SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Trainer: Kelvin Teh
(By Oriontrain Sdn. Bhd.)
Training Provider: PSDC

Engagement Model

Factories

All industries

Training

Customised Trainings According to Demands and Requirements

Training Center

Technical Consultant Panel

ORIONTRAIN

- Design training outline
- Design training content
- Conduct training session

Training Solutions

Industrial Solutions

Customised Solutions According to Requirements

Technical Solution Consultant & Provider

ORIONPLEX

- Solution Consultant
- Solution Designer
- Total Solution Provider

ORIONPLEX ORIONTRAIN

SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Module 0: Introduction and Overview

Module 0

MODULE 0: INTRODUCTION AND OVERVIEW

- Course Objectives
- Training Agenda

Course Objectives

By successfully completing this course, participants will learn:

- Principles of Signal Integrity
- Principles of Power Integrity
- Practical Design Techniques

Training Agenda (1/10)

Module 1: Introduction to Signal Integrity & Power Integrity (SIPI)

1) What is Signal Integrity?
2) What is Power Integrity?
3) Classification of Signal Integrity Problems
4) Sources of SIPI Issues
5) Signal Integrity and EMC/EMI
6) Analogue and Digital Signals
7) Sampling and Aliasing
8) Analogue to Digital Conversion (ADC) and Quantisation
9) Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Training Agenda (2/10)

Module 2: SIPI in Analogue Circuits (Properties of Electrical Systems)

1) Impedance and Electrical Models
2) Physical Basis of Resistance, Capacitance and Inductance
3) Lumped versus Distributed Systems
4) Introduction to Transmission Lines
5) Bandwidth and Rise Time
6) Knee Frequency

Training Agenda (3/10)

Module 3: Time and Frequency Domains

1) The Time Domain
2) Sine Waves in the Frequency Domain
3) Fourier Transform
4) Spectral Content of Digital Signals
5) Filtering

Module 4: Transmission Lines

1) Ideal Transmission Lines
2) Propagation Delay
3) Characteristic Impedance
4) Popular Transmission Line Configurations
5) Lossy Lines and Rise-Time Degradation

Training Agenda (4/10)

Module 5: Reflections and Termination Techniques

1) Reflections and Its Consequences
2) Effects of Source and Load Impedance
3) Reflections at Impedance Changes
4) Controlling Reflections
5) Termination Techniques
   - End Source Termination
   - Termination Strategies
   - Rise-Time Implications
   - Power Dissipation

Training Agenda (5/10)

Module 6: Cross Talk

1) Coupling Mechanisms
2) Common Path Noise
3) Inductive and Capacitive Coupling
4) Reference Plan Splits
5) Near-End and Far-End Cross-Talk
6) Guard Traces
7) Connector Cross-Talk

Training Agenda (6/10)

Module 7: Integrated Circuits

1) Package Types
2) Lead Inductance
3) Ground Bounce
4) Synchronous Switching Noise (SSN)
5) Input Buffers and Output Drivers
6) On-Die Termination (ODT)
7) IBIS Simulations

Training Agenda (7/10)

Module 8: Differential Signalling

1) Differential Pairs and Their Advantages
2) Return Current Distribution
3) Differential Impedance
4) Common Impedance
5) Termination Techniques

Module 9: Power Distribution Networks (PDNs)

1) Power Distribution
2) Voltage Reference Distribution
3) Frequencies of Interest
4) Target PDN Impedance
5) Bypass Capacitors

Training Agenda (8/10)

Module 10: Grounding Strategies

1) What is “Ground”?
2) Ground Plane Topologies
3) Single versus Split Ground Planes
4) Ground Loops

Training Agenda (9/10)

Module 11: Power Integrity Analysis

1) What is Power Integrity Analysis
2) IR Drop Analysis
3) Noise Analysis
4) Decoupling Analysis
5) Power-Ground Plane Resonance
6) Voltage Fluctuation and Target Impedance
7) Transient Current and Target Impedance

Training Agenda (10/10)

Module 12: Common SIPI Issues and Prevention in Electronic Circuits

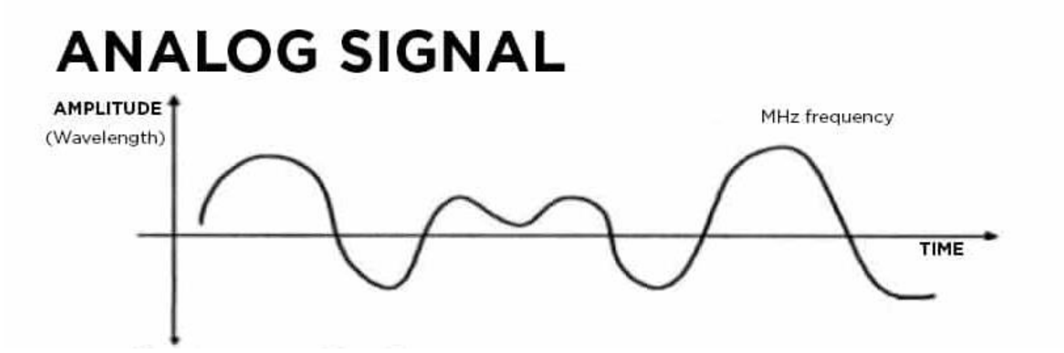
1) SIPI of Electronic Circuits
2) Rise/Fall Times of Digital Signals
3) Noise in Sensor Signals and Algorithms for Removing Them
4) Image Filtering and Denoising
5) Frequency of Events/Interrupts (Risk of Missing Events)
6) Error Detection/Correction Codes

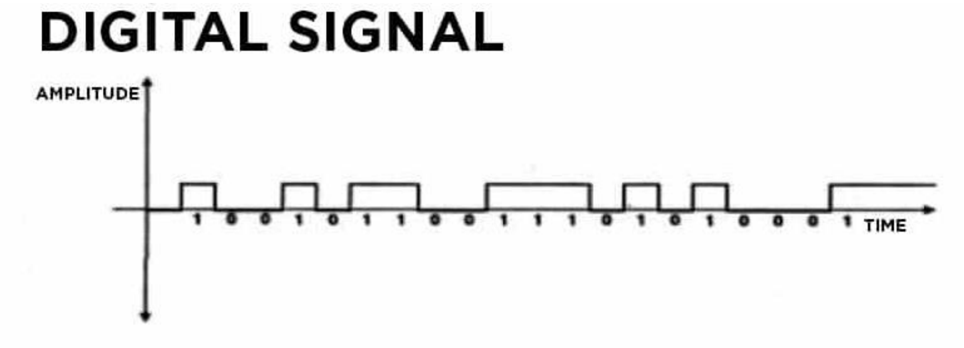
SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Module 1: Introduction to Signal Integrity & Power Integrity (SIPI)

What is Signal Integrity?

What is Signal Integrity?



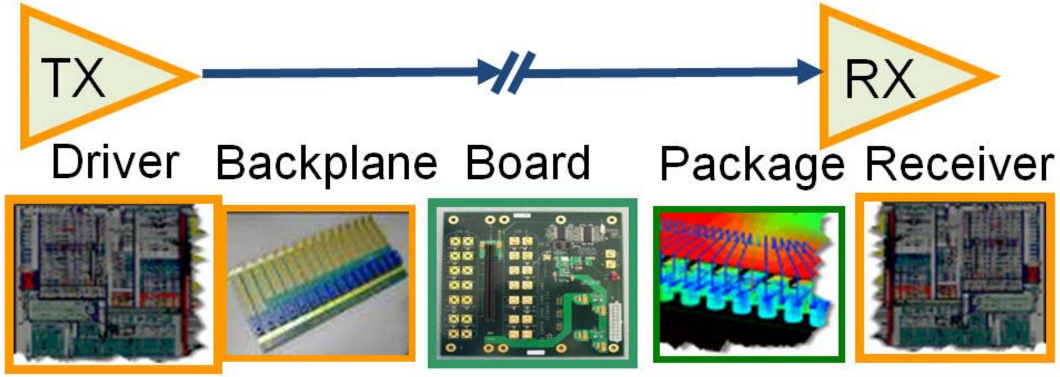


- In a system, signals travel through various kinds of interconnections (e.g., from chip to package, package to board trace, and trace to high-speed connectors).

- Signal integrity problems arise from the physical nature of interconnecting wires. Unlike a connection line drawn on a schematic, a real wire has resistance, capacitance to ground and other wires, and inductance. At higher frequencies, capacitance and inductance can cause the wire to act as a transmission line, and antenna effects can result in crosstalk and EMI.

- Signal integrity problems cause systems to fail or work only intermittently, producing "bad" data. As such, signal integrity issues are particularly important to find early in a design cycle because intermittent failures are very difficult to debug on prototypes.

What is Signal Integrity?



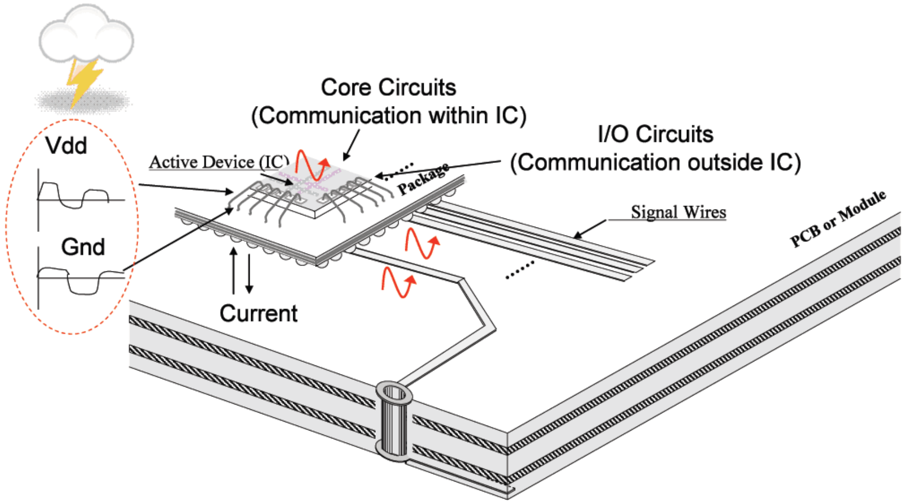
What could possibly go wrong?

If care is not taken to ensure a high level of signal integrity when designing the PCB layout, then manufacturing problems can occur in that:

1. It will cause the design to work incorrectly in some cases, but not all cases.
2. The design might actually fail completely.
3. The design might operate slower than expected (and required).

What is Power Integrity?

What is Power Integrity?

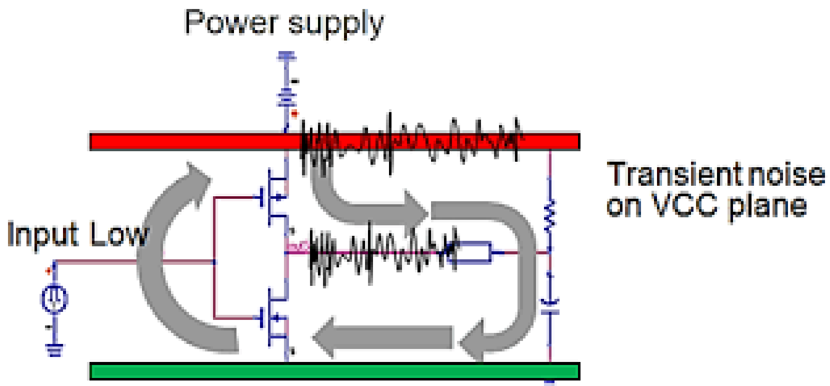


- Circuit boards today will still have power and ground planes in them, but managing the power delivery network is much more complex than simply connecting an ICs power and ground pins to them.

