Chapter

5

Resistors, Capacitors, MOSFETs

This chapter provides more information and examples related to the layout of resistors, capacitors, and MOSFETs. Layout using the poly2 layer and how poly2 is used to make poly-poly capacitors will be covered. We'll also introduce some fundamental layout techniques including using unit cells, layout for matching, and the layout of long length and wide MOSFETs. The temperature and voltage dependence of resistors and capacitors will also be covered.

5.1 Resistors

The resistors and capacitors in a CMOS process have values that change with voltage and temperature. The change is usually listed as ppm/°C (parts per million per degree C). The ppm/°C is equivalent to a multiplier of 10^{-6}/°C.

Temperature Coefficient (Temp Co)

As temperature goes up, so does the value of a resistor, Fig. 5.1 (in general). Generally, the value of a resistor, $R(T_0)$, is specified at room temperature, T_0. To characterize the resistor we use

$$R(T) = R(T_0) \cdot (1 + TCR1 \cdot (T - T_0)) \qquad (5.1)$$

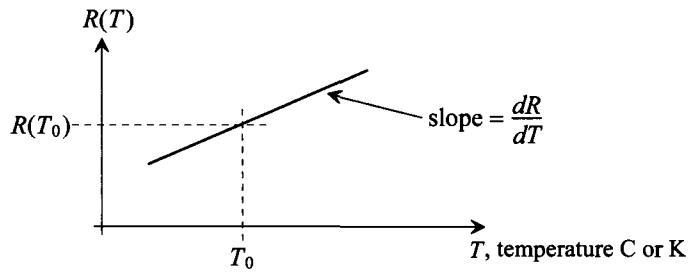


Figure 5.1 Resistor's value change with temperature.

where T is the actual temperature of the resistor. Because of the difference in the equation, it doesn't matter if Celsius or Kelvin are used for the units of T as long as the units match the units used for T_0. The first-order temperature coefficient of a resistor, $TCR1$, is given by

$$TCR1 = \frac{1}{R} \cdot \frac{dR}{dT} \tag{5.2}$$

Notice that the $TCR1$ changes with temperature. For most practical applications, the "temp co" is assumed to be linear. Typical values of $TCR1$ for different layers in the CMOS process are given in Table 4.1. SPICE uses a quadratic, in addition to this first order term, to model the temperature dependence of a resistor

$$R(T) = R(T_0) \cdot [1 + TCR1 \cdot (T - T_0) + TCR2 \cdot (T - T_0)^2] \tag{5.3}$$

For hand calculations, we assume $TCR2$ is 0.

Example 5.1

Using the values in Table 4.1, estimate the minimum and maximum resistance of an n-well resistor with a length of 100 and a width of 5 over a temperature range of 0 to 100 °C. Verify the hand calculations using SPICE.

The n-well resistor sheet resistance is approximately 500 Ω/square (at 27 °C). The value of the resistor in this example (20 squares) is 10 k at 27 °C. The temp co is 2400 ppm/°C (= 0.0024) The minimum resistance is then determined, using Eq. (5.1) by

$$R_{\min} = 10,000 \cdot [1 + 0.0024 \cdot (0 - 27)] = 9.35\ k\Omega$$

and the maximum resistance is determined by

$$R_{\max} = 10,000 \cdot [1 + 0.0024 \cdot (100 - 27)] = 11.75\ k\Omega$$

The SPICE simulations and netlist are shown in Fig. 5.2. ■

Polarity of the Temp Co

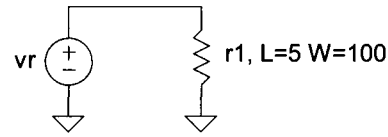
We made the comment that in general the temp co of a resistor is positive, which means that the resistor's value increases with increasing temperature. In other words, the resistivity of the silicon material used to fabricate the resistor increases with increasing temperature. If we were to look at the electron concentration in an n-well as temperature is increased, for example, we would see an exponential increase in the number of thermally generated carriers. More carriers, less resistivity, right? Looking at the carrier concentration alone, we would expect silicon's resistivity to decrease with increasing temperature. With the increase in the carrier concentration, however, we get a reduction in the carrier's mobility. The carrier-to-carrier interactions increase the scattering (the average distance a carrier can travel before colliding with some other carrier decreases with a larger number of free carriers). The material parameter "mobility" characterizes how easily carriers can move through a material. The mobility of an electron or hole, μ_n or μ_p respectively, is a material parameter that relates the applied electric field across the material to the velocity the carriers can drift through the material. High velocity indicates that the carriers have high-mobility and can move through the material quickly. Mobility is a measured parameter and is given by

```
*** Figure 5.2 CMOS: Circuit Design, Layout, and Simulation ***

.control
destroy all
set temp=0
run
set temp=27
run
set temp=100
run
let iref0=-dc1.i(vr)
let iref27=-dc2.i(vr)
let iref100=-dc3.i(vr)
let r0=vr/iref0
let r27=vr/iref27
let r100=vr/iref100
plot r0 r27 r100
.endc

.dc Vr 0.9 1.1 1m
vr vr 0 DC 0

r1 vr 0 10k rmod L=5 W=100
.model rmod R RSH=500 TNOM=27 TC1=0.0024
.end
```



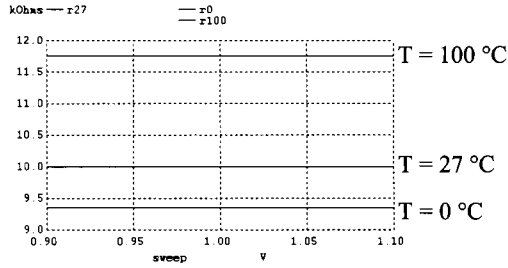


Figure 5.2 Simulating the temperature dependence of an n-well resistor.

$$\mu_{n,p} = \frac{\text{Average velocity of carriers, } cm/s}{\text{Applied electric field, } V/cm} \tag{5.4}$$

noting mobility's units are $cm^2/V \cdot s$. The resistivity of the material depends on the number of free carriers (electrons/cm^3, n, and holes/cm^3, p) or

$$\rho = \frac{1}{q(\mu_n n + \mu_p p)} \tag{5.5}$$

where q is the electron charge (see Eqs. [2.2] and [2.3]). This equation is important and shows why we can have a negative or positive resistor temperature coefficient. If the mobilities decrease faster than the carrier concentrations increase, we get a positive temp co. The resistor's value goes up with increasing temperature because the mobility is dropping faster than the carrier concentration is increasing. However, if the increase in carrier concentration with temperature is faster, the resistivity goes down with increasing temperature and we get a negative temp co (the resistor's value goes down with increasing temperature). At room temperature, a positive temperature coefficient is normally observed.

Voltage Coefficient

Another important contributor to a changing resistance is the voltage coefficient of the resistor given by

$$VCR = \frac{1}{R} \cdot \frac{dR}{dV} \tag{5.6}$$

where V is the average voltage applied to the resistor, that is, the sum of the voltages on each end of the resistor divided by two. The resistance as a function of voltage is then given by

$$R(V) = R(V_0) \cdot (1 + VCR1 \cdot (V - V_0) + VCR2 \cdot (V - V_0)^2) \qquad (5.7)$$

The value $R(V_0)$ is the value of the resistor at the voltage V_0. A typical value of $VCR1$ is 8000 ppm/V for an n-well resistor. The main contributor to the voltage coefficient is the depletion layer width between the n-well and the p-substrate. The depletion layer extends into the n-well, resulting in an effective change in the sheet resistance. The thickness of the n-well available to conduct current decreases with increasing potential (reverse bias) between the n-well and the substrate.

Example 5.2

Estimate the average resistance of an n-well resistor with a typical value of 10k at 27 °C, for an average voltage across the resistor of 0, 5, and 10 V.

$$R(0) = 10,000 \cdot (1 + 0.008 \cdot 0) = 10\ k\Omega$$

$$R(5) = 10,000 \cdot (1 + 0.008 \cdot 5) = 10.4\ k\Omega$$

$$R(10) = 10,000 \cdot (1 + 0.008 \cdot 10) = 10.8\ k\Omega$$

This is a small change compared to the change due to temperature. However, as the next example will show, the change due to the applied voltage can have a greater influence on the circuit performance than the temperature. ■

Example 5.3

Compare the change in V_{out} in the following circuit due to the VCR with the change due to the TCR.

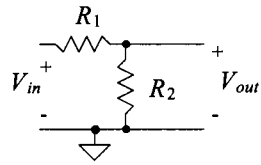


Figure 5.3 Comparing a resistive divider's temperature performance to its voltage performance.

