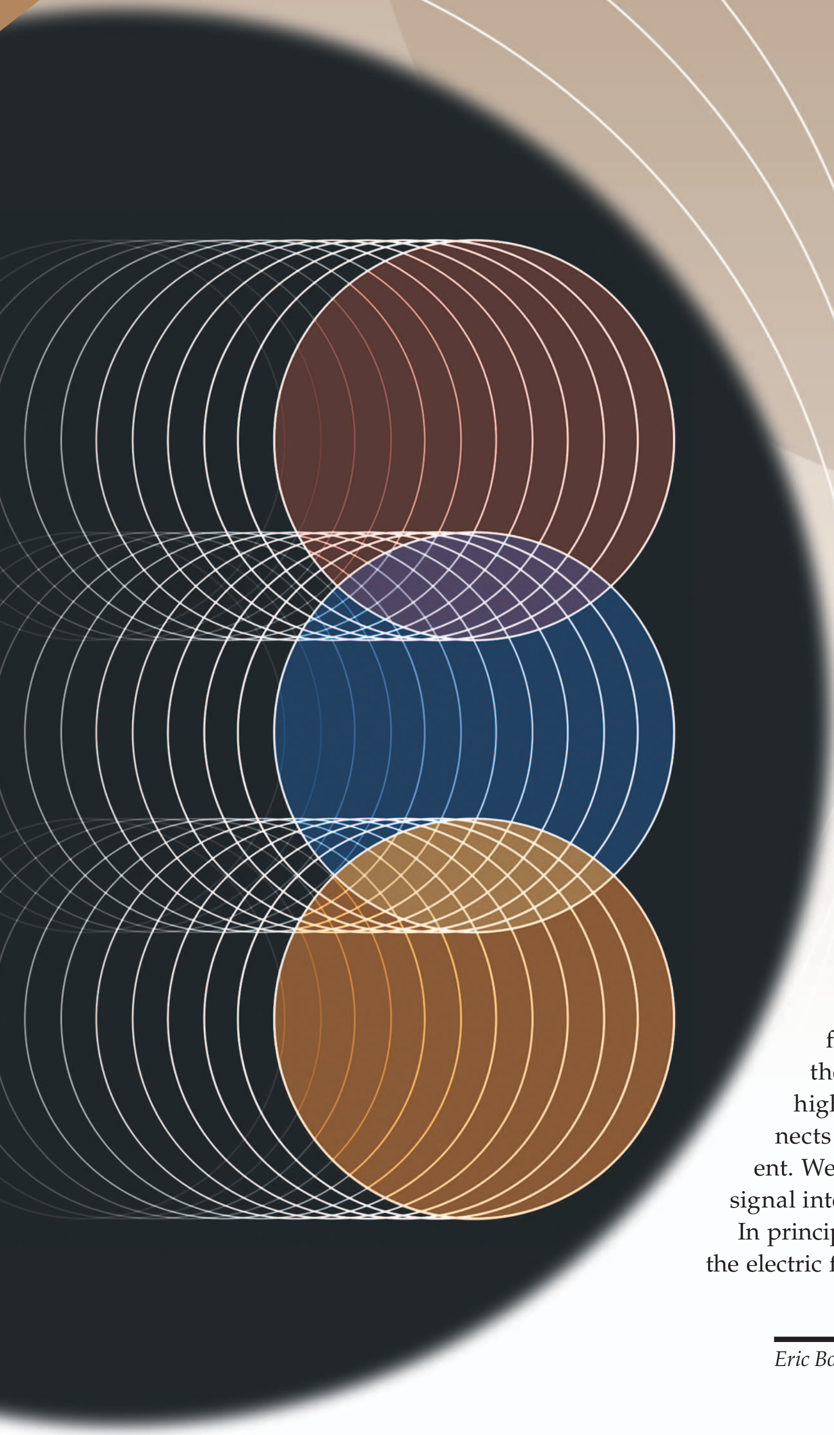
FOCUSED ISSUE FEATURE

# Essential Principles of Signal Integrity

*Eric Bogatin*





All electronic products have one common theme, there is a driver outputting a signal to a receiver with a defined signal and noise sensitivity. Between the transmitter and receiver are the interconnects. The purpose of the interconnects is to transmit the signal while keeping the noise to an acceptable level. The interconnects, shown in Figure 1, the packages, solder balls, connectors, circuit board traces, cables, and even the component leads, can never improve performance over what the active devices can deliver, but they can dramatically reduce performance, especially at higher clock frequencies. It is the analog impact interconnects have on digital signals that make them not transparent. We lump these analog effects in the general category of signal integrity [1].

