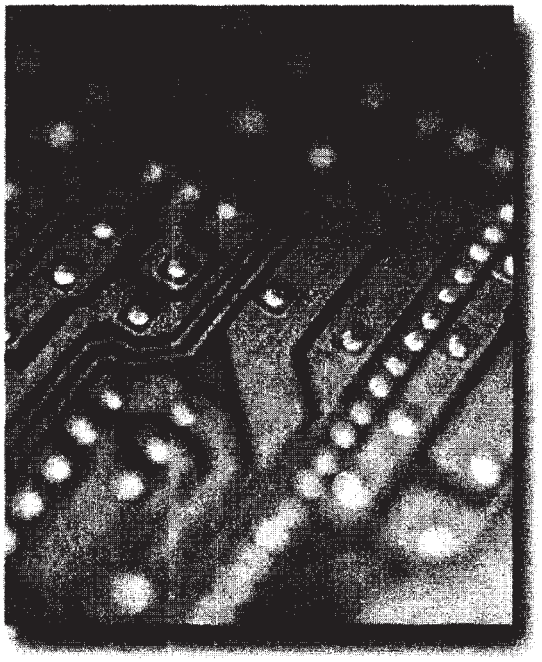
PART 2



SIGNAL INTEGRITY ISSUES

Ten or 15 years ago most people treated printed circuit boards as totally passive devices for connecting components together. As long as the traces were connected to the correct pins, boards almost never had a negative impact on circuit performance. In more recent years, this is becoming less and less true. Many board designers now need to worry about the parasitic elements of traces (resistance, capacitance, and inductance), the interaction between individual traces, and even the interaction between traces and the outside world. Part 2 explains why this is so and what the designer needs to do about it. Many different topics are covered, including EMI, crosstalk, transmission lines and reflections, and power system decoupling. You will learn what a fundamental role signal rise time has for all of these issues, and then what design rules you need to consider as a result.

INTEL-1010
U.S. Patent 7,151,430



8 SIGNAL INTEGRITY OVERVIEW

One time I served on a panel where I was asked, "At what point does signal integrity become a problem?" The intent of the question was to ask at what *frequency* signal integrity becomes an issue for board designers, but as stated, the question was more general. That got me thinking about how broad the term *signal integrity* really is, even though we have come to think of it in terms of more narrowly defined high-speed issues.

A not-really-facetious answer to the question is that we have a signal integrity problem whenever the signal begins to lose its integrity. And this is not necessarily related to its frequency. Two of the more obvious and common ways a signal can lose integrity is when it becomes distorted or when the signal-to-noise ratio (S/N) begins to degrade.

Signal distortion typically means that the waveform of interest begins to change shape. The degree to which this can happen before it becomes a problem depends very much on the application. Digital signals, which we typically think of as being rectangularly shaped pulses, usually carry one bit of information per clock cycle. They can often withstand a fair amount of distortion without obscuring the bit state they are in. Analog signals, on the other hand, such as we find in video and audio systems, can be very sensitive to distortion. A change in the waveform will often be seen or heard. Your home stereo system, for example, probably has a specification for harmonic distortion, which relates to the "purity" of the audio signal as it is processed.

In my lifetime, home entertainment harmonic distortion specifications have improved greatly, and today's audio and home entertainment systems are, for all practical purposes, distortion free. The most common way to make a distortion-free system is to design it so that the gain is absolutely linear over the frequency range of interest.

INTEL-1010
U.S. Patent 7,151,430

S/N issues come into play whenever an unwanted noise becomes detectable or interferes with the signal we are concerned with. For example, if you hear a 60-Hz hum (which is, of course, very low frequency) from your home audio system, you have a S/N or a signal integrity problem. Signal integrity issues and solutions are not new. People who have spent part of their careers designing power or electromagnetic switching systems have understood the importance of decoupling and separation of power supply and grounding systems for decades. Problems associated with reflections, and their transmission line solutions, have also been around for years. Radio frequency engineers (even with "low" AM transmitting frequencies below one MHz) have needed to understand transmission line techniques since the broadcasting industry began. Signal integrity issues at the circuit design level have been around for a long, long time.

In our industry (circuit board design) we have come to equate *signal integrity* with *high speed* only in recent years. Up until now the circuit board has been a purely passive device with virtually no circuit impact (unless you were among the very few designers putting RF or microwave circuits on the board). However, in recent years, the board itself has begun to cause S/N degradation. The reason is almost exclusively because the rise times of the signals we work with have decreased. These decreased rise times, in conjunction with trace inductance and capacitance, now cause noise signals of greater magnitude than in the past.

Let's use inductance for our illustration. Traces have *always* had inductance. That is important for new board designers to understand. Inductance isn't something that just raised its ugly head in the last few years. The voltage generated across an inductor can be approximated by $V = L \times di/Tr$, where L is, in this case, the inductance of the trace (measured in henrys), di is the change in current (in amps) being switched, and Tr is the rise (fall) time (in seconds) associated with the switching current. V gets larger as (among other things) Tr gets smaller. V, in this case, can be considered a noise voltage. Thus, the S/N ratio gets worse as V gets bigger (or, as Tr gets smaller). Therefore, signal integrity gets worse (because the S/N ratio degrades) as Tr gets smaller (i.e., rise time gets faster).

This illustration used inductance. We can use stray trace capacitance to illustrate exactly the same point. There is stray (parasitic) inductance and capacitance all over our boards.

So, back to the initial question: At what point does signal integrity become a problem? For board designers, the answer is: When the rise time decreases to the

INTEL-1010
U.S. Patent 7,151,430

point where the parasitic inductances and capacitances on the board begin to result in noise signals and transients that become troublesome. When is that? Well, it depends, of course, on the circuit specifics, so it is very difficult to generalize on a specific value. Some typical symptoms that might accompany signal integrity problems might include:

1. Design fails FCC compliance testing (a sure sign!).
2. Design fails intermittently, often without discernable pattern.
3. Design *did* work reliably, now it doesn't, after:
   Nothing.
   Changing IC supplier.
   Changing board fabricator.
4. Boards must be selected to work together.
5. Design is sensitive to power supply variations.
6. Design is sensitive to temperature variations.
7. Design works in the lab, but is intermittent in the field.

Symptom 3 is a particularly subtle problem, but one that is becoming increasingly common. Consider the following situation: A component (IC) supplier improves its processes. The process results in smaller die, better yields, lower cost, and improved performance, especially with regard to rise time. No announcement is made regarding these improvements, which are introduced in-line without fanfare or part number modification. It is hard to be critical of this—after all, everybody is a winner here, right? Well, almost everyone.

Consider a design that has been proven over the years and one that has been highly reliable. Suddenly, reliability problems begin to appear and they are difficult to diagnose, in part because they are so unexpected. Sometimes problems like this have been traced to unannounced improvements in the designs of the components used. The faster rise times caused signal integrity problems that were not there before. Board design techniques that were acceptable with lower rise time circuits were no longer acceptable. Similarly, problems like this can occur when switching from one component supplier to another. The parts may be considered substitutes from the standpoint of form, fit, and function, but may have subtle rise-time or timing differences that cause them to react differently in circuits.

Circuit board materials can also play a significant role in signal integrity. Factors like board thickness are obvious. But many people don't understand the im-

INTEL-1010
U.S. Patent 7,151,430

portant role that other parameters, such as relative dielectric coefficient, can play in controlling signal integrity. If a company changes board fabricators, and appropriate control over fabrication parameters, including relative dielectric, is not maintained, unexpected problems can be introduced during this type of changeover.

In general, signal integrity problems on circuit boards fall into four areas, all of which are related to rise time (or related frequency harmonics). They are:

- EMI (radiations beyond the board or susceptibility to radiations from outside the board)
- Reflections on a single net
- Crosstalk between two or more nets, in many ways a special case of EMI
- Power system stability during component switching

We are going to talk about each of these areas in this part of the book. We will cover the nature of the problem (i.e., what causes it), how the problem is manifested in circuits, and what design solutions can be employed to minimize its effects.

You may be surprised at the number of solutions we are going to offer. There aren't very many, because, in part, there aren't very many *causes* to the problems. The primary cause is the decreasing rise (or fall) time of our signals and its interaction with the reactive characteristics of the stray inductances and capacitances on our boards.

Thus, you will notice that the same figure sometimes pops up in several different chapters. There are certain design "errors" we can make that have multiple consequences. For example, a slot in a plane can cause EMI problems (Chapter 9), reflection issues (Chapters 10 and 11), and crosstalk problems (Chapters 12 and 13).

One aspect to the solution to signal integrity problems is usually a real eye opener for designers just learning about these issues. Three of the important electronic principles covered in Part 1 of this book are:

1. Current is the flow of electrons.
2. Current flows in a closed loop.
3. Current is a constant everywhere in a closed loop.

This has fairly basic consequences for designers:

INTEL-1010
U.S. Patent 7,151,430

1. Every signal has a return.
2. You need to know where it is!

When you stop to think about it, designers often spend an enormous amount of time and effort planning and strategizing how to route 50% of the signals, and then leave the other 50% (the returns) to chance. *One of the primary causes of signal integrity issues is the lack of control over signal return currents.* This will be a common theme in almost every chapter.

Control over signal return currents is almost totally achieved by using planes. One solution that's common to every area is the use of reference planes. It is pretty much axiomatic that (for most of us) good signal integrity performance *requires* reference planes. Planes play an integral role in:

- Controlling EMI
- Stabilizing trace impedance, an integral part of controlling signal reflections
- Controlling crosstalk
- Providing high-frequency decoupling for our power systems.

Another solution that is common to at least two areas is the minimization of the distance between the trace and the reference plane (i.e., the height of the trace above the plane). Minimizing this distance (a) minimizes EMI and (b) minimizes crosstalk.

A third solution relates to the general problem of minimizing inductance. This is especially important around connections to devices (pads, vias, soldered connections, etc.). Minimizing inductance is generally even more important around power and ground planes.