- Each IC will need power supplied to it that is strictly controlled to comply with its power requirements. The goal of a good power delivery network in a circuit board is to provide stable voltage references, and to distribute power to all of the devices with acceptable noise and tolerance levels.

- This careful management of power to all of the devices on the board to keep a consistent voltage level is what is known as **Power Integrity**.

What is Power Integrity?

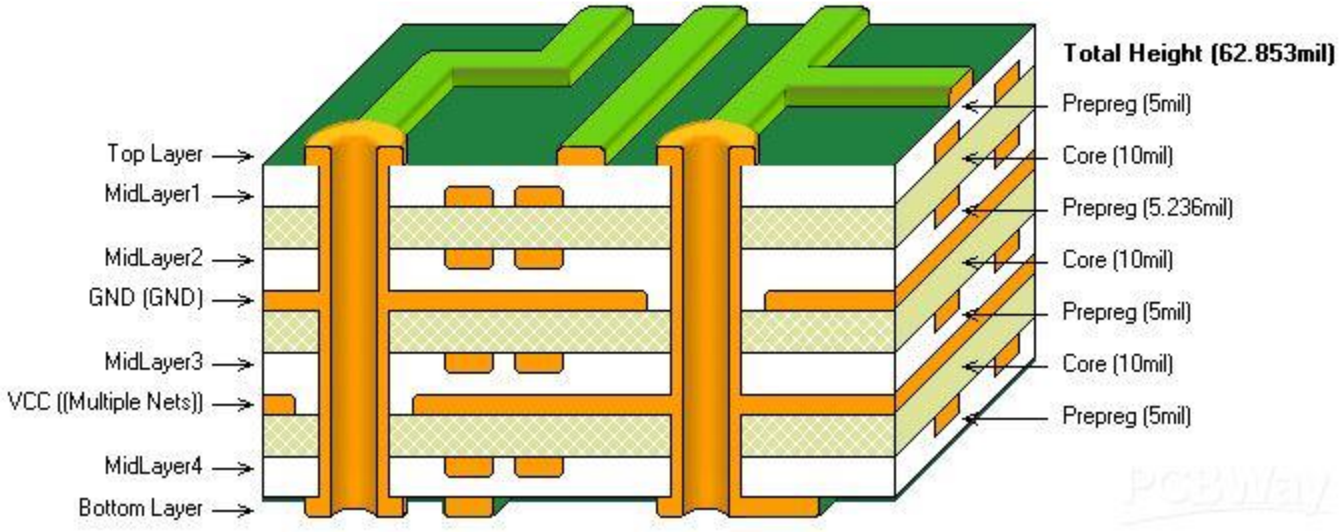


- Power integrity can be defined as the assurance that an electrical system and all of its elements have the needed power such that operation can occur adequately or as intended.

- For design and analysis, this requires not only supplying adequate power to active components but also maintaining power levels and minimizing losses of all signals that your board processes.

- To achieve power integrity, all components or modules must be supplied with power at the level required for operation, which is not simply making sure that voltages are at, or above, an acceptable level. The major threat to power integrity for your Power Distribution Network involves line variations or transients from your power supply that may cause fluctuations in the quality of your power profile.

What is Power Integrity?



- Today, PCBs are typically small and densely packed with components. This closeness of components, especially between signal traces and power paths can be a source of electromagnetic interference (EMI) or noise for your signals and impact power signal stability.
- In most cases, your boards are comprised of one or more high power components that require power dissipation to prevent adversely affecting other board elements. Optimal management of your power integrity requires that these higher frequency RF signals are isolated from your PDN.
- To a great extent, the power integrity of your board depends upon your application of design techniques and choices for its manufacture to mitigate these potential issues and promote power and signal integrity.

Classification of Signal Integrity Problems

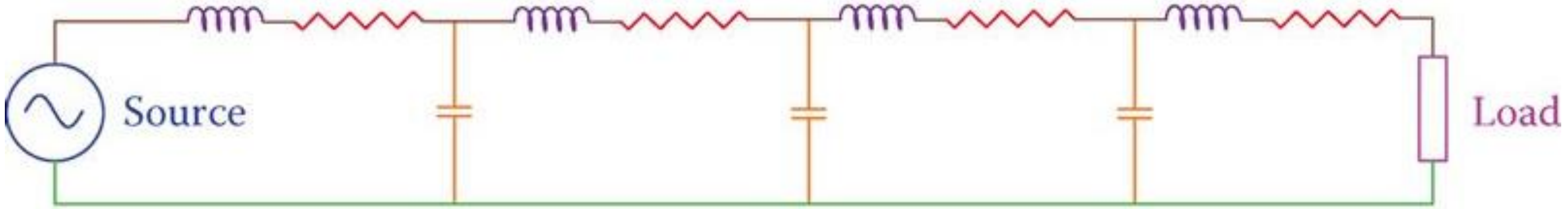
ORIONPLEX | ORIONTRAIN

Classification of Signal Integrity Problems

There are six main areas of circuit design and layout that must be taken into consideration to ensure that the signal integrity of a board or circuit design are maintained:

1. Transmission line effects
2. Impedance matching
3. Simultaneous switching effects
4. Crosstalk
5. Electromagnetic Interference (EMI)
6. Ground Bounce

Classification of Signal Integrity Problems



Transmission line effects

- At low frequencies a length of track may be considered purely by its DC characteristics. However, as frequencies rise, effects including the capacitance and inductance associated with the track start to have a significant impact on the performance of the line. Accordingly it is necessary to consider the tracks as transmission lines, and treat them accordingly, looking at aspects such as the line impedance.

- In order to ensure that the transmission lines are treated correctly. First it is necessary for the lines to have a ground plane underneath them. It is also necessary to calculate the impedance of the line. This is determined from a combination of the line thickness, the distance between the line and the ground plane, and the dielectric constant of the board.

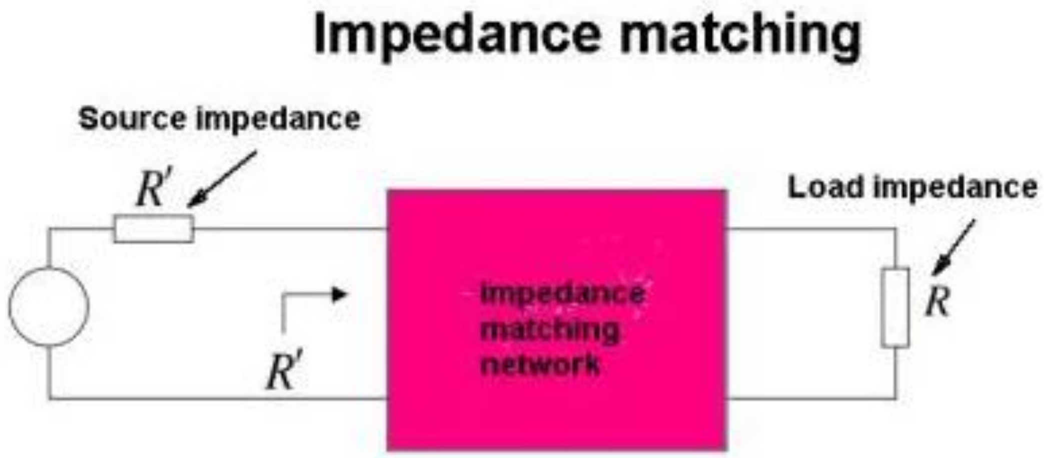
- If as often may happen, the line needs to traverse between layers and therefore the distance between the line and the ground plane changes. It will be necessary to ensure the line impedance remains the same, possibly by changing the line thickness.

Classification of Signal Integrity Problems

Impedance matching

- It is necessary to consider the way in which the impedances need to be matched to ensure good signal integrity. When there is a mismatch between the line and the load, not all the energy of the waveform is absorbed by the load. That which is not absorbed is reflected back along the line where it may again not be absorbed if there is a mismatch between the transmitter and the line.

- This can cause overshoot and ringing which leads to poor signal integrity and giving rise to signal errors. To overcome this problem it is necessary to match the transmission line to the line drivers or transmitters and the line receivers. Many drivers and receivers exits that have suitable input and output impedances.



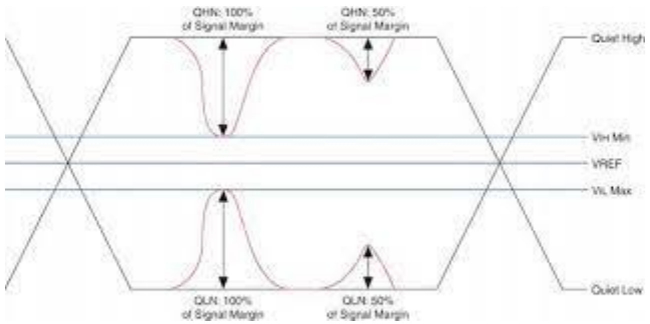
Classification of Signal Integrity Problems

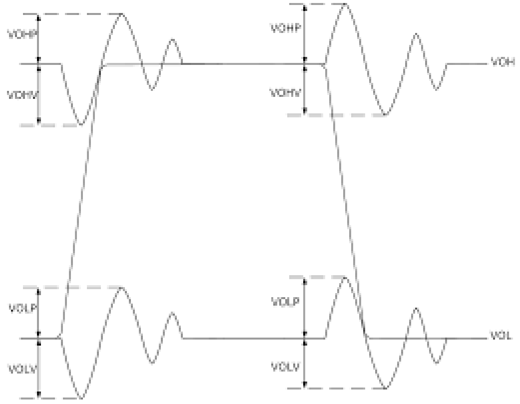
Simultaneous switching effects

- One effect that can disrupt the signal integrity on a circuit board occurs when several output lines are switched simultaneously. As stored charge on the outputs needs to be discharged, this gives rise to high levels of transient currents.

- While the levels of transients are normally adequate for single outputs changing, if several lines are switched simultaneously, especially on the same chip, the transient currents are larger, and this can give rise to problems. Problems with signal integrity arise because a voltage arises between the device ground and the board ground.

- To overcome this problem there are a number of measures that can be incorporated. One is to ensure that simultaneous switching does not occur, but this is not always possible, especially when circuits are operated in a synchronous manner. Good grounding is essential: a ground plane must be used to ensure a low resistance ground return.





Classification of Signal Integrity Problems

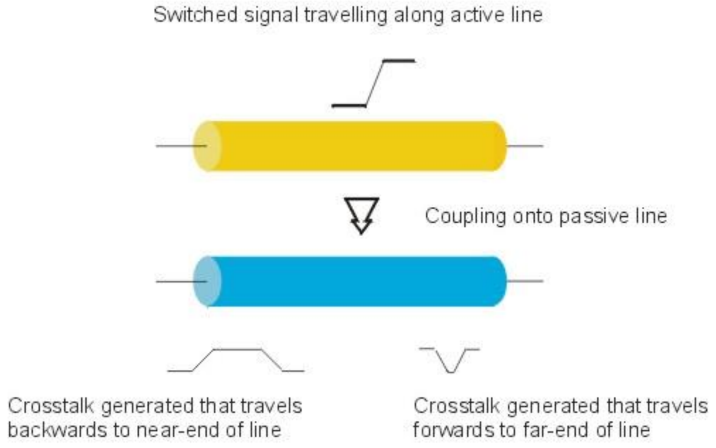
Crosstalk

- This aspect of signal integrity arises from the fact that signals appearing on one line appear on nearby lines. This can result in spurious spikes and other signal appearing on nearby lines. This can cause erroneous data or clocking pulses to appear, and these can be very difficult to track down in some circumstances. Poor signal integrity from crosstalk arises from two causes, namely **mutual inductance**, and **mutual capacitance**.