In principle, all we have to do is solve Maxwell's equations for the electric field, $E$, and the magnetic field, $H$, given the boundary

*Eric Bogatin (eric@bethesignal.com) is with Bogatin Enterprises, 26235 W. 110th Terr., Olathe, Kansas 66061 USA.*

*Digital Object Identifier 10.1109/MMM.2011.941411*
*Date of publication: 7 July 2011*

 1527-3342/11/$26.00©2011 IEEE 

**Performance is best when all interconnects are uniform transmission lines and discontinuities are as short as possible.**

conditions of the conductors and distribution of the dielectrics, and we will know everything about how the signals interact with the interconnect. This might result in an accurate display of the *E* and *H* fields in some combination of time, space, or frequency domains, but using this information to make practical design decisions would be daunting [2], [3]. In practice, no tools allow this level of analysis except for small structures, and the interpretation of the performance of the *E* and *H* fields is still relegated to more advanced engineers.

Instead, as a practical approach, we approximate the real world of fields and boundary conditions into a simpler world of voltages, currents, and circuit elements. If we can describe the interconnects as circuit elements and how the signals, as described by voltages and currents, interact with them, we can sometimes get to an acceptable answer faster.

Using just nine essential principles, based on the concepts of voltage, currents, and circuit elements with a little field theory thrown in, problems such as ground bounce, electromagnetic compatibility (EMC), cross talk, ringing, collapse of the eye, and power distribution network (PDN) noise can be easily understood and eliminated from your next design.

## #1 All Interconnects Are Transmission Lines

The first and most important approximation to the real world is that all interconnects behave like transmission lines. A transmission line is nothing more than two conductors, which we usually label as the signal conductor and the return conductor. This is illustrated in Figure 2. It's a good habit to forget the word ground and use the term return path.

There are uniform transmission lines, which have a constant cross section down their length, and nonuniform transmission lines, which we generally call a discontinuity. As will become apparent, signals propagate with little change down uniform transmission lines but are dramatically distorted by discontinuities. Performance is best when all interconnects are designed, wherever possible, as uniform transmission lines and discontinuities are kept as short as possible.

## #2 Signals Are Dynamic

In this view of all interconnects as transmission lines, signals are voltages between the signal and return path conductor. At any instant of time, we can imagine walking up and down a transmission line and mapping the voltage distribution between the signal and return conductors, as shown in Figure 3.

This voltage pattern, frozen in time, may vary up and down the length of the line. As we advance time, this pattern will also advance down the line. Reflections, noise sources, or other drivers may generate other signals propagating in the opposite direction on the same line. No matter what their origin, voltages on a transmission line will always be in constant motion. They are unstoppable.

Signals propagate in the transmission lines at the speed of light in the material, which is reduced from the speed of light in air, by the dielectric constant of the medium, given by

$$v = \frac{c}{\sqrt{Dk}} = \frac{29.9\ {}^{\text{cm}}/_{\text{ns}}}{\sqrt{4}} = 15\ {}^{\text{cm}}/_{\text{ns}} = 6\ {}^{\text{in}}/_{\text{ns}},$$

where *v* is the speed of the signal on the transmission line, *c* is the speed of light in air, and *Dk* is the dielectric constant of the medium. For most circuit board

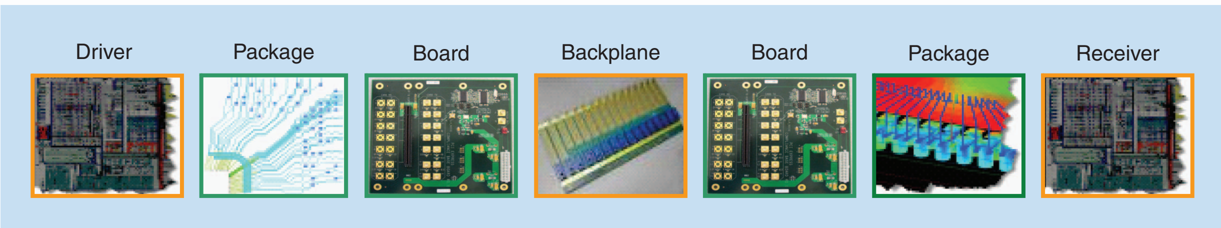



**Figure 1.** *The signal path has interconnect elements that are not transparent.*

**No matter what their origin, voltages on a transmission line will always be in constant motion.**

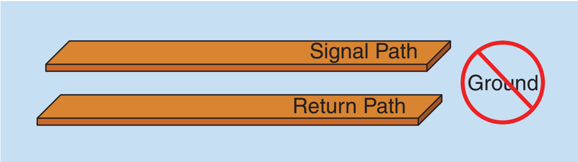



**Figure 2.** *A transmission line is composed of two conductors, a signal conductor and a return conductor. Avoid calling the return conductor ground.*