A few other solutions will be offered at various specific places. But the point remains. It may seem like there are a lot of potential signal integrity issues with our boards, but there are only a few causes. Therefore, there are only a relatively few required solutions. You will find this out by the end of the book.

A student in one of my classes once observed (and I am paraphrasing here), "I can look at boards from (and he named some large computer and memory manufacturers) that seem to violate everything you and the other instructors are teaching in these seminars. How come?" This is an excellent question. To start with, all the potential signal integrity problems highlighted in this book and in the various semi-

INTEL-1010
U.S. Patent 7,151,430

nars are just that, *potential* problems. The rules we provide are ways to solve these potential problems if they do occur. Will the problems always occur? No, but they often do. Will these solutions always work? Usually, but not every single time.

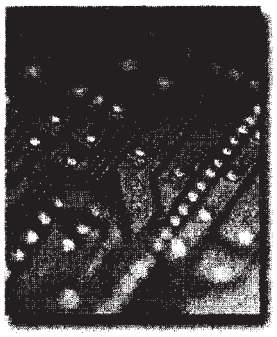
Here is another viewpoint: If (and this is a big if) you are thoroughly familiar with every aspect of every signal on the board, then you will know, for example, if two traces will or will not crosstalk to each other during operation. Therefore, you will know whether the crosstalk guidelines need apply to these particular traces. If you have control over every aspect of the design of the system, then you can optimize the pin outs from the various devices so that signal integrity problems are minimized. If you have unlimited resources, you can have several people make several design iterations, testing different concepts, until you find an optimum set of rules for your particular design (recognizing that several of the iterations are likely not to work well at all).

A friend told me once that his company went through *thirteen* design iterations before they finalized their board design for production. He proudly explained that the last 12 were just "tweaks." After all, the system *did* boot up with the first try. They had the resources to keep trying alternative strategies until they truly optimized the design.

But if you live in the more real world that the rest of us live in, then for all practical purposes you have only one or two shots to get the design right. In that case it is often easier to simply follow good design rules, just to be sure. A less-than-optimum design is usually more cost effective than multiple design iterations. If compromises must be made, knowing the proper design rules will help you troubleshoot a board later to determine which compromises turned out not to work and need to be corrected.

We often make trade-offs in our designs. That is not uncommon, and it is acceptable. What's important is that we do so with our eyes open, so we are alert to potential problems in the future and how to correct them. It is for these reasons that studying books like this and taking as many seminars as you can are important.

INTEL-1010
U.S. Patent 7,151,430



9 ELECTROMAGNETIC INTERFERENCE (EMI)

BACKGROUND

Current is the flow of electrons. When electrons move down a trace or a wire, current flows. Electrons are negatively charged particles. Therefore, we can envision that there is an electrical (negative) charge around a wire as current flows, and the magnitude of this charge depends on the number of electrons (magnitude of the current). We call this charge an *electric field.* Furthermore, as the current flow changes in magnitude, this electric field changes in intensity.

Also, as electrons flow, a magnetic field is generated around the wire or trace. Andre Marie Ampere, for whom the measure of current is named, is credited as the first person to state this law. This is the basic relationship behind an electromagnet. The strength of the magnetic field is related to the magnitude of the current flow, so the magnitude of the magnetic field will change with changing current levels.

Therefore, around any wire or trace carrying a current, there will be an electrical field (often designated by the letter E) and a magnetic field (often designated by the letter H). Together, these two fields form an electromagnetic field, as described in Chapter 2. The two components of the field (electrical and magnetic) must flow together at the same speed. The speed at which they can travel is determined by the medium they travel through, the medium surrounding the current flow, and specifically the relative dielectric constant of that medium. Thus, the propagation speed of a signal is determined by how fast the electromagnetic field can travel through the surrounding medium (see Chapter 2).

INTEL-1010
U.S. Patent 7,151,430

In 1831 Michael Faraday published Faraday's Law of Magnetic Induction. This law states that a *changing* magnetic field (itself caused by a changing current flow) can induce an electrical current in an adjacent wire. So if we send a changing (AC) current down a wire or trace, it can induce a similar current in an adjacent wire or trace. This can be a good thing: It is the basic principle behind radio and television transmission and reception. But it can also be a bad thing: It is the cause of crosstalk and FCC compliance testing problems. As board designers, our challenge is to design boards that maximize this effect when we want to transmit a signal and minimize this effect when we don't. Transmissions we *don't* want are called EMI (in this chapter), or crosstalk (in Chapters 2, 12, and 13.)

In one sense, all of our traces are antennas. A good transmitting antenna is also a good receiving antenna. Therefore, any trace that is a good "radiator" is also a good receiver. Designs that *emit* EMI are also more *susceptible* to EMI. In this chapter we will look at design techniques that help reduce EMI. We will see later that the very same techniques also help minimize crosstalk.

FIELDS AND CANCELLATIONS

Figure 9-1 illustrates the magnetic field lines around a wire when the current is flowing out of the page. The direction of the magnetic flux lines can be determined by the right-hand rule. Point the thumb of your right hand in the direction of the current flow. The magnetic flux lines curl in the direction of your fingers.

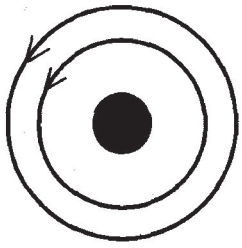


Figure 9-1 Magnetic flux lines curl around a wire following the right-hand rule.

Now consider Figure 9-2. Here there are two conductors, one with current coming out of the page and one with current going into the page. If these two wires

INTEL-1010
U.S. Patent 7,151,430

are carrying a signal and its return current, then the currents will be equal and opposite. Therefore the magnetic fields will also be equal and opposite. Think of yourself as positioned off to the right-hand side of this page. One of these wires will be closer to you than the other (the one with current going into the page). Thus the EMI radiation from that wire will be greater than that from the other. You will be able to measure this EMI radiation, and under the right conditions this radiation might be significant.

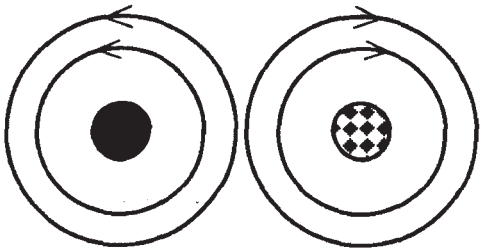


Figure 9-2 A signal and its return have magnetic fields that approximately cancel.

But if we move the two wires closer together, then you, as an observer off to the right, would begin to see radiations from each wire that are more equal. And since they are opposite in phase, or polarity, they tend to cancel. In the limit, if the wires were very close together, their EMI radiations would completely cancel and would not be measurable. Herein is one of our most basic principles: *If you want to minimize EMI radiation, keep the signals and their returns close together.* That, of course, may be easier to say than to do.

This is why wires formed into a twisted pair work fairly well in AC environments. Consider Figure 9-3. The twisted pair of wires carries a signal and its return to and from an operational amplifier. Again, position yourself off to the right of the page. At any given point along the twisted pair, one wire will be closer to you than the other, so the EMI radiation from it will be slightly larger than from the other. But since the wires are twisted, on average they are the same distance from you, so on average, the radiation from each one cancels the other. Therefore, you receive very little net radiation.

INTEL-1010
U.S. Patent 7,151,430

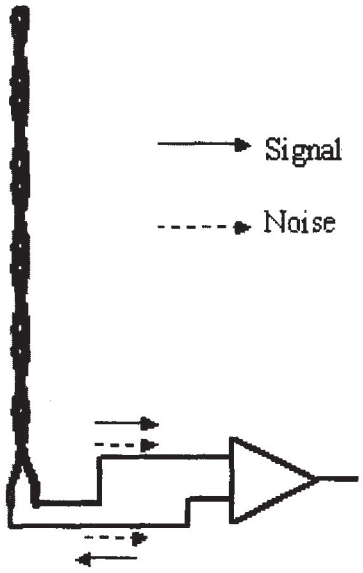


Figure 9-3 A twisted pair has desirable features from an EMI standpoint.

Of course, the wires in a twisted pair can act as an antenna and receive noise, also. Twisted pairs usually feed into differential circuits that are (a) very sensitive to the difference in signal between the pair, but (b) relatively insensitive to any (common mode) signal that exists on both wires. The common mode rejection specification is one measure of a circuit's insensitivity to common mode signals. We explore this more fully in Chapter 14.

SOME BASIC "TRUTHS"

There are a few truths you should always keep in mind. The first, and perhaps most important, is that current always flows in a closed loop. Remember, current is the flow of electrons, and if current does not flow in a closed loop, where do the extra electrons come from or go? Because we know we don't have to deal with buildups of stray electrons on our boards, it seems reasonably intuitive that currents really do flow in a closed loop.

INTEL-1010
U.S. Patent 7,151,430

That means every signal has a return, somewhere. And you (as the designer) need to know where that is. It sometimes amazes us to discover that we designers spend an inordinate amount of time worrying about, considering, and strategizing about 50% of the signals, and then leave the other 50% (the return signals) somehow to chance. So the first truth is that every signal has a return and you need to worry about and control where it is.

A second rule is that the return current will always follow the path of least impedance. We all have heard the rule that signals follow the path of least resistance, but that is a special case, not the general one. It only applies to low-frequency signals. The general rule, for all signals, and particularly the high-speed signals we worry about, is that return currents follow the path of least impedance. This may not always be where you think it is.

SIGNAL COUPLING

Consider Figure 9-4. A changing current i1 is flowing down Trace 1. The changing current creates a changing magnetic field around Trace 1. That changing magnetic field couples into Trace 2 with a coupling coefficient whose value is k. The coupling generates a current in Trace 2 whose value is k x i1 and whose direction (by Lenz's Law, see Appendix B) is opposite to i1. This is exactly how a transformer works, and is the fundamental basis behind all transformer designs.