We start by writing the voltage divider equation and then substitute the temperature or voltage dependence. For the temperature dependence

$$V_{out} = V_{in} \cdot \frac{R2(T)}{R1(T) + R2(T)} = V_{in} \cdot \frac{R2 \cdot [1 + TCR(T - T_0)]}{(R1 + R2) \cdot [1 + TCR(T - T_0)]} = V_{in} \cdot \left[\frac{R2}{R1 + R2}\right]$$

showing that the divider is independent of temperature. For the voltage dependence

$$V_{out} = V_{in} \cdot \frac{R2(V)}{R1(V) + R2(V)} = V_{in} \cdot \frac{R2 \cdot (1 + VCR \cdot V_2)}{R1 \cdot (1 + VCR \cdot V_1) + R2 \cdot (1 + VCR \cdot V_2)}$$

where $V_1 = \frac{Vin + Vout}{2}$ and $V_2 = \frac{Vout}{2}$. These results show that the voltage divider has no temperature dependence but that it does have a voltage dependence. ■

Using Unit Elements

The preceding example shows the advantage of ratioing components. For precision design over large temperature ranges, this fact is very important. Usually, a *unit resistor* is laid out with some nominal resistance. Figure 5.4a shows a unit cell resistor layout, say 5 kΩ, using the n-well. Figure 5.4b shows how the nominal 5 kΩ unit cell resistor implements a 4/5 voltage divider. Using the divider in Fig. 5.3 as a reference, the resistor R_1 is 5k and the resistor R_2 is 20k (notice the layout of the 20k resistor is 4 unit cells). The errors due to corners and differing perimeters using a serpentine pattern, see Fig. 2.28, are eliminated when using unit cell layout techniques. Also, the exact calculation of the resistor's value isn't important. Changes in the nominal resistance value because of process shifts, temperature, or how the number of squares is calculated only affect the current flowing in the divider (the power dissipation) and not the output voltage. Amplifier circuits, such as those using op-amps with feedback, are examples of circuits where ratioing is important.

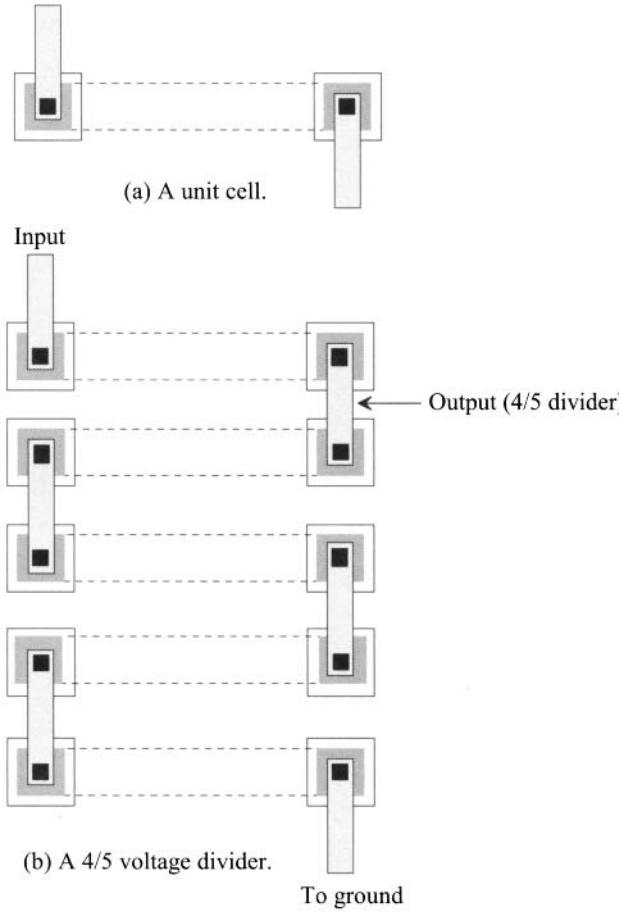


Figure 5.4 (a) Layout of a unit resistor cell, and (b) layout of a divider.

Guard Rings

Any precision circuit is susceptible to substrate noise. Substrate noise results from adjacent circuits injecting current into one another. The simplest method of reducing substrate noise between adjacent circuits is to place a p+ implant (a substrate contact for a p-substrate wafer) between the two circuits. The substrate contact removes the injected carriers and holds the substrate, ideally, at a fixed potential (ground). Figure 5.5 shows the basic idea for a resistor. The p+ implant guards the circuit against carrier injection. Because the implants are laid out in a ring, they are often called *guard rings*.

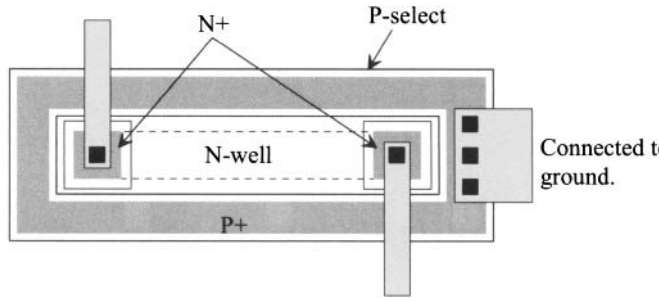


Figure 5.5 Guard ringing an n-well resistor.

Interdigitated Layout

Matching between two different resistors can be improved by using the layout shown in Fig. 5.6. These resistors are said to be interdigitated. Process gradients, in this case changes in the n-well, n+, or p+ doping at different places on the die, are spread between the two resistors more evenly. Notice that the orientation of the resistors is consistent between unit cells; that is, all cells are laid out (here) vertically. Also, each resistor has essentially the same parasitics.

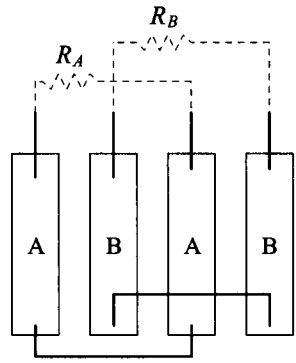
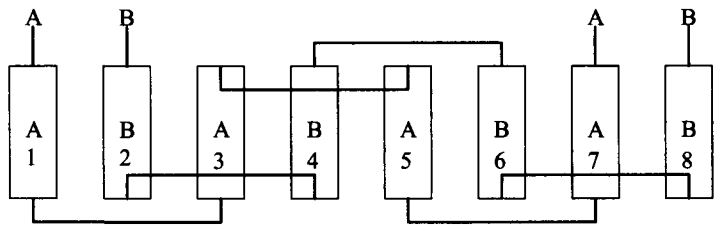


Figure 5.6 Layout of interdigitated resistors.

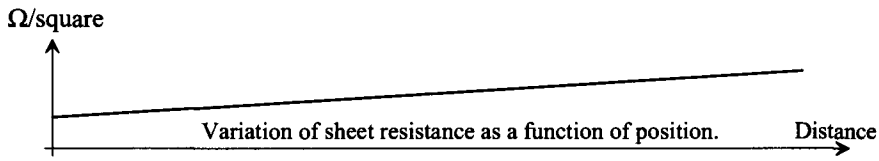
Common-Centroid Layout

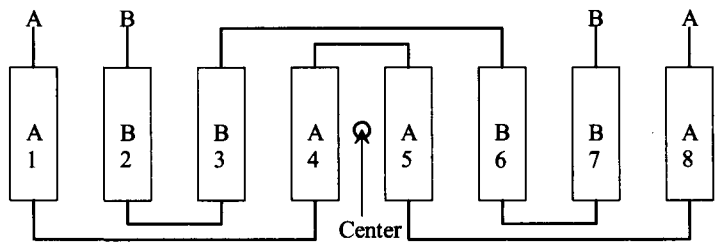
Common-centroid (common center) layout helps improve the matching between two resistors (at the cost of uneven parasitics between the two elements). Consider the arrangement of unit resistors shown in Fig. 5.7a. This interdigitated resistor is the same layout style as just discussed in Fig. 5.6. Consider the effects of a linearly varying sheet resistance on the overall value of each resistor. If we assign a normalized value to each unit resistor, as shown, then resistor A has a value of 16 and resistor B has a value of 20. Ideally, the resistor values are equal.

Next consider the common-centroid layout shown in Fig. 5.7b, noting that resistors A and B share a common center. The value of either resistor in this figure is 18. In other words, the use of a common center (ABBAABBA) will give better matching than the interdigitized layout (ABABABAB). Figure 5.8 shows the common-centroid layout (two different possibilities) of four matched resistors. Note that common-centroid layout can also improve matching in MOSFETs or capacitors.



(a) Interdigitated layout





(b) Common-centroid layout

Figure 5.7 (a) Interdigitated layout and (b) common-centroid layout.

