

Interconnect Design

CHAPTER OUTLINE

- 10.1 Introduction
- 10.2 Interconnect *RC* Delays
- 10.3 Buffer Insertion for Very Long Wires
- 10.4 Interconnect Coupling Capacitance
- *10.5 Interconnect Inductance
- *10.6 Antenna Effects
- 10.7 Summary

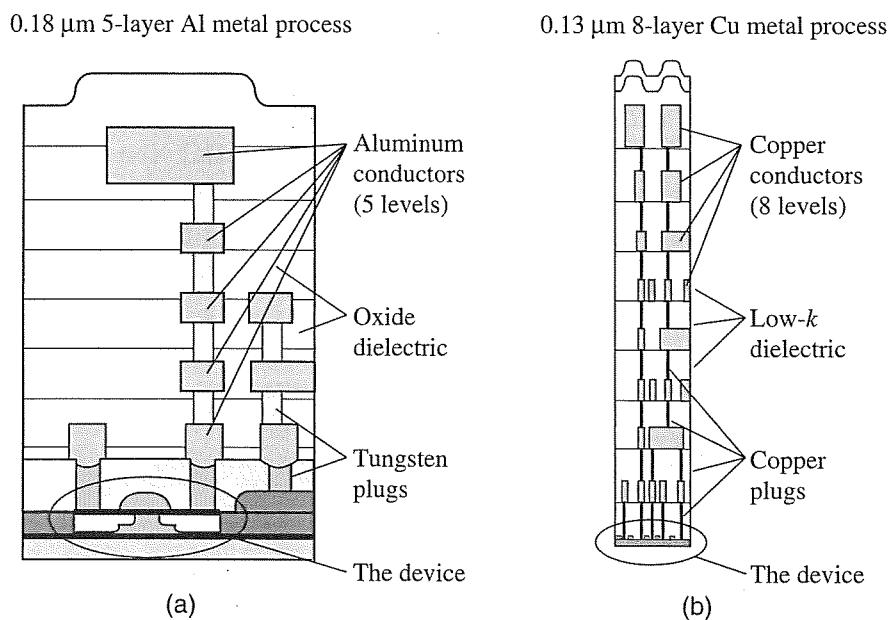
References

Problems

10.1 Introduction

This chapter is devoted to the study of interconnect issues that the IC designer faces when designing in deep submicron technologies. The relative importance of interconnect in deep submicron cannot be overstated. In fact, the topic of interconnect should actually appear much earlier in the text to stress its importance, but it is more appropriate at this stage since the discussion of transistor level circuit design has been completed. Initial coverage of interconnect topics was provided in Chapters 1 and 3. This chapter addresses the issues associated with *RLC* aspects of wires in detail. We begin by reexamining the *RC* delay calculation using the Elmore delay, and address the issue of buffer insertion in long wires. Then, we examine the capacitances associated with 3D interconnect. This topic is rather complex but we provide a simplified treatment to convey the most important concepts. This leads to a discussion of the effect of coupling capacitance on delay and noise injection in logic circuits. The chapter continues with a discussion of inductance effects and concludes with a brief look at antenna effects.

“Interconnect is everything” was an often-used expression starting in the mid-1990s to characterize the importance of interconnect in deep submicron technologies.

**Figure 10.1**

Cross sections of aluminum and copper processes.

While device sizes were shrinking with each technology generation, multilevel metal structures rose higher and higher above the surface of the silicon and soon began to dominate the landscape of the integrated circuit. This is illustrated in the cross-sectional views of $0.18 \mu\text{m}$ aluminum and $0.13 \mu\text{m}$ copper technologies in Figure 10.1. The number of layers of metal has increased from a few layers in the 1980s, to four or five layers in the 1990s, to eight or more layers today. The fabrication of these superstructures required significant advances in integrated circuit processing. However, the impact on design was even more significant. Suddenly, interconnect controlled timing, power, noise, design functionality, and reliability. It controlled all of these important electrical characteristics on the chip, even though the tiny devices (buried under a multilevel structure of wires and vias) defined the actual logic functions. Today interconnect still controls the design process in advanced technologies and will continue to do so for the foreseeable future.

The dimensions associated with a cross section of interconnect are shown in Figure 10.2. The rectangular wires are W wide and T thick, as shown. The separation between layers is a distance H (insulator thickness), and from other conductors is a distance S . The length of the wire, L , is shown only for Metal 2. The Metal 1 wires run into the page and so their lengths are not shown. It is important to realize that the vertical dimensions H_1 , T_1 , H_2 , and T_2 are determined during the fabrication process. The designer has no control over these quantities. On the other hand, the horizontal dimensions W_1 , S_1 , L_1 , L_2 , etc. are all under the control of the designer.

There are a number of reasons to emphasize the importance of interconnect. First, as the width W of wires is decreased, the resistance increases. This increase in wire resistance causes an RC delay phenomenon that had not been experienced before

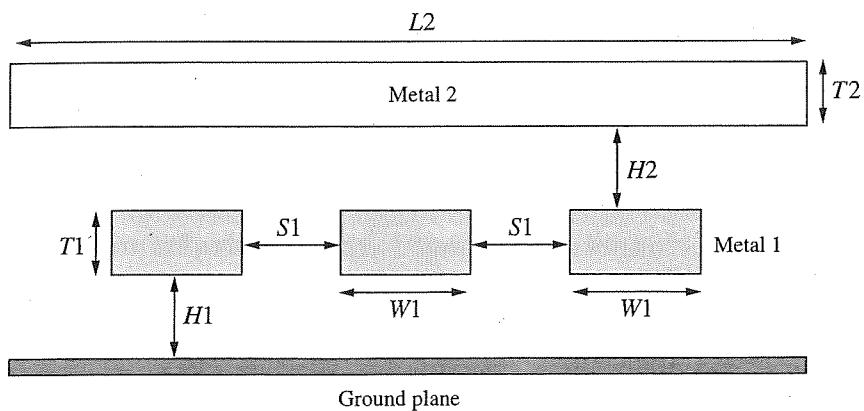


Figure 10.2

Wire dimensions for resistance and capacitance calculations.

1990. By the mid-1990s, a noticeable increase in wire delay led to widespread concern that it would soon be greater than gate delay. Second, the spacing S between wires has been decreasing to the point where the coupling between wires is significant. The resulting capacitive coupling introduces additional delay and noise effects that can cause failures in the design, requiring respins of silicon¹ in order to fix the problem. The overall term for all these problems is *signal integrity*. Recent issues of inductance in wires have been included in the growing list of signal integrity problems.

Other problems have surfaced due to the scaling of interconnect. The increase in resistance in the power supply lines due to scaling causes voltage drop along interconnect, referred to as *IR drop*. This affects the timing and functionality of gates connected to the power lines if the drop becomes too significant. A similar issue concerns the Ldi/dt drop in the power grid due to inductance. Still other effects such as charge storage on metal lines during manufacture, called *antenna effects*, lead to gate oxide breakdown or a threshold voltage shift. Further reliability problems exist due to the migration of the aluminum or copper material in the wires as high levels of current flow through them. Eventually, cracks and other imperfections in the metal cause material erosion until opens or shorts occur. This phenomenon is called *electromigration*, a long-term reliability issue.

Indeed, interconnect is more important than devices in many respects. This chapter and the next should be studied carefully as they contain valuable information pertaining to interconnect design issues. The material on interconnect in the open literature is rather extensive. Our purpose here is to convey the major issues and provide useful insight into the design methods to circumvent these issues. This chapter considers resistance, coupling capacitance, and inductance effects on delay and noise. Antenna effects are also described. The next chapter considers power grid

¹A spin of silicon is the process of designing a chip and then having it fabricated. When errors are found in the design after fabrication, the whole process must be repeated. This is a very expensive and time-consuming exercise. Often several "spins of silicon" are needed to produce a working design for large complex chips.

and clock tree design in the context of interconnect issues. For further information, the reader is encouraged to consult the references at the end of the chapter.

10.2 Interconnect RC Delays

10.2.1 Wire Resistance

The delay due to interconnect has changed considerably due to deep submicron effects. It is our objective to develop models that can be used for accurate delay calculation when interconnect effects must be taken into account. So far, we have used a simple lumped capacitance to model the wire as shown in Figure 10.3a. We now consider the effect of wire resistance on delay. As the width dimension of the wires is reduced, the increased resistance has a pronounced effect on the delay. This was first observed at the $0.8 \mu\text{m}$ technology node at a number of semiconductor companies, but became a widespread issue at the $0.35 \mu\text{m}$ technology node. The interconnect model shown in Figure 10.3b is required when interconnect resistance is important.

The resistance of a material is given by

$$R = \frac{\rho L}{A} = \frac{\rho L}{TW} \quad (10.1)$$

where ρ is the resistivity of the material in $\Omega\text{-cm}$, L is the length of the wire, T is the thickness of the wire, and W is the wire width. The leading term in Equation (10.1) is an important one:

$$R_{sq} = \frac{\rho}{T} \quad (10.2)$$

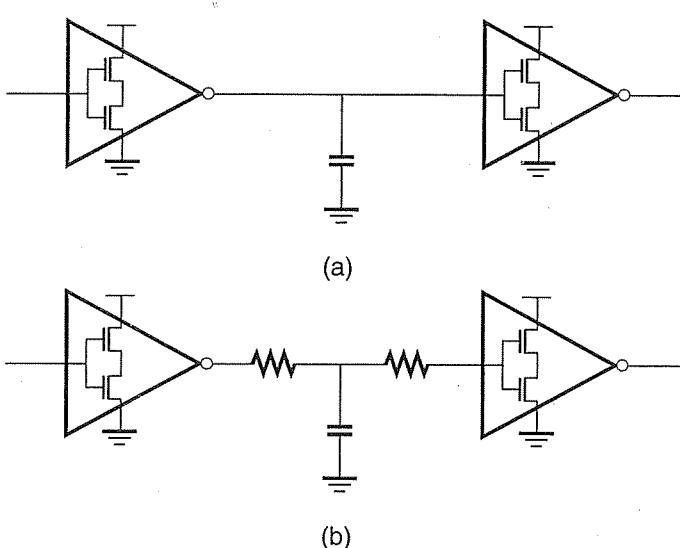


Figure 10.3

(a) Lump capacitance model. (b) *RC* interconnect model.

where R_{sq} is the *sheet resistance* of the material and has the units of *ohms per square* (Ω/\square). The sheet resistance of any metal layer can readily be computed from the resistivity and the thickness, T . To keep resistance relatively low in each technology generation, the thickness has not been scaled at the same rate as the minimum line width. However, the thickness of the metal lines will vary, depending on the metal layer. To simplify our description, we assume that two different thicknesses are used, one for the lower layers and one for the upper layers. Because the upper layers carry much larger currents over longer distances, they require a lower resistance, and therefore a higher thickness.

The two most common materials for interconnect are aluminum, with $\rho = 2.7 \mu\Omega\text{-cm}$, and copper, with $\rho = 1.7 \mu\Omega\text{-cm}$.² The metal hierarchies for a $0.18 \mu\text{m}$ aluminum process and a $0.13 \mu\text{m}$ copper process were shown in Figure 10.1. The approximate sheet resistances for a $0.18 \mu\text{m}$, 5-layer aluminum process assuming that metals 1 to 4 have a thickness of $0.5 \mu\text{m}$ and metal 5 has a thickness of $1.0 \mu\text{m}$ are as follows:

$$R_{sq} = \frac{\rho}{T} = \frac{2.7 \mu\Omega\text{-cm}}{0.5 \mu\text{m}} = 54 \text{ m}\Omega/\square \quad \text{Al metals 1-4} \quad (10.3)$$

$$R_{sq} = \frac{\rho}{T} = \frac{2.7 \mu\Omega\text{-cm}}{1.0 \mu\text{m}} = 27 \text{ m}\Omega/\square \quad \text{Al metal 5}$$

The approximate sheet resistances for a $0.13 \mu\text{m}$, 8-layer copper process assuming that metals 1–6 have a thickness of $0.4 \mu\text{m}$ and metals 7–8 have a thickness of $0.8 \mu\text{m}$ are as follows:

$$R_{sq} = \frac{\rho}{T} = \frac{1.7 \mu\Omega\text{-cm}}{0.4 \mu\text{m}} = 42 \text{ m}\Omega/\square \quad \text{Cu metals 1-6} \quad (10.4)$$

$$R_{sq} = \frac{\rho}{T} = \frac{1.7 \mu\Omega\text{-cm}}{0.8 \mu\text{m}} = 21 \text{ m}\Omega/\square \quad \text{Cu metals 7-8}$$

Given these parameters, the actual resistance of a segment of wire is:

$$R = R_{sq} \left(\frac{L}{W} \right) \quad (10.5)$$

The ratio of L/W is referred to as the *number of squares* of wire. It is the aspect ratio of the wire in terms of length and width. Designers count squares of interconnect as a way of obtaining a relative resistance value for a wire since R_{sq} is fixed for a given technology. As we know by now, the number of squares of resistance has continued to increase over the years since L is increasing and W is decreasing for the longest wires on the chip. Therefore, these are the wires that cause significant RC delay problems.

² As a point of comparison, tungsten plugs have a much higher resistivity of $5.5 \mu\Omega\text{-cm}$.

10.2.2 Elmore Delay Calculation

In long wires, the increased resistance presents a distributed RC load to a driving gate, as shown in Figure 10.4. For simplicity, the driver is represented as an ideal voltage source. The distributed RC line has been converted to a lumped RC ladder structure with n segments. The total resistance for the line is the sum of all the individual resistances, $R_{\text{wire}} = nR_W$. The same holds for the total capacitance, $C_{\text{wire}} = nC_W$.

The wire capacitance in the delay calculation is computed using the capacitance per unit length of interconnect, C_{int} . In modern technologies, there are multiple levels of metal and each one has a different C_{int} . In fact, as we shall see later, interconnect capacitance is distributed in nature and highly dependent on the layout. Today, there is more coupling capacitance between lines than there is capacitance to ground. For the moment, we use an average capacitance value and assume that all layers have the same C_{int} for hand calculations. The effective wire capacitance for metal interconnect is $C_{\text{int}} = 0.2 \text{ fF}/\mu\text{m}$ for densely packed lines ($3\text{--}4\lambda$ apart) and of the order of $C_{\text{int}} = 0.1 \text{ fF}/\mu\text{m}$ for loosely packed lines ($30\text{--}40\lambda$ apart). We will often use $C_{\text{int}} = 0.2 \text{ fF}/\mu\text{m}$ for the capacitance for an interconnect line to be somewhat pessimistic.