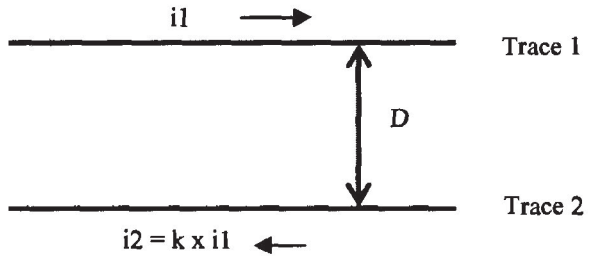


Figure 9-4 Traces can and do couple into each other.

The coupling coefficient will have a value somewhere between 0 and 1.0. Zero means no coupling whatsoever, which can occur if (a) the traces are very far

apart, or (b) the current is not changing (i.e., is DC). A coupling coefficient of 1.0 means perfect coupling, which of course cannot be achieved. The value of k, for our purposes, depends primarily on (a) how fast the current is changing, or the rise time (di/dt), and (b) how closely the traces are spaced (D in the figure). Fast rise times and closely spaced traces lead to higher coupling coefficients. Realistic values for k on circuit boards where the traces are close together are in the range of .4 to .6.

Consider the special case where i1 is a signal and i2 is the signal return. Therefore, i2 = –i1. The coupled signal, k x i1 actually helps boost the return signal. But the return signal, i2, also couples into Trace 1, boosting the primary signal, i1. So, the mutual coupling of the signal and its return work to reinforce, or boost, each other. We can interpret the effect of this two ways: (a) It takes less force (voltage) to push the same amount of current through the circuit, since these currents reinforce each other; or (b) more current can flow for the same voltage, since reinforcing currents are generated on each trace. By Ohm's Law, V = iZ. No matter which interpretation you prefer, it means that the impedance goes down when there is coupling of this type. Furthermore, the impedance gets progressively lower as the coupling increases, or as k increases. And k increases as (a) the distance between the traces decreases, or as (b) the rise time decreases (gets faster).

Now consider Figure 9-5, which illustrates the simple case of a microstrip trace placed over a plane. Assume there is a current flowing down the trace and the return current is on the plane. A good question is: "Where is the return current?" Assume the return current is at Point a. Following the preceding discussion, there will be some mutual coupling between the signal and its return current, resulting in some mutual reinforcement. Now let's assume the return current is, instead, at Point b. Here, the coupling is stronger, because the signal and its return are closer together. Since the coupling is stronger, the overall impedance will be less. Therefore, Point b is a lower impedance path than Point a. Point c represents a signal return path that is even closer to the signal, so the coupling will be even stronger at Point c and the impedance will be even less. Point c represents the closest point between the signal and its return—it is directly under the signal trace—and therefore the point of maximum coupling and lowest impedance.

This is why high-speed signals want to return directly underneath the signal trace: It is the path of lowest overall impedance. Be careful to note, however, that this argument depends completely on the coupling coefficient between the two signals. The coupling coefficient depends, among other things, on di/dt, or rise time. If there is no rise time (e.g., the signals are DC power signals) there is no coupling and

INTEL-1010
U.S. Patent 7,151,430

the return signal might be anywhere. In fact, the DC return will truly follow the path of least resistance anywhere on the circuit board.

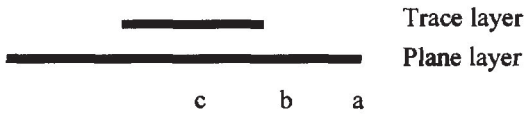


Figure 9-5 Cross-sectional view of a trace and underlying plane. A return current will "want" to follow the path of least impedance, Path (c).

To summarize, DC signals might return anywhere on the board. But AC signals want to return directly underneath the trace carrying the signals. This tendency increases as the rise time decreases (gets faster).

LOOP AREA

Consider a simple configuration with a trace directly over an uninterrupted plane. The signal travels down the trace and the return signal comes back on the plane directly under the trace. Conceptually, we can define a loop that the signal travels as being the length of the trace from the driver to the receiver, down through the receiver to the plane, back under the trace, and back through the driver again. In simple terms, the area enclosed by this loop would be the length of the trace multiplied by the height of the trace above the plane (Figure 9-6). We call this the *loop area* of the signal. The loop area is the area defined by the signal as it travels down the trace and returns back to the source. (This is one reason you need to know where the return signal is on your boards.)

Now loop area is important for this reason. From a practical standpoint, for high-speed signals, EMI is related to loop area. *If you want to minimize EMI, you must minimize loop area.* It's as simple as that. Most of us are intuitively aware that traces, even microstrip traces, routed close to planes generally perform well from an EMI standpoint. That's because their loop areas are small. For the next few pages we'll look at illustrations where loop areas might get out of control, perhaps unexpectedly so.

INTEL-1010
U.S. Patent 7,151,430

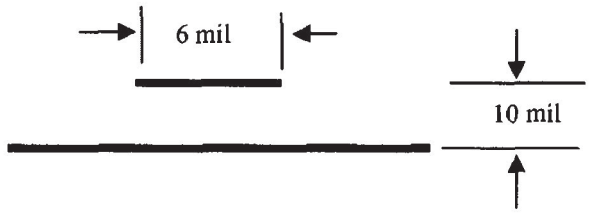


Figure 9-6 Loop area, in this illustration, is 10 x length mil^2.

Slots in Planes 1 Figure 9-7 is an illustration you will see three times in this book. There are many reasons you don't do what is shown there. The figure shows a trace crossing a slot or discontinuity in the plane.

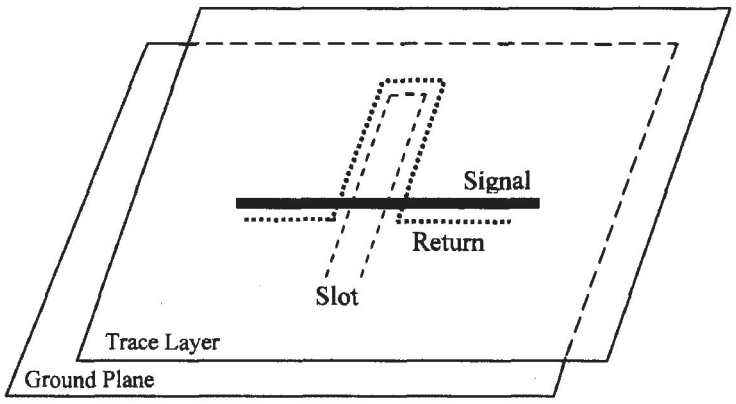


Figure 9-7 If a trace is routed across a slot, the return current must go around the slot, creating a larger loop area.

What might cause a slot or discontinuity in a plane? Perhaps you have finished the design and then your engineer comes to you with "one more little part" that needs to be added. Making provisions for the routing of this new part may involve a significant amount of rerouting of already routed nets. But, perhaps you could just put a little slot in one of the planes, route one or more of the new traces in that slot, and finish the board quickly. Or, perhaps the plane has been partitioned in some way, creating a discontinuity. Or, perhaps there is a special component requirement nearby that results in a small discontinuity.

INTEL-1010
U.S. Patent 7,151,430

No matter what the cause of this slot or discontinuity, if a high-speed trace is routed over it, the return signal will be unable to travel underneath that trace. When the return signal hits the discontinuity, it must travel around it and then come back to its position under the trace again. This trip around the slot or discontinuity creates a loop. Since EMI is related to loop area, you have now created a possible EMI problem where none existed previously.

Return Pathways Figure 9-8 illustrates a case where two boards are connected through a connector. The signal between IC1 and IC2 travels through connector Pin A. What provision should be made for the return signal?

If we use Pin B for the return signal, the loop area will be pretty small. EMI performance should be satisfactory. But if we use Pin E for the return signal, a significant loop will result. This loop may cause an EMI problem.

This illustration highlights the importance of pin assignments in our devices and connectors. It should also be noted that some connectors are better than others with regard to their internal paths. Some poorly designed connectors introduce loop areas even between adjacent pins.

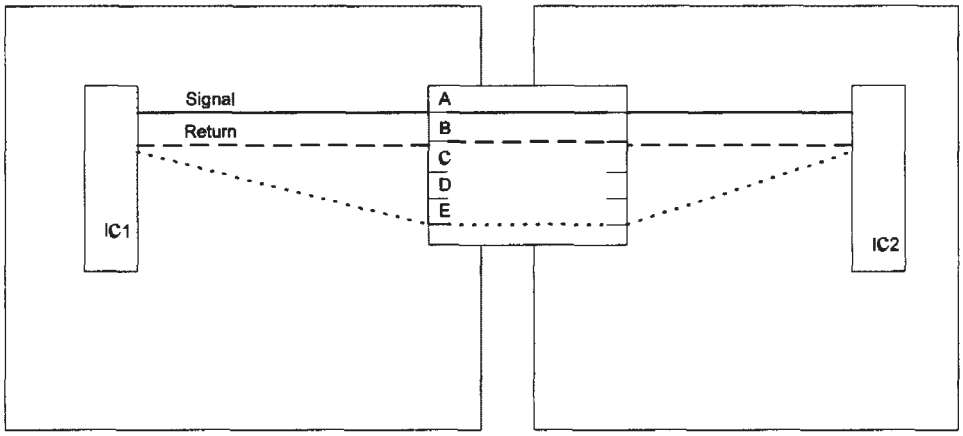


Figure 9-8 If provision is not made for a signal return (ground) through a connector, a larger loop area might be created.

Figure 9-9 illustrates another common problem. The figure shows a row of clearance holes for pins for perhaps a component or connector. In Figure 9-9a and

Figure 9-9b, a signal trace connects to one pin and the return (ground) pin is on the back side. In Figure 9-9a the clearance holes are wide enough to remove all the copper from the plane under the connector or the device. The return signal, traveling under the trace, must then travel around the copper void area to get to the ground pin. This creates a loop that might become an EMI problem.

