



Architecture of Computers and Parallel Systems

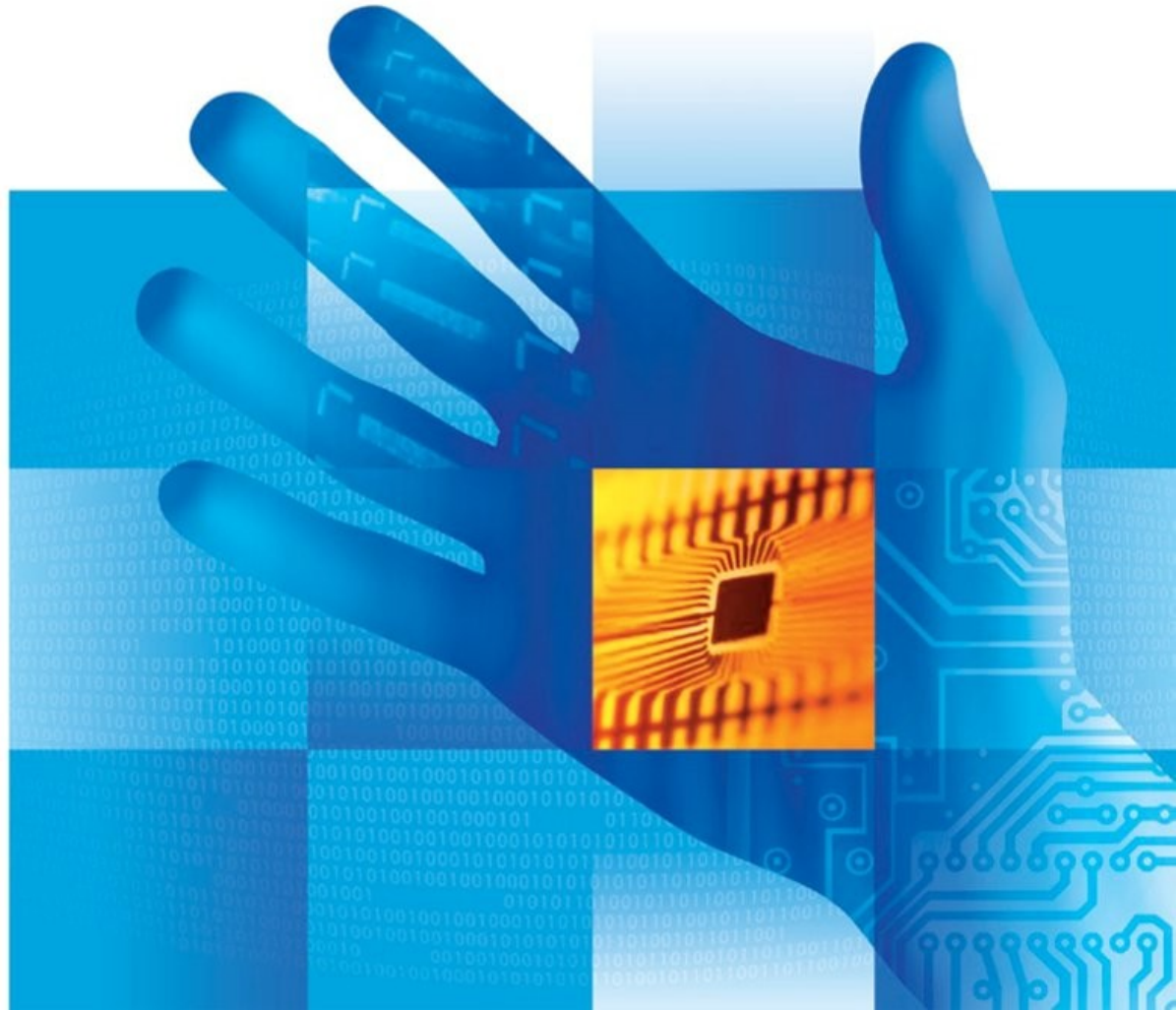
Part 9: Digital Circuits

Ing. Petr Olivka

petr.olivka@vsb.cz

Department of
Computer Science

FEI VSB-TUO





Architecture of Computers and Parallel Systems

Part 9: Digital Circuits

Ing. Petr Olivka

petr.olivka@vsb.cz

Department of Computer Science

FEI VSB-TUO



European
social fund in the
Czech Republic



EUROPEAN UNION

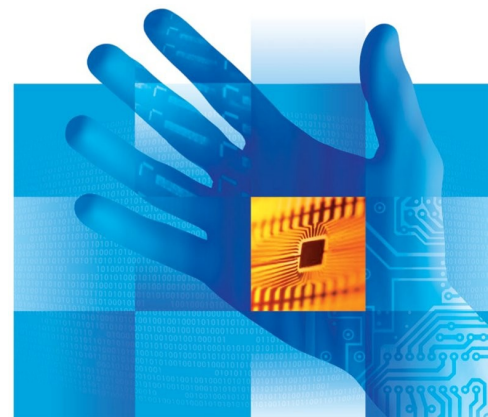


MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT





Architecture of Computers and Parallel Systems

Part 9: Digital Circuits

Ing. Petr Olivka

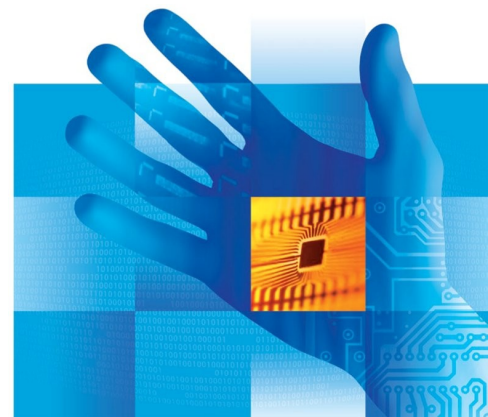
petr.olivka@vsb.cz

Department of Computer Science

FEI VSB-TUO



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Integrated Circuits

The first integrated circuits (IC) were produced just after the transistor had been invented. These circuits integrate all necessary circuit components in one package. In computers, ICs are called Digital Circuits, because they work only with two signal levels. ICs can be divided according to the level of integration:

