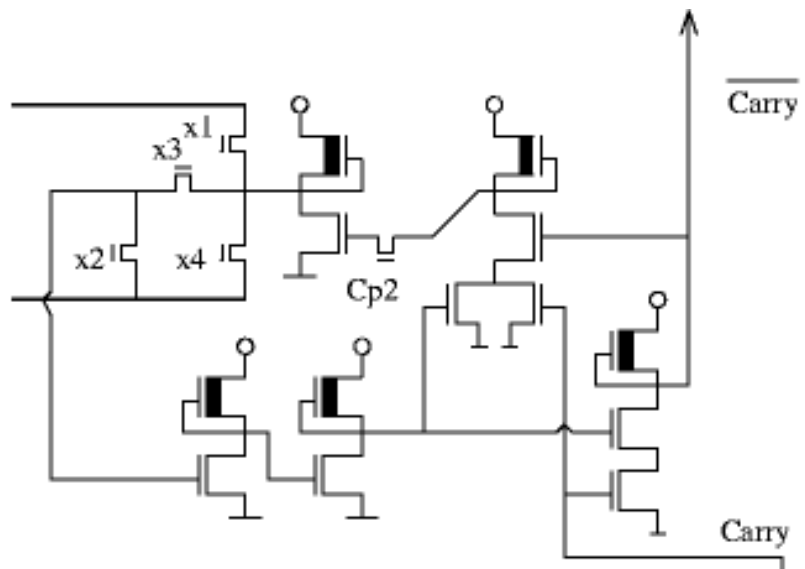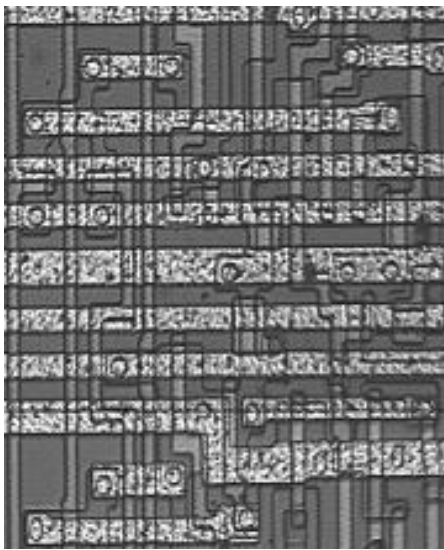The aim of the description below is to provide help for people with little knowledge of electronics. This detailed explanation gives an idea of how integrated circuits work. Thus, using these instructions any layout can be analysed.


Technological introduction

Integrated circuits are formed / developed on a one-crystal silicon slice. Some addictive is given to the silicon, and thus it becomes contaminated –due to the different substances added in a hot furnace– to influence, for instance, conductivity and other features. Diagram forming is made by photo-lithography (photosensitive material with a resist on the surface, illuminated through a mask, solubility of the resist depends on whether it got UV light or not, then milling).

For wiring, poly-crystal silicon and a metal layer is used (this is mostly aluminum or nowadays copper). To divide the different components, an insulating layer is needed, in this case silicon-dioxide. When such a circuit is looked at – either with naked eye or under a microscope – its colourful appearance can be seen at once. Moreover, these colours change when looked at from different angles but silicon is single-coloured, that is, grey.

The explanation of the phenomenon is the following: during production the components are formed in a given order – thin oxide, polysilicon wires, transistor gates, active zones, and the metal layer. Since the different elements are kept in the furnace for different times, depending on the lower substance, the final oxide-layers vary in thickness. The lights reflected from the oxide surface and from the oxide-silicon that divides the surface interfere; the colour depends on the thickness of the oxide and on the angle from which the plate is looked at. The dividing line between the components is clearly seen and this greatly contributes to the analysis of the circuit.