- The mutual inductance is the effect that is used in transformers. It arises from the fact that a current in one track sets up a magnetic field. Changes in this field then induce a current in a nearby track.

- Mutual capacitive occurs as a result of the coupling of the electric fields between two tracks. A voltage appearing on one track creates an electric field which can couple to a second line.

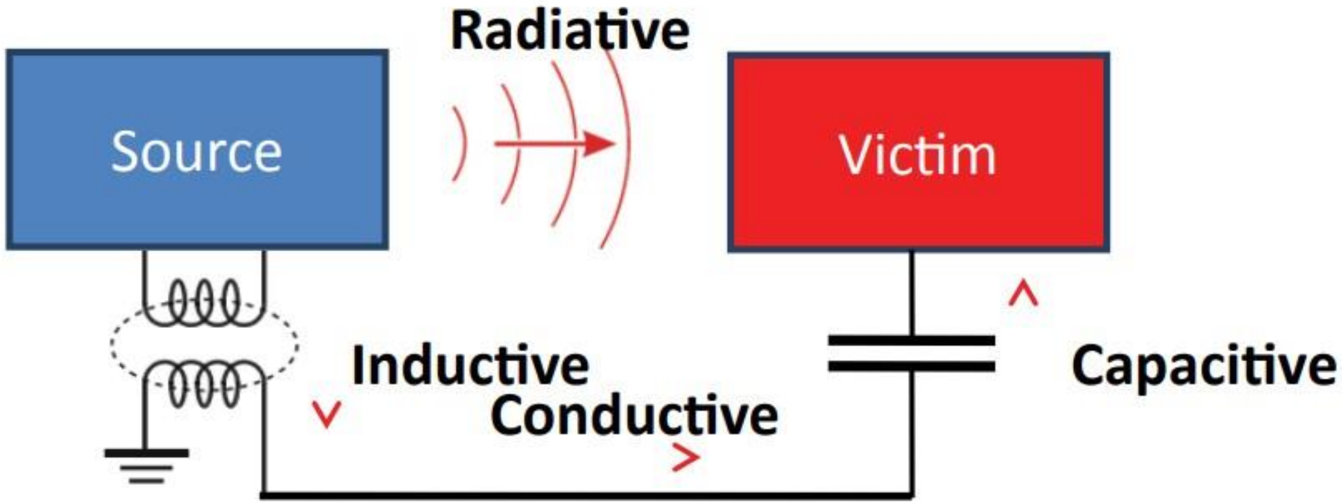
- There are several techniques that can be used to overcome these effects. As poor signal integrity from crosstalk arises from mutual inductance and capacitance, the solutions involve taking steps to reduce them.



Classification of Signal Integrity Problems

Electromagnetic Interference (EMI)

- Traces with higher frequency signals can behave like antennas and radiate EMI if they aren’t routed carefully.
- These traces include signal return paths on the reference plane, which must be kept clear of splits
- Additionally, sensitive high-speed signals such as clock lines should be kept separated from other traces and routed cleanly on a single layer as much as possible without 90° corners. Another good habit is to avoid traces or via configurations that are stubs which can create signal reflections and act as an antenna.

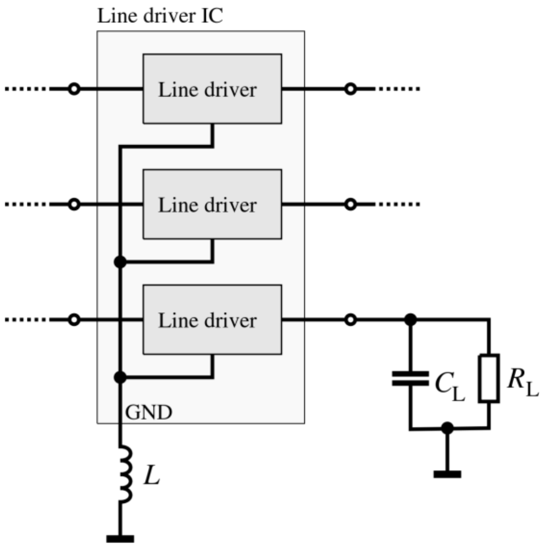


Classification of Signal Integrity Problems

Ground Bounce

- With the high-speed switching of the components on the board, the voltages may not return to the ground level and instead “bounce” above it. This bouncing leads to signal pulses being miss-interpreted by receivers and producing false results.

- To avoid this, make sure to place the decoupling capacitors as close as possible to their associated power supply pin on their assigned ICs. This practice will reduce the current spikes during the switching. Also, make sure to connect each ground pin to the ground plane individually.



Sources of SIPI Issues

Sources of SIPI Issues

Power integrity is also tied in with some other concerns of the power delivery network. Some of these include:

- **Ground bounce**: Also known as simultaneous switching noise or SSN, ground bounce happens when a lot of signals switch at the same time. This can happen when a processor writes to memory, and the data signals all switch at the same time. If the signals don't return all the way to their reference ground level due to the speed of the switching, they "bounce" above it. This ground bounce noise can create false switching and potentially disrupt or shut down the device.

- **Power ripples**: The switching characteristics of a power supply can cause power ripples. These ripples may create crosstalk in adjacent circuitry compromising the accuracy of the signals. Once again, this can cause disruption of the circuits.

Sources of SIPI Issues

- **Electromagnetic interference (EMI)**: How the power planes in a board layer stackup are arranged can contribute to an EMI problem if not configured correctly. The power and ground planes can help prevent EMI from affecting the performance of signal layers by shielding them, but only if the signal layers are between the planes.

- **Return paths**: The planes of the board need to be designed in order to preserve the best signal return paths. Unfortunately the power requirements often leave a plethora of holes for connecting vias and stitching vias, and the multiple power supplies can dictate the need for split planes. All of these can put obstacles in the way of a good return path for the signals which will lead to poor signal integrity. All of these concerns must be considered along with the analysis of the board for good power integrity. Many of the problems outlined above can be taken care of with strategically placed capacitors and resistors as well as carefully laid out power planes.

Signal Integrity and EMC/EMI

Signal Integrity and EMC/EMI

- Signal integrity becomes more important in electronic design as circuit speeds increase. Faster data rates and shorter rise/fall times make it more challenging to transmit a signal from point A to point B.

- Signal distortion and degradation simultaneously have adverse effects on electromagnetic compatibility. Circuit radiation and circuit susceptibility both increase as signal integrity decreases.

How Does SI Cause EMI?

Using EMI Analyst software, the connection between SI and EMI is easy to see. The example provides a concrete illustration of how changing just one signal property, skew, affects cable radiation, a critical EMC characteristic for many systems.

Signal Integrity and EMC/EMI

SI affects EMC

- SI is concerned with the analog characteristics of digital circuits.
- EMC is concerned with making sure that circuit operation does not produce excessive interference and that circuits are not susceptible to interference.
- SI is essential a time domain subject. EMC, except for transient events, is primarily a frequency domain discipline.

EMC Affects SI

- A design engineer could be proficient in SI but know little about EMC. However, his design decisions might have a significant effect on EMC.
- Likewise, an EMC engineer could know little about SI and still be proficient in most aspects of EMC. Both engineers would benefit from having some knowledge about the other's field of expertise.

Signal Integrity and EMC/EMI

SI and EMC Common Ground

Signal integrity and EMC overlap. Poor signal integrity produces greater radiated emissions from circuit traces and signal wires.

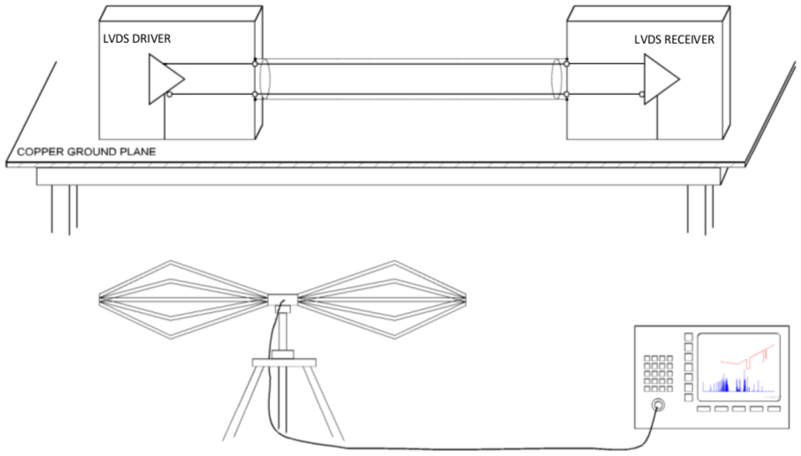
Simultaneously, a circuit with poor signal integrity has less margin than a circuit with good signal integrity. Add electromagnetic interference to the system, and the circuit with poor SI is more susceptible.

Signal Integrity and EMC/EMI

Effect of Signal Skew on Radiated Emissions

The example below illustrates one way that SI affects EMC. Cable radiation produced by a balanced differential LDVS signal is calculated for two SI conditions. First, the skew of the LVDS signal is zero; then it is changed to 200 picoseconds. The change in radiation levels at some frequencies is dramatic.

The example uses a standard radiated emissions test setup. An antenna placed a fixed distance from the equipment containing the circuitry measures the emitted field strength over a specified frequency range, in this case, 10 MHz to 2 GHz.



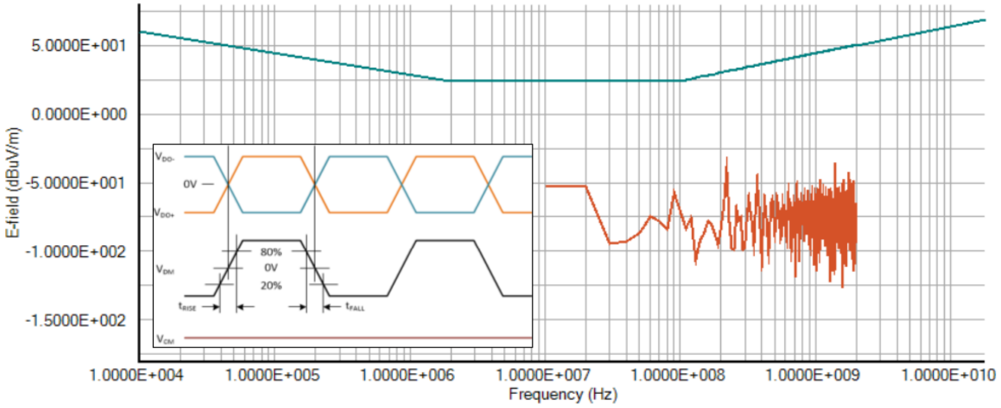
Radiated Emissions Test Setup

Signal Integrity and EMC/EMI

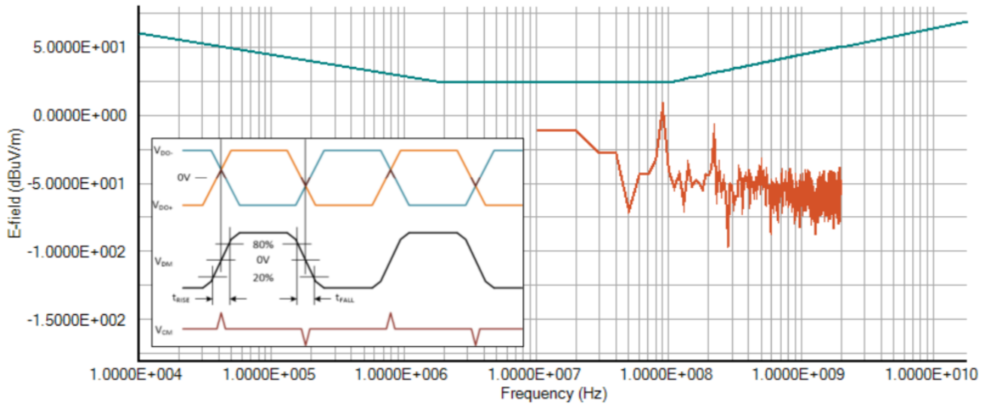
Predicted Radiation

The graphs below show the effect of signal skew on radiated emissions for this circuit. Predicted cable radiation is indicated by the reddish-brown line. The radiation limit is the blueish-green line.First, radiated emissions are computed for the case where the circuit has no signal skew. The positive and negative driver outputs transition at the same time and the rise and fall times are perfectly matched. Then, radiated emissions are calculated when the circuit has just 200 picoseconds of signal skew. The skew introduces about 70 mV common mode voltage that pulses very briefly each time the circuit changes state.

