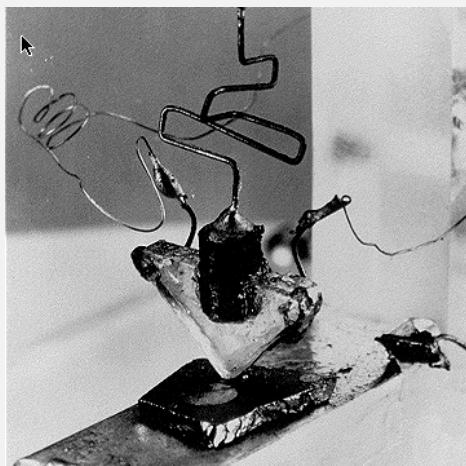
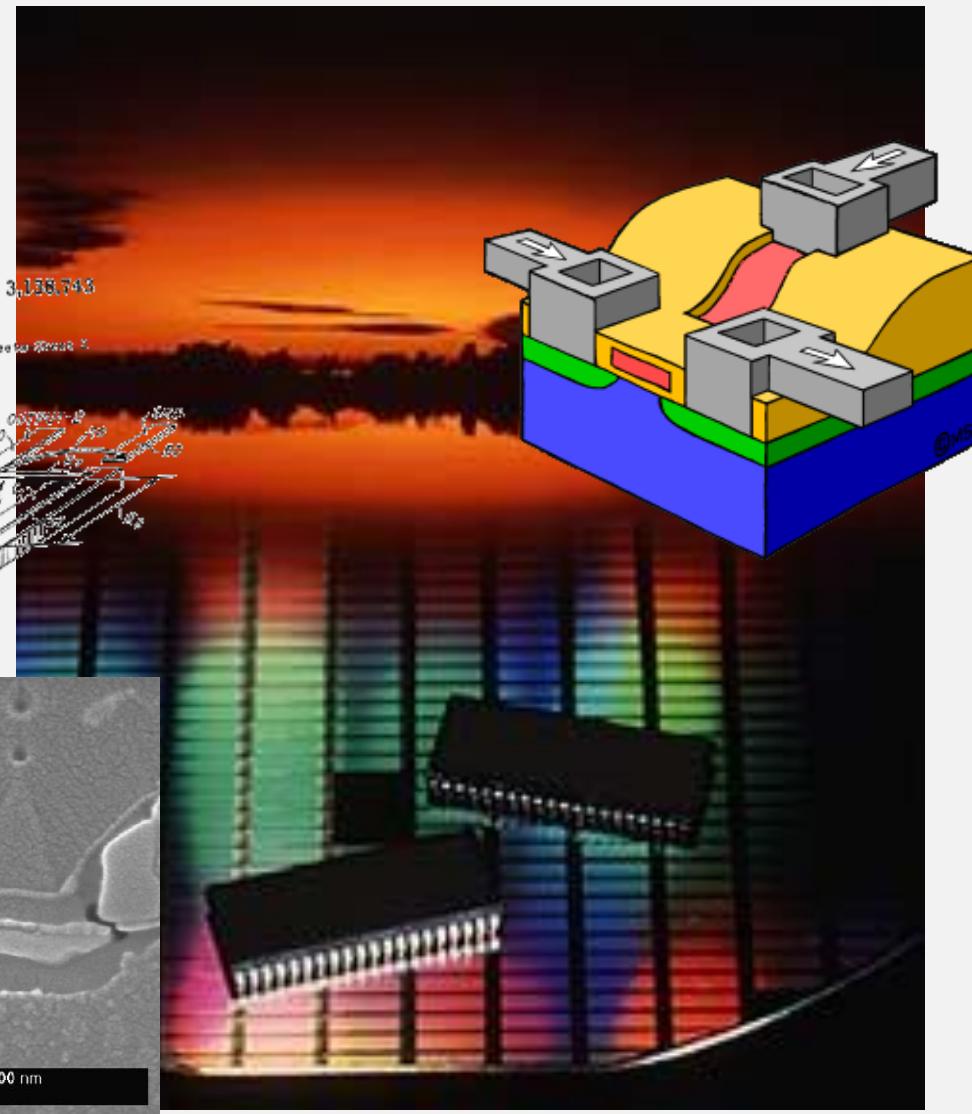




June 23, 1964
Filed Feb. 6, 1969

J. S. KILBY
MINIATURIZED ELECTRONIC CIRCUITS



1.1 Historical Review of Semiconductor Development

The Long Dawning of Hell-Made Semiconductors

The Impact of Bell Labs

From First Transistors to Today's ICs

1.2 Basic Device Principles and Family Tree

Basic Device Principles

Family Tree

1.3 MOSFETs as Mainstream Devices

Needs of Information Society

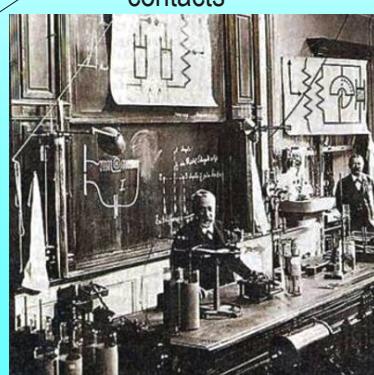
Advantages of MOSFETs for Integration

Future Calculability of MOSFETs

1752 Franklin: Lightning rod experiments
1798 Galvani: Frog's leg experiments

Current transport in gases
1860 Plücker: current transport in gases since 1860 Maxwell: theory of elmag.waves
1870 Hittorf: deflection of current due to fields -> current = neg. particles
1883 Edison: explanation of glow-emission 1886 Hertz: creation and proof of elmag waves
1887 Hertz/Hallwachs: Photo effect 1892 Wien: e/m 1895 Thomson: Identification of the electron 1900 Planck: "quantum constant" 1905 Einstein: explanation of photo effect
 Quantum Physics

Applications
1800 Volta: Battery
1806 Sömmerring: electrical telegraph with wire connection
1833 Weber/Gauß: el.mag. telegraph with wire connection
1854 Goebel: electric bulb with carbon filament
1860 Reis: telephon with wire connection
1866 Siemens: Dynamo
1876 Bell: microphon-telephon with wire connection
1878 Edison: Phonograph
1895 Braun: cathode ray tube
1895 Marconi: Telegraph with elmag.waves
1904 Fleming: discovery of rectifying effect in tube diode
1906 DeForest: tube triode
1913 Arnold: vacuum triode
Telephon-Radio-Tube-Market

Current transport in solid-states
1826 Ohm: $R=V / I$
1831 Faraday: El.mag. induction
1874 Braun: rectifying metal/semiconductor contacts


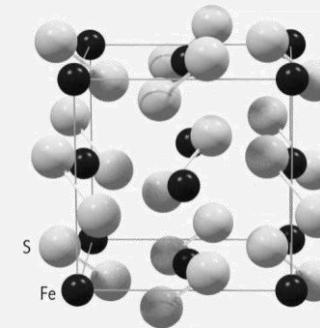
1914-1918: World War I

Advanced MOSFETs and Novel Devices

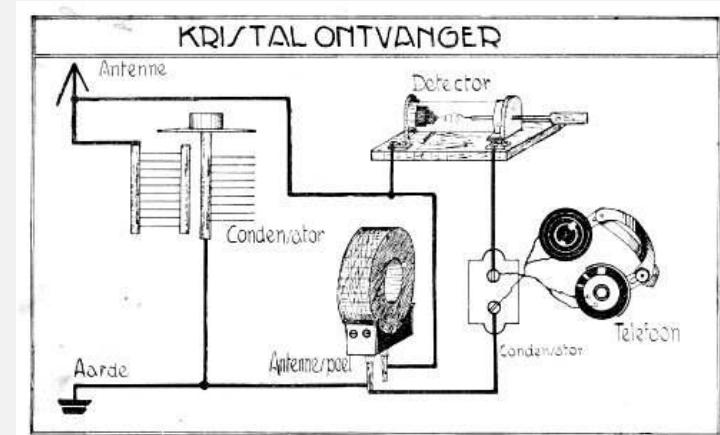


The mineral **pyrite**, or **iron pyrite**, is an iron sulfide with the formula FeS_2 . This mineral's metallic luster and pale-to-normal, brass-yellow hue have earned it the nickname **fool's gold** due to its resemblance to gold. The color has also led to the nicknames **brass**, **brazzle** and **brazil**, primarily used to refer to pyrite found in coal.

Pyrite is a natural semiconductor with a bandgap of 0.8 - 1.8 eV (depending on contamination). During the early years of the 20th century, pyrite was used as a mineral detector in radio receivers, and is still used by 'crystal radio' hobbyists. Until the vacuum tube matured, the crystal detector was the most sensitive and dependable detector available- with considerable variation between mineral types and even individual samples within a particular type of mineral. The most sensitive mineral was galena, which was very sensitive also to mechanical vibration, and easily knocked off the sensitive point; the most stable were perikon mineral pairs; and midway between was the pyrites detector, which is approximately as sensitive as a modern 1N34A diode detector.



Galena is the natural mineral form of lead sulfide. It is the most important lead ore mineral. Galena is one of the most abundant and widely distributed sulfide minerals. It crystallizes in the cubic crystal system often showing octahedral forms. Galena is a semiconductor with a small bandgap of about 0.4 eV which found use in early wireless communication systems. For example, it was used as the crystal in crystal radio sets, in which it was used as a point-contact diode to detect the radio signals. The galena crystal was used with a safety pin or similar sharp wire, which was known as a "cat's whisker".



1. Introduction

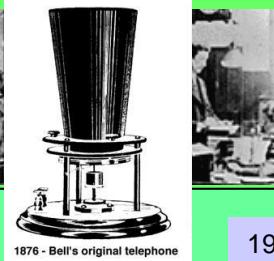
1.1. Historical Review

The Tube Quarrel

1875: Foundation of AT&T (Bell) (Telephon interconnections)



1877: Foundation of Bell-Company (fabrication of telephones)



1910: order for intercont. telephon cable

1913: AT&T buys deForest-patent

1913: Arnold develops optimized triode (vacuum, oxide-coated cathode, grid position)

1906: deForest invents tube-triode (with gas filling), -> amplification



1886: Foundation of Westinghouse Electric Company (Electric light)



1887: Foundation of General Electric Company (Electric light, power generators)

1887: Foundation of Marconi Wireless Telephone

1904: Fleming discovers tube-diode (-> rectification)



1913: Langmuir: tube-triode with vacuum + Hg-vapor + space charge theory

1919: GE takes over Marconi and founds with Westinghouse the RCA (Radio Company of America)

1920: First tube quarrel

AT&T
cable telegraphy/-telephony + radio tubes in radio telephony

New
„Entertainment Broadcasting“ (Radio)
(Using cable + wireless)

1922: 60'000 \$US market

.....

RCA
wireless telecommunication (mail, navigation) + tube-based end devices

The looser: AT&T

1925: Second tube quarrel

The winner: RCA

1925: Foundation of Bell-Labs

New tubes

Solid-state alternatives



1929: 850 Mill. \$US market

Advanced MOSFETs and Novel Devices

The lesson to be learned:

The foundation of Bell-Labs was a strategic „must“
for a big player not to lose a big commercial market



UNITED STATES PATENT OFFICE
EXCELSIOR, WASHINGTON, DISTRICT OF COLUMBIA.
TRADE-MARK FOR GRAMOPHONES.

No. 14,830
TRADE-MARK.
EMIL BERLIER,
GRAMOPHONES.
Registered July 10, 1900.

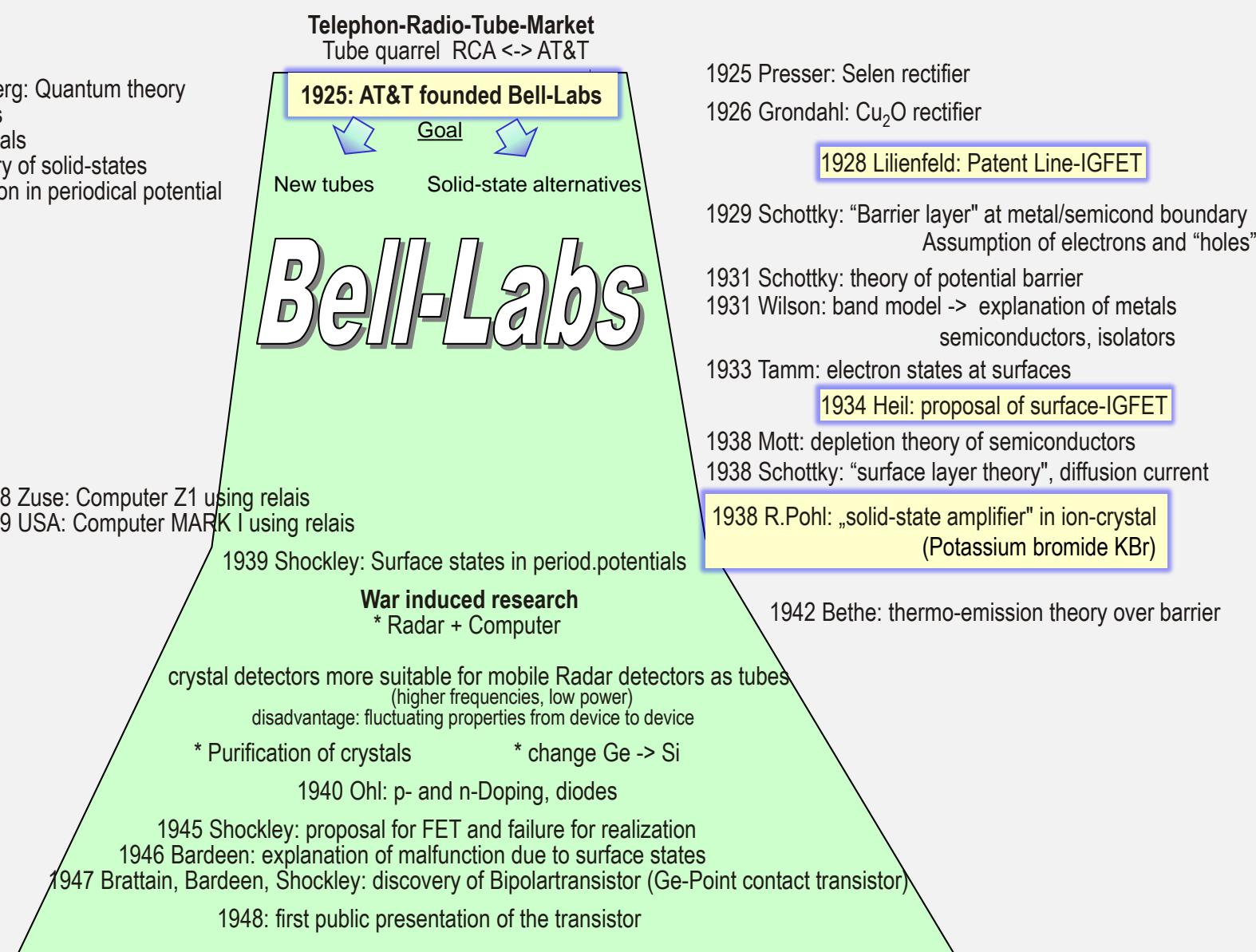
Emil Berliner
Manufacturing Co., Inc.



Painting by Francis Barraud, 1899
"His Master's Voice"



1925 Born, Schrödinger, Heisenberg: Quantum theory
 1926 Fermi, Dirac: Fermi-statistics
 1927 Pauli: Electron theory of metals
 1928 Sommerfeld: Quantum theory of solid-states
 1929 Bloch: solution of SG-equation in periodical potential



The First Solid-State Amplifier: invented 1938 in Germany

Zeitschrift für Physik, 111 (1938) 399-408

406

R. Hilsch und R. W. Pohl,

KBr-Kristall mit Steuergitter. Der KBr-Kristall mit nur einer enden Elektrode, also das Modell einer Sperrsicht, entspricht ektroden-Vakuumrohr mit einer glühenden Elektrode. Das wird durch Einbau einer dritten Elektrode, kurz Gitter genannt, gan.

rechender Weise haben wir in das KBr-Sperrsichtmodell er eingeschaltet und es dadurch in ein Steuerorgan verwandelt. stand aus einem Platindraht von 0,2 mm Dicke etwa 2 mm oide. Der Draht war in den Kristall eingeschmolzen. Das Ver wir schon vor Jahren veröffentlicht¹⁾: Ein elektrisch geheizter ringt genau so in KBr-Kristalle ein, wie ein warmer Kupfern Eisklotz. Der Einkristall bleibt dabei erhalten. tigung dieses Dreielektrodenkristalles ist aus Fig. 6 ersichtlich. de der Anodenstrom i_A und der Gitterstrom i_G in Abhängigkeit vorspannung P_g . Diese Messungen wurden bei zwei verschiedenen

AnodenSpannungen ausgeführt, nämlich $P_A = 100$ Volt und $= 150$ Volt. Die Messungen sind in Fig. 7 und 8 graphisch dargestellt. Die Kurven gleichen durchaus den bekannten Kennlinien eines Dreielektrodenrohrs. Man beschreibt die Ergebnisse am besten mit den für Elektronenröhren entwickelten Begriffen.

Zur Steuerung ist eine Gittervorspannung von etwa + 10 Volt erforderlich. Die obere Kennlinie besitzt eine Steilheit

$$S = \left(\frac{\partial i_A}{\partial P_g} \right)_{P_A = 150 \text{ Volt}} = 0,028 \text{ Milliampere/Volt.}$$

Beide Kennlinien zusammen ergeben als Durchgriff

$$D = \left(\frac{\partial P_g}{\partial i_A} \right)_{i_A = 2,5 \cdot 10^{-4} \text{ Ampere}} = 7,3 \%$$

und den inneren Widerstand

$$R_i = \left(\frac{\partial P_g}{\partial i_A} \right)_{P_g = 10 \text{ Volt}} = 4,6 \cdot 10^5 \text{ Ohm,}$$

$S \cdot D \cdot R_i$ ergibt 0,95 statt 1.

Amplification:
20 - 100

Dieser als Beispiel gewählte Dreielektrodenkristall steuert einen Strom von 0,4 Milliampere mit einem Gitterstrom von 0,02 Milliampere. Man hat also hier eine 20fache Verstärkung. Das ist keineswegs eine obere Grenze. Wir haben auch Verstärkungen über 100fach hergestellt. Dazu braucht man höhere Werte von δ . Doch wird bei diesen Verstärkungen

¹⁾ Naturwissensch. 20, 932, 1932.

Steuerung von Elektronenströmen mit einem Dreielektrodenkristall usw. 407

die Trägheit selbst für einen Modellversuch zu groß, wenn man nicht den Elektrodenabstand verkleinern will.

