Chapter

3

The Metal Layers

The metal layers in a CMOS integrated circuit connect circuit elements (MOSFETs, capacitors, and resistors). In the following discussion we'll discuss a generic CMOS process with two layers of metal. These levels of metal are named metal1 and metal2. The metal in a CMOS process is either aluminum or copper. In this chapter we look at the layout of the bonding pad, capacitances associated with the metal layers, crosstalk, sheet resistance, and electromigration.

3.1 The Bonding Pad

The bonding pad is at the interface between the die and the package or the outside world. One side of a wire is soldered to the pad, while the other side of the wire is connected to a lead frame, as was seen in Fig. 1.3. Figure 3.1 shows a close up of a bonding pad and wire. In this chapter we will not concern ourselves with electrostatic discharge (ESD) protection, which is an important design consideration when designing the pad.

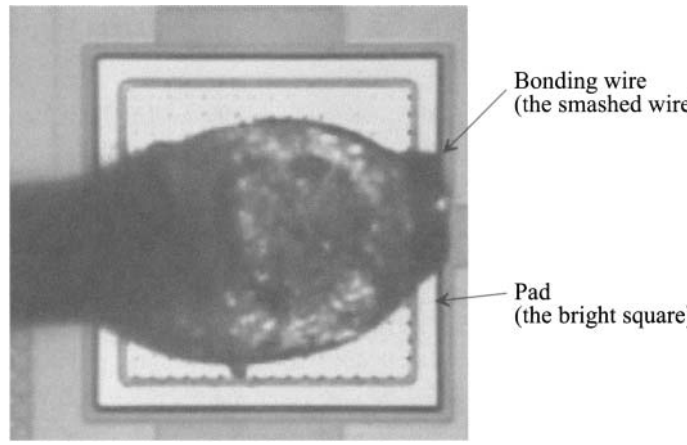


Figure 3.1 The bonding wire connection to a pad.

3.1.1 Laying Out the Pad I

The basic size of the bonding pad specified by MOSIS is a square 100 μm x 100 μm (actual size). For a probe pad, used to probe the circuit with a microprobe station, the size should be greater than 6 μm x 6 μm. In production chips the pads may vary in size (e.g., 75 x 100, or 50 x 75, etc.) depending on the manufacturer's design rules. *The final size of the pads are the only part of a layout that doesn't scale as process dimensions shrink.* The layout of a pad that uses metal2 is shown in Fig. 3.2. Notice, in the cross-sectional view, the layers of insulator (SiO_2 in most cases) under and above the metal2. These layers are used for isolation between the other layers in the CMOS process.

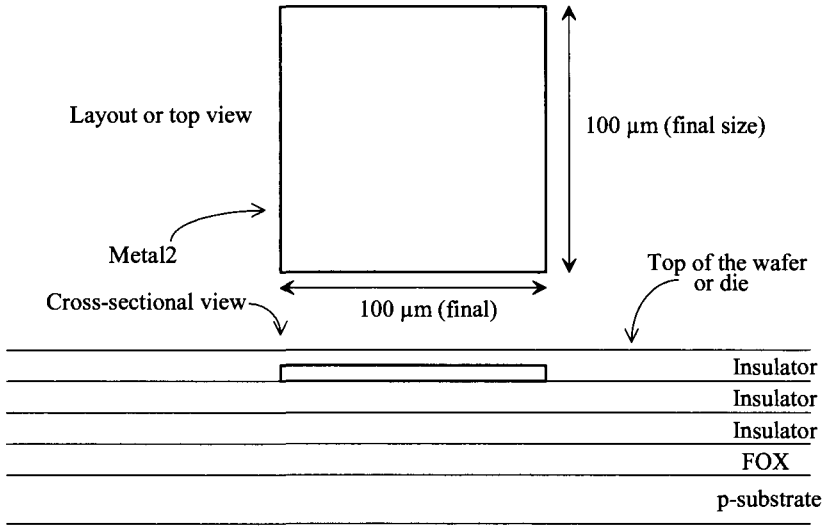


Figure 3.2 Layout of metal2 used for bonding pad with associated cross-sectional view.

Capacitance of Metal-to-Substrate

Before proceeding any further, we might ask the question, "What is the capacitance from the metal2 box (pad) in Fig. 3.2 to the substrate?" The substrate is at ground potential and so, for all intents and purposes, it can be thought of as an equipotential plane. This is important because we have to drive this capacitance to get a signal off the chip. Table 3.1 gives typical values of parasitic capacitances for a CMOS process. Consider the following example.

Example 3.1

Estimate the parasitic capacitance associated with the pad in Fig. 3.2.

The capacitance associated with this pad is the sum of the plate (or bottom) capacitance and the fringe (or edge) capacitance. We can write

$$C_{pad,m2\rightarrow sub} = area \cdot C_{plate} + perimeter \cdot C_{fringe} \qquad (3.1)$$

The area of the pad is 100 μm² square (100 μm by 100 μm), while the perimeter of the pad is 400 μm. Using the typical values of capacitance for metal2 to substrate in Table 3.1 gives

$$C_{pad,m2\rightarrow sub} = 10,000 \cdot 14\ aF + 400 \cdot 81\ aF = 172,400\ aF = 172.4\ fF = 0.172\ pF$$

A significant on-chip capacitance. ■

Table 3.1 Typical parasitic capacitances in a CMOS process. Note that while the physical distance between the layers decreases, as process technology scales downwards, the dielectric constant used in between the layers can be decreased to keep the parasitic capacitances from becoming too significant. The values are representative of the parasitics in both long- and short-channel CMOS processes.

	Plate Cap. aF/μm²			Fringe Cap. aF/μm		
	min	typ	max	min	typ	max
Poly1 to subs. (FOX)	53	58	63	85	88	92
Metal1 to poly1	35	38	43	84	88	93
Metal1 to substrate	21	23	26	75	79	82
Metal1 to diffusion	35	38	43	84	88	93
Metal2 to poly1	16	18	20	83	87	91
Metal2 to substrate	13	14	15	78	81	85
Metal2 to diffusion	16	18	20	83	87	91
Metal2 to metal1	31	35	38	95	100	104

Example 3.2

The pad layout in Fig. 3.2 is the actual size. However, when we lay out the pad with the other circuit components, it must also be scaled when the layout is streamed out (see Sec. 1.2.3). If the scale factor in a design is 50 nm, what is the size of the box used for a pad that we draw with the layout program? Does the capacitance calculated in Ex. 3.1 change?

Because we want a final pad size of 100 μm by 100 μm, the drawn layout size of the box with a scale factor of 50 nm is

$$\frac{100\ \mu m}{0.05\ \mu m} = 2,000 \text{ (drawn size)}$$

Each side of the pad, in Fig. 3.2, is drawn with a size of 2,000 for a final (actual) size of 100 μm by 100 μm.

The capacitance calculated in Ex. 3.1 doesn't change. We can rewrite Eq. (3.1) as

$$C_{pad,m2\rightarrow sub} = area_{drawn} \cdot (scale)^2 \cdot C_{plate} + perimeter_{drawn} \cdot (scale) \cdot C_{fringe} \quad (3.2)$$

to use the drawn layout size. At this point there should be no confusion between the terms "drawn layout size" and "actual or final layout size." ■

Passivation

Because an insulator is covering the pad (the piece of metal2) in Fig. 3.2, we can't bond (connect a wire) to it. The top layer insulator on the chip is also called passivation. The passivation helps protect the chip from contamination. Openings for bonding pads are called cuts in the passivation. To specify an opening or cut in the glass (insulator) covering the metal2, we use the overglass layer. The MOSIS rules specify 6 μm distance between the edge of the metal2 and the overglass box, as seen in Fig. 3.3. The drawn distance between the smaller overglass box and the larger metal2 box, with a scale factor of 50 nm, is 6/0.05 or 120.