material with a *Dk* = 4, the speed of a signal is about 15 cm/ns (6 in/ns). If a signal has a rise time of 0.1 ns, it will have a spatial extent of about 15 cm/ns (6 in/ns) × 0.1 ns = 1.5 cm (0.6 in) as it propagates down the line. This 1.5-cm (0.6-in) long edge will interact with features of the transmission line and respond.

It is the dynamic propagation of signals on a transmission line that causes the finite delay between a driver and receiver, the changing voltage pattern of received signals, the impact of discontinuities in one part of the circuit affecting other parts, and even the specific noise signature seen in near- or far-end cross talk.

## #3 Signals See an Instantaneous Impedance

The most important property of the interconnect a signal cares about is the impedance it encounters, each step along the path. The impedance the signal sees determines the current associated with the voltage signal and whether the signal continues undistorted or if there will be a reflection.

Impedance, fundamentally, no matter what, is the ratio of the voltage across a device to the current through it. As a signal propagates down a line, it is constantly probing this impedance between the signal and return path by looking at the voltage applied and the current driven required to support the propagating voltage.

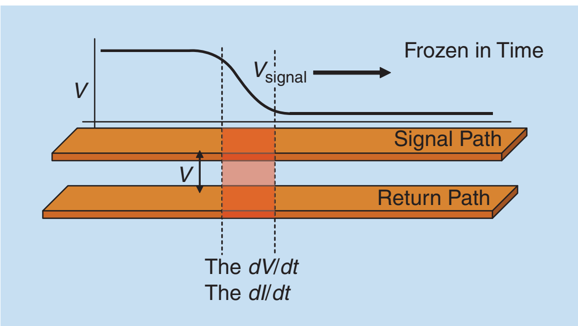



**Figure 3.** *Voltage distribution on a transmission line frozen in time.*

As the changing voltage, which is the signal, propagates down the transmission line, it leaves in its wake excess charge between the signal and return path. This is after all what causes the voltage difference between the signal and return conductors.

In a uniform line, there is a constant capacitance per length between the signal and return paths and, for a constant voltage down the line, represents a uniform distribution of excess charge. As the signal propagates down the line, the signal is carrying with it a constant current to successively charge up each region of the line. With a constant current flowing down the line to charge up each successive regions of the line, the signal will see a constant impedance.

We refer to this impedance the signal sees at each step as the instantaneous impedance. As long as the transmission line is uniform, the instantaneous impedance is constant down the transmission line. This impedance is, fundamentally, the ratio of the voltage to the current driven down the line and is given by

$$Z = \frac{V}{I} = \frac{V}{vC_L V} = \frac{1}{vC_L},$$

where $Z$ is the instantaneous impedance the signal sees, $V$ is the voltage across the signal and return conductors at any location, $I$ is the current flowing between the signal and return conductors, $v$ is the speed of the signal, and $C_L$ is the capacitance per length of the transmission line [1].

For example, when the capacitance per length of the transmission line is 3.4 pF/in in a typical circuit board interconnect with a *Dk* = 4, the instantaneous impedance of the transmission line is

$$Z = \frac{1}{vC_L} = \frac{1}{6\,^{\text{in}}/_{\text{ns}} \times 3.4\,^{pF}/_{\text{in}}} = 49\,\Omega.$$

The instantaneous impedance of the transmission line is the most important property of a line to which a signal reacts. It defines the primary signal quality of the line and is an important metric determining the distortions of signals.

When the instantaneous impedance the signal sees is constant down the line, the signal will propagate undistorted. When the instantaneous impedance changes, reflections occur and the propagating signal is distorted. This is why designing uniform transmission lines is so important.

To understand the impact on signal quality from a specific interconnect, it is useful to "be the signal" in a Zen sort of way, to imagine what instantaneous impedance the signal will encounter each step along its way. This is illustrated in Figure 4. The more constant this instantaneous impedance is, through all of the twists and turns, ups and downs, and ins and outs, the fewer reflections and less the signal distortion.