To determine the delay of an RC ladder, consider an aluminum wire in $0.18 \mu\text{m}$ technology. Assume a wire width of $W = 0.4 \mu\text{m}$, and $C_{\text{int}} = 0.2 \text{ fF}/\mu\text{m}$ and $R_{sq} = 0.054 \Omega/\square$. Then for a length of wire 1mm , the total resistance is $R_{\text{wire}} = 0.054 \Omega/\square \times 1000/0.4 = 125 \Omega$ and the total capacitance is $C_{\text{wire}} = 0.2 \text{ fF}/\mu\text{m} \times 1000 \mu\text{m} = 200 \text{ fF}$. It is tempting to compute the total RC delay as $0.7 R_{\text{wire}} C_{\text{wire}} = (0.7)125 \Omega \times 200 \text{ fF} = 17.5 \text{ ps}$. However, the actual delay is closer to 10 ps . Clearly, we need an accurate way to compute the delay of such RC ladders.

By multiplying R_{wire} and C_{wire} together, we assumed that the response was exponential in nature, as shown in Figure 10.5a. In actual fact, the response of a distributed RC line is given in the graph of Figure 10.5b. The key to developing a convenient way of computing the delay for a ladder network is to examine the area under the two graphs and establish a relationship between the two responses.

The area underneath the curve in Figure 10.5a is called the *first moment of the impulse response* and is given by

$$\text{Area} = \int_0^\infty e^{-t/RC} dt = R_{\text{wire}} C_{\text{wire}}$$

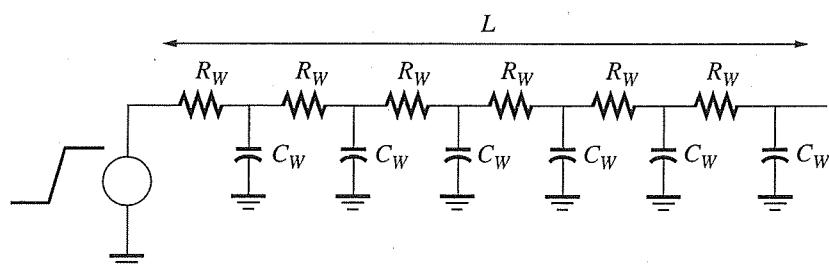
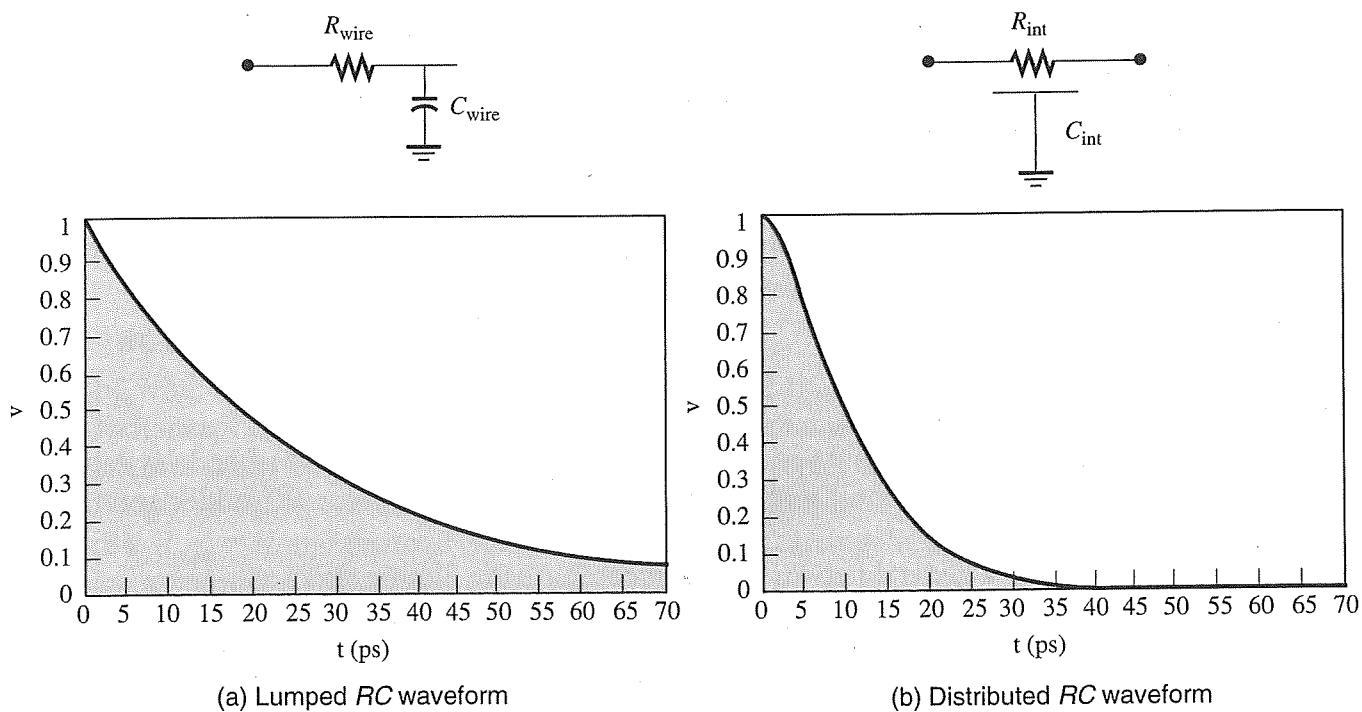


Figure 10.4

Distributed RC line as a lumped RC ladder.

**Figure 10.5**

Lumped RC versus distributed RC output waveforms.

This gives us exactly the time constant for a lumped RC circuit. Apparently the area under the curve provides first-order delay information. If we apply this observation to the waveform for the distributed RC given by Figure 10.5b, we find that the area under this curve is some value which we call τ given by

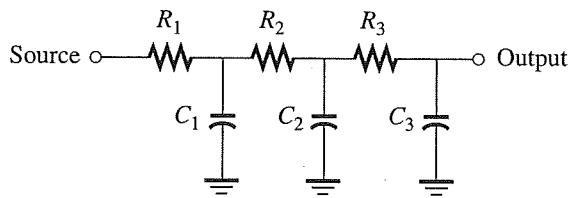
$$\text{Area} = \int_0^\infty v(t) dt = \tau$$

This quantity is also the first moment of the impulse response but is more commonly referred to as the *Elmore delay* by circuit designers. It can be shown that this is the dominant time constant of the network and is a good estimate of the delay. While this value is difficult to compute in general, it is fairly straightforward for RC networks. The Elmore delay for an RC ladder involves visiting each node in the circuit and multiplying the capacitance at that node by the sum of all the resistances from the source node to the node being visited.

To further illustrate this procedure, consider the circuit in Figure 10.6. It has an Elmore delay of

$$\tau = R_1 C_1 + (R_1 + R_2) C_2 + (R_1 + R_2 + R_3) C_3$$

We can examine two extreme cases of this delay value to see if the Elmore delay is consistent. Specifically, if $C_1 = C_2 = 0$, then the Elmore delay and the RC time constant are equal to $(R_1 + R_2 + R_3)C_3$. Similarly if $R_2 = R_3 = 0$, then the Elmore delay and the RC time constant are both $R_1(C_1 + C_2 + C_3)$. Since both cases are

**Figure 10.6**

RC ladder for Elmore delay calculation.

correct, we have some confidence the Elmore delay is a reasonable delay estimator. In fact, it is the most popular form of delay calculation for *RC* ladder and tree circuits.

For a general network, we can compute the Elmore delay as

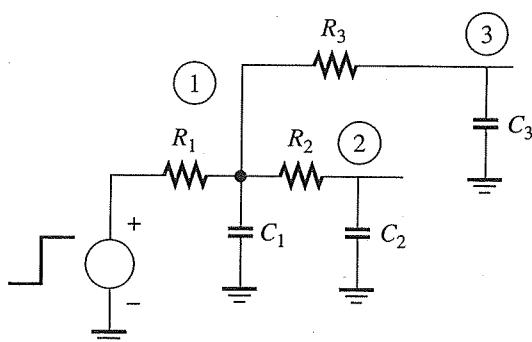
$$\tau_i = \sum_k (C_k \times R_{ik}) \quad (10.6)$$

where the node of interest is node i , C_k is the capacitance at node k , and R_{ik} is the sum of all the *resistances in common* from the source to node i and the source to node k . In other words, we visit each node and multiply the capacitance at that node by a resistance value that is the sum of the resistances that are common when we draw two paths: one from the source to node i and another from the source to node k . The process is best illustrated with an example.

Example 10.1 Elmore Delay of a Simple *RC* Tree

Problem:

Compute the Elmore delay from the input to nodes 1, 2, and 3 in the following *RC* tree. The Elmore delay is not accurate for internal nodes such as node 1. However, we are not usually interested in the delay from the input to an internal node in *RC* trees. In any case, write the expression for the time constant due to Elmore for each node.



Solution:

$$\tau_1 = R_1 C_1 + R_1 C_2 + R_1 C_3$$

$$\tau_2 = C_1 R_1 + C_3 R_1 + C_2 (R_1 + R_2)$$

$$\tau_3 = C_1 R_1 + C_2 R_1 + C_3 (R_1 + R_3)$$

10.2.3 RC Delay in Long Wires

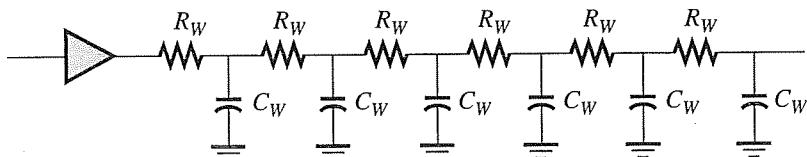
For a long wire, we can compute the *RC* delay in terms of the total resistance and total capacitance. Consider Figure 10.4 once again. If the total wire length is L , and each segment is ΔL in length, then $L = n\Delta L$. For convenience, set the wire resistance per unit length to $R_{\text{int}} = r$ and the wire capacitance unit length to $C_{\text{int}} = c$. The total resistance is $R_{\text{wire}} = rL$ and the total capacitance is $C_{\text{wire}} = cL$. Using the Elmore delay calculation assuming n segments, we obtain

$$\begin{aligned}\tau_{\text{Elmore}} &= (r\Delta L)(c\Delta L) + 2(r\Delta L)(c\Delta L) + \cdots + n(r\Delta L)(c\Delta L) \\ &= (\Delta L)^2 rc(1 + 2 + \cdots + n) \\ &= (\Delta L)^2 rc(n)(n + 1)/2 \\ &\approx (\Delta L)^2 rcn^2/2 = L^2 rc/2 = R_{\text{wire}} C_{\text{wire}}/2\end{aligned}\tag{10.7}$$

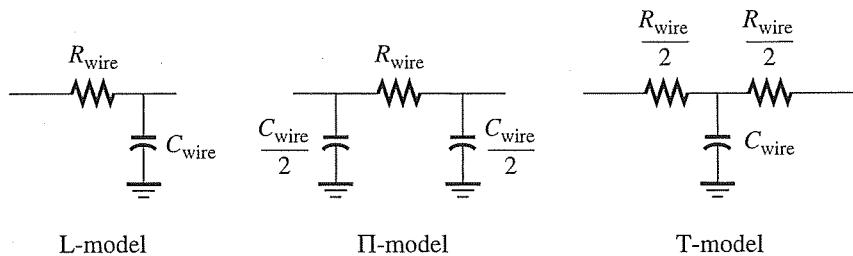
The actual propagation delay value is closer to $0.38R_{\text{wire}}C_{\text{wire}}$ which is slightly less than what Elmore predicts, so Elmore can be viewed as being somewhat pessimistic. Note that, according to Equation (10.7), the *delay increases as the square of the length*. This alerts us to the fact that long wires will create significant delays.

In practice, there must be a buffer or inverter driving the interconnect as shown in Figure 10.7. The driver can be represented as a source voltage and a source resistance. The source voltage is either Gnd or V_{DD} depending on the direction of the switching event. The source resistance is the on-resistance of the buffer.

We should always compute the total delay with the buffer attached to the interconnect. In addition, we must include all fanout loading capacitances on the interconnect for accurate delay estimation. To simplify the calculation, we replace the distributed *RC* line with a lumped *RC* model after computing R_{wire} and C_{wire} of the line. There are three forms of *RC* models as shown in Figure 10.8. While they all

**Figure 10.7**

Driver and interconnect.

**Figure 10.8**

Three possible lumped RC models.

appear to be similar in form, they are not all suitable for the purposes of delay calculation. We can evaluate each one by computing its corresponding time constant.

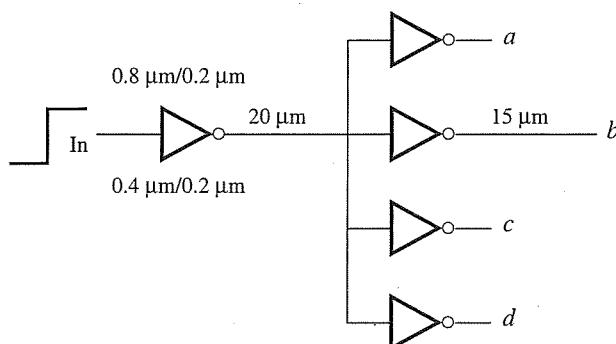
First, the L-model produces a time constant of $R_{\text{wire}} C_{\text{wire}}$. We already know from Equation (10.7) that the delay should be $R_{\text{wire}} C_{\text{wire}}/2$ so this model should not be used. The second model is the π -model where the capacitance has been divided in half and assigned to the input and output. The time constant is $R_{\text{wire}} C_{\text{wire}}/2$. The third model is a T-model where the resistance has been divided in half and assigned to the input and output. Here, the time constant is also $R_{\text{wire}} C_{\text{wire}}/2$. Therefore, the L-model is not an accurate model for long distributed RC lines, whereas the π -model and the T-model produce the correct answer, so they are both suitable for the delay calculation. The T-model has an extra node that may increase the number of calculations. As a result, the π -model is the most popular lumped model for a distributed RC line.

Example 10.2 Delay with and without Resistance Effects for Short Wires

Problem:

What is the propagation delay through the first inverter shown below, ignoring the wire resistance? Assume all inverters are equal in size. Include all capacitances and use $0.18 \mu\text{m}$ technology parameters. Recompute the delay including interconnect resistance. Assume that the metal is Al and that the width of the wire is $0.2 \mu\text{m}$. Is wire resistance significant?