There may be another layer in the MOSIS setups for the layout tool called the PAD layer. This layer has no fabrication significance but rather is used by the machine that bonds the chip to the lead frame to indicate the location of the pads. Since MOSIS takes designs of varying sizes and shapes, the locations of the pads change from one project to the next. This layer isn't really necessary since we can use the overglass layer (ensuring the overglass box has a drawn size of 1,760 square or a final size of 88 μm square) to indicate the location of the pads. We won't use the PAD layer in our layouts here.

An Important Note

Here we are using a CMOS process with (only) two layers of metal. In most modern CMOS processes, more than two layers of metal are used. If the process has five layers of metal, then the top layer (just like the top floor in a five-story building) is metal5. Therefore, metal5 is the layer the bonding wire is connected to.

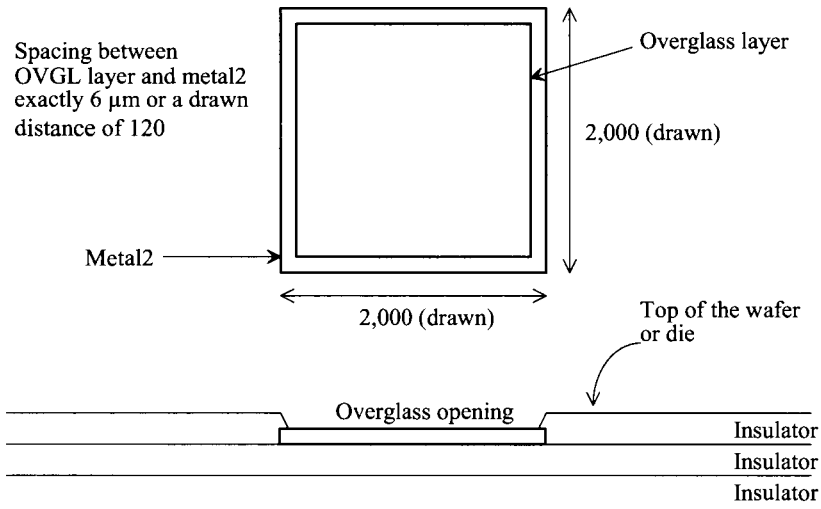


Figure 3.3 Layout of a metal2 pad with pad opening for bonding connection in a 50 nm (scale factor) CMOS process.

3.2 Design and Layout Using the Metal Layers

As mentioned earlier, the metal layers connect the resistors, capacitors, and MOSFETs in a CMOS integrated circuit. So far, in this book, we've learned about the layout layers n-well, metal2, overglass, and pad. In this section we'll also learn about the metal1 and via1 layers and the associated parasitic resistances and capacitances of these layers.

3.2.1 Metal1 and Via1

Metal1 is a layer of metal found directly below metal2. Figure 3.4 shows an example layout and cross-sectional view. The via1 layer connects metal1 and metal2. The via layer specifies that the insulator be removed in the location indicated. Then, for example, a tungsten "plug" is fabricated in the insulator's opening. When the metal2 is laid down, the plug provides a connection between the two metals. Note that if we were to use more than two layers of metal, then via2 would connect metal2 to metal3, via3 would connect metal3 to metal4, etc.

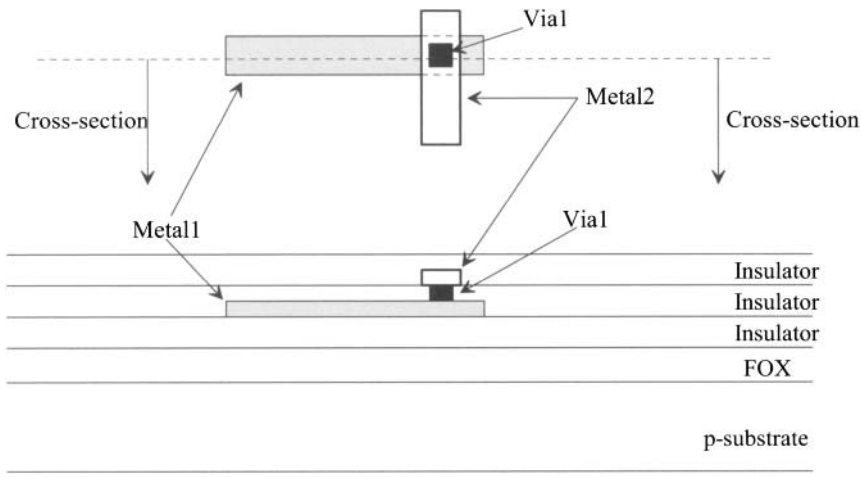


Figure 3.4 Layout and cross-sectional views.

An Example Layout

Figure 3.5 shows an example layout using the n-well, metal1, via1, and metal2 layers. It's important that, before proceeding, this layout and the associated cross-sectional view are understood. For example, how would our cross-sectional view change if we moved the cross-sectional line used in Fig. 3.5 down slightly so that it only intersects the n-well and the metal2 layers? Answer: the cross-sectional view would be the same as seen in Fig. 3.5 except that the metal1 and via1 layers wouldn't be present.

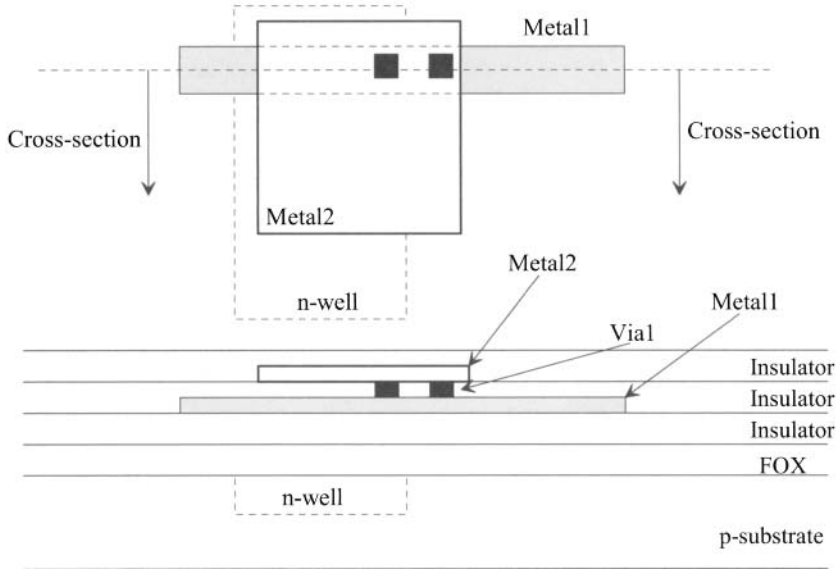


Figure 3.5 An example layout and cross-sectional view using including the n-well.

3.2.2 Parasitics Associated with the Metal Layers

Associated with the metal layers are parasitic capacitances (see Table 3.1) and resistance. Like the n-well in the last chapter, the metal layers are characterized by a sheet resistance. However, the sheet resistance of the metal layers is considerably lower than the sheet resistance of the n-well. For the sake of examples in this book, we'll use metal sheet resistances of **0.1 Ω/square**. Also, there is a finite contact resistance of the via. The following examples illustrate some of the unwanted parasitics associated with these layers.

Example 3.3

Estimate the resistance of a piece of metal1 1 mm long and 200 nm wide. What is the drawn size of this metal line if the scale factor is 50 nm? Also estimate the delay through this piece of metal, treating the metal line as an RC transmission line. Verify your answer with a SPICE simulation.

The drawn size of the metal line is 1 mm/50 nm (= 20,000) by 200/50 (= 4). Figure 3.6 shows the layout of the metal wire (not to scale). The line consists of 1,000/0.2 = 20,000/4 = 5,000 squares of metal1.