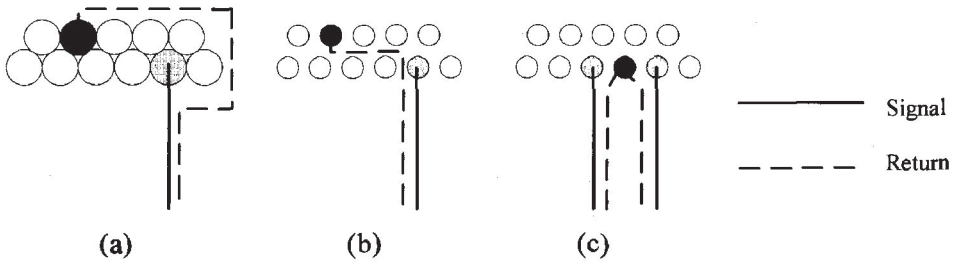


Figure 9-9 Good practice includes providing copper paths between through hole pins and assigning pins to minimize loop areas.

A better solution is shown in Figure 9-9b. The clearance holes are smaller, so that they do not remove all of the copper from the plane in this area. Now there is a path for the return signal to travel between the pins to the return (ground) pin. This approach is marginally better than Figure 9-9a and results in a smaller loop area. And this illustrates why it is usually preferable to provide a copper path on the planes through all areas like this.

However, Figure 9-9c shows an even better solution. If you have the flexibility to do so, it is best to assign pins so that there is always a ground (return) pin beside every signal pin. This provides the absolute minimum loop area for the signal and its return, resulting in the best EMI performance.

Power Plane Returns Now is the time to introduce an interesting complication to all this. We earlier pointed out that the return signal will follow the path of least impedance. That is the point on the plane closest to the signal trace. What if that plane happens to be a power plane rather than a ground plane? How does the return signal get back from the power plane to the ground pin of the device?

First, DC voltage has no meaning to a high-speed return signal. The return signal can be just as happy on a power plane as it is on a ground plane. From an AC

INTEL-1010
U.S. Patent 7,151,430

standpoint there is no difference between a power and a ground plane. The most obvious evidence of this point is all the bypass capacitors we have connecting the two planes together. From an AC standpoint the planes are shorted together all over the place.

So, what happens if the return is on a power plane? The exact answer may not be satisfying or may not even be known, but the practical answer is that the return signal probably transitions from the power plane to the ground plane through a nearby bypass cap. And if we have very high frequency harmonics, the transition may be through the planar capacitance between the two planes themselves, if such planar capacitance has been designed into the board.

It may not be a satisfying answer that the signal passes through a nearby bypass cap. After all, that leaves the definition of the loop area somewhat up to chance. I have heard a few people, for example Howard Johnson and Henry Ott (two highly respected experts in the field), sometimes advocate the placement of bypass caps near connectors or devices such as that shown in Figure 9-9 simply to provide a path for the return current. The bypass capacitor would not be intended to provide a decoupling function at all. Its sole purpose would be to control the loop area for the return currents on the power plane and therefore control EMI.

A few people advocate never routing such a signal adjacent to a power plane. They advocate routing all such signals adjacent (referencing) to only ground planes. Most people (including us at UltraCAD) consider this position too extreme. If your engineer (customer) insists on this constraint, however (after all, the customer *is* the boss), EMI and loop area are two possible reasons.

Changing Trace Layers Figure 9-10 illustrates another interesting case. Figure 9-10a shows a simple signal referenced to a plane. Figure 9-10b illustrates the trace transitioning to the other side of the same plane. In this case the return signal would simply transition to the other side of the same reference plane, and there should be no effect on loop area.

What about the case shown in Figure 9-10c? Here the signal transitions to an entirely different trace layer and reference plane. A legitimate question is, "How does the return signal get back under the trace on the new signal layer?"

INTEL-1010
U.S. Patent 7,151,430

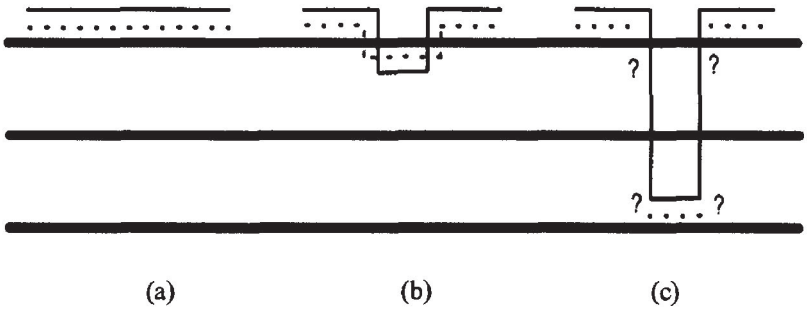


Figure 9-10 Some engineers are concerned about the location of the return path, and possible loop area problems, if signals are allowed to transition to different trace layers.

Again, the not very satisfying answer is through a nearby bypass cap. Again, some people advocate placing a bypass cap near each via for the single purpose of providing a path for the return signal and reducing loop area. A few circuit design engineers even prohibit transitions of traces from one trace layer to a completely different one because of the potential for increased loop areas and EMI problems (although most people would call this an extreme position to take).

On the designs we do at UltraCAD there are typically quite a few bypass capacitors. We generally take no particular special precautions when transitioning signals to different trace levels, nor do we pay particular attention to which plane (power or ground) the return signal is on. We are not aware of any customer design failures that can be traced to either of these practices.

Unrelated Planes Many boards have separate, well-defined areas on them for different power systems. There may be both 3-volt and 5-volt power supplies, for example, or perhaps separate analog and digital power supplies (or even transmitter and receiver power supplies, etc.). Often, in such cases, we define planar areas for these supplies.

Figure 9-11 illustrates a situation in which there are separate analog and digital power supplies and planes. Consider a trace that must be routed between the digital ICs 1 and 2. Ideally, the trace would be routed exclusively over the digital power plane.

INTEL-1010
U.S. Patent 7,151,430

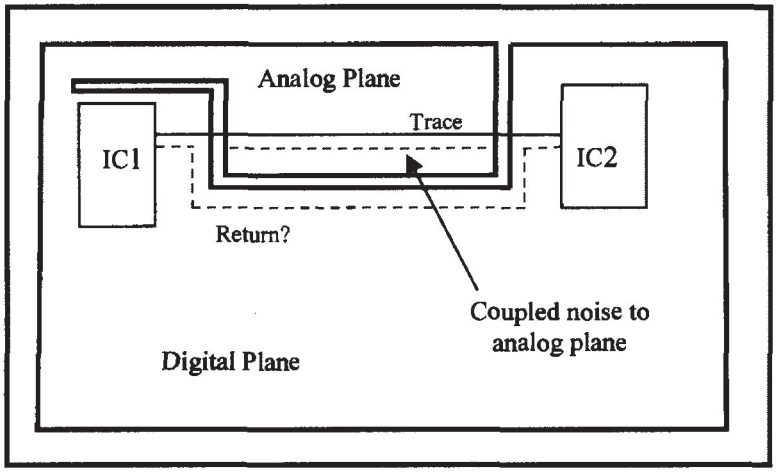


Figure 9-11 Never route a trace over an unrelated plane.

What if we took the shorter path and routed it over the analog plane? The question then is, "Where is the return signal?" There are two possible answers to that question, both of which are bad.

One possible answer is that the return signal stays on the digital plane and flows around the analog plane. But this creates a wider loop area and therefore potential EMI problems. Another possibility is that the return signal somehow finds a path onto the analog plane and continues on under the trace. We may not have an EMI problem in this case, but what we do have now is a digital signal on the analog plane. This digital signal might interfere with other analog signals in the area and create a crosstalk or ground noise problem. After all, the whole reason we have separate power systems is to keep the analog and digital signals separated in the first place. Either of these possibilities is bad. So the right answer is: *Don't route traces over unrelated planes.*

Stripline We intuitively know that a shielded wire (e.g., a coaxial cable) performs well from an EMI standpoint. We probably are aware that another defense against EMI is a shielded enclosure. So we can understand that traces shielded between planes probably perform well, too. This is true, within limits.

The single most important rule for EMI control is to control loop area. Even in a stripline environment (where traces are shielded between two planes) a large loop area can cause EMI problems both externally (EMI radiation) and internally

INTEL-1010
U.S. Patent 7,151,430

(crosstalk). But all other things being equal, a stripline environment adds an extra measure of EMI control. So, even though microstrip traces can be designed to perform fully satisfactorily from an EMI standpoint, designers sometimes opt to place all critical signal traces in a stripline environment for an added measure of safety.

STUBS

Stubs are short trace segments that extend, unterminated, from (an often controlled impedance or transmission line) main trace. Most commonly they connect a pad, and therefore a device, to a net. Figure 9-12 illustrates four stubs, S1 through S4. S1 and S3 are normal stubs that we all use. Stub S2 is too long and should be avoided in high-speed designs. Route the trace closer to the pad as shown for Stub S1.

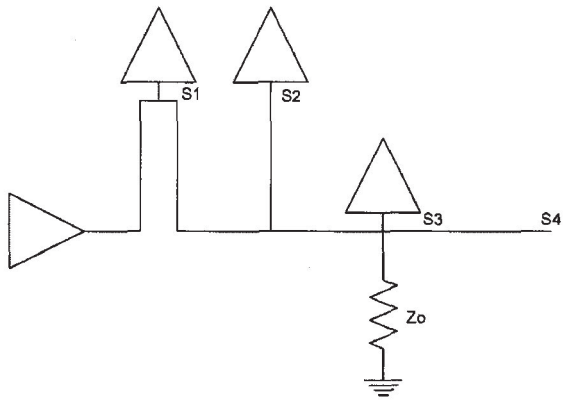


Figure 9-12 Various examples of stubs.

There are two reasons for avoiding long stubs. One relates to transmission lines and is covered in Chapter 10. The other relates to EMI.

Look at your cell phone. There is probably a short antenna extending out from the top. Now look at Stub S2 in Figure 9-12. This unterminated stub looks a lot like the antenna that extends out of your cell phone. Stubs like this can radiate just like antennas and should be avoided.

INTEL-1010
U.S. Patent 7,151,430

The stub shown as S4 is particularly harmful. This might have been created when a trace was rerouted and the designer forgot to remove a short segment from the original route. This *is* an antenna. Its results might not be harmful, depending on the circumstances, but they are never beneficial. Never allow such stubs to exist on your board.