Parameters: $C_{\text{eff}} = 1 \text{ fF}/\mu\text{m}$, $C_{\text{int}} = 0.2 \text{ fF}/\mu\text{m}$, $C_g = 2 \text{ fF}/\mu\text{m}$.



Solution:

Fanout capacitance: inverter is driving four identical inverters.

$$C_{\text{fanout}} = 4 \times C_g(W_N + W_P) = 4(2 \text{ fF}/\mu\text{m})(0.8 \mu\text{m} + 0.4 \mu\text{m}) = 9.6 \text{ fF}$$

Self-capacitance:

$$C_{\text{self}} = C_{\text{eff}}(W_N + W_P) = (1 \text{ fF}/\mu\text{m})(0.8 \mu\text{m} + 0.4 \mu\text{m}) = 1.2 \text{ fF}$$

Wire capacitance:

$$C_{\text{wire}} = C_{\text{int}}L_{\text{wire}} = (0.2 \text{ fF}/\mu\text{m})(20 \mu\text{m}) = 4 \text{ fF}$$

Total capacitance:

$$C_{\text{load}} = C_{\text{fanout}} + C_{\text{self}} + C_{\text{wire}} = 9.6 + 1.2 + 4 = 14.8 \text{ fF}$$

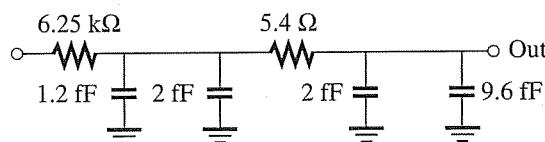
Based on this total capacitance, the delay assuming a ramp input is given by

$$\begin{aligned} \tau_{\text{driver}} &= R_{\text{eff}}C_{\text{load}} = (12.5k)\left(\frac{L}{W}\right)(14.8 \text{ fF}) \\ &= (12.5k)\left(\frac{1}{2}\right)(14.8 \text{ fF}) = 6.25k(14.8 \text{ fF}) = 92.5 \text{ ps} \end{aligned}$$

To include wire resistance, we first compute the resistance for the 20 μm wire:

$$R_{\text{wire}} = R_{\text{sq}}\left(\frac{L}{W}\right) = 54 \text{ m}\Omega/\mu\text{m} \times \left(\frac{20 \mu\text{m}}{0.2 \mu\text{m}}\right) \approx 5.4 \Omega$$

Next, use π -model for the wire:



Use Elmore delay calculation:

$$\begin{aligned} \tau_{\text{driver}} &= (6.25 \text{ k}\Omega)(3.2 \text{ fF}) + (6.25 \text{ k}\Omega + 5.4 \Omega)(11.6 \text{ fF}) \\ &= 92.6 \text{ ps} \end{aligned}$$

Conclusion: the inclusion of wire resistance made no appreciable difference to the solution.

The above example illustrates that short wires can (and should) be handled as a lumped capacitance with little or no error in the calculations. The definition of a short wire depends on the technology but typically wires that are a few microns long or as high as 20 μm should be handled as a lumped capacitance. Over 90% of the wires in a chip are short wires and, therefore, can be computed in this way.

While the analysis above is appropriate when the length of interconnect is relatively short, it is not appropriate when the interconnect line is long. In this case,

the wire should be treated as a distributed RC line rather than a simple lumped capacitance. Consider the example below that illustrates the impact of a long wire.

Example 10.3 Delay for Inverter Driving a Long Wire

Problem:

Consider an isolated metal 5 Al wire that is 20 mm long and driven by a 100X inverter (i.e., 100 times larger than the minimum size inverter). Since the wire is isolated, its capacitance is $0.1 \text{ fF}/\mu\text{m}$. Estimate the total line resistance and the total line capacitance assuming that the width of the wire is $0.5 \mu\text{m}$. What is the propagation delay as calculated by Elmore for the π -model from the input of the inverter to the end of the wire?

Solution:

The wire resistance is

$$R_{\text{wire}} = R_{\text{int}}L = \left(\frac{0.027 \Omega/\square}{0.5 \mu\text{m}} \right) (20,000 \mu\text{m}) = 1080 \Omega$$

The wire capacitance is

$$C_{\text{wire}} = C_{\text{int}}L = (0.1 \text{ fF}/\mu\text{m})(20,000 \mu\text{m}) = 2 \text{ pF}$$

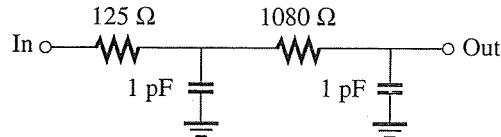
The inverter on-resistance is

$$R_{\text{eff}} = R_{\text{eqn}}/100 = 12.5 \text{ k}\Omega/100 = 125 \Omega$$

The inverter output capacitance is

$$C_{\text{self}} = C_{\text{eff}}(2W + W)100 = (1 \text{ fF}/\mu\text{m})(0.6 \mu\text{m})100 = 60 \text{ fF}$$

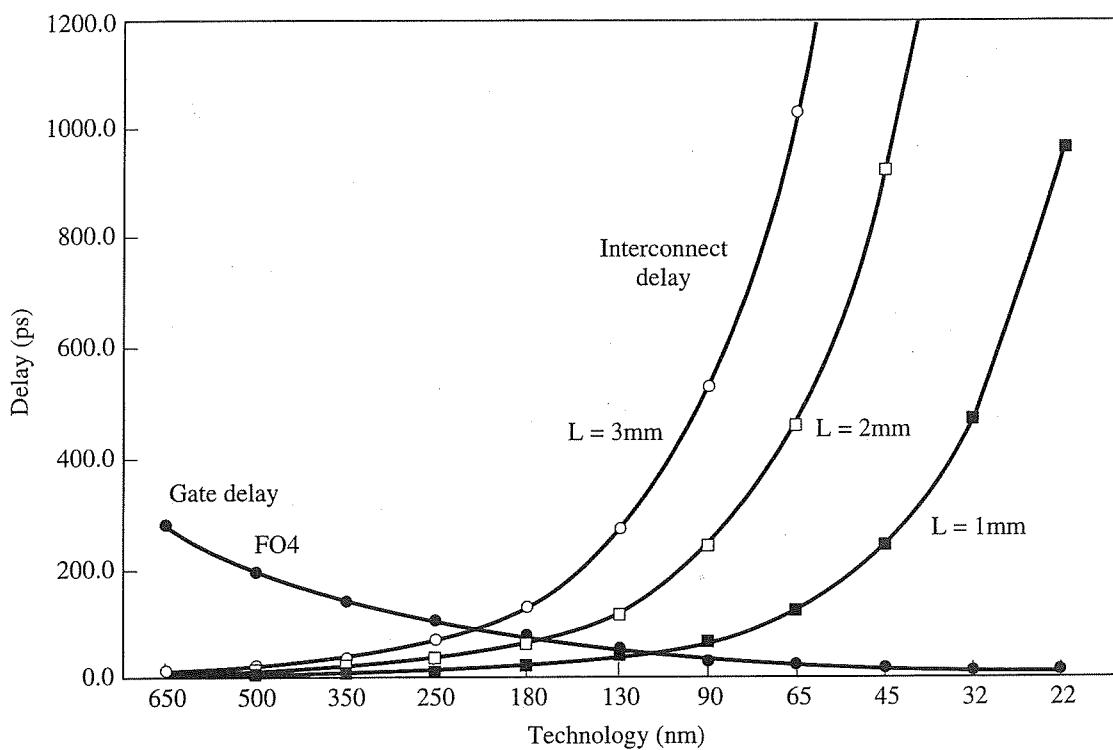
We can ignore the junction capacitance since it is rather small. Using the π -model, the wire capacitance is split in half while the full wire resistance is used in the model.



The resulting Elmore delay is

$$\tau_{\text{Elmore}} = (125 \Omega)(1 \text{ pF}) + (125 \Omega + 1080 \Omega)(1 \text{ pF}) \approx 1.3 \text{ ns}$$

In order to appreciate the large delays associated with long wires, we plot the FO4 gate delay and the wire delay for a number of different wire lengths as technology scales in Figure 10.9. As reference points, the FO4 delay for 350 nm is approximately 150 ps; for 130 nm, the FO4 delay is reduced to approximately 40 ps; for 65 nm, the value reduces to about 25 ps. On the other hand, for $L = 1 \text{ mm}$ the interconnect delay increases from 3 ps, to 27 ps, to 116 ps, respectively. For $L = 2 \text{ mm}$ the delay

**Figure 10.9**

Gate delay versus interconnect delay.

values are 12 ps, 110 ps, and 460 ps, respectively. For $L = 3$ mm the delay values are 30 ps, 240 ps, and 1035 ps, respectively. The wire delay doubles with each technology node and increases quadratically as a function of wire length. Clearly, when a single buffer drives a long wire, the interconnect delay will eventually dominate the FO4 delay as technology scales or as the length increases.

We can define a rule of thumb for long wires by setting the FO4 delay equal to the interconnect delay in each technology. A critical wire length can be derived as follows:

$$\text{FO4} = 0.38R_{\text{int}}C_{\text{int}}L_{\text{crit}}^2$$

$$\therefore L_{\text{crit}} = \sqrt{\frac{\text{FO4}}{0.38R_{\text{int}}C_{\text{int}}}}$$

For example, in a 0.18 μm technology, the value of L_{crit} can be computed using this equation or estimated from the graph in Figure 10.9 to be approximately 2.5 mm. Of course, this value will vary depending on the values of R_{int} and C_{int} for a given wire.

10.3 Buffer Insertion for Very Long Wires

For very long wires, the quadratic delay characteristics described above cannot be tolerated in any design. The interconnect delay would be so long that it would dominate the gate delay. In ICs, global wires tend to be very long since they must be distributed

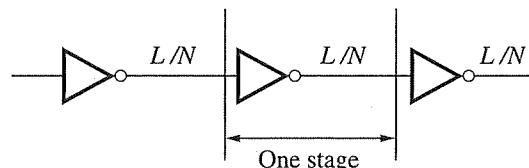
all over the chip. These wires are associated with clocks, busses, and other major signals in the design. Since the delay for a long wire is quadratic with respect to L , the standard solution is to insert *repeaters* or *buffers* periodically along the wire. This effectively shortens the wire into smaller segments. The buffers act to convert the quadratic delay to a more linear delay. Of course, if we insert too many buffers, the delay increases due to the buffers themselves. But with too few buffers, we are prone to the quadratic effect. Finding the optimum number of equal-sized buffers needed to minimize the overall delay is referred to as the *buffer insertion problem*.

Consider a wire of length L that has an unacceptably long delay without any buffers driving it. If the per unit length resistance and capacitance are given by R_{int} and C_{int} , respectively, the total resistance and capacitance of the wire is given by

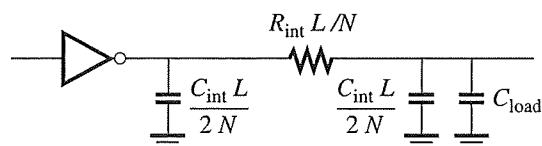
$$R_{\text{wire}} = \frac{\rho}{T} \frac{L}{W} = R_{\text{sq}} \frac{L}{W} = R_{\text{int}} L$$

$$C_{\text{wire}} = C_{\text{int}} L$$

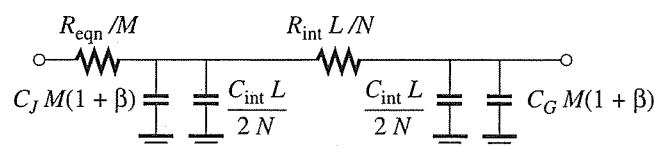
To reduce the delay we seek to insert a number of buffers that produce the minimum delay. Let N be the number of such buffers (including the first buffer) and L/N be the corresponding wire lengths between buffers as shown in Figure 10.10a. Each segment consists of a buffer driving a wire load and a downstream gate load. A π -model may be used for the interconnect in one segment as shown in Figure 10.10b.



(a) Wire of length L with N buffers inserted



(b) One segment of buffer and interconnected



(c) RC model for one segment

Figure 10.10

Buffer insertion for very long wires.

Assuming that a minimum size buffer is 1X, the corresponding input and output capacitances are as follows:

$$C_{\text{out}} = C_{\text{eff}}W(1 + \beta) = C_J(1 + \beta) \quad (\text{in fF})$$

$$C_{\text{in}} = C_gW(1 + \beta) = C_G(1 + \beta) \quad (\text{in fF})$$

Here, β is the ratio of the PMOS to NMOS device size. The value of C_J is $C_{\text{eff}} \times W$, where W is the size of the NMOS device for a 1X inverter. Similarly, C_G is $C_g \times W$.

Let the multiplier M represent the optimal buffer size for the buffer insertion problem. Then, for a buffer which is M times larger than a 1X buffer:

$$R_{\text{eff}} = R_{\text{eqn}}/M$$

$$C_{\text{out}} = C_{\text{eff}}W(1 + \beta)M = C_J(1 + \beta)M$$

$$C_{\text{in}} = C_gW(1 + \beta)M = C_G(1 + \beta)M$$

In Figure 10.10c, we illustrate the RC model used for each stage during delay calculation. Using the Elmore delay, the delay for each segment is given by

$$\begin{aligned} t_{\text{segment}} &= \frac{R_{\text{eqn}}}{M} \left(C_JM(1 + \beta) + \frac{C_{\text{int}}L}{2N} \right) \\ &\quad + \left(\frac{R_{\text{eqn}}}{M} + \frac{R_{\text{int}}L}{N} \right) \left(\frac{C_{\text{int}}L}{2N} + C_GM(1 + \beta) \right) \end{aligned}$$

The total delay is N times the segment delay. After some rearrangement, we find that it produces a quadratic equation with respect to L .

$$\begin{aligned} t_{\text{total}} &= N \times t_{\text{segment}} = N(C_G + C_J)R_{\text{eqn}}(1 + \beta) \\ &\quad + \left(C_G(1 + \beta)R_{\text{int}}M + \frac{C_{\text{int}}R_{\text{eqn}}}{M} \right)L + \left(\frac{C_{\text{int}}R_{\text{int}}}{2N} \right)L^2 \quad (10.8) \end{aligned}$$

When we examine each delay term, we find that the first term is due solely to the buffer while the last term is due solely to the interconnect. The middle term has two components: one is due to the wire resistance driving the input capacitance of the next buffer, and a second RC term is due to the buffer driving the wire capacitance.