Wir haben mit den Versuchen keinerlei technische Ziele verfolgt, uns interessierte nur die grundsätzliche Seite der Frage. Für den Steuervorgang sind die Einzelheiten der Elektronenbewegung ohne Belang. Die langsame

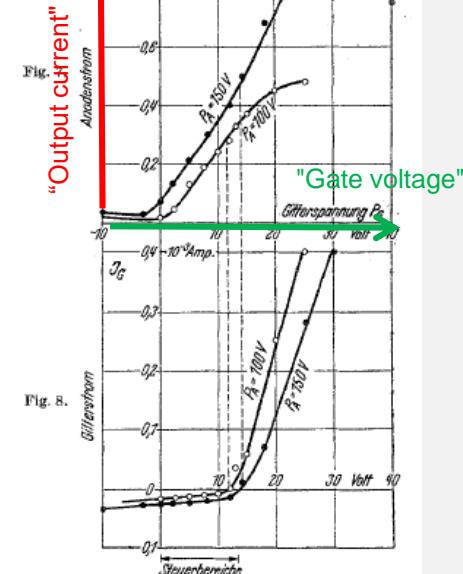
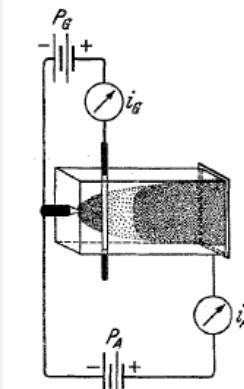


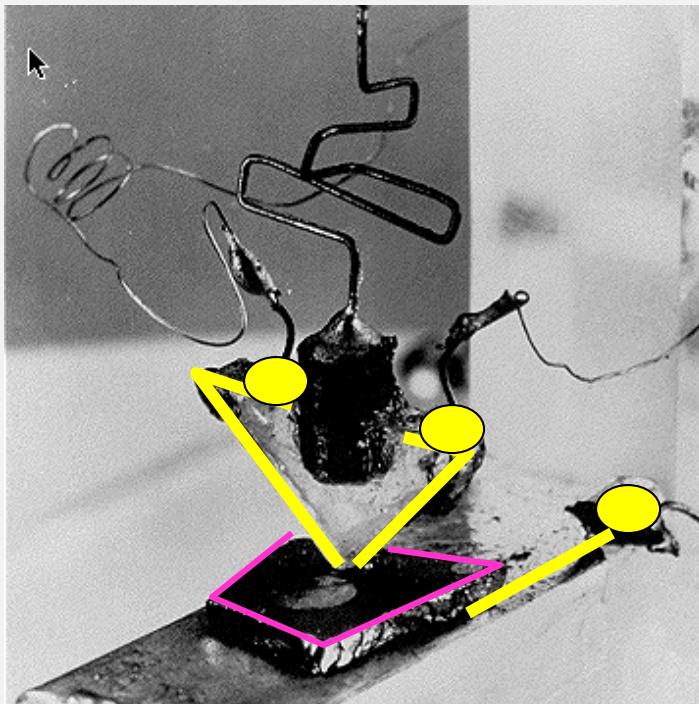
Fig. 6. Schema und Schaltung eines Dreielektrodenkristalles. Gezeichnet ist die durch Farbzentren sichtbar gemachte Elektronenverteilung während einer Abnahme des Anodenstromes. Die negative Aufladung des Gitters ist vergrößert worden, der Anodenstrom hat aber noch nicht seinen stationären Wert erreicht. (Kristallabmessungen $2 \times 5 \times 10$ mm).

Fig. 7 und 8. Die beiden Kennlinien eines Dreielektrodenkristalles mit 20facher Verstärkung. $T = 400^\circ \text{C}$.

Diffusion der Elektronen im Kristall wirkt nicht anders als ihre beschleunigte Bewegung im Hochvakuum. Wesentlich ist nur, daß die Elektronen den überwiegenden Anteil des Stromes tragen und (wenigstens überwiegend) nur der einen Elektrode entstammen.

Die Trägheit der Ströme in einem langen Dreielektrodenkristall ist für Schauversuche ein Vorteil. Sie gibt die Möglichkeit, den zeitlichen Ablauf der Erscheinungen sowohl mit dem Strommesser wie mit dem Auge

In 1947, during their investigations for the replacement of tubes the researcher Brattain, Bardeen und Shockley discovered an amplification effect in the semiconductor material Germanium.



A small plastic triangle was wrapped with a gold foil and cut with a razor blade at the tip.

On top two contacts were soldered.

The whole construction was pressed on a piece of Germanium (the base), which was also contacted.

At both gold contacts the power supply was applied, without allowing any current to flow.

Not until a bias was applied to the Germanium, a current flowing between the gold contacts was observed, which was about 18 times larger than the applied current in the Germanium.

A tiny current amplifier was invented with variable transfer resistance (transfer resistor)

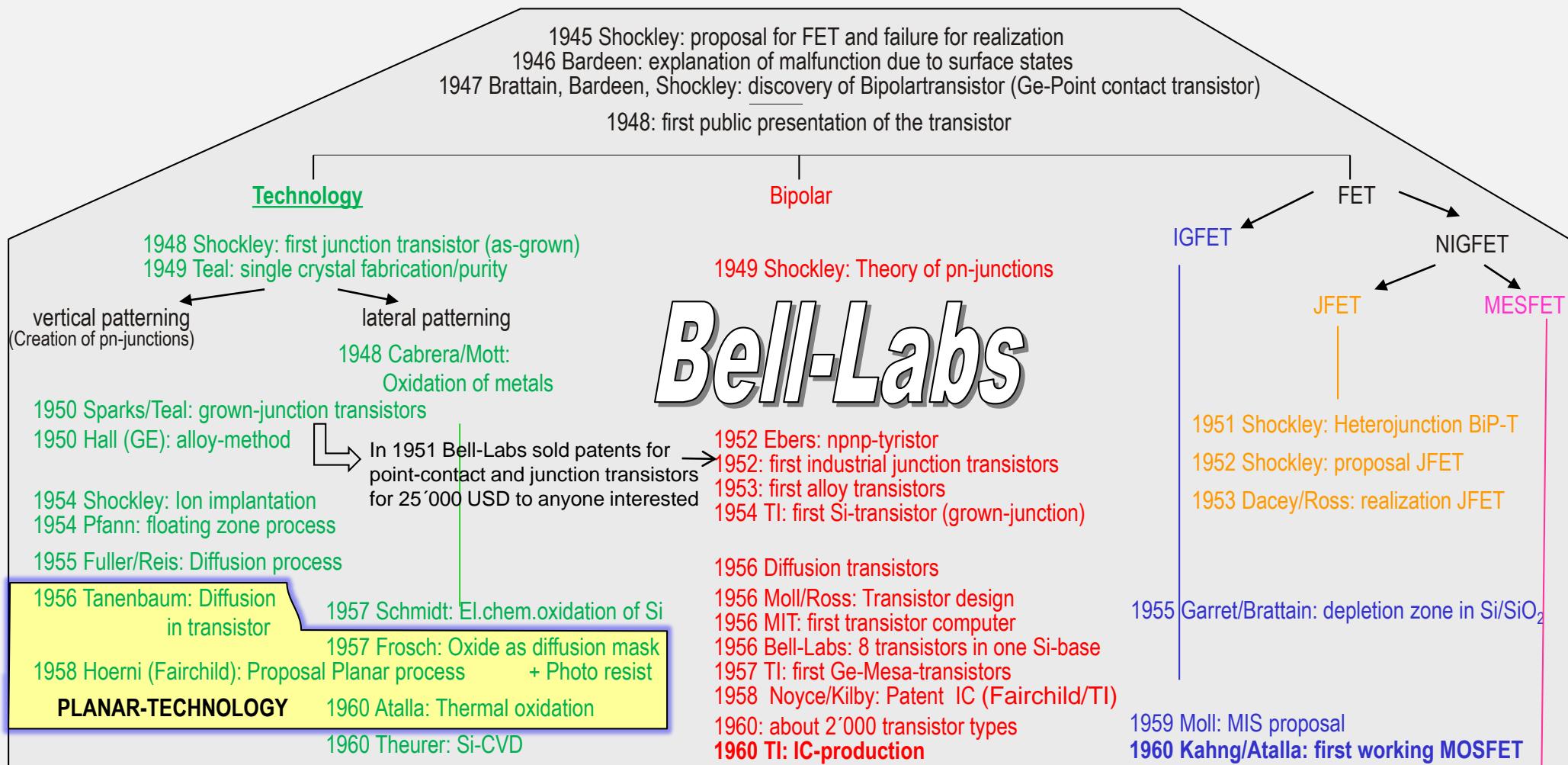
The Transistor was invented

Transistors consumed immediately 100x less current than tubes

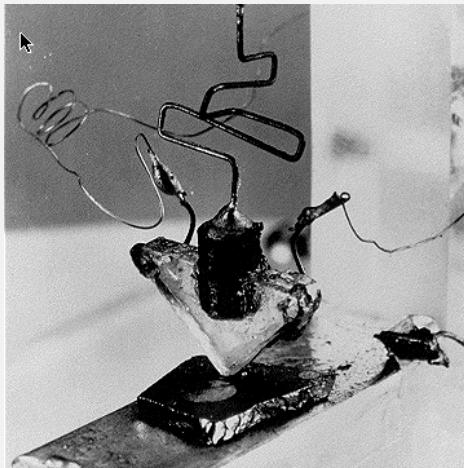
Transistors were immediately 100x smaller than tubes

Transistors were immediately 100x faster than tubes

Transistors were immediately 10x more expensive than tubes (Ge was more expensive than Gold)

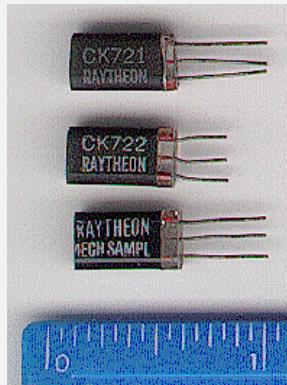


Nice details on transistor history see for example: <http://www.pbs.org/transistor/>

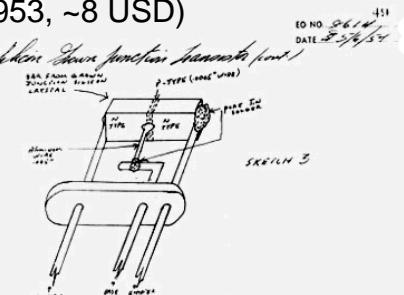
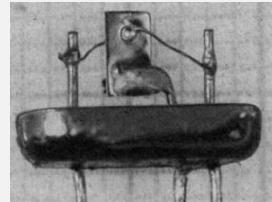


First industrial fabricated transistor
(Point-contact transistor, Raytheon 1952)

Laboratory design of the first transistor (Bell-Labs, 1948, amplification 18)



Large volume production of transistors
(Germanium, alloyed junctions)
(Raytheon 1953, ~8 USD)



First transistors in silicon
(grown-junction, TI, 1954, 60 USD)



First transistors in Germany
by SAF, Telefunken, Siemens
in 1953 (~8 USD)

1960: 2'000 various transistor types, market: ~ 8 Mill. \$US

1970: 20'000 transistor types, 15'000 IC types, market: ~ 14 Mill. \$US

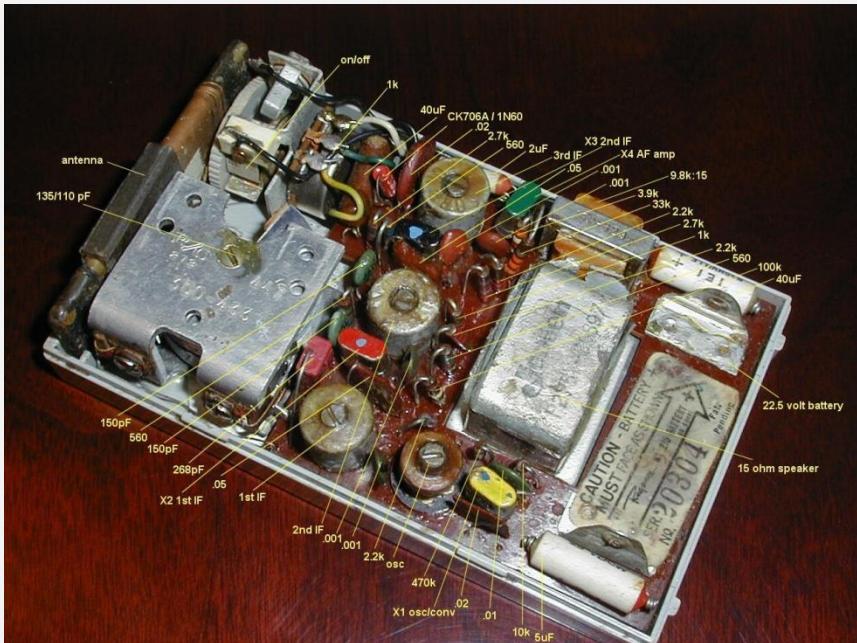
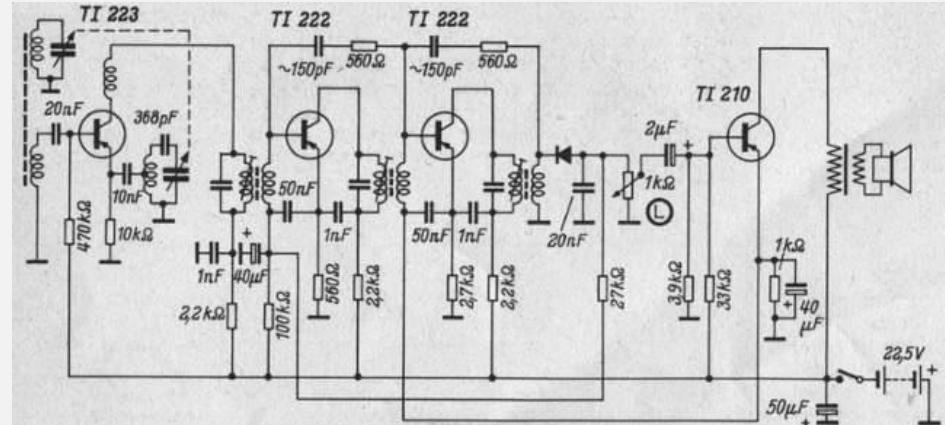
First Applications



First Transistorradio TR-1
Nov., 1954 by Regency & TI

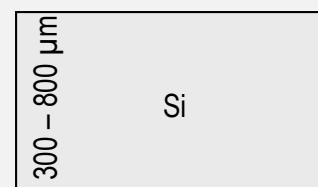
The demand was:
4 transistors, each 2.50 \$US
selling price TR-1: 49.95 \$US

for one year appr. 100'000 units were sold

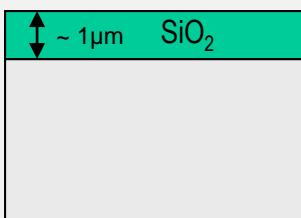


SONY, TR-55, 1955
(Japan's first transistor radio,
employing five transistors
developed in-house)





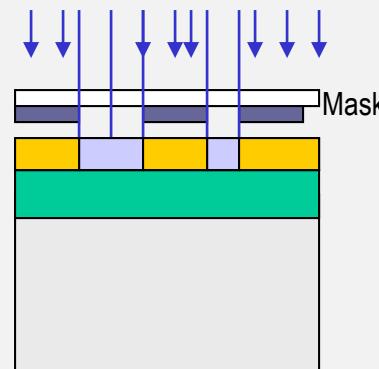
1. Defect-free, single-crystal semiconductor



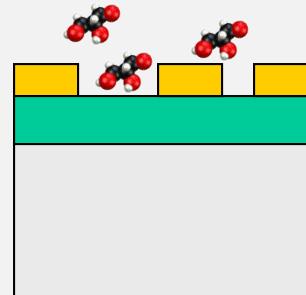
2. Creation of a mask layer, e.g. thermal oxide



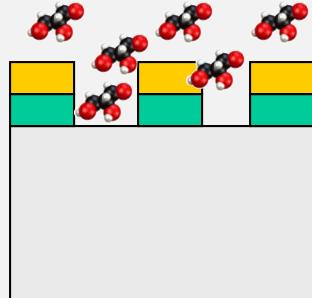
3. Deposition of photo resist



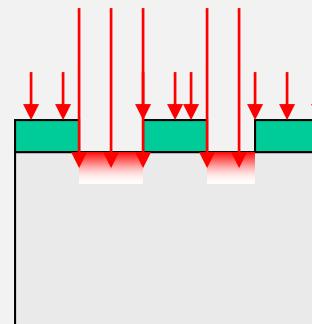
4. Exposure



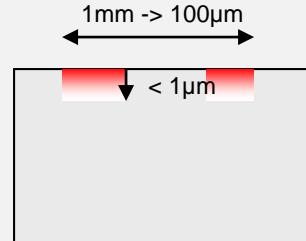
5. Development of resist



6. Patterning of mask layer



7. Process, e.g. doping



8. Removal of mask layer

Geometrical characteristics:

Structures are wide ($>100\mu\text{m}$) compared to height ($\sim 1\mu\text{m}$)
-> planar

Technological advantage:

All devices are fabricated simultaneously (parallel)
-> all devices exhibit same performance

Economical advantage:

- * All devices are fabricated simultaneously (parallel)
- * As smaller the devices are as more can be fabricated simultaneously (nearly constant processing costs)

Planartechnology paved the roads
for two major inventions:



- the „Integrated Circuit“
- the MOS-Transistor

1948 - 1952:

From the invention of the (bipolar) transistor in 1948 until commercializing the transistor in 1952 the theoretical fundamentals were developed within the Bell-Labs

1952 - 1960:

In 1952 the industry was given two bipolar transistor types:

the point-contact transistor and the grown-junction transistor

Until 1960 the industry developed further transistor types (alloyed-junction transistor, diffused transistor, silicon transistor, mesa transistor, ...) in about 2000 modifications.

Using transistors space- and power-saving and reliable electronic equipment could be developed.