Antenna theory is a very complex area that is not well understood by most people. There are a huge number of variables that affect whether a wire or trace can function well as an antenna. Some of these include the environment the wire or trace is in—the degree to which it is shielded by the planes, its length, the surrounding material and its relative dielectric coefficient, the specific shape of the wire or trace, and so on. Another variable is the frequency of the signal (or its harmonics). So it is difficult to look at a stub and say with certainty whether it will create a problem. But the general rule is that stubs create problems, so minimize them. Be sure to take the time to trim back any trace that is left "hanging" and not connected to anything.

COMMON MODE

Loop areas are one of the two most common causes of EMI. The other cause is referred to as common mode problems. The problem is that common mode is a poorly understood and frequently misused term. Before we can talk about this part of the EMI issue, we must spend a little time talking about the term and its misuse.

One of the most frequent explanations of common mode is made in contrast to another term, *differential mode* (not to be confused with differential traces or signals). Differential mode is the mode we are usually familiar with. That is, a signal travels down a well-defined path and returns, in the opposite direction, on another well-defined path. Since the signal and its return flow in opposite directions, the term differential is used. Following this logic, *common mode* then refers to the case where the signal flows in the same direction on both the trace and the return.

Since all signals, by Kirchhoff's Laws, must flow in closed loops, this explanation becomes troublesome when you really start to think about it. If common mode currents flow in the same direction, how do they return to their source? Then there is the confusion caused by the term differential mode in this context, and differential mode in the context of differential signaling and differential traces. We

INTEL-1010
U.S. Patent 7,151,430

will find later (in Chapter 14) that the term differential mode is actually misused in the latter case.

We are going to leave the differential signal case to Chapter 14 and talk about common mode EMI issues here only in the context of a single-ended, individual trace. But the extension of the problem to the case of differential traces, and then how to deal with it, will be somewhat apparent, even if the actual causes of the problem will not.

Let's define common mode a little differently from the definition we are normally familiar with. Let's use the term differential mode (in the context of a single-ended trace) to mean the currents that flow where we expect them to flow based on the schematic and the layout. If we are worried at all about EMI, we have already learned that we want to minimize loop areas and route traces directly over, and close to, reference planes. Therefore, the differential mode signals travel down the trace and return on the plane directly under the trace.

Common mode currents are those that flow anywhere else.

This raises several obvious questions. (a) How can this happen? (b) Even if it happens, what's the problem? (c) What's the relationship between this definition and the normal one generally used? (d) What, if anything, can be done about it?

How Can This Happen? Figure 9-13 illustrates one example of common mode currents based on our definition. A signal travels down the trace and returns on the ground plane directly underneath the trace. This is the differential mode signal current, id. Another part of the signal travels down the trace and returns through a completely different unintended path. This is called the common mode signal current, ic.

This can happen for several reasons. One is that the ground plane is not a perfect conductor. There is some inductance (admittedly small) associated with it. There can be a voltage divider action involved where the return current splits along the intended path (id) and some other paths (ic).

INTEL-1010
U.S. Patent 7,151,430

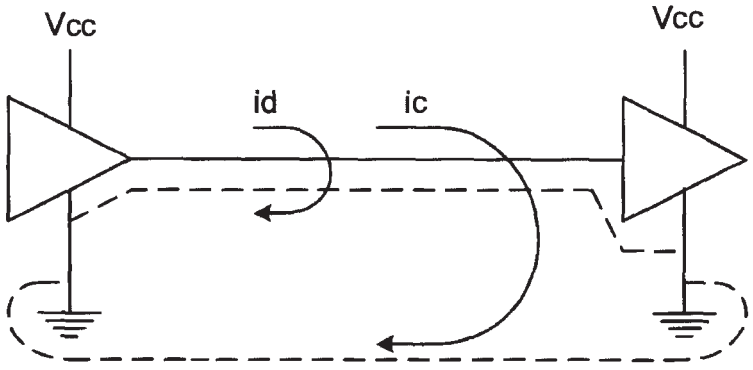


Figure 9-13 Differential mode, id, and common mode, ic, currents.

Or, perhaps there is a wire, stub, or cable shield (e.g., a coaxial cable) attached to the trace. Very fast rise-time transients may travel along such paths and radiate to other traces or surfaces. The characteristic impedance of a wire in space is generally known to be on the order of 370 Ω, which means the impedance along another unintended path could be in this range.

If the impedance along the plane is a fraction of an ohm and a radiated signal is on the order of 370 Ω, the amount of current flowing in the alternative path would only be a tiny fraction of the primary (differential) current. But it would not be zero. And that is the point. There can be unintended (although very small) common mode currents flowing in places we don't expect.

So What's the Problem? We have already learned in this section that to control EMI we want to minimize loop areas. The problem with common mode currents is we don't know where they are flowing. They may be flowing in very (relatively) large loop areas. If so, they may be radiating very significantly. So a common mode signal may only be a tiny fraction of the magnitude of a primary signal, but if it is radiating many times more loudly, it can still become a significant EMI problem.

Relationship to Conventional Definition Figure 9-14 illustrates the conventional definition of common mode current (ic). It is usually shown as appearing on both the trace and ground, flowing toward the receiver.

First, there is no necessary reason that it flows toward the receiver; common mode could just as easily exist on the trace and plane flowing away from the

INTEL-1010
U.S. Patent 7,151,430

receiver (though we might have trouble envisioning how such a signal might be generated).

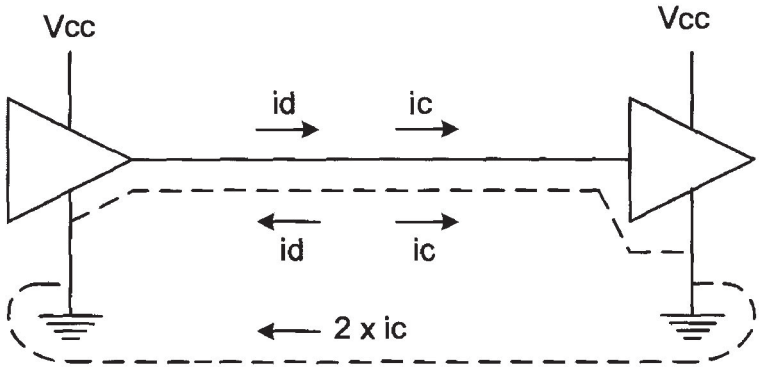


Figure 9-14 Any signal can be separated into differential and common mode components.

We suggested that common mode current is any current that flows in the circuit along a path other than the one we intended. *Any* combination of signal and return current can be broken into differential and common mode components suggested in Figure 9-14, assuming common mode currents exist at all. For example, assume that the signal current on the trace is 10,000 μA (10 mA) and the return signal on the plane under the trace is slightly smaller, say 9,950 μA. This, then, means that there is 50 μA flowing back to the driver along some other path. The differential and common mode components of Figure 9-14, therefore, under these assumptions, are id = 9,975 μA and ic = 25 μA.

What Do We Do About It? Since common mode currents are unintended, it seems like they would be difficult to control. Indeed, that can be the case. There are a few basic ground rules for controlling common mode problems:

1. Use good high-speed design techniques everywhere in your design.
2. In particular, maintain a solid, continuous reference plane for every trace.
3. Common mode problems are smaller for stripline traces contained between planes. If possible, route all critical traces in stripline environments.

INTEL-1010
U.S. Patent 7,151,430

4. Minimize the presence of any stubs or extraneous trace lengths associated with any high-speed trace. Try to eliminate any path for a signal to flow on other than the intended path.
5. Since all signals ultimately flow from the power plane through a driver to a receiver to the ground plane and back to the power plane, minimize this path length. One way to do this is to use power and ground plane pairs in your board stackup. Then ensure there is good capacitive decoupling between these two power sources.

THE 20-H RULE

A rule that is sometimes proposed has been nicknamed the 20-H rule. It has been credited originally to W. Michael King and a thorough explanation can be found in one of Mark Montrose's books.[1] The rule argues as follows: If there are high-speed currents on the board, there are electromagnetic fields associated with them. At the edge of the planes (presumably at the edge of the board) these fields will fringe outward from the board as shown in Figure 9-15a. But if one of the planes is recessed, as the power plane is in Figure 9-15b, the fringing tends to be directed downward to the other plane instead of outward. Thus, outward EMI radiation is reduced and there is less chance for an external EMI problem.

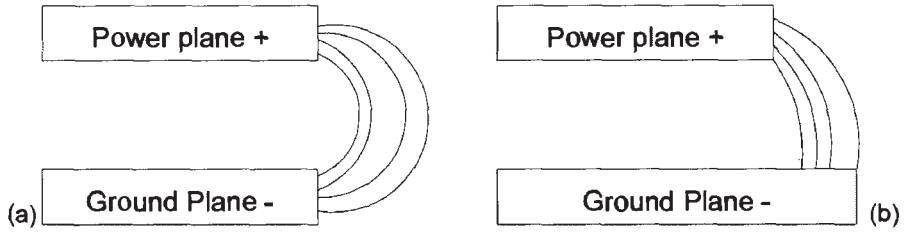


Figure 9-15 Recessing a plane (b) may reduce outward fringing.

The reduction in fringing is directly related to the amount one plane is recessed from the other. The argument goes that the reduction in field intensity is about 70% if the recess is 20 times the distance between the planes (H); hence the name of the rule.

[1] *Printed Circuit Board Design Techniques for EMC Compliance*, Montrose, Mark; IEEE Press, 1996, p. 26.

INTEL-1010
U.S. Patent 7,151,430

A few studies have tried to examine or measure this effect. In general they all seem to come to the same conclusion: The 20-H rule is not particularly effective.[2]

There is a certain intuitiveness to this conclusion. The high-speed signals are carried on the traces. On well-designed boards, the signal returns are confined to the planes directly under the traces. Therefore, there should not be a great deal of radiation at the edge of the board anyway, assuming we have done everything else well. At UltraCAD we have never used the 20-H rule on a board and we have had no particular EMI problems. Our position is that if this is the worst problem you have with your design, you are already in pretty good shape.

