASP's.
- In 2007 gross margins were negatively impacted by declining ASP's and higher start-up costs for the new 45nm process technology.
- At the end of 2007, Intel had roughly $20B cash.

Intel 2008

- In 2008 the average selling price for all products continued to decline
- The revenues for the mobility group as contrasted to the digital enterprise group continued to increase



Percentage of Revenue
(Dollars in Millions)

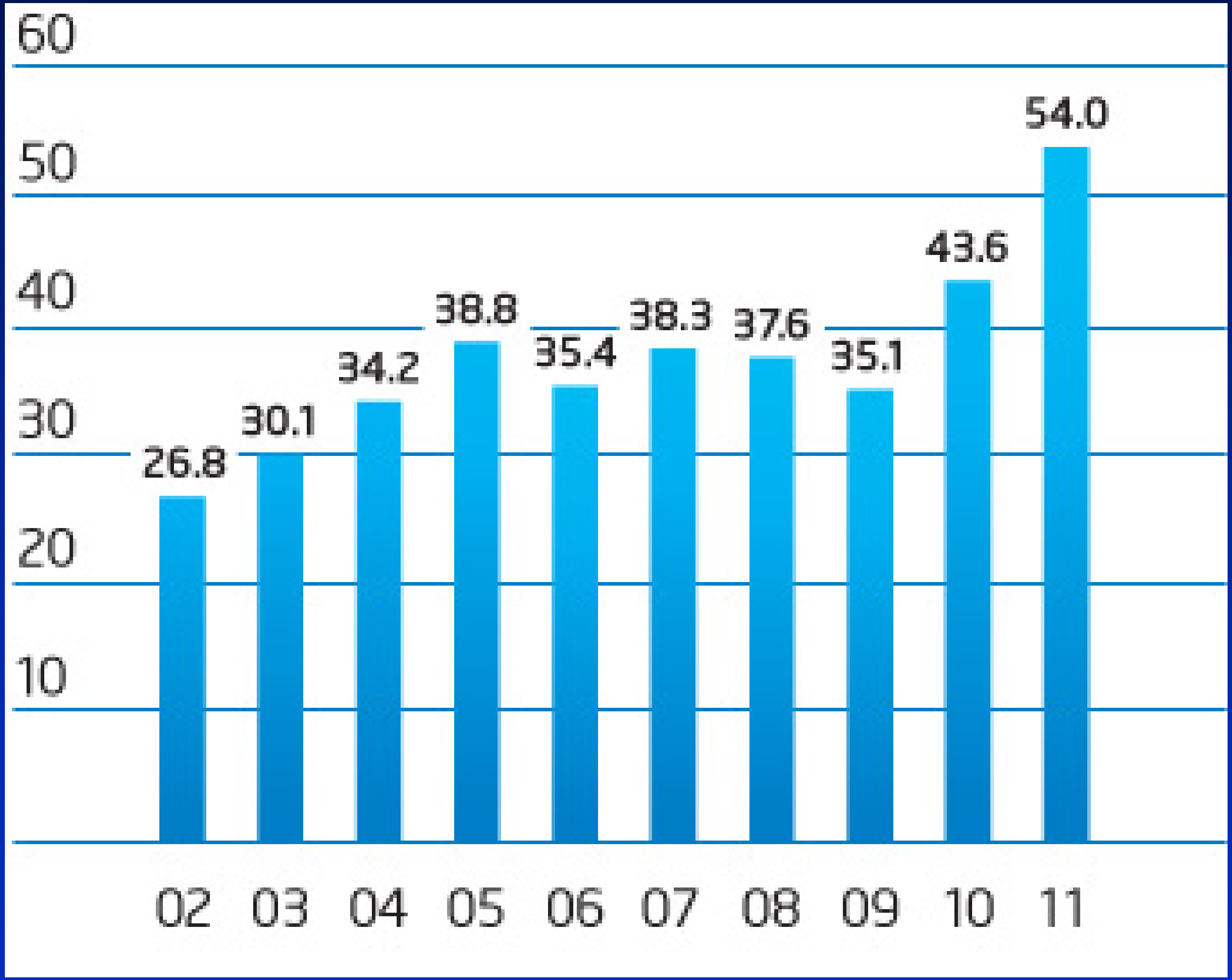
Intel 2011

- A new fab costs approximately $3-4B or more
- Should Intel Continue to Invest In Creating New Fabrication Facilities?