To calculate the resistance of the metal line, we use Eq. (2.3)

$$R = \frac{0.1\ \Omega}{\text{square}} \cdot \frac{20,000}{4} = 0.1 \cdot 5,000 = 500\ \Omega$$

To calculate the capacitance, we use the information in Table 3.1 and either Eq. (3.1) or (3.2)

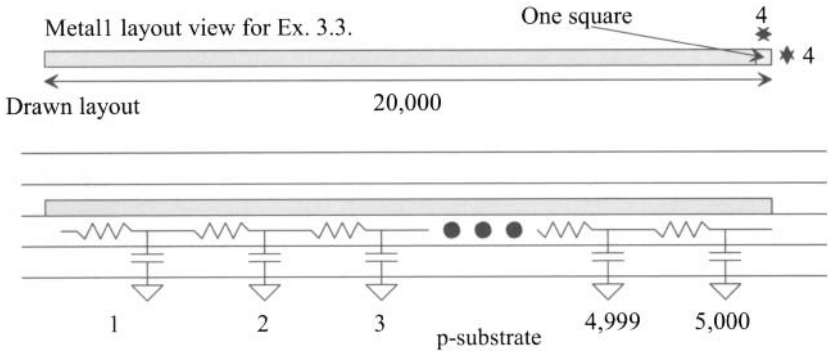


Figure 3.6 Layout and cross-sectional view with parasitics for the metal line in Ex. 3.3.

$$C = (1{,}000 \cdot 0.2) \cdot 23\ aF + (2{,}000 \cdot 4) \cdot 79\ aF = 162\ fF$$

or the capacitance for each 200 nm by 200 nm square (4 by 4) of metal1 is

$$C_{square} = \frac{162\ fF}{5{,}000} = 32\ aF/\text{square}$$

The delay through the metal line is, using Eqs. (2.32) or (2.33)

$$t_d = 0.35 \cdot R_{square} C_{square} \cdot l^2 = 0.35(0.1)(32\ aF)(5{,}000)^2 = 28\ ps$$

or

$$t_d = 0.35RC = 0.35 \cdot 500 \cdot 162\ fF = 28\ ps$$

The delay of a metal1 line (with nothing connected to it) is 28 ps/mm when the parasitic capacitance and resistance are the limiting factors. The SPICE simulation results are seen in Fig. 3.7. ■

Intrinsic Propagation Delay

The result of this example (a metal delay of 28 ps/mm) should be compared to the intrinsic delay of a signal propagating in a material with a relative dielectric constant, ε_r (no parasitic resistance). The velocity, v, of the signal in this situation is related to the speed of light, c, by

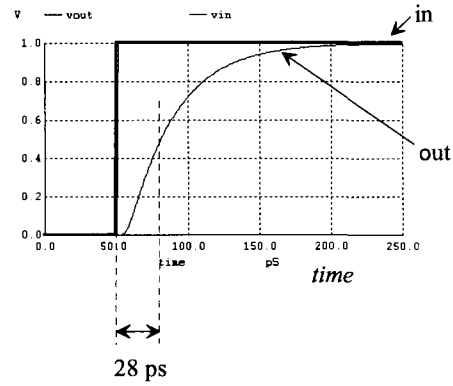
$$v = \frac{c}{\sqrt{\varepsilon_r}} \text{ (meters/second)} \tag{3.3}$$

If we assume the signal is propagating in silicon dioxide (SiO_2) with a relative dielectric constant of roughly 4, then we can estimate the delay of the metal line as

$$\frac{t_d}{\text{meter}} = \frac{1}{v} = \frac{\sqrt{\varepsilon_r}}{c} = \frac{2}{3 \times 10^8\ m/s} = \frac{6.7\ ns}{\text{meter}} \tag{3.4}$$

or a delay of 6.7 ps/mm. For any practical integrated circuit wire in bulk CMOS, the parasitics (RC delay) dominate the propagation delays.

Note that increasing the width of the wire decreases its resistance and increases its capacitance (resulting in the delay staying relatively constant).




```
*** Figure 3.7 CMOS: Circuit Design, Layout, and Simulation ***
.control
destroy all
run
plot vin vout
.endc
.tran 1p 250p

O1 Vin 0 Vout 0 TRC
Rload Vout 0 1G
Vin vin 0 DC 0 pulse 0 1 50p 0
.model TRC ltra R=0.1 C=32e-18 len=5k
.end
```

Figure 3.7 Simulating the delay through a 1 mm wire made using metal1.

Example 3.4

Estimate the capacitance between a 10 by 10 square piece of metal1 and an equal-size piece of metal2 placed exactly above the metal1 piece. Assume a scale factor of 50 nm. Sketch the layout and the cross-sectional views. Also sketch the symbol of a capacitor on the cross-sectional view.

The plate capacitance, from Table 3.1, between metal1 and metal2 is typically 35 aF/μm^2, while the fringe capacitance is typically 100 aF/μm. The two layers form a parallel plate capacitor, Fig. 3.8. The capacitance between the plates is given by the sum of the plate capacitance and the fringe capacitance, or

$$C_{12} = 100 \cdot (0.05)^2 \cdot 35\ aF + 40 \cdot (0.05) \cdot 100\ aF = 209\ aF$$

■

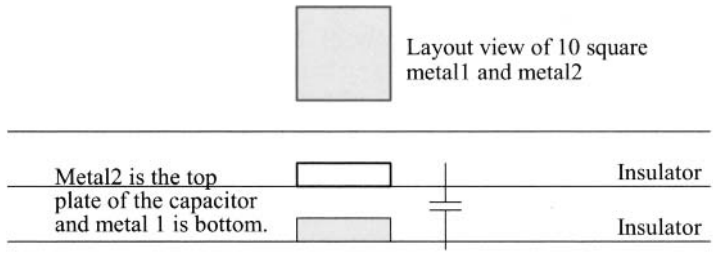


Figure 3.8 Capacitance between metal1 and metal2.

Example 3.5

In the previous example, estimate the voltage change on metal1 when metal2 changes potential from 0 to 1 V. Verify the result with SPICE.

The capacitance from metal2 to metal1 was calculated as 209 aF. The capacitance from metal1 to substrate is given by

$$C_{1sub} = 100 \cdot (0.05)^2 \cdot (23) + 40 \cdot (0.05) \cdot 79 = 164\ aF$$

The equivalent schematic is shown in Fig. 3.9. Since charge must be conserved we can write

$$C_{12} \cdot (\Delta V_{metal2} - \Delta V_{metal1}) = C_{1sub} \cdot \Delta V_{metal1}$$

The change in voltage on C_{1sub} (metal1) is then given by

$$\Delta V_{metal1} = \Delta V_{metal2} \cdot \frac{C_{12}}{C_{12} + C_{1sub}} = 1 \cdot \frac{C_{12}}{C_{12} + C_{1sub}} = \frac{209}{209 + 164} = 560\ mV$$

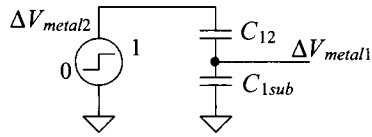
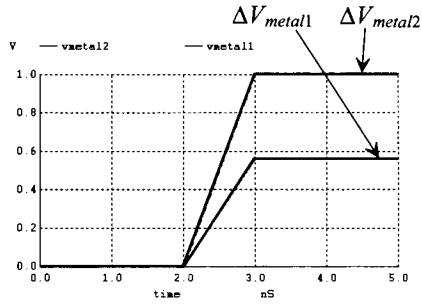


Figure 3.9 Equivalent circuit used to calculate the change in metal1 voltage, see Ex. 3.5.