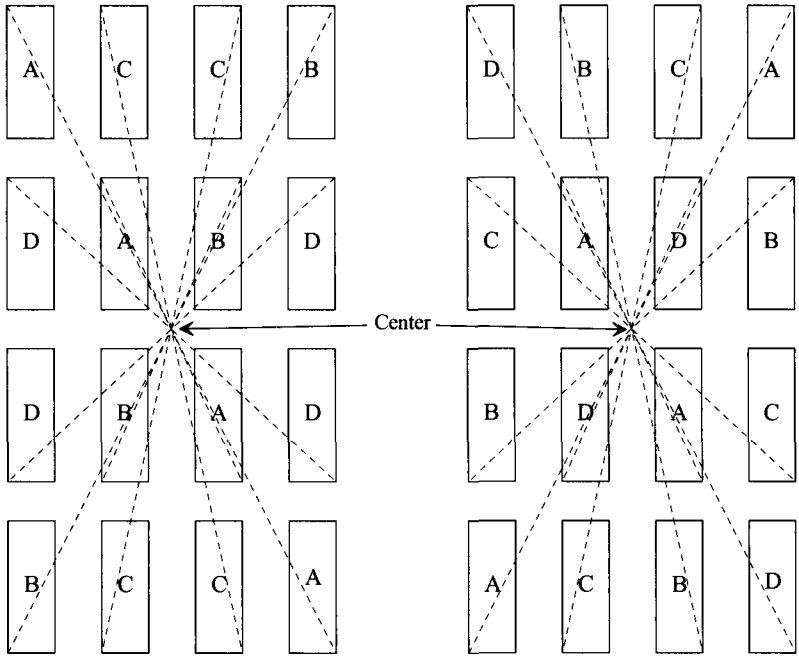


Figure 5.8 Common-centroid layout of four matched resistors (or elements).

Example 5.4

Suppose that the nominal value of the unit resistor, A, on the left side of the layouts in Fig. 5.7 is 5k (a nominal value because we know the actual sheet resistance varies with process shifts and temperature). If the sheet resistance linearly varies across the layout, from the left to the right, and the final resistor's value is 5.07k, compare the interdigitated layout to the common-centroid layout.

The farthest left resistor in the layout has a value of 5k (given). The second resistor's value, because of the linear variation in the sheet resistance from the left to the right in the layout, is 5.01k. The third resistor's value is 5.02k, etc. For the interdigitated layout, the value of resistor A is

$$R_A = 5.0 + 5.02 + 5.04 + 5.06 = 20.12\ k\Omega$$

and the value of resistor B is

$$R_B = 5.01 + 5.03 + 5.05 + 5.07 = 20.16\ k\Omega$$

For the common-centroid layout, resistor A's value is

$$R_A = 5.0 + 5.03 + 5.04 + 5.07 = 20.14\ k\Omega$$

and resistor B's value is

$$R_B = 5.01 + 5.02 + 5.05 + 5.06 = 20.14\ k\Omega$$

exactly the same result. If we had laid out all of the unit elements for resistor A on the left and the all of the unit elements for resistor B on the right, we would get $R_A = 20.06\ k\Omega$ and $R_B = 20.22\ k\Omega$. This shows that the interdigitated layout does help with matching. ■

Dummy Elements

Another technique that improves the matching between two or more elements is the use of dummy elements. Consider the cross-sectional views of the three n-wells shown in Fig. 5.9a. The final amount of diffusion under the resist, on the edges, is different between the outer and the inner unit cells. This is the result of differing dopant concentrations at differing points on the surface (during the diffusion process). This difference results in a mismatch between unit resistors. To compensate for this effect, dummy elements can be added (see Fig. 5.9b) to an interdigitated or common-centroid layout. The dummy element does nothing electrically. It simply ensures that the unit resistors of matched resistors see the same adjacent structures. Normally, these dummy elements are tied off to either ground or *VDD* rather than left floating.

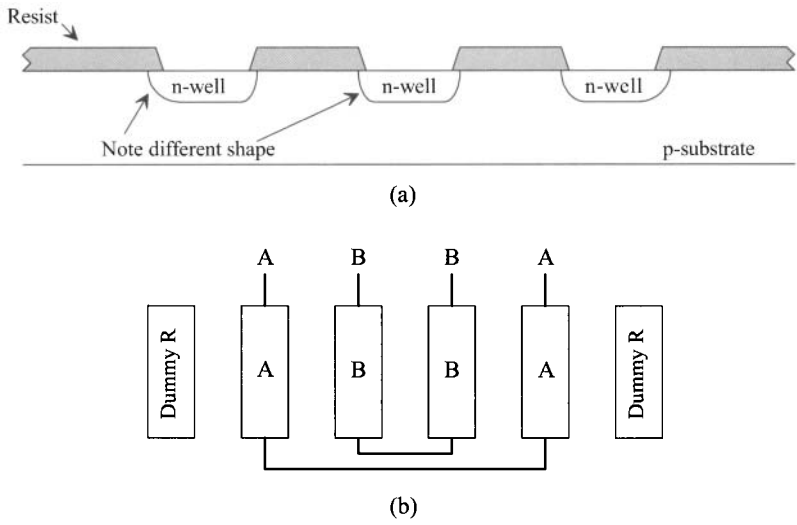


Figure 5.9 (a) Edge effects and (b) a common-centroid layout with dummy R.

5.2 Capacitors

An added layer of polysilicon, called poly2, can be present in a CMOS process for capacitor formation (called a poly-poly capacitor), for MOSFET formation (we can use poly2 instead of poly1 for the gate of a MOSFET) and for creating a floating gate device (see Ch. 16's discussion of Flash memory technology for example). In this section we discuss the layout of capacitors using poly2, the parasitics present, and the temperature behavior of the poly-poly capacitor. Many of the layout techniques discussed in the last section can be used when laying out capacitors.

Layout of the Poly-Poly Capacitor

Layout and cross-sectional views of a capacitor using the poly1 and poly2 layers are shown in Fig. 5.10. The silicon dioxide dielectric between the two layers of poly is roughly the same thickness as the gate oxide (GOX), t_{ox} in a MOSFET (see Fig. 4.3b). Table 5.1 shows **typical values for the t_{ox} that we'll use in the 1 μm and 50 nm processes, referred to as long- and short-channel CMOS processes, in this book.** Also seen in the table is the oxide capacitance per area calculated using

$$C'_{ox} = \frac{\varepsilon_r \cdot \varepsilon_0}{t_{ox}} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{5.8}$$

where $\varepsilon_0 = 8.85 \times 10^{-18}\ F/\mu m = 8.85\ aF/\mu m$ and the relative dielectric constant of SiO_2 is 3.97 ($= \varepsilon_r$). To calculate the value of a capacitor, we look at the area where poly1 and poly2 intersect, A, or

$$C_{ox} = C'_{ox} \cdot A \tag{5.9}$$

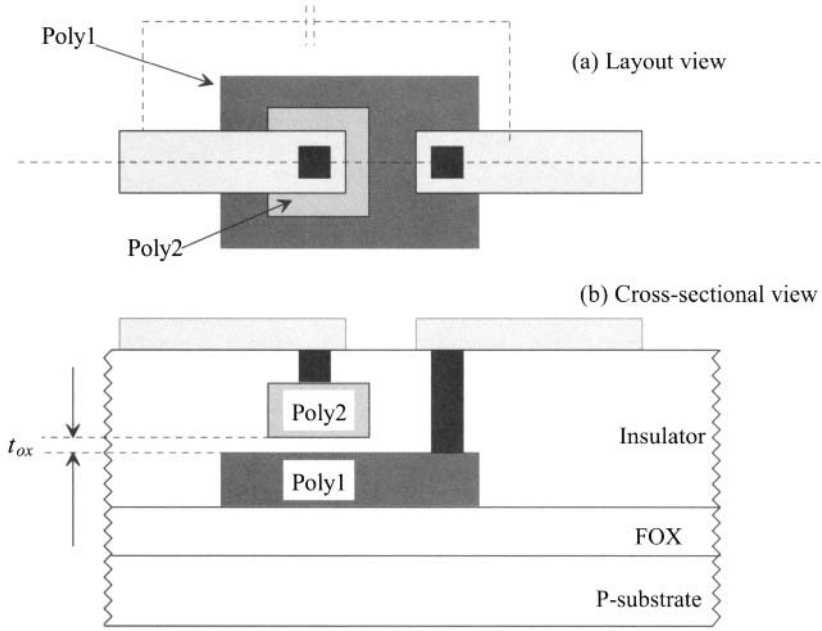


Figure 5.10 Layout and cross-sectional view of a poly-poly capacitor.

If the scale factor is included, then we can write this equation as

$$C_{ox} = C'_{ox} \cdot A \cdot (scale)^2 \tag{5.10}$$

If, in the 50 nm process, a poly-poly capacitor is formed with an intersection of the poly1 and poly2 that measures 10 by 20 then the capacitor's value is

$$C = C_{ox} = 25\ fF/\mu m^2 \cdot 200 \cdot (0.05\ \mu m)^2 = 12.5\ fF \tag{5.11}$$

Table 5.1 Oxide thicknesses and oxide capacitances for the long- and short-channel CMOS processes used in this book.

CMOS technology	Oxide thickness, t_{ox}	C'_{ox}
1 μm (long channel)	200 Å	1.75 $fF/\mu m^2$
50 nm (short channel)	14 Å	25 $fF/\mu m^2$

Note that when the poly2 is used to form a floating gate device (see Ch. 16) poly2 can be called an electrode (see the MOSIS design rules). Also note that the practical minimum capacitor value one should try using (laying out) in the long- and short-channel processes described in Table 5.1 is approximately 100 fF and 10 fF, respectively.