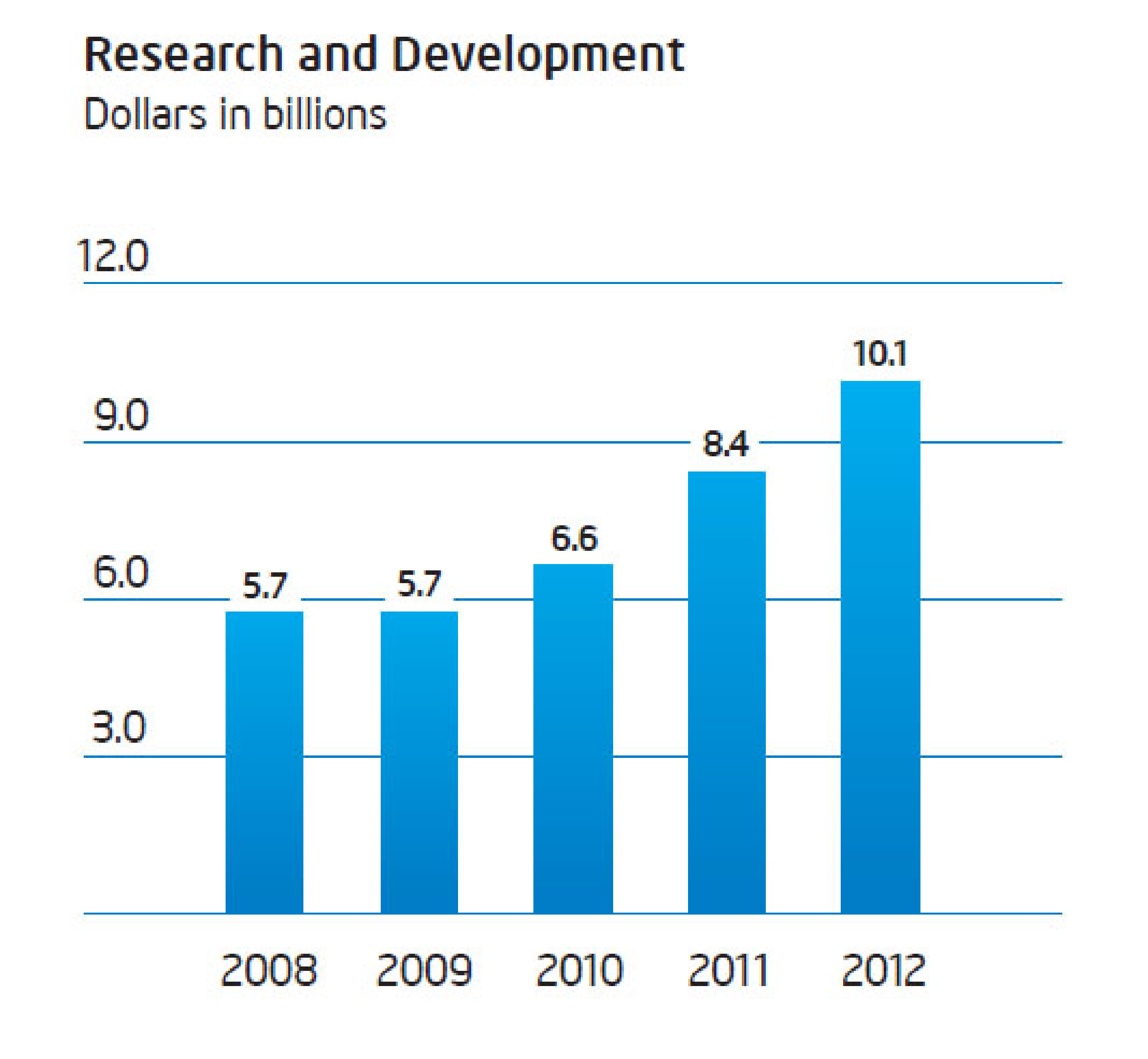
Intel’s In a Sweet Spot 2011

- Having invested in its 32nm fab, Intel achieved higher than expected efficiencies and introduced new chips faster than expected.
- Sandy Bridge, their latest microprocessors was introduced in 2011.
- AMD, even if it designed better chips, was stuck with its 45nm production and couldn’t compete. Their chips were more expensive to produce.
- Intel’s new chips possibly eroded the graphics market for competitors (nVidia & AMD) as PC makers no longer needed stand-alone graphics processors.