A displacement current flows through the capacitors, causing the potential on metal1 to change by 560 mV. This may seem significant at first glance. However, one must remember that most metal lines in a CMOS circuit are being driven from a low-impedance source; that is, the metal is not floating but is being held at some potential. This is not the case in some dynamic circuits or in circuits with high-impedance nodes or long metal runs. Figure 3.10 shows the SPICE simulation results and netlist. Notice how we used the "use initial conditions" (UIC) in the transient statement. This sets all nodes that aren't driven by a source to, initially (at the beginning of the simulation), zero volts. Note that older versions of SPICE don't recognize "a" as atto so we used "e-18." ■




```
*** Figure 3.10 CMOS: Circuit Design, Layout, and Simulation ***

.control
destroy all
run
plot vmetal2 vmetal1
.endc

.tran 10p 5n UIC

vmetal2 vmetal2 0 DC 0 pulse 0 1 2n 1n

C12 vmetal2 vmetal1 209e-18
C1sub vmetal1 0 164e-18

.end
```

Figure 3.10 Simulating the operation of the circuit in Fig. 3.9.

3.2.3 Current-Carrying Limitations

Now that we have some familiarity with the metal layers, we need to answer the question, "How much current can we carry on a given width or length of metal?" The factors that limit the amount of current on a metal wire or bus are metal electromigration and the maximum voltage drop across the wire or bus due to the resistance of the metal layer.

A conductor carrying too much current causes metal electromigration. This effect is similar to the erosion that occurs when a river carries too much water. The result is a change in the conductor dimensions, causing spots of higher resistance and eventually failure. If the current density is kept below the metal migration threshold current density, J_{Al}, metal electromigration will not occur. Typically, for aluminum, the current threshold for migration J_{Al} is $1 \rightarrow 2 \frac{\text{mA}}{\mu\text{m}}$.

Example 3.6

Assuming a scale factor of 50 nm, estimate the maximum current a piece of metal1 with a drawn width of 3 can carry. Also estimate the maximum current a 100 by 100 μm^2 bonding pad can receive from a bonding wire. Assume that the metal wires are fabricated in aluminum.

The actual width of the metal1 wire in this example is 150 nm. Assuming that $J_{AL} = 1 \frac{\text{mA}}{\mu\text{m}}$, the maximum current on a 0.15 μm wide aluminum conductor is given by

$$I_{max} = J_{Al} \cdot W = 10^{-3} \cdot 0.15 = 150\ \mu A$$

The maximum current through a bonding pad is then 100 mA. ■

Example 3.7

Estimate the voltage drop across the conductor discussed in the previous example when the length of the conductor is 1 cm and the current flowing in the conductor is 150 μA (I_{max}).

The sheet resistance of metal1 is 0.1 Ω/square. The voltage drop across a metal1 wire that is 3 (0.15 μm) wide and 10,000 μm (1 cm) long carrying 150 μA is

$$V_{drop} = (0.1\ \Omega/\text{square}) \cdot \frac{10{,}000}{0.15} \cdot 150\ \mu A = 1\ V$$

or a significant voltage drop. If this conductor were used for power, we would want to increase the width significantly; however, if the conductor is used to route data, the size may be fine. ■

In general, the higher levels of metal (metal2, metal3, etc.) should be used for power routing. Metal2 is approximately twice as thick as metal1 and, therefore, has a lower sheet resistance. Metal3 is thicker than metal2, etc. When routing power, the more metal that is used, the fewer problems, in general, that will be encountered. If possible, a ground or power plane should be used across the entire die (entire levels of metal are used for *VDD* and ground). The more capacitance between the power and ground buses, the harder it is to induce a voltage change on the power plane; that is, the DC voltages will not vary.

3.2.4 Design Rules for the Metal Layers

The design rules for the metal1, via1, and metal2 layers are seen in Fig. 3.11. Note that the via1 size must be exactly 1.5 by 1.5. Also note that the minimum allowable spacing between two wires using metal1 is 1.5, while the spacing between wires using metal2 is 2. There isn't a spacing rule between metal1 and metal2 because wires made with metal1 and metal2 are isolated by an insulator (sometimes called an interlayer dielectric, ILD).

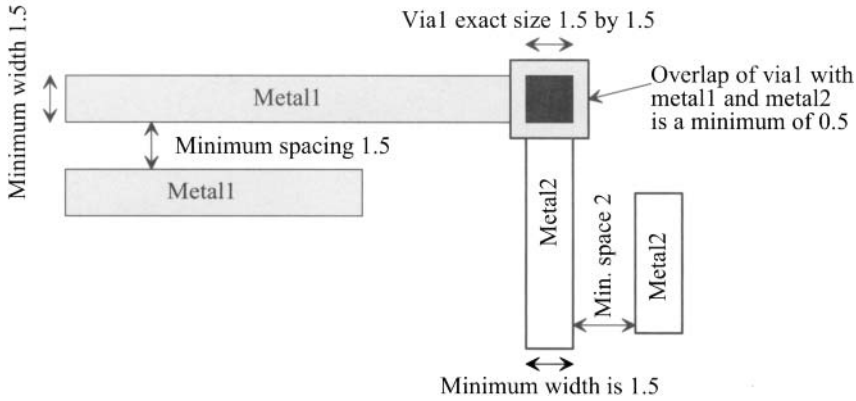
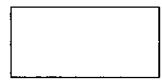


Figure 3.11 Design rules for the metal layers using the CMOSEDU rules.

Layout of Two Shapes or a Single Shape

When learning to do layout, one may wonder about the equivalence of the two layouts seen in Fig. 3.12. In (a) two boxes are used while in (b) a single box is used. When the masks are made the layouts are equivalent.

(a) Layout using two boxes



(b) Layout using a single box

Figure 3.12 Equivalence of layouts drawn with a different number of shapes.

A Layout Trick for the Metal Layers

Notice that the size of the via is exactly 1.5 by 1.5 and that the minimum metal surrounding the via is 0.5. In order to save time when doing layout, a cell can be made called "via1." Instead of drawing boxes on the via1, metal1, and metal2 layers each time a connection between metal1 and metal2 is needed, we simply place the "via1" cell into the layout, Fig. 3.13.

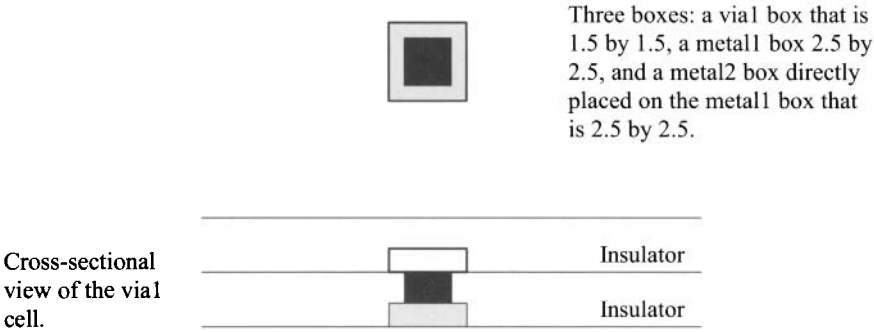


Figure 3.13 Via1 cell with a rank of 1.

3.2.5 Contact Resistance

Associated with any contact to metal (or any other layer in a CMOS process for that matter) is an associated contact resistance. For the examples using metal layers in this book, we'll use a contact resistance of **10 Ω/contact**. Consider the following example.

Example 3.8

Sketch the equivalent electrical schematic for the layout depicted in Fig. 3.14a showing the via contact resistance. Estimate the voltage drop across the contact resistance of the via when 1 mA flows through the via. Repeat for the layout shown in Fig. 3.14b.

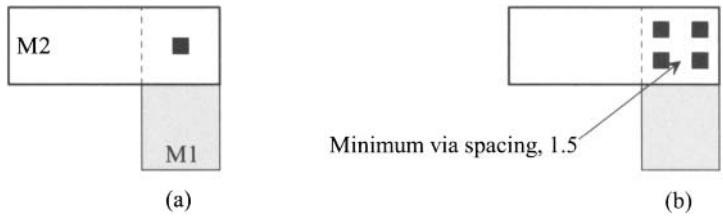


Figure 3.14 Layouts used in Ex. 3.8.