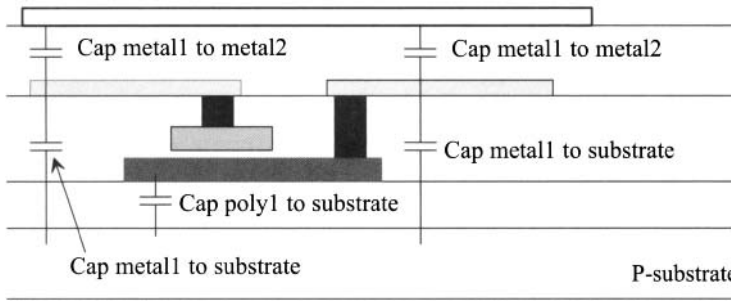


Figure 5.11 The parasitic capacitances of the poly-poly capacitor.

Parasitics

As with any layout structure, we must be concerned with the parasitics. Figure 5.11 shows the parasitics associated with the poly-poly capacitor. The most important (largest) parasitic is the capacitance from poly1 to substrate. This capacitance is called the **bottom plate parasitic** capacitance. Reviewing Table 3.1, this parasitic (plate) capacitance is 58 $aF/\mu m^2$ (there are fringe capacitances as well that can be considered). Keeping in mind that the poly1 area is larger than the poly2 layout area (and the intersection of poly1 and poly2 is the desired capacitance), the bottom plate capacitance can be 20% (or more) of the desired capacitance value. In analog circuits the bottom plate of a capacitor is indicated, as seen in Fig. 5.12, see Ch. 25.

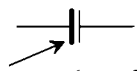


Figure 5.12 The bottom plate parasitic of a capacitor.

Temperature Coefficient (Temp Co)

The first-order temperature coefficient of a capacitor, *TCC*, is given by

$$TCC = \frac{1}{C} \cdot \frac{dC}{dT} \tag{5.12}$$

A typical value of *TCC* for a poly-poly capacitor is 20 ppm/°C. Matching of large area poly-poly capacitors on a die is typically better than 0.1% with good layout techniques. The capacitance as a function of temperature is given by

$$C(T) = C(T_0) \cdot [1 + TCC \cdot (T - T_0)] \tag{5.13}$$

where $C(T_0)$ is the capacitance at T_0.

Voltage Coefficient

The voltage coefficient of a capacitor is given by

$$VCC = \frac{1}{C} \cdot \frac{dC}{dV} \tag{5.14}$$

The voltage coefficient of the poly-poly capacitor is in the neighborhood of 10 ppm/V (for a long-channel process). The capacitance as a function of voltage is given by

$$C(V) = C(V_0) \cdot (1 + VCC \cdot V) \tag{5.15}$$

where $C(V_0)$ is the capacitance between the two poly layers with zero applied voltage, and V is the voltage between the two plates.

5.3 MOSFETs

We introduced the layout of MOSFETs in the last chapter. In this section we discuss some electrical parameters of interest and how to lay out MOSFETs to minimize parasitics or with a long length or width.

Lateral Diffusion

When the drain and source regions are implanted, some of the implant dose laterally diffuses out underneath the gate poly, as depicted in Fig. 5.13. If the drawn length is called L_{drawn}, we can write the effective length as

$$L_{effective} = L_{drawn} - 2L_{diff} \tag{5.16}$$

where L_{diff} is the length of the lateral diffusion. To minimize the lateral diffusion, a spacer is normally deposited adjacent to the poly after a light implant (see Figs. 4.7f and g) and before the heavy source/drain implants. This type of structure is called a lightly doped drain (LDD).

Oxide Encroachment

There are also imperfections associated with the width of the MOSFET, Fig. 5.13. When the active area is defined, Fig. 4.3, the FOX won't be precisely patterned as specified by the active mask layer. The oxide may encroach on the active area (called oxide encroachment) and reduce the active opening area. The drawn width, W_{drawn}, of the MOSFET will be different from the effective width, $W_{effective}$ by $2W_{enc}$. To compensate for oxide encroachment, the layout may be bloated before making the active mask.

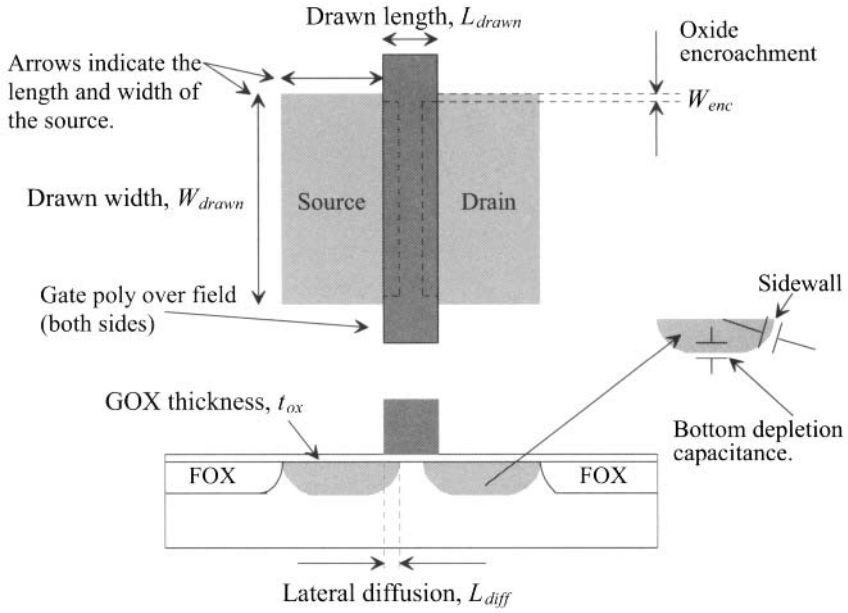


Figure 5.13 Lateral diffusion in a MOSFET.

Source/Drain Depletion Capacitance

As we saw in Ex. 2.3, a pn junction fabricated in the bulk has a depletion capacitance consisting of a bottom component and a sidewall component. For the source and drain junction (depletion) capacitances, this can be written in terms of SPICE parameters as

$$C_{js,d} = \frac{cj \cdot A_{S,D} \cdot (scale)^2}{\left(1 + \frac{V_{S,DB}}{pb}\right)^{mj}} + \frac{cjsw \cdot P_{S,D} \cdot (scale)}{\left(1 + \frac{V_{S,DB}}{pbsw}\right)^{mjsw}} \tag{5.17}$$

where *cj* is the zero-bias bottom depletion capacitance, $A_{S,D}$ is the area of the source or drain implant, *scale* is the scale factor (1 μm for the long-channel process and 50 nm for the short-channel process), *cjsw* is the zero-bias depletion capacitance for the sidewalls, *pb* and *pbsw* are the built-in potentials for the bottom and sidewall components, respectively, *mj* and *mjsw* are the grading coefficients for the bottom and sidewall components, respectively, and finally $V_{S,DB}$ is the potential from the source or drain to the MOSFET's body (the substrate for the NMOS and the well for the PMOS).

It's very common to call the source/drain depletion capacitance (incorrectly) *diffusion capacitance*. As discussed in Sec. 2.4.3, a pn junction only exhibits diffusion capacitance when it becomes forward biased. When CMOS technology was first introduced, the source/drain regions were formed using a diffusion process step (perhaps the reason for the incorrect labeling). CMOS technology developed from, approximately, the mid-80s has used implantation to form the source/drain regions. In any case, this diode junction capacitance should be called either a "junction capacitance" or "depletion capacitance," unless the diode is forward biased.

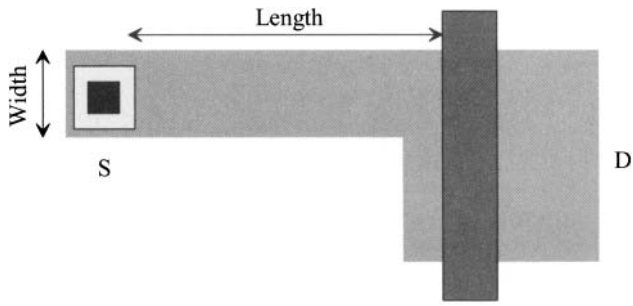


Figure 5.14 Determining a resistance in series with the drain of the MOSFET.

Source/Drain Parasitic Resistance

Examine the layout seen in Fig. 5.14. The active area, in this layout, is used to move the contact to the MOSFET's source or drain away from the gate poly. This can be useful if metal needs to be run over the area directly next to the gate poly. The length of active adds a parasitic resistance in series with the MOSFET. This resistance is determined by specifying the number of squares in the source/drain (NRD/NRS). An estimate for the number of squares, assuming that the extension is on what we label the source of the MOSFET, for the layout in Fig. 5.14, is

$$NRS = \frac{Length}{Width} \tag{5.18}$$

For the other side of the MOSFET (here we call this side the drain), the length is less than the width so we can set *NRD* to zero (or not specify it). To calculate the resistance in series with a MOSFET's source or drain, Fig. 5.15, we use

$$R_S = NRS \cdot R_{sh} \text{ or } R_D = NRD \cdot R_{sh} \tag{5.19}$$

In a SPICE model the paramter rsh is used to specify sheet resistance of the n+ (for the NMOS model) or p+ (for the PMOS model).