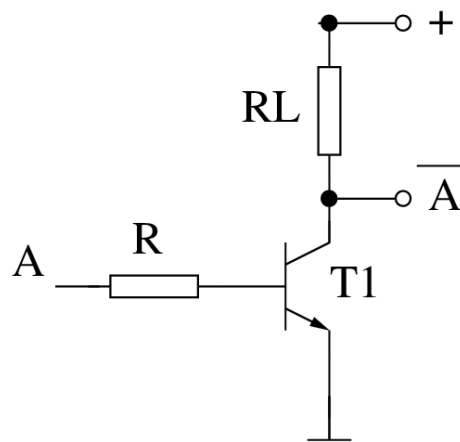
- IC – a few single transistors connected on a small board.
- SSI – Small scale integrity – up to 30 parts in package.
- MSI – Medium SI – up to 1000 parts in package.
- LSI – Large SI – up to 100000 parts in package.
- VLSI – Very Large SI – up to 10 million parts in package.
- ULSI – Ultra Large SI – up to 1 billion parts in package.
- GSI – Gigantic SI – more than ULSI.



Bipolar Technology, RTL

The bipolar technology uses bipolar transistors – NPN and PNP. This technology can achieve the high speed, but it has the high power consumption and it allows less integration.

The first example of the simplest technology is RTL. It uses only resistors and transistors to create two levels of signal (voltage). A simple inverter is shown on the scheme below. The input signal of this circuit is labeled A. The high level of this signal can open transistor T1.



The opened transistor T1 connects the output \bar{A} to the ground – the output is in low level.

When the input signal is in low level, transistor T1 is closed and resistor R_L directly passes the supply voltage to output – \bar{A} is in the high level.

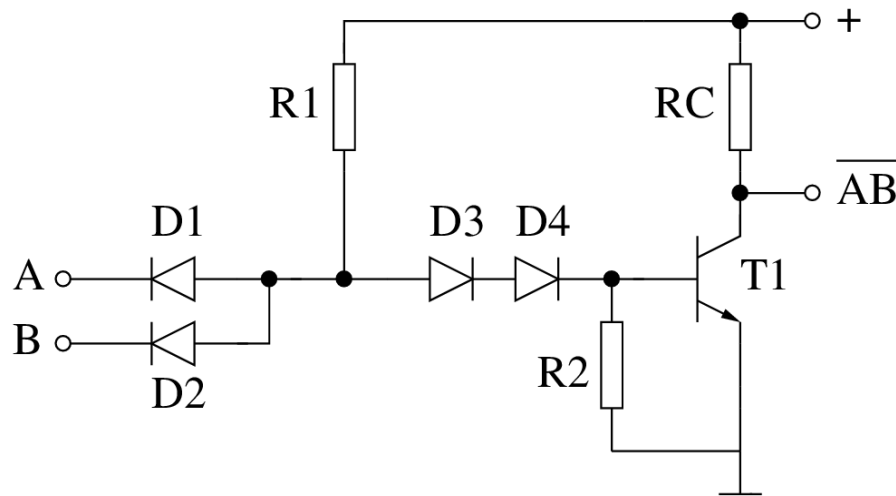
The output is inverted in both cases. This simple schema is very often used in various installations.



DTL (Diode TL)

The second technology DTL was used earlier in the digital circuit construction. The technology uses diodes in addition to the transistors and resistors. The simple example of the NAND circuit is below. The transistor T1 together with the resistor RC works as the inverter.

When A and B are disconnected or they are in log. 1, the voltage over the resistor R1 opens transistor T1 and output is in log. 0. When inputs A or B or both are connected to log. 0, the voltage between