In a uniform transmission line, the signal sees an instantaneous impedance each step of the way, from one end to the other end of the uniform line. There is only one value of the instantaneous impedance for a uniform transmission line. The transmission line can be characterized with one value of instantaneous impedance. We label the value of this one instantaneous impedance that characterizes the transmission line, the "characteristic impedance" of the transmission line.

By definition, characteristic impedance only applies to a uniform transmission line. If the interconnect is not uniform, the instantaneous impedance will vary down its length and there is no one value of the instantaneous impedance that characterizes it. The characteristic impedance of a uniform transmission line is one of the two terms that describe it, its time delay being the second term.

**Impedance is the ratio of the voltage across a device to the current through it.**

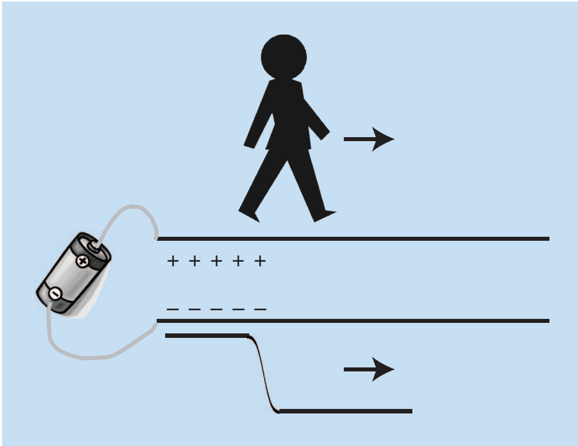



**Figure 4.** *In a Zen sort of way, "be the signal" and walk down the transmission line to see the instantaneous impedance the signal would see.*

## #4 The Return Current Is Just as Important as the Signal Current

A signal is the propagating voltage between the signal and return conductors. Since this signal sees an instantaneous impedance each step along the way, there will also be a current associated with the signal.

This current flows down the signal conductor and back to the source through the return conductor. However, the current always flows in a complete loop at any instant in time. This means that, at the edge of the propagating signal, there will be a current loop completing the circuit and flowing between the signal and return conductors, coincident with the signal edge. As the voltage edge propagates down the transmission line, the current loop flowing between the signal and return conductors propagates down the transmission line as well.

How is it possible for the current to flow through the insulating dielectric between the signal and return conductors at the wave front edge? The mechanism is exactly the same as how current flows through a capacitor. After all, there is insulating dielectric between two conductors in a capacitor, yet current flows between them.

However, the only time current flows through a capacitor is when the voltage between the conductors change. The changing voltage means the electric field between the conductors is changing. James Clerk Maxwell coined the term "displacement current" to refer to the apparent current that flows through even empty space when the electric field changes [3].

The voltage across a capacitor increases when extra positive charges are added to one conductor and extra negative charges are added to the other conductor. The added positive charges are positive currents flowing into one conductor and the extra negative charges are really a positive current flowing out of the other one. This appears to the outside world as a positive current flowing into one conductor and out the other through the capacitor but only when the voltage between the conductors changes.

In the same way, as the rising or falling edge of the signal propagates down the transmission line, the *dV*/*dt*, or changing voltage of the edge, drives displacement current between the signal and return conductors. The total current that flows across the rising edge wave front is the total signal swing divided by the instantaneous impedance of the transmission line. A 1-V signal into a 50-Ω impedance line will have a 20-mA current loop flowing between the signal and return paths, as shown in Figure 5. This 20-mA current loop wave

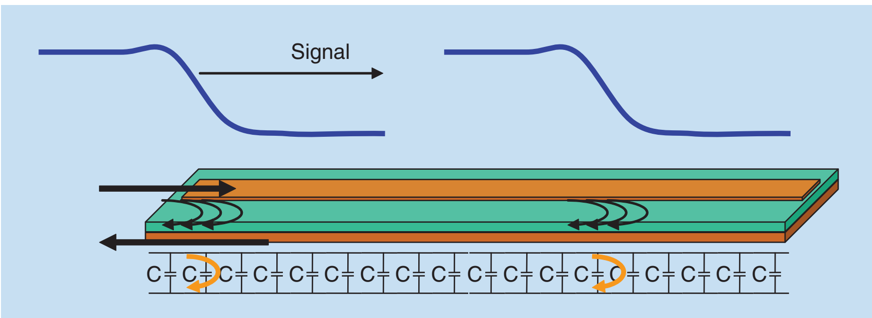



**Figure 5.** *The return current flowing between the signal and return paths at two instants in time: when the signal launches into the line and when it is midway down the line.*