PICKET FENCES (FARADAY SHIELDS)

Some people advocate placing closely spaced vias between two ground planes to form a type of shield around the perimeter of the board. This technique is sometimes called a *picket fence* because of the obvious resemblance. The idea is that if any stray EMI is generated within the board, it can't get past the shield formed by the vias and radiate into the outside world.

People differ on how closely the vias should be placed, but you can hear estimates as close as 1/20 wavelength (which could be roughly every inch for a 1 ns rise time signal). Wavelength is discussed in Chapter 2.

Not all engineers like the idea of picket fences. UltraCAD has never made such a design for any of our customers. A picket fence isn't needed if there is no EMI generated in the first place. My philosophy and advice is to concentrate design efforts on preventing EMI generation in the first place, rather than in controlling it after it is generated.

[2]See, for example, "Effects of 20-H Rule and Shielding Vias on Electromagnetic Radiation From Printed Circuit Boards," Chen, Huabo and Fang, Jiayuan, Dept. of Electrical Engineering, University of California at Santa Cruz. They conclude, "For the two-plane structure, 20-H rule yields much more radiation than the normal structure. For the multiple plane case, no significant change in radiation is found if the 20-H rule is applied to the internal planes."

INTEL-1010
U.S. Patent 7,151,430



10 REFLECTIONS AND TRANSMISSION LINES

COMMUNICATIONS MODEL

*F*igure 10-1 illustrates what we sometimes call a general communications model. Any time there is a communication, we can think in terms of there being (a) a sender of that communication, (b) a message that is being sent, (c) the media over which the message is being sent, and (d) the (intended) receiver of the message.



Figure 10-1 A general communications model.

For example, at this moment you are reading something "sent" by Doug Brooks, author of this book. The message is what is written on this page. The media is the page you are reading, and the receiver is you, the reader.

On PCBs, the sender is the driver, the message is typically the change in state from a high to a low signal or from a low to a high signal, the media is the trace, and the receiver is the receiving circuit.

INTEL-1010
U.S. Patent 7,151,430

The evening news is "sent" by the news anchor (and his or her team behind the scenes). The message is what you see on TV, and might be designed to be entertaining as well as informative. The media can be thought of in simple terms as the television system. In reality it can be a much more complex system of microphones, cameras, switchers, satellites, retransmitters, and so on. The receivers are those watching the program.

In advertising, we think of the sender being the advertiser (e.g., Nike). The message is the advertising message. The media is the media chosen. The receivers are those targeted to receive the message. When we think about advertising in this context, we begin to think about some of the issues that advertisers face:

- How do you select which receivers to target?
- Where are they?
- What media will reach those targeted receivers most efficiently?
- How do we structure our message so that the receivers "hear" it while being bombarded with thousands of other messages at the same time from other senders trying to get their attention?
- How do we make sure the receiver "hears" the same message the sender thinks he or she is sending? It is fairly common for targets of advertising messages to perceive a different message than that which the advertiser intends to send.

Think about this communication problem: Picture yourself sitting down at the end of a high school gym at assembly time. The football coach is at the other end handing out the team awards. He is speaking through a handheld public address (PA) system. The floor is made of hardwood, the walls are concrete, the ceiling is metal, and the echoes are awful. It is almost impossible to understand what the coach is saying. The media is the primary cause of this particular communications problem.

Now, this time, picture yourself positioned along a PCB trace. The driver is sending a signal down the trace and you are trying to "hear" it. But here is the problem: Any time a signal travels down a wire or a trace, it reflects. Every time—(well, almost). Notice what I said and what I didn't say. I did not say "often" reflects. I did not say "sometimes" reflects. I did not say "under certain circumstances" it reflects. It *always* reflects. (Well, it turns out there is one very special case when it doesn't.

INTEL-1010
U.S. Patent 7,151,430

But this special case doesn't happen by accident, as you will see. It is this special case that is, in fact, the point of this chapter.)

So, in both our examples, the high school gym and the PCB trace, there are reflections and echoes that can interfere with the receiver understanding the message being sent. Here are some solutions to these communications problems:

1. We can encode the message, so we can pull it from the noise level. There is an entire science related to encoding and decoding messages in noisy environments, and this certainly a legitimate solution. It's not particularly practical in the gym environment, however, and probably not in our PCB environment, either.
2. We can listen "harder," or become better listeners. In the gym, we can concentrate harder on what the coach is saying. On our boards we can use better, more selective, but probably more expensive receivers. Again, these are legitimate solutions, but perhaps not desirable.
3. We can shorten the distance the message travels. In the gym, we can get up and move closer to the coach. We can shorten the trace on the board. On the circuit board, this is usually a highly desirable alternative, but we probably have already designed the traces about as short as they can be, so further shortening them is usually no longer an alternative.
4. A particularly interesting solution is to slow down the message. If the coach will speak more slowly it will clearly be easier to understand him.[1] The analogy on our boards is to use signals with slower rise and fall times. This, too, is a desirable solution. Many people recommend using circuits with the slowest rise and fall times we can get away with to avoid reflection (and most other high-speed design) problems. The problem is (a) we may not have control over the rise time of the circuits we are using and (b) we may have already chosen the fastest rise time circuits available to meet other circuit requirements.

These are four legitimate solutions to our communications problem, but none of them may be practical for our specific situation. But there is one more:

[1]See Appendix G for information on how to download three audio files that illustrate this point.

5. We can acoustically engineer the gym to absorb reflections. Next time you are in a hotel or convention center conference room or hall, look closely at the surroundings. Chances are very good the room has been acoustically engineered to absorb echoes and reflections. It will have a soft carpet, cloth-paneled walls, and sound-absorbing ceiling tiles. Speakers are probably placed where there will be no phase shift between the sound coming from the presenter (wherever he or she is) and where you are sitting.

Similarly, we can (not acoustically, but) electrically engineer our trace to absorb reflections. Then there will be no reflections going back and forth on the trace to interfere with the receiver's ability to receive and understand the message the sender is sending.

The problem is this: In general, we don't know how to do that. We do know how to acoustically engineer a room, but we don't know how to electrically engineer a wire or trace to absorb reflections—with one very special exception. We do know what to do if we are dealing with what we call a *transmission line.*

TRANSMISSION LINES

So what makes transmission lines so special? Consider a long, straight wire or trace with its return wire or trace nearby. The wire has some inductance along its length. There is also some capacitive coupling between the wire and its return (see Figure 10-2). Figure 10-2 shows what we call a lumped model of the wire pair. It is called a lumped model because we show the capacitors and inductors as individual, lumped components. In reality, the inductance and capacitance are spread continuously along the wires. We don't know how to show that in a drawing, so we approximate it with a lumped approximation. On the other hand, this is fine because we are going to think of this wire as being infinitely long. Therefore, the part shown in Figure 10-2 is just an infinitely small part of the total length.

INTEL-1010
U.S. Patent 7,151,430

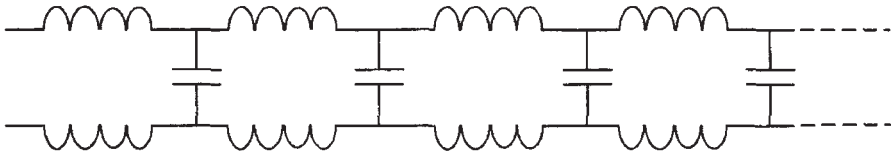


Figure 10-2 A transmission line is made up of an infinitely long network of capacitors and inductors.

Now, if these wires are infinitely long, there will be no reflection. At least if there is a reflection, it will take an infinitely long time for it to return, so we can assume it doesn't exist.

It takes at least one other thing to avoid reflections. The wires must be absolutely uniform. The little Ls and Cs must be identical everywhere along the wire. If the wires are not uniform, then we can consider them to be multiple sets of wires, each with different characteristics. Individual sets of wires will not be infinitely long, and therefore there will be reflections from each individual set. Therefore, the way to avoid reflections is to use an infinitely long, absolutely uniform wire or trace pair. We give such a wire or trace pair the special name *transmission line.*

It can be shown mathematically that if we look into the front of this wire pair there is an input impedance we can calculate. We give it the symbol Zo, and call it the *intrinsic impedance* of the line. If we could calculate the "lumped" values of inductance (L) and capacitance (C), the impedance would be calculated as shown in Equation 10-1:

$$Zo = \sqrt{L/C} \qquad [10\text{-}1]$$

Now here is a clever twist. If you look into the transmission line at the front, it looks like it has an impedance of Zo. Let's take our infinitely long transmission line and break it in two parts. If we look into the second part, it also looks like an infinitely long transmission line with an input impedance of Zo. What if we simply replaced the second part of the line with an impedance equal to Zo (see Figure 10-3)? From the front of the first line, it still looks exactly like an infinitely long line. It turns out it behaves that way, too. *A transmission line of finite length, terminated in its characteristic impedance, Zo, looks like an infinitely long transmission line.* Therefore, even though it has a finite length, it still will have no reflections. In this case it is not the infinite length that makes reflections irrelevant. It is the fact

that the energy traveling down the line is exactly absorbed (dissipated) in the termination. There is no energy left to reflect back. The net effect is the same thing. There are no reflections to worry about.

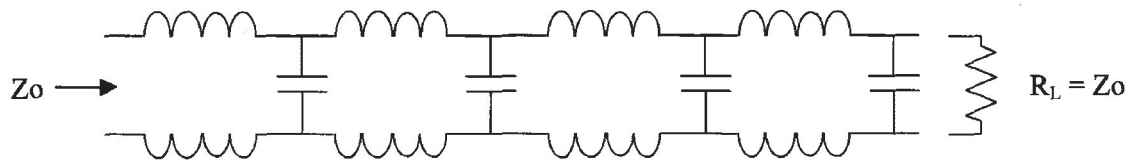


Figure 10-3 If we terminate a finite-length transmission line in its characteristic impedance, it looks infinitely long.