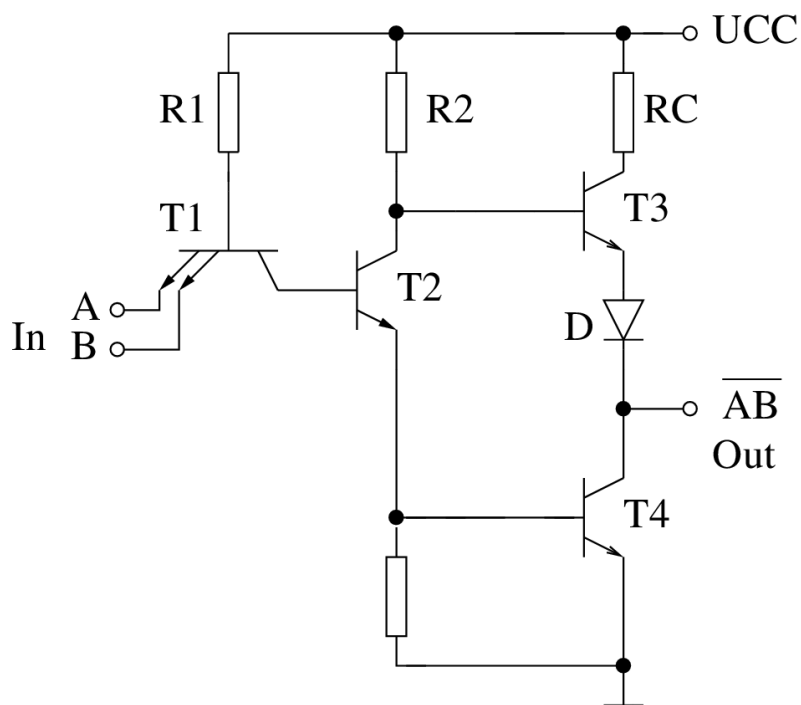


D1 and D3 falls down and the transistor T1 is closed and output is in logical 1. The resistor R2 helps to close transistor and diodes D3 and D4 increase noise immunity. The idea of diodes on input pins is often used to extend some input pins of ICs.



TTL – Transistor -Transistor Logic

The TTL is one of the older technologies used today. Of course its design has been updated. It is used for simple and auxiliary circuits. The figure below shows an example of the NAND gate.



The input is made from the multi-emitter transistor. It works in a similar way as diodes in DTL.

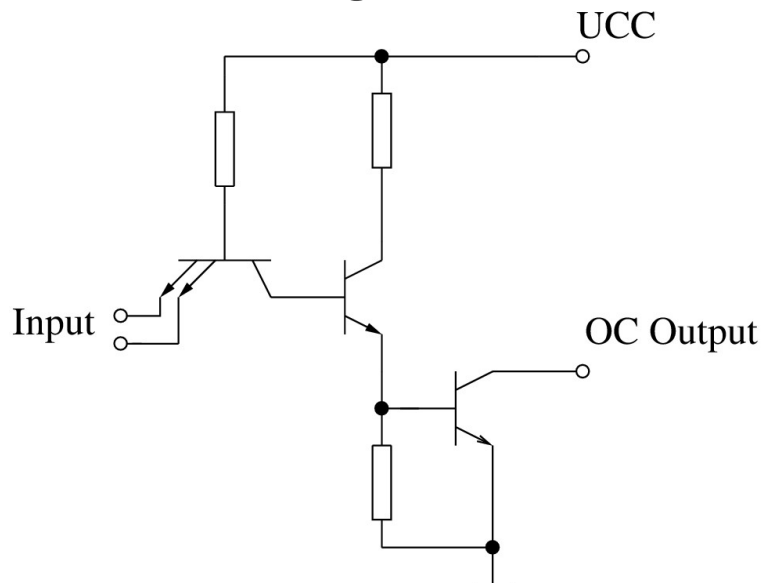
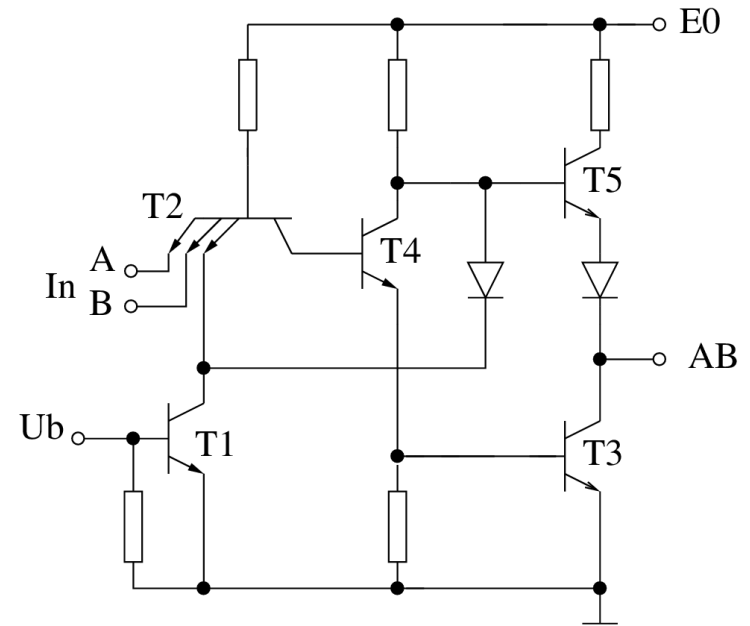
The output is created from two transistors and it has better output parameters, especially the switching speed. Its output voltage in level 1 is not influenced by the size of the load.

The transistor T2 accelerates the transition between output levels from 0 to 1 and back.