The equivalent schematics are shown in Fig. 3.15a and (b) for the layouts in Figs. 3.14a and (b) respectively. If the via contact resistance is 10 Ω, and 1 mA flows through the via in (a), then a voltage drop of 10 mV results. Further, the reliability of the single via will be poor with 1 mA flowing through it due to electromigration effects. A "rule-of-thumb" is to allow no more than 100 μA of current flow per via. The four vias shown in Fig. 3.14(b) give an effective contact resistance of 10/4 or 2.5 Ω because the contact resistances of each of the vias are in parallel. The voltage drop across the vias decreases to 2.5 mV with 1 mA flowing in the wires. Increasing the metal overlap and the number of vias will further decrease the voltage drop (and electromigration effects). ■

(a) The contact resistance of the via in Fig. 3.14a.

(b) The contact resistance of the four vias in Fig. 3.14b.

Figure 3.15 The schematics of the contact resistances for the layouts in Fig. 3.14.

3.3 Crosstalk and Ground Bounce

Crosstalk is a term used to describe an unwanted interference from one conductor to another. Between two conductors there exists mutual capacitance and inductance, which give rise to signal feedthrough. Ground bounce (and *VDD* droop) are terms describing local variations in the power and ground supplies at a circuit. While crosstalk is only a problem for time-varying signals in a circuit, ground bounce can be problematic for both time varying and DC signals.

3.3.1 Crosstalk

Consider the two metal wires shown in Fig. 3.16. A signal voltage propagating on one of the conductors couples current onto the conductor. This current can be estimated using

$$I_m = C_m \frac{dV_A}{dt} \tag{3.5}$$

where C_m is the mutual capacitance, I_m is the coupled current, and V_A is the signal voltage on the source conductor. Treating the capacitance between the two conductors in this

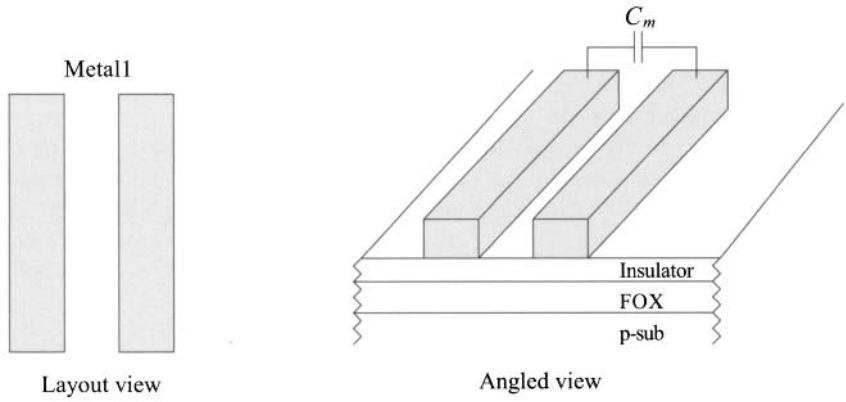


Figure 3.16 Conductors used to illustrate crosstalk.

simple manner is useful in most cases. Determining C_m experimentally proceeds by applying a step voltage to one conductor while measuring the coupled voltage on the adjacent conductor. Since we know the capacitance of any conductor to substrate (see Table 3.1), we can write

$$\Delta V = V_A \cdot \frac{C_m}{C_m + C_{1sub}} \tag{3.6}$$

where ΔV is the coupled noise voltage to the adjacent conductor and C_{1sub} is the capacitance of the adjacent conductor (in this case metal1) to ground (the substrate).

The adjacent metal lines shown in Fig. 3.16 also exhibit a mutual inductance. The effect can be thought of as connecting a miniature transformer between the two conductors. A current flowing on one of the conductors induces a voltage on the other conductor. Measuring the mutual inductance begins by injecting a current into one of the conductors. The voltage on the other conductor is measured. The mutual inductance is determined using

$$V_m = L_m \frac{dI_A}{dt} \tag{3.7}$$

where I_A is the injected (time-varying) current (the input signal), V_m is the induced voltage (the output signal), and L_m is the mutual inductance.

Crosstalk can be reduced by increasing the distance between adjacent conductors. In many applications (e.g., DRAM), the design engineer has no control over the spacing (pitch) between conductors. The circuit designer then attempts to balance the signals on adjacent conductors (see, for example, the open and folded architectures in Ch. 16 concerning DRAM design).

3.3.2 Ground Bounce

DC Problems

Consider the schematic seen in Fig. 3.17a. In this schematic a circuit is connected to *VDD* and ground through two wires measuring 10,000 μm (10 mm) by 150 nm (with a resistance of 6.67 kΩ). Next consider, in (b), what happens if the circuit starts to pull a DC current of 50 μA. Instead of the circuit being connected to a *VDD* of 1 V the actual *VDD* drops to 667 mV. Further, the actual "ground" connected to the circuit increases to 333 mV. The voltage dropped across the circuit is the difference between the applied *VDD* and ground or only 333 mV (considerably less than the ideal 1 V). The obvious solution to making the supplied *VDD* and ground move closer to the ideal values is to increase the widths of the conductors supplying and returning currents to the circuit. This reduces the series resistance. The key point here is that *VDD* and ground are not fixed values; rather, they can vary depending how the circuit is laid out.

AC Problems

It is common, in CMOS circuit design, for a CMOS circuit to draw practically zero current in a static state (not doing anything). This is why, for example, it's possible to use solar power in a CMOS-based calculator. In this situation, conductors with small widths, as those in Fig. 3.17, may be fine. However, consider what happens if the circuit, for a short time, pulls 50 μA. As discussed above, the ground bounces up and *VDD* droops

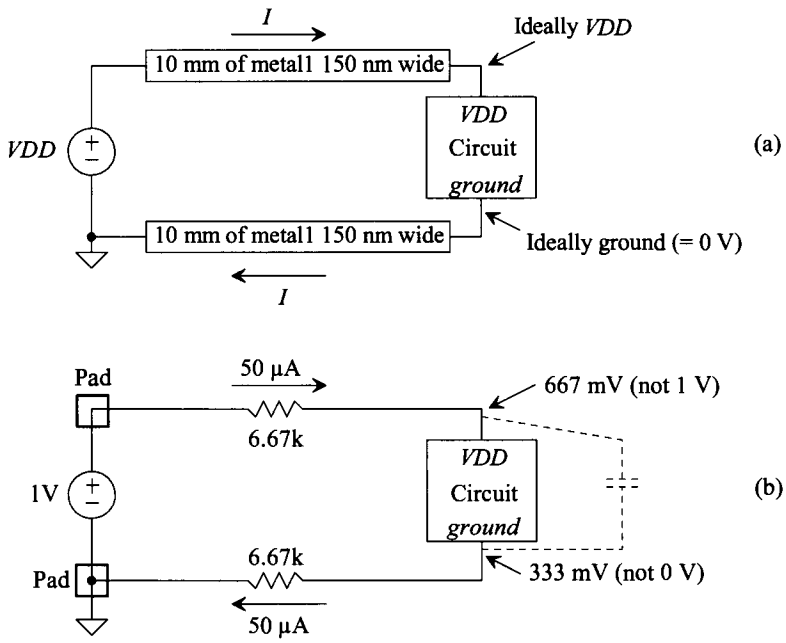


Figure 3.17 Illustrating problems with incorrectly sized conductors.

down during this short time. The average current supplied by *VDD* may be well under a microamp; however, the occasional need for 50 μA still creates or causes problems. To circumvent these, consider adding an on-chip *decoupling* capacitor physically at the circuit between *VDD* and ground (see dotted lines in Fig. 3.17b). The added capacitor supplies the needed charge during the transient times and keeps the voltage applied across the circuit at *VDD*. Note that a decoupling capacitor should be used external to the chip as well. The capacitor is placed across the *VDD* and ground pins of the chip.

Example 3.9

Suppose that the circuit in Fig. 3.17b needs 50 μA of current for 10 ns. Estimate the size of the decoupling capacitor required if the voltage across the circuit should change by no more than 10 mV during this time.