The main user of transistors was the US military and space program.

since 1960:

Until 1960 the Bell-Labs developed the planar technology. Using the planar technology it was possible:

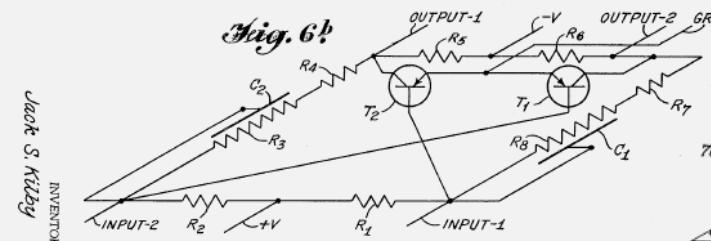
- mass fabrication of reproducible and reliable transistors
- the fabrication of MOSFETs (successful fabrication 1960)
- the fabrication of Integrated Circuits (invented 1959)

First Integrated Circuits

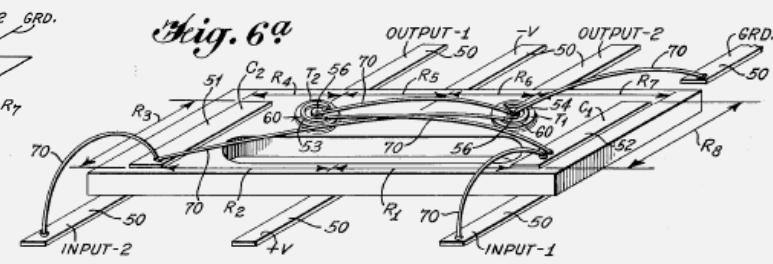


J. Kilby (TI): first IC 1958,
patent submission Feb.1959,
granted in 1964

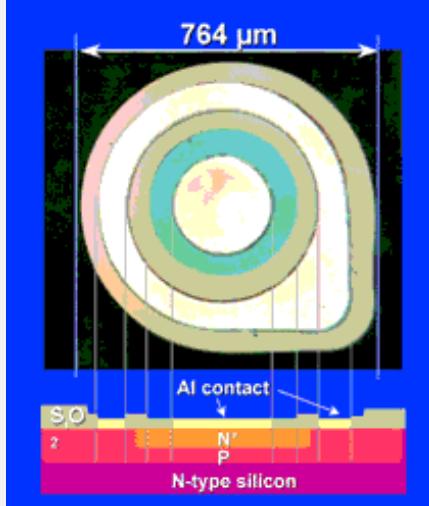
Nobel price 2000



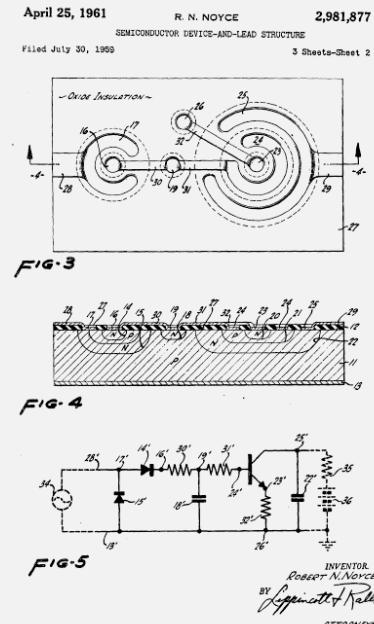
Realization of a multivibrator:
2 transistors + 8 resistors + 2 capacitors



IC with Ge-Mesa transistors,
but **interconnections with external wiring**



Silicon planar transistor
Hoerni (Fairchild) 1959



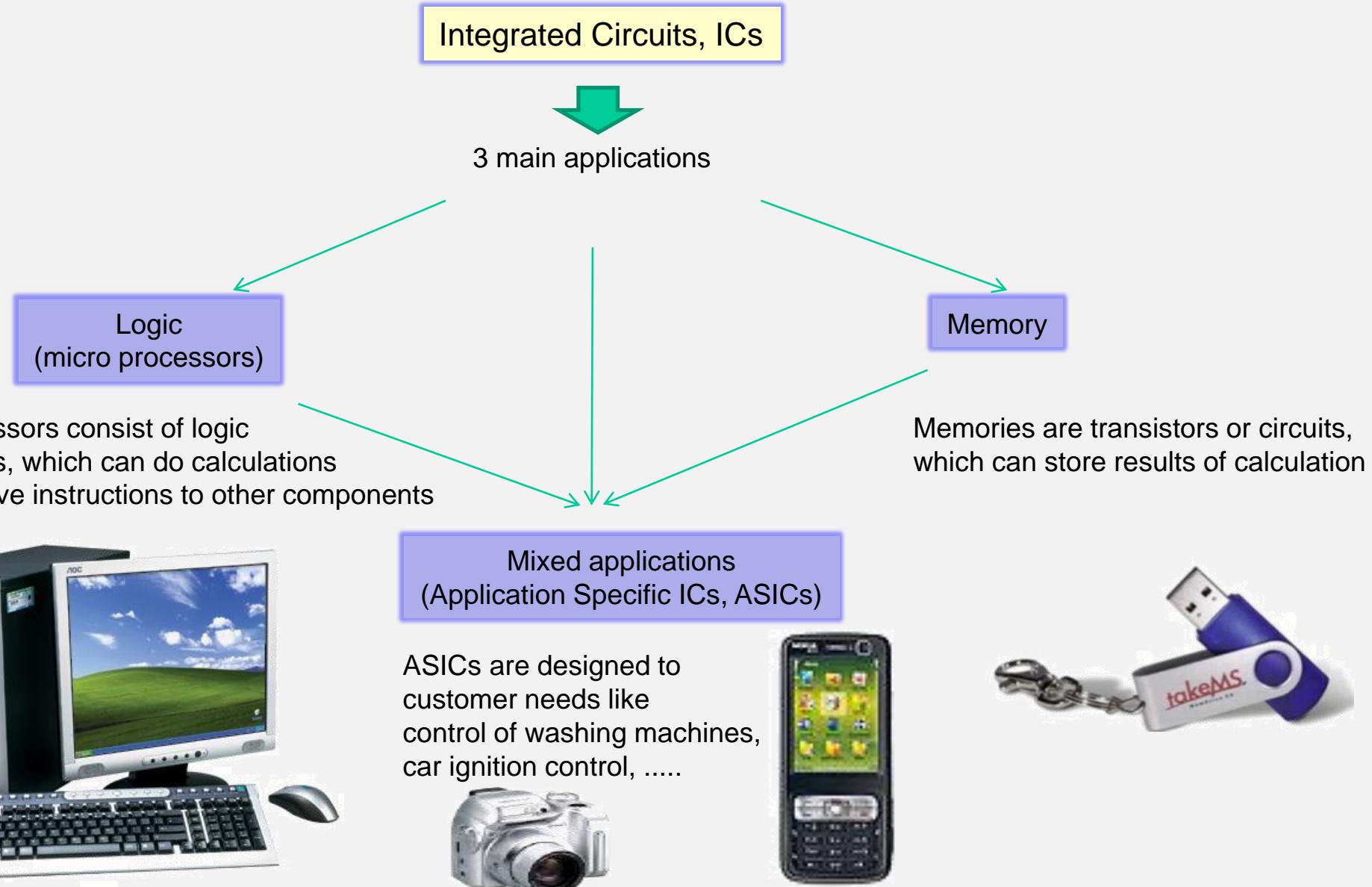
Noyce (Fairchild): patent submission July 1959,
granted in 1961
IC with **Si planar transistors including wiring**

1961: start of IC mass production (~4'000/week)
(logic: flip-flops, counters, NAND, XOR, ..)
up to 1968: US space program buys 95% of IC production
-> price decrease from 1'000 \$US / IC -> 30 \$US

1970: First DRAM (1kb, INTEL)
1971: First microprocessor 4004 (2kb, INTEL)

...

2006-2011: daily production ~ 10^8 ICs



1. Introduction

1.1. Historical Review

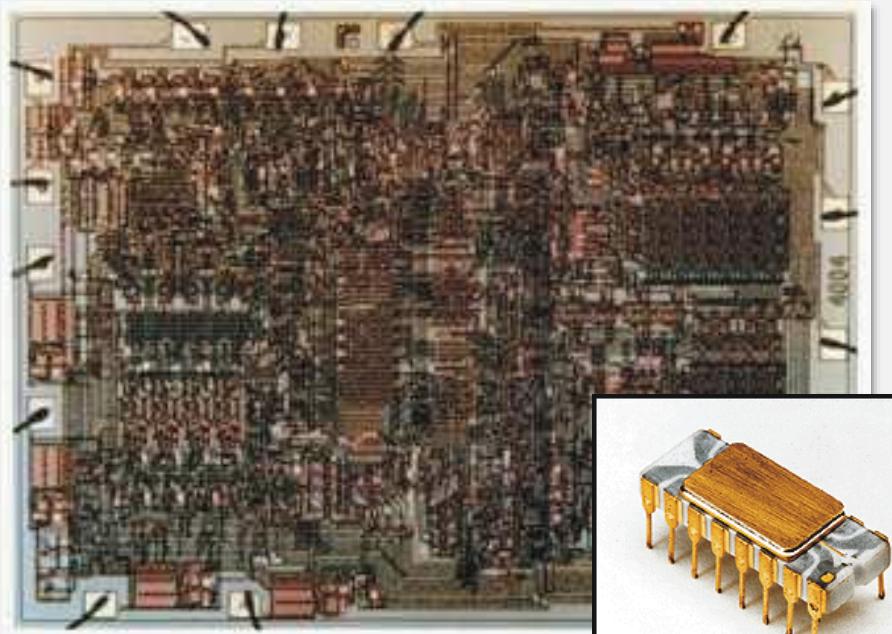
Logic and First Micro-Processor

1968: Noyce and Moore left Fairchild and founded INTEL (see e.g. <http://de.wikipedia.org/wiki/Intel>)

1969: Nippon Calculating Machines (Japan) asked INTEL for some RAM ICs for making their industrial calculators (Busicom) more flexible. Hoff (INTEL) was assigned for the project, but came up with a even more flexible solution designing one central logic chip (NMC-ARU) with 3 additional ICs (NMC-RAM, NMC-ROM, NMC-SHR).

1970: Intel offered Busicom a lower price for the chips in return for securing the rights to the microprocessor design and the rights to market it for non-calculator applications, allowing the Intel 4004 microprocessor to be advertised in the November 15, 1971 issue of Electronic News.

It's then that the Intel 4004 became the first general-purpose microprocessor on the market.



1971: Model 4004, 2'500 Transistors, 8MHz

The Intel 4004, it was supposed to be the brains of a calculator. Instead, it turned into a general-purpose micro-processor as powerful as ENIAC.



courtesy of Ruud Dingemans

1960: SSI <100 devices

1961: TTL (PacificMicrotel, 25µm feature size)

1962: ECL (Motorola)

1966: MSI < 1000 devices

1967: First MOS-logic ICs

1969: LSI < 10'000 devices, CMOS

1971: First Microprocessor INTEL 4004, 2'500 devices

....

2008: > 2 Bill. devices in logic circuits
(INTEL, CPU Tukwila)

2012: INTEL, IvyBridge, 22 nm technology,
1.4 Bill. devices, chip size 160 mm²

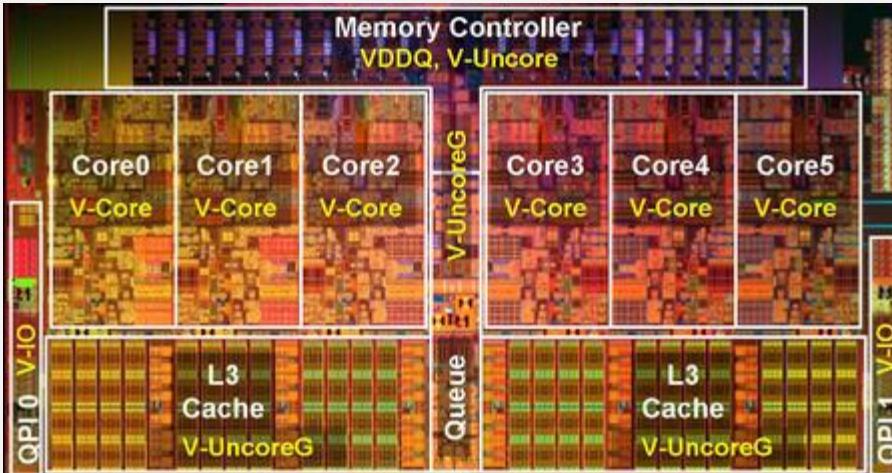
2017: INTEL, Coffee Lake, 14 nm technology,
~3 Bill. devices, chip size 174 mm²

2010 INTEL:

All processors with 32 nm technology are called "Westmere" with variations: Arrandale, Clarkdale, Gulftown
New architecture in this new technology (= tick) are called "Sandy-bridge" (=tock)

2016 INTEL:

The two-phase "tick-tock" development model is being replaced with a three-phase model: Process(technology), Architecture, Optimization



Feb 2010:
Gulftown
6 cores,
chip size: 240 mm²
transistors: $1.17 \cdot 10^9$

Feb 2012:
Ivy-Bridge
4 cores,
chip size: 160 mm²
transistors: $1.4 \cdot 10^9$

Aug 2015:
Skylake
10 cores,
chip size: 122 mm²
transistors: $1.7 \cdot 10^9$



Computing Power:

2012: "Super-MUC" LRZ München
~3 Petaflops, assembled with ~19'000 CPUs (INTEL XEON, 8-10 core)
main memory 340 TB, hard disc 14 PB (= $14 \cdot 10^{15}$),
weight: 100 t, energy consumption: 3 MW, costs: 83 Mill. €

Fastest supercomputer 2020:

2018: "Summit or OLCF-4", IBM, Oak Ridge National Laboratory, USA
~200 Petaflops, main memory > 10 PB, hard disc 250 PB (= $250 \cdot 10^{15}$),
energy consumption: 13 MW

Cannon Lake stumbles into the market: The IdeaPad 330-15ICN is the first laptop with a 10-nm-CPU

Intel Cannon Lake once was a very highly anticipated CPU-release. After massive delays, the first Intel CPU generation that is manufactured in the advanced 10 nanometer process has finally arrived – and it enters the market silently, in a low-budget device.

by Benjamin Herzig, 2018/05/13

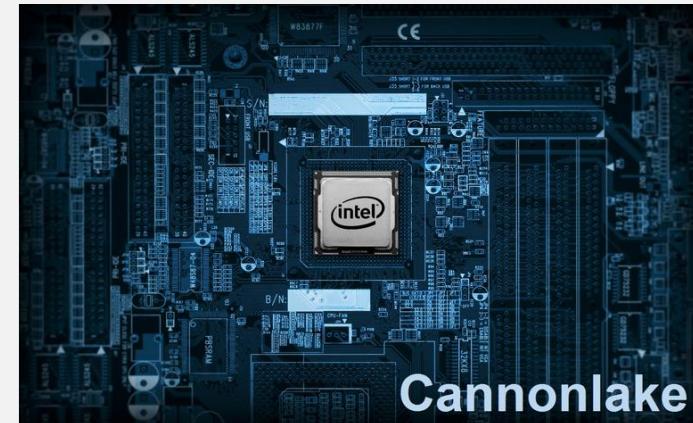
Cannon Lake Ultrabook Laptop Coffee Lake AMD

The first 10 nm laptop has been announced: When Lenovo introduced a couple of new IdeaPad laptops [a few days ago](#), this release also included the IdeaPad 330-15ICN – "ICN" stands for "Intel Cannon Lake". For those not in the know, Cannon Lake was once upon a time announced as the next big step in Intels release schedule, the first CPU generation to be manufactured in the 10 nm process. Originally, its release was expected in late 2016, almost two years ago.

<https://www.notebookcheck.net/Cannon-Lake-stumbles-into-the-market-The-IdeaPad-330-15ICN-is-the-first-laptop-with-a-10-nm-CPU.303330.0.html>



Cannon Lake stumbles into the market: The IdeaPad 330-15ICN is the first laptop with a 10-nm-CPU



<https://www.tweaktown.com/news/59209/intel-delays-10nm-cpu-tech-third-time-late-2018/index.html>

Because of yield problems, mass production of the new Cannon Lake processors started in 2019

It is expected that Cannon Lake is a test drive for 10 nm production, will be replaced by Ice Lake in 2020

1. Introduction

1.1. Historical Review

A View on Actual CPU Performance

TUM Asia

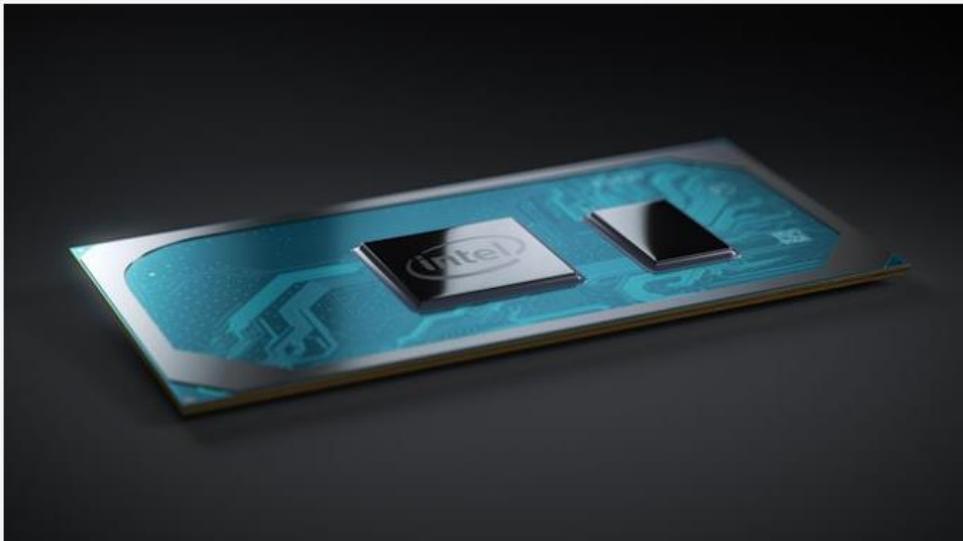
Intel Begins Commercial Shipments of 10nm Ice Lake CPUs to OEMs

by Anton Shilov on July 26, 2019 5:30 PM EST

Posted in CPUs Intel 10nm Ice Lake

115
Comments

+ Add A Comment



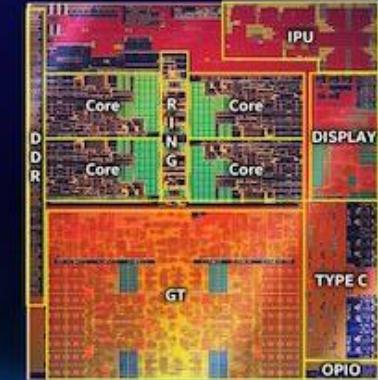
Intel has begun shipments of its 10th generation Core "Ice Lake" processors as of the second quarter, according to the company in an earnings call this week. Made using Intel's 10nm process technology, these laptop CPUs were qualified by OEMs earlier in 2019 and are on track to reach the market inside mobile PCs by the holiday season.