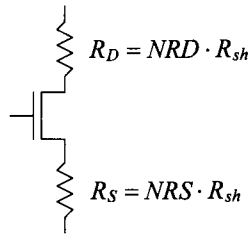


Figure 5.15 How source/drain series parasitic resistance is modeled in SPICE.

Example 5.5

For the MOSFET layout seen in Fig. 5.16, write the SPICE statement including the areas and number of squares in the source/drain regions with or without a scale factor of 50 nm.

Estimates for the areas of the drain/source are

$$A_D = 40 \text{ and } A_S = 45$$

with units of area squared. Note how we didn't include the small area directly adjacent to the gate on the source side. Again, we are estimating (or, more appropriately, approximating) the area. The actual values of capacitance (or resistance) vary with process shifts.

The perimeters of the drain/source are estimated as

$$P_D = 28 \text{ and } P_S = 36$$

The length of the MOSFET is 1 and the width is 10, or,

$$L = 1 \text{ and } W = 10$$

The number of squares of n+ in series with the drain/source is

$$NRD = \frac{4}{10} \rightarrow 0 \text{ and } NRS = \frac{11}{3} \approx 4$$

The SPICE statement for the MOSFET is

```
M1  D G S B NMOS L=1 W=10 AD=40 AS=45 PD=28 PS=36 NRD=0 NRS=4
```

The MOSFET device name always starts with an M (a voltage source device name always starts with a V, a resistor R, etc.) The drain, gate, source, and bulk nodes (in this statement) are labeled D, G, S, B, respectively. Note that in an n-well process the bulk is always tied to ground (which is universally zero, 0, in SPICE). The MOSFET's model name is NMOS (same as the technology to make

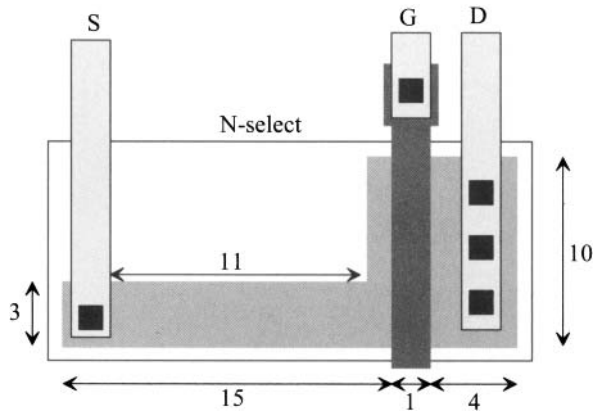


Figure 5.16 Layout of the MOSFET used in Ex. 5.5.

it easier to remember). The statement above is used if we employ a scale factor in the layouts and simulations. Using this MOSFET specification, we would also have to include, in the SPICE netlist,

```
.options scale=50n
```

If we didn't include this statement, then the MOSFET specification would be

```
M1  D G S B NMOS L=50n W=500n AD=100f AS=112.5f PD=1.4u
+ PS=1.8u NRD=0 NRS=4
```

noting that the "+" symbol in the first column of a line indicates that the previous line is continued on the line. **We use drawn sizes in the layout and circuits in this book.** If a SPICE netlist using MOSFETs with drawn sizes doesn't include the .options statement with the scale parameter, then the simulation output is likely flawed. ■

Layout of Long-Length MOSFETs

Figure 5.17 shows the layout of a MOSFET with a long length. The active layer is "snaked" back-and-forth under the poly. Each side of this active is contacted to metal1 for the source and drain connections. The width of the MOSFET is equal to the width of the active under the poly. The length of the MOSFET is estimated by looking at the length of the active underneath the poly between the drain and source. In other words, we know that the drain current flows from the drain to the source. Further we know that the drain current flows in the active area. The poly controls the drain current but is isolated from it by the gate oxide. The length is determined by the intersection of the poly and active between the source and drain contacts (the length of the dotted arrows in the figure). Long-length MOSFETs have, as we'll see, a higher effective switching resistance.

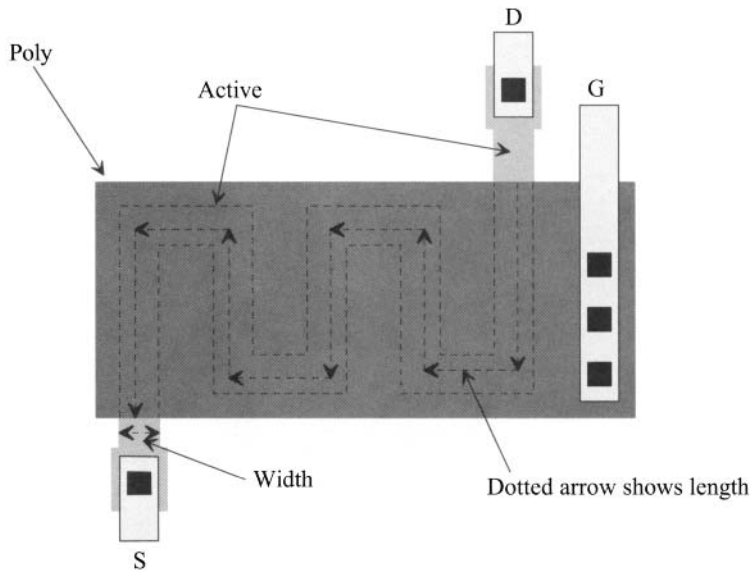


Figure 5.17 Layout of a long-length MOSFET.

Layout of Large-Width MOSFETs

Figure 5.18 shows the layout of a large MOSFET. The length of the MOSFET, L, is still set by the length of poly, as seen in the figure. The MOSFET's overall width is set by the width of poly over active, W, times the number of poly "fingers"

$$\text{Width of MOSFET} = (\text{number of fingers}) \cdot W \tag{5.20}$$

This layout minimizes area by sharing drain and source connections between MOSFETs. Notice how the widths of MOSFET's laid out in parallel (with the gates tied together) add to form an equivalent (to the sum) width MOSFET. This concept can also be used for MOSFET's laid out in series, Fig. 5.19. MOSFETs laid out in series (with their gates tied together) form a MOSFET with an effective length equal to the sum of the individual MOSFET's lengths. We use these concepts often when doing design. For example, we can lay out a MOSFET with a

$$\text{Width-to-length ratio} = W/L = 10/2 \tag{5.21}$$

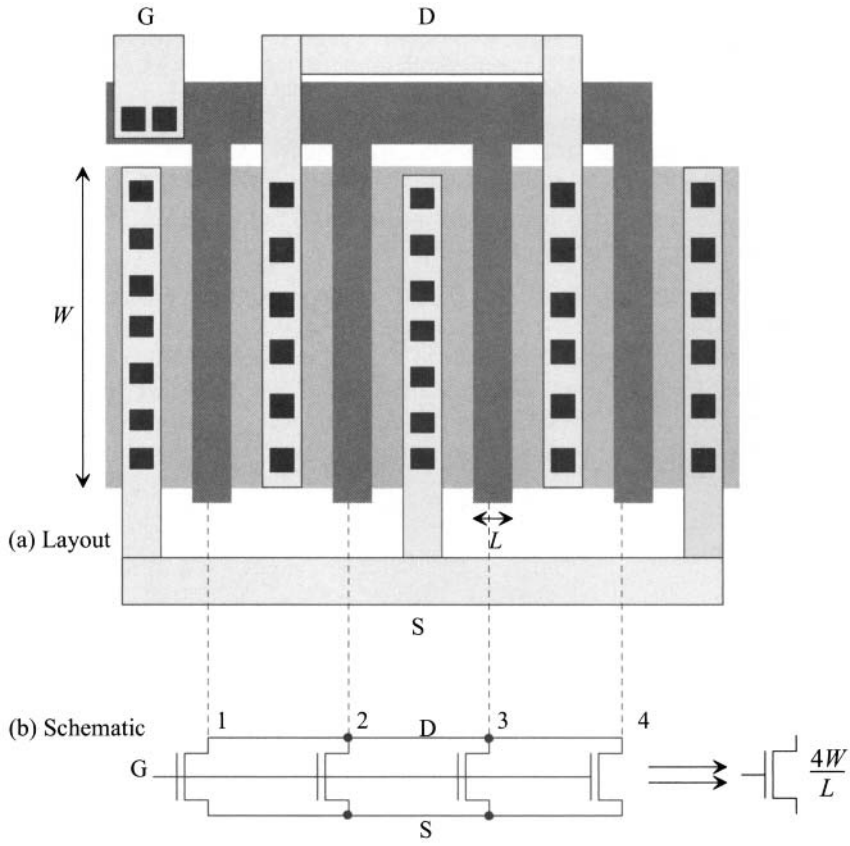


Figure 5.18 Layout and equivalent schematic of a large-width MOSFET.