These are the two actions we need to take to control reflections on PCBs:

1. We need to make our traces look like transmission lines.
2. We need to terminate them in their characteristic impedances, Zo.

Of course, we can also make our traces short enough that reflections don't interfere with the receiver's ability to hear the signal in the first place, or we can slow down the rise time enough that the receiver can hear through the echoes.

There are certain types of transmission lines that are commonly used around us everyday. The coaxial cable leading to our cable TV is a 75-Ω transmission line. If you use 10Base2 coaxial cable for networking, that is a 50-Ω transmission line. The 300-Ω "twin lead" cable from your TV "rabbit ear" antenna to your TV is a transmission line. And it is no accident that those high-power electrical lines from the power generating plants to our cities are called transmission lines, strung along transmission line towers. Even power system engineers need to worry about reflections if the lines are long enough (about 480 miles or so).

The rest of this chapter is concerned with how to make our traces look like transmission lines, and what the implications are if we don't do that.

INTEL-1010
U.S. Patent 7,151,430

CRITICAL LENGTH

I have suggested that reflections (echoes) are not too bad a problem if the receiver is close enough to the sender. There is also less of a problem if the message is slowed down (the rise time is slowed down, or lengthened). If we think of a signal traveling down a trace, these two things are actually equivalent. The issue is this: What is the length of the signal path (in propagation time) relative to the rise (or fall) time of the signal?

For example, consider Figure 10-4. There are two cases shown, Driver A is driving a signal to Receiver B1, or to both Receivers B1 and to B2. B2 is much further away from the driver than B1.

The driver signal propagates down the trace. The horizontal axis represents both distance and time, because all signals on a particular trace layer (usually) propagate at the same, constant speed. Pick a spot (x) on the lower portion of the rising edge of the driver signal. Now move horizontally from that spot to the slope representing B1. This horizontal distance represents the propagation time between the driver and B1. Move horizontally an equal distance again, and then move up vertically to the driver slope (Point y). Notice that the driver is still driving at this point in time. We can think of the driver having momentum and overpowering the reflection, so the reflection has no effect on the circuit. (This is not an entirely accurate view of the dynamics going on here, but this helps to visualize what is happening.) B1 is "close" to the driver in this example.

On the other hand, move horizontally from the same point, x, to the point representing B2. This represents the propagation time to B2. Now, extend beyond B2 another similar increment in time, and then rise up vertically to the driver signal. By now, the driver signal has long since stabilized at its upper value, and the reflection may be significantly apparent on the trace, possibly disrupting the operation of the circuit. B2 is a "long" way away from the driver in this example.

The distance from the driver to B1 is intuitively short in this example, and the distance to B2 is intuitively long. Reflections have minimal impact on short traces and may have significant impact on long traces. How do we define the difference between a short trace and a long trace? The general rule of thumb accepted by most people goes like this:

INTEL-1010
U.S. Patent 7,151,430

- If the propagation time down the trace and back again is less than the rise time of the signal, then we can consider the trace to be short.

- If the propagation time down the trace and back again is longer than the rise time of the signal, then we must consider the trace to be long, and we must therefore consider whether terminations might be necessary.

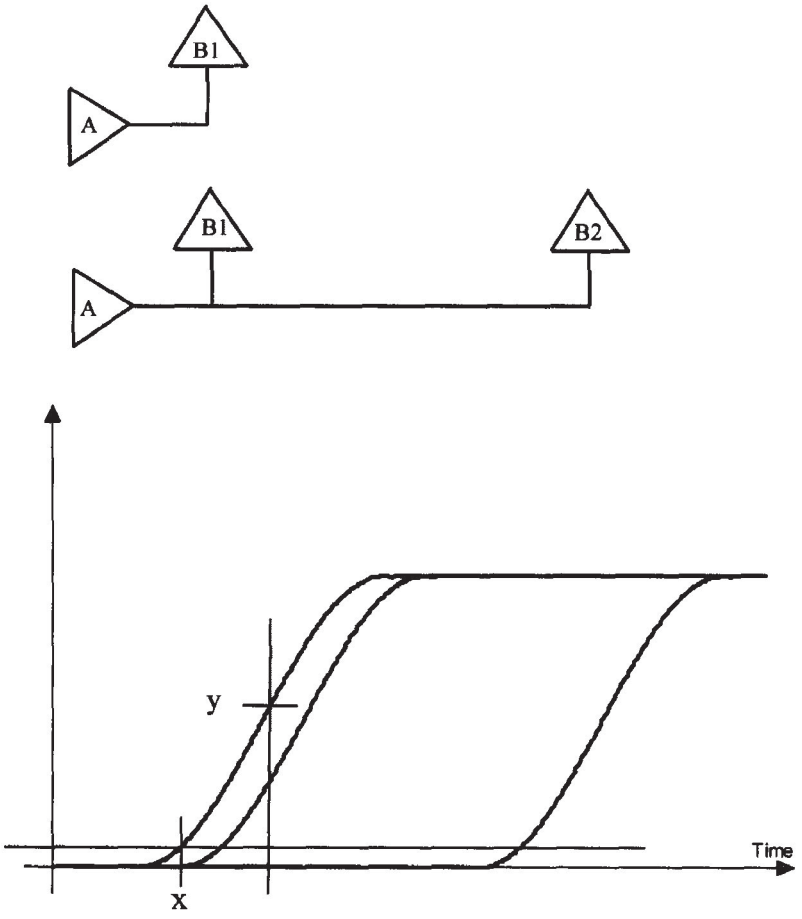


Figure 10-4 Receiver B1 is within the critical length, but B2 is beyond it.

The boundary between these two lengths is called the *critical length* and is often defined as the length where the two-way delay of the line is less than the rise time of the pulse.

REFLECTION COEFFICIENTS

What causes reflections and what do they look like? If we construct our traces to look like transmission lines, we call them *controlled impedance* traces. We then should be able to control reflections. But if we don't control them correctly, reflections may still exist. The question is why and how large might they be.

If we have (tried to) use controlled impedance traces (transmission lines), at least two things might still cause a reflection in the system: (a) an impedance discontinuity along the line, and (b) an improper termination. Recall that the characteristic impedance of a transmission line is defined by its geometry (and the dielectric coefficient of the material surrounding the line). Any change in the geometry, or in the dielectric coefficient, will cause a change in the characteristic impedance of the line at that point. That may (and probably will) cause a reflection at that point. It is imperative that the geometry of our transmission lines be held constant everywhere along the line. (The implications of having lines change layers will be covered in a minute.) Also, if we want to control the characteristic impedance of a trace over several board fabrication runs, we need to fully specify (ideally in conjunction with the board fabricator) the materials to be used and their relative dielectric coefficient. If the geometry remains constant, but the materials change from run to run, the impedance will also change from run to run. At a minimum, this will require changing the termination values from run to run, something that might not even be possible.

Suppose we take as a given that the trace has been designed and constructed carefully with a constant and controlled characteristic impedance. If it is terminated correctly, there will be no reflection from the far end. But if the terminating resistor is not selected properly, there will be a reflection. The magnitude of the reflection is determined by a measure we call the *reflection coefficient*, ρ. In the case of a single, parallel terminating resistor at the far end of the trace, the voltage reflection coefficient is defined as shown in Equation 10-2:

$$\rho = \frac{RL - Zo}{RL + Zo} \qquad [10\text{-}2]$$

where RL is the terminating, or load, resistor.

Note that if RL = Zo, then the magnitude of the reflection coefficient goes to zero, and there is no reflection. If the trace is left unterminated (i.e., RL is infinite), then the reflection coefficient becomes RL/RL = 1. This means the reflection is 100% and it is positive in sign. So, if we send a 3-volt signal down an unterminated trace, the initial reflection back will be 3 volts for a total of 6 volts on the trace (at least at that moment).

Note also that if we short-circuit the trace, so that the terminating resistance is zero, then the reflection coefficient will become –Zo/Zo = –1. This means the reflection will be 100% but with the opposite sign. So, if we send the same 3-volt signal down a shorted trace, the initial reflection will be –3 volts, and the net signal on the trace will be zero volts (at least at that precise moment). (You might ask how you can put a voltage down a shorted trace. Remember that there is a propagation time before the signal reaches the far end. When the signal starts down the trace, it simply sees the characteristic impedance of the trace, Zo. It doesn't know the trace is shorted. It doesn't find that out until it reaches the far end. We're dealing with very short transient times here.)

The value of the reflection coefficient can range between –1 and +1. It is +1 for an open trace and –1 for a shorted trace. A reflection coefficient value of zero means the trace is "perfectly" terminated.

VISUALIZING REFLECTIONS

It can be difficult to visualize what reflections look like. One reason is because it depends on where you are looking at them. For example, do you want to see their impact along the trace, at the far end, or back at the near end? Do you want to look at them over time or at a single point in time? What does the basic signal look like before the reflections begin interfering with it? Are there multiple signals involved that are all interacting? UltraCAD has created a Transmission Line Simulator that you can download from the Web, described in Appendix F (Figure 10-5). This is one tool that can be used to help visualize the effects of various termination situations and problems.