<https://www.anandtech.com/show/14679/intel-begins-commercial-shipments-of-10-nm-ice-lake-cpus-to-oems>

Mass production of Ice Lake started in late 2019

7nm node is planned for mass production in 2021

10NM ICE LAKE CLIENT
Shipping in June



INTEL, 2019

INVESTOR MEETING

OUR GAME PLAN... INVESTING IN PROCESS LEADERSHIP

EXTEND 14NM

Build Capacity
to Support
Customer Growth

vs. TSMC 10NM

RAMP 10NM

Client Systems on
Shelf for 2019
Holiday Season
Server in 1H'20

vs. TSMC 7NM

ACCELERATE TO 7NM

Production and
Launch in 2021

vs. TSMC 5NM

WORLD CLASS PACKAGING TECHNOLOGY COMPLEMENTS PROCESS LEADERSHIP

intel INVESTOR MEETING

der Bundeswehr

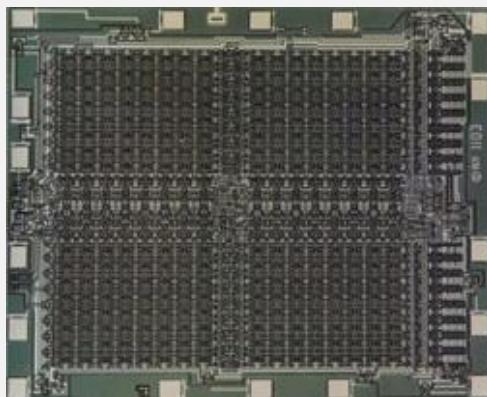
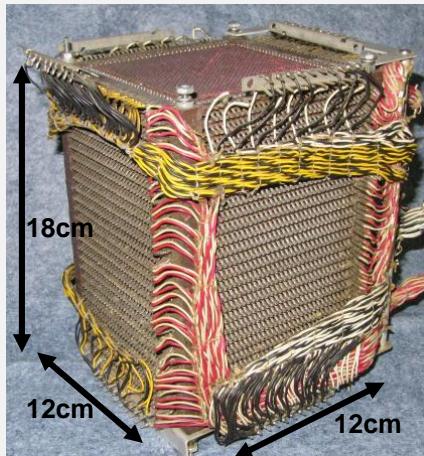
1. Introduction

1.1. Historical Review

Memory

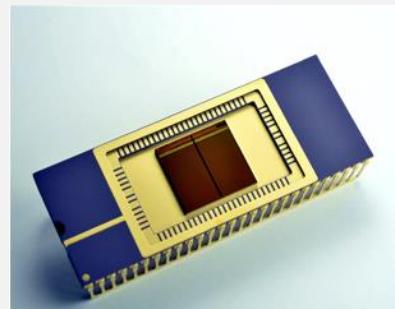


Magnetic core memory (1954 -1970)
- up to 1 kB



1970: The INTEL i1103 (**1kB DRAM**) was manufactured on a 6 mask silicon gate PMOS process with 8 μ m minimum features. The resulting product had a 2,400 μ m² memory cell size, a die size just under 10mm² and sold for around \$21. It was the best selling memory chip in 1972.

www.icknowledgecom/trends/8086_8088B1.jpg



2010: SAMSUNG
Producing Industry's First Higher-performing
20nm-class NAND Flash Memory, 64 Gb
Seoul, Korea on Apr 19, 2010

www.koreaittimes.com

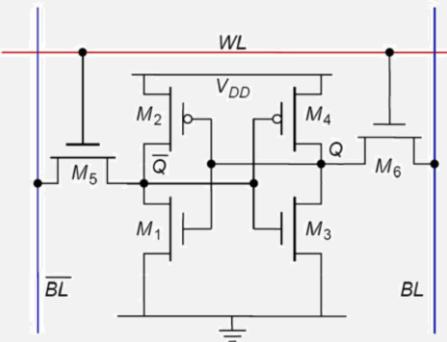
2018: SAMSUNG
Producing ~20nm-class 3D-Vertical NAND Flash Memory,
=> 512 Gb on one chip !
also: Toshiba, Micron similar products

Have a look at: <http://www.theinquirer.net/inquirer/feature/2286446/micron-bets-on-3d-nand-flash-for-the-future-of-storage>

SRAM

(Static Random Access Memory)

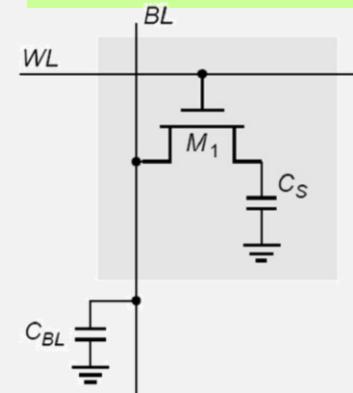
6 transistor cell, 1 bit



DRAM

(Dynamic Random Access Memory)

1 transistor cell, 1 bit
(+ 1 capacitor)



-> disadvantage:

- * volatile, when power switched off
- * big footprint -> expensive

-> advantage: very fast (few nsec)



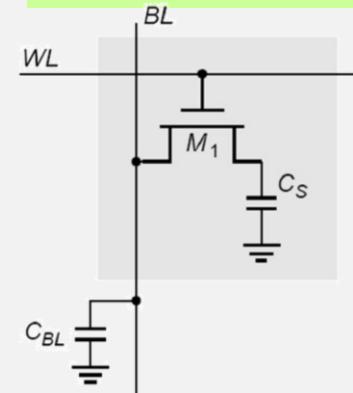
used as a small cache memory on processor chips

no power supply -> storage lost

DRAM

(Dynamic Random Access Memory)

1 transistor cell, 1 bit
(+ 1 capacitor)



-> advantage:

small footprint -> cheap
fast (~10 nsec)

-> disadvantage:

* volatile, even with power on



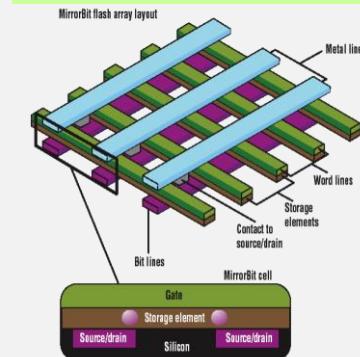
used as main memory for processors in duty

no power supply -> storage lost

Flash (EEPROM)

(Electrically Erasable Programmable Read Only Memory)

1 transistor cell, multi bits



-> advantage:

- * non-volatile
- * very small footprint -> cheap

-> disadvantage:
very slow (10 - 100 μ sec)



used as mass storage

no power supply
-> storage remains for 10 years

Goal of
Traditional Semiconductor Companies
(since starting ~1950)

imagine what the market needs,
create it, manufacture it
and then sell it on the open market
to multiple customers.



* few applications
-> calculators (need logic and memory ICs)
=> General Purpose Market

Invention of new electronic applications
-> TV, camera, printer, washing machine, ...

market needs more "Systems"
-> each product, each company needs
own, application-specific ICs
-> but tiny volume (say 100'000 chips/year)
-> extremely high fabrication costs

Goal of
New System Companies
(starting since ~1980)

-> try to find a IC-design,
which is highly standard, but can be customized



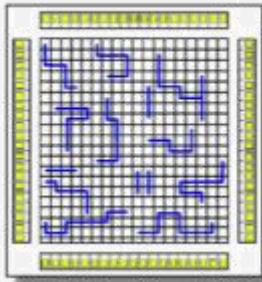
Gate-Arrays

-> many transistors not connected
-> function created by last metal interconnects
-> cost-effective up to ~1'000 transistors

Standard-Cells

-> highly integrated functional blocks
-> cost-effective up to millions of transistors

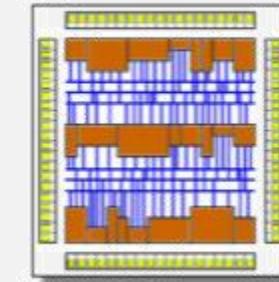
Gate Arrays



Gate arrays consisted of regular arrays of unconnected transistors. The chips were complete except for metal layers needed to connect the transistors. The gate array vendor would mass-produce these chips. Software provided by the gate array vendor, called a place-and-route tool, would map the logic gates of the customer's design – NANDs, NORs, flip-flops, and so forth – to specific transistors on the chip and determine a good way to put metal traces on the chip to connect the transistors so as to complete the design. Because all customers used the same basic platform, the costs associated with creating these base wafers were shared among all the customers. Thus, the initial NRE (non recurring expense) for a gate array was fairly cheap, though it could still be in the thousands or tens of thousands of dollars.

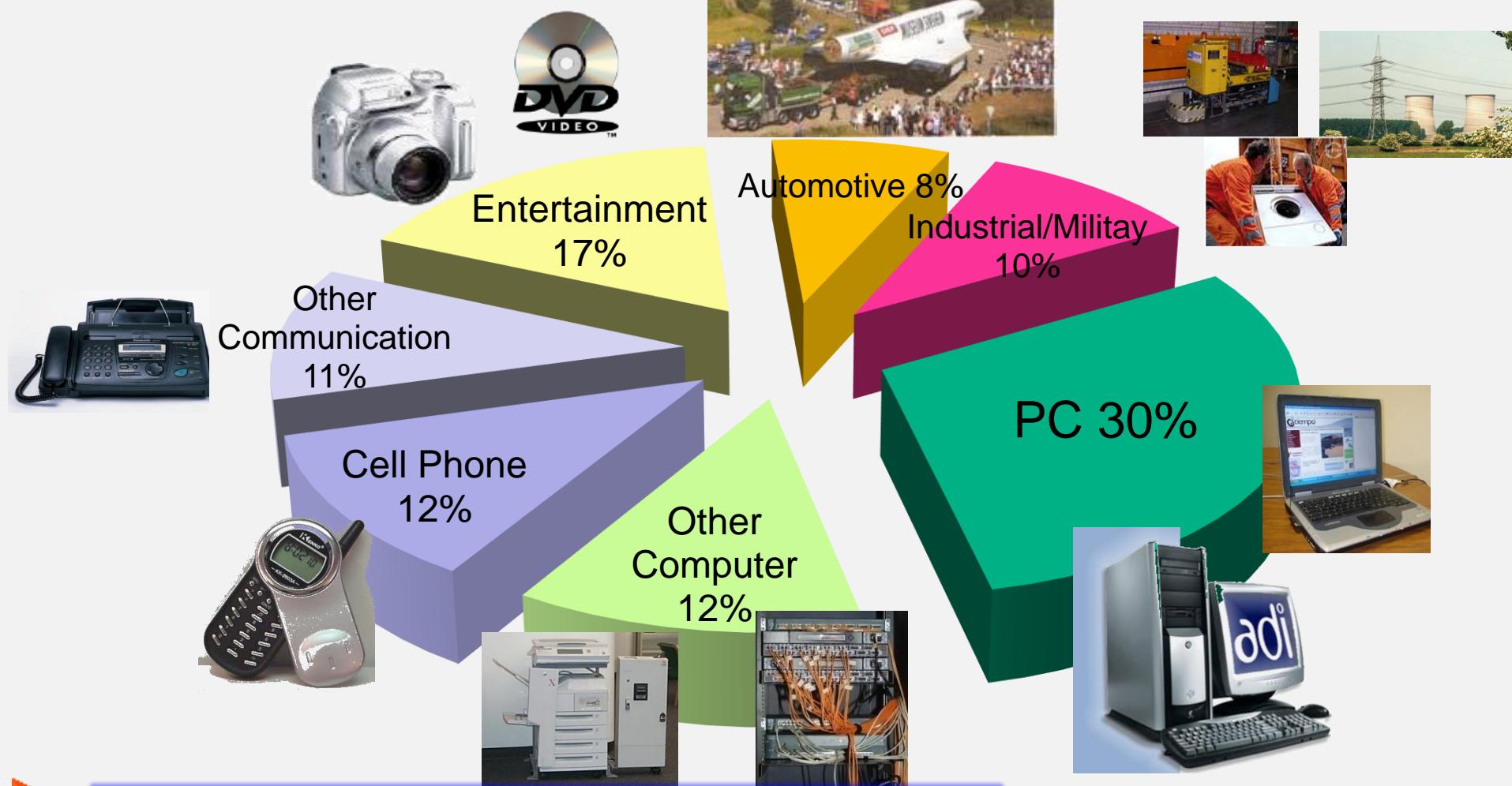
Gate array vendors offered families of gate arrays with a fixed number of transistors. A vendor might offer a 5,000 transistor chip and a 10,000 transistor chip. If your design required 5,001 transistors you were forced to put the design into the 10,000 transistor chip. Utilization of the transistors in a gate array was thus inefficient. Since it costs a fixed amount of money to process a single wafer, the more chip die that fit on that wafer – the cheaper the cost per chip. Conversely, an underutilized gate array meant that the customer paid for all transistors on the chip, even the unused ones. Thus the cost per piece for a gate array was relatively high.

Standard Cells



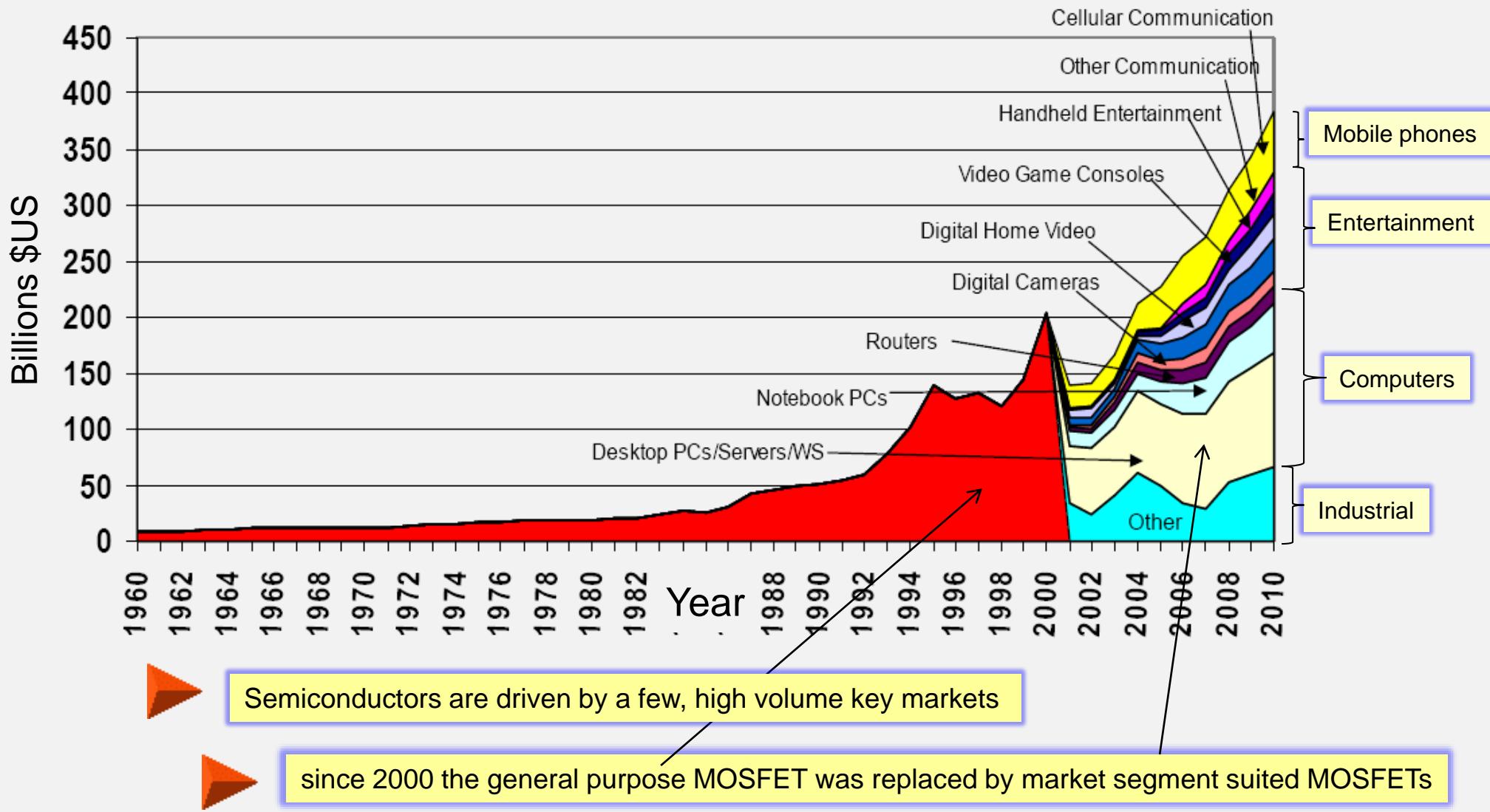
The other ASIC architecture was the standard cell, which started as a blank die. The standard cell vendor created a library of logic gates and higher-level logic functions called cells. The customer created a design from this library. The standard cell vendor provided place-and-route software that placed each cell onto the die and wired them all together to create the customer's design. Because only cells that the design specifically required were placed on the chip, the chips were smaller and thus cost less per piece than a gate array. The tradeoff was that the initial costs were not shared by other customers because each customer's standard cell chip was completely unique. Thus the standard cell chip had a large NRE.

The cost structure was such that gate arrays were used by companies requiring fewer volumes of chips while standard cells were used by companies requiring higher volumes of chips.



► There exist almost no electrical equipment without semiconductors

► The average semiconductor inventory is about 20% of the value of the equipment



1.1 Historical Review of Semiconductor Development

The Long Dawning of Hell-Made Semiconductors
The Impact of Bell Labs
From First Transistors to Today's ICs

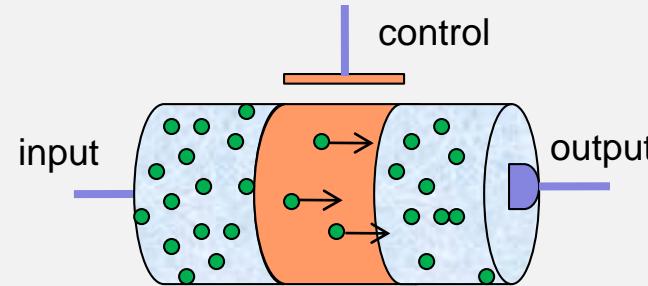
1.2 Basic Device Principles and Family Tree

Basic Device Principles
Family Tree

1.3 MOSFETs as Mainstream Devices

Needs of Information Society
Advantages of MOSFETs for Integration
Future Calculability of MOSFETs

The basic device structure consists of an input, output and control electrode, which manipulates the charge carrier transport:



The movement of charge carriers is based on various transport mechanisms:

1

Thermal Movement

Due to collisions with vibrating lattice atoms the electrons reach a thermal velocity of $\sim 10^7 \text{ cm/sec}$ in random direction

2

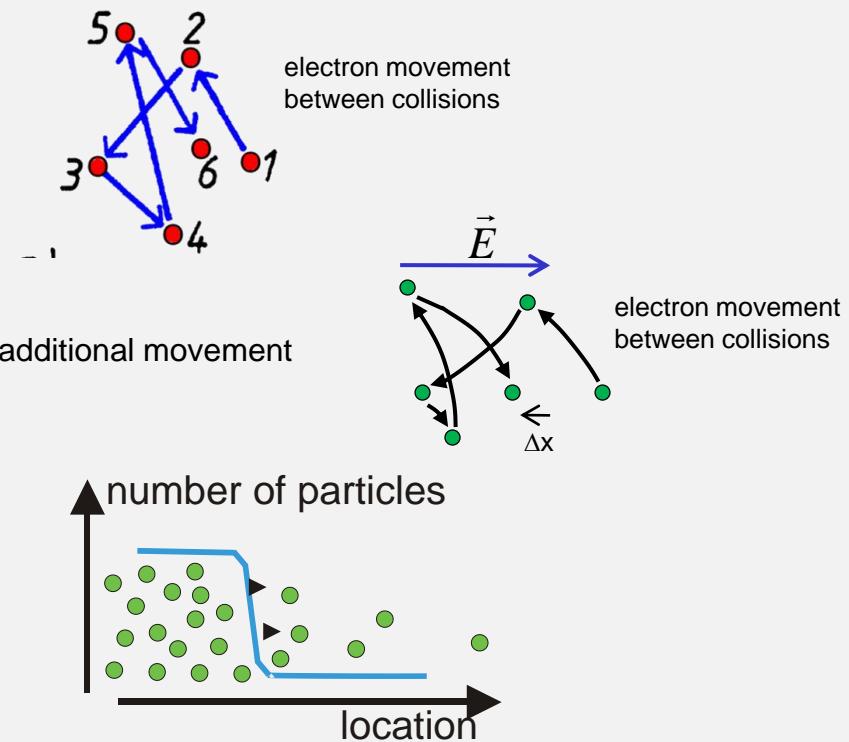
Drift-Movement

Within an electrical field the electrons receive an additional movement in direction of the electrical field