We can write the charge supplied by the capacitor as

$$Q = I \cdot \Delta t = (50\ \mu A) \cdot 10\ ns = 500 \times 10^{-15}\ Coulumbs$$

The decoupling capacitor must supply this charge

$$\Delta V \cdot C = Q \rightarrow C \geq \frac{Q}{\Delta V} = \frac{I \cdot \Delta t}{\Delta V} = \frac{500 \times 10^{-15}}{10\ mV} \rightarrow C \geq 50\ pF \qquad (3.8)$$

A reasonably large capacitor. ■

Example 3.10

To drive off-chip loads, an output buffer (see Ch. 11) is usually placed in between the on-chip logic and the large off-chip load, Fig. 3.18. If *VDD* is 1 V and it is desirable to drive the 30 pF off-chip load in Fig. 3.18 to 900 mV in 1 ns, estimate the size of the decoupling capacitor required. Assume that ground variations are not a concern.

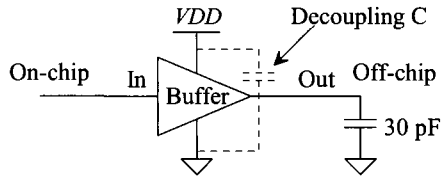


Figure 3.18 Estimating the decoupling capacitance needed in an output buffer.

The charge supplied to the 30 pF capacitor (the load capacitance) by the output buffer is

$$Q = (900\ mV) \cdot (30\ pF) = 27\ pC$$

This charge is supplied by the decoupling capacitor (assuming that the conductors powering the buffer are narrow). Initially, the decoupling capacitor is charged to *VDD* (1 V). If *VDD* (actually the voltage across the decoupling capacitor) drops to 900 mV, then using Eq. (3.8) we can calculate the size of decoupling capacitor as

$$C \geq \frac{27\ pC}{100\ mV} = 270\ pF\ !!!$$

Not a practical value for an on-chip capacitor in most situations. The solution to this problem is to supply *VDD* and ground to the output buffers through wide conductors. Often separate power and ground pads (and very wide wires) are used to power the output buffers separately from the other on-chip circuitry. Using separate pads reduces the size of the decoupling capacitor required and eliminates the noise (ground bounce and *VDD* droop) from interfering with the operation of the other circuitry in the chip. Off-chip decoupling capacitors should still be used across the power and ground pins for the output buffers.

Note that if this buffer is running at 500 MHz (a clock period of 2 ns), the average current supplied to the load is

$$I_{avg} = \frac{27\ pC}{2\ ns} = 13.5\ mA$$

A significant value for a single chip output. ■

A Final Comment

It should be clear that some thought needs to go into the sizing of the metal layers and the number of vias used when transitioning from one metal layer to the next. Ignoring the parasitics associated with the wires used in an IC is an invitation for disaster.

3.4 Layout Examples

In this section we provide some additional layout examples. In the first section we discuss laying out a pad and a padframe. In the following section we discuss laying out test structures to measure the parasitics associated with the metal layers.

3.4.1 Laying out the Pad II

Let's say we want to lay out a chip in a 50 nm process. Further let's say that the final die size (chip size) must be approximately 1 mm on a side with a pad size of 100 μm square (again, the pads can be smaller depending on the process). From the MOSIS design rules, the distance between pads must be at least 30 μm. Further let's assume a two-metal process (so metal2 is the top layer of metal the bonding wire drops down on). Table 3.2 summarizes the final and scaled sizes for our pads.

Table 3.2 Sizes for an example 1 mm square chip with a scale factor of 50 nm.

	Final size	**Scaled size**
Pad size	100 μm by 100 μm	2,000 by 2,000
Pad spacing (center to center)	130 μm	2,600
Number of pads on a side (corners empty)	6	6
Total number of pads	24	24
Overglass opening	88 μm by 88 μm	1,760 by 1,760

Let's start out by laying out a cell called "via1" like the one seen in Fig. 3.13. The resulting cell is seen in Fig. 3.19. We'll use this cell in our pad to connect metal1 to metal2. The bond wire will touch the top metal2. However, we'll place metal1 directly beneath the metal2 so that we can connect to the pad using either metal1 or metal2. The layout of the pad is seen in Fig. 3.20. The spacing between the pads is a minimum of 30 μm. We use an outline layer (no fabrication significance) to help when we place the pads together to form a padframe. We've assumed the distance from the pad metal to the edge of the chip is 15 μm.

Figure 3.21 shows the detail of how the overglass layer is placed in the pad metal area. Also seen in Fig. 3.21 is the placement of the via1 cell in Fig. 3.19 around the perimeter of the pad. This ensures metal1 is solidly shorted to metal2, Fig. 3.22.

Next let's calculate, assuming we want a chip size of approximately 1 mm on a side, the number of pads we can fit on the chip. The size of the pad in Fig. 3.20 is 130 μm square. To determine the number of pads we take the length of a side and divide by the size of a pad or

$$\#\text{ of pads} = \frac{1\ mm}{130\ \mu m} \approx 8 \tag{3.9}$$

However, the corners don't contain a pad so the actual number of pads on a side is six as seen in Fig. 3.23. The CMOS circuits are placed in the area inside the padframe, while

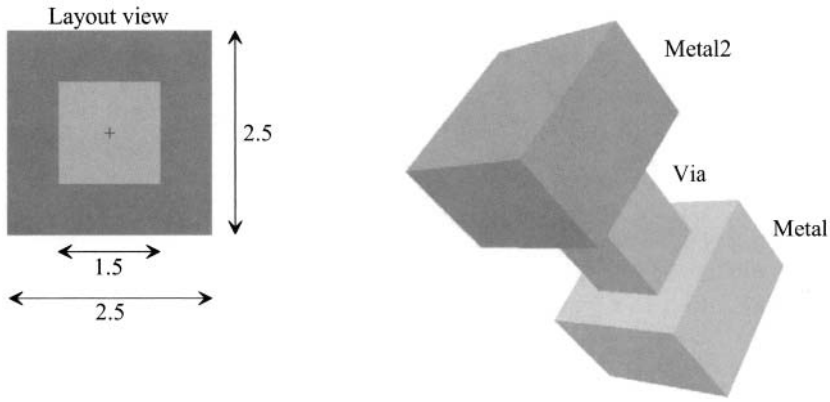


Figure 3.19 Layout of a Via1 cell.

outside the padframe, the scribe, gets cut up when the chips are separated as seen in Fig. 1.2. Note that while this discussion assumed a two metal CMOS process it can be extended to a CMOS process with any number of metals.

The procedure to lay out probe pads, those pads that are not connected to a bonding wire but rather used for probing signals in, generally, unpackaged chips, follows basically the same procedure. The differences are that probe pads can reside anywhere on the chip and that they are smaller than bonding pads.

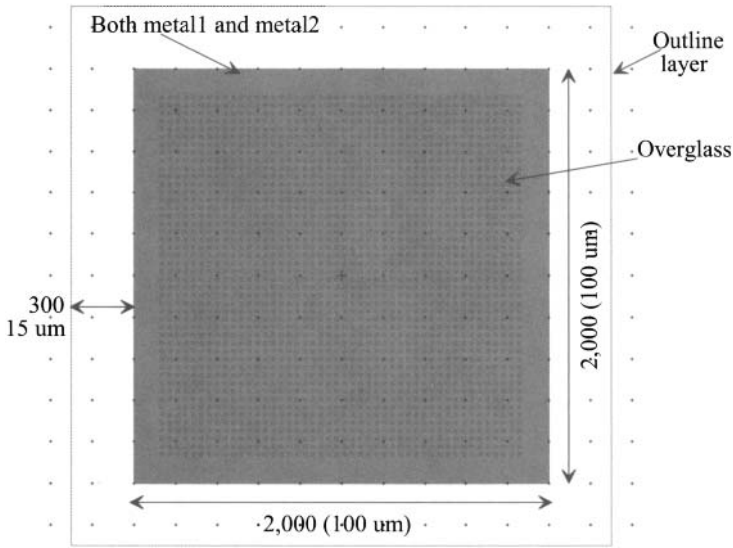


Figure 3.20 Layout of the bonding pad.