Intel Net Revenue 2011

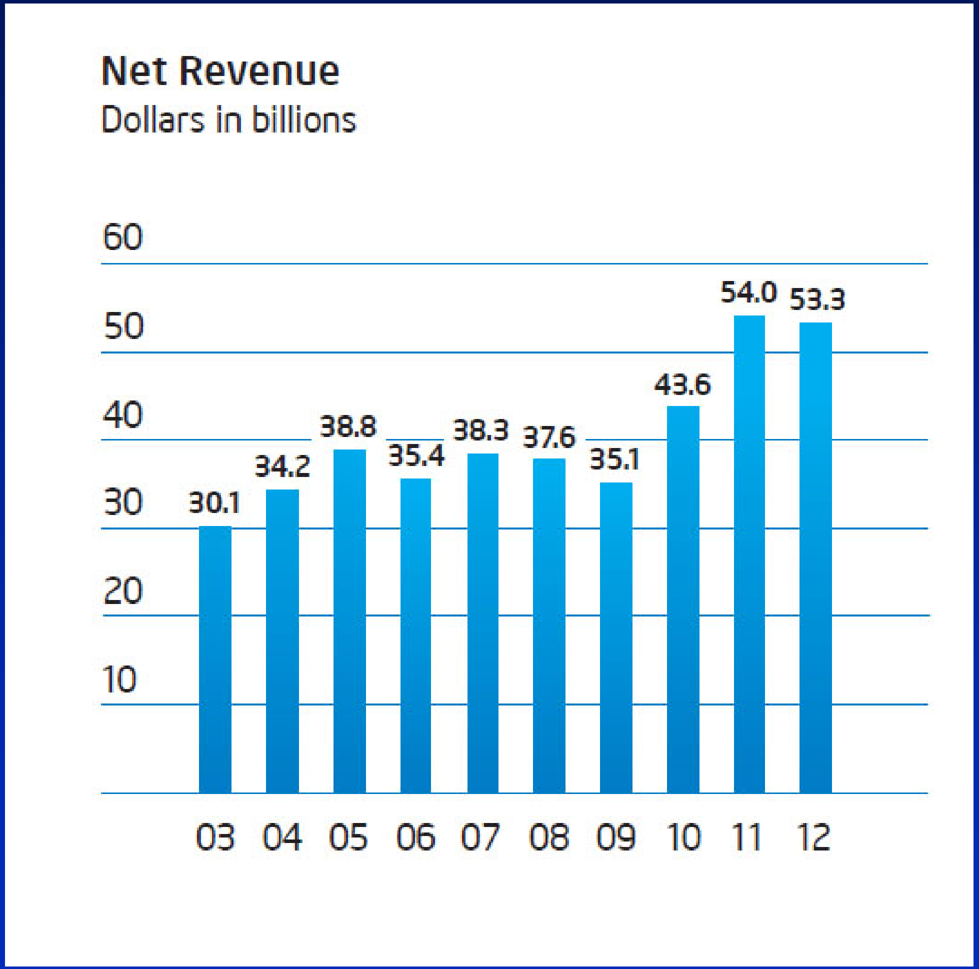


Intel Research & Development 2012

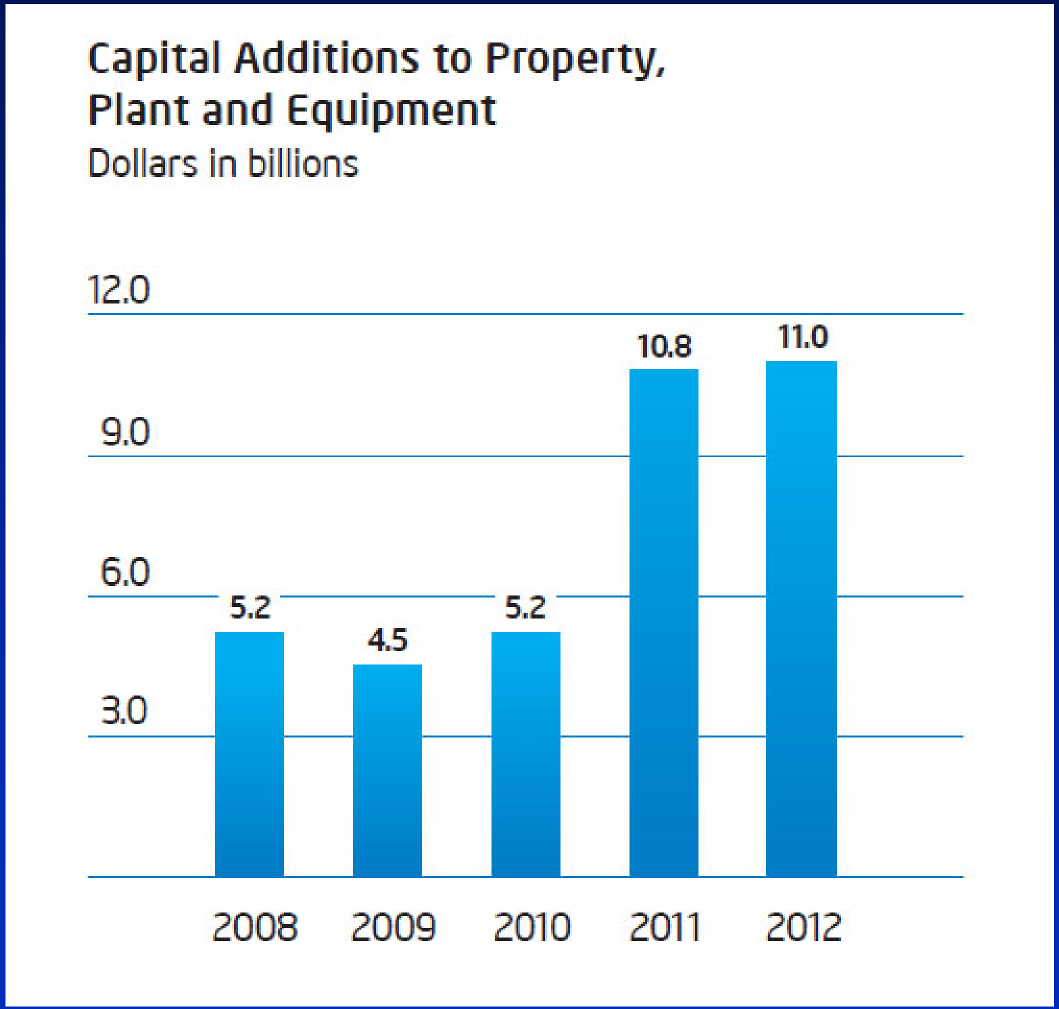


Obama at Intel 2012

Intel Net Revenue 2012



Intel 2012



Capital Additions to Property, Plant and Equipment
Dollars in billions
12.0
9.0
6.0
3.0
5.2
4.5
5.2
10.8
11.0
2008
2009
2010
2011
2012

Intel 2012

- Intel announced that it would spend $9B to upgrade four fabrication plants to move to 22nm technology (one in Israel).
- ARM and IBM announced a joint agreement to move to 14nm technology.

Computer Industry Problem 2013

- The high price servers are representing a much smaller percentage of revenue stream
- The prices of laptops and netbook computers are continuing to decrease
- Competition and price wars in the mobile computing segments (mobile phones, smart devices, tablets) are fierce

Intel cancels 14nm Fab 42 in AZ, due to increasing competition from ARM

January 2014, ExtremeTech.com

Intel 2015

- Intel again delays 10nm technology. It will depend on revenue increase from Windows 10 and its new Skylake processor.
- The second generation of 14nm production technology had significant yield improvements.
- At the same time, Intel moved to purchase Altera so it could shift from PC's to mobile devices.