INTEL-1010
U.S. Patent 7,151,430

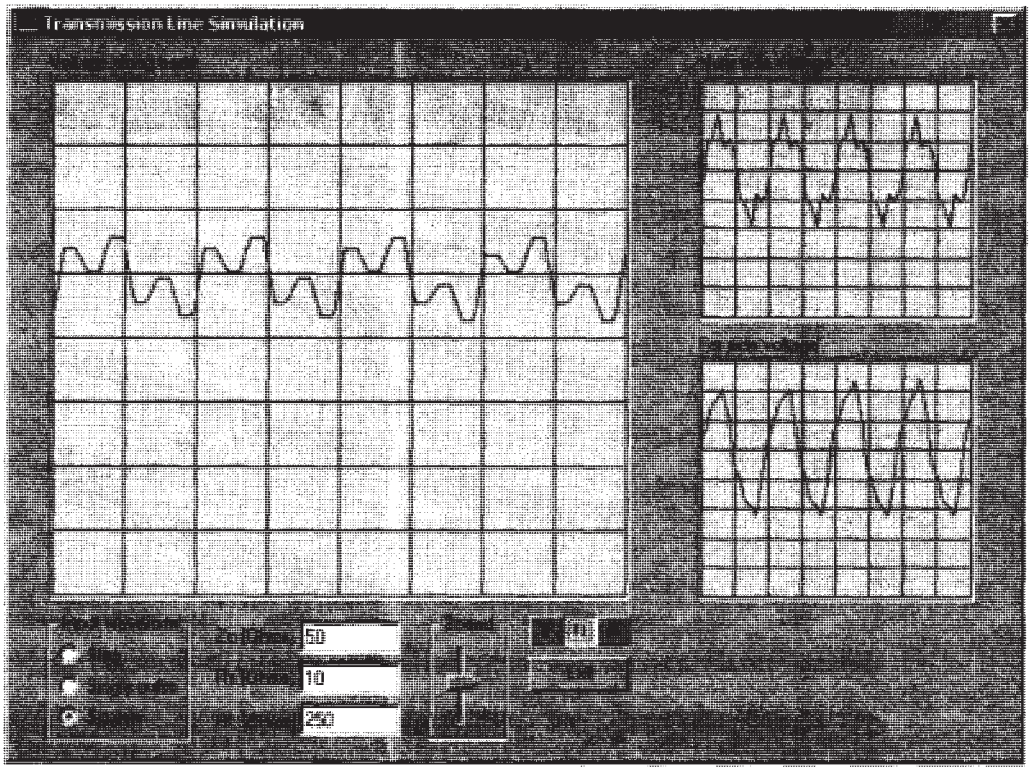


Figure 10-5 UltraCAD's Transmission Line Simulator, illustrating distortion along a trace cause by reflections. The driving input signal is a square wave.

Various commercial simulators exist to model transmission lines. One of them, Signal Vision, is used to illustrate various termination situations in the next chapter. Figures 10-6 and 10-7 illustrate simulations of terminations that are too low and too high, from both the Signal Vision tool and from the Hyperlynx LineSim tool. Both tools are products of Mentor Graphics.

DETERMINING TRACE IMPEDANCE

With the exception of the effect of the material's relative dielectric coefficient (ε_r), the characteristic impedance of a trace is determined purely by its geometry. That means the board designer has complete control over it (at least in theory). The formulas for determining and defining impedance are very complex. The

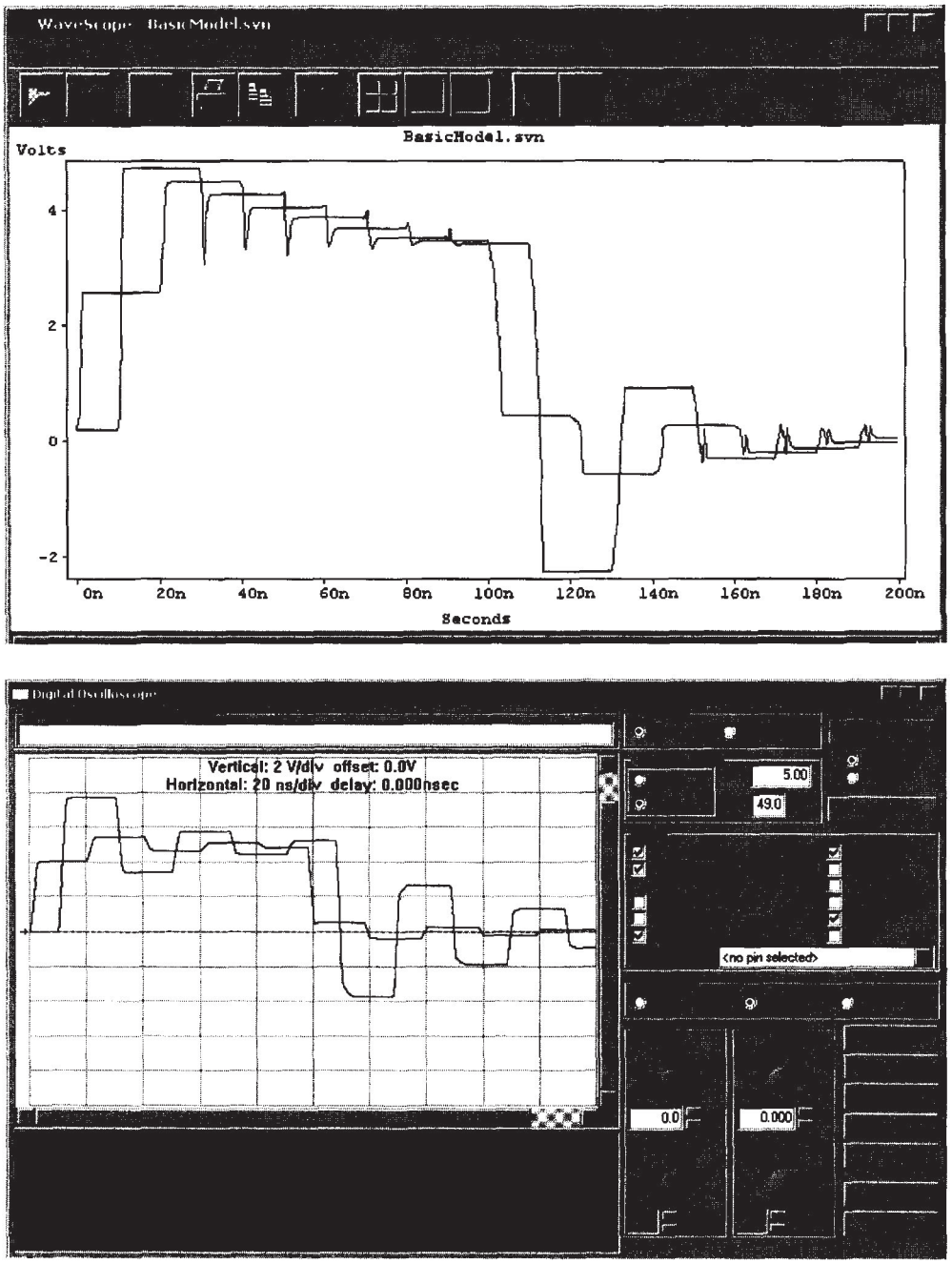


Figure 10-6 Simulation of a driver driving a 50-Ω transmission line with 1,000-Ω termination. Upper illustration is from the Signal Vision tool, lower one from the Hyperlynx LineSim tool.

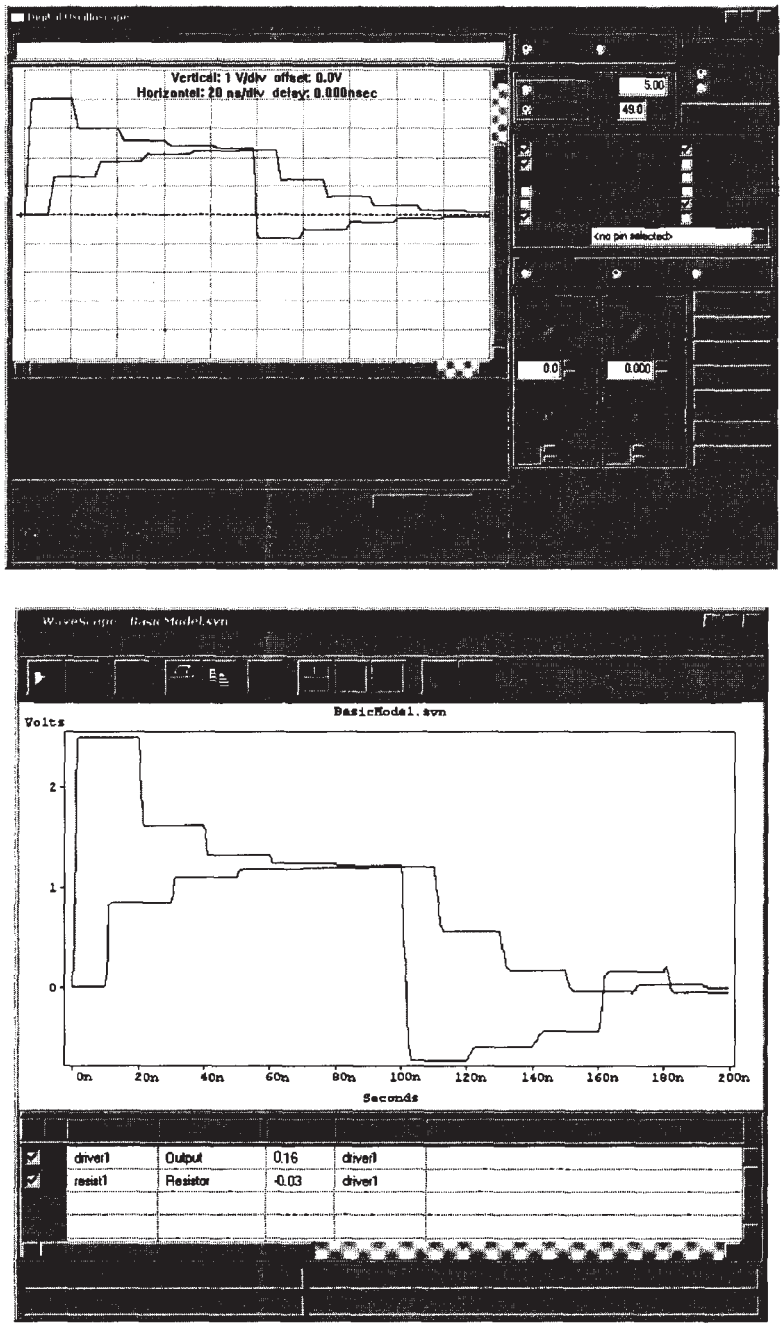


Figure 10-7 Simulation of a driver driving a 50-Ω transmission line with 10-Ω termination. Upper illustration is from the Signal Vision tool, lower one from the Hyperlynx LineSim tool.