3

Diffusion

Due to their thermal movements charge carriers move from areas of high concentration into areas of low concentration



The basic control of an electronic device is by applying external electric fields
 -> in first order the drift movement of the charge carrier should be optimized

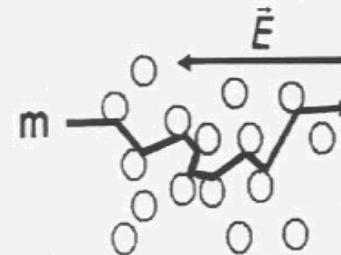
Drift-Movement

Due to permanent collisions with lattice atoms the charge carriers reach an average velocity v , which is proportional to the electrical field E

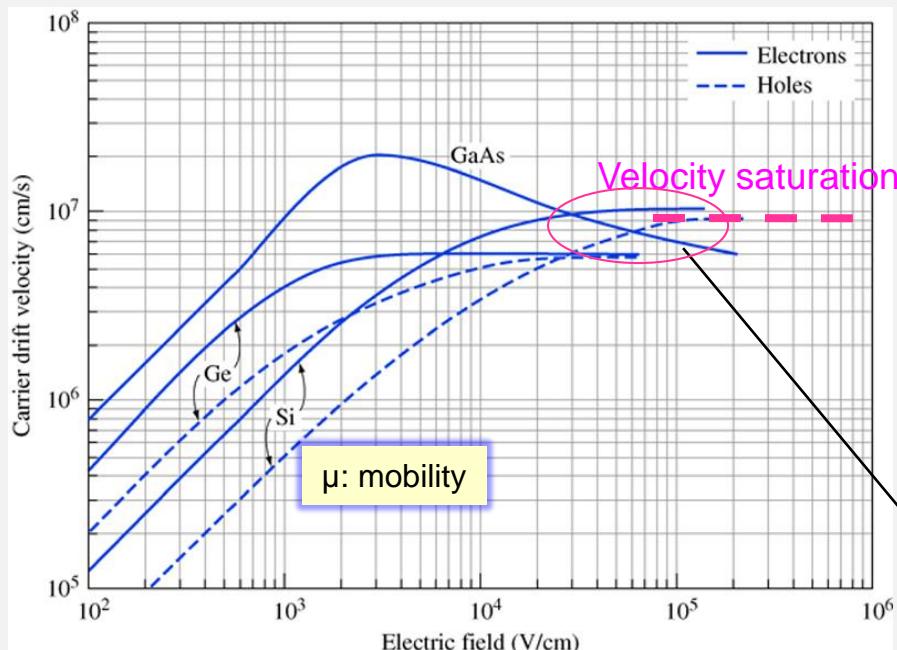
$$\vec{v}_D = -\frac{1}{2} \cdot \frac{e}{m^*} \cdot \tau \cdot \vec{E} = -\mu \cdot \vec{E}$$

details see chap.5

μ : mobility



e: electron charge,
 m^* = electron mass in solid,
 τ = time between collisions
 electron with Si-atoms



For same electrical field:

velocity of electrons is higher than for holes

velocity in silicon is lower than in Germanium than in GaAs

Due to the carrier mobilities are:

- Bipolar npn-transistors are faster than pnp
- FET n-channel devices faster than p-channel FETs

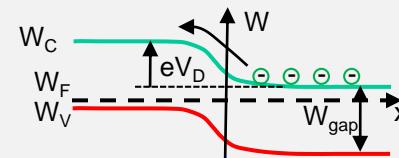
Maximum carrier speed: $v_{max} \sim 10^7$ cm/sec



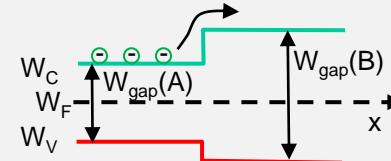
- Current control in semiconductor devices is realized as in any other control unit:
 - creation of barriers for charge carriers and manipulation of barrier height by external fields

Barrier Creation

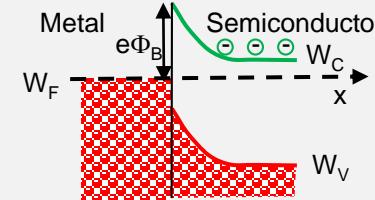
pn-junction



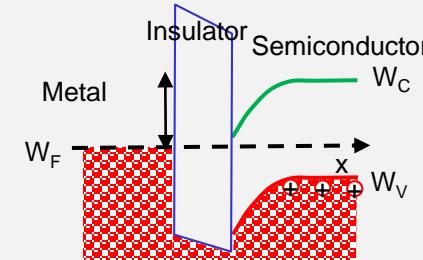
Hetero-junction



Schottky-Contact

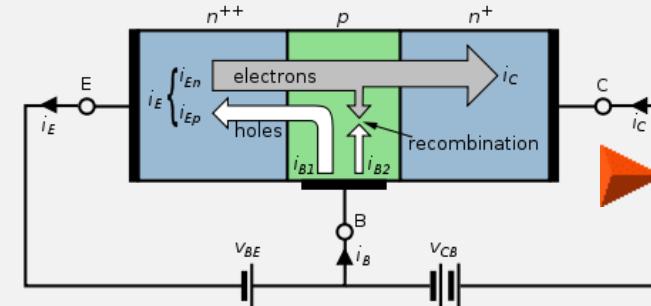


MIS-Sandwich



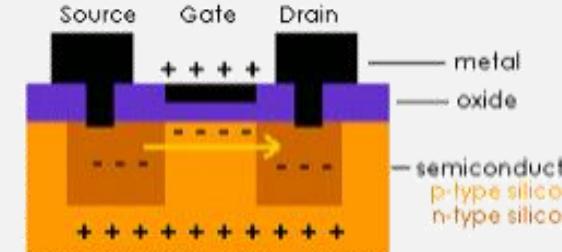
Control

Charge carrier injection

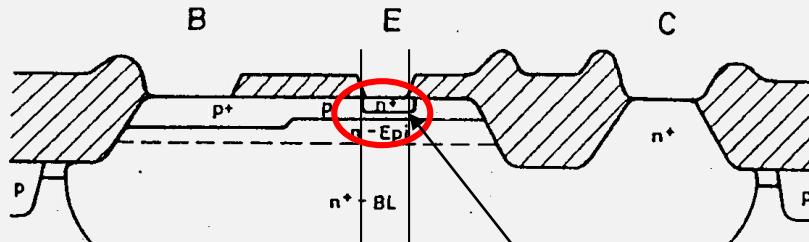


Bipolar Transistors

Field effect



Field-Effect Transistors



Minimum lateral feature size
(-> determines area consumption)

Minimum vertical feature size
(-> determines device performance)

Vertical feature size is small ($< 1\mu\text{m}$):

- > to ensure device bipolar working principle
(diffusion and life time of minority carriers in base region)
- > easy to make small due to fabrication technology (dopant diffusion)

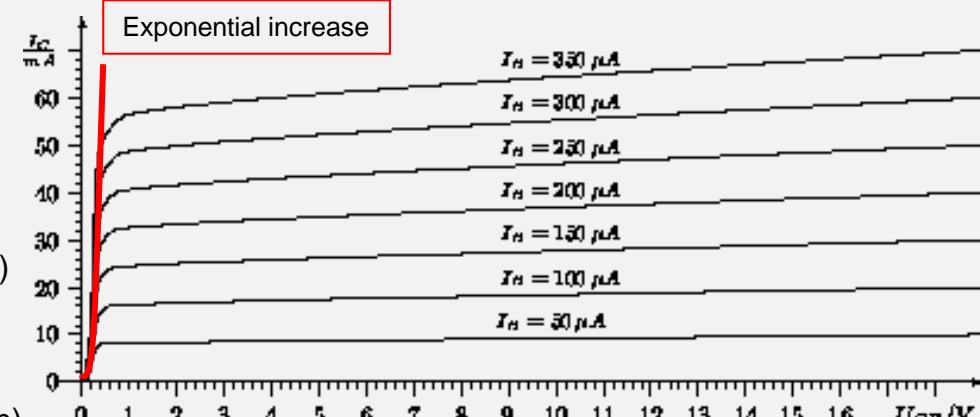
But:

- > not easy to make very small ($< 100\text{nm}$) due to dopant profiles
- > not allowed to make very small to ensure device performance (gain)

Classical Bipolar:

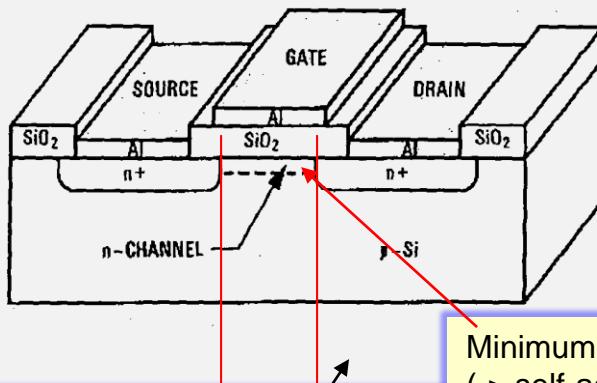
- # majority and minority carriers take part in current transport
- consumes power in on-state
- + exponential I-V characteristics -> fast switching

$$I_C = I_{reverse} \cdot \exp \left[\frac{V_{BE}}{(kT/q)} \right]$$



Classical bipolar devices were fast in switching since their historical invention, but could not be improved very much

(-> a little due to better process technology (e.g. doping techniques) and due to shrinking lateral dimensions (parasitic RC))
-> a lot due to new bipolar device concepts (Si -> Ge -> GaAs, BiP -> HBT, HEMT)



Minimum lateral feature size

(-> determines device performance + area consumption)

Minimum vertical feature size
(-> self-adjusting, not important)

MOS:

- # only majority carriers
- quadratic I-V characteristics -> slower switching
- + CMOS consumes no static power

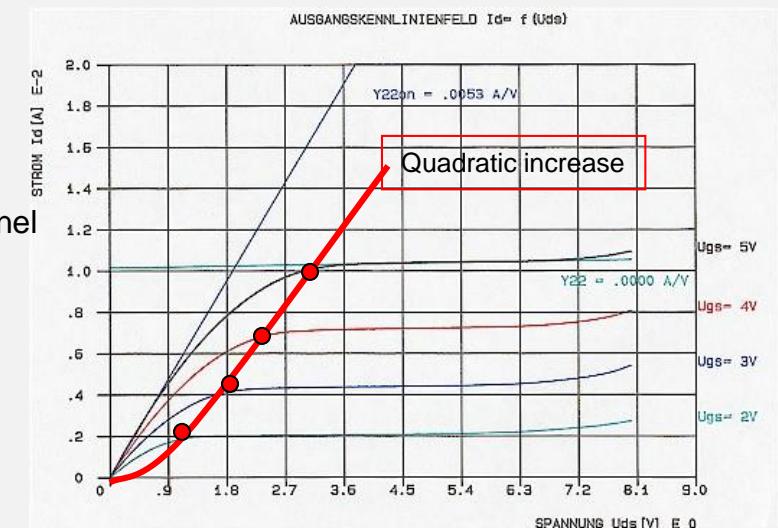
$$I_D = k \cdot (V_{GS} - V_{th})^2$$

Vertical feature size is very, very small (~ 5nm):

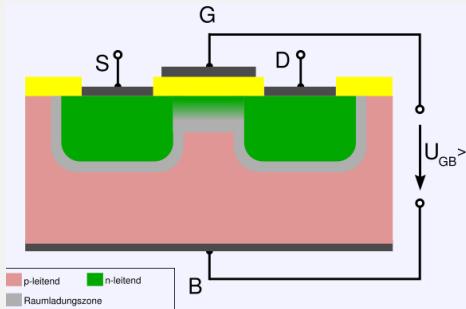
-> MOS device working principle induces an inversion layer as conducting channel
(thickness of channel fixed by physics)

But: Lateral feature size

- > not easy to shrink due to process technology (lithography)
- > very important influence to device performance



MOSFETs are getting faster with shrinking dimensions

**MOSFET:**

Advantage:

- + no gate current in any case -> the input resistance is essentially infinite
- + depletion and enhancement types can be fabricated very easily -> easy circuit design (only one voltage polarity is needed)

Disadvantage:

- The carrier mobility - also referred to as surface mobility - is less than half of the mobility of bulk material.

MOSFETs are best suited for ultimate high integration (scalability, low power, footprint,...)

JFET:

Advantage:

- + the voltage amplification is almost linear
- + the input resistance is very high (10^8 - 10^{10} Ω). This minimizes the interference with other sources -> low noise
- + the reversed biased gate-junction can take a considerable amount of radiation damage

Disadvantage:

- the voltage amplification is low (50 - 300) compared to bipolar transistors
- the fabrication of turned-off devices without gate voltage (enhancement type) is highly sophisticated (-> low yield), because the thickness and doping levels of all layers must be controlled very exactly (some ten nanometers)
- using enhancement type JFETs the usable voltage range is only in the best case from 0 - 0.7 V

JFETs are not frequently used, sometimes in low noise, very linear small signal amplifiers

MESFET:

Advantage:

- + higher mobility (x2) of the carriers in the channel as compared to the MOSFET. The higher mobility leads to a higher current, transconductance and transit frequency of the device.
- + Compared to the JFET the MESFET exhibits lower parasitic gate capacitances (low C_{GD} , C_{GS}). This results in higher frequencies.

Disadvantage:

- The Schottky metal gate limits the forward bias voltage on the gate to the turn-on voltage of the Schottky diode. This turn-on voltage is typically 0.7 V. The threshold voltage therefore must be lower than this turn-on voltage.
- No easy shrinking due to not self-aligned metal gate

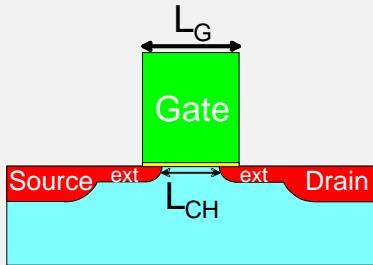
Further improvement in speed by using compound semiconductors (GaAs: 2x higher carrier mobility)

Due to the surface reduced mobility of charge carriers in MOSFET channels the MESFET and JFET with bulk channels are faster switching. Compared to the JFET the MESFET exhibits lower parasitic capacitances, which improves speed in addition.

To use the speed advantage of the MESFET usually MESFETs are fabricated in compound semiconductor material (GaAs, InP).



1) Switching time of a single MOSFET



Transit time:

$$\tau_{transit} = \frac{L_{ch}}{v} = \frac{L}{\mu \cdot E} \approx \frac{L^2}{\mu \cdot V_{DS}}$$

Example: 100nm channel length MOSFET

$$\tau_{transit} = \frac{L_{ch}}{v} = \frac{100\text{nm}}{10^7 \text{cm/sec}} = \frac{10^{-5} \text{cm}}{10^7 \text{cm/sec}} = 10^{-12} \text{sec} = 1\text{ps}$$

2) Switching time of a MOSFET in a circuit

Gate delay (Fanout: 1 or 3):

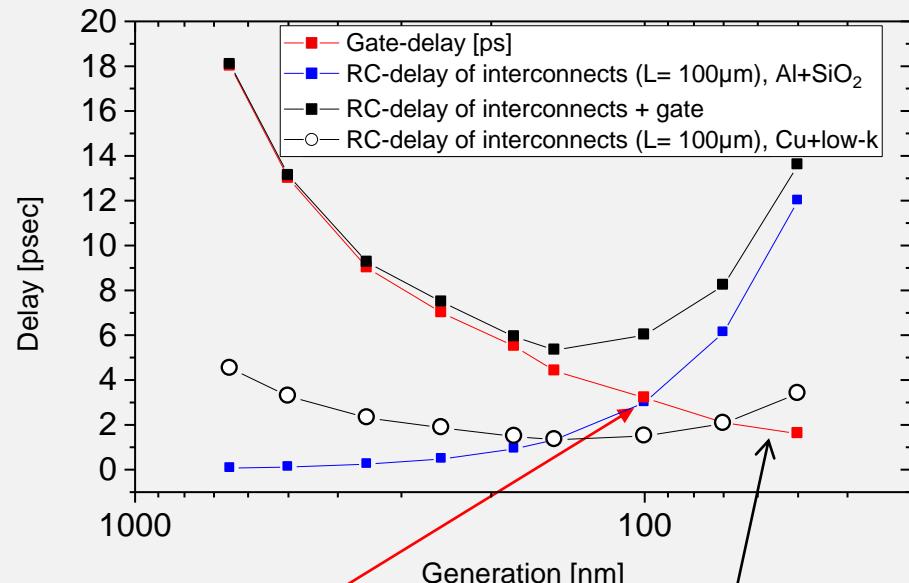
Gate capacitance: $C_g = \frac{\epsilon_0 \epsilon_{SiO2} \cdot w \cdot L}{t_{ox}}$ \rightarrow charge at gate: $Q_g = C_g \cdot V$

Gate-delay: $\tau_{gate} = \frac{C_g \cdot V}{I}$

$I=Q/t$ charging current of loading MOSFET
(~600 μ A/ μ m from roadmap, year 2000, see page 1-44)

Example: 100nm channel length MOSFET

$$\tau_{gate} = \frac{C_g \cdot V}{I} = \frac{\frac{\epsilon_0 \epsilon_{SiO2} \cdot w \cdot L}{t_{ox}} \cdot V}{I} = \frac{\frac{8.85 \cdot 10^{-14} F/cm}{1.7nm} \cdot 3.9 \cdot 100nm \cdot 100nm}{60\mu A} \cdot 1.0V = 3.4 \cdot 10^{-12} \text{ sec}$$