## Any interconnect feature that affects the return current will also affect the signal current.

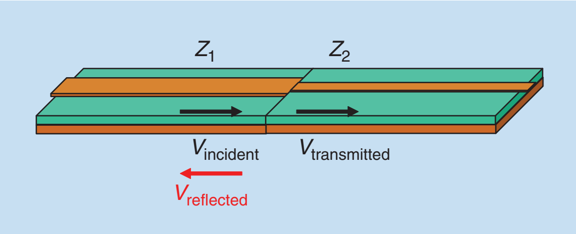



**Figure 6.** *A reflected voltage is created at the interface between two impedances to match the boundary conditions of continuous voltage and equal current going into and out of the interface.*

front, between the signal and return conductors, flows down the transmission line, coincident with the *dV/dt* edge of the signal.

The current loop at the wave front edge really has two directions associated with its motion. It propagates on the transmission line in the direction of the signal: left or right, for example. In addition, the current loop has a direction of circulation.

At the wave front, a positive signal, propagating from left to right, is a current loop circulating in the clockwise (CW) direction. A negative signal, propagating from left to right is a current loop flowing from the return, up to the signal conductor, circulating in the counterclockwise (CCW) direction.

Our model of a signal now shows it as really both a voltage wave propagating down a transmission line and a circulating current loop at the wave front. The instantaneous impedance the signal sees is the ratio of the voltage to the current at the wave front. Any interconnect feature that affects the return current will also affect the signal current.

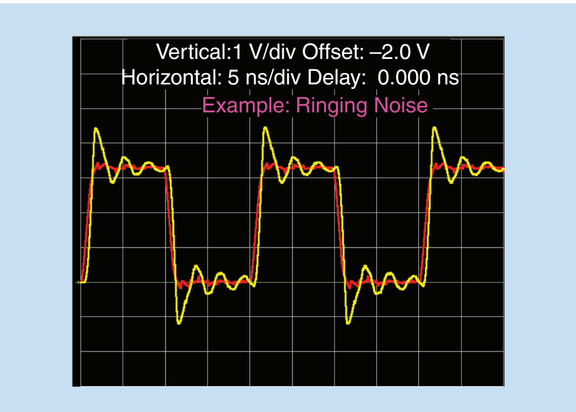



**Figure 7.** *Typical ringing noise from multiple reflections between the two unterminated ends of a transmission line.*

Controlling the impedance the signal sees is just as much about controlling the current in the return path as the signal path.

### #5 Reflections Occur Whenever the Instantaneous Impedance Changes

As the voltage edge, and its circulating current loop, propagate down the transmission line, it sees the instantaneous impedance of each step. If the instantaneous impedance remains constant, the signal will continue propagating at roughly 15 cm/ns (6 in/ns), undistorted. This is why uniform transmission lines are such important interconnects. However, if the instantaneous impedance changes, for whatever reason, some of the signal will reflect and what continues will be distorted. This has to happen to keep the universe from exploding.

At the interface, where the instantaneous impedance changes, two boundary conditions must be met: the voltage between two adjacent points must be slowly varying, and the net current flowing into one interface must equal the net current flowing out. If these two conditions are not met, either an electric field will reach an infinite level or charge will build up to an infinite level, either of which will destroy the universe.

To prevent this, a reflected wave is created at the boundary so that the incident voltage and reflected voltage on one side of the boundary create a net voltage that exactly matches the net voltage on the other side of the boundary. This is illustrated in Figure 6. And the net current flowing into the interface from the incident signal is exactly equal to the current flowing out of the boundary in the reflected wave and the transmitted wave.

The reflection coefficient, $\rho$, and the transmission coefficient, $t$, can be derived as

$$\rho = \frac{V_{\text{reflected}}}{V_{\text{incident}}} = \frac{Z_2 - Z_1}{Z_2 + Z_1}$$

$$t = \frac{V_{\text{transmitted}}}{V_{\text{incident}}} = \frac{2Z_2}{Z_2 + Z_1}.$$

This happens at every boundary and to every instantaneous voltage point in the wave front that propagates down the line. Of course, if there are multiple interfaces, these reflections will happen at each one. Keeping track of all the reflections becomes very tedious [4]. If the distance between the boundaries becomes short compared to the spatial size of the leading edge, the reflections still happen, but they can appear to be smeared out. The only way to keep track of the reflections in such cases is by either using a scope to measure them or using a circuit simulator to simulate them.