To find the optimal values for N and M , we apply the derivative with respect to each of the variables, as follows. The number of wire segments is

$$\begin{aligned} \frac{\partial t_{\text{total}}}{\partial N} &= 0 = R_{\text{eqn}}(C_J + C_G)(1 + \beta) - \frac{R_{\text{int}}C_{\text{int}}L^2}{2N^2} \\ \therefore N &= \sqrt{\frac{R_{\text{int}}C_{\text{int}}L^2/2}{R_{\text{eqn}}(C_J + C_G)(1 + \beta)}} = \sqrt{\frac{t_{\text{wire}}}{\text{FOI}}} \quad (10.9) \end{aligned}$$

The buffer size, assuming all buffers are equal, is

$$\frac{\partial t_{\text{total}}}{\partial M} = 0 = R_{\text{int}}LC_G(1 + \beta) - \frac{R_{\text{eqn}}C_{\text{int}}L}{M^2}$$

$$\therefore M = \sqrt{\frac{R_{\text{eqn}}}{C_G(1 + \beta)}} \frac{C_{\text{int}}}{R_{\text{int}}} \quad (10.10)$$

Using these values of N and M , we can obtain the optimal delay through the interconnect. We can further understand these values by examining the total delay equation provided in Equation (10.8). The variable N exists in the first and last term of t_{total} ; the terms are due to the buffer and interconnect, respectively, each one acting as if the other did not exist. If we set these two terms to be equal to one another, we arrive at the equation for the optimal N . The expression under the square-root sign in Equation (10.9) can be viewed as the interconnect delay (t_{wire} according to Elmore) divided by the inverter delay when driving an identical inverter (FO1 delay).

The middle term of the total delay expression in Equation (10.8) is dependent on M . If we set the two expressions in the middle term to be equal to one another, we arrive at the optimal M . This value does not depend on the length of the wire but rather the physical characteristics of the transistors and wires. In fact, this quantity depends only on technology parameters of the wire and the minimum-size inverter. Therefore, we can compute this value once the process parameters have been determined. Note that precise values of the terms in the expressions are needed to produce accurate results.

It is useful to compare and contrast the buffer insertion problem to the logical (LE) effort problem described in Chapter 6. For buffer insertion, we are concerned with driving a long wire so we break up the wire into segments and insert very large equal-sized buffers between the wire segments. For logical effort, the objective is to drive a very large output capacitance starting from a small input capacitance. The solution is obtained by determining the number of buffers and then sizing up the buffers by a given factor, typically by 4X. In LE, the delay through each buffer stage is the same at the optimal solution. In the case of buffer insertion, the delay due to the buffer and the wire are the same at the optimal point.

Example 10.4

Buffer Insertion Example

Problem:

From Example 10.3, consider the 20 mm Al metal 5 wire again, but this time reduce the delay using optimal buffer insertion. What is the new delay through the wire after buffer insertion?

Solution:

Using the following values:

$$R_{\text{int}} = 54 \text{ m}\Omega/\mu\text{m} \quad C_{\text{int}} = 0.1 \text{ fF}/\mu\text{m} \quad L = 20,000 \mu\text{m}$$

$$R_{\text{eqn}} = 12.5 \text{ k}\Omega/\square \quad C_g = 2 \text{ fF}/\mu\text{m} \quad C_{\text{eff}} = 1 \text{ fF}/\mu\text{m} \quad W = 0.2 \mu\text{m} \quad \beta = 2$$

The number of segments in the wire is

$$N = \sqrt{\frac{R_{\text{int}} C_{\text{int}} L^2 / 2}{R_{\text{eqn}} (C_G + C_J) (1 + \beta)}} = 6.9 \approx 7$$

The buffer size, assuming all equal buffers, is

$$M = \sqrt{\frac{R_{\text{eqn}} C_{\text{int}}}{C_G (1 + \beta) R_{\text{int}}}} \approx 140$$

Then, the buffer resistance is

$$R_{\text{eff}} = \frac{12.5k\Omega}{140} \approx 90\Omega$$

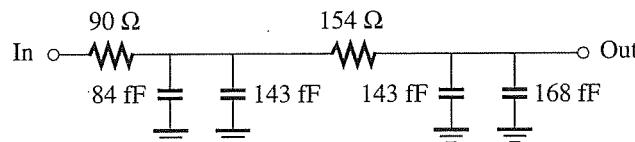
The buffer output capacitance is:

$$C_{\text{self}} = C_{\text{eff}}(2W + W)140 = (1\text{fF}/\mu\text{m})(0.6\text{\mu m})140 = 84\text{fF}$$

The buffer input capacitance is

$$C_{\text{fanout}} = C_g(2W + W)140 = (2\text{fF}/\mu\text{m})(0.6\text{\mu m})140 = 168\text{fF}$$

Each segment produces this RC delay model.



The total Elmore delay is

$$\begin{aligned}\tau_{\text{Elmore}} &= 7 \times [(90\Omega)(84\text{fF} + 143\text{fF}) + (90\Omega + 154\Omega)(143\text{fF} + 168\text{fF})] \\ &= 0.67\text{ ns}\end{aligned}$$

The total delay has been cut in half using buffer insertion.

10.4 Interconnect Coupling Capacitance

In this section, the issues associated with coupling capacitance are described. The goal is to understand the various components of wire capacitance and to develop formulas for hand calculation. In reality, IC interconnect structure is three-dimensional in nature as seen in Figures 1.1 and 1.2 of Chapter 1. However, for hand calculations, we can develop simple formulae that work well in practice. Thus far, we have been using $0.2\text{ fF}/\mu\text{m}$ for wire capacitance without really knowing its origin or range of validity. This section attempts to clarify its origin and justify the use of this value from a modeling perspective.

10.4.1 Components of Coupling Capacitance

In modern integrated circuits, interconnect capacitance modeling is a complex task. The capacitance of such wires is dependent on their topology, on the distances from other wires above or below, and on the separation between wires. To acquire a basic

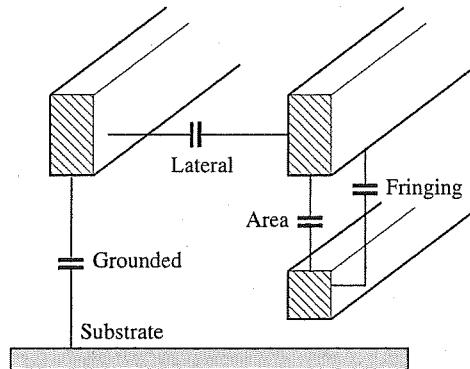


Figure 10.11

Area, lateral, and fringing components.

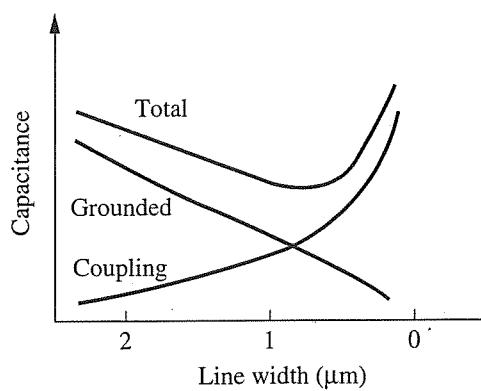
understanding of the problem, we can use simple first-order models to describe interconnect capacitances and derive useful constants for design purposes.

Before beginning the analysis, we describe the different components of capacitance. These components are all shown in Figure 10.11. The first capacitance worth noting is between a line and the substrate, called the *grounded* capacitance. This was the dominant capacitance in the past, but its value has been diminishing due to scaling effects. The next capacitance to note is between metal lines on the same layer. This is called the *lateral* capacitance. This is the main coupling capacitance that is the source of delay and noise issues. This capacitance has been increasing over the years. There are two other capacitances that may give rise to coupling between lines. One is referred to as the *fringing* capacitance which occurs between various edges and surfaces between two crossing lines. One of many such fringing capacitances is shown in the figure.³ The other type of coupling capacitance occurs when two wires on different metal layers overlap. It is called the *area* capacitance.

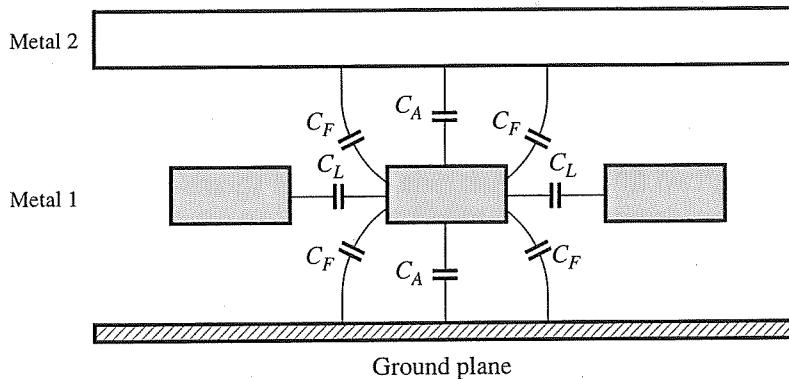
Over the years, coupling capacitance has been increasing while the grounded capacitance has been decreasing. The general trend is depicted in Figure 10.12, where the coupling capacitance due to area, lateral, and fringing components is overtaking the grounded component as the line widths shrink. In a typical $0.13 \mu\text{m}$ design, over 70% of the wire capacitance at a node may be due to coupling between wires. This is usually the case for long nets, although short nets may also have significant coupling to their neighbors. Most nets have significant junction and gate capacitances due to the logic gates that increase the overall grounded capacitance. However, the coupling to other nodes is still significant.

To obtain a quantitative view of coupling capacitance, consider the interconnect structure shown in Figure 10.13. We would like to determine the capacitance of the center wire on a component basis so that it can be used for delay and noise calculations. The capacitances associated with the middle conductor are given in Figure 10.13. The dimensions of the wires were shown earlier in Figure 10.2. The total values of area, lateral, and fringing capacitances are all indicated in the figure. The grounded capacitance is treated as an area capacitance since the

³ Actually, there are many other fringe capacitances that are not shown in the figure to keep it simple.

**Figure 10.12**

Coupling capacitance, grounded capacitance, and total capacitance.

**Figure 10.13**

Capacitances associated with middle conductor.

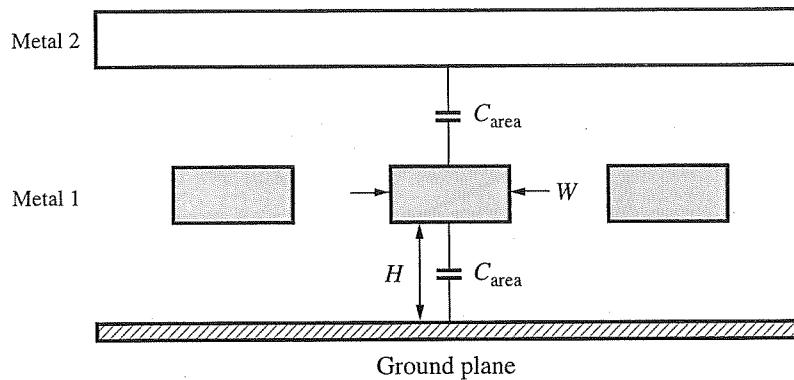
calculation will be the same for both. While the capacitance cannot be exactly partitioned as shown, we model it this way for convenience. Once computed, we sum them together to compute the total wire capacitance. We now compute each one in units of capacitance per unit length.

The easiest component to compute is the area capacitance per unit length. There are two such capacitances, one from the conductor to ground and the other from the conductor to Metal 2 above it. This is illustrated separately in Figure 10.14. Assuming silicon dioxide with a dielectric constant of 4 is used as the insulating material, each of these capacitances is computed as

$$C_{\text{area}} = \epsilon_{\text{ox}} \frac{W}{H} \quad (10.11)$$

The numerical value of this capacitance is

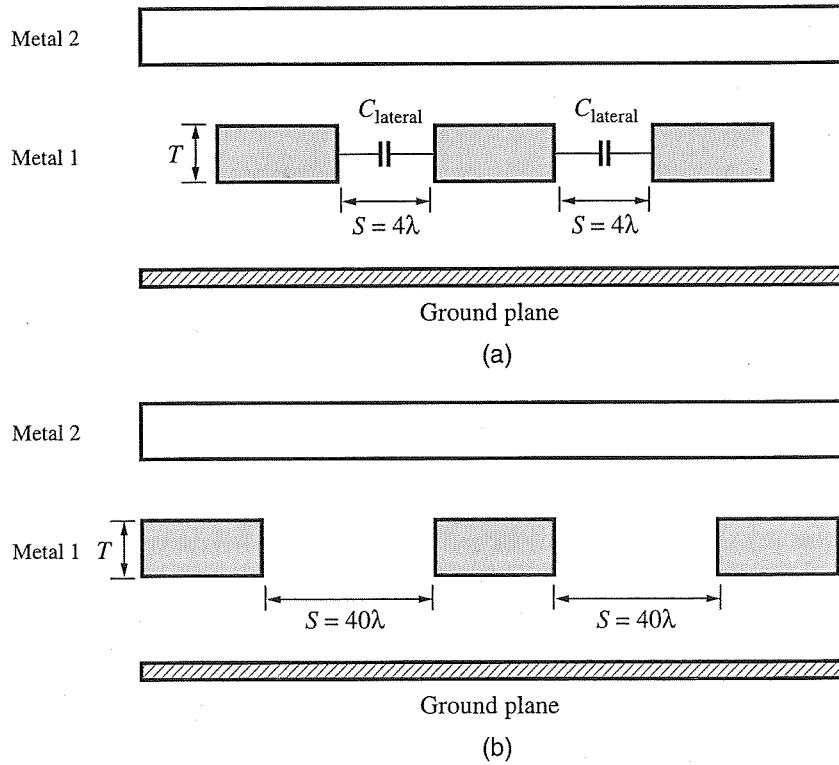
$$C_{\text{area}} = \epsilon_{\text{ox}} \frac{W}{H} = 4\epsilon_0 \frac{W}{H} = 4(88.5 \times 10^{-4} \text{ fF}/\mu\text{m}) \frac{W}{H} = 0.035 \frac{W}{H} \text{ fF}/\mu\text{m}$$

**Figure 10.14**

Modeling area capacitance.

As W shrinks in value with technology scaling, the area capacitance reduces linearly. In $0.18 \mu\text{m}$ technology, the upper layer of metal would have $W = 0.4 \mu\text{m}$ and $H = 0.5 \mu\text{m}$. Therefore, C_{area} is approximately $0.03 \text{ fF}/\mu\text{m}$.