Faster switching due to shrinking dimensions

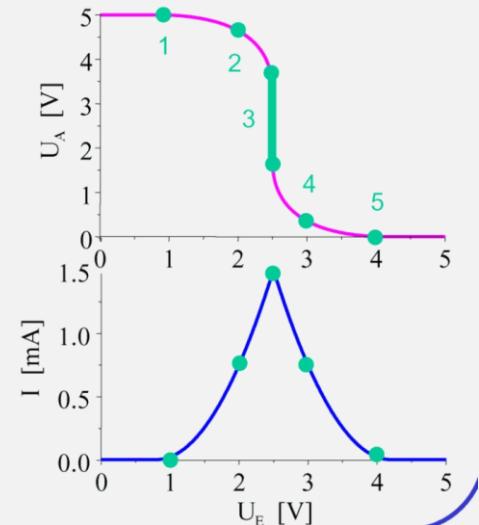
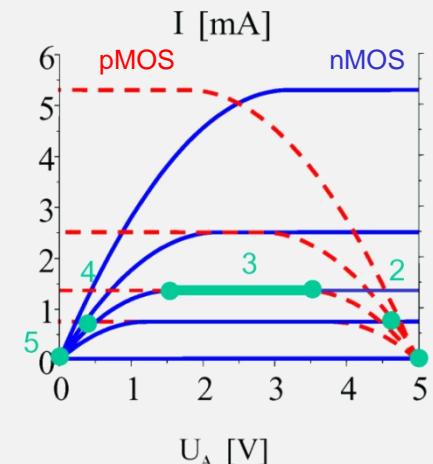
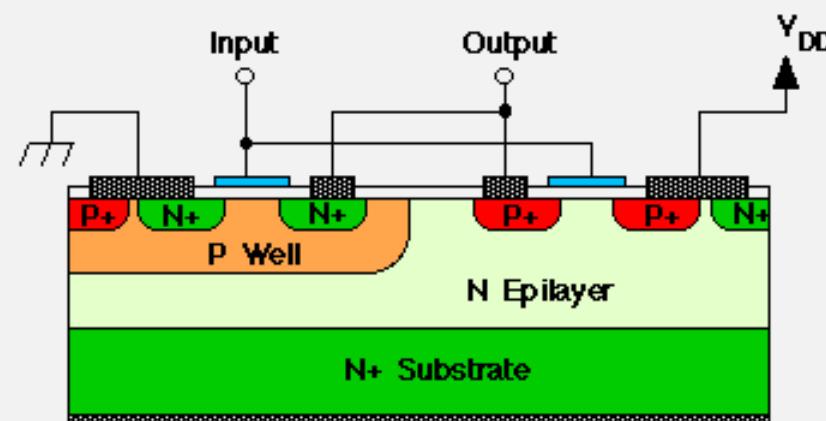
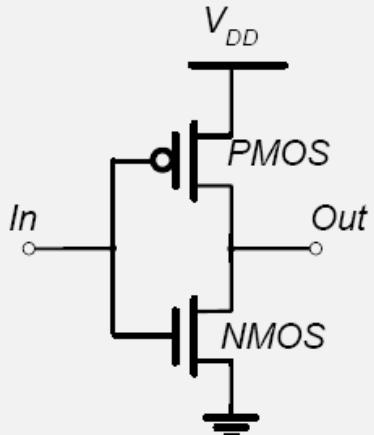


CMOS

lowest power consumption technology

best suited for ultimate integration

urgent needs for mobile information- / communications society



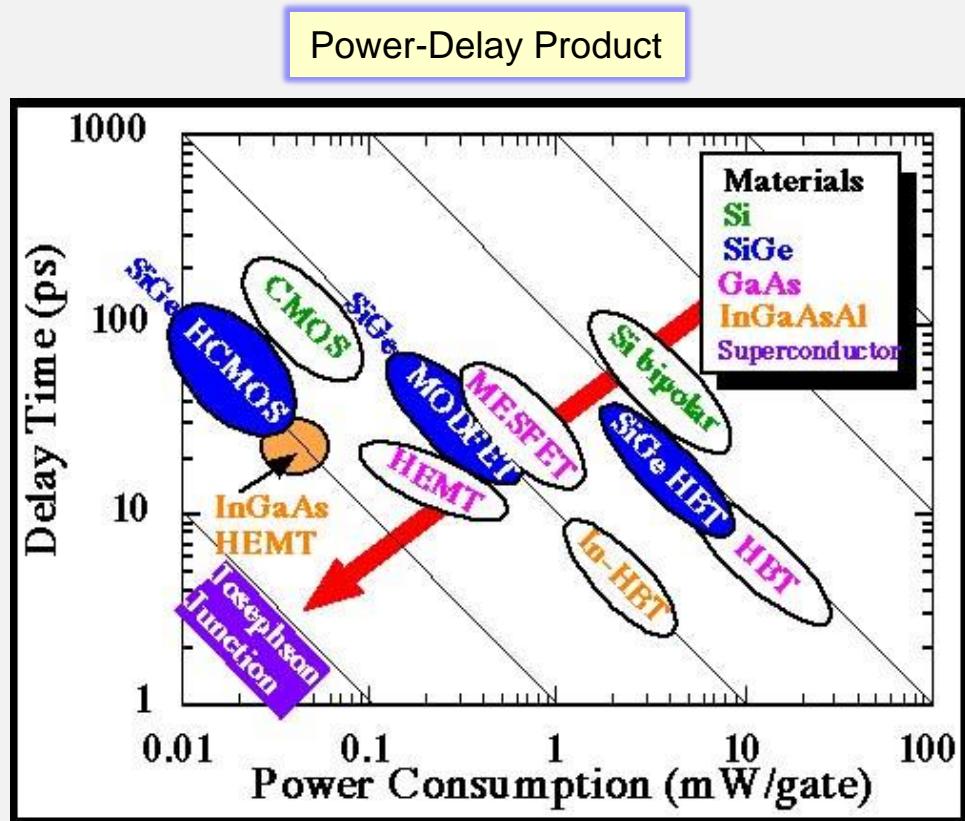
The inverter is one of two basic elements in any logical circuit

An inverter in CMOS-technique only consumes power during switching (no static power)



CMOS devices are best suited for high integration and mobile applications

Different advantages of bipolar Transistors and Fieldeffect-Transistors:

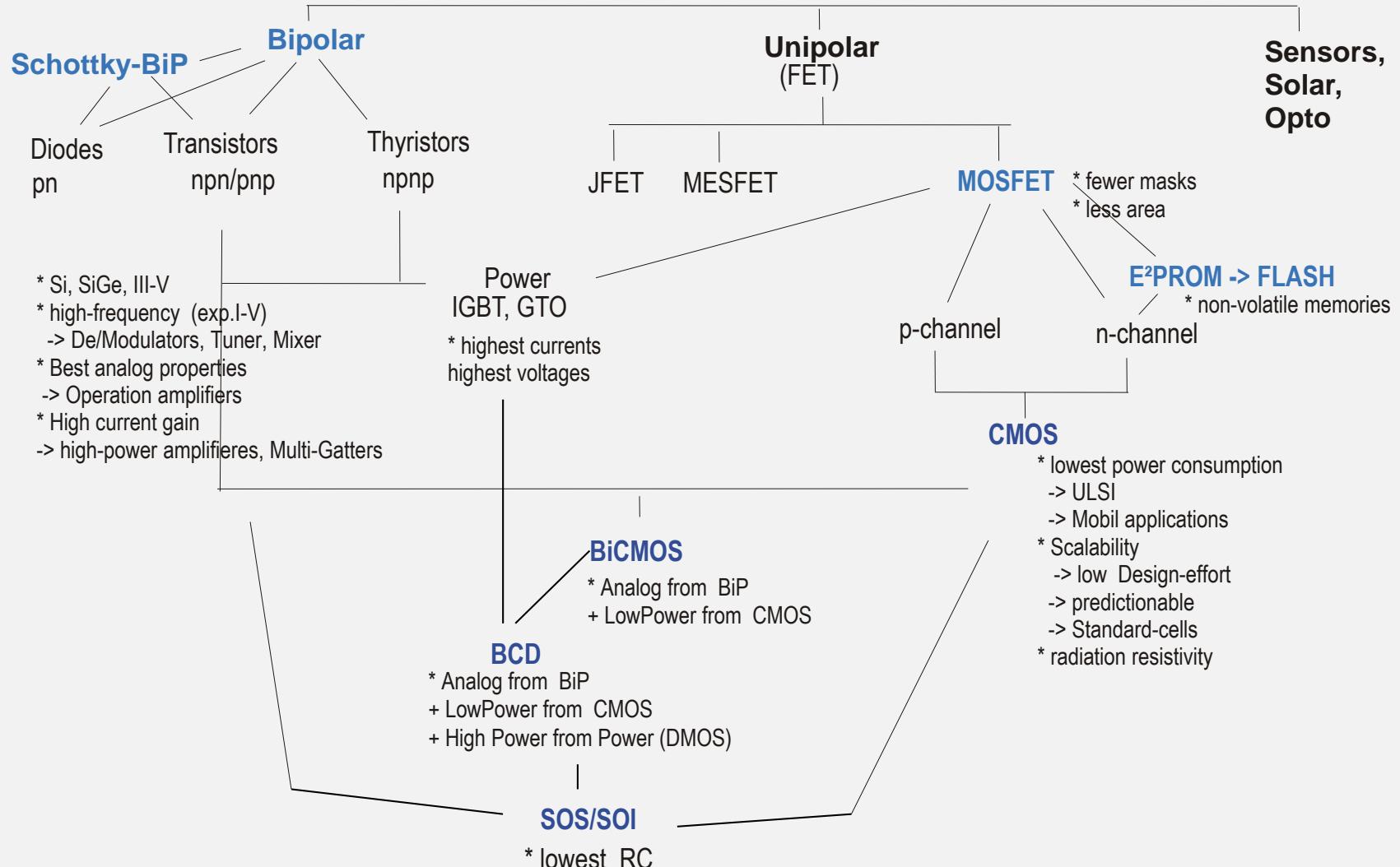


<http://www.sp.phy.cam.ac.uk/~dp109/Roadmap.html>

Collecting facts:

(only concerning speed and power consumption)

- All devices are getting faster as more power they consume (MOS, BiP,...)
- Classical bipolar devices may be a factor 10 faster than MOSFETs (using the same minimum lateral dimensions)
- Classical bipolar devices may consume a factor 10 more power than MOSFETs (using the same minimum lateral dimensions)
- Several (high sophisticated) new bipolar concepts are under development

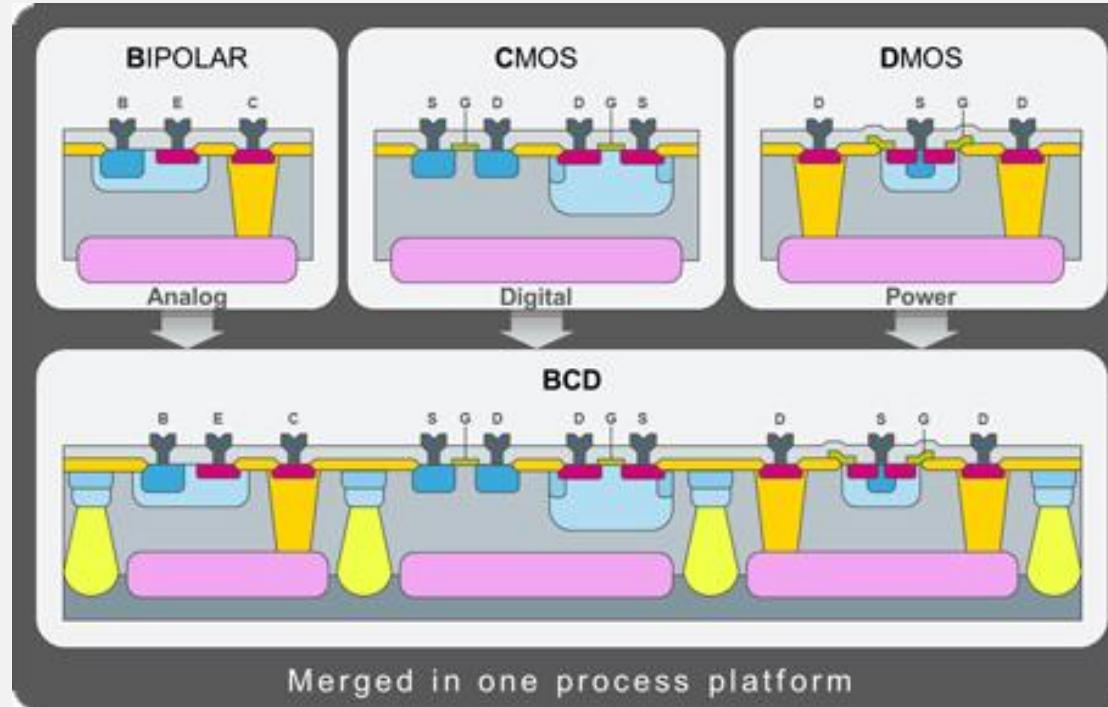


Depending on application focus various mainstream device families were developed

Example, how various technologies may be introduced in one chip.



This is a challenge for technology engineers and IC design engineers



courtesy:
ST-microelectronics

BCD (BIPOLAR-CMOS-DMOS) is a key technology for power ICs. BCD is a family of silicon processes, each of which combines the strengths of three different process technologies onto a single chip: Bipolar for precise analog functions, CMOS (Complementary Metal Oxide Semiconductor) for digital design and DMOS (Double Diffused Metal Oxide Semiconductor) for power and high-voltage elements.

This combination of technologies brings many advantages: Improved reliability, reduced electromagnetic interference and smaller chip area. BCD has been widely adopted and continuously improved to address a broad range of products and applications in the fields of power management, analog data acquisition and power actuators -> main application: automotive

1.1 Historical Review of Semiconductor Development

The Long Dawning of Hell-Made Semiconductors

The Impact of Bell Labs

From First Transistors to Today's ICs

1.2 Basic Device Principles and Family Tree

Basic Device Principles

Family Tree

1.3 MOSFETs as Mainstream Devices

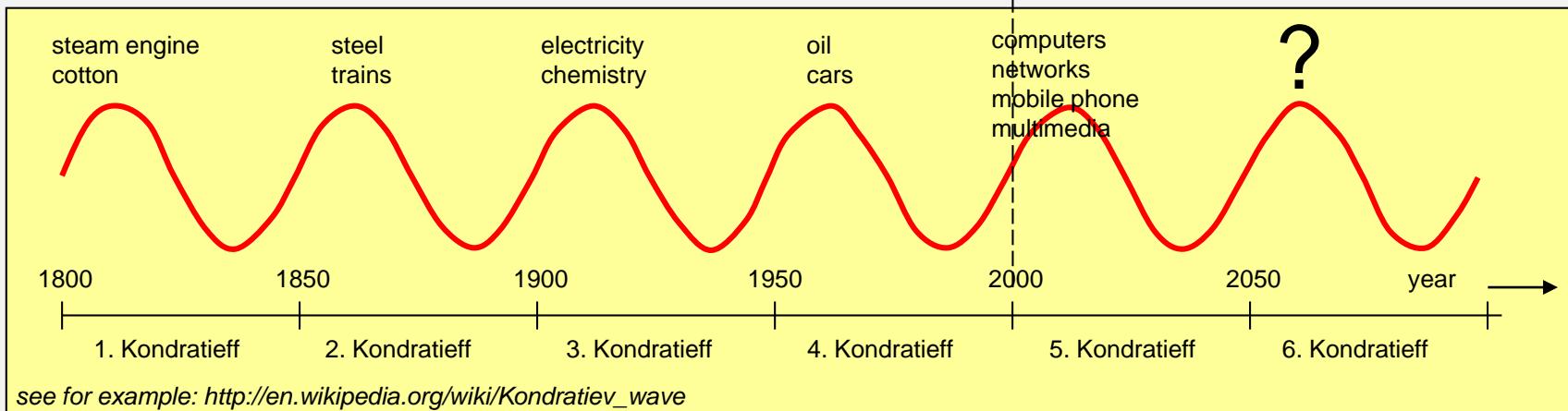
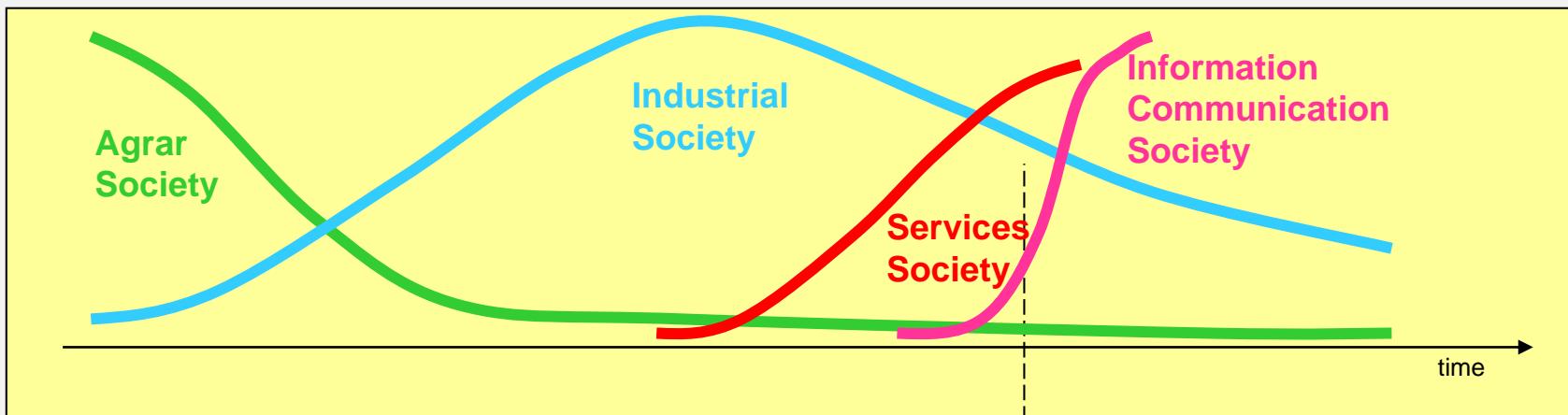
Needs of Information Society

Advantages of MOSFETs for Integration

Future Calculability of MOSFETs

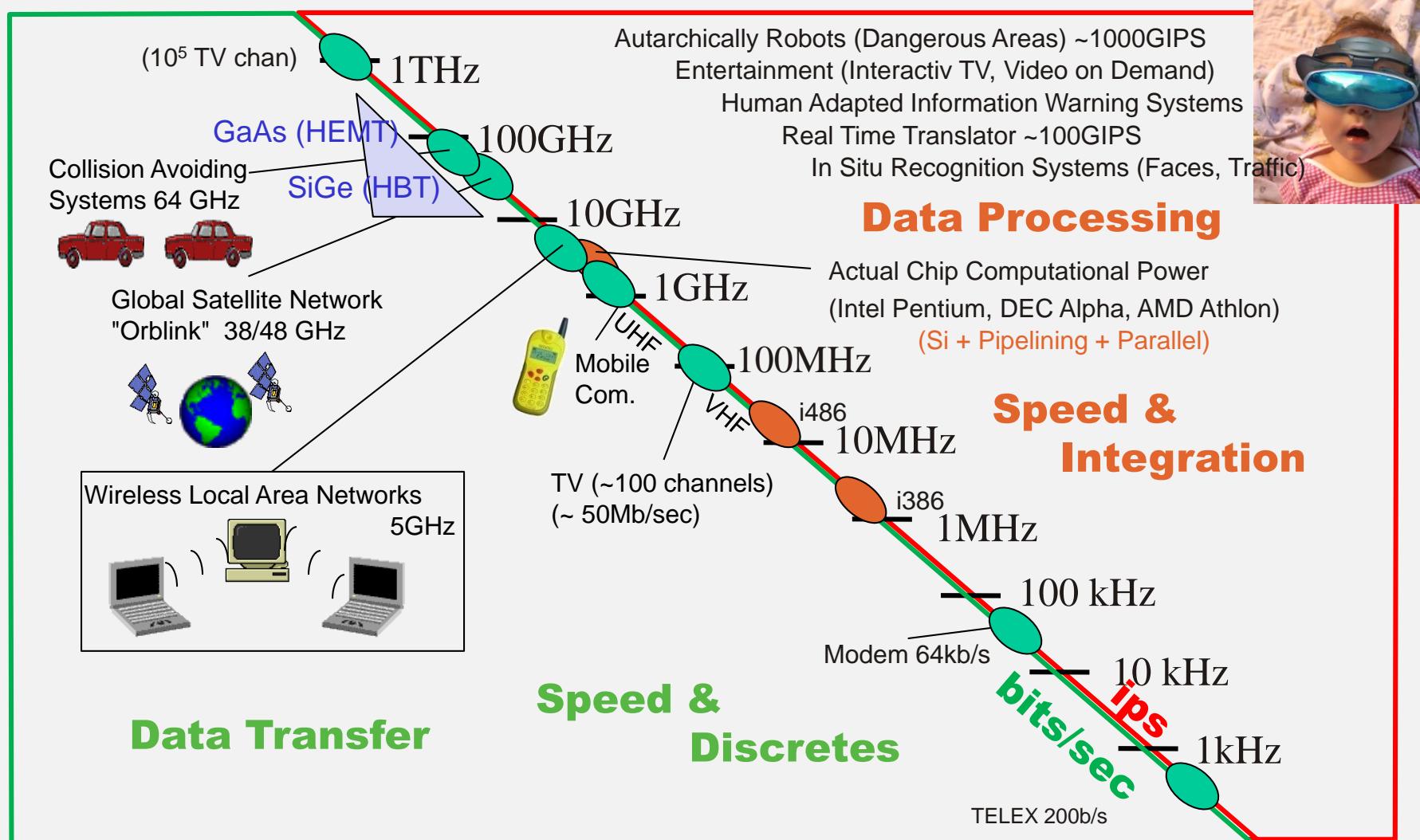
Why are MOSFETs the dominant transistors ?