Although both analyses show that radiated emissions are below the limit, notice that signal skew causes the cable radiation to jump by 55 dB at 90 MHz. When the additive effects of common mode noise from other circuits in the system are present, radiated emissions could easily exceed the limit, resulting in an EMI test failure.



Radiated Emissions with No Signal Skew



Radiated Emissions with 200 picosecond Signal Skew

Signal Integrity and EMC/EMI

Take Away

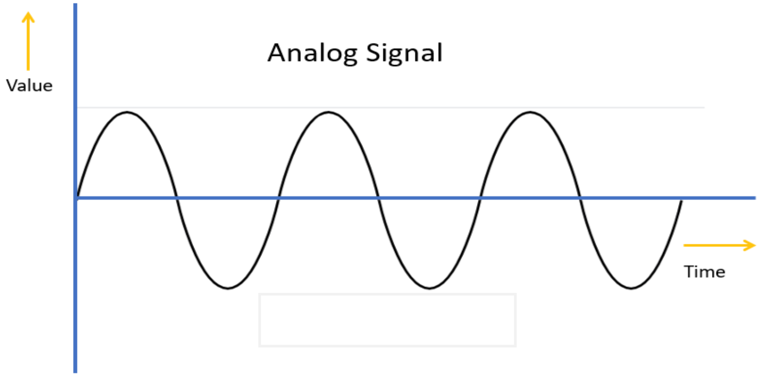
- Signal skew can have a dramatic effect on cable radiation because it induces common mode voltage on the signal lines.

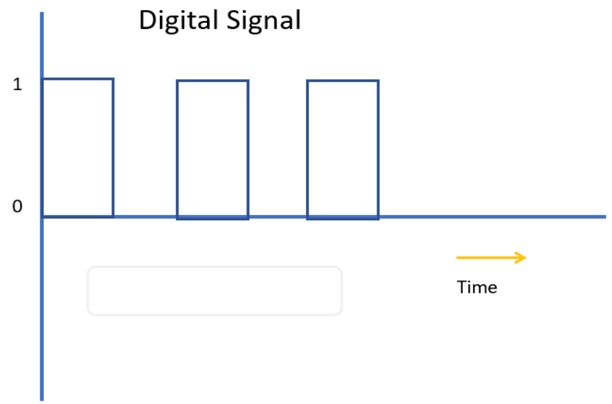
- Signal skew is just one signal integrity property. Other SI properties, such as impedance mismatch, crosstalk, return loss, and propagation delay also affects EMC.

- While SI is not the same thing as EMC, SI affects EMC. The quality of the digital signals carried on system electrical cabling has a direct effect on electromagnetic interference emissions and susceptibility.

Analogue and Digital Signals

Analogue and Digital Signals





- A signal is an electromagnetic or electrical current that is used for carrying data from one system or network to another. The signal is a function that conveys information about a phenomenon.

- In electronics and telecommunications, it refers to any time-varying voltage that is an electromagnetic wave which carries information. A signal can also be defined as an observable change in quality such as quantity. There are two main types of signals: Analog signal and Digital signal.

- A digital signal is a signal that is used to represent data as a sequence of separate values at any point in time. It can only take on one of a fixed number of values.

- This type of signal represents a real number within a constant range of values. Now, let's learn some key difference between Digital and Analog signals.

Analogue and Digital Signals

Characteristics of Analog Signal & Digital SIgnal

Here, are essential characteristics of Analog Signal

- These type of electronic signals are time-varying
- Minimum and maximum values which is either positive or negative.
- It can be either periodic or non-periodic.
- Analog Signal works on continuous data.
- The accuracy of the analog signal is not high when compared to the digital signal.
- It helps you to measure natural or physical values.
- Analog signal output form is like Curve, Line, or Graph, so it may not be meaningful to all.

Here, are essential characteristics of Digital signals

- Digital signal are continuous signals
- This type of electronic l signals can be processed and transmitted better compared to analog signal.
- Digital signals are versatile, so it is widely used.
- The accuracy of the digital signal is better than that of the analog signal.

Analogue and Digital Signals

Advantages of Analog Signals & Digital Signals

Here, are pros/benefits of Analog Signals

- Easier in processing
- Best suited for audio and video transmission.
- It has a low cost and is portable.
- It has a much higher density so that it can present more refined information.
- Not necessary to buy a new graphics board.
- Uses less bandwidth than digital sounds
- Provide more accurate representation of a sound
- It is the natural form of a sound.

Here, are pros/advantages of Digital Signals:

- Digital data can be easily compressed.
- Any information in the digital form can be encrypted.
- Equipment that uses digital signals is more common and less expensive.
- Digital signal makes running instruments free from observation errors like parallax and approximation errors.
- A lot of editing tools are available
- You can edit the sound without altering the original copy
- Easy to transmit the data over networks

Analogue and Digital Signals

Disadvantages of Analog Signals

Here are cons/drawback of Analog Signals:

- Analog tends to have a lower quality signal than digital.
- The cables are sensitive to external influences.
- The cost of the Analog wire is high and not easily portable.
- Low availability of models with digital interfaces.
- Recording analog sound on tape is quite expensive if the tape is damaged
- It offers limitations in editing
- Tape is becoming hard to find
- It is quite difficult to synchronize analog sound
- Quality is easily lost
- Data can become corrupted
- Plenty of recording devices and formats which can become confusing to store a digital signal
- Digital sounds can cut an analog sound wave which means that you can't get a perfect reproduction of a sound
- Offers poor multi-user interfaces

Here are cons/drawback of Digital Signals:

- Sampling may cause loss of information.
- A/D and D/A demands mixed-signal hardware
- Processor speed is limited
- Develop quantization and round-off errors
- It requires greater bandwidth
- Systems and processing is more complex.

Analogue and Digital Signals

Analog	Digital
• An analog signal is a continuous signal that represents physical measurements.	• Digital signals are time separated signals which are generated using digital modulation.
• It is denoted by sine waves	• It is denoted by square waves
• It uses a continuous range of values that help you to represent information.	• Digital signal uses discrete 0 and 1 to represent information.
• Temperature sensors, FM radio signals, Photocells, Light sensor, Resistive touch screen are examples of Analog signals.	• Computers, CDs, DVDs are some examples of Digital signal.
• The analog signal bandwidth is low	• The digital signal bandwidth is high.
• Analog signals are deteriorated by noise throughout transmission as well as write/read cycle.	• Relatively a noise-immune system without deterioration during the transmission process and write/read cycle.
• Analog hardware never offers flexible implementation.	• Digital hardware offers flexibility in implementation.
• It is suited for audio and video transmission.	• It is suited for Computing and digital electronics.
• Processing can be done in real-time and consumes lesser bandwidth compared to a digital signal.	• It never gives a guarantee that digital signal processing can be performed in real time.
• Analog instruments usually have s scale which is cramped at lower end and gives considerable observational errors.	• Digital instruments never cause any kind of observational errors.
• Analog signal doesn't offer any fixed range.	• Digital signal has a finite number, i.e., 0 and 1.

Sampling and Aliasing

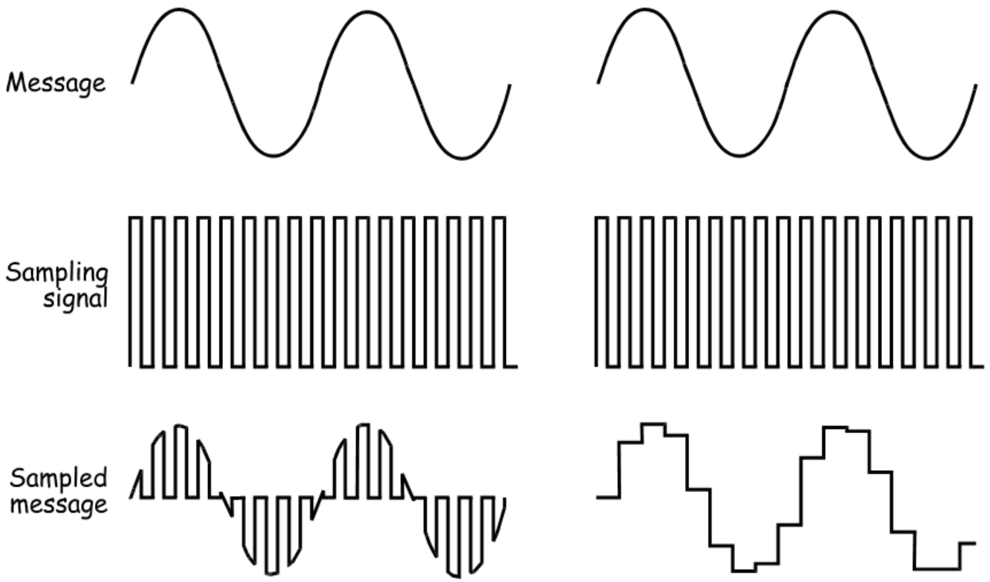
Sampling and Aliasing

Sampling Theorem

The sampling theorem can be defined as the conversion of an analog signal into a discrete form by taking the sampling frequency as twice the input analog signal frequency. Input signal frequency denoted by Fm and sampling signal frequency denoted by Fs.

Sampling frequency Fs=1/Ts

The output sample signal is represented by the samples. These samples are maintained with a gap, these gaps are termed as sample period or sampling interval (Ts). And the reciprocal of the sampling period is known as “sampling frequency” or “sampling rate”. The number of samples is represented in the sampled signal is indicated by the sampling rate.



ORIONPLEX ORIONTRAIN

Analogue and Digital Signals

Sampling Theorem Statement

Sampling theorem states that “continues form of a time-variant signal can be represented in the discrete form of a signal with help of samples and the sampled (discrete) signal can be recovered to original form when the sampling signal frequency Fs having the greater frequency value than or equal to the input signal frequency Fm.

$$F_s \geq 2F_m$$

If the sampling frequency (Fs) equals twice the input signal frequency (Fm), then such a condition is called the Nyquist Criteria for sampling. When sampling frequency equals twice the input signal frequency is known as “**Nyquist rate**”.