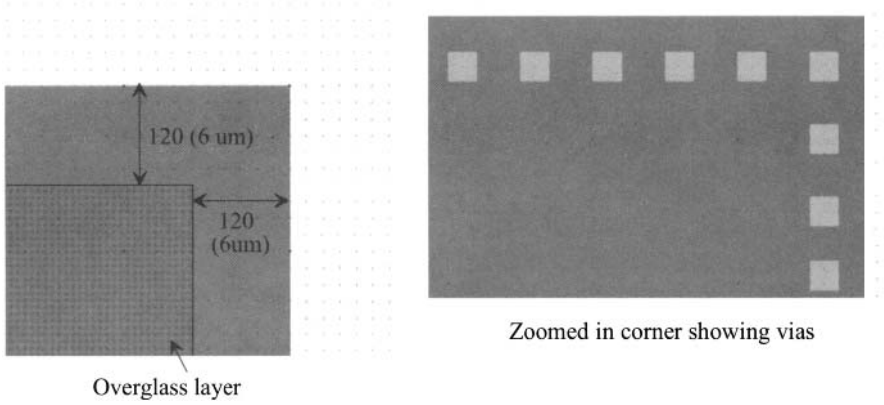


Figure 3.21 Corner detail for the pad in Fig. 3.20.

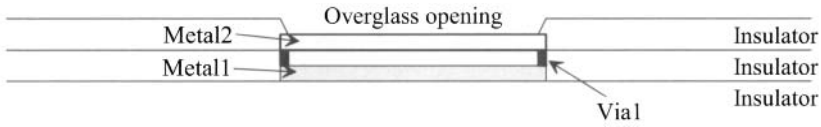


Figure 3.22 Simplified cross-sectional view of the bonding pad discussed in this section.

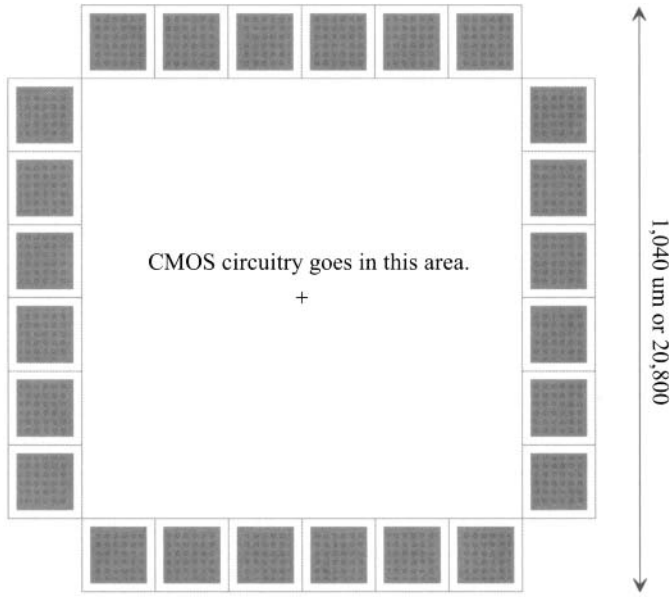


Figure 3.23 The layout of a padframe.

3.4.2 Laying Out Metal Test Structures

To characterize the sheet resistance, plate capacitance, fringe capacitance, and mutual capacitance associated with the metal layers, layouts called "test structures" are used. These layouts take one of two basic shapes. The first is the serpentine pattern seen in Fig. 3.24a (long perimeter while minimizing the area). This pattern is used for measuring sheet resistance or, with two serpentine layouts, mutual capacitance (see Figs. 3.16 and 3.24c). While it can be better to use a very long, straight, length of metal to measure resistance instead of a serpentine pattern (to avoid corners) the length of metal is generally limited by the chip size. By "snaking" the layout the length of metal can be made quite long. The layout seen in Fig. 3.24b, large area minimizing the perimeter, is useful for measuring plate capacitance. We don't use this type of layout to measure resistance because of the error associated with the connections to the metal.

To understand this last statement in more detail consider making a connection to points A and B in (a) to measure the line's resistance. A current is sent flowing in the metal line (say from A to B) and the voltage drop across the line is measured (hence why we can't use too short of a line [resistor], that is, the voltage drop would then be difficult to measure). At the source of current contact point, A, the current will spread out and then flow uniformly down the line where it converges to collection at the receiving contact point, B. In figure (b) the same thing happens; however, the height of the metal line is larger and thus allows for larger spreading/contraction resulting in more measurement error. Also, the wider width in (b) decreases the resistance, between points A and B, lowering the measured voltage and increasing the difficulty of the measurement.

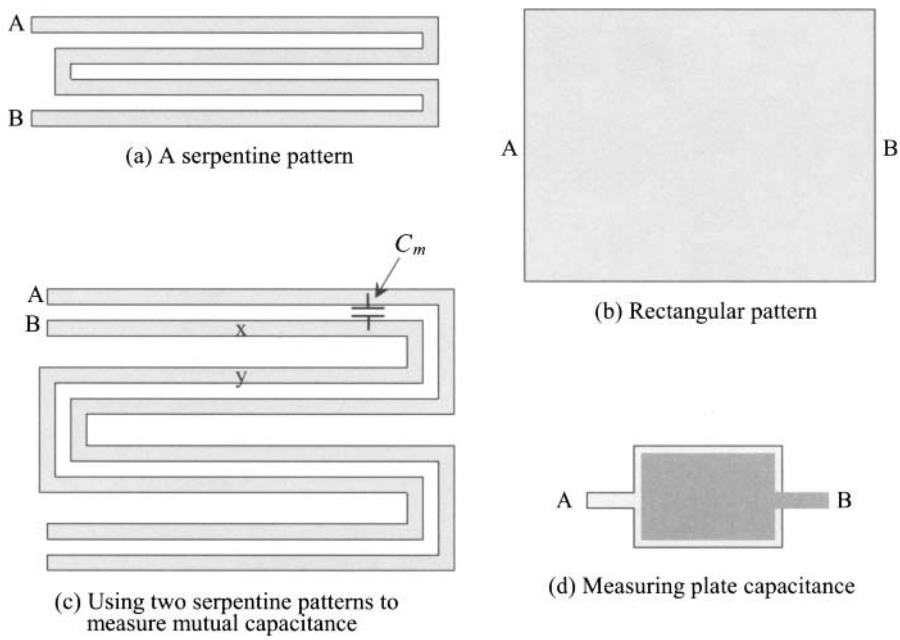


Figure 3.24 Showing the layout of various patterns for measuring parasitics.

In Fig. 3.24c two serpentine patterns are laid out adjacent to each other. This test structure is used to measure the mutual capacitance between like layers (as seen in Fig. 3.16). Again, a serpentine pattern is used to increase the measured variable (capacitance) between points A and B. Minimum spacing is used between the metal lines to maximize the capacitance and because this is the spacing where mutual capacitance has the most influence. To measure the capacitance a low frequency AC voltage is applied between A and B while the displacement current is measured. We need to use a low frequency source to avoid the distributed effects of the metal lines (the delay through the metal lines). Note that at low frequencies points x and y are at the same potential. The result is that the current we measure is restricted to (only) the displacement current between conductors A and B.

The test structure seen in Fig. 3.24d can be used to measure plate capacitance (the capacitance is measured between points A and B). Again large area structures are used to minimize the effects of the perimeter (fringe) capacitance. The test structure can be drawn so that both layers are the same size. To measure the fringe capacitance between two layers the rectangles in (d) are replaced with serpentine structures.