Let us examine now the cross-section picture of an nMOS transistor:
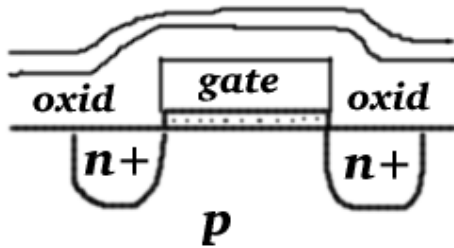
In the picture there are areas with $n+$ additives on a p type carrier; among them a thin oxide layer on silicon (see dotted area), on top of it polycrystal silicon. All this is covered by an oxide layer and a protective layer. The two $n+$ areas are the Source and Drain electrodes of the transistors (their sequence is insignificant here as the structure is symmetrical); the polysilicon is the Gate electrode.
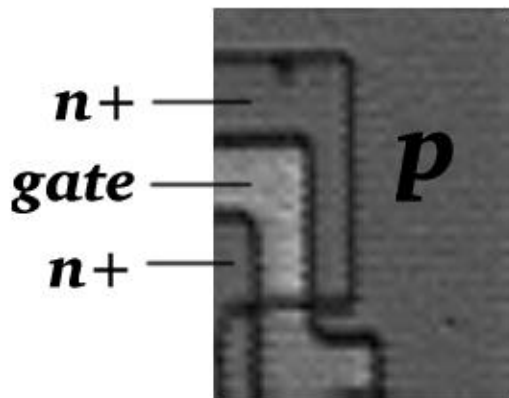

This is what one sees when looking at the above transistor under a microscope. The dark line between areas p and $n+$, which signals the boundary of the active zone (area $n+$), can be clearly seen. Well perceptible is the polysilicon wire winding in a Z-shaped line, which functions as a transistor gate at the same time. This is the first and most important rule of the layout analysis: a transistor can only be found where the active zone is crossed by a polysilicon wire. Areas p and $n+$ are of the same colour but their boundary is unambiguous. By applying a certain rule, which will be made known later, recognising it will cause no problem at all.
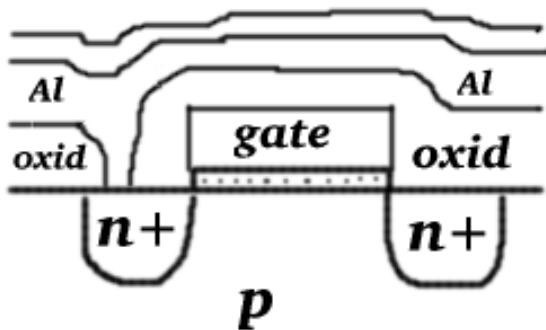
Let us now take the aluminum wire into the cross-section picture. The oxide layer above one of the electrodes is interrupted; there is a passing through the transistor-electrode and the aluminum wire. The aluminum wire and the polysilicon wire cross each other without any pass over; this is impeded by the oxide layer.
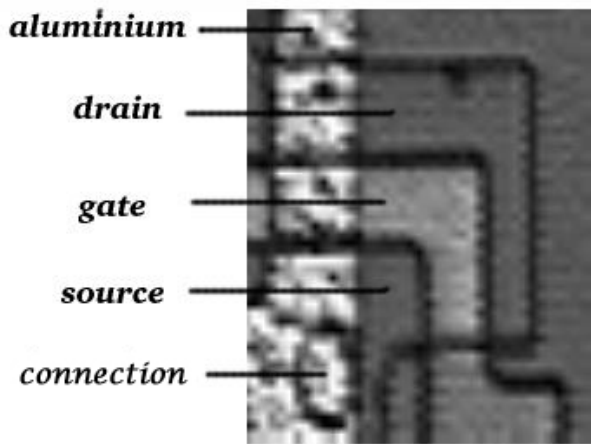
The light grey vertical wire is the aluminum, which crosses the transistor and is connected to the Source electrode through a contact. The circular area is the contact with an active zone under it, but this can only be deduced from the zone boundary, not from the colour. Sometimes it is difficult to analyse such areas, which are covered by large-surface metal layer.
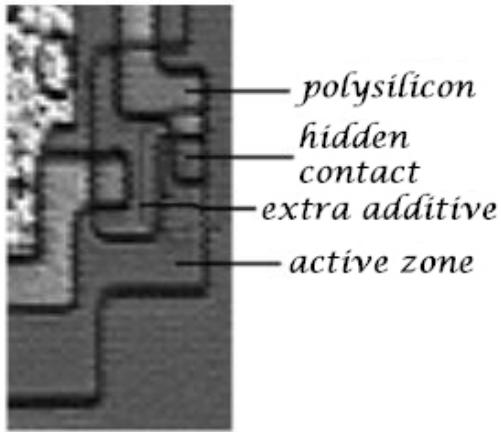
There is another type of contact, the hidden one, which provides a pass over between the active zone and the polysilicon. An extra additive is used to avoid the rectifier feature of such a transition; this can be observed in the layout photo.