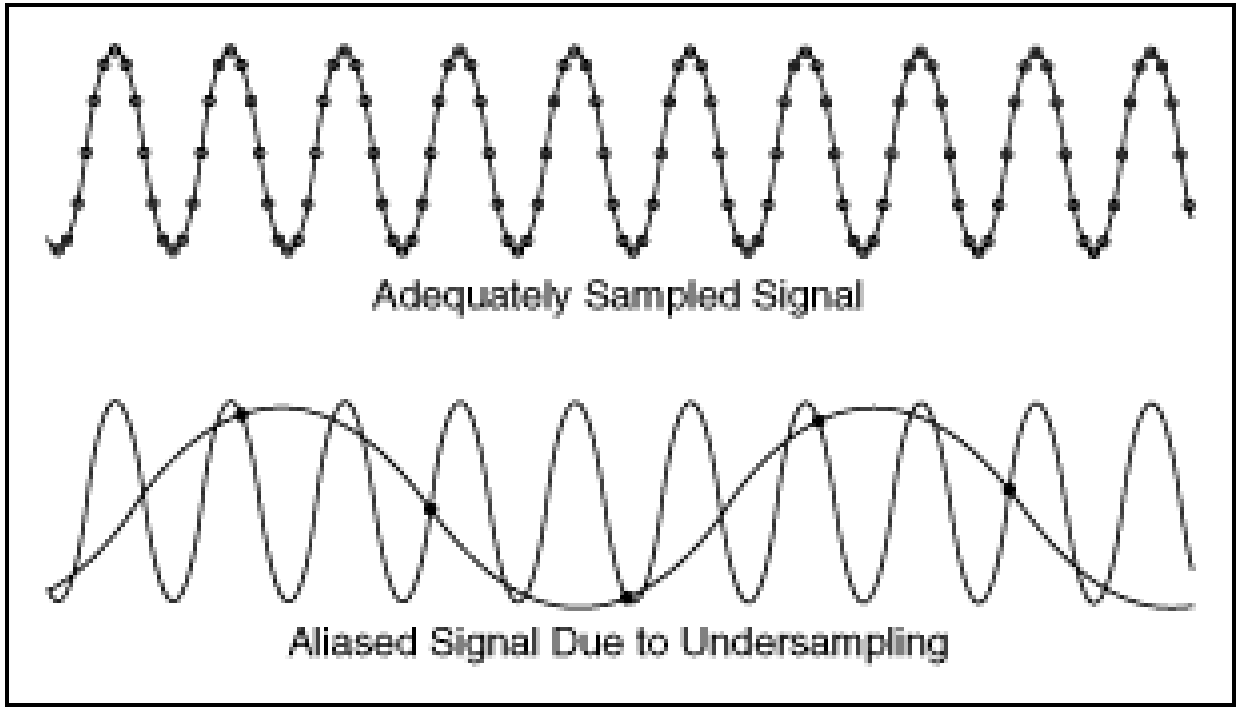
$$F_s = 2F_m$$

If the sampling frequency (Fs) is less than twice the input signal frequency, such criteria called an **Aliasing effect**.

$$F_s < 2F_m$$

Analogue and Digital Signals

Aliasing is the presence of unwanted components in the reconstructed signal. These components were not present when the original signal was sampled. In addition, some of the frequencies in the original signal may be lost in the reconstructed signal. Aliasing occurs because signal frequencies can overlap if the sampling frequency is too low. Frequencies "fold" around half the sampling frequency - which is why this frequency is often referred to as the folding frequency.



Analogue and Digital Signals

Nyquist Sampling Theorem

In the sampling process, while converting the analog signal to a discrete version, the chosen sampling signal is the most important factor. And what are the reasons to get distortions in the sampling output while conversion of analog to discrete? These types of questions can be answered by the “Nyquist sampling theorem”.

Nyquist sampling theorem states that the sampling signal frequency should be double the input signal’s highest frequency component to get distortion less output signal. As per the scientist’s name, Harry Nyquist this is named as Nyquist sampling theorem.

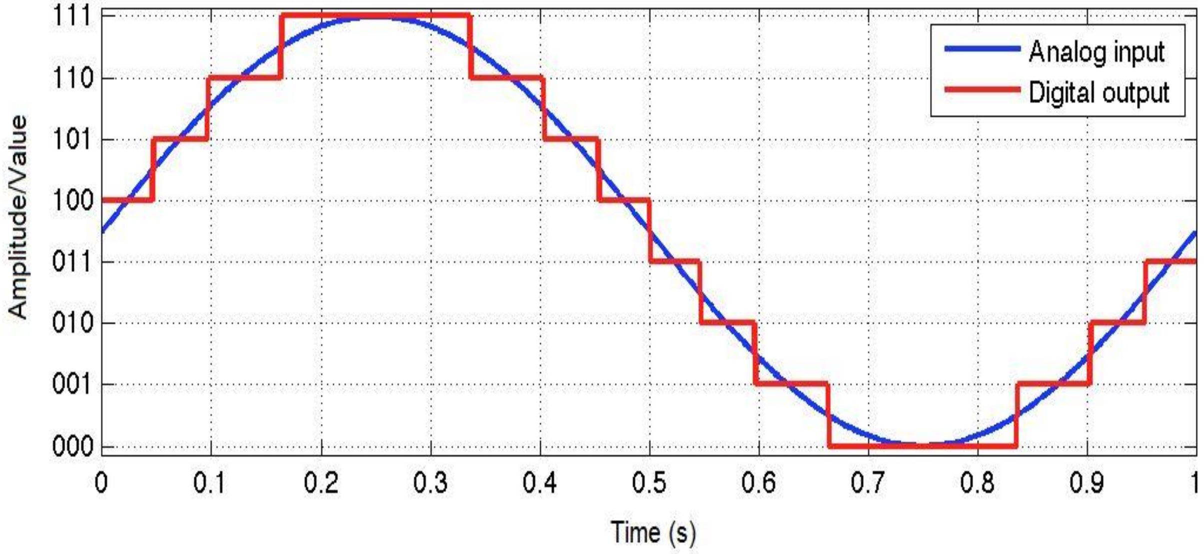
$$F_s = 2F_m$$

Analogue to Digital Conversion (ADC) and Quantisation Noise

Analogue to Digital Conversion (ADC) and Quantisation Noise

- Analog to Digital Converter (ADC) is a very useful feature that converts an analog voltage on a pin to a digital number. By converting from the analog world to the digital world, we can begin to use electronics to interface to the analog world around us.

- Analog-to-digital converters (ADCs) are an important component when it comes to dealing with digital systems communicating with real-time signals. With IoT developing quickly to be applied in everyday life, real-world/time signals have to be read by these digital systems to accurately provide vital information.



ORIONPLEX ORIONTRAIN

Analogue to Digital Conversion (ADC) and Quantisation Noise

Sampling Rate/Frequency

The ADC's sampling rate, also known as sampling frequency, can be tied to the ADC's speed. The sampling rate is measured by using "samples per second", where the units are in SPS or S/s (or if you're using sampling frequency, it would be in Hz). This simply means how many samples or data points it takes within a second. The more samples the ADC takes, the higher frequencies it can handle.

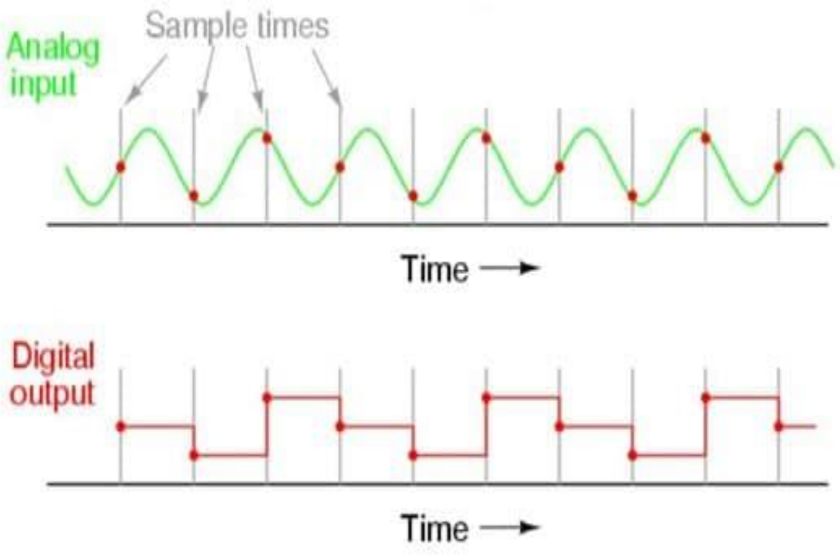
One important equation on the sample rate is:

$$f_s = 1/T$$

Where,

f_s = Sample Rate/Frequency

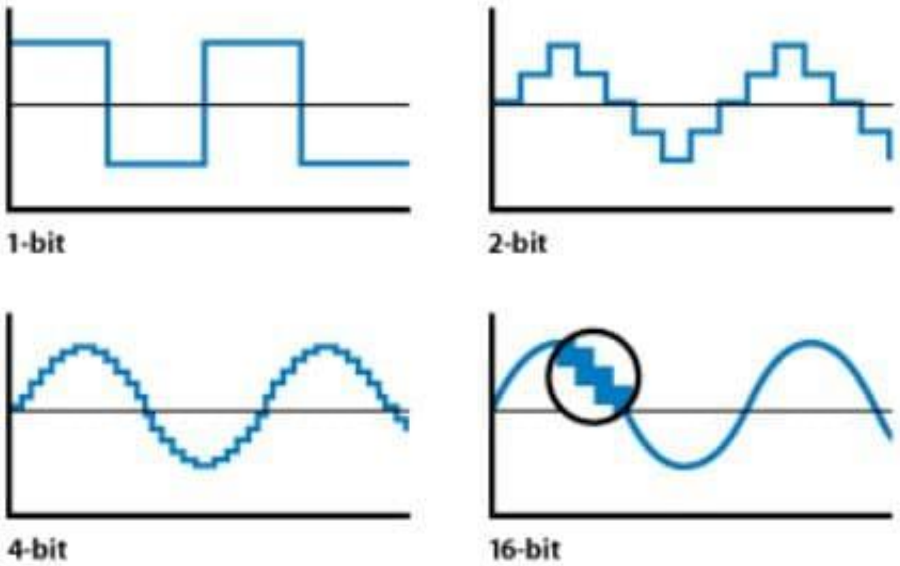
T = Period of the sample or the time it takes before sampling again



Analogue to Digital Conversion (ADC) and Quantisation Noise

- ADC's resolution can be tied to the precision of the ADC. The resolution of the ADC can be determined by its bit length. A quick example on how it helps the digital signal output a more accurate signal is shown in Figure Below. Here you can see that the 1-bit only has two "levels".

- As you increase the bit length, the levels increase making the signal more closely represent the original analog signal.

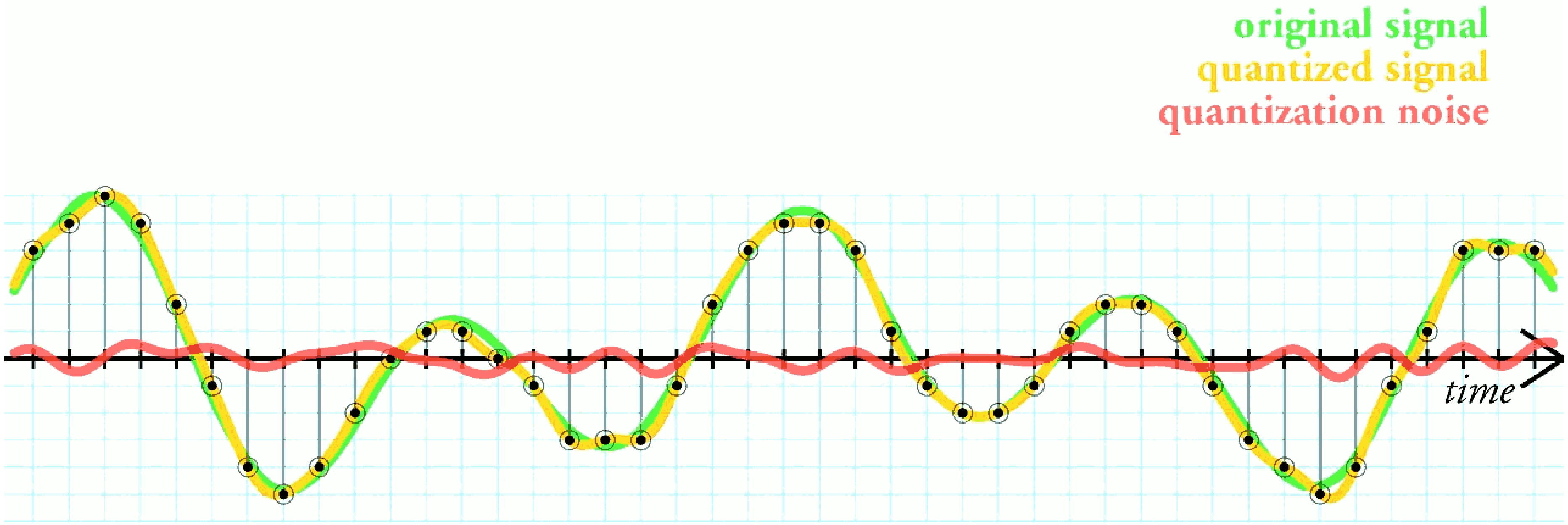


Bit Length	Levels	Step Size (5V Range)
8-bits	256	19.53 mV
10-bits	1024	4.88 mV
12-bits	4096	1.22 mV
16-bits	65536	76.29 µV
18-bits	262144	19.07 µV
20-bits	1048576	4.76 µV
24-bits	16777216	0.298 µV