The most common signal integrity problem is ringing, which is due to multiple reflections of the signal between unterminated ends. An example of ringing is shown in Figure 7. To manage the reflections at these ends, we fool the signal into seeing no change in the instantaneous impedance by adding resistors at one or both ends.

In addition, the routing topology also causes discontinuities. Whenever the signal line branches, there is a change in the instantaneous impedance, as the signal sees two paths in parallel. Reflections from branches fundamentally limit the highest data rate possible in any branched topology to about 2 Gb/s. To achieve higher data rates, it is essential to route in a point to point topology. Many routing topology decisions are based on controlling reflections.

### #6 Inductance Is Fundamentally About How Efficient a Conductor Is in Generating Rings of Magnetic Field Lines

No other topic in signal integrity is as confusing and as important as inductance. Its series impedance will increase with higher frequency. This means that, as rise times decrease and clock frequencies increase, the series impedance from an inductance will increase and, when it exceeds the acceptable impedance for the environment it is in, will dominate performance. This is why the physical design decisions we make that influence inductance are so important [5].

While inductance is well defined mathematically in terms of the integrated flux density around a conductor per amp of current, translating this into practical engineering terms is confusing in many of the textbooks from which we are introduced to inductance.

The physical origin of inductance is based on the presence of rings of magnetic field lines that circulate around any current. Figure 8 shows the rings of magnetic field lines that surround a current loop. Double the current and the number of rings of field lines double.

Performing an integral of the flux density around a conductor is really counting the number of rings of magnetic field lines that circulate around the conductor. We count individual rings of field lines in units of Webers.

Inductance is fundamentally the number of rings of field lines around the current per amp of current through the conductor. The ratio of Webers of field lines per amp of current we give the special name Henrys and is a measure of inductance. This is described by

$$L = \frac{\Psi}{I} = \frac{1}{I}\iint_{\text{area}} \vec{B} \cdot d\vec{a}\,,$$

where $\Psi$ is the flux of field lines around the conductor loop, $I$ is the current in the loop, $B$ is the magnetic field density in the area within the current loop, and $da$ is the small surface area through which the flux lines pass.

When not designing controlled impedance transmission lines, it is usually important to design an

**Inductance plays a critical role in the generation of ground bounce and reducing the impedance of the power distribution network.**

interconnect that minimize the inductance. This means spoiling the efficiency of generating field lines.

While each specific geometry has a slightly different approximation for how the physical features translate into a loop inductance [5], for the special case of wide, flat conductors, the loop inductance is approximately

$$L = \frac{\mu_0}{4\pi} h \times \frac{len}{w} = 32\,{}^{\text{pH}}\!/_{\text{mils}}\, h \times \frac{len}{w},$$

where $L$ is the loop inductance of two long, flat, wide conductors; $h$ is the spacing between the two wide, flat conductors; $len$ is the length of the wide conductors; and $w$ is the width of the wide conductors.

This relationship illustrates the three design terms that influence inductance and enable lower inductance interconnects: shorter lengths, wider conductors, and closer proximity of the signal to return path. This is the design guide for low inductance structures [5].

Inductance plays a critical role in two signal integrity problems in particular, the generation of ground bounce and reducing the impedance of the PDN.

### #7 Current in a Conductor Redistributes at Higher Frequency Driven by Minimizing Loop Inductance

The current distribution in a conductor always takes the path of lowest impedance. At dc, every path through the cross section of the signal and return has equal impedance, so current is uniformly distributed through the signal and return conductors.

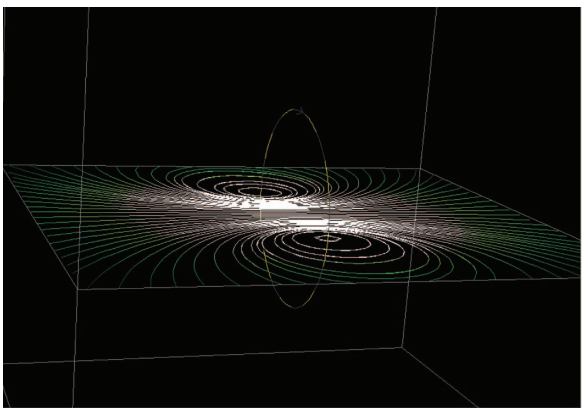


**Figure 8.** *Rings of magnetic field lines surrounding a current loop. Inductance is a measure of the total number of rings of magnetic field per amp of current. We count the rings of magnetic field surrounding the current loop by integrating the field density inside the current loop.*