... TTL, 3-state, OC

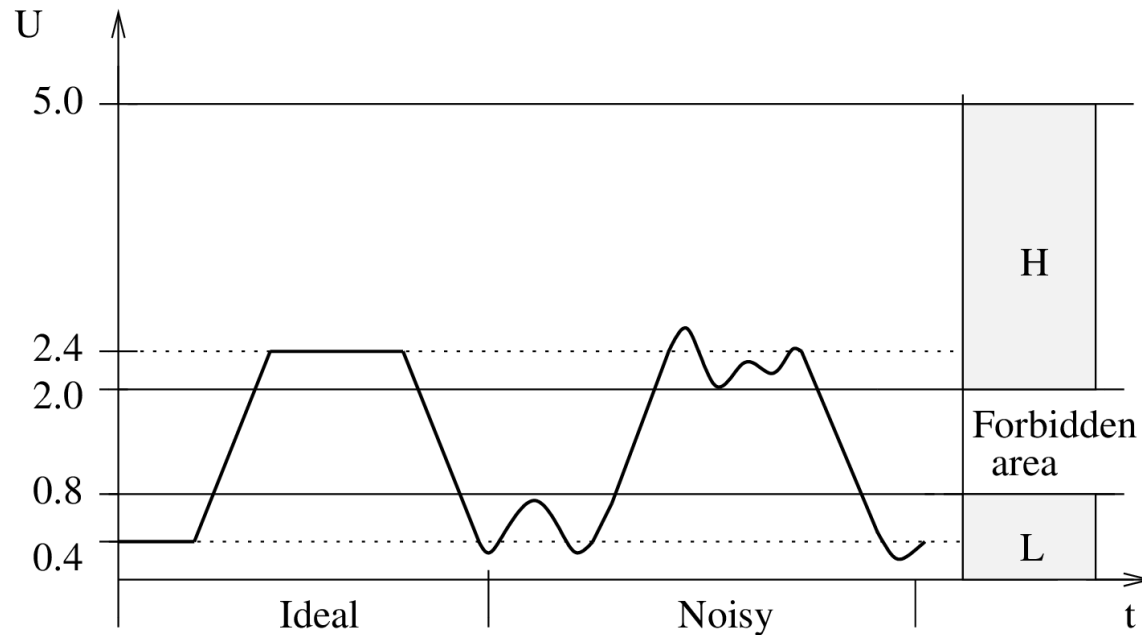
The TTL output part consists of two transistors, which allow a fundamental design feature: **Three state output**. This feature is absolutely necessary for digital circuits in computers. In this way, it is possible to connect more outputs together on the bus. You can see the scheme on right.



For many application another type of the output is used – **Open Collector**. In the output part there is only the lower transistor and its collector is directly drained out. This output can be used to control the signal on buses (many signals are active in 0) or to switch higher voltage than the supply is.



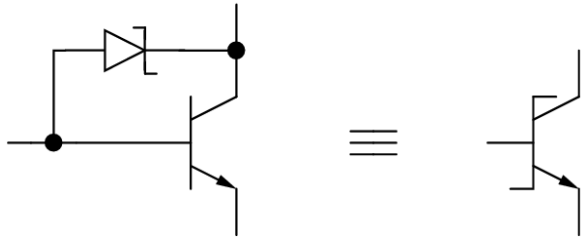
... TTL



TTL digital circuits operate in precisely defined voltage levels. You can see them in the diagram above. The high level signal – logical 1 is above 2.0V. The low level – logical 0 is under 0.8V. Forbidden area is between these both levels. When the signal changes its value, it has to pass through this area so quickly as it is possible. The narrowed zone around the forbidden zone – 0.4V - is the zone of the noise immunity.



... STTL



To speed up transistor switching the Schottky diode (metal-semiconductor) can be implemented in circuit between transistor base and collector. The diode does not allow full saturation of the transistor. Thus transistor can close faster.

Overview of TTL technologies:

	Delay [ns]	Consumption [mW]	Output current [mA]	Input current [mA]
TTL	22	10	16	1.6
HTTL	13	22	30	3
LTTL	35	2	8	0.8
STTL	5	20	50	5
LSTTL	20	3	8	0.8
ASTTL	1.7	8	7	0.7
ALSTTL	15	2	4	0.4



Unipolar Technology

Today the unipolar technology is one of the most important technologies used for IC production. This technology is used in all processors, chip-sets, peripherals, memories, etc...

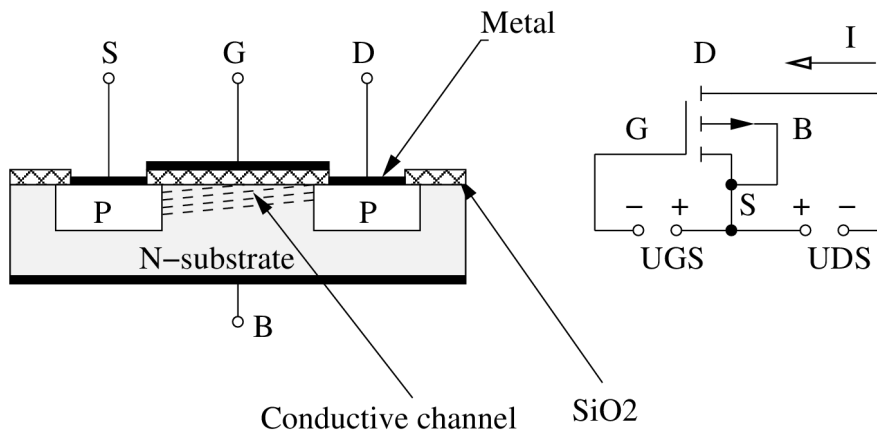
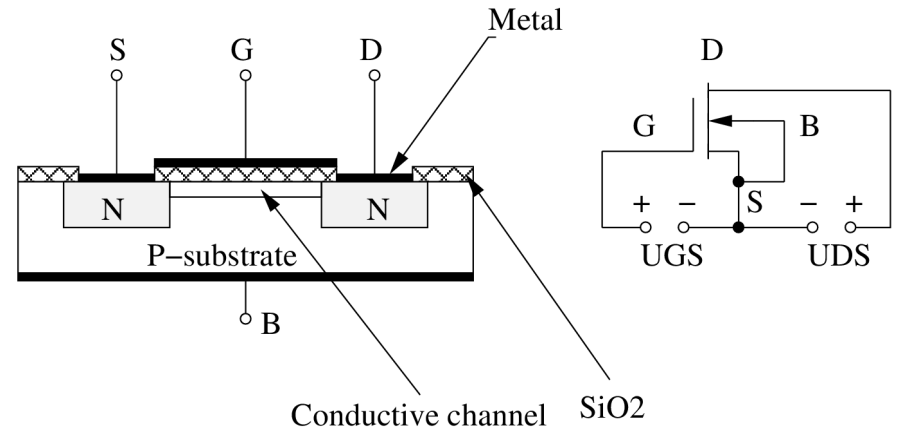
The elementary part of this technology is unipolar transistor. It is better known under the acronym FET – Field Effect Transistor. The FET transistor has three electrodes: S-Source (emitter), D-Drain (collector) and G-Gate (base). We distinguish transistors according to their conductive channel – transistors with N-channel and P-channel.