Analogue to Digital Conversion (ADC) and Quantisation Noise

Quantization

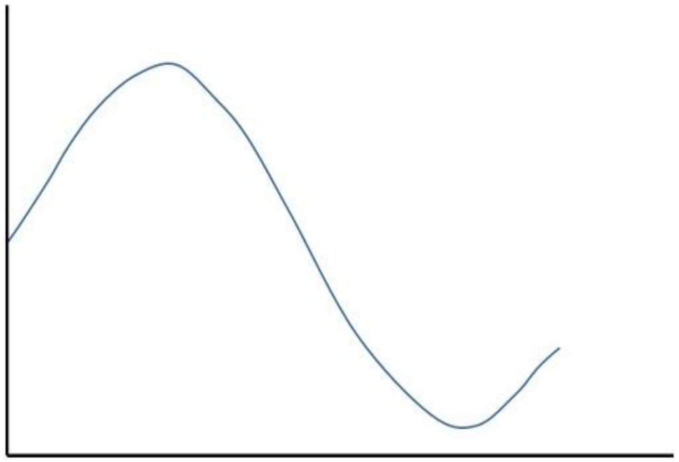
The digitization of analog signals involves the rounding off of the values which are approximately equal to the analog values. The method of sampling chooses a few points on the analog signal and then these points are joined to round off the value to a near stabilized value. Such a process is called as Quantization.

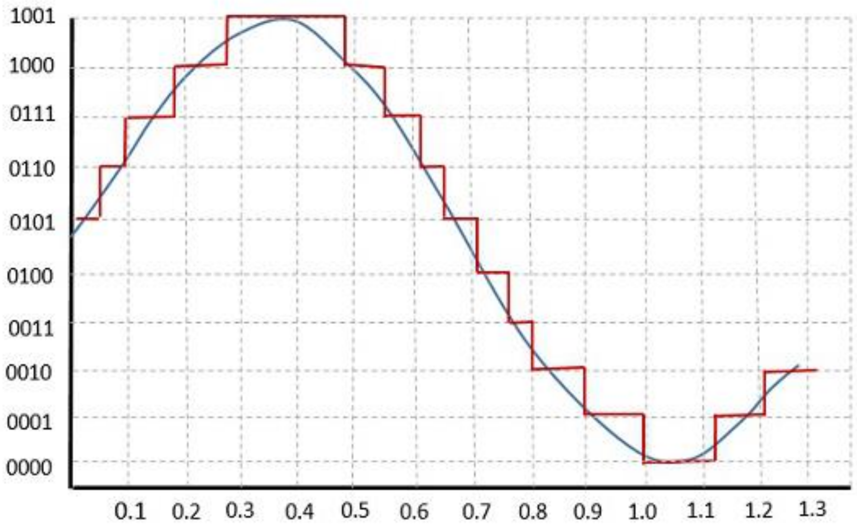


Analogue to Digital Conversion (ADC) and Quantisation Noise

Quantizing an Analog Signal

The analog-to-digital converters perform this type of function to create a series of digital values out of the given analog signal. The following figure represents an analog signal. This signal to get converted into digital, has to undergo sampling and quantizing.



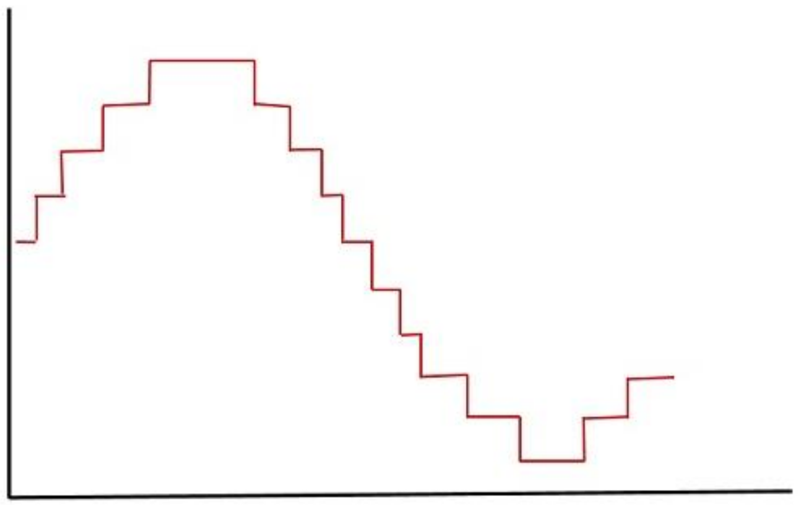


ORIONPLEX ORIONTRAIN

Analogue to Digital Conversion (ADC) and Quantisation Noise

- The quantizing of an analog signal is done by discretizing the signal with a number of quantization levels. Quantization is representing the sampled values of the amplitude by a finite set of levels, which means converting a continuous-amplitude sample into a discrete-time signal.

- Both sampling and quantization result in the loss of information. The quality of a Quantizer output depends upon the number of quantization levels used. The discrete amplitudes of the quantized output are called as representation levels or reconstruction levels. The spacing between the two adjacent representation levels is called a quantum or step-size.



Analogue to Digital Conversion (ADC) and Quantisation Noise

There are two types of Quantization - **Uniform Quantization** and **Non-uniform Quantization**.

The type of quantization in which the quantization levels are uniformly spaced is termed as a Uniform Quantization. The type of quantization in which the quantization levels are unequal and mostly the relation between them is logarithmic, is termed as a Non-uniform Quantization.

There are two types of uniform quantization. They are Mid-Rise type and Mid-Tread type. The following figures represent the two types of uniform quantization

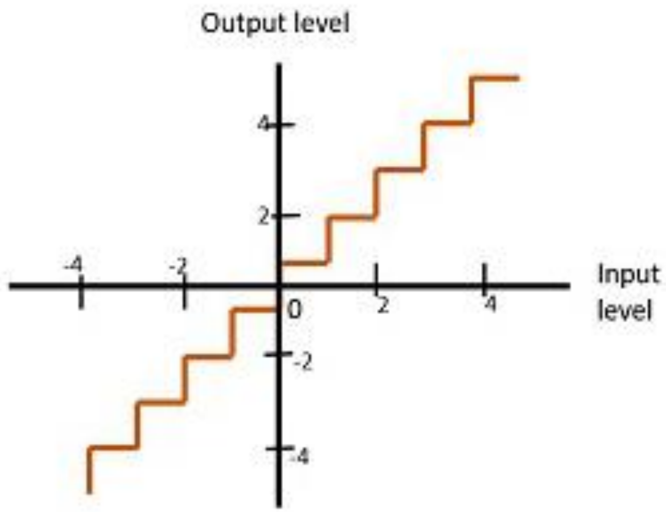


Fig 1 : Mid-Rise type Uniform Quantization

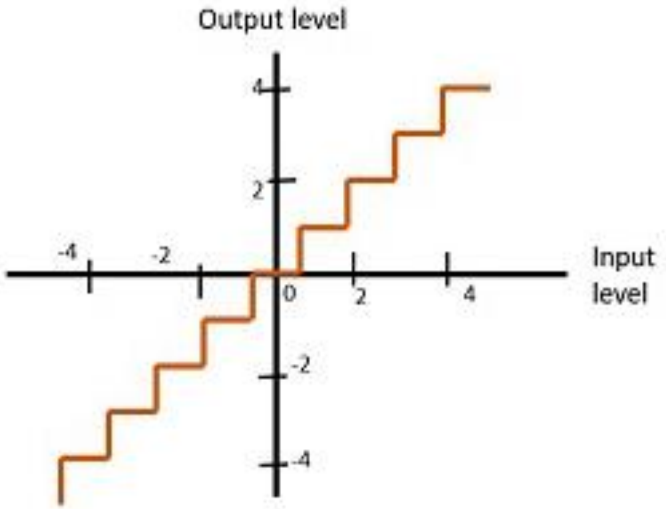


Fig 2 : Mid-Tread type Uniform Quantization

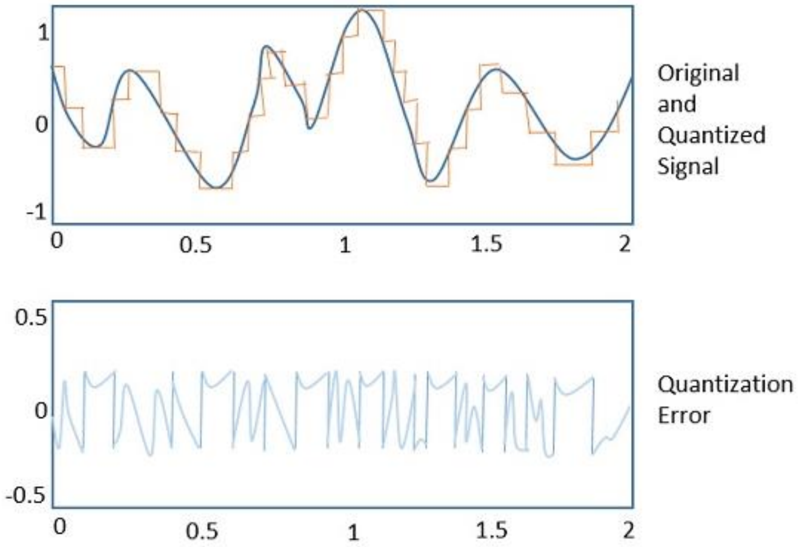
Analogue to Digital Conversion (ADC) and Quantisation Noise

Quantization Error

For any system, during its functioning, there is always a difference in the values of its input and output. The processing of the system results in an error, which is the difference of those values.

The difference between an input value and its quantized value is called a Quantization Error. A Quantizer is a logarithmic function that performs Quantization rounding off the value. An analog-to-digital converter (ADC) works as a quantizer.

The following figure illustrates an example for a quantization error, indicating the difference between the original signal and the quantized signal.



ORIONPLEX ORIONTRAIN

Analogue to Digital Conversion (ADC) and Quantisation Noise

Quantization Noise

It is a type of quantization error, which usually occurs in analog audio signal, while quantizing it to digital. For example, in music, the signals keep changing continuously, where a regularity is not found in errors. Such errors create a wideband noise called as Quantization Noise.

