Fundamentals of Computer Systems

Transistors, Gates, and ICs

Stephen A. Edwards

Columbia University

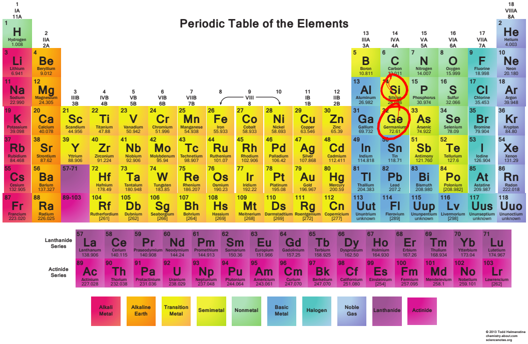
Summer 2021

Semiconductor

sem·i·con·duc·tor

noun

1. A substance, such as silicon or germanium, with electrical conductivity intermediate between that of an insulator and a conductor
2. A semiconductor device



Sand into Silicon

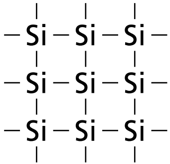
Silica a.k.a. SiO_2 a.k.a. Quartz

$$SiO_2 + 2\ C \rightarrow Si + 2\ CO$$

Elemental, amorphous silicon

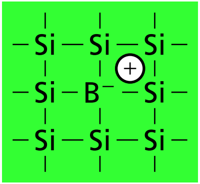
Monocrystalline
Silicon Ingot

Doping Silicon Makes It a Better Conductor



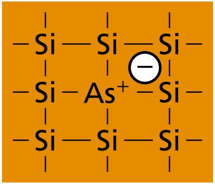
Undoped (pure)
silicon crystal

Not a good
conductor



p-type (doped)
silicon:

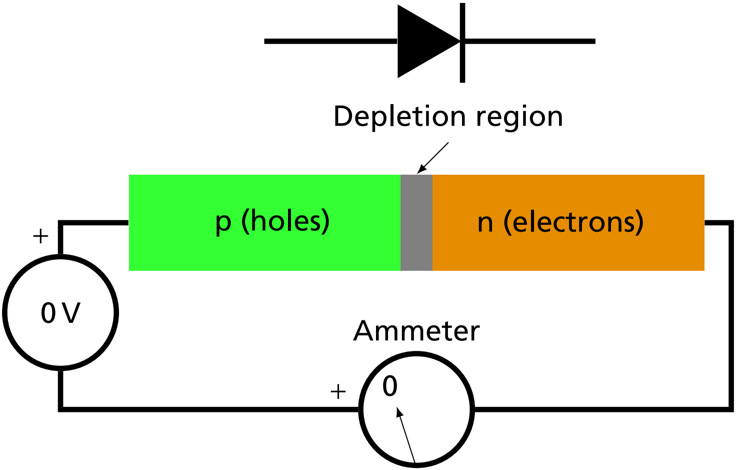
boron atom steals
a nearby electron



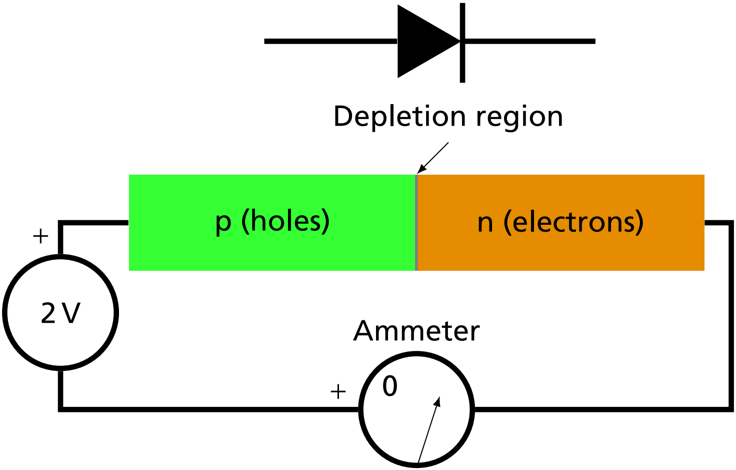
n-type (doped)
silicon:

arsenic’s extra
electron jumps loose

A PN Junction aka A Diode

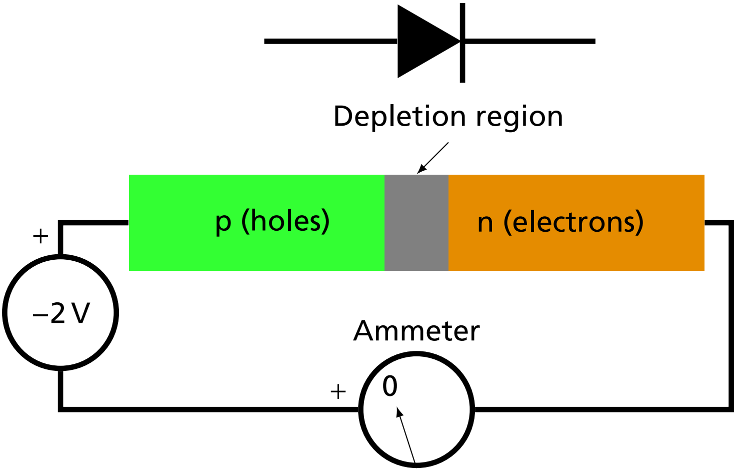


A PN Junction aka A Diode



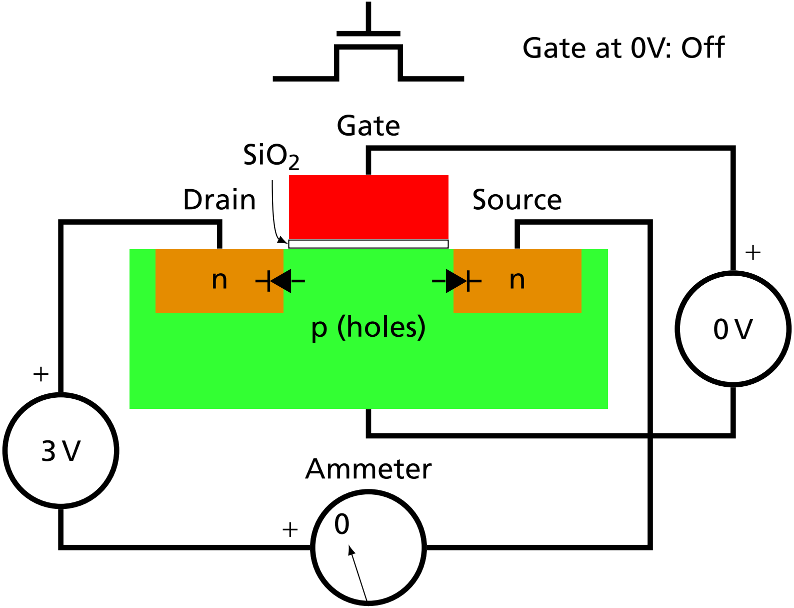
Forward biased: current flows

A PN Junction aka A Diode

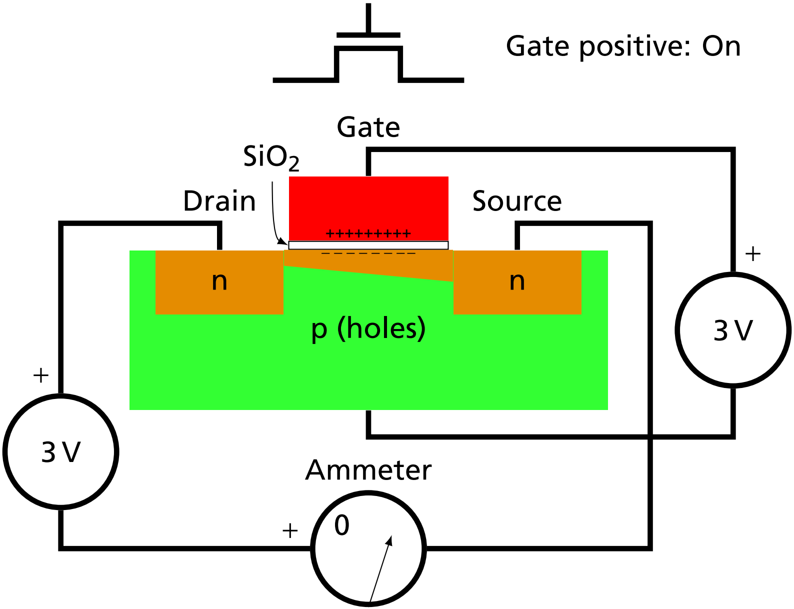


Reverse biased: no current flow

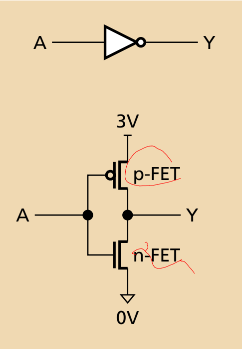
An N-Channel MOS Transistor



An N-Channel MOS Transistor



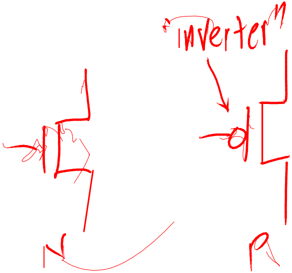
The CMOS Inverter



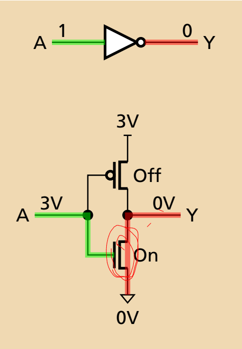
An inverter is built from two MOSFETs:

An n-FET connected to ground

A p-FET connected to the power supply



The CMOS Inverter



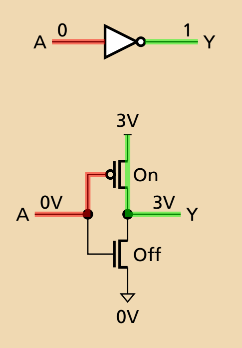
When the input is near the power supply voltage (“1”),

the p-FET is turned off;

the n-FET is turned on, connecting the output to ground (“0”).

n-FETs are only good at passing 0’s

The CMOS Inverter