The FET controls the flow of electrons (or “holes”) from the source to the drain by affecting the size and shape of a “conductive channel” created and influenced by the voltage (or lack of voltage) applied across the source terminal to gate. This conductive channel is the “stream” through which electrons flow from source to drain. The transistor FET with SiO_2 insulation layer is known as MOSFET (Metal Oxide Semiconductor FET).



MOSFET with N & P Channel

The cross section on the right is a FET with the P channel working in the depletion mode. The principle of control is based on narrowing the conductive channel between the S and D by voltage on the Gate.

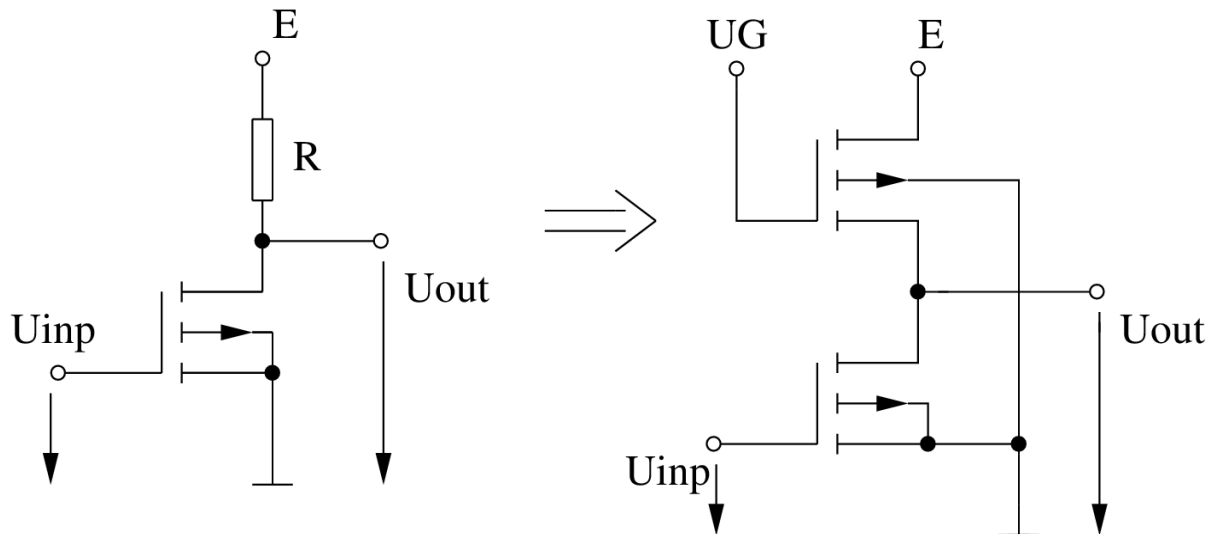


The cross section on the left is a FET with the N channel working in the enhancement mode. The principle of the control is based on the creation and expansion of the conductive channel between the S and D by the voltage on the Gate.



PMOS

The PMOS technology is based on the MOSFET transistor with P channel. The example of an inverter is on the scheme below (one with resistor and second with two transistors). The PMOS is the oldest technology used in digital circuits. The PMOS circuit has low power consumption thanks to FET transistors. But transistors with the P channel are slow and they use supply voltage of $-10 \div -30\text{V}$. The PMOS was also incompatible with TTL.





NMOS, HMOS

The NMOS technology is based on the MOSFET transistor with N channel. Gates are implemented similarly to the PMOS technology. The transistor with N channel is three times faster than the one with the P channel. NMOS circuits are easily compatible with TTL ones and they use supply voltage of +5V.

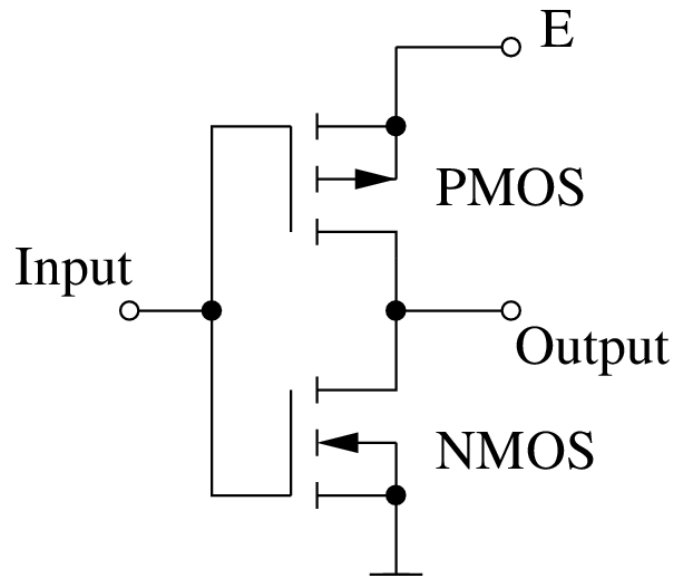
The HMOS technology is based on the fact that the product of the delay and power dissipation is approximately proportional to the cube of dimension of the basic structure. This means that while keeping the value of power dissipation constant, and reducing the structure size by 50%, achieving 8 times faster performance is possible. Or at the same speed it is possible to reduce the energy dissipation 8 times.