Note: The 'hidden contact' could also be called 'buried' and not hidden in the sense that they were trying to deceive or hide from view.
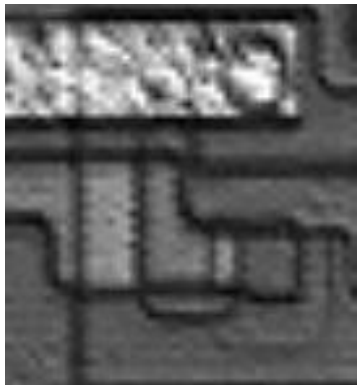
In the end, the last configuration needed to analyse the nMOS layout: depletion mode transistor. This type of transistor can be found in every logical gate. Its gate is always connected to its Source electrode and the Drain is connected to the power supply. In this layout the aluminum wire is +5V, and the gate is reconnected to the Source through a buried contact.

The recognition of the figures shown above is enough to analyse an nMOS integrated circuit, which contains depletion mode gates. The illustrations are from processor No. 6502.

Processor No. 286 is made with identical technology, though the form of its layout is somewhat different.

Control No. 8255 is also an nMOS but with triode load gates; the gate of the load transistor is not reconnected but it is attached to the +12V, the second is to the power supply.

Digital base switches

Although the MOS transistor is used in analogue switches, its real application is in the digital technique. To understand the operation of the MOS transistor and the switches built from it, there is no need of either knowledge of atomic physics, or voltage-electricity characteristics. In the first approach, the MOS transistor is a switch: the voltage on the gate determines whether electricity can flow between the other two electrodes.

Processor No. 6502 and No. 286 contain no other components but the enhancement and the depletion mode transistors and the wires. Their circuit diagram is the following:
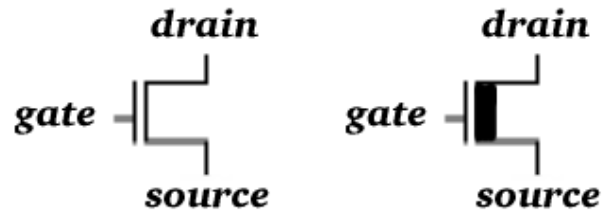
The most important basic switch is the inverter, which is the most frequent gate in processors. It is used not only for inversions but also for the following:

- Delaying
- Signal regeneration (even-numbered inverters on a long wire)
- Storage (inverters attached crosswise)
- Amplification (push-pull inverter on the IC pins)
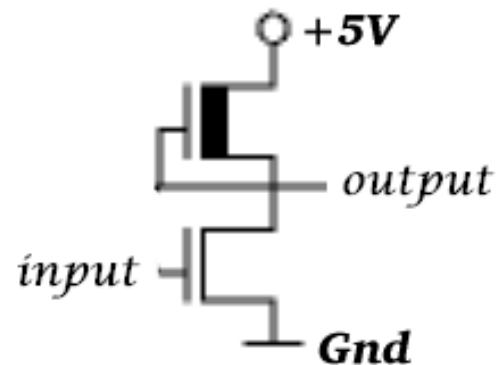
The essence of the operation of the inverter as a two-transistor switch is whether the lower (driver) transistor pulls down the output to 0 potential or not. This depends on the input. If the input is logical 1 (e.g. 5V), the driver opens, that is, it forms a short circuit between the output and the ground; thus the output becomes logical 0.