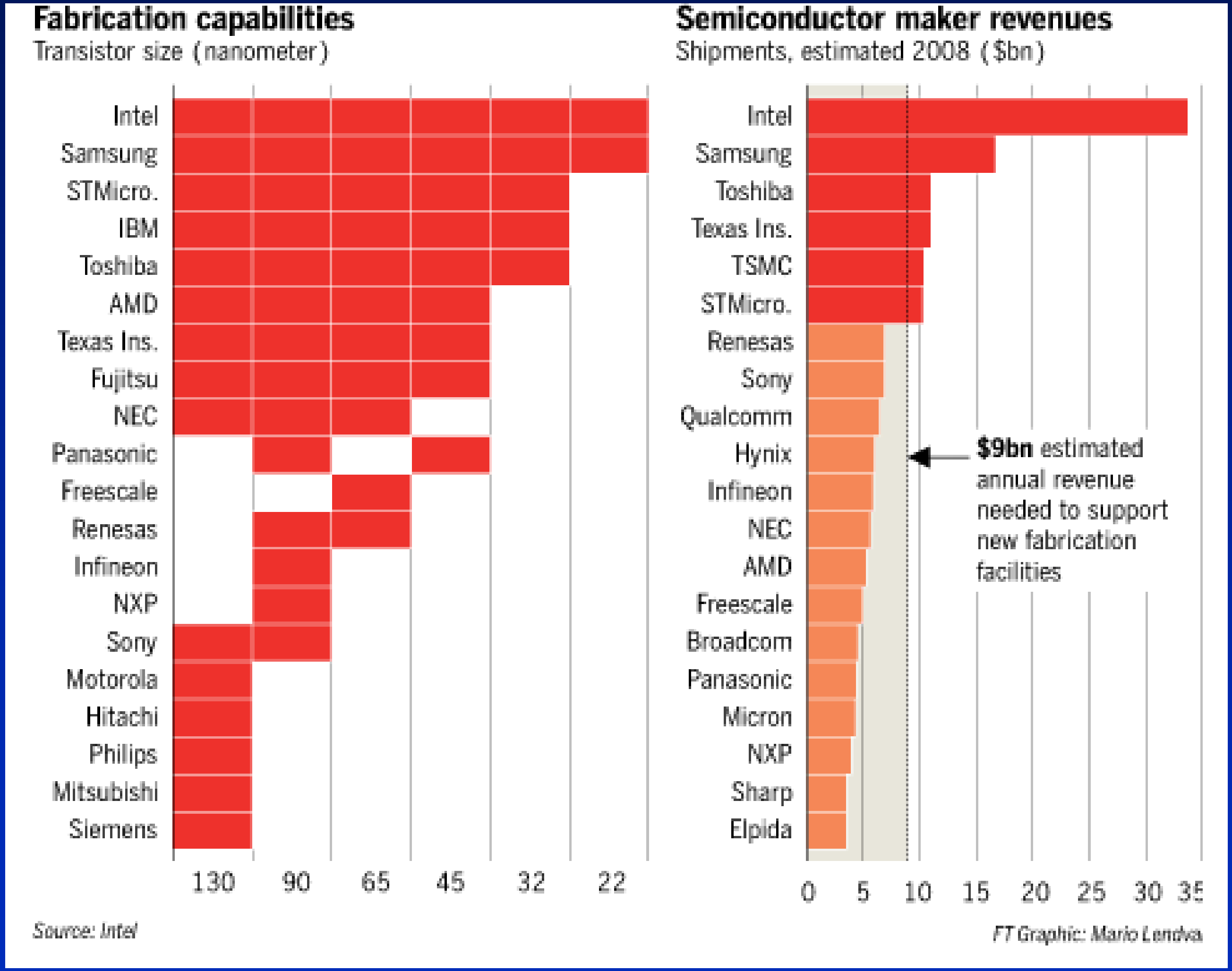
Intel's $7B Investment

Wall Street Journal, Feb. 8th, 2017

Potential Plans

- 7 nanometer chip technology
- 5 G Networks
- Drones

Fewer companies can deliver smaller and more powerful chips July 20, 2009



Foundry Model

- Many companies (Integrated Device Manufacturers, IDMs) design and manufacture integrated circuits (efficiency through vertical integration)
- Today, there are many companies that:
  - only design devices (fabless semiconductor companies),
  - as well as ***merchant foundries*** that only manufacture devices.
- The ***foundry model*** is a business vision that seeks to optimize productivity.
- In 1987, the world's first dedicated merchant foundry opened its doors: Taiwan Semiconductor Manufacturing Company (TSMC)