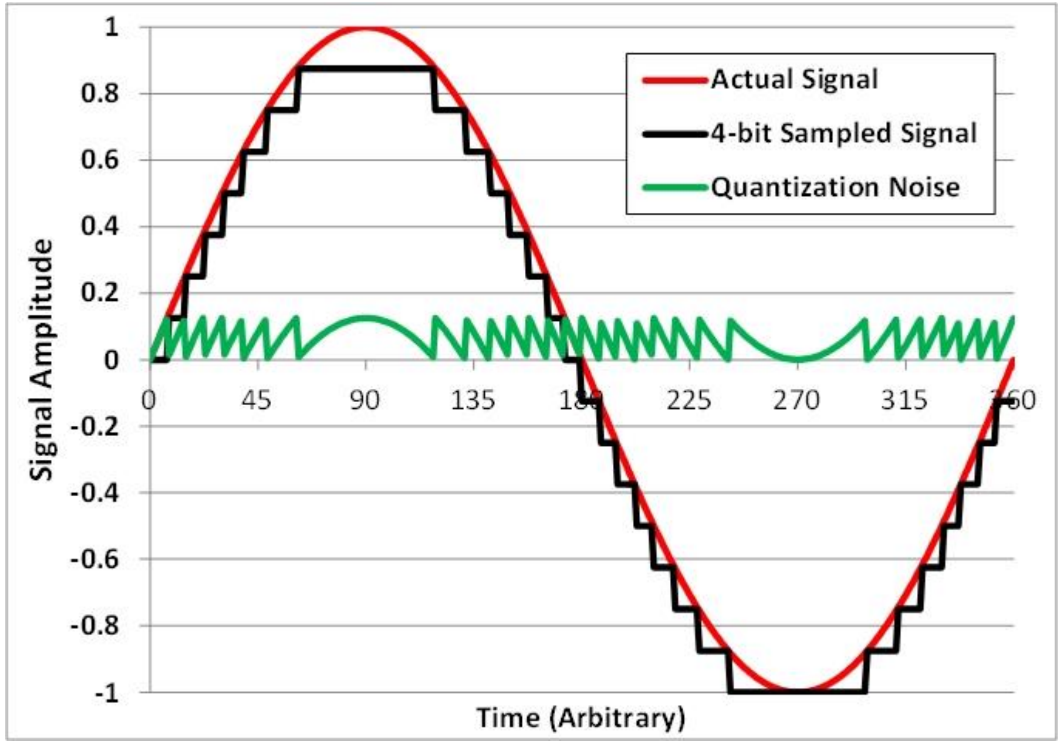


Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Image Noise

- In digital photography, image noise can be compared to film grain for analogue cameras. Image noise usually manifests itself as random speckles on a smooth surface and it can seriously affect the quality of the image. Sometimes however, it can be helpful to increase the apparent sharpness of a digital image.

- The level of noise usually increases depending on the length of exposure, the physical temperature, and the sensitivity setting of the camera. The amount of certain types of image noise present at a given setting varies for different camera models and is related to the sensor technology.

Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Three Types of Image Noise

- The main types of image noise are random noise, fixed pattern noise, and banding noise. Random noise is shown by fluctuation of the colors above the actual intensity of the image.
- Fixed pattern noise appears after long exposures and high temperatures.
- Finally, banding noise is introduced when the camera is reading data from the sensor and it directly related to in-camera technology factors.

Random Noise
Short Exposure
High ISO Speed

Fixed Noise
Long Exposure
Low ISO Speed

Banding Noise
In-Camera
Brightened Shadows

Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Noise in Digital Image Processing

Noise is always presents in digital images during image acquisition, coding, transmission, and processing steps. It is very difficult to remove noise from the digital images without the prior knowledge of filtering techniques.

A(x,y) = H(x,y) + B(x,y) Where, A(x,y)= function of noisy image, H(x,y)= function of image noise , B(x,y)= function of original image.But in theoretical terms, a picture that we look at is a function of image intensity at a particular position in the image. I.e I(x,y) is an image function where I = Intensity at position (x,y) in an image.

ORIONPLEX ORIONTRAIN

Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

Types of digital images:

There are typically three types of digital images.

1. Binary Images
2. Gray Scale Images
3. Color Images

We can describe image as a function f where x belongs to [a,b] and y belongs to [c,d] which returns as output ranging between maximum and minimum pixel intensity values.

So, it can be stated as, f: [a,b] * [c,d] -> [min,max]

Image Noise and Integrity (Out of Focus, Motion Blur, etc.)

1.Binary images

f: [a,b] * [c,d] -> 0 or 255 (For binary images, the output of the function is either the brightest pixel 255 or the darkest pixel 0)

2.Gray Scale images

f: [a,b] * [c,d] -> [min,max](For gray-scale images, the output of the function is a range of possible values from the brightest pixel 255 to the darkest pixel 0)

3. Color Images

For color images they are three functions stacked together as a “vector valued function. Those function represent red , blue and green pixel values.

$$f(x,y) = \begin{bmatrix} r(x,y) \\ b(x,y) \\ g(x,y) \end{bmatrix}$$

SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Module 2: SIPI in Analogue Circuits (Properties of Electrical Systems)

Impedance and Electrical Models

Impedance and Electrical Models

- In high-speed digital systems, where signal integrity plays a significant role, we often refer to signals as either changing voltages or a changing current.
- All the effects that we lump in the general category of signal integrity are due to how analog signals (those changing voltages and currents) interact with the electrical properties of the interconnects. The key electrical property with which signals interact is the impedance of the interconnects.
- **Impedance is defined as the ratio of the voltage to the current**. We usually use the letter **Z** to represent impedance. The definition, which is always true, is

$$\mathbf{Z = V/I.}$$

Impedance and Electrical Models

Describing Signal-Integrity Solutions in Terms of Impedance Each of the four basic families of signal-integrity problems can be described based on impedance.

1. **Signal-quality problems** arise because voltage signals reflect and are distorted whenever the impedance the signal sees changes. If the impedance the signal sees is always constant, there will be no reflection and the signal will continue undistorted. Attenuation effects are due to series and shunt-resistive impedances.
2. **Crosstalk** arises from the electric and magnetic fields coupling between two adjacent signal traces (and, of course, their return paths). The mutual capacitance and mutual inductance between the traces establishes an impedance, which determines the amount of coupled current.
3. **Rail collapse of the voltage supply** is really about the impedance in the power-distribution system (PDS). A certain amount of current must flow to feed all the ICs in the system. Because of the impedance of the power and ground distribution, a voltage drop will occur as the IC current switches. This voltage drop means the power and ground rails have collapsed from their nominal values.
4. **The greatest source of EMI** is driven by voltages in the ground planes, through external cables. The higher the impedance of the return current paths in the ground planes, the greater the voltage drop, or ground bounce, which will drive the radiating currents.

Physical Basis of Resistance, Capacitance and Inductance

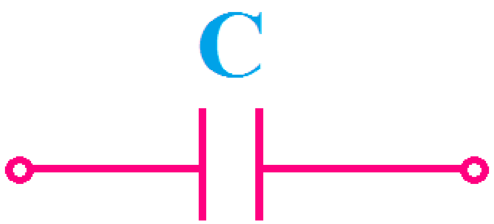
Physical Basis of Resistance, Capacitance and Inductance

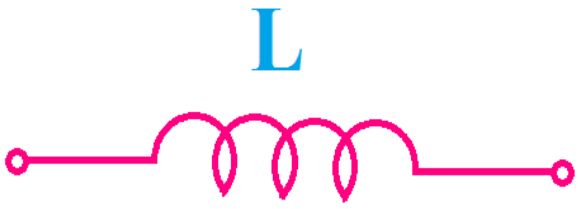
A Resistor is an element which can oppose the flow of current in an electrical or electronic circuit.

The resistor is called a passive element because it has no need of power supply or biasing for its operation. An active element like Transistor need a power supply or biasing for its operation but passive elements have no need of power supply for their operation.

The Capacitor is an element which can store electrical energy for a short time in form of charge.The capacitor is a passive element which does not require a power supply or biasing for its operation.The capacitor is also a bidirectional element which allows the flow of current in both directions through it.

The Inductor is nothing but a coil of wire which opposes any change in current which flows through it. The Inductor is also a passive element which does not require a power supply or biasing for its operation. The inductor is a bidirectional element which allows the flow of current in both directions through it.





ORIONPLEX ORIONTRAIN

Physical Basis of Resistance, Capacitance and Inductance

Resistance

According to the Ohm's law if the temperature and other physical quantity is constant then the resistance formula is,

$$R = \frac{V}{I}$$

Here,

R = resistance

V = Voltage across the Resistor

I = Current flowing through the resistor

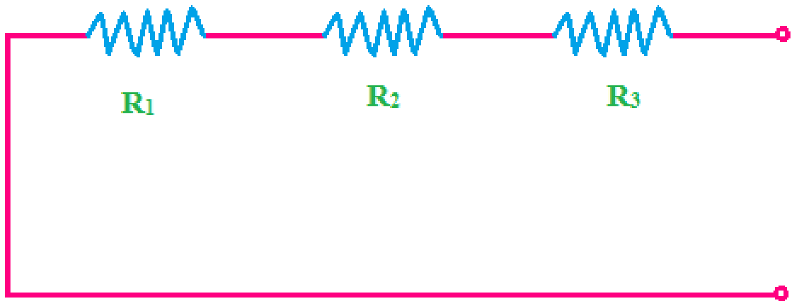
So the formula of the voltage drop across the resistor is,

$$V = IR$$

$$R_{eff} = R1 + R2 + R3 + \ldots\ldots\ldots + R_n$$

Resistors in series:

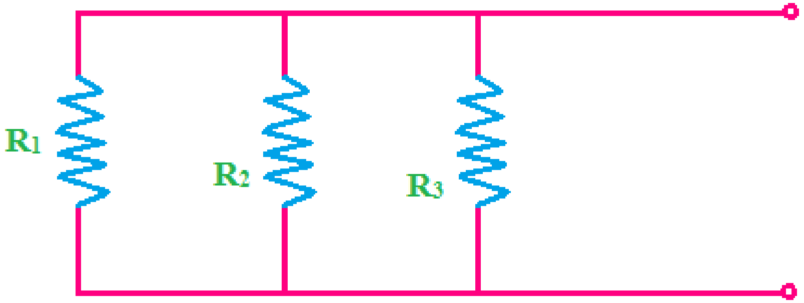
If a no. of resistors are connected in series then the formula of resistance will be,



$$\frac{1}{R_{eff}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots\ldots\ldots + \frac{1}{R_n}$$

Resistors in Parallel:

If a no. of resistors are connected in parallel then the formula of resistance will be,



Physical Basis of Resistance, Capacitance and Inductance

Capacitance

One(1) Coulomb of charge stored per One Volt applied voltage.

When we apply an electrical voltage to the capacitor then the capacitor stores the electrical energy in form of electrical charge.