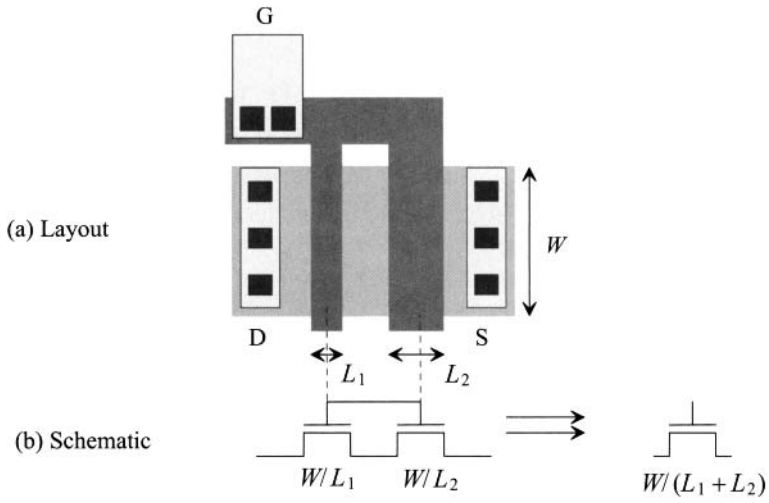
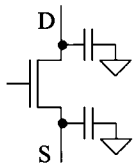


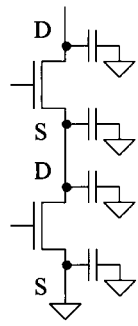
Figure 5.19 MOSFETs in series with gates tied together behave as a single MOSFET with the sum of the lengths.

as a MOSFET with a width of 10 and a length of 2 or as two MOSFETs in series with lengths of 1 and widths of 10. **Notice** that when we write, for a MOSFET size, 10/2, $\frac{15}{3}$, 25/10, or 10/100 the first or top number always specifies the width of the MOSFET, while the second number specifies the MOSFET's length.

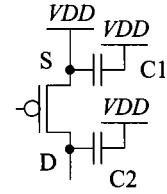
Notice how, when there is an even number of fingers in a large width MOSFET (Fig. 5.18), the area of the active is larger on one side of the poly (on one side of the MOSFET) than on the other side. This leads to a larger parasitic depletion capacitance. Consider the MOSFET schematics in Fig. 5.20. In (a) of this figure, we've drawn the parasitic capacitance on the MOSFET symbol. Generally, for high-speed design, we want to place the MOSFET terminal with the larger parasitic capacitance closest to ground (for



(a) Showing MOSFET drain/source depletion capacitance.



(b) Series connection of MOSFETs.



(c) PMOS device showing depletion capacitances.

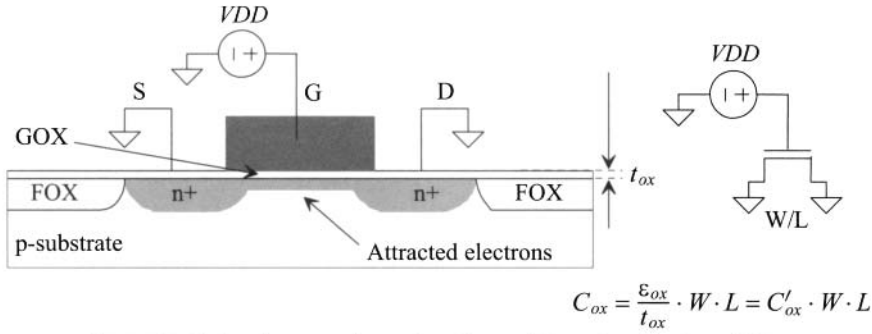
Figure 5.20 Showing depletion capacitance on the MOSFET symbols.

the NMOS) or *VDD* for the PMOS. To understand this, consider the MOSFETs seen in Fig. 5.20b. Thinking of the two NMOS devices as switches, we see that when both switches turn on, the top parasitic capacitance is discharged through both switches. The middle two parasitic capacitors discharge through one switch and the bottom parasitic doesn't charge or discharge (both sides are always tied to ground). The smallest parasitic capacitance should be at the top of the switches because it has the highest resistance discharge path (in this example through two MOSFETs). For the PMOS device in Fig. 5.20c, we would want the larger of C1 or C2 to be called the source terminal and connected to *VDD*.

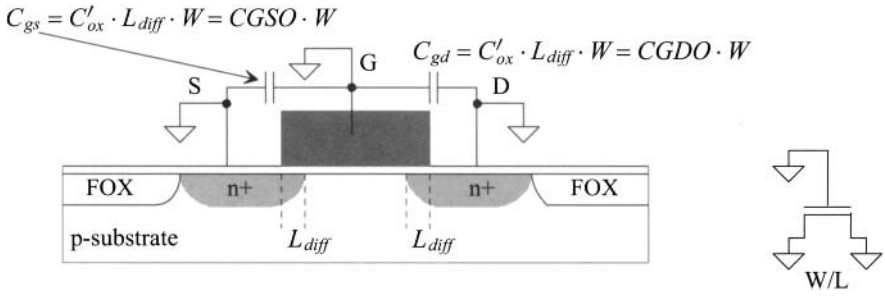
A Qualitative Description of MOSFET Capacitances

Consider the NMOS device in Fig. 5.21a. Here we apply a voltage to the gate of the NMOS device (the most positive voltage in the circuit, the power supply voltage, *VDD*) while holding the source and drain at the same potential as the substrate (ground). The application of this voltage attracts electrons under the gate oxide. This creates a continuous channel of electrons, effectively shorting the drain/source implants. The capacitance from the gate terminal to ground is calculated as

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \cdot W \cdot L = C'_{ox} \cdot W \cdot L \tag{5.22}$$



(a) NMOS device in strong inversion (channel formed under the oxide).



(b) NMOS device in depletion (no channel formed under the oxide).

Figure 5.21 Qualitative description of MOSFET capacitances.

The device is said to be operating in the strong inversion region. The surface under the GOX is p-substrate. When we apply a large positive potential to the gate, we change this material from p-type to n-type (we invert the surface). Note that the oxide capacitance, C_{ox}, does not depend on the extent of the lateral diffusion.

Next examine the configuration in Fig. 5.21b. In this figure all of the MOSFET terminals are grounded. No channel is formed under the GOX. Because of contact potentials, discussed in the next chapter, the area under the GOX is depleted of free carriers. Under these conditions, the MOSFET is operating in the depletion region. The source and drain are not connected as they were in Fig. 5.21a. The capacitance from the gate to the source (or drain) depends on the lateral diffusion and is given by

$$C_{gs} = C'_{ox} \cdot L_{diff} \cdot W = C_{gd} = CGDO \cdot W = CGSO \cdot W \qquad (5.23)$$

The parameter *CGDO* (or *CGSO)* is SPICE parameter called the gate-drain (gate-source) overlap capacitance and is given by

$$CGDO = CGSO = C'_{ox} \cdot L_{diff} \qquad (5.24)$$

When the MOSFET is off, as it is in Fig. 5.21b, the overlap of the gate over the source/drain implant region (the overlap capacitance) is an important component of the capacitance at the MOSFET's gate terminal.

Before leaving this section let's show an SEM photo of a cross-sectional view of a MOSFET (actually 3 MOSFETs), Fig. 5.22. For more information about MOSFET formation see Fig. 6.18 and the associated discussion.

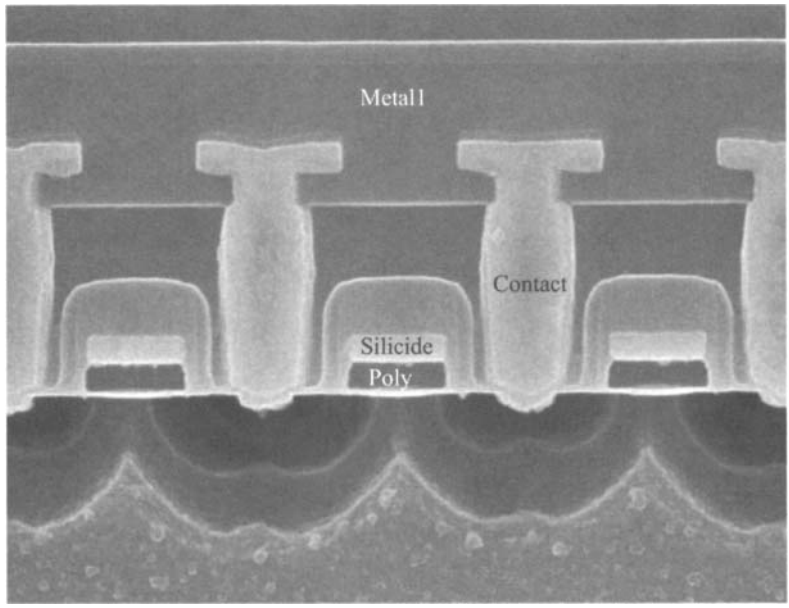


Figure 5.22 SEM image of the cross-section of three MOSFETs.