Let's have a look for the applications and their semiconductor needs



► New inventions may change the society

► The invention of the transistor and the IC enabled the mobile information- and communication society



 Data transfer is done by bipolar discrete devices, data processing is done by CMOS Integrated Circuits

Discretes

if: $v = 10^7 \text{ cm/sec}$
 $f = 1 \text{ THz}$
 $\rightarrow x < 100 \text{ nm}$

Fast devices must be small

Performance driver:

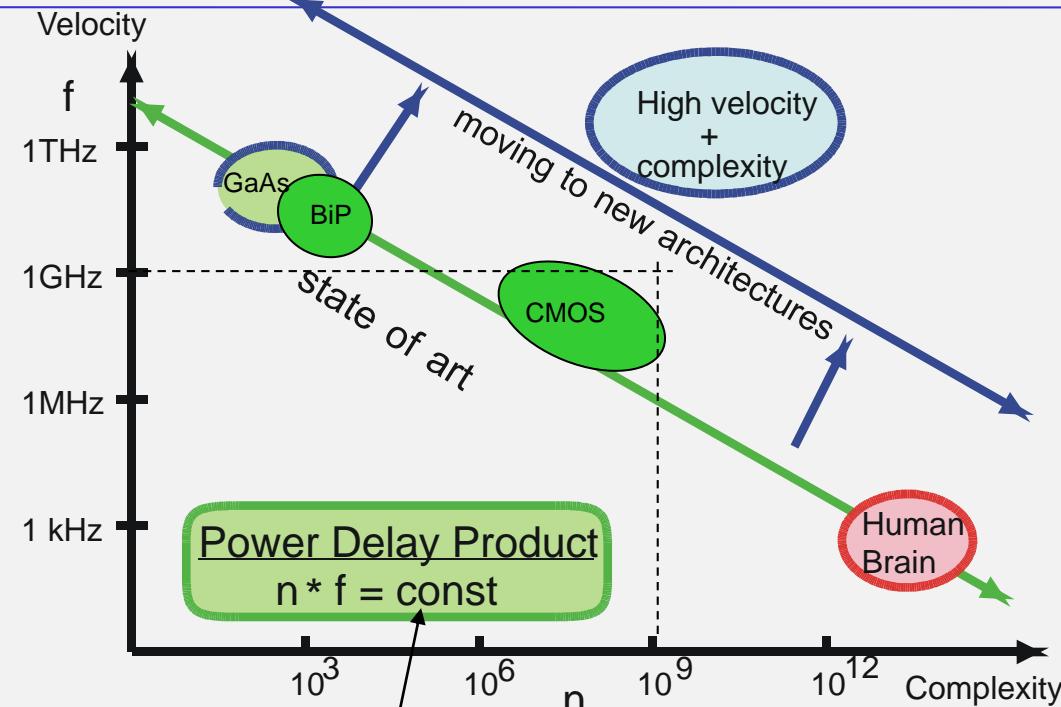
► Increasing speed (data transfer) and/or integration requires shrinking dimensions

► The limitation of maximum power dissipated requires a trade-off between speed and integration



Which are the best technologies for the smallest devices ?

Speed vs. Integration

Integration

$$10 \text{ Gb} = 10^{10} \text{ dev/cm}^2$$

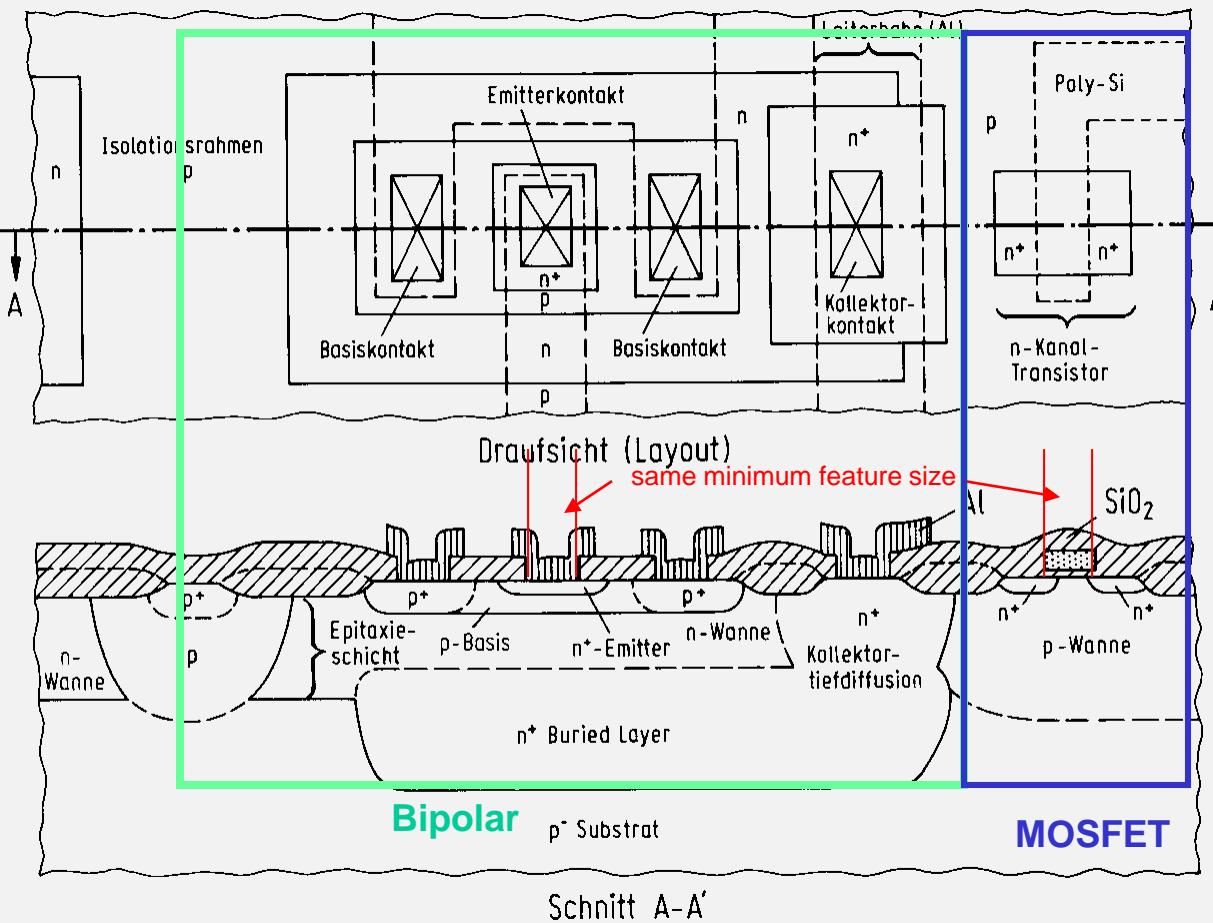
$$10^5 \text{ dev}$$

$$1\text{cm}^2$$

Ultimate Integrated Devices must be small

$$\rightarrow x < 100 \text{ nm}$$





Integration means:

many devices side by side must be switched on/off separately

- > Isolation

- > all contacts on top

Basic device design:

Bipolar: 6 masks

- > large area

- > expensive

- > low yield

MOS: 4 masks

- > smaller area

- > cheaper

- > higher yield

about 4-10 MOSFETs can be placed on the same area of one BiP using the same technology

- > scalable -> safety for shrinking

- > easy design

 MOS is much more suitable for integration than bipolar

Roughly speaking:

Bipolar devices allow faster switching than MOSFETs (due to device physics)
if the IC should be fast, it should have only a few devices (due to power consumption)
if the IC has only a few devices, the IC can be done in bipolar technology (due to costs and yield)

Bipolar

The advantage of **bipolar** is in fast discrete or small scale integration,
the disadvantage is all-time power consumption and large area costs

In commercial applications the number of bipolar devices is low

The advantage of **CMOS** is in high-density integration, because of low cost and low power consumption
the disadvantage is in lower speed by using the same design rules as in bipolar

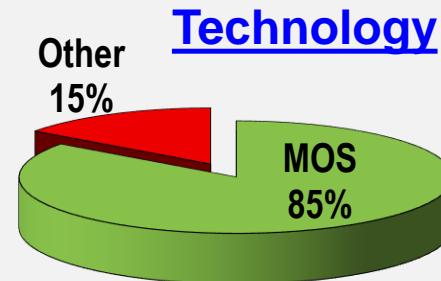
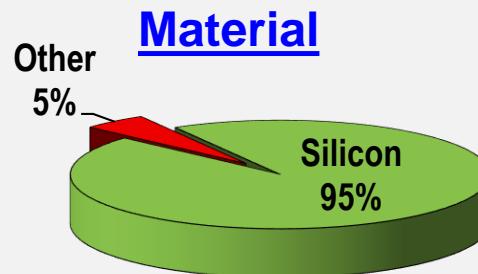
MOS

In commercial applications the number of CMOS devices is immense
(each day $\sim 10^{16}$ MOSFETs are fabricated worldwide)



CMOS is the far dominant device technology where the mainstream money is in

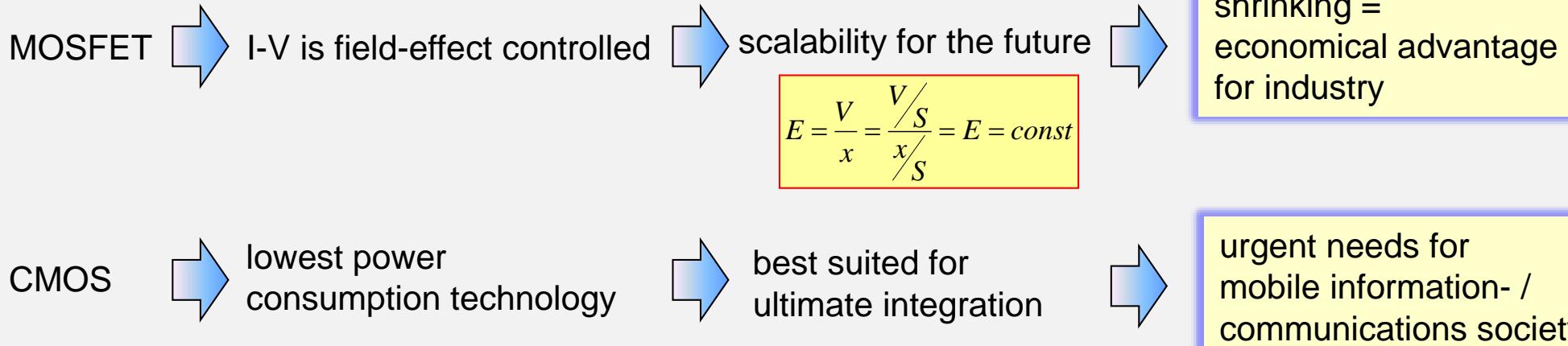
The semiconductor market today:



source: ICE 2009

economical values, not pieces

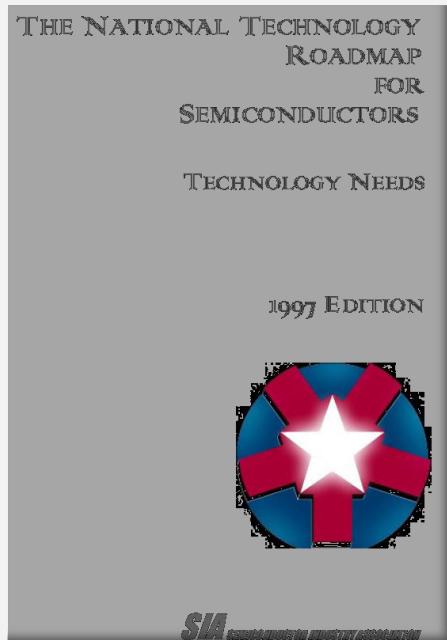
because CMOS offers significant advantages:



1 Introduction

1.3 Mainstream MOSFET

Early roadmap from 1997:



Year of First Product Shipment Technology Generation	1997 250 nm	1999 180 nm	2001 150 nm	2003 130 nm	2006 100 nm	2009 70 nm	2012 50 nm
Min. Logic V_{dd} (V) (desktop)	2.5-1.8	1.8-1.5	1.5-1.2	1.5-1.2	1.2-0.9	0.9-0.6	0.6-0.5
V_{dd} Variation	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$
T_{ox} Equivalent (nm)	4-5	3-4	2-3	2-3	1.5-2	< 1.5	< 1.0
Equivalent Maximum E-field (MV/cm)	4-5	5	5	5	> 5	> 5	> 5
Max I_{off} @ 25°C ($\mu A/\mu m$) (For minimum L device)	1	1	3	3	3	10	10
Nominal I_{on} @ 25°C ($\mu A/\mu m$) (NMOS/PMOS)	600/280	600/280	600/280	600/280	600/280	600/280	600/280
Gate Delay Metric (CV/I) (ps)*	16-17	12-13	10-12	9-10	7	4-5	3-4
V_T 3σ Variation ($\pm mV$) (For minimum L device)	60	50	45	40	40	40	40
L_{gate} 3σ Variation (For nominal device)	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$
...							
S/D Extension Junction Depth, Nominal (nm)	50-100	36-72	30-60	26-52	20-40	15-30	10-20
Total Series Resistance of S/D (% of channel resistance)	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$	$\leq 10\%$
Gate Sheet Resistance (Ω/sq)	4-6	4-6	4-6	4-6	4-6	< 5	< 5
Equivalent oxide thickness T_{ox} (nm)	4-5	3-4	2-3	2-3	1.5-2	< 1.5	< 1.0
Thickness control (% 3σ)	± 4	± 4	± 4	$\pm 4-6$	$\pm 4-8$	$\pm 4-8$	$\pm 4-8$
Drain structure	Drain Extension				+ Elev. S/D	Elev. Single Drain	
Contact X_j (nm)	100-200	70-140	60-120	50-100	40-80	15-30	10-20
X_j @ channel (nm)	50-100	36-72	30-60	26-52	20-40	15-30	10-20
Drain extension conc. (cm^{-3})	1×10^{18}	1×10^{19}	1×10^{19}	1×10^{19}	1×10^{20}	1×10^{20}	1×10^{20}
Channel conc. for $W_{depletion} < 1/4L_{eff}$ (cm^{-3})	1×10^{18}	2×10^{18}	2.5×10^{18}	3×10^{18}	4×10^{18}	8×10^{18}	1.4×10^{19}
DRAM 1st Year Electrical D _r @ 60% Yield/ 3rd Year @ 80% Yield (d/m^2)	2080 / 1390	1455 / 985	1310* / 875*	1040 / 695	735 / 490	520 / 350	370 / 250
MPU 1st Year Electrical D _r @ 60% Yield/ 3rd Year @ 80% Yield (d/m^2)	1940 / 1310	1710 / 1150	1510* / 1025*	1355 / 910	1120 / 760	940 / 640	775 / 525
Solutions Exist Solutions Being Pursued No Known Solution							

Year 2012 -> 50nm Generation
Supply Voltage: (0.5 - 0.6) V

$I_{off} < 10 \text{ nA}/\mu\text{m}$ } $\Delta I = 5 \text{ decades}$
 $I_{on} : 600 \mu\text{A}/\mu\text{m}$
 $\Delta V_G = 315 \text{ mV}$
(with ideal $S = 63 \text{ mV/dec}$)

Equivalent oxide thickness for Long Channel Criterion

S/D junction depth for Long Channel Criterion

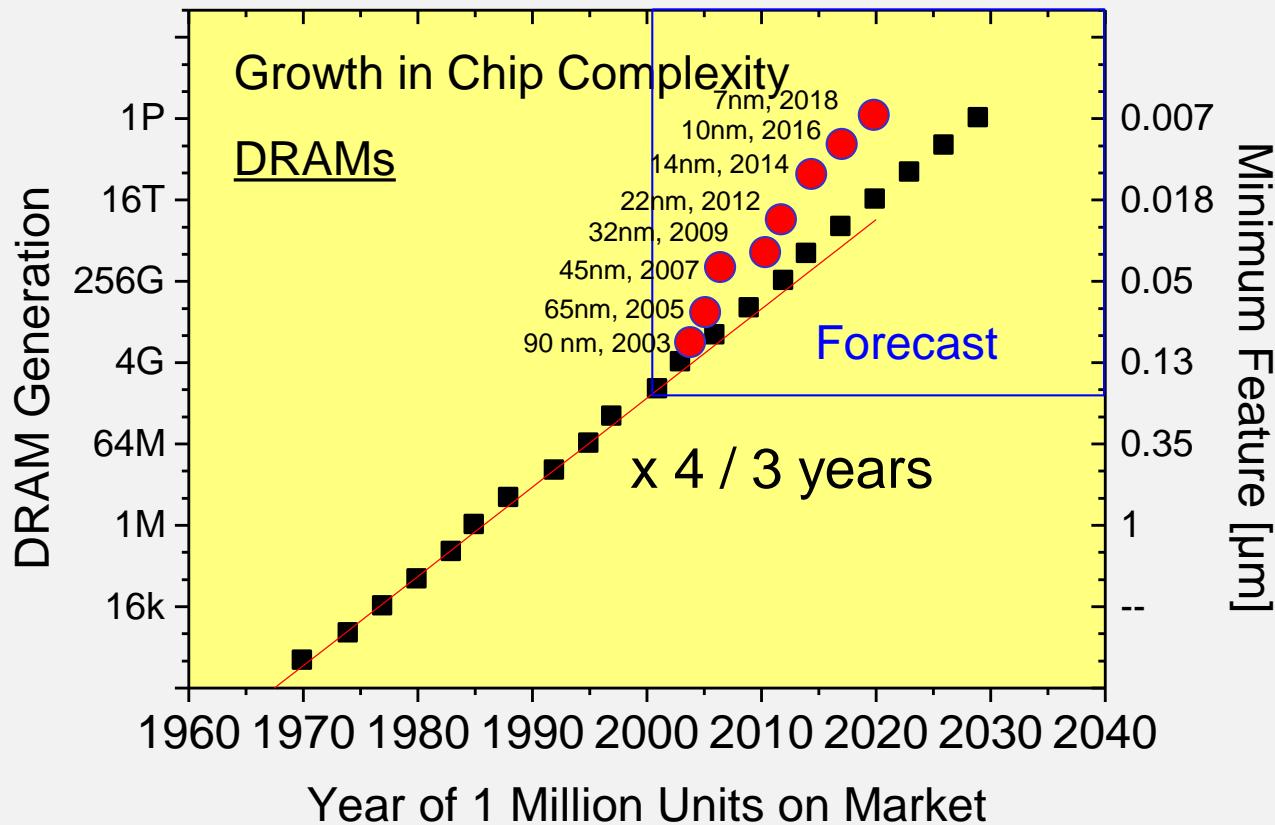
Channel doping for Long Channel Criterion