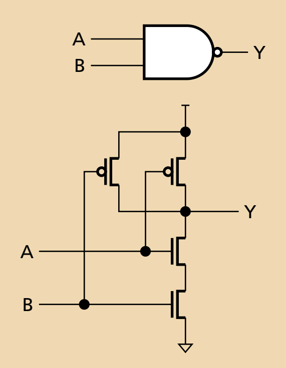
When the input is near ground (“0”),

the p-FET is turned on, connecting the output to the power supply (“1”);

the n-FET is turned off.

p-FETs are only good at passing 1’s

The CMOS NAND Gate

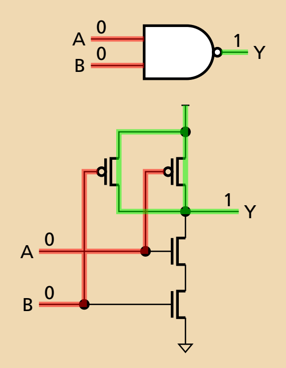


Two-input NAND gate:

two n-FETs in series;

two p-FETs in parallel

The CMOS NAND Gate

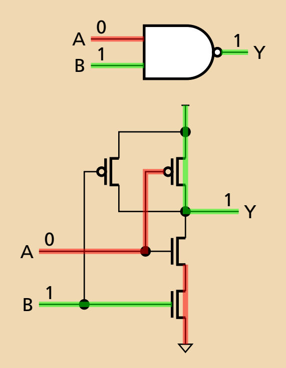


Both inputs 0:

Both p-FETs turned on

Output pulled high

The CMOS NAND Gate



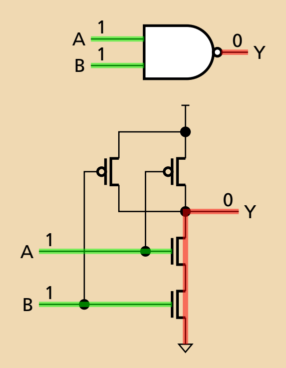
One input 1, the other 0:

One p-FET turned on

Output pulled high

One n-FET turned on, but does not control output

The CMOS NAND Gate



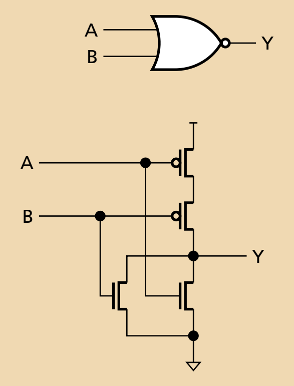
Both inputs 1:

Both n-FETs turned on

Output pulled low

Both p-FETs turned off

The CMOS NOR Gate



Two-input NOR gate:

two n-FETs in parallel;

two p-FETs in series.