Next, we examine the lateral capacitance. This value can be complicated to compute but there are two limiting situations that lend themselves to a straightforward formula. These two situations are shown in Figure 10.15. When the conductors are

**Figure 10.15**

Modeling lateral capacitance.

very closely spaced by 4λ , as shown in Figure 10.15a, the capacitance can be computed using the parallel-plate formula:

$$C_{\text{lateral}} = \epsilon_{\text{ox}} \frac{T}{S} \quad (10.12)$$

The capacitance is inversely proportional to the spacing, S . Applying the dielectric constant value to the equation:

$$C_{\text{lateral}} = \epsilon_{\text{ox}} \frac{T}{S} = 4\epsilon_0 \frac{T}{S} = 4(88.5 \times 10^{-4} \text{ fF}/\mu\text{m}) \frac{T}{S} = 0.035 \frac{T}{S} \text{ fF}/\mu\text{m}$$

In our generic $0.18 \mu\text{m}$ technology, the upper layer metal has $T = 0.8 \mu\text{m}$ and $S = 0.4 \mu\text{m}$. Therefore, C_{lateral} is $0.070 \text{ fF}/\mu\text{m}$. When the conductors are spaced out by 40λ as in Figure 10.15b, then C_{lateral} is $0.007 \text{ fF}/\mu\text{m}$ and can effectively be ignored.

The last capacitance to consider is the fringing capacitance. Again, its complexity prohibits any sort of simplified analysis. However, for hand calculations, it is sufficient to consider two cases as shown in Figure 10.16. The first case in Figure 10.16a consists of densely packed wires where the lateral capacitance dominates. For this case, C_{fringe} is approximately zero. However, if the conductors are widely spaced, as in Figure 10.16b, there is a fringe capacitance on each edge of approximately

$$C_{\text{fringe}} = \epsilon_{\text{ox}} \ln\left(1 + \frac{T}{H}\right) \quad (10.13)$$

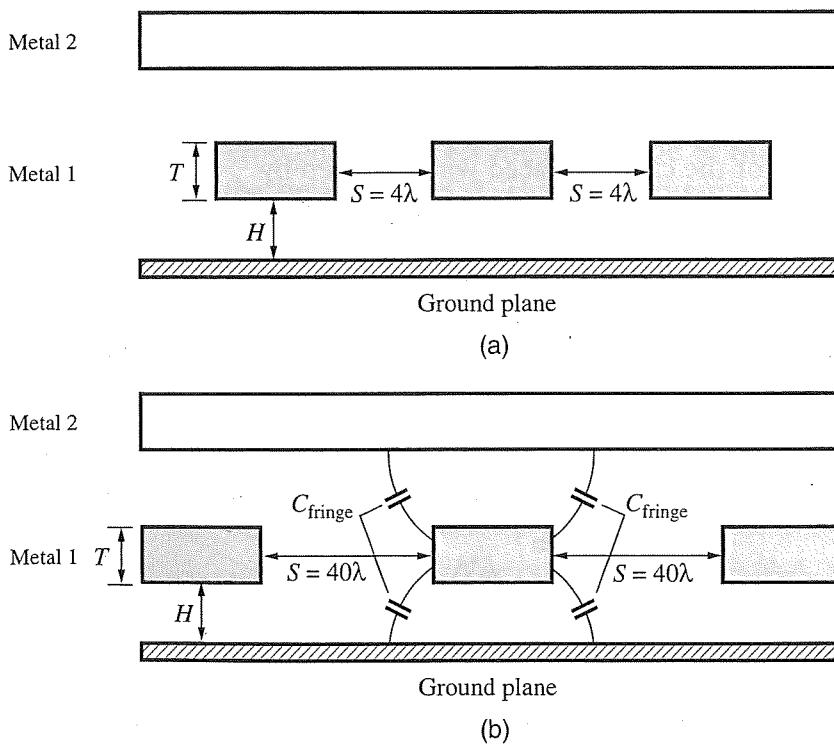


Figure 10.16

Modeling fringing capacitance.

Since T and H are fixed by the technology, this capacitance is almost independent of the topology. While this is not strictly true, it is sufficient for our modeling purposes here. We can compute the value of the fringing component as follows:

$$\begin{aligned} C_{\text{fringe}} &= \epsilon_{\text{ox}} \ln \left(1 + \frac{T}{H} \right) = 4(88.5 \times 10^{-4} \text{ fF}/\mu\text{m}) \ln \left(1 + \frac{0.5}{0.5} \right) \\ &= 0.025 \text{ fF}/\mu\text{m} \end{aligned}$$

Now that each of the components has been derived, we are ready to compute the quantity C_{int} .

The total capacitance for the middle wire depends on whether the wires are closely spaced or widely spaced. The case for the closely spaced wires on the upper layers is given by

$$\begin{aligned} C_{\text{int}} &= 2C_{\text{area}} + 2C_{\text{lateral}} + 2C_{\text{fringe}} \\ &\approx 2(0.03) + 2(0.070) + 2(0.0) \text{ fF}/\mu\text{m} \\ &= 0.2 \text{ fF}/\mu\text{m} \end{aligned}$$

If the wires are widely spaced, we can use the following formulation:

$$\begin{aligned} C_{\text{int}} &= 2C_{\text{area}} + 2C_{\text{fringe}} + 2C_{\text{lateral}} \\ &\approx 2(0.03) + 2(0.025) + 2(0.0) \text{ fF}/\mu\text{m} \\ &\approx 0.1 \text{ fF}/\mu\text{m} \end{aligned}$$

Now consider the closely spaced and widely spaced cases for the lower level metal lines. The case for the closely spaced wires is given by

$$\begin{aligned} C_{\text{int}} &= 2C_{\text{area}} + 2C_{\text{lateral}} + 2C_{\text{fringe}} \\ &= 2(0.015) + 2(0.035) + 2(0.0) \text{ fF}/\mu\text{m} \\ &= 0.1 \text{ fF}/\mu\text{m} \end{aligned}$$

If the wires are widely spaced, we can use the following formulation:

$$\begin{aligned} C_{\text{int}} &= 2C_{\text{area}} + 2C_{\text{fringe}} + 2C_{\text{lateral}} \\ &\approx 2(0.015) + 2(0.025) + 2(0.0) \text{ fF}/\mu\text{m} \\ &= 0.08 \text{ fF}/\mu\text{m} \approx 0.1 \text{ fF}/\mu\text{m} \end{aligned}$$

To summarize, we have considered the 3D capacitances of deep submicron interconnect and found that upper layers of metal typically exhibit total capacitance of about $0.2 \text{ fF}/\mu\text{m}$ for closely spaced lines and $0.1 \text{ fF}/\mu\text{m}$ for widely spaced lines. The

capacitance of wires that are spaced between the two extremes would lie between $0.1 \text{ fF}/\mu\text{m}$ and $0.2 \text{ fF}/\mu\text{m}$ but the calculations are much more complex. However, capacitance extraction tools are available to obtain these components accurately. The lower levels of metal exhibit wire capacitances of about $0.1 \text{ fF}/\mu\text{m}$ in both situations. Individual values of the coupling capacitances may be estimated using Equations (10.11), (10.12), and (10.13). These first-order equations provide valuable insight into the origin and nature of the 3D effects. Worst-case capacitances lie in the range of $0.2 - 0.3 \text{ fF}/\mu\text{m}$. We now consider the effect of coupling capacitance on delay and noise.

10.4.2 Coupling Effects on Delay

The RC equations given in Section 10.3 using the Elmore delay did not properly take the coupling effects into account. A more realistic situation for delay calculation is shown in Figure 10.17 where neighboring lines are coupled together. The grounded capacitances shown include self-capacitance, fanout capacitance, and wire capacitance (to substrate). The coupling capacitances due to area, lateral, and fringing effects are actually distributed along the nets. The fact that the coupling capacitors exist between all neighboring lines implies that delay is a function of not only what is happening on the line of interest, but also what is happening on the neighboring lines.

To understand this effect, consider the case of two adjacent buffers driving their respective lines with a single coupling capacitance, as shown in Figure 10.18. To simplify the problem, we ignore the wire resistance and concentrate on the effects of coupling capacitance on delay. The line of interest is called the *victim* net while the neighboring line is called the *aggressor* for reasons that will become clear in the descriptions that follow. Consider the delay associated with the switching of net 2. It is coupled to net 1 with a large capacitor C_c . How should the coupling capacitance be handled? One is tempted to connect the other side of the capacitance to ground while computing the delay for net 2. However, this may not be correct depending on what is happening at net 1. The actual delay depends on whether net

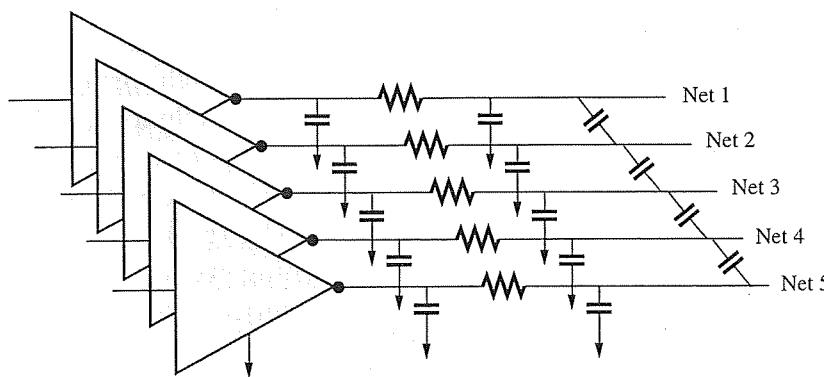
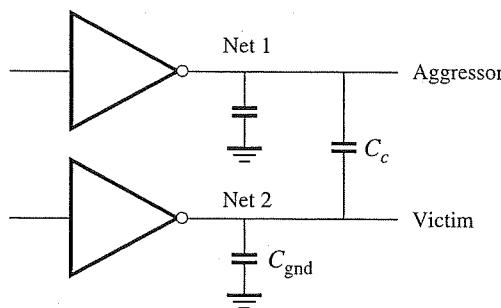


Figure 10.17

Coupling capacitance between lines.

**Figure 10.18**

Circuit for analysis of coupling delay and noise effects.

1 is switching from high to low or low to high, or perhaps not switching at all. If it is not switching at all, then it is appropriate to ground the other side of the capacitance. This would make the total capacitance at net 2 equal to $C_{\text{gnd}} + C_c$. That covers only one of the three possible cases that may occur.

To complete the story, we examine the total loading capacitance, C_L , of net 2 for a number of different cases. Again, if the neighboring net is stationary, then

$$C_L = C_{\text{gnd}} + C_c \quad \text{aggressor not switching} \quad (10.14)$$

since the adjacent net is not switching and the coupling capacitor behaves like a grounded capacitor. If the two nets are switching in the same direction, then

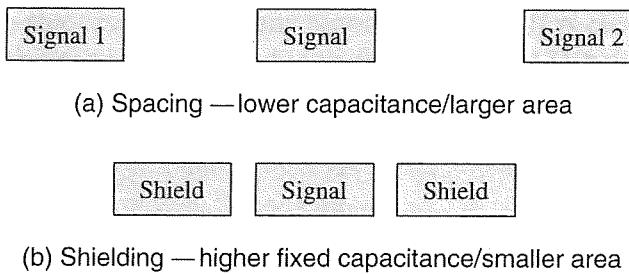
$$C_L = C_{\text{gnd}} \quad \text{both switching together} \quad (10.15)$$

since net 2 does not have to charge the coupling capacitance. However, if net 1 is switching in the opposite direction to net 2, then we should include the so-called Miller effect. Since net 2 must supply twice the charge on the coupling capacitor to compensate for the switching process:

$$C_L = C_{\text{gnd}} + 2C_c \quad \text{both switching opposite} \quad (10.16)$$

Even more combinations are possible if there are more neighboring nets. For example, in Figure 10.17, net 3 is coupled to net 2 and net 4. Obviously, there are many combinations of switching that result in different values of its C_L . In this example, the proper value of C_L must be determined after considering 3^2 possible cases. For n coupled nets, there are 3^n possible combinations. The result is that the delay for a net is unpredictable as it may vary from one switching event to another. This is one of the reasons why timing is so difficult to characterize accurately in deep submicron technologies.

A number of layout-based approaches have been developed to cope with this unpredictable delay, as shown in Figure 10.19. One approach is to *space out* the wires so that the lateral capacitance term decreases. Equation (10.12) tells us that

**Figure 10.19**

Spacing versus shielding to reduce coupling effects.

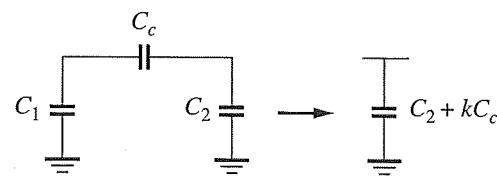
there is a $1/S$ relation between lateral capacitance and spacing, so we can reduce the amount of coupling by *spacing* signals apart from one another as in Figure 10.19a. Of course this will cost area, but it is effective when applied to a subset of the signals. Another approach is to interleave switching signals with nonswitching signals (such as power and ground) wherever possible. This is referred to as *shielding* as in Figure 10.19b. We can place well-behaved signals on either side of a signal so that its delay may be more predictable. However, in this case the total capacitance increases since there will always be a fixed capacitance to ground due to the shields.

Process engineers are also working to reduce the effects of coupling. Typically SiO_2 is used as the insulator in most CMOS processes. It has a relative dielectric constant of approximately 4. If the relative dielectric constant of the insulating material could be reduced from 4 to 2, the coupling capacitance would be cut in half. The pursuit of these so-called *low-k* dielectrics has been ongoing for years. Currently, the relative constant is roughly 3 but there are many issues to resolve to reach the target value of 2.

Timing verification tools can be used to obtain more accurate delay information since they have some notion of when signals switch. These tools require input from capacitance extraction tools that provide detailed coupling information. One way to characterize the effective load capacitance for these tools is to create a pre-multiplier for the coupling capacitance called a *k*-factor.⁴ Each coupling capacitor can be multiplied by this factor to adjust its value according to switching activity. This *k*-factor is nominally in the range 0 to 2.

To handle coupling effects, the timing verifier would ground all coupling capacitances after taking the *k*-factors into account. While this is a simplistic solution, it is one of the more practical approaches to coupling analysis. Figure 10.20 shows this process for a pair of adjacent nets. The load capacitance for net 2 is adjusted by the *k*-factor depending on the switching of net 1. One complication is that the *k*-factors may vary over time due to the switching conditions associated with adjacent nets. Fortunately the designer is primarily interested in the worst-case values of *k* since it

⁴This is not the same as the *k* used in low-*k* dielectrics. This one is used to adjust capacitance values taking into account the switching behavior. To distinguish them, we will use the terms low-*k* and *k*-factor when referencing them.