SEM View of Metal

Figure 3.25 shows an SEM image of a portion, a layout view, of a CMOS memory chip. The brighter areas of the image are metal1 while the bright, and circular-shaped, objects are contacts (discussed in the next chapter). Notice that none of the sections of metal or contacts are square or rectangular. While the layout can be square or rectangular the actual fabricated metal layers show rounding (tools such as optical proximity correction, OPC, are used for corrections). If the reader looks closely at the image, the misalignment between the metal and the contacts is seen (hence the reason for design rules).

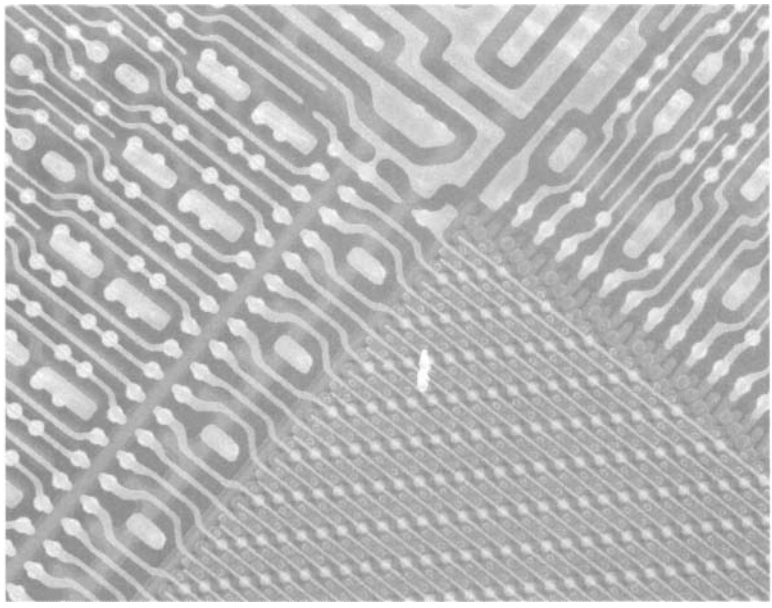


Figure 3.25 SEM photo showing patterned metal layers.

ADDITIONAL READING

[1] R. S. Muller, T. I. Kamins, and M. Chan, *Device Electronics for Integrated Circuits*, John Wiley and Sons Publishers, 2002. ISBN 0-471-59398-2

[2] J. D. Plummer, M. D. Deal, and P. B. Griffin, *Silicon VLSI Technology, Fundamentals, Practice, and Modeling*, Prentice-Hall Publishers, 2000. ISBN 0-13-085037-3

PROBLEMS

Unless otherwise indicated, use the data from Table 3.1, a metal sheet resistance of 0.1 Ω/square, and a metal contact resistance of 10 Ω.

3.1 Redraw the layout and cross-sectional views of a pad, similar to Fig. 3.2, if the final pad size is 50 μm by 75 μm with a scale factor of 100 nm.

3.2 Estimate the capacitance to ground of the pad in Fig. 3.20 made with both metal1 and metal2.

3.3 Suppose a parallel plate capacitor was made by placing a 100 μm square piece of metal1 directly below the metal2 in Fig. 3.2. Estimate the capacitance between the two plates of the capacitor (metal1 and metal2). Estimate the capacitance from metal1 to substrate. The unwanted parasitic capacitance from metal1 to substrate is often called the **bottom plate** parasitic.

3.4 Sketch the cross-sectional view for the layout seen in Fig. 3.26.

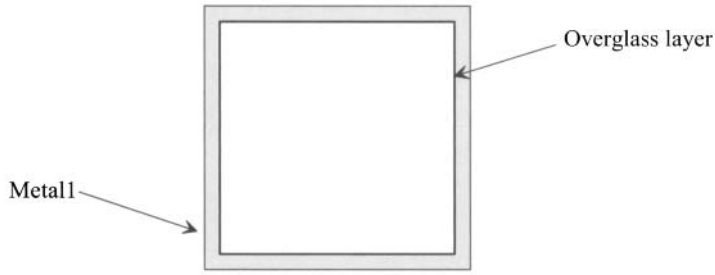


Figure 3.26 Layout used in Problem 3.4.

3.5 Sketch the cross-sectional view, at the dashed line, for the layout seen in Fig. 3.27. What is the contact resistance between metal3 and metal2?

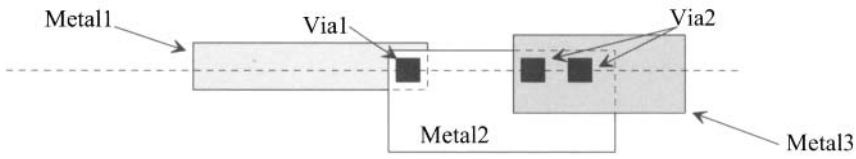


Figure 3.27 Layout for Problem 3.5.

3.6 The insulator used between the metal layers (the interlayer dielectric, ILD) can have a relative dielectric constant well under the relative dielectric constant of SiO_2 (= 4). Estimate the intrinsic propagation delay through a metal line encapsulated in an ILD with a relative dielectric constant of 1.5. What value of metal sheet resistance, using the values from Ex. 3.3, would be required if the RC delay through the metal line is equal to the intrinsic delay?

3.7 Using $CV = Q$, rederive the results in Ex. 3.5.

3.8 For the layout seen in Fig. 3.28, sketch the cross-sectional view (along the dotted line) and estimate the resistance between points A and B. Remember that a via is sized 1.5 by 1.5.

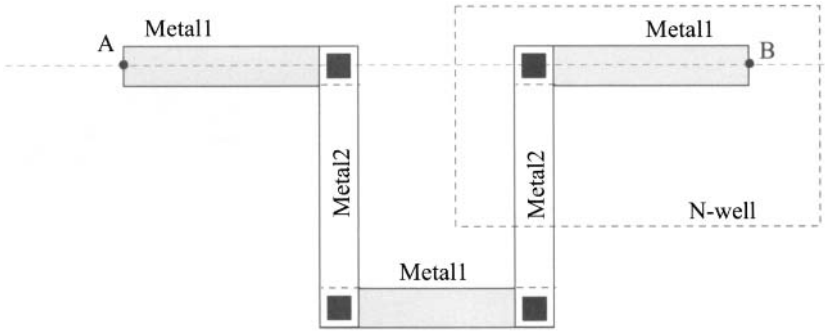


Figure 3.28 Layout for Problem 3.8.

3.9 Laying out two metal wires directly next to each other, and with minimum spacing, for a long distance increases the capacitance between the two conductors, C_m. If the two conductors are *VDD* and ground, is this a good idea? Why or why not?

3.10 Consider the schematic seen in Fig. 3.29. This circuit can be used to model ground bounce and *VDD* droop. Show, using SPICE, that a decoupling capacitor can be used to reduce these effects for various amplitude and duration current pulses.

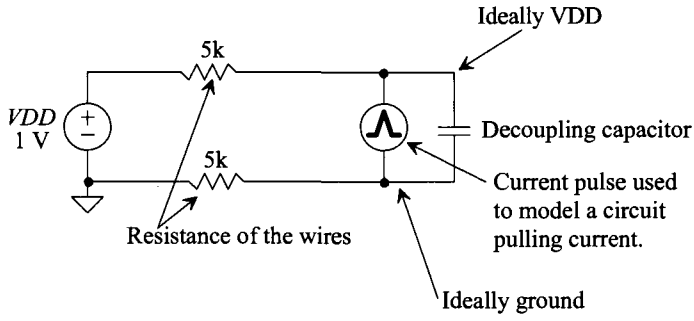


Figure 3.29 Circuit used to show the benefits of a decoupling capacitor.

3.11 Lay out the padframe specified by the information in Table 3.2. Assume a 3 metal CMOS process is used. Comment on how the scale factor affects the (drawn) layout size.

3.12 Propose, and lay out, a test structure to measure the sheet resistance of metal3. Comment on the trade-offs between accuracy and layout size. Using your test structure provide a numerical example of calculating sheet resistance for metal3.