Not as fast as the NAND gate because n-FETs are faster than p-FETs

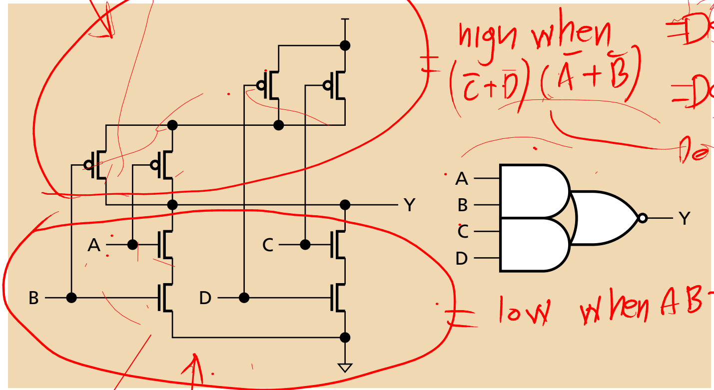
A CMOS AND-OR-INVERT Gate

True network

AND = parallel

OR = series

$\overline{AB + CD}$



high when $(\bar{C}+\bar{D})(\bar{A}+\bar{B})$

De Morgan says $\overline{AB + CD}$

low when $AB + CD$

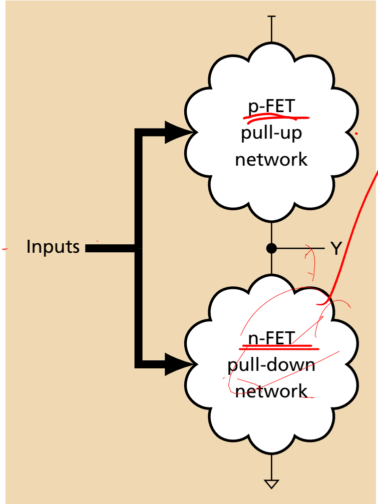
$\overline{AB + CD}$

False network

AND = series

OR = parallel

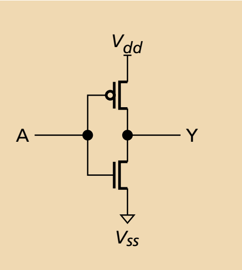
Static CMOS Gate Structure

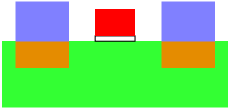


Pull-up and Pull-down networks must be complementary; exactly one should be connected for each input combination.

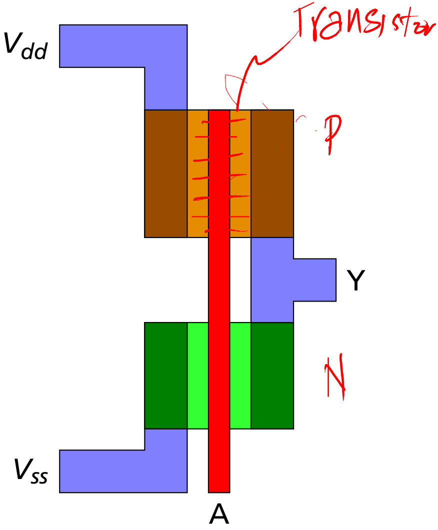
Series connection in one should be parallel in the other

CMOS Inverter Layout





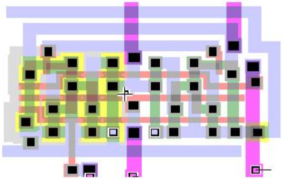
Cross Section Through
N-channel FET



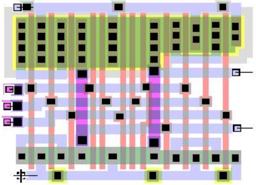
Top View

Full Adder Layouts

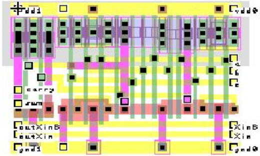
fa_ly_mini_jk size: 60 · 40µm (1.2µmCMOS)