source: https://en.wikipedia.org/wiki/Foundry_model

TSMC's Customers

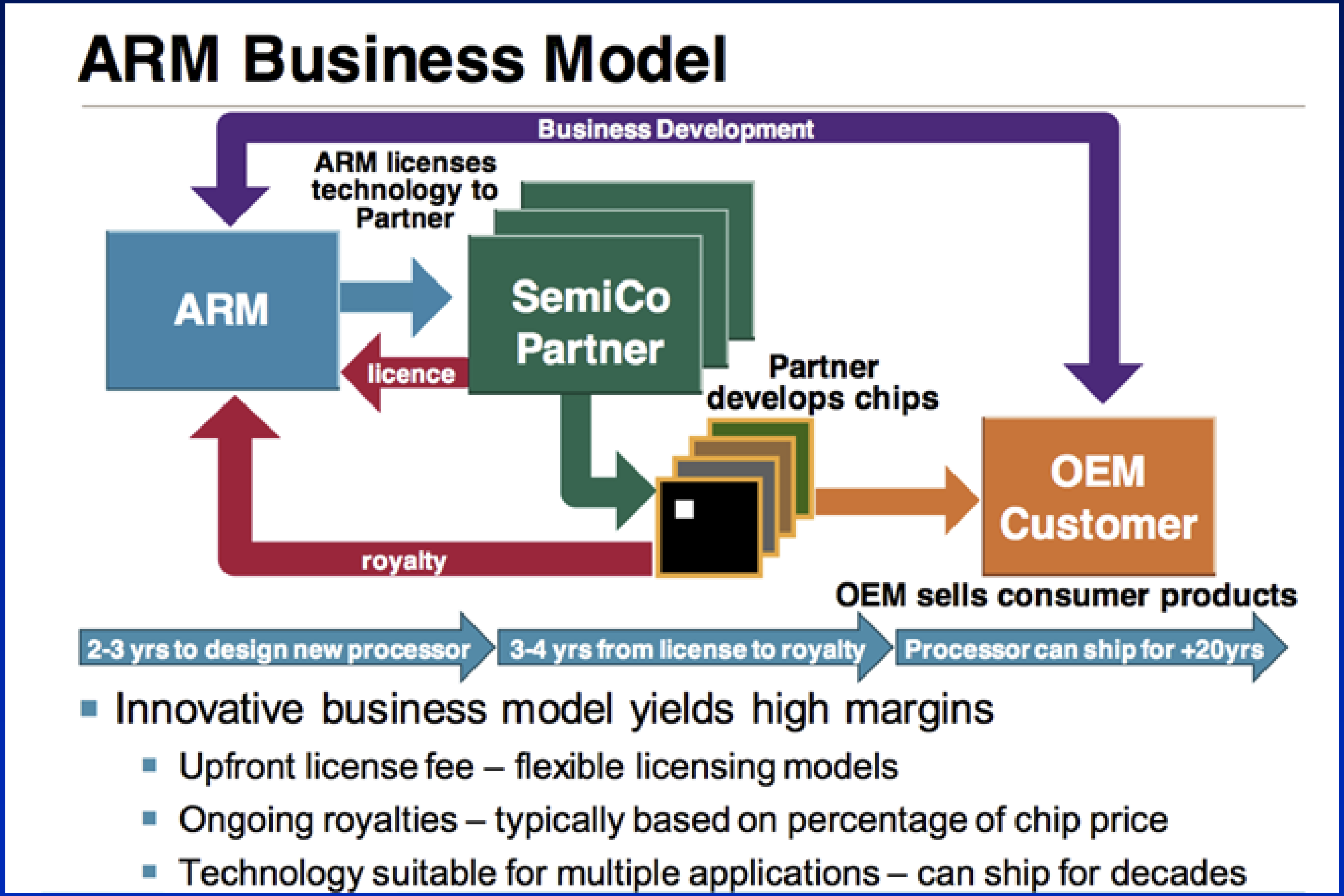
- Manufacture's chips for
  - Qualcomm
  - Nvidia
  - Advanced Micro Devices (AMD)
  - Broadcom, Altera
    - > (even some for Intel & Texas Instruments)
  - Apple's A5, A6 for iPad & iPhone
  - Apple's new A8

source: https://en.wikipedia.org/wiki/TSMC

TSMC's Fabrication Plants 2014

- TSMC had four 300mm wafer plants in Taiwan
- TSMC had four 200mm wafer plants in Taiwan
- TSMC had one 200mm wafer plant in Shanghai, Washington State, Singapore, and other smaller plants.

ARM Holdings - Business Model



Shimpi, Anand Lal. "The ARM Diaries, Part 1: How ARM's Business Model Works." *Anand Tech*. N.p., 28 June 2013. Web. 31 Aug. 2015.

ARM Holdings

- Original name was Acorn Computers
- In 1990 a new customer arrived, Apple: and company was renamed Advanced RISC Machines (ARM)

“Watts are more important then MIPS of FLOPS”

- George Gilder

ARM’s Customers

- Apple (iPhone 5, iPad, iPhone 5s, iPhone 6, etc.)
- Samsung (Galaxy S4, S5, etc.)
- Qualcomm (Snapdragon)

Japan’s Softbank Purchased ARM For $32B 2016

- Influenced by the growing “Internet of Things” (IOT)
- Price was greater than 40% over the closing stock price

Predicting Demand

How do you predict what the technology, manufacturing cost, market demand, market supply, and competition will be five years in the future?

CASE STUDY 1:
The Great Chip Glut: Economist August 11, 2001

- East Asia did not understand the industry's woes
  - Oversupply
  - Taiwan's "foundries"
  - TSMC
  - UMC

  ⇒ Operating at 30% of capacity (from 70%)

  - Singapore – Charted Semiconductor
  - Korea's Hynix (Hyundai) - $1B loss in 2Q01
  - Malaysia – new fab, 1st Silicon + 2 more
  - China – Shanghai alone, 2 fabs under construction

    2 more on drawing board

    12 more planned

Case Study #2

Intel's MMX Introduction

Microprocessor Report, July 1997

Marketing & Advertising Strategies in the Computer Industry

- *In a fast moving technology, how do you market your product?*
- *How do you get brand name recognition?*
- *When do you start advertising?*