5.4 Layout Examples

In this section we provide some additional layout examples with a focus on implementing capacitors using (only) metal. Metal-only capacitors are important in CMOS processes that only have one layer of poly (common in most digital CMOS processes). We end the chapter with some discussions about laying out polysilicon resistors.

Metal Capacitors

One method of forming capacitors in a single-poly CMOS process uses the metal layers. Consider the cross-sectional view of a parallel plate capacitor shown in Fig. 5.23. If the plate capacitance between the metal1 and metal2 dominates because the metals have a large layout area (that is, the fringe capacitance contribution is small), then the capacitance can be estimated using

$$C_{12} = \text{Area} \cdot (\text{capacitance per area}) \tag{5.25}$$

If the capacitance per area is 50 aF/μm^2, then it would take an area of 100 μm by 200 μm to implement a 1 pF capacitor. While large area is a problem, it isn't the main problem with a metal parallel-plate capacitor. The main problem occurs from the extremely large bottom plate parasitic capacitance, that is, the capacitance from metal1 to substrate. This parasitic capacitance can be anywhere from 80 to 100% of the desired capacitance. Further it usually slows the circuit response and results in a waste of power (see Fig. 25.16 and the associated discussion).

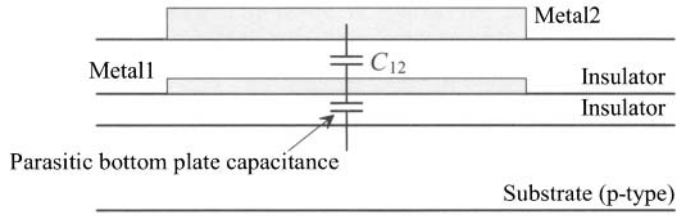


Figure 5.23 Parallel-plate capacitor using metal1 and metal2.

To help decrease the bottom plate's percentage of the desired capacitor value, consider the cross-sectional view shown in Fig. 5.24, where four layers of metal implement a capacitor. The capacitance of this structure can be estimated using

$$C = C_{12} + C_{23} + C_{34} \tag{5.26}$$

If plate capacitance between each metal layer is, again, 50 aF/μm^2, then the area required to implement a 1 pF capacitor is 100 μm by 66 μm. The area needed is reduced by one-third of the area used in the metal1/metal2-only capacitor. While we used the same plate capacitance value in between each level the actual value will vary because of the differing thickness in between the metals. The absolute value of the capacitors, in most circuit design situations (as we'll see later in the book), isn't important but rather the ratio of capacitors is the important parameter (see for example, Eq. [25.19] and the associated discussion). Also notice how the thickness of the metals (made most often now with copper) increases as we move away from the substrate.

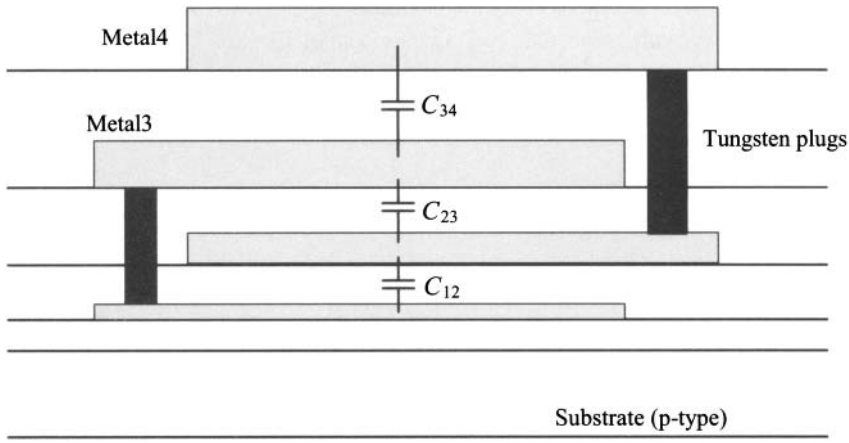


Figure 5.24 Cross-sectional view of a parallel plate capacitor using metal1-metal4.

The value of the capacitors in Figs. 5.23 and 5.24 was set by the areas of the metals and the corresponding plate capacitance. We assumed the perimeter of the metals and the resulting fringe capacitance was a small contributor. Figure 5.25 shows typical minimum sizes and distances between pieces of metal1 where the fringe capacitance dominates. We can make a capacitor using the two pieces of metal1 shown in this figure. A typical value of capacitance per length is 50 aF/μm. The parasitic bottom plate capacitance is half of this value or 25 aF/μm. Since, as seen in the figure, the electric fields can terminate on the close adjacent metal, the bottom component is a smaller percentage than it was when the plate capacitance dominated.

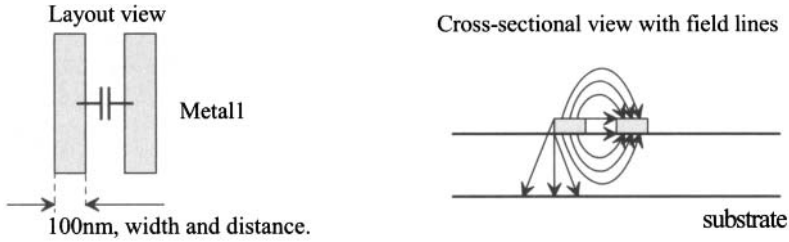


Figure 5.25 Typical size when fringing capacitance dominates.

Next consider the use of metal2 and via1 in the implementation of a capacitor as seen in Fig. 5.26. While the fringe capacitance is still a major component in this capacitor because of the addition of the via between the metals, it is sometimes called a lateral capacitor (there exists a "plate" capacitance between the vias). A typical value of capacitance for this structure is 500 aF/μm. The bottom plate capacitance remains approximately 25 aF/μm. Using additional vias and metal layers will increase the capacitance but generally not linearly. The higher levels of metal, e.g. metal4 or metal5, generally have larger spacing and width design rules than do the lower levels of metal.

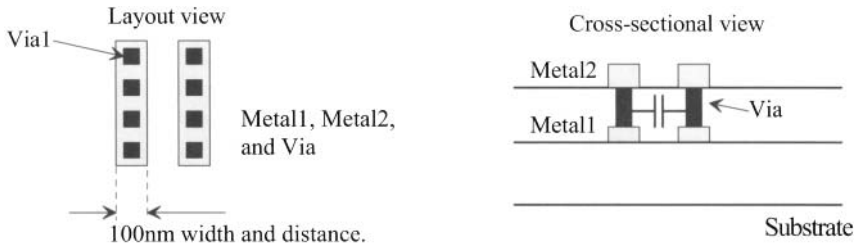


Figure 5.26 Using two layers of metal and the via to implement a lateral capacitor.

It's important to note that while we've concentrated on the bottom plate parasitic, it is also possible to have a top plate parasitic. Often, to avoid coupling noise into the relatively large area occupied by the capacitor, a ground plate is placed above the capacitor. This would allow noisy digital signals to be routed above the capacitor, as seen in Fig. 5.27.

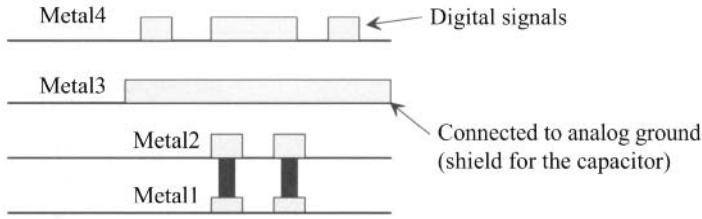


Figure 5.27 Using a metal3 shield to isolate the lateral capacitor.

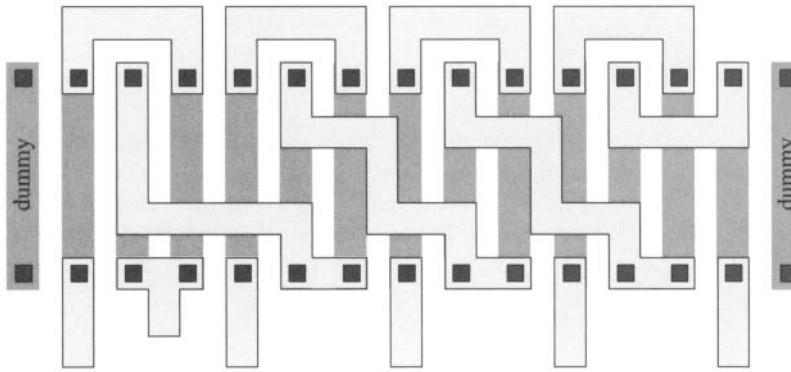
Polysilicon Resistors

In general, polysilicon resistors offer the best performance when precision resistive ratios are required. This is mainly due to the fact that the polysilicon material sits on top of the FOX while the other resistive materials reside in the bulk (and thus form a pn junction that gives rise to a large voltage coefficient.) Since the matching characteristics, temperature behavior, and voltage coefficient are, overall, better for the polysilicon resistors, they are generally preferred in the implementation of precision circuits such as data converters.