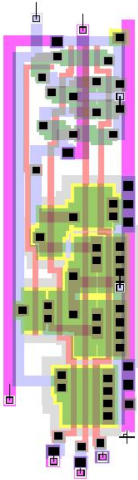
fa_ly_opt1 size: 63 · 50µm (1.2µmCMOS)



Fulladd.L size: 37 · 26 µm (0.5µCMOS)



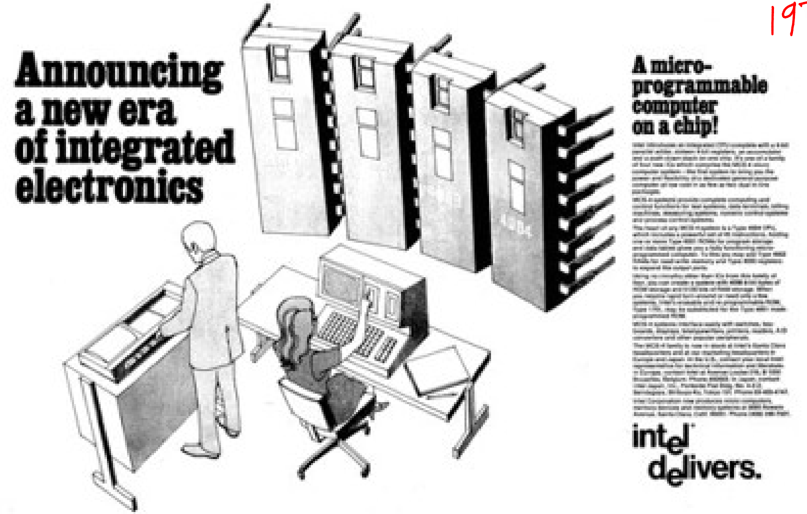
fa_ly_itt size: 117 · 31 µm (1.2µCMOS)



From http://book.huihoo.com/design-of-vlsi-systems/

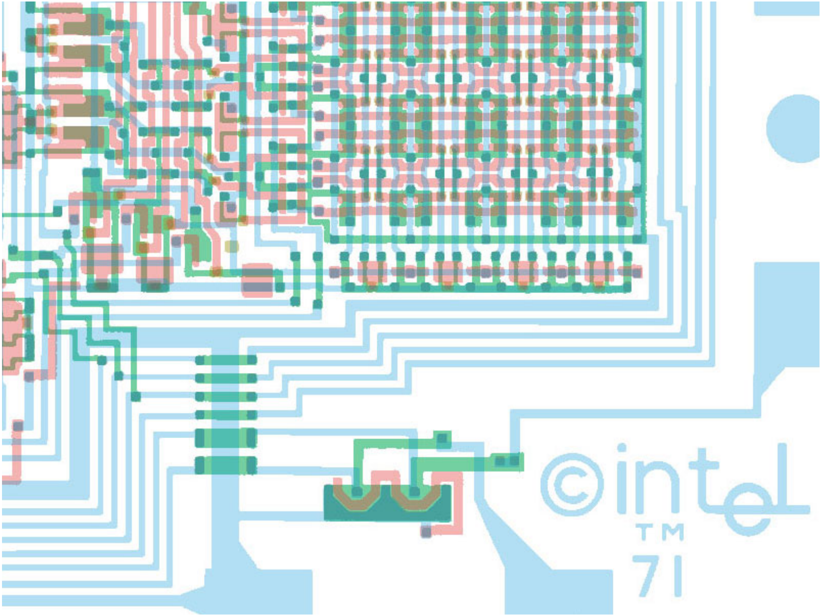
Intel 4004: The First Single-Chip Microprocessor

1971



4001: 256-byte ROM + 4-bit IO port
4002: 40-byte RAM
4003: 10-bit shift register
4004: 740 kHz 4-bit CPU w/ 45 instructions (2300 transistors)

Intel 4004 Masks



Intel 4004 Die Photograph