Starting Yield: 60 %

Roadmaps are only a list of required device properties in time to ensure the prosperity of semiconductor industry

-> Roadmaps are based on planar technology and shrinking
-> for a lot of requirements today no realization possibility (red walls) exist

Moore's Law



The prosperity of the semiconductor market (annual growth by ~17%) is mainly due to cost reduction by shrinking devices in planar technology (see next chapter)

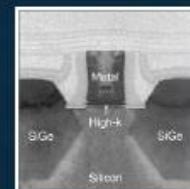
To ensure prosperity in future so-called roadmaps (of shrinking) exist

Only CMOS allows shrinking at lowest cost because of its scalability

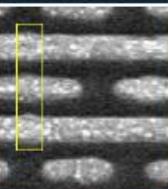
INTEL INNOVATION LEADERSHIP



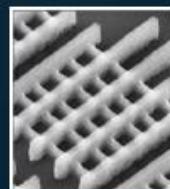
Strained Silicon



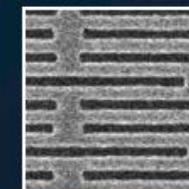
High-k Metal Gate



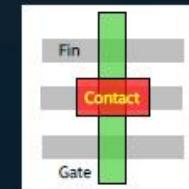
Self Align Via



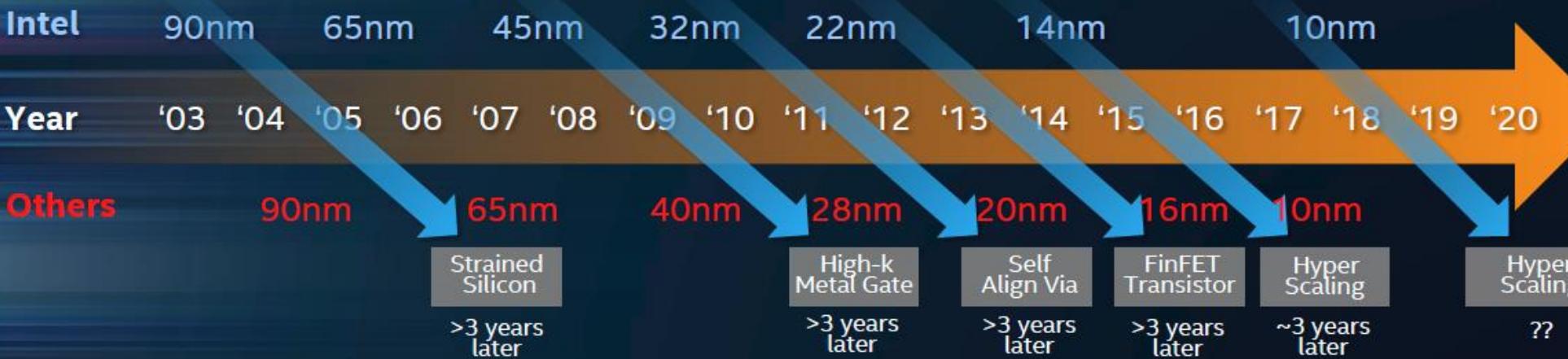
FinFET Transistor



Hyper Scaling



Hyper Scaling



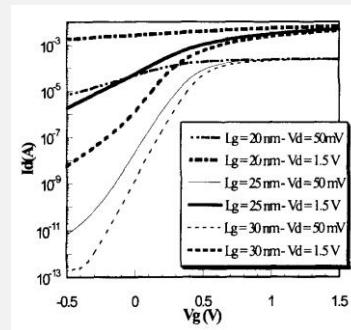
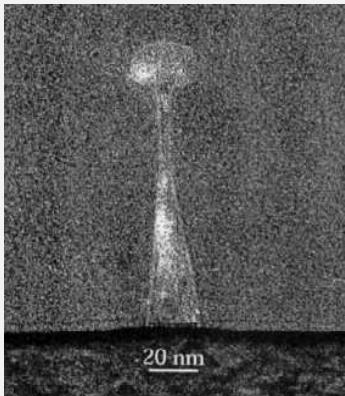
INTEL, 2017



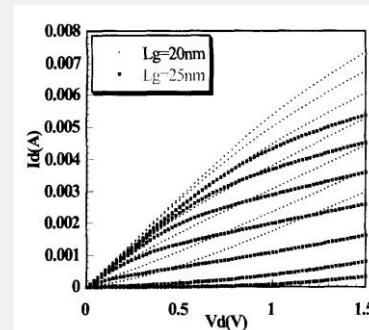
In history INTEL wanted to be the most innovative semiconductor manufacturer

20/7 nm MOSFET (Experiment):

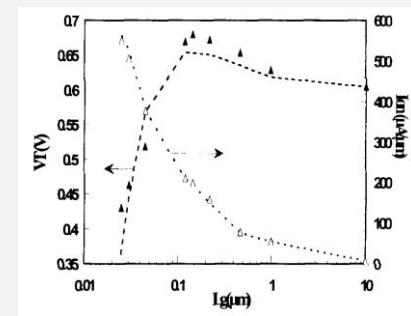
Deleonibus et al., EDL 21 (2000) 173



Ion/loff ~20



Control of punch-through and avalanche

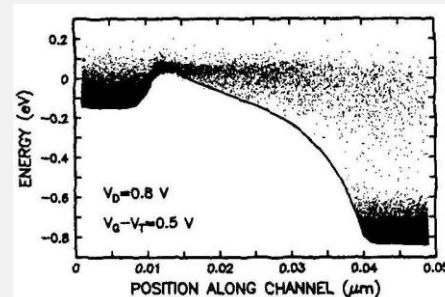
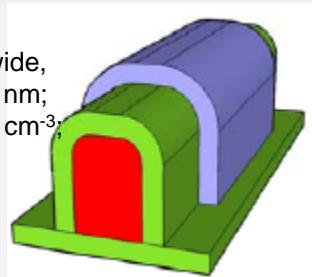


Vth > 0.4V

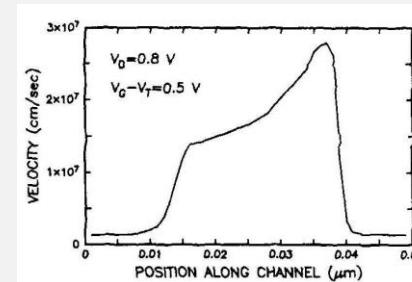
10 nm MOSFET (Simulation):

Frank et al., IEDM'92

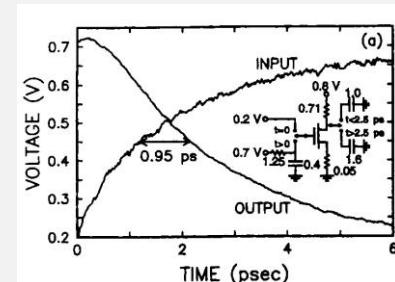
$L_G=10\text{nm}$;
~4nm high, 3nm wide,
S/D extension = 5 nm;
S/D doping = 10^{20} cm^{-3} ;
 $t_{ox}=1\text{nm}$



Ballistics



Overshoot



Inverter: 1psec

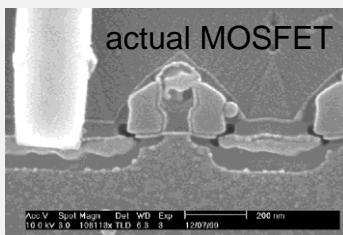
Long term research is almost 20 years before production



Alternatives for CMOS need:

- close to CMOS (=low risk)
- smaller footprint (=cheaper)
- better performance
- more reliable

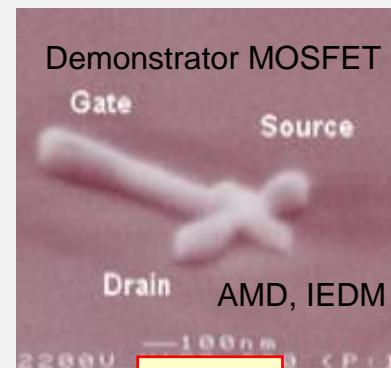
Bulk -CMOS



70 nm
in the middle
of Nano-Roadmap

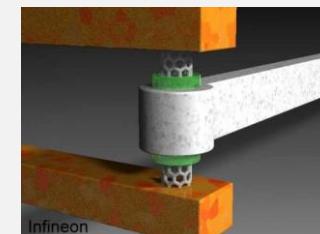
2000

Double-Gate MOSFET



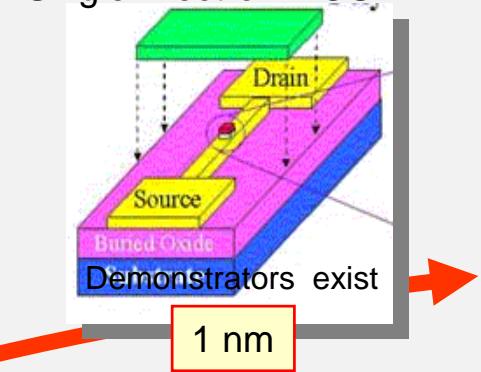
10 nm

close to the end
of CMOS-Roadmap



2020

Single-Electron MOSFET

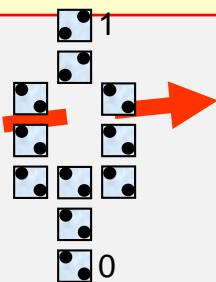


at the end or one step behind
of Roadmap

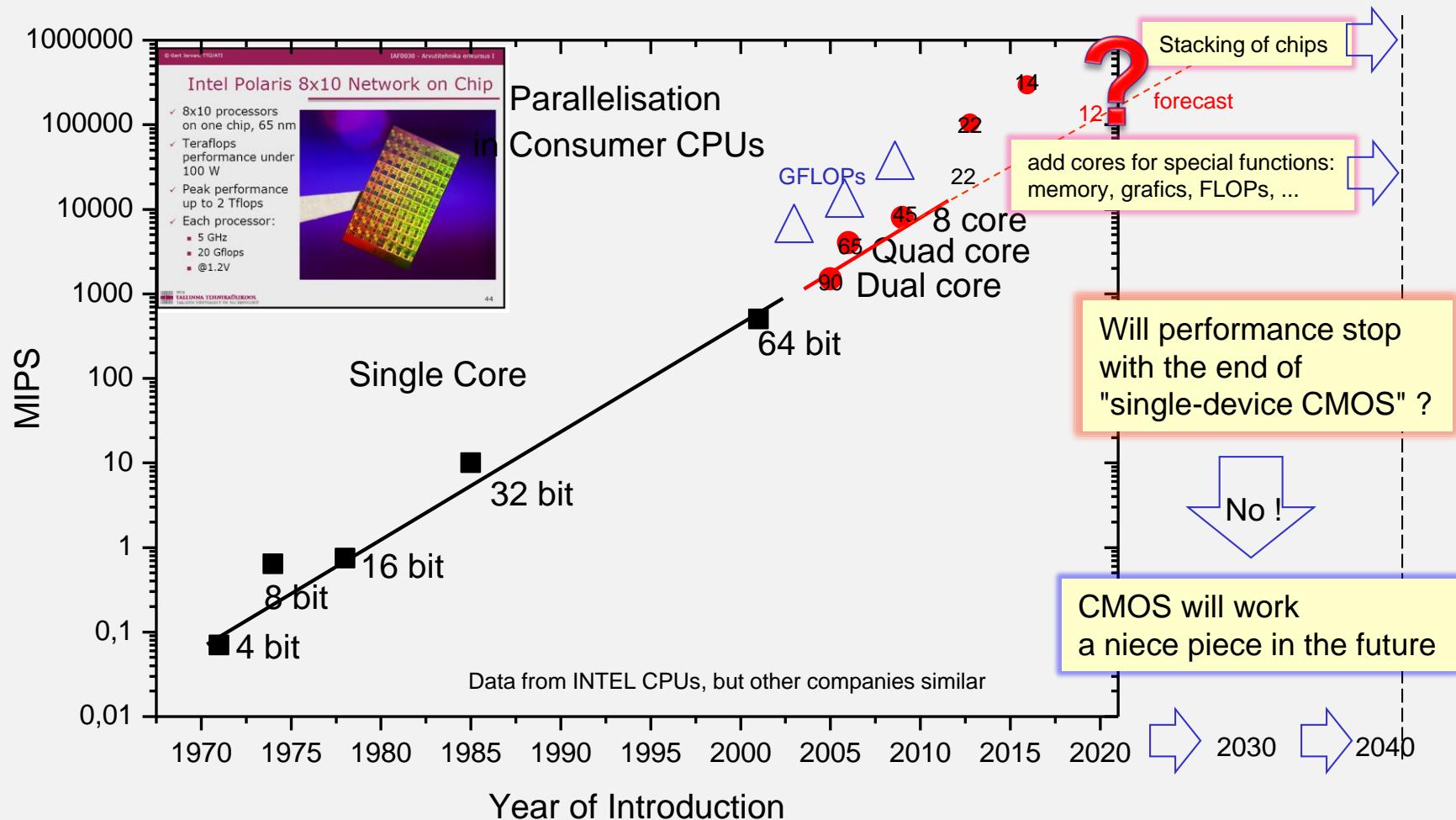


Molecule switches
Nanotubes

Single
Electronics



2040

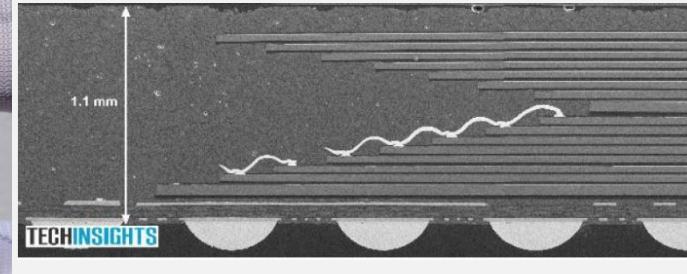
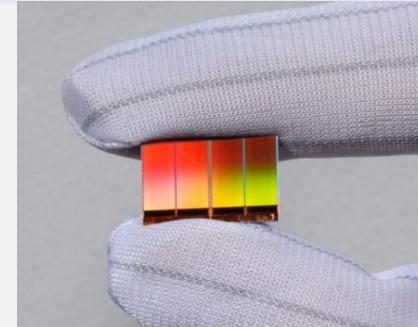
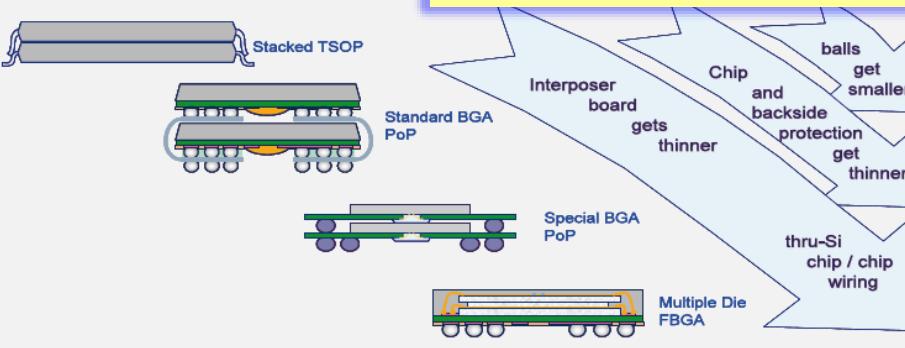


Parallelization of developed technology drives performance, technology is just the water on the mill



No "More Moore" necessary and no "Beyond CMOS" necessary

These chip-stacking technologies are already working !



Wiring around packages
(Package on/in Package)

Wiring around chips
(Multi Die Packages)

Wiring through chips
(Thru-Si-Stacking)

"New" Technologies

2013: SAMSUNG
Producing ~10nm-class NAND Flash Memory,
128 Gb -> 8 chips stacked =>
128 Gb on one chip !
also: Toshiba, Micron similar products

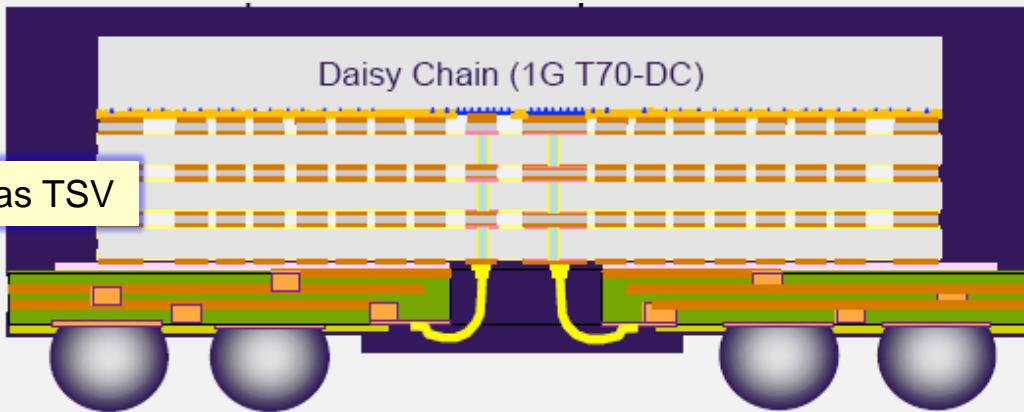
2016: SAMSUNG
Producing ~21nm-class 3D-Vertical NAND Flash Memory,
256 Gb -> 16 chips stacked =>
256 Gb on one chip !

Chip Thinning

Daisy Chain (1G T70-DC)

Through Silicon Vias TSV

Disposer Technology



Again: Performance Improvement with parallelization of conventional CMOS



1.1 Historical Review of Semiconductor Development

The Long Dawning of Hell-Made Semiconductors

The Impact of Bell Labs

From First Transistors to Today's ICs

1.2 Basic Device Principles and Family Tree

Basic Device Principles

Family Tree

1.3 MOSFETs as Mainstream Devices

Needs of Information Society

Advantages of MOSFETs for Integration

Future Calculability of MOSFETs

End
of
Chapter 1