

Chapter 1: MOS Devices and Circuits

Copyright © 1978, C.Mead, L.Conway

Sections:

The MOS Transistor - - - The Basic Inverter - - - Inverter Delay - - - Parasitic Effects - - - Driving Large Capacitive Loads - - - Space vs Time - - - Basic NAND and NOR Logic Circuits - - - Super Buffers - - - A Closer Look at the Electrical Parameters - - - Depletion Mode vs Enhancement Mode Pullups - - - Delays in Another Form of Logic Circuitry - - - Pullup/Pulldown Ratios for Inverting Logic Coupled by Pass Transistors - - - Transit Times and Clock Periods - - - Properties of Cross Coupled Circuits - - - Effects of Scaling Down the Dimensions of MOS Circuits and Systems

In this chapter we begin with a discussion of the basic properties of the n-channel, metal-oxide-semiconductor (MOS), field effect transistor (FET). We then describe and analyze a number of circuits composed of interconnected MOS field effect transistors. The circuits described are typical of those we will commonly use in the design of integrated systems. The analysis, though highly condensed, is conceptually correct and provides a basis for the solution of most system problems typically encountered.

Integrated systems in MOS technology contain three levels of conducting material separated by intervening layers of insulating material. Proceeding from top to bottom, these levels are termed the *metal*, the *polysilicon*, and the *diffusion* levels respectively. Patterns for paths on these three levels, and the locations of contact cuts through the insulating material to connect certain points between levels, are transferred into the levels during the fabrication process from *masks* similar to photographic negatives. The details of the fabrication process will be discussed in chapter 2.

In the absence of contact cuts through the insulating material, paths on the metal level may cross over paths on the polysilicon or diffusion levels with no significant functional effect. However, wherever a path on the polysilicon level crosses a path on the diffusion level, a transistor is created. Such a transistor has the characteristics of a simple switch, with a voltage on the polysilicon level path controlling the flow of current in the diffusion level path. Circuits composed of such transistors, interconnected by patterned paths on the three levels, form our basic building blocks. With these basic circuits, we will architect integrated systems, to be fabricated on the surface of monolithic crystalline chips of silicon.

The MOS Transistor

An MOS transistor will be produced on the integrated system chip wherever a polysilicon path crosses a diffusion path, as shown in figure 1a. The electrical symbol used to represent the MOS transistor in our circuit diagrams is shown in figure 1b, along with symbols and polarities of certain voltages of interest. Note that the source and drain terminals of the device are physically symmetrical. For the n-channel MOSFETs, these terminal labels are assigned such that V_{DS} is normally positive. A more detailed view of the rectangular region called the gate, where the polysilicon (poly) crosses the diffusion, is given in figure 1c. During fabrication the diffusion paths are formed after the poly paths are formed, as explained more fully in chapter 2. The poly gate and thin layer of oxide under masks the region under the gate during diffusion, thus interrupting the diffusion path under the gate. Therefore, there is no direct diffusion connection between the source and drain terminals of the transistor. Notice in this discussion that metal, poly, and diffusion paths all conduct electricity well enough to be considered "wires" until further notice.

In the absence of any charge on the gate, the drain to source path through the transistor is like an open switch. The gate, separated from the substrate by the layer of thin oxide, forms a capacitor. If sufficient positive charge is placed on the gate so that V_{GS} exceeds a *threshold voltage* V_{th} , electrons will be attracted to the region under the gate to form a conducting path between drain and source. Most of the transistors we will use in our systems have threshold voltages greater than zero. These are called *enhancement mode* MOSFETs, and their threshold voltage typically equals $\sim 0.2(V_{DD})$, where V_{DD} is the positive supply voltage for the particular technology.

The basic operation performed by the MOS transistor is to use charge on its gate to control the movement of negative charge between source and drain through the channel under the gate. The current from source to drain equals the charge induced in the channel divided by the transit time or average time required for an electron to move from source to drain. The transit time itself is the distance the electron has to move divided by its average velocity. In semiconductors under normal conditions, the velocity is proportional to the electric field driving the electrons. The relationship between drain to source current I_{DS} , drain to source voltage V_{DS} , and gate to source voltage V_{GS} is sketched in figure 1d. For small V_{DS} , the transit time τ is given by equation 1.