$k = 0$ switching in same direction

$k = 1$ aggressor not switching

$k = 2$ switching in opposite directions

Figure 10.20

Handling coupling using k -factors.

will produce the worst-case delay values. This also simplifies the analysis significantly because it would be impractical to analyze the circuit exhaustively with case-dependent k -factors. For example, setup and hold time checks are important for the timing verification of flip-flops. When checking setup time, the designer uses the maximum k -factor on signal nets and the minimum on the clock net. The opposite settings are applied for hold time violations.

If this were not complicated enough, it turns out that the exact timing of the aggressor and victim signals also plays a role in the observed delay. Figure 10.21 illustrates the effect of signal timing in the presence of wire coupling in a SPICE transient analysis. The uncoupled waveforms of the victim (net 2) and aggressor nets (net 1) are shown along with a second plot of the victim waveform when many aggressors switch against it. Because of the early timing of the aggressors, the coupling actually causes capacitive feedthrough in the wrong direction before the proper transition begins. The resulting propagation delay in this example doubles

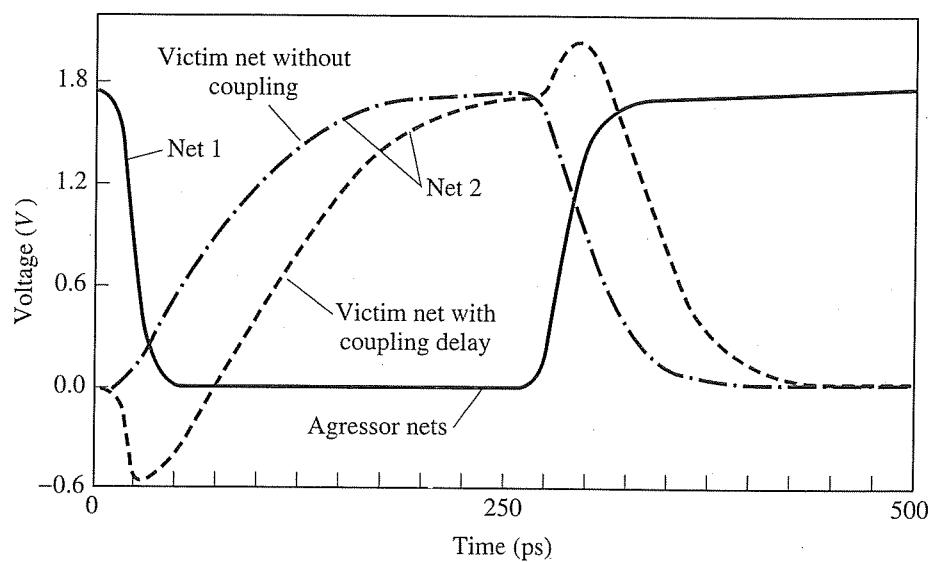


Figure 10.21

Transient analysis of coupled nets.

when compared to the case without any signal coupling. It is clear that the k -factor can lie outside of the range of 0 to 2 due to feedthrough.⁵ For the designer, the use of structured layout to avoid coupling wherever possible, and careful timing verification using k -factors are the two main approaches to cope with this signal integrity problem.

10.4.3 Capacitive Noise or Crosstalk

The second type of problem is noise injection due to capacitive coupling. This problem arises when there is a nonswitching net, again the victim net, surrounded by aggressor nets that are switching. When the aggressor nets switch, noise is injected into the victim line due to capacitive feedthrough effects. We have already observed this effect during a transition in Figure 10.21. Here, we consider the case where the victim line is quiet while aggressor lines are switching and injecting unwanted noise. Depending on the size and duration of the noise, it may act to upset the functionality of a design—that is, a “soft” error may occur. Coupling noise at latch or flip-flop inputs and in dynamic logic circuits is particularly troublesome. Specifically, incorrect values can be captured in flip-flops, or voltage levels at dynamic nodes may be upset due to noise injection. Noise may also propagate through logic gates, assuming that the noise level is high enough. This can cause problems for downstream logic circuits, particularly for dynamic logic. Tracking down this type of problem in a multimillion transistor chip has consumed many engineers for months, often without success. It can also increase the power consumption unnecessarily.

To understand coupling noise, consider Figure 10.22 which depicts our aggressor and victim nets connected only by a coupling capacitance, C_c . The feedthrough equation for the noise injected from net 1 into net 2 is

$$\Delta V_2 = \frac{C_c \Delta V_1}{(C_{\text{gnd}} + C_c)} \quad (10.17)$$

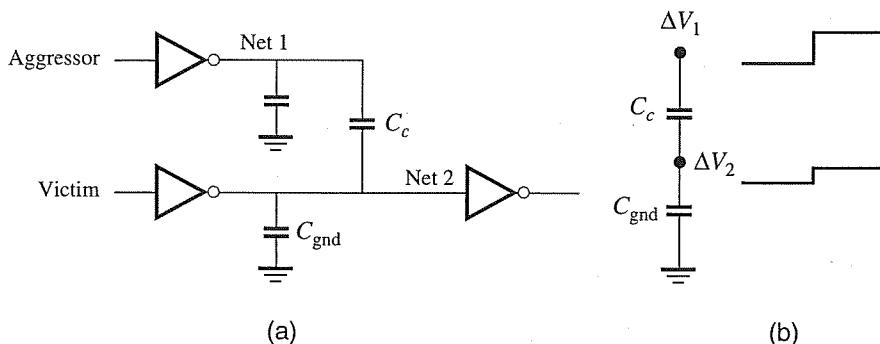


Figure 10.22

Coupling noise analysis.

⁵ The k -factor actually lies in the range of -1 to 3 when feedthrough is taken into account.

where ΔV_1 is the voltage transition at net 1 during a switching event. This equation was described in Chapter 7 on dynamic logic. If the change on the aggressor net is V_{DD} , then the injected noise is

$$\Delta V_2 = \frac{C_c V_{DD}}{(C_{\text{gnd}} + C_c)}$$

This first-order analysis is useful in obtaining an idea of the relative magnitude of noise due to coupling capacitance. To compute the noise level more accurately, a number of other factors must be taken into account such as the driver resistances, wire resistance, switching times of the aggressor signals, the switching direction of the signals, the slew rates of the signals, resistive shielding of driver, etc. Of course, once the noise is injected, the victim acts to restore the original level as quickly as possible. Complex situations that include these factors require detailed transient simulation in SPICE to obtain accurate results. Linear RC simulators are also useful for this purpose.

In practice, designers try to minimize coupling to neighboring signals to reduce noise. This can be accomplished by spacing the signals apart to reduce the C_c term, or by shielding signals with V_{DD} or Gnd, thereby eliminating the C_c term. These two approaches were illustrated in Figure 10.19.

To summarize, the delay and noise effects are difficult to analyze by hand, except for the smallest circuits. Designers must use structured design methodologies to avoid coupling effects as much as possible, especially when critical nets are involved. CAD tools must be used extensively to identify problem areas and apply the spacing and shielding approaches to resolve the problems. Process engineers must continue to develop *low-k* dielectric material to further reduce the effect of coupling capacitance. Ultimately, signal integrity-aware CAD tools, good design practices, and improved dielectrics are all needed to cope with these problems.

*10.5 Interconnect Inductance

A relatively new and even more complex issue of inductance is lurking around the corner as we scale technology and increase the clock frequency. Unfortunately, it may become a problem soon if not addressed properly from a design perspective. It is interesting to note the order in which we encountered DSM interconnect issues: first we experienced the increase in resistance; then, coupling capacitance; and now inductance. This order of appearance was quite fortunate since each one is progressively more difficult to understand, extract, or analyze than the previous one. Interconnect resistance is easy to understand and is confined to the metal conductor experiencing the current flow. Coupling capacitance is an electrostatic phenomenon between wires in a localized region around the wire of interest. It is clearly more complex to analyze than resistance, but still a tractable problem. Inductance, on the other hand, is an electromagnetic phenomenon and its effect is not necessarily confined to a localized region. It is more complicated to understand and almost impossible to extract and analyze, except for small sections. Nevertheless, we must understand its effects and explore options to avoid its impact.

The first step is to review inductance from an electromagnetic point of view. When current flows through a conductor, it sets up an electromagnetic field around it with a flux that stores energy. This flux is linearly related to the current, as follows:

$$\phi = Li \quad (10.18)$$

The proportionality constant is obviously the inductance, L , which is measured in *Henrys* (H). As the current changes, the magnetic flux acts on the wire itself and induces a voltage drop. For a linear inductor, we can take the time derivative of Equation (10.18) to obtain

$$\Delta V = L \frac{di}{dt} \quad (10.19)$$

This equation implies that the voltage drop is zero for a dc current. That is, the inductor can be treated as a short circuit for low-frequency operation. As the current demand of a chip increases, the voltage drop associated with the inductor increases. In high-speed designs, the rate of change of current with respect to time is increasing rapidly. Therefore, inductance effects are especially important at the package pins and in the power distribution system. We will examine this issue in more detail in Chapter 11.

Signal lines are also beginning to exhibit increased effects of inductance. It is important to realize that interconnect is really a distributed *RLC* line, not simply a distributed *RC* line. As such, it behaves like a transmission line under certain circumstances. We now elaborate these circumstances and describe how the inductance effects modify the delay equation. We will use R_{int} , C_{int} , and L_{int} to represent the resistance, capacitance, and inductance per unit length, respectively. An estimate of the inductance per unit length can be obtained from a well-known relationship in electromagnetic theory

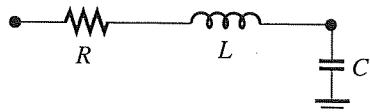
$$C_{int}L_{int} = \epsilon_{ox}\mu_{ox} \quad (10.20)$$

where the insulating material is assumed to be oxide with permittivity, ϵ_{ox} , and permeability, μ_{ox} . This equation only applies to a conductor that is completely surrounded by a uniform dielectric, but it is useful in obtaining first-order inductance values. Assuming that $C_{int} = 0.1 \text{ fF}/\mu\text{m}$, $\epsilon_{ox} = (4)(8.85 \times 10^{-14} \text{ F/cm})$ and $\mu_{ox} = 12.6 \times 10^{-7} \text{ Wb/amp}$, then $L_{int} = 0.45 \text{ pH}/\mu\text{m}$. This is a very crude estimate but does provide the correct order of magnitude for inductance.

In Figure 10.23, a lumped model of the *RLC* line is shown. We can analyze this simple circuit to develop some intuition before proceeding on to the distributed line. First, consider the complex impedance associated with the resistance and inductance of the signal line

$$Z_{RL} = R + j\omega L \quad (10.21)$$

where $\omega = 2\pi f$ is the signal frequency in radians/second. If $R \gg j\omega L$, then Equation (10.21) tells us that inductance is not important. If $R < j\omega L$, then inductance

**Figure 10.23**

Lumped model of RLC interconnect.

effects will be observed in the signal line. In the past, the resistance was larger than the reactive component so that RC effects were dominant. In recent years ω has been increasing significantly in very high-speed designs to the point where the reactance term is important. At the same time, process engineers and designers have been working to reduce wire resistance. First, the resistance of the metal lines has been reduced with the introduction of copper as a replacement for aluminum. Second, certain signals use very wide wires to lower the resistance significantly. This reduction in resistance combined with an increase in the operating frequency has made the reactive term non-negligible in certain signals. Interestingly, while inductance values have not increased over the years, the increase in frequency and decrease in resistance have combined to raise concerns about inductance.

The above analysis is somewhat localized to the interconnect itself. When considering the significance of inductance, it is important to consider both wire resistance and the on-resistance of the buffer driving the wire. In most cases, the buffer resistance is very large (on the order of kilo-ohms) and will dominate. That is,

$$R_{\text{buffer}} + R_{\text{wire}} \gg j\omega L$$

Therefore, we should be able to ignore self-inductance for most signals on the chip. Of course, if the buffer is large and the wire is wide, we would have

$$R_{\text{buffer}} + R_{\text{wire}} < j\omega L \quad (10.22)$$

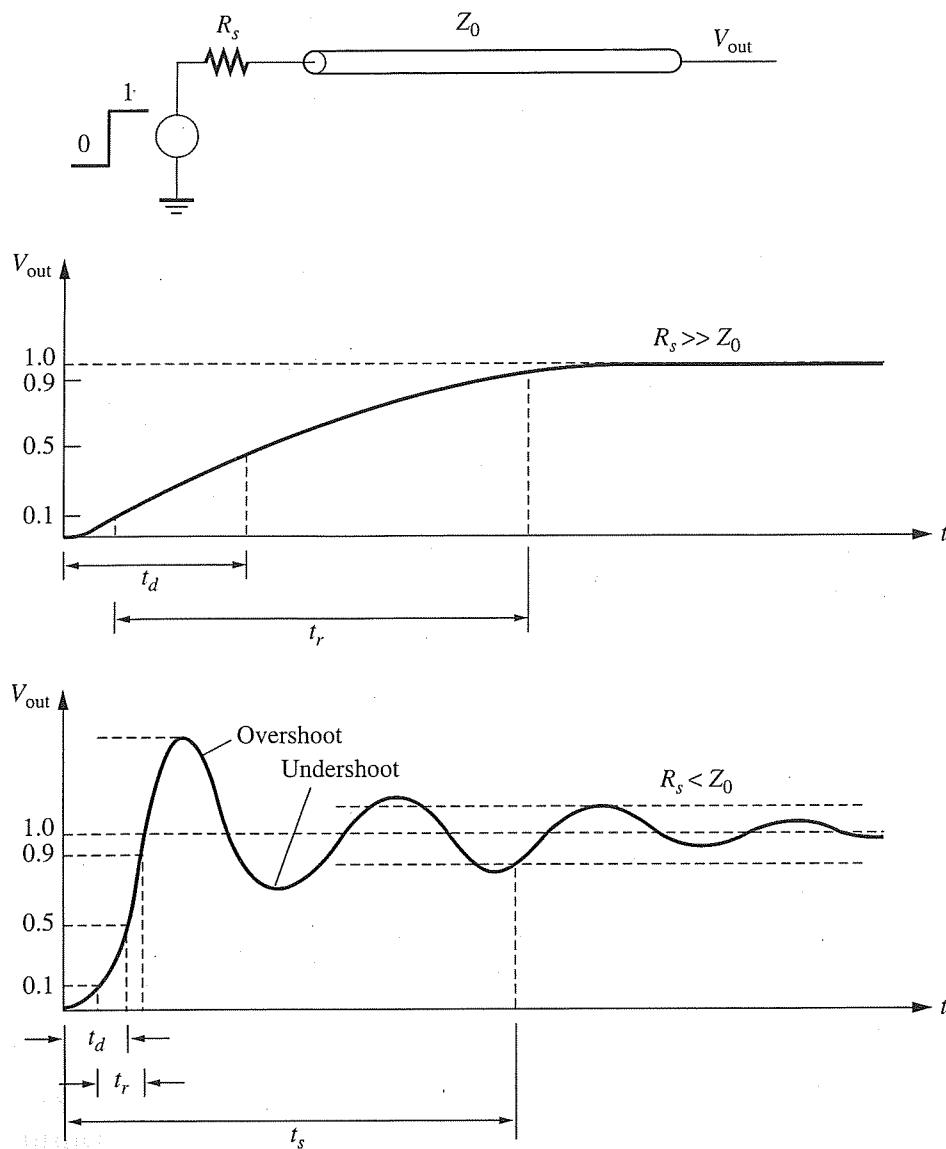
and inductance would be noticeable. This is most prominent in clock lines. Therefore, the delay should be computed using RLC delay equations.