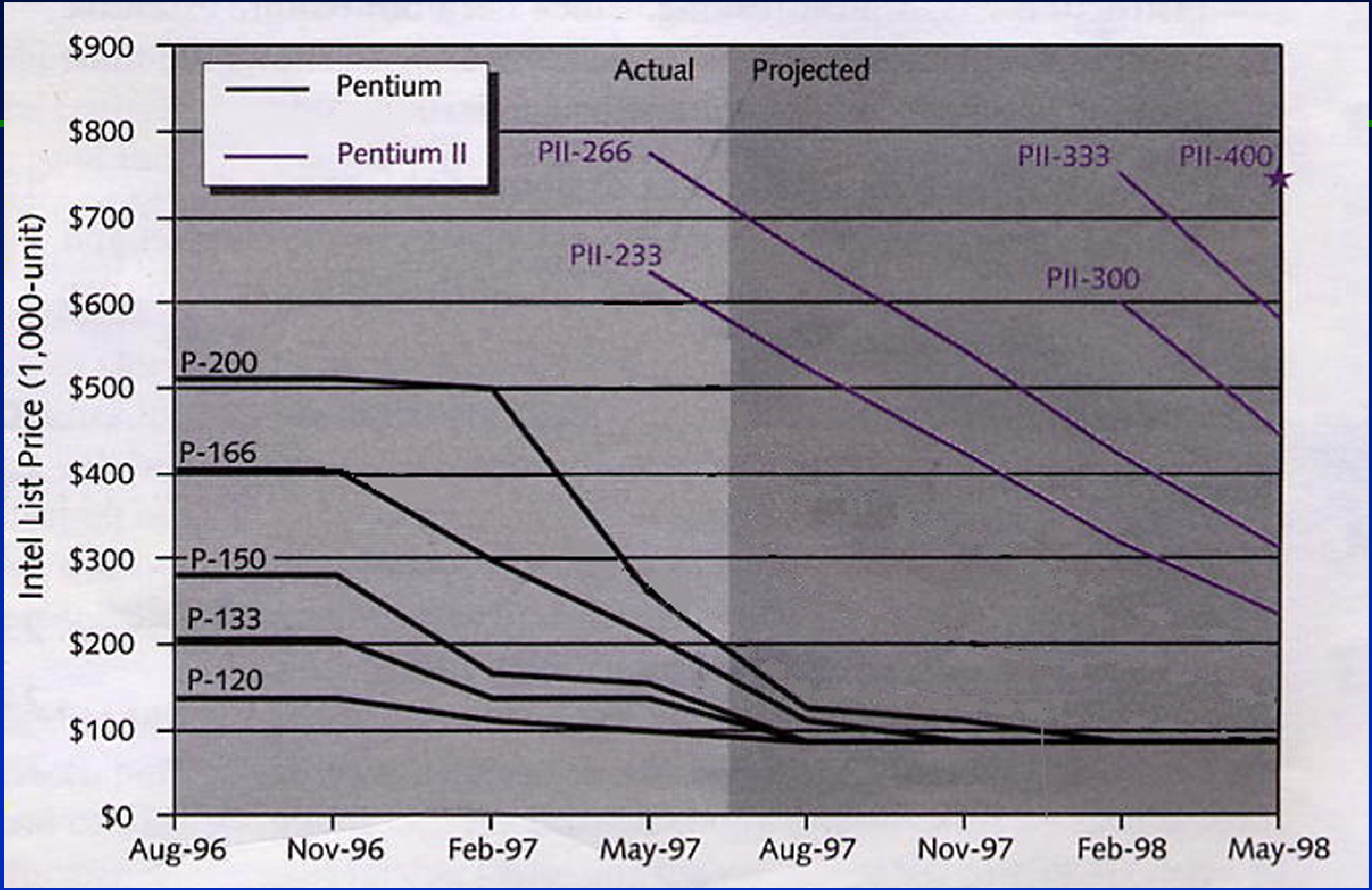
What is MMX?

- First major extension to x86 instruction set since 1985
- 57 new instructions to accelerate:

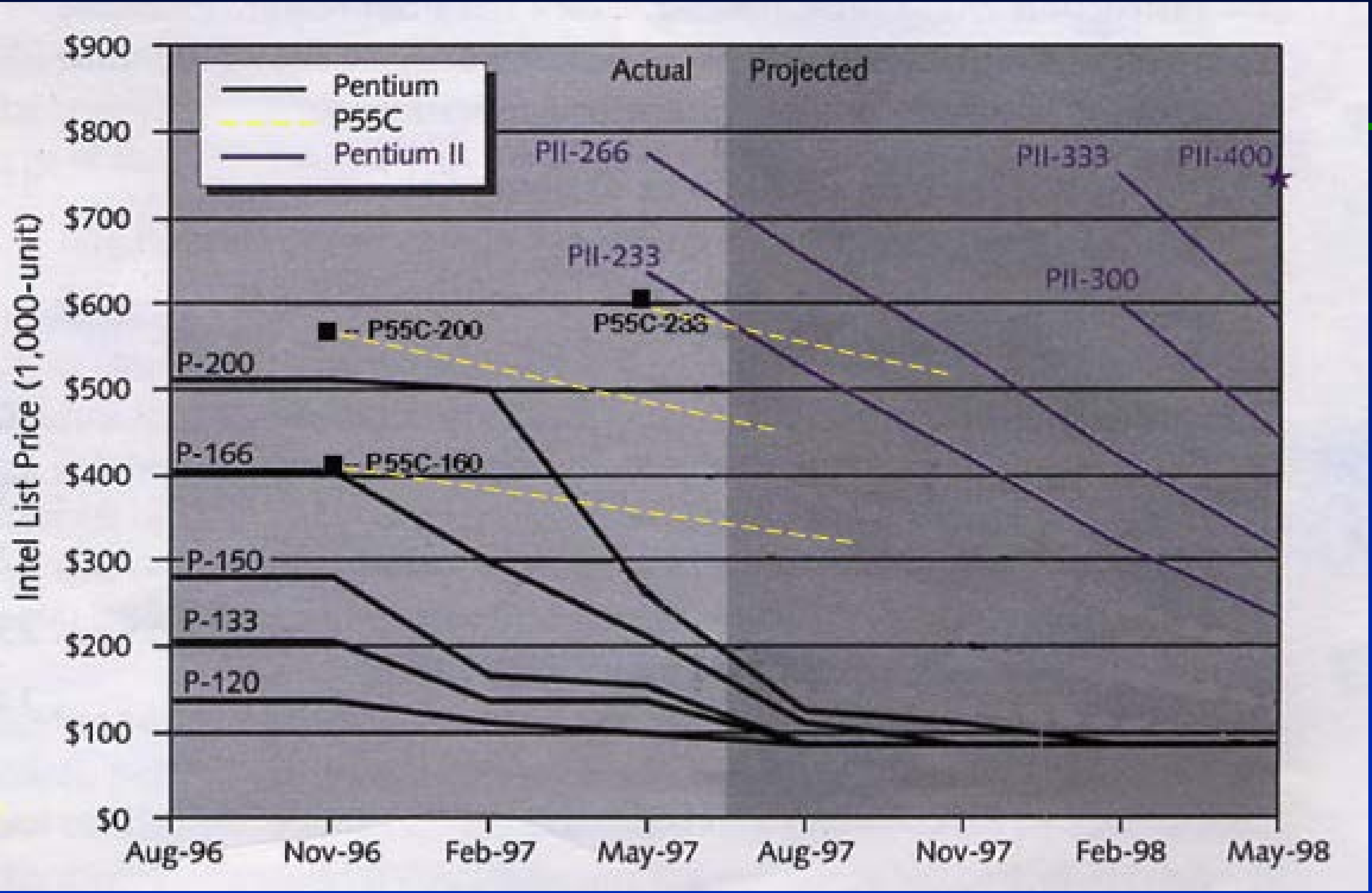
  2D & 3D graphics

  Video

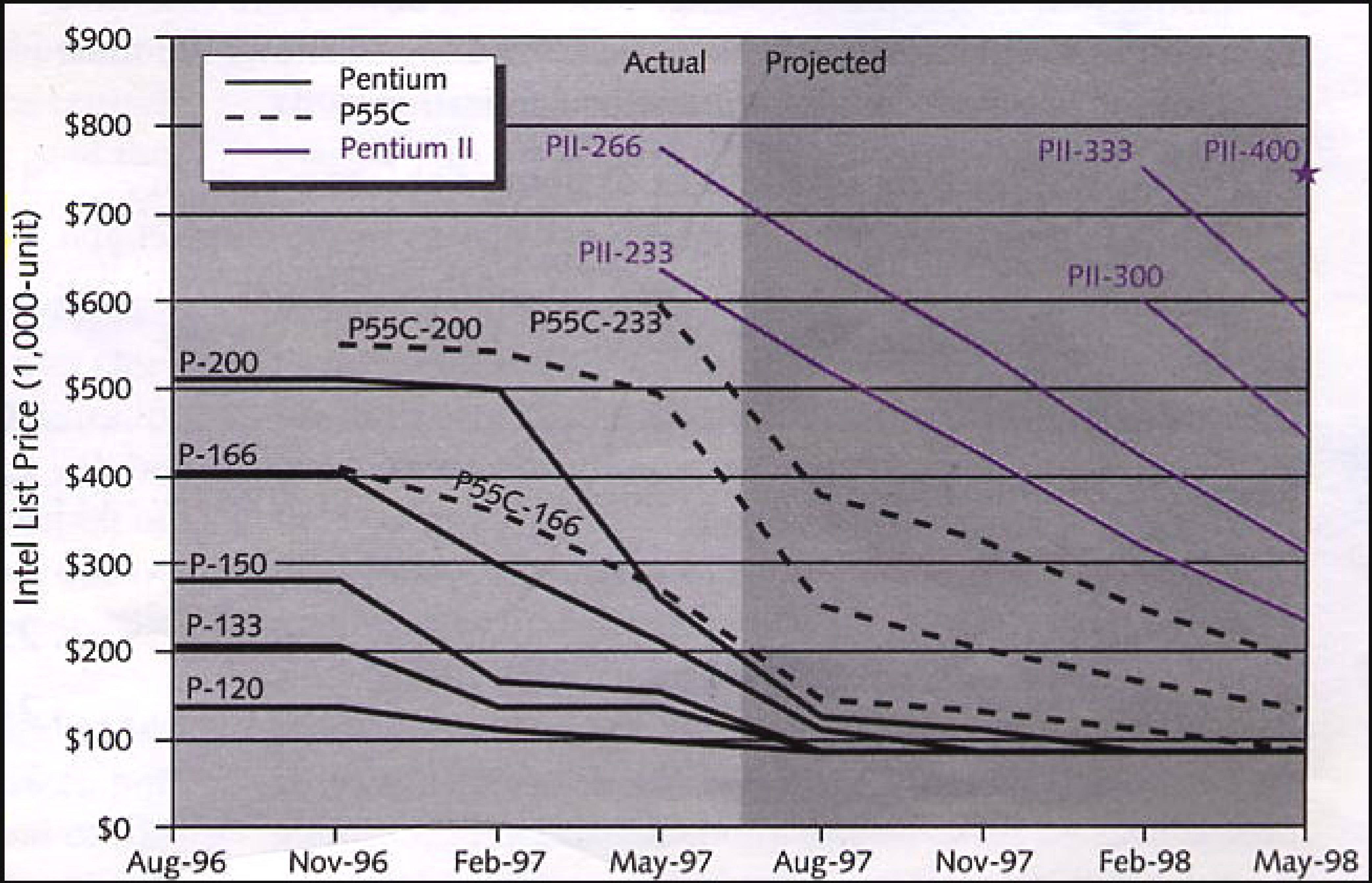
  Speech synthesis and recognition



Pentium
Pentium II
Actual
Projected
PII-266
PII-233
PII-333
PII-400
PII-300
P-200
P-166
P-150
P-133
P-120
Intel List Price (1,000-unit)
$900
$800
$700
$600
$500
$400
$300
$200
$100
$0
Aug-96
Nov-96
Feb-97
May-97
Aug-97
Nov-97
Feb-98
May-98