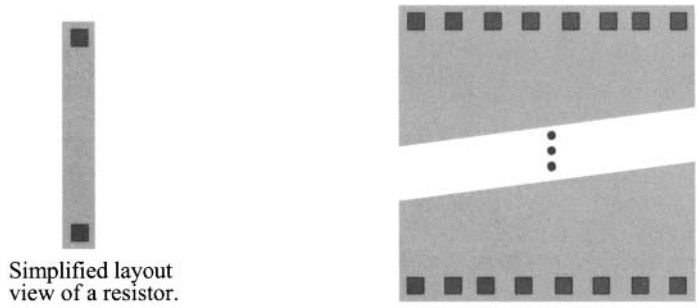
In general, a resistor's width and length should be at least 10 and 100 times the minimum feature size of the process, respectively. For example, if L_{min} is 50 nm, then the minimum width of the resistor should be 500 nm (or wider!). Using larger widths and lengths for the resistors is important both for matching and to ensure that the self-heating, which occurs because of the different current densities flowing in the resistors, doesn't cause any noticeable differences in linearity. In simple terms, the larger resistor area dissipates heat better than the same valued resistor in a smaller area.

Figure 5.28a shows the conceptual layout of an *R*-2*R* resistor string in a minimum area (common resistor topology in an R-2R data converter, see Fig. 29.5.) Figure 5.28b

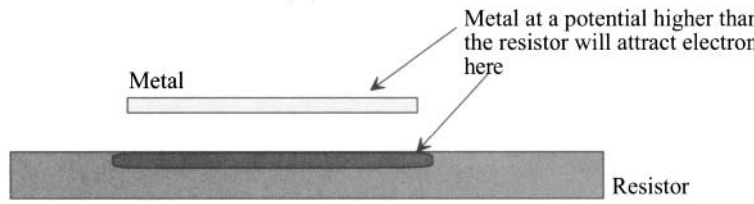
shows the actual layout of the resistors having large width and length along with a large number of contacts to reduce metal/resistive material contact resistance. Figure 5.28c shows the problem of laying out metal over the resistive material, that is, resistor *conductivity modulation*. The figure shows what happens when a metal, having a potential greater than the potential of the underlying resistor is laid out directly over a resistor. Electrons are attracted towards the surface of the resistor causing spots of lower



(a)



(b)



(c) Cross sectional view of metal over resistor.

Figure 5.28 (a) Minimal layout of R-2R string, (b) actual layout of resistor, and (c) conductivity modulation of the resistor value.

resistivity. The solutions to avoiding or reducing conductivity modulation are (1) avoiding running metal over the resistors, (2) using higher levels of metal to route the resistive signals so as to increase the distance between the resistor and the overlaying metal (remembering vias and contacts must be plentiful to avoid adding unwanted series resistance), or (3) inserting a conducting "shield" connected to analog ground and made with metal1 between the resistors and the routing wires above the *R*-2*R* resistor array.

Finally, to conclude this subsection, we ask, "What is the best method of laying out the resistors in an *R*-2*R* string to avoid process gradients and achieve good matching?" While there are no absolute answers, we will discuss a possibility where layout area is a concern. In other words, we won't discuss methods that use a large amount of layout area to average out process variations but will limit our averaging to at most twice the layout area of the *R*-2*R* string shown in Fig. 5.28.

Figure 5.29 shows one possibility for averaging process gradients using a common-centroid configuration with two *R*-2*R* strings connected in series. In this figure we are assuming that the process variations change linearly with position. For example, the first resistor in the string may have an effective value of 1k, while the second's value may be 1.01k, and the third's value is 1.02k, etc. The normalized change in the resistance value is shown in the figure using numbers. However, we could show that the process gradients average out no matter what numbers are used, when using this layout topology, as long as the sheet resistance varies linearly with position. For example, the MSB 2*R* in

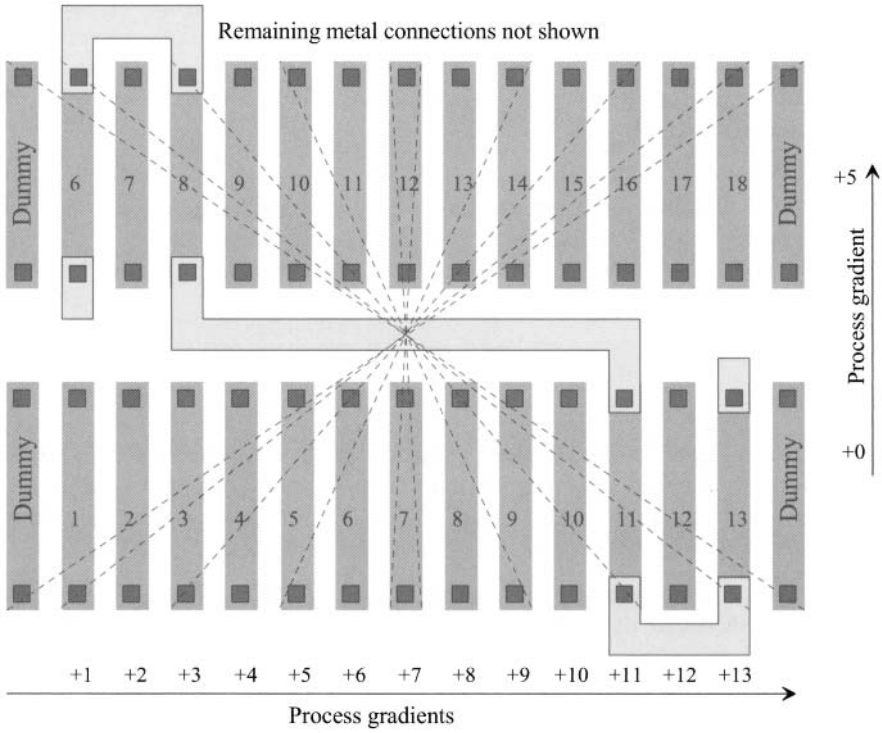


Figure 5.29 Two-string layout for improving matching of R-2Rs. Assumes the resistors are connected together on higher levels of metal to avoid conductivity modulation.

the top string of Fig. 5.29 (on the left) has a value of 14 (6 + 8). The MSB $2R$ in the bottom string (on the right) has a value of 24. Adding the values of the two resistors, by connecting them in series, results in a resistor value of 38 ($2R$ = 38 while R = 19). The middle resistor value in the top string has a value of 12, while the bottom resistor has a value of 7. Again, adding the two resistors results in a value of 19. Fundamentally, the limiting factor in matching then becomes the voltage and temperature (because of the different current densities through the resistors) coefficients of the resistors.

ADDITIONAL READING

[1] R. A. Pease, J. D. Bruce, H. W. Li, and R. J. Baker, "Comments on Analog Layout Using ALAS!" *IEEE Journal of Solid-State Circuits*, vol. 31, no. 9, September 1996, pp. 1364–1365.

[2] D. J. Allstot and W. C. Black, "Technology Design Considerations for Monolithic MOS Switched-Capacitor Filtering Systems," *Proceedings of the IEEE*, vol. 71, no. 8, August 1983, pp. 967–986.

PROBLEMS

5.1 Suppose a current in a circuit is given by

$$I = \frac{V_{REF}}{R}$$

If the voltage, V_{REF}, comes from a precision voltage reference and doesn't change with temperature, determine the temperature coefficient of the current in terms of the resistor's temperature coefficient. If the resistor is fabricated using the n-well plot, similar to Fig. 5.1, the current's change with temperature. Use the TCR1 given in Table 4.1.

5.2 Suppose a silicided n+ resistor with a value of 100 Ω is used. Using the data from Table 4.1, sketch the layout and cross-sectional views of the resistor. The current in the resistor flows mainly in the silicide. Suppose the mobility of free carriers in the silicide is constant with increasing temperature. Would the temperature coefficient of the resistor be positive or negative? Why?

5.3 Using a layout program, make a schematic and layout for the 1/5 voltage divider seen in Fig. 5.4 if the resistor's value is 5k. Use n-well resistors and DRC/LVS the final layout and schematic.

5.4 Using a layout program, make a schematic and layout for an RC lowpass circuit where the resistor's value is 10k and the poly-poly capacitor's value is 100 fF. Use the 50 nm process (see Table 5.1). DRC/LVS the final layout and schematic. Simulate the operation of the circuit with a pulse input (see Fig. 1.27).

5.5 Estimate the areas and perimeters of the source/drain in the layout seen in Fig. 5.18 if the length of the device, L, is 2 and the width of a finger, W, is 20.

5.6 Provide a qualitative discussion for the capacitances of the PMOS device similar to the discussion associated with Fig. 5.21 for the NMOS device. Make sure the descriptions of operation in the strong inversion and depletion regions are clear. Draw the equivalent (to Fig. 5.21) figure for the PMOS devices.