Next, we include capacitance in the analysis. The lossless transmission line of Figure 10.24 has inductance and capacitance but no resistance. It is driven by a voltage source V_S with a source resistance of R_S . Since the interconnect has no resistance, it can be represented by a characteristic impedance of Z_0 given by

$$Z_0 = \sqrt{\frac{L_{\text{int}}}{C_{\text{int}}}}$$

When we apply a step input to the interconnect, the response depends on the relative values of R_S and Z_0 . When the resistance is much greater than the characteristic impedance, i.e., $R_S \gg Z_0$, the RC effect will dominate the transient response. We expect an exponential response given by

$$V_{\text{out}}(t) = 1 - e^{-t/RC}$$

**Figure 10.24**

Responses of RLC interconnect.

As shown in the figure, the propagation delay, t_p , and rise time, t_r , are rather large.

However, if $R_s < Z_0$, inductance effects will be important and we obtain an oscillatory response within an exponential envelope. The response has a general form given by

$$V_{\text{out}}(t) = 1 - re^{-t/a} \cos(bt + c)$$

where r , a , b , and c depend on the RLC values. The figure illustrates the nature of the response. Note that t_d and t_r are much shorter in this case, but the tradeoff is a relatively large overshoot and undershoot of the waveform, and a longer settling time, t_s . These delay values can be computed from the response equation above.

Finally, we consider a lossy transmission line where distributed effects are taken into account. For this case, the length, l , of a distributed RLC line plays a role in whether it exhibits the effects of inductance. In particular, as a signal propagates down the wire, it experiences a certain degree of attenuation due to wire resistance. When it reaches the end of the wire, it reflects off the load impedance and propagates back toward the source. This reflected signal is also attenuated as it travels back down the wire. The interaction of the forward and backward propagating signals leads to the ringing effects observed in Figure 10.24. If the wire is very long, then inductance effects will not be observed due to signal attenuation. A standard derivation can be used to find the maximum length at which inductance effects occur

$$l_{\max} < \frac{2}{R_{\text{int}}} \sqrt{\frac{L_{\text{int}}}{C_{\text{int}}}} \quad (10.23)$$

Qualitatively, this tells us that as the resistance per unit length increases, the length over which inductance is important decreases.

The transition time of the signal at the output of the buffer also controls the transmission line effect. If the round-trip of the signal down the line and back to the buffer is short relative to the rise time, it does not have a significant transmission line effect. The round-trip time is given by

$$t_{\text{round-trip}} = 2T_0 = 2l\sqrt{L_{\text{int}}C_{\text{int}}} \quad (10.24)$$

The quantity T_0 is referred to as the time-of-flight. If the rise time (or fall time) is shorter than twice the time-of-flight, we will observe the transmission line effects. Therefore, we require that

$$t_r < 2l\sqrt{L_{\text{int}}C_{\text{int}}} \quad (10.25)$$

In terms of l , this translates to a minimum length of

$$l_{\min} > \frac{t_r}{2\sqrt{L_{\text{int}}C_{\text{int}}}} \quad (10.26)$$

Combining the results of Equations (10.23) and (10.26), we find that

$$\frac{t_r}{2\sqrt{L_{\text{int}}C_{\text{int}}}} < l < \frac{2}{R_{\text{int}}} \sqrt{\frac{L_{\text{int}}}{C_{\text{int}}}} \quad (10.27)$$

This result tells us that the wire must be larger than the length specified by $t_r/2\sqrt{L_{\text{int}}C_{\text{int}}}$ but smaller than the value specified by $(2/R_{\text{int}})\sqrt{L_{\text{int}}/C_{\text{int}}}$ for inductance to be important in delay calculation. We should view these results as being more qualitative than quantitative since the terms in the equations are difficult to compute with great precision. Furthermore, the effect of buffer resistance must also be considered to obtain accurate results.

What makes the inductance problem particularly difficult is that, fundamentally, inductance can only be defined for a closed current loop, not simply for a seg-

ment of wire. The inductance of the loop is proportional to the area of the loop. Knowing the forward path of the current is not sufficient; the *return* path must also be identified. This is not easy to determine for most signals. In the past, the return path was assumed to be in the substrate. For high-speed designs, the return path often occurs in the power and ground distribution systems, and even in other signal lines. Worse yet, there may be multiple return paths, some through power/ground rails, and others through signal lines that are not close to the signal of interest. The extraction of inductance values is difficult since the effect of inductance is not confined to a localized region around the conductor.

More problematic is *mutual inductance* since it causes noise and delay effects on unsuspecting neighboring lines in unexpected ways. This type of problem is a growing concern for most integrated circuit designers, as it is difficult to extract and analyze. Mutual inductance refers to time-varying current in one line giving rise to a voltage drop in a second line. Two *RLC* lines are coupled via mutual inductance, as follows

$$\begin{aligned} v_1 &= L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} \\ v_2 &= M \frac{di_1}{dt} + L_2 \frac{di_2}{dt} \end{aligned} \quad (10.28)$$

where L_1 and L_2 are the self-inductances of the two wires and M is the mutual inductance between the two wires. When a current flows in one line, it induces a return current in the other line due to its mutual inductance. The voltage drop on each line is the sum of the drop due to its self-inductance and its mutual inductance with the other line. This effect can occur over distances of 100 μm or more and create functional failures that are intermittent in nature. The details of mutual inductance effects, while important in future technologies, are beyond the scope of this book. The interested reader is encouraged to consult the references at the end of this chapter for further information.

To summarize, self and mutual inductance effects have been observed most frequently in clocks and busses. Clocks are prone to self-inductance effects whereas busses are more likely to experience mutual inductive effects. To limit the effects of self-inductance, wire lengths can be kept within the boundaries suggested by Equation (10.27). For mutual inductance, shielding techniques of the form shown in Figure 10.19 can be used to limit its effects.

*10.6 Antenna Effects

Another important issue associated with interconnect is the *antenna effect*. This is a reliability issue that may destroy the thin-oxide of transistors during fabrication or modify the threshold voltage of devices. During the fabrication process, charge builds up on each metal layer during plasma etching, sputtering, or chemical-mechanical polishing (CMP). This phenomenon is referred to as an antenna effect since charge is attracted to each metal layer like an antenna. Since $Q = CV$, the accumulated charge

creates a potential difference across the gate oxide of a MOS transistor. When too much charge accumulates on the gate, the voltage drop across the oxide damages the transistor by shorting the gate to the substrate. A less dramatic, but still undesirable, effect is a shift in the V_T due to injection of mobile carriers into the oxide through Fowler-Nordheim tunneling.

The antenna effect manifests itself as an interconnect design issue since the metal layer selection and routing determine the impact of this phenomenon. When Metal 1 is processed, its charge buildup is rather small since the lengths of the wires associated with Metal 1 tend to be short. As we add more and more layers, the area tends to increase since the wires are longer and thicker. Higher levels of metal have more surface area and therefore accumulate more charge, so they tend to be more troublesome with respect to the antenna effect. If a transistor survives the process on a given layer, that layer does not play as significant a role in the antenna effect during subsequent processing operations. Therefore, we usually assess antenna problems for each metal segment in each layer separately.

The *antenna rules* specify maximum allowable ratios of metal (or poly) area to gate area for each layer.

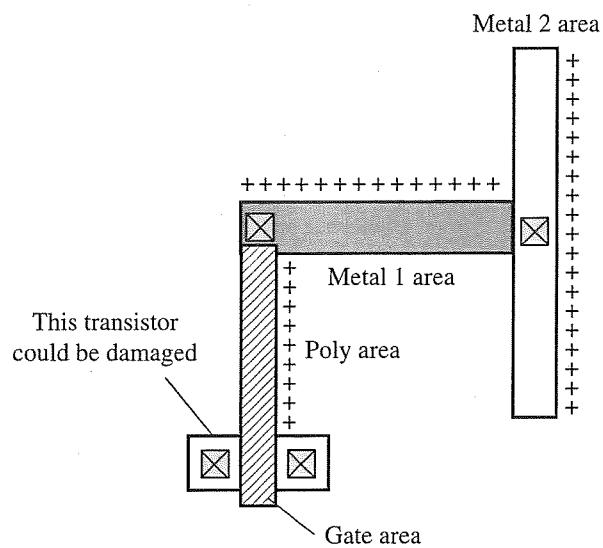
$$\text{Antenna Ratio} = \frac{\text{Area}_{\text{metal}}}{\text{Area}_{\text{gate}}} \quad (10.29)$$

For example, if the antenna ratio is set to 100, the area of a metal line would have to be 100 times larger than the gate area for an antenna rule violation to occur. There is a different ratio for each metal line. The design-rule checking (DRC) tools can check to see if each wire segment on each layer causes a violation, and flag those that are potential problems. For threshold voltage shifting, all metal layers connected to a gate need to be combined to obtain $\text{Area}_{\text{metal}}$ in Equation (10.29).

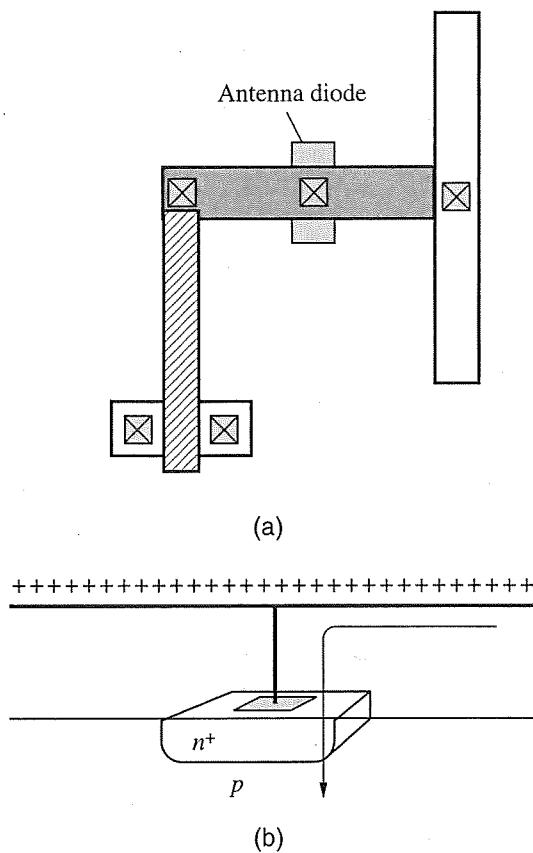
An example of the antenna effect is shown in Figure 10.25. Relative to the indicated device, poly, Metal 1, or Metal 2 may act to damage the gate region. For example, when Metal 1 is deposited and polished, charge develops on it and builds up a potential difference that could damage the oxide of the indicated transistor. If it tolerates the voltage level, then the corresponding Metal 1 segment is not considered to be a problem in the rest of the processing sequence. However, when the next layer of metal is applied, the corresponding charge on this second layer alone is checked against its antenna ratio. A more conservative approach recognizes that process induced damage is a cumulative effect. In that case, the total area of poly + Metal 1 + Metal 2 is divided by the gate area for the antenna check.

A number of different approaches have been used to avoid antenna effects. First, the insertion of diodes along the wire can be used to discharge the metal lines during the processing sequence. This approach is illustrated in Figure 10.26 where a *pn* junction has been connected to Metal 1. The job of this diode is to discharge each layer of metal as it is deposited so that the transistor does not see any significant voltage drop. It is reverse-biased and harmlessly removes charge as it accumulates.

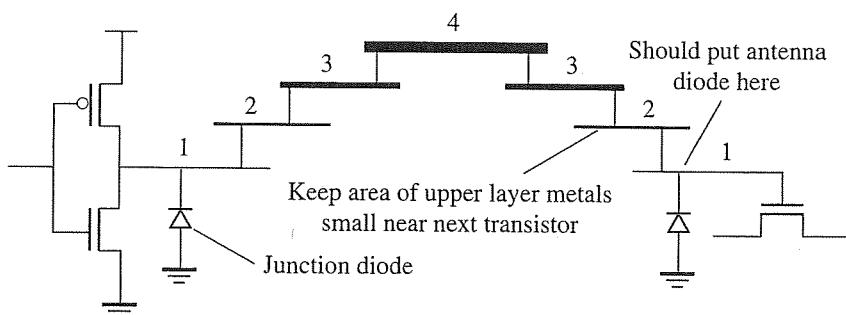
The use of antenna diodes is very effective but presents some practical problems. Diodes cost area and their location must be decided based on the metal structure. Ideally, the diode should be near the transistor gate as shown in Figure 10.27. There is already a reverse-biased junction at the output of the inverter, but the

**Figure 10.25**

Antenna effects during IC processing.

**Figure 10.26**

Antenna diodes to discharge metal lines.