**The leakage current, which turns into heat, increases proportional to the frequency of the electric field.**

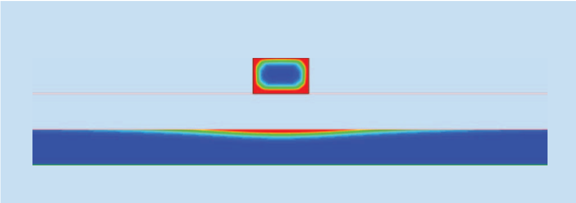


**Figure 9.** *Current distribution in the signal and return conductors at 100 MHz for copper traces is driven by two forces: current within each conductor wants to spread out as much as possible, while the signal and return current want to get as close together as possible.*

The impedance of any path depends on the series resistance and the loop inductance of that path. As frequency goes up, the impedance of the loop inductance dominates the current distribution and current will distribute to minimize the loop inductance. This translates into spreading out as far as it can within each conductor while at the same time moving as close together as possible between the signal and return currents.

The current flows within a thin region of the surface, labelled the "skin depth." This thickness depends on the conductor resistivity and frequency and is approximately

$$\delta = \pi\sqrt{\frac{1}{\sigma\pi\mu_0\mu f}} = 2.1\mu\sqrt{\frac{1}{f}},$$

where $\delta$ is the skin depth in $\mu$m, $\sigma$ is the conductivity of the metal, $\mu_0$ is the permeability of free space, $\mu$ is the relative permeability of the conductor, and $f$ is the frequency in GHz. For copper, at 1 GHz, it is about 2 $\mu$m. As frequency goes up, the skin depth thickness goes down proportional to the square root of frequency.

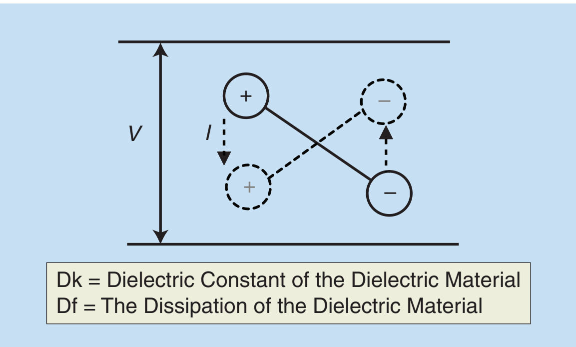



**Figure 10.** *Motion of dipoles in the external field of the signal causes them to rotate, converting electric energy into heat.*

There are two important consequences of this current redistribution or skin depth effect. The return current distribution in a wide plane will always be concentrated underneath the signal conductor in order to minimize the inductance of the signal-return current loop. For frequencies above about 100 kHz, the return current will directly follow the signal path as it meanders over the board. The resulting current distribution for the special case of a microstrip is shown in Figure 9.

At frequencies above about 10 MHz, the skin depth of copper is thinner than the geometrical thickness of 0.5 oz copper, which is about 17 $\mu$m. This means the cross sectional area through which the current flows will decrease above 10 MHz. This will increase the series resistance at higher frequencies and contribute to frequency dependent loss. This is one of the causes of rise time degradation, which results in intersymbol interference (ISI), data dependent jitter, and collapse of the eye.

## #8 Dielectric Materials Absorb Electric Field Energy and Cause Attenuation

The second loss mechanism in transmissions lines, which is also frequency dependent, is from the dielectric.

Inside all dielectrics are small dipoles that have some ability to rotate. This gives rise to the well-known dielectric constant. In addition, the rotation of the dipoles in a changing electric field will result in friction and convert some of the electric field energy into heat.

A dipole rotating to align in an external electric field is a slight, momentary current. When the field changes and it rotates back, there is another small burst of current. This orientation is illustrated in Figure 10. The fraction of the current in phase with the voltage is a leakage current between the signal and return path.

The faster the field changes, the more bursts of current happen per second and the higher the leakage current. This means that the leakage current, which turns into heat, increases proportional to the frequency of the electric field. The leakage conductivity is given by:

$$\sigma = 2\pi f \times \varepsilon_0 \times Dk \times Df,$$

where $\sigma$ is the leakage conductivity of the dielectric, $f$ is the frequency of the applied electric field, $\varepsilon_0$ is the permittivity of free space, $Dk$ is the dielectric constant of the dielectric material, and $Df$ is the dissipation of the dielectric material.