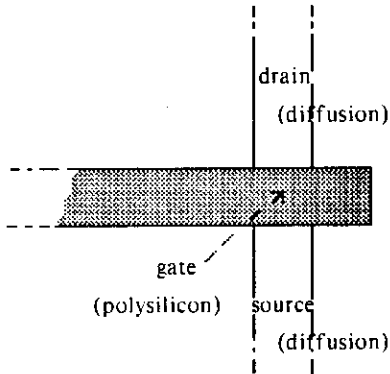


Fig. 1a. MOS Transistor (top view)

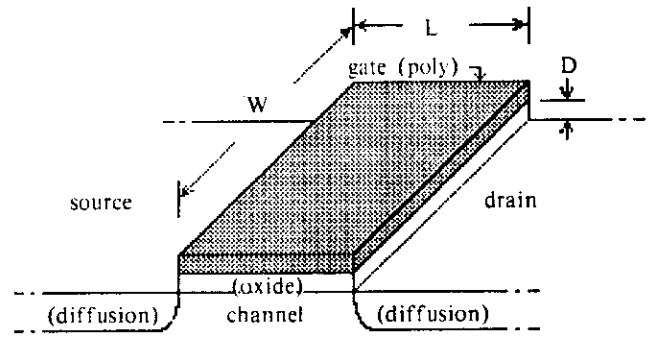


Fig. 1c. MOSFET Gate Dimensions

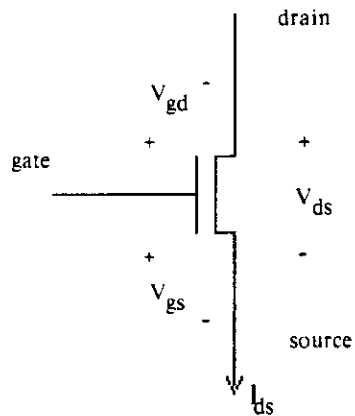


Fig. 1b. MOS Transistor Symbol (subscripts in + to - direction sequence)

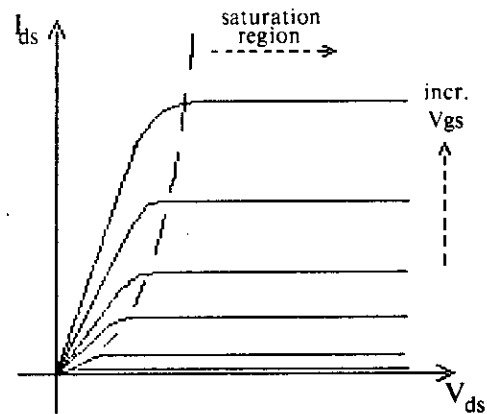


Fig. 1d. Current vs Voltage

Transit time:

$$\tau = L/\text{velocity} = L/\mu E = L^2/[\mu V_{ds}] \quad [\text{eq.1}]$$

The proportionality constant μ is called the *mobility* of the charge carriers, in this case electrons, under the influence of an electric field in the conducting material of the channel region. It is a velocity per unit electric field ($\text{cm}^2/\text{volt-sec}$). We shall see that the transit time is the fundamental time unit of the entire integrated system.

The amount of negative charge in transit is just the gate capacitance times the voltage on the gate in excess of the threshold voltage. The capacitance of two parallel conductors of area A , separated by insulating material of thickness D , equals $\epsilon A/D$. The proportionality constant ϵ is called the permittivity of the insulating material, and has a simple interpretation. It is the capacitance of parallel conductors of area $A = 1 \text{ cm}^2$, separated by a thickness $D = 1 \text{ cm}$ of the insulator material, and is in the units farad/cm. Therefore, the gate capacitance equals $\epsilon WL/D$. Thus the charge in transit is given by eq. 2, and the current is given by eq. 3.

Charge in transit:

$$Q = - C_g(V_{gs} - V_{th}) = - \frac{\epsilon WL(V_{gs} - V_{th})}{D} \quad [\text{eq.2}]$$

Current:

$$I_{ds} = - I_{sd} = - \frac{\text{charge in transit}}{\text{transit time}} = \frac{\mu \epsilon W(V_{gs} - V_{th})(V_{ds})}{LD} \quad [\text{eq.3}]$$

Note that for small V_{ds} , the drain current is proportional to the source-drain voltage and also to the gate voltage above threshold. Any device with a current through it proportional to the voltage across it, may be viewed as a resistor, and in the case of an MOS device with *low* drain to source voltage, the resistance is controlled by the gate voltage as given in eq. 3a.

$$V_{ds}/I_{ds} = R = L^2/[\mu C_g(V_{gs} - V_{th})] \quad [\text{eq.3a}]$$

In both equations 2 and 3a, C_g is the gate to channel capacitance of the turned on transistor. In the simple case where this transistor is driving the gate of another one identical to it, the time response of the system will be an exponential with a time constant RC_g , given in equation 4. This time constant is identical to the transit time τ given in equation 1.