The typical delay of the HMOS element is then about 1 ns. The next generation of the HMOS – HMOSII and HMOSIII were used for processors with a few hundred thousand transistors. HMOS gates have a delay in the range from 0.4 to 0.2 ns.



CMOS

The CMOS (Complementary MOS) is the technology based on two technologies: the PMOS and NMOS. All gates in circuits are implemented by both technologies. On the scheme below is the example of the CMOS inverter. The input signal is connected directly to Gates of PMOS and NMOS transistors.

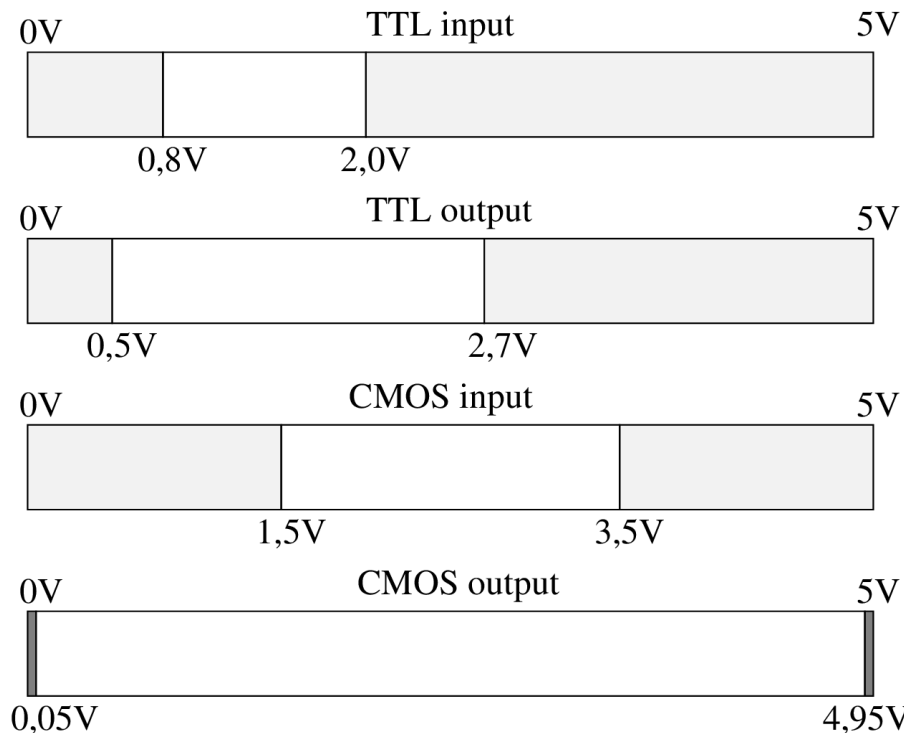


When the Input is in logical 1, the PMOS transistor is closed and the NMOS opened. The output is connected to the ground and its level is in logical 0. And vice-versa. One transistor is always open and the second one is used as high resistance (load) in closed state. The current does not flow in both states from power supply to ground. Gates do not need current either. The consumption of gates is thus zero. This is the main advantage of CMOS!



... CMOS

The CMOS technology is used in all computer parts today. Without this technology we could not achieve such a high performance computing we're used to. The CMOS technology is compatible with the TTL and has very good noise resistance. The comparison between CMOS and TTL voltage levels is in the diagram below:



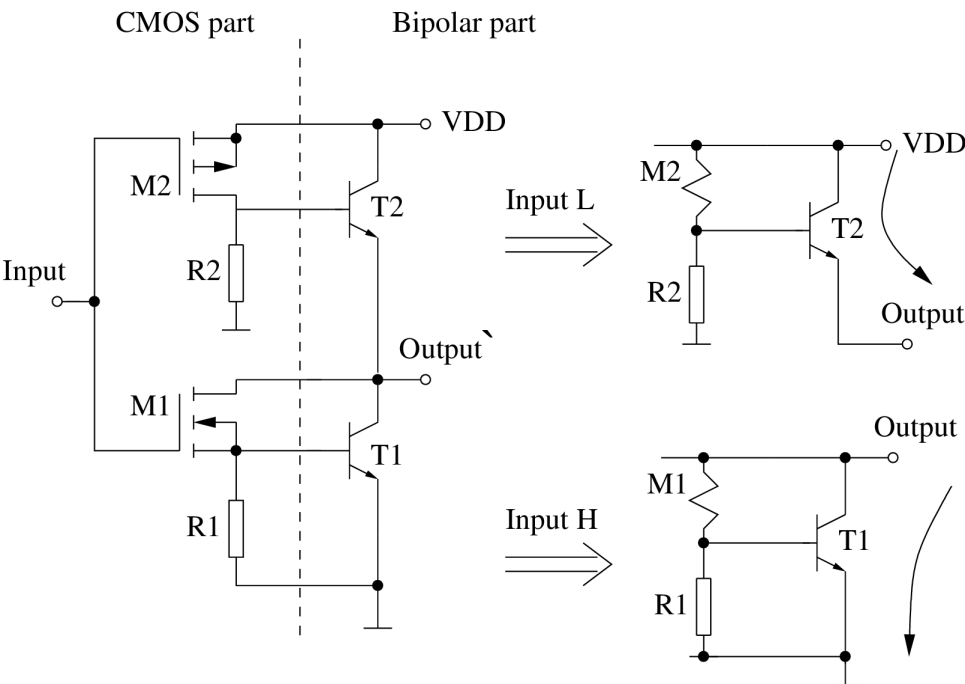


BiCMOS

The BiCMOS technology is a combination of the Bipolar and CMOS technology. The CMOS technology offers less power dissipation, higher packing density and smaller noise margins. The bipolar technology, on the other hand, ensures fast switching and high I/O speed and good noise immunity. The BiCMOS technology accomplishes both - improves the speed over CMOS and has lower

power dissipation than the bipolar technology. The main drawback of BiCMOS technology is the higher cost.

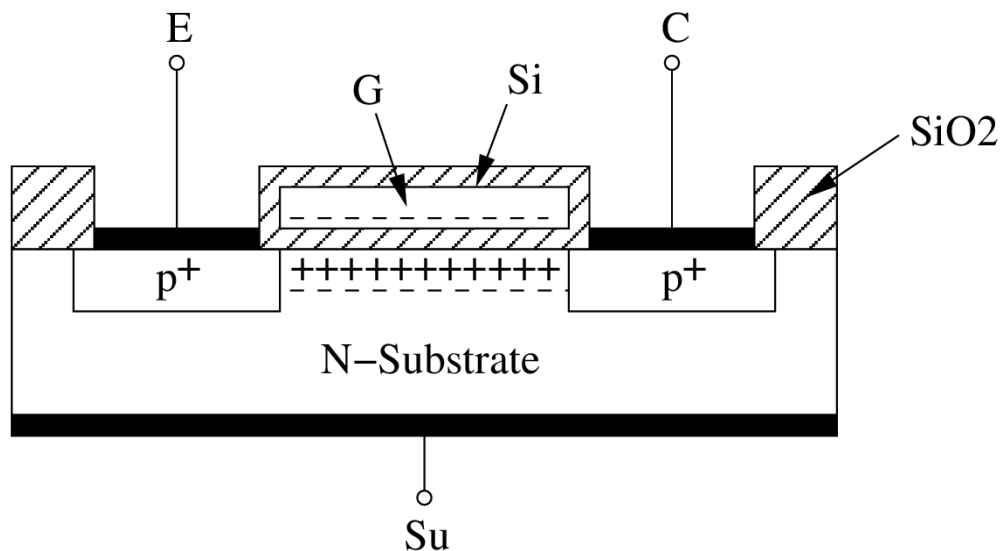
The scheme on the left shows the example of BiCMOS inverter and its behavior with input states H and L.





FAMOS

The FAMOS (Floating-gate Avalanche-injection MOS) is used in the nonvolatile EPROM memory. The special transistor with the insulated gate is used as the single memory cell. The scheme of the transistor is below. The Gate is programmed by “electron injection”. The higher voltage between C and E creates avalanche breakdown and charges the gate. The programming takes about 1 ms. The transistor is in conductive state after programming.



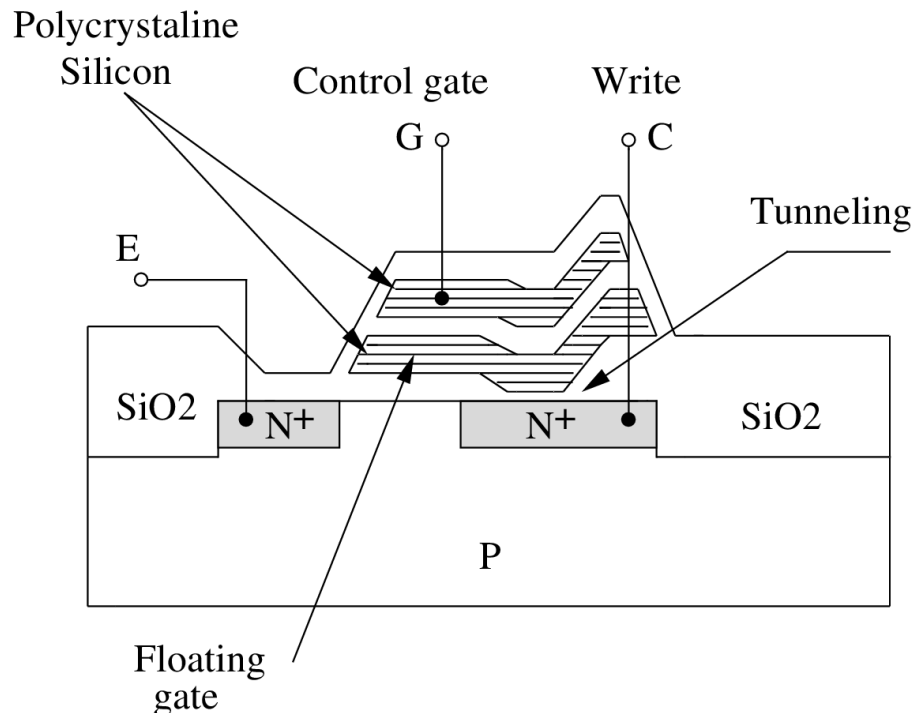
The charged gate can be “cleared” by the UV light. Every erasing process little by little destroys the control gate. The number of erasing and programming cycles is limited.