As the dipoles rotate in the external field, they constantly rub against other molecules in their polymer backbone and transfer the electric field energy into heat. A microwave oven operates on the same

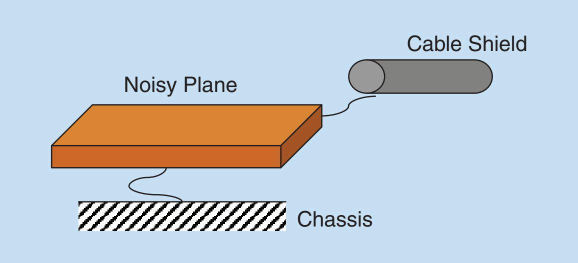



**Figure 11.** *Common currents on external cables, driven by ground bounce voltages on circuit boards or on connectors, are the chief cause of EMC failures.*

principle. The 2.45-GHz electric field inside the oven drives mainly water molecules to rotate back and forth, causing them to rub against other molecules and produce friction. This heats up the soup.

The amount of leakage current flowing between the signal and return conductors in a transmission line is proportional to the number of dipoles in the material, how large their dipole moment, and how much they can move. The dissipation factor of a material is a metric of these qualities. A high dissipation factor means the material has many large dipoles and they can move a lot, turning electric field energy into heat.

The frequency dependent dielectric loss will contribute to rise time degradation and collapse of the eye. The primary way of reducing the signal degradation from dielectric loss is to select a laminate material with the lowest dissipation factor you can afford.

## #9 Common Currents in Conductors Radiate and Often Cause EMC Failures

It only takes 10 nW of radiated power into the 120 kHz bandwidth of the typical detector to fail an EMC certification test, such as the U.S. Federal Communications Commission (FCC) part 15 class B test [6]. This is why it is often said, "there are two kinds of designers, those who are designing antennas on purpose and those who aren't designing them on purpose."

Using a simple electric dipole antenna model, we can estimate the far field radiate electric field strength as:

$$E = 4\pi \times 10^{-7} \frac{f \times I \times Len}{R},$$

where $f$ is the frequency of the current, $I$ is the current in the dipole antenna, $Len$ is the length of the radiating antenna, and $R$ is the distance from the antenna where the $E$ field is measured.

For the special case, shown in Figure 11, of a 1-m long cable, at 100 MHz, the radiated field strength for a 3-uA current, 3 m distant is 120 uV/m. This is just enough to fail a class B test. This is a very tiny amount of common current compared to the signal current, which has to flow in the cable. After all, in a 1-V signal in a 50-Ω cable, there is 20 mA of signal-return current flowing. If only 0.01% of this is unbalanced and is common to both the signal and return conductors, the product will fail compliance and it will not be allowed for sale in any country with an EMC certification requirement.

**Common currents on external cables are the chief reason products fail EMC certification tests.**

Common currents on external cables, while not the only reason, is the chief reason products fail EMC certification tests. They have three common origins: driven by ground bounce on ground planes, driven by ground bounce in poor connectors, and created from mode conversion of differential to common signals.

Many of the EMC problems with high-performance products can be dramatically reduced by reducing these three problems: ground bounce in the planes, asymmetric connectors, and mode conversion.

## Conclusions

High-speed digital design is driven by the analog effects of interconnects. These are fundamentally driven by the overlap of Maxwell's equations and board design. While simulations can help optimize a design, it is still fundamentally an engineer's design intuition that establishes the starting place. The earlier in the design process problems can be identified and designed out, the shorter the development time, the lower the development cost, and the greater the chance of an acceptable design on the first pass. The stronger your signal integrity design intuition, the more effective you will be in this era of high-speed products.

## References

[1] E. Bogatin, *Signal and Power Integrity—Simplified*. Englewood Cliffs, NJ: Prentice-Hall, 2009.

[2] S. H. Hall and H. L. Heck, *Advanced Signal Integrity for High-Speed Digital Designs*. Hoboken, NJ: Wiley, 2009.

[3] P. G. Hurray, *The Foundations of Signal Integrity*. Hoboken, NJ: Wiley, 2010.

[4] W. M. Blood, *Motorola MECL System Design Handbook*, 4th ed., Motorola Semiconductor Products, Austin, TX, 1988.

[5] C. R. Paul, *Inductance Calculations*. Hoboken, NJ: Wiley, 2010.

[6] H. Ott, *Elecptromagnetic Compatibility Engineering*. Hoboken, NJ: Wiley, 2010.