If the input is 0, the driver transistor is closed as if it was not there at all. In this case the output becomes 1 through the enhancement mode transistor (since this is always open due to the gate-source connection). Of course, the enhancement mode transistor is open even if the input is 1 (and the driver also opens). Designers make is possible that the two transistors do not strain against each other by measuring out the channel width of transistors properly. Thus, there is a definite switch if the input switches over 1 from 0, that is, the driver has much power.

All the other logical gates operate by the same principle. The output of each gate is determined by a depletion load (all of them with a negated output) and the logical function is determined by the switch of the transistors controlled by the input. These transistors can have either a series or a parallel connection. Therefore, if the two inputs are bound to transistors connected in parallel, for the output to become 0, one OR the other input has to be 1. Consequently, the gate is a NOR gate due to the negated output. If the inputs are on transistors that have series connection, one AND the other transistor has to open for the output to be 1; this is the NAND gate.

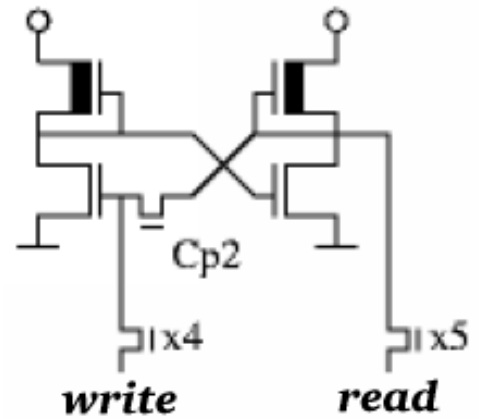$$\overline{A*B}$$

NAND

$$\overline{A+B}$$

NOR
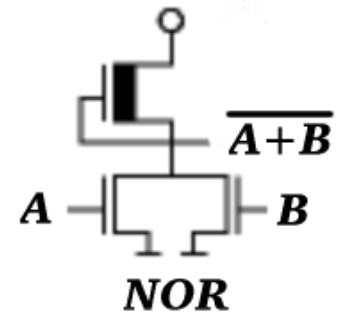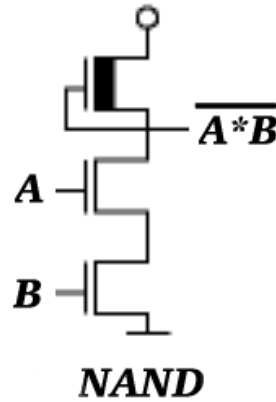
There are so-called complex gates too, with multiple inputs connected in series and in parallel, the functioning of these can easily be deduced by the NAND and NOR logic.

On a dataline, the flow of data can be switched by the help of a transfer gate: this transistor is connected in the bus by its source-drain pin. If the gate gets logical 1 (e.g.5V), it behaves as a short circuit and thus does not disturb the flow of data. At logical 0 there is a break, that is, the status of the output side does not change. (Of course, this is also symmetrical. Therefore, to decide which is the output, one has to see which side is connected to the input of the gate.) More than three transfer gates are never connected in series because it damages the signal level. (Opening voltage of the MOS transistor.) If more than three are needed, even-numbered inverters have to be connected in the bus. (Signal regeneration.)

transzfer gate
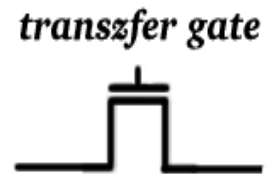
The 1 bit register seen in the picture is from processor No. 6502. Storage is done by a cross-connected pair of inverters; the input of one of them has the information, then, after two inversions, the unchanged bit gets on the same input through feedback. In the feedback branch there is a transfer gate controlled by a Cp2 to ensure that the feedback operates only in the second half of the clock phase and not continually. Otherwise, the overturn of the storage would be difficult; the inverters would work against each other. Writing and reading the storage can be done by switching x4 and x5 transfer gates on the given wire. Storage function can be created not only with inverters but also with an optional complex gate, the output of which has a feedback on its input by adequate timing. (For instance, in 6502 flags work in this way.)