$900
$800
$700
$600
$500
$400
$300
$200
$100
$0
Intel List Price (1,000-unit)
Pentium
P55C
Pentium II
Actual
Projected
PII-266
PII-233
PII-333
PII-400
PII-300
P55C-200
P55C-233
P55C-160
P-200
P-166
P-150
P-133
P-120
Aug-96
Nov-96
Feb-97
May-97
Aug-97
Nov-97
Feb-98
May-98



Pentium
P55C
Pentium II
Actual
Projected
PII-266
PII-333
PII-400
PII-233
PII-300
P55C-200
P55C-233
P-200
P-166
P55C-166
P-150
P-133
P-120
Intel List Price (1,000-unit)
$900
$800
$700
$600
$500
$400
$300
$200
$100
$0
Aug-96
Nov-96
Feb-97
May-97
Aug-97
Nov-97
Feb-98
May-98

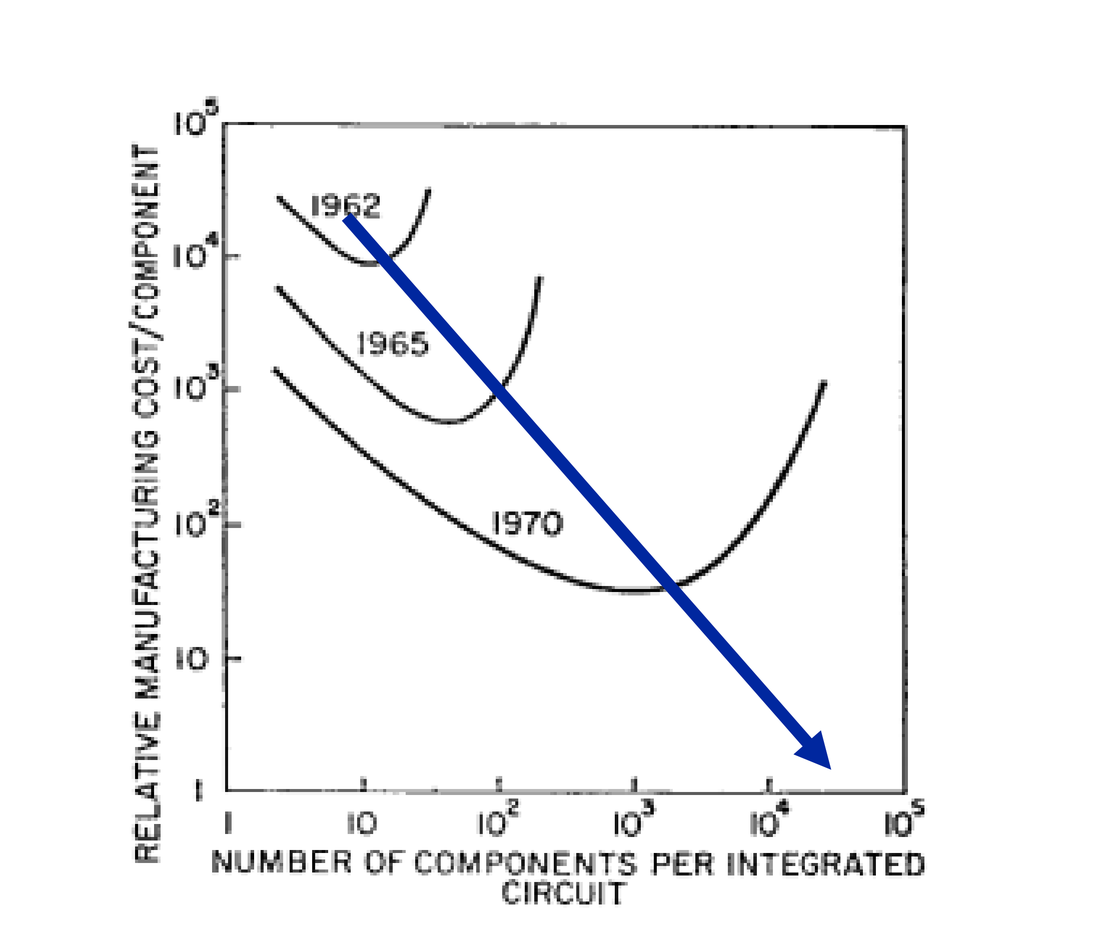
Lessons Learned?

- Need to completely integrate new product development, production capacity, advertising and marketing
- New products need to be introduced frequently to keep ASP constant or at high levels
- Case explains the drive for continually shrinking technology

Moore's Original Article 1965

Source: http://web.eng.fiu.edu/npala/eee6397ex/gordon_moore_1965_article.pdf

Moore’s Original Prediction 1965



"Every economic era is based on a key abundance and a key scarcity."

George Gilder,
Forbes ASAP, 1992

END. . .