**Figure 10.27**

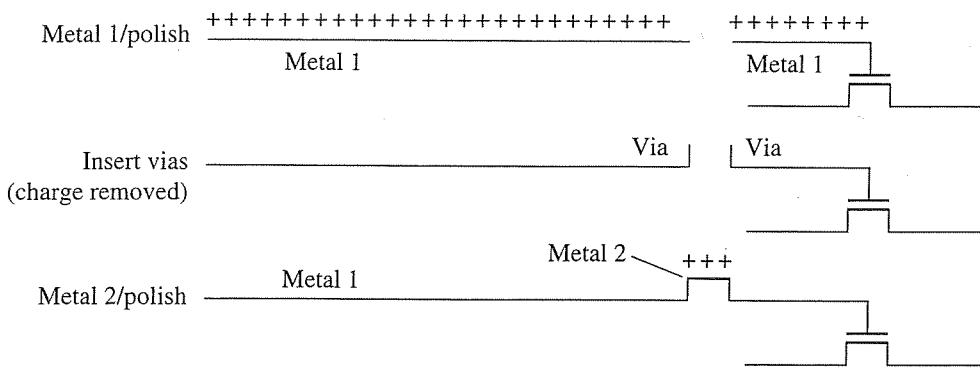
Proper placement of antenna diodes.

antenna diode is needed near the transistor gate as shown. Note that as each layer in the example is deposited, the far-end transistor must be able to handle all the charge accumulation on each successive layer in this configuration. Therefore, the far-end metal lines must be kept short or we must place diodes near the transistor. However, it would be too expensive in terms of area to place diodes near all transistor gate inputs, or embed the antenna diodes inside cell libraries. It should only be used where absolutely necessary. This involves a complete DRC verification of the entire layout to identify the locations that are appropriate. The second issue is to define the area of the diode. This depends on the area of the metal segments that it supports. If large diodes are used unnecessarily, area may be compromised.

A second approach is to break up long wires with buffer insertion. This will reduce the total area available for charge accumulation and increase the transistor gate area, thereby reducing the total charge on each metal line. Of course, such a process is already carried out as part of reducing the RC delay, but we obtain a side benefit of a reduction of antenna effects. We must still verify that no segment creates a problem, but we will definitely reduce the number of violations after buffer insertion.

A third approach is to create metal jumpers from one level to the next to break up the lines near the far-end of a wire. This is illustrated in Figure 10.28. In this case, we show a Metal 1 wire that violates the antenna ratio. A seemingly unnecessary jog to Metal 2 is introduced near the gate of the transistor to avoid the total charge of Metal 1. Recall that, when the processing steps are applied to introduce the vias, the charge of Metal 1 is removed. Therefore, the transistor will only experience the small charge from the Metal 2 segment. The additional resistance and capacitance of the vias and Metal 2 segment are usually negligible.

In summary, the proper solution to the antenna problem depends on the physical design methodology, tools, and engineering tradeoffs associated with the design. All of the above solutions are viable if carried out with careful considerations of the tradeoffs. In advanced processes, diode and buffer insertion appear to be most helpful. Hopefully, tools of the future will automatically take care of this issue on behalf of the designer.

**Figure 10.28**

Metal line jogs to avoid antenna effects.

10.7 Summary

Interconnect data:

- (a) $0.18 \mu\text{m}$ 5-layer aluminum process: $\rho = 2.7 \mu\Omega\text{-cm}$; Metals 1–4, $T = 0.5 \mu\text{m}$, $H = 0.5 \mu\text{m}$, $W_{\min} = 0.3 \mu\text{m}$; and Metal 5, $T = 1 \mu\text{m}$, $H = 0.5 \mu\text{m}$, $W_{\min} = 0.4 \mu\text{m}$
- (b) $0.13 \mu\text{m}$, 8-layer copper process: $\rho = 1.7 \mu\Omega\text{-cm}$; Metals 1–6, $T = 0.4 \mu\text{m}$, $H = 0.5 \mu\text{m}$, $W_{\min} = 0.2 \mu\text{m}$; and Metals 7–8, $T = 0.8 \mu\text{m}$, $H = 0.5 \mu\text{m}$, $W_{\min} = 0.4 \mu\text{m}$

Wire resistance calculations:

$$R = \frac{\rho L}{A} = \frac{\rho L}{TW} \quad R_{sq} = \frac{\rho}{T} \quad R = R_{sq} \left(\frac{L}{W} \right)$$

$$R_{sq} = \frac{\rho}{T} = \frac{2.7 \mu\Omega\text{-cm}}{0.5 \mu\text{m}} = 54 \text{ m}\Omega/\square \quad \text{Al} \quad \text{Metals 1–4}$$

$$R_{sq} = \frac{\rho}{T} = \frac{2.7 \mu\Omega\text{-cm}}{1.0 \mu\text{m}} = 27 \text{ m}\Omega/\square \quad \text{Al} \quad \text{Metal 5}$$

$$R_{sq} = \frac{\rho}{T} = \frac{1.7 \mu\Omega\text{-cm}}{0.4 \mu\text{m}} = 42 \text{ m}\Omega/\square \quad \text{Cu} \quad \text{Metals 1–6}$$

$$R_{sq} = \frac{\rho}{T} = \frac{1.7 \mu\Omega\text{-cm}}{0.8 \mu\text{m}} = 21 \text{ m}\Omega/\square \quad \text{Cu} \quad \text{Metals 7–8}$$

Elmore delay calculation:

$$\tau_i = \sum_k (C_k \times R_{ik})$$

Optimal buffer insertion:

$$\text{Number of stages } N = \sqrt{\frac{R_{\text{int}} C_{\text{int}} L^2 / 2}{R_{\text{eqn}} (C_J + C_G) (1 + \beta)}}$$

$$\text{Size of buffers } M = \sqrt{\frac{R_{\text{eqn}} C_{\text{int}}}{C_G (1 + \beta) R_{\text{int}}}}$$

Coupling capacitance:

$$C_{\text{area}} = \varepsilon_{\text{ox}} \frac{W}{H}$$

$$C_{\text{lateral}} = \varepsilon_{\text{ox}} \frac{T}{S}$$

$$C_{\text{fringe}} = \varepsilon_{\text{ox}} \ln\left(1 + \frac{T}{H}\right)$$

Coupling capacitances in delay calculation:

$$C_L = C_{\text{gnd}} + C_c \quad \text{aggressor not switching}$$

$$C_L = C_{\text{gnd}} \quad \text{both switching together}$$

$$C_L = C_{\text{gnd}} + 2C_c \quad \text{both switching opposite}$$

Crosstalk noise:

$$\Delta V_2 = \frac{C_c \Delta V_1}{(C_{\text{gnd}} + C_c)}$$

Self-inductance:

$$\Delta V = L \frac{di}{dt}$$

Mutual inductance:

$$\nu_1 = L_1 \frac{di_1}{dt} + M \frac{di_2}{dt}$$

$$\nu_2 = M \frac{di_1}{dt} + L_2 \frac{di_2}{dt}$$

Conditions where inductance should be considered in delay calculation:

$$R_{\text{buffer}} + R_{\text{wire}} < j\omega L$$

$$\frac{t_r}{2\sqrt{LC}} < l < \frac{2}{R} \sqrt{\frac{L}{C}}$$

Antenna rule:

$$\text{Antenna Ratio} = \frac{\text{Area}_{\text{metal}}}{\text{Area}_{\text{gate}}}$$

REFERENCES

1. M. Celik, L. Pileggi, and A. Odabasioglu, *IC Interconnect Analysis*, Kluwer Academic Publishers, Boston, MA, 2002.
2. Y. Ismail and E. Friedman, *On-chip Inductance in High-Speed Integrated Circuits*, Kluwer Academic Publishers, Boston, MA, 2001.
3. H. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley Publishing, Reading, MA, 1990.
4. N. Arora, K. V. Raol, R. Schumann, and L. Richardson, "Modeling and Extraction of Interconnect Capacitances," *IEEE Transactions on Computer-Aided Design*, vol. 15, no. 1, pp. 58–67, Jan. 1996.

PROBLEMS

- P10.1.** Compare the resistance of a 20 mm Metal 5 wire in 0.18 μm technology with a 20 mm Metal 8 wire in 0.13 μm technology (a non-scaled wire). Use the information provided in the Summary section. Compute the delay for each wire assuming an ideal source and a wire capacitance of $C_{\text{int}} = 0.1 \text{ fF}/\mu\text{m}$.
- P10.2.** Compare the resistance of a 20 mm Metal 5 wire in 0.18 μm technology with a 14 mm Metal 8 wire in 0.13 μm technology (a scaled wire). Use the information provided in the Summary section. Compute the delay for each wire assuming that the wire capacitance is $C_{\text{int}} = 0.1 \text{ fF}/\mu\text{m}$.
- P10.3.** Two adjacent Metal 5 wires in 0.18 μm technology are spaced out from 4λ to 40λ in 4λ increments. Plot the capacitance per unit length of these wires as a function of spacing, assuming that they are shielded above and below. Use minimum width dimensions for the wire.
- P10.4.** Simulate a distributed *RC* line as an *RC* ladder circuit in SPICE with 1 stage, 3 stages (as shown in Figure P10.4), and 20 stages. You can choose your own sizes but ensure that the total resistance and total capacitance are the same in all three cases. Measure the propagation delay for the three cases when a step input is applied at the source. Derive a delay formula of the form:

$$\tau = \alpha R_{\text{wire}} C_{\text{wire}}$$

and compute α for the three cases. Comment on the results.

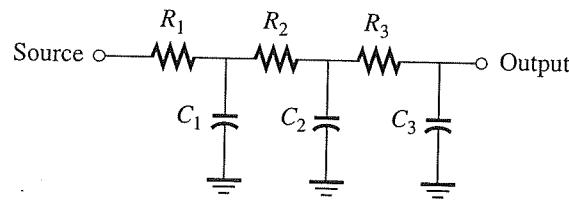


Figure P10.4

- P10.5. Repeat the same experiment as described in Problem P10.4 but this time use each of the following three models shown in Figure P10.5 in place of the RC ladder. Which model produces accurate results?

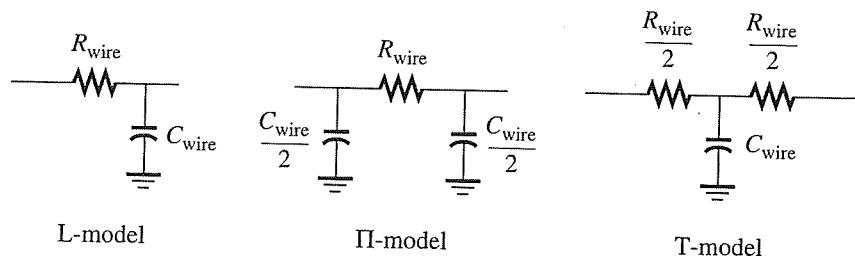


Figure P10.5

- P10.6. Simulate the RC circuit shown in Figure P10.6, using SPICE. Set $R_1 = R_2 = R_3$ and $C_1 = C_2 = C_3$. Compare the Elmore delay with the delay derived from SPICE. How accurate are the results for each node using Elmore? Explain why any particular result is not accurate.

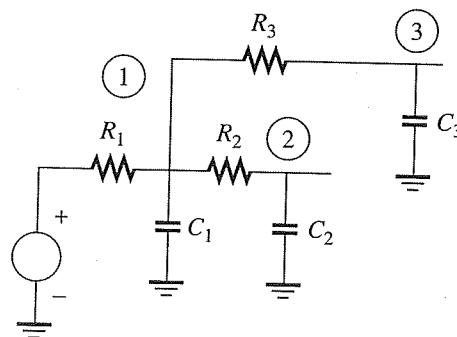


Figure P10.6

- P10.7. Wires can be placed into short, medium, long, and very long categories. Establish rules of thumb for a Metal 5 wire in $0.18 \mu\text{m}$ technology that quantify each of these categories based on the following definitions:

- (a) Short wires do not require any resistance or capacitance calculations if they contribute less than 5% to the delay. What is the maximum short wire length?
- (b) Medium wires must include capacitance in the delay calculation, but do not require any resistance information. What is the maximum medium wire length?
- (c) Long wires require both resistance and capacitance information but do not require buffer insertion. What length of wire can remain unbuffered?
- (d) Very long wires require buffer insertion.
- P10.8.** Figure P10.8 shows a circuit that has two interconnect lines with a large coupling capacitance between them.
- Estimate the peak value of noise on the victim line when the aggressor line switches by 1.8V.
 - What is the effective capacitance loading on the victim for delay calculation if the aggressor switches in the opposite direction of the victim?

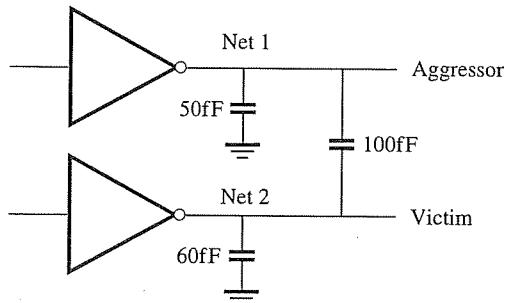


Figure P10.8

- P10.9.** For a $0.13 \mu\text{m}$ technology, estimate the range over which inductance is important assuming a Metal 8 wire. You can assume typical numbers in the given technology. How does this range compare against the critical length in the same technology? (See Section 10.2.3.)
- P10.10.** The amount of noise injected on a wire depends on the strength of the driver, rise time of the aggressor, as well as the percent of the total capacitance that is due to coupling. Create a simple RC model based on Figure P10.8 that you can use to help you estimate the size of noise caused by the coupling.
- P10.11.** Compare the number of repeaters required in a $0.18 \mu\text{m}$ technology versus a $0.13 \mu\text{m}$ technology for a 10 mm wire. Use the top layer of metal and the standard technology parameters to compute the numbers.

- P10.12.** Consider an 18 mm Metal 7 wire in a $0.13 \mu\text{m}$ technology. Assume the wire is $0.4 \mu\text{m}$ wide and $0.8 \mu\text{m}$ thick with a spacing to adjacent wires of $2 \mu\text{m}$. The height above and below to Metal 8 and Metal 6 is $0.5 \mu\text{m}$. Assume Cu for interconnect ($\rho = 0.017 \Omega\cdot\mu\text{m}$) and a low- k dielectric material ($\epsilon_r = 3.0$). Assume a worst-case coverage of Metal 6 below, Metal 8 above, and Metal 7 for adjacent lines that are shielding the wire of interest.
- Compute the resistance per unit length and capacitance per unit length for this wire.
 - Estimate the delay of the distributed RC wire, assuming that the driver is a perfect voltage source and that the load capacitance is 50 fF .
 - Assume that the wire is being driven by a $25X$ inverter. What is the new delay?
 - Now, assume that the inductance is $0.5 \text{ pH}/\mu\text{m}$. Will the inductance be a factor in the delay calculation for this buffer size? Compare hand analysis with results in SPICE.
 - Now remove the $25X$ buffer and perform buffer insertion on this wire using the standard formulas. Should inductance be a factor for the new wire lengths? Compare your hand analysis with results from SPICE.