Information will be safely stored in the memory for the next few decades.



FLOTOX

The FLOTOX (FLOating–gate Tunnel OXide cell) is the nonvolatile memory semiconductor technology. It is a successor of the FAMOS and is used for EEPROM memories. The improvement of this technology is in the ability to clear contents of the memory electrically. No UV light is necessary. The cross section of the cell is shown below. The transistor uses two insulated gates.



The upper gate is used to control the programming process, the Floating gate stores charge and creates the conductive channel.

The charging and discharging of the floating gate is carried by the Tunnel between C and Floating gate.

The memory will safely store information for decades too.



Flash Memory

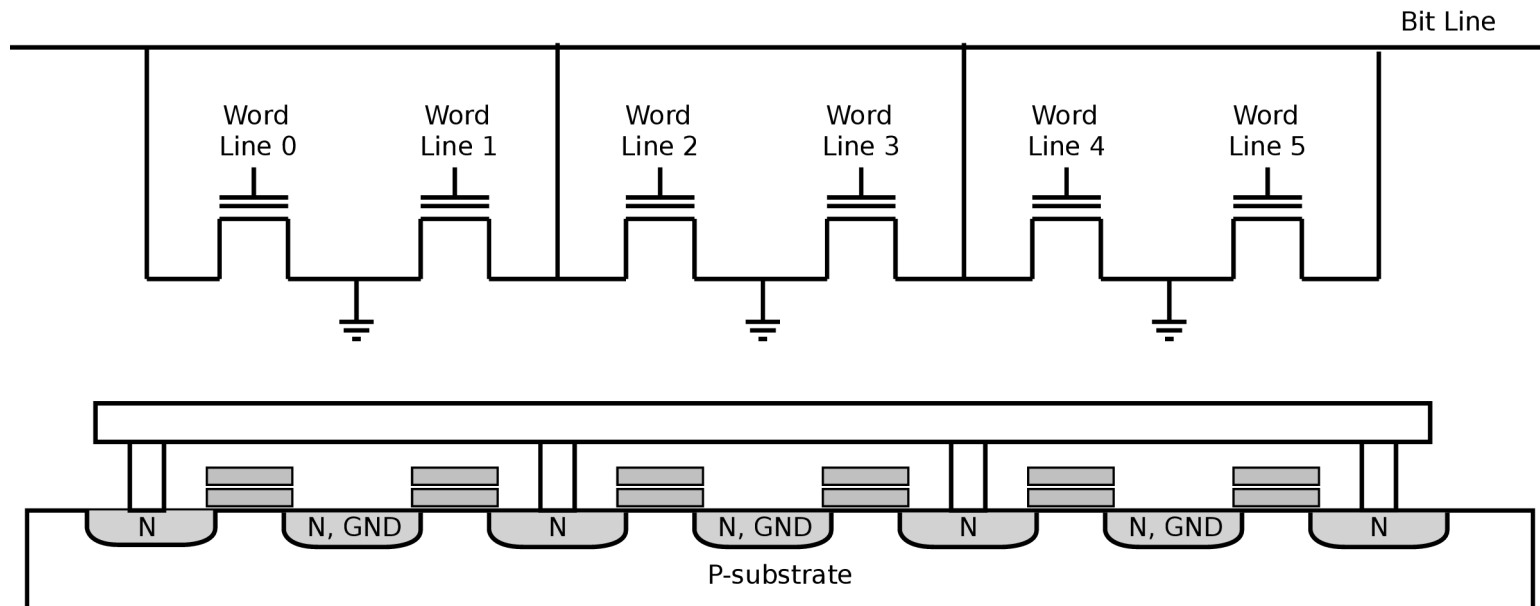
The Flash memory removes one major disadvantage of the EEPROM and that is the higher voltage needed for the programming and erasing the memory. It also reduces the time required for programming the content of memory. It also allows the much higher number of rewrites.

The scheme of the transistor is very similar to the FLOTOX technology, it also uses two gates. But today's transistors are significantly smaller. Different materials are used too. One example of the flash memory technology is SONOS (Silicon-Oxide-Nitride-Oxide-Silicon). In 2005, the improvement of flash technology with the MLC (Multi-level cells) was introduced. The floating gate of the transistor is able to store four levels of the charge. Thus one transistor can store 2 bits. The cell with 2 levels of charge is called SLC - Single Level Cell. The MLC technology usually allows 100 thousand rewrites. But it is little slower than the SLC technology. The capacity of the flash memory of today is up to 4GB per single chip.



NOR Flash Memory

The flash memory is not only classified by the SLC and MLC technology. The internal structure of the chip is important too. The first example of its organization is the NOR flash. See below. In the NOR flash, each cell has one terminal connected directly to the ground, and the other one connected directly to a bit line. When one of word lines (connected to the cell's CG) is brought to high level, the corresponding storage transistor acts to pull the output bit line low. The NOR is suitable for embedded devices due to low latency.





NAND Flash Memory

The NAND flash also uses floating-gate transistors, but they are connected in a way that resembles a NAND gate: several transistors are connected in series, and only if all word lines are pulled to high level, the bit line is pulled low. These groups are then connected via additional transistors to a NOR-style bit line array in the same way that single transistors are linked in NOR flash. Despite the additional transistors, the reduction of ground wires and bit lines allows a denser layout and the greater storage capacity per chip.