So we can write.

$$C = \frac{Q}{V}$$

Where

C= capacitance

Q= charge

V= Voltage or potential difference

Based on the physical characteristics of the capacitor, the capacitance formula is,

$$C = \frac{k\varepsilon_0 A}{d}$$

Here,

k= relative permittivity

ε0 = permittivity of free space

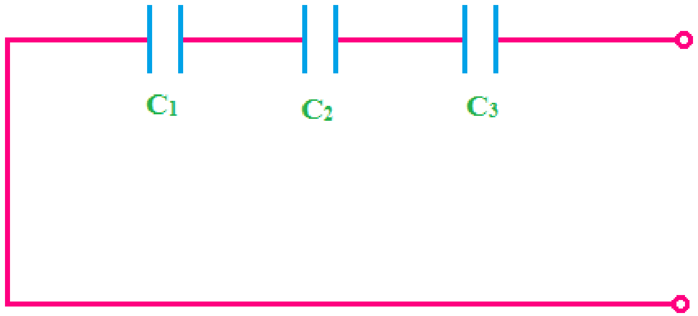
A = surface area of plates

d=distance between plates

Capacitors connected in Series:

When a no. of capacitors are connected in series then the formula of capacitance will be

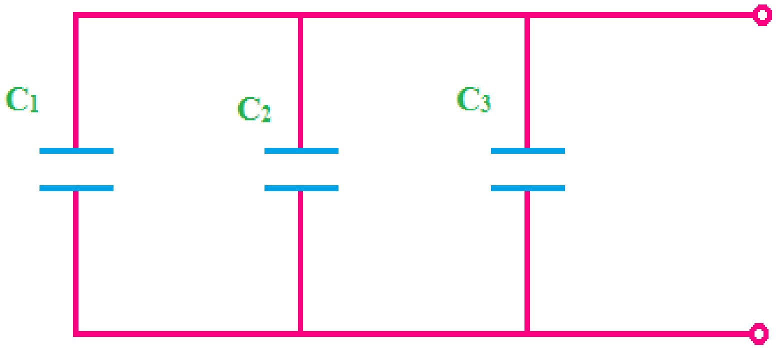
$$\frac{1}{C_{eff}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_n}$$



Capacitors connected in Parallel:

When a no. of capacitors are connected in parallel then the formula of capacitance will be

$$C_{eff} = C_1 + C_2 + C_3 + \cdots + C_n$$



Physical Basis of Resistance, Capacitance and Inductance

The Formula of self-inductance

$$L = \frac{\mu N^2 A}{l}$$

Here, L= self-inductance

N= No. of turns

l (small L)= length of the wire

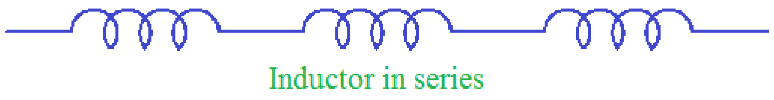
A= area of each turn

$$\text{Area,}\quad A = \pi r^2$$

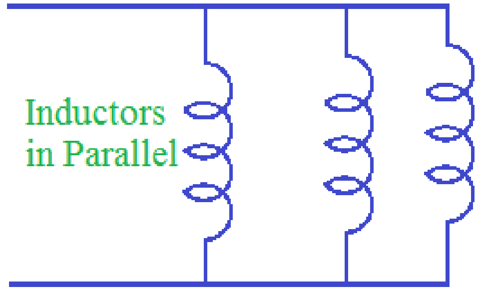
So we can write,

$$L = \frac{\mu N^2 \pi r^2}{l}$$

$$L_{eq} = L_1 + L_2 + L_3 + \cdots + L_n$$



$$\frac{1}{L_{eq}} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \cdots + \frac{1}{L_n}$$



Lumped versus Distributed Systems

Lumped versus Distributed Systems

Lumped Elements

If the size of an element is smaller than the wavelength of the applied signals, then it is a lumped element. In lumped elements, the effect of wave propagation can be neglected. The physical dimensions of lumped elements make it so that signals do not vary over the interconnects interfacing them. There are only minimal phase differences between the input and output signals in lumped elements.

Lumped inductor transformers and baluns are commonly used in RF and microwave applications. Other examples of lumped elements include:

- Resistors
- Capacitors
- Inductors

Lumped versus Distributed Systems

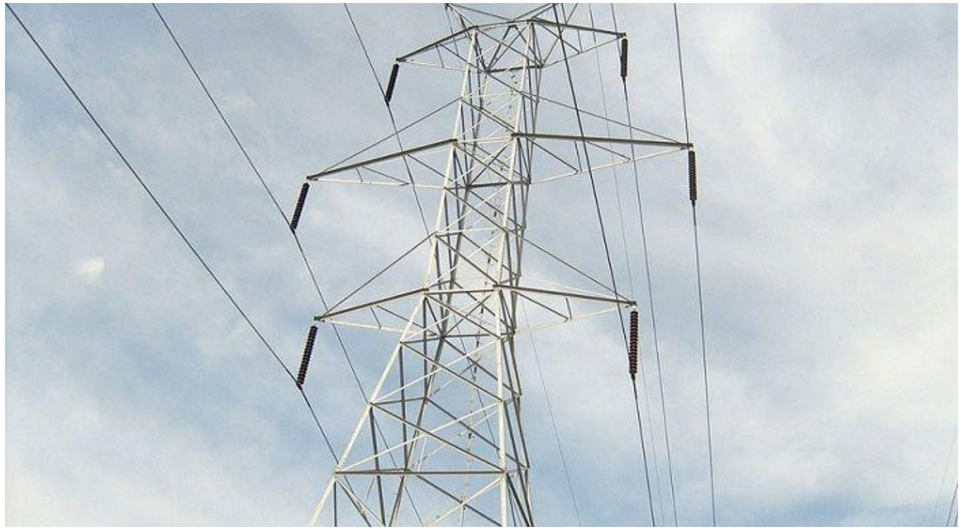
Distributed Elements

- The physical dimension of distributed elements is comparable with the operating wavelength. They are distributed over lengths in an RF or microwave circuit.
- When conventional lumped elements are difficult to implement at microwave frequencies, distributed-elements are used instead. They perform the same functions as lumped elements, but the signals vary along the lines and between the elements.
- Distributed element circuits can be built by forming the medium itself into specific patterns. The significant advantage of distributed element circuits is that they can be manufactured at a lower cost. Distributed elements can be found in printed circuit boards and coaxial and waveguide formats. Other examples of distributed elements used in microwave circuits such as filters, power dividers, and circulators include:
    - Stubs
    - Coupled lines
    - Cascaded lines

Introduction to Transmission Lines

Introduction to Transmission Lines

A transmission line is used for the transmission of electrical power from generating substation to the various distribution units. It transmits the wave of voltage and current from one end to another. The transmission line is made up of a conductor having a uniform cross-section along the line. Air act as an insulating or dielectric medium between the conductors

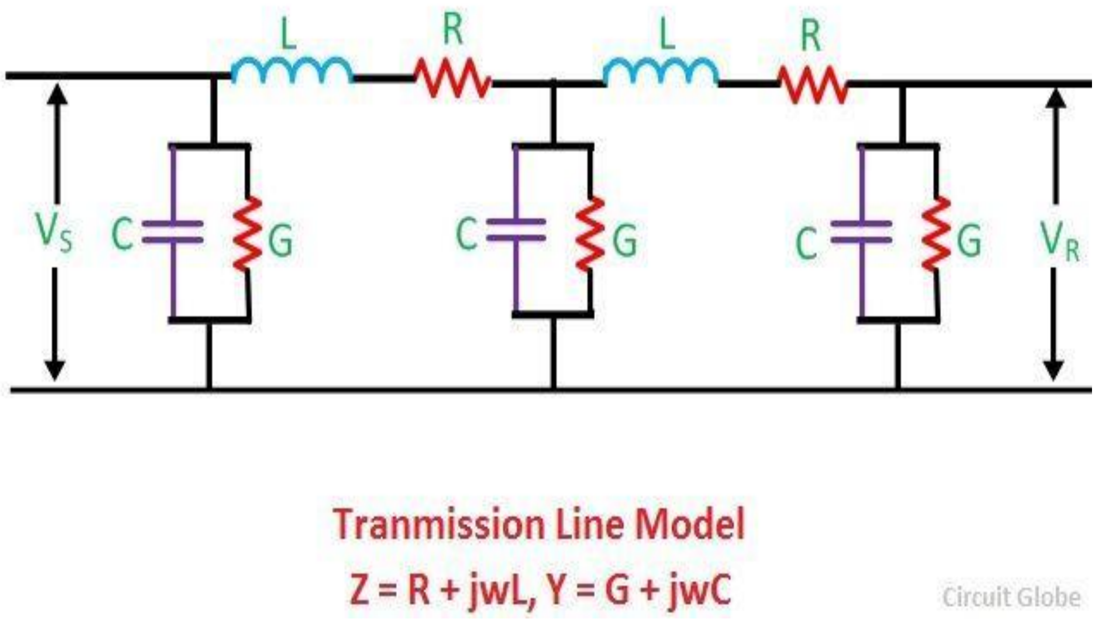


For safety purpose, the distance between the line and ground is much more. The electrical tower is used for supporting the conductors of the transmission line.Tower are made up of steel for providing high strength to the conductor. For transmitting high voltage, over long-distance high voltage direct current is used in the transmission line.

Introduction to Transmission Lines

Parameters of transmission line

The performance of transmission line depends on the parameters of the line. The transmission line has mainly four parameters, resistance, inductance, capacitance and shunt conductance. These parameters are uniformly distributed along the line. Hence, it is also called the distributed parameter of the transmission line.



Introduction to Transmission Lines

Line inductance – The current flow in the transmission line induces the magnetic flux. When the current in the transmission line changes, the magnetic flux also varies due to which emf induces in the circuit. The magnitude of inducing emf depends on the rate of change of flux. Emf produces in the transmission line resist the flow of current in the conductor, and this parameter is known as the inductance of the line.

Line capacitance – In the transmission lines, air acts as a dielectric medium. This dielectric medium constitutes the capacitor between the conductors, which store the electrical energy, or increase the capacitance of the line. The capacitance of the conductor is defined as the present of charge per unit of potential difference.

Capacitance is negligible in short transmission lines whereas in long transmission; it is the most important parameter. It affects the efficiency, voltage regulation, power factor and stability of the system.

Shunt conductance – Air act as a dielectric medium between the conductors. When the alternating voltage applies in a conductor, some current flow in the dielectric medium because of dielectric imperfections. Such current is called leakage current. Leakage current depends on the atmospheric condition and pollution like moisture and surface deposits.

Introduction to Transmission Lines

Performance of transmission lines

The term performance includes the calculation of sending end voltage, sending end current, sending end power factor, power loss in the lines, efficiency of transmission, regulation and limits of power flows during steady state and transient conditions. Performance calculations are helpful in system planning. Some critical parameters are explained below

Voltage regulation – Voltage regulation is defined as the change in the magnitude of the voltage between the sending and receiving ends of the transmission line.

$$\%\,Voltage\ regulation = \frac{Sending\ end\ voltage - Receiving\ end\ voltage}{Sending\ end\ voltage} \times 100$$

The efficiency of transmission lines – Efficiency of the transmission lines is defined as the ratio of the input power to the output power.

$$\%\,Transmission\ line\ efficiency = \frac{Power\ delieverd\ at\ receiving\ end}{Power\ sent\ from\ the\ sending\ end} \times 100$$

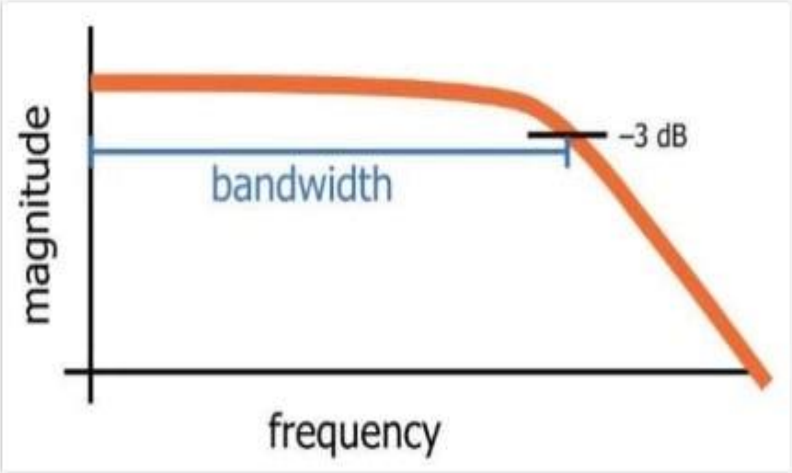
Bandwidth and Rise Time

Bandwidth and Rise Time

Bandwidth of a Digital Signal

- **Bandwidth** is a common frequency domain parameter used to describe the behavior of a circuit. For example, we usually consider a 3-dB bandwidth to describe the frequency response of a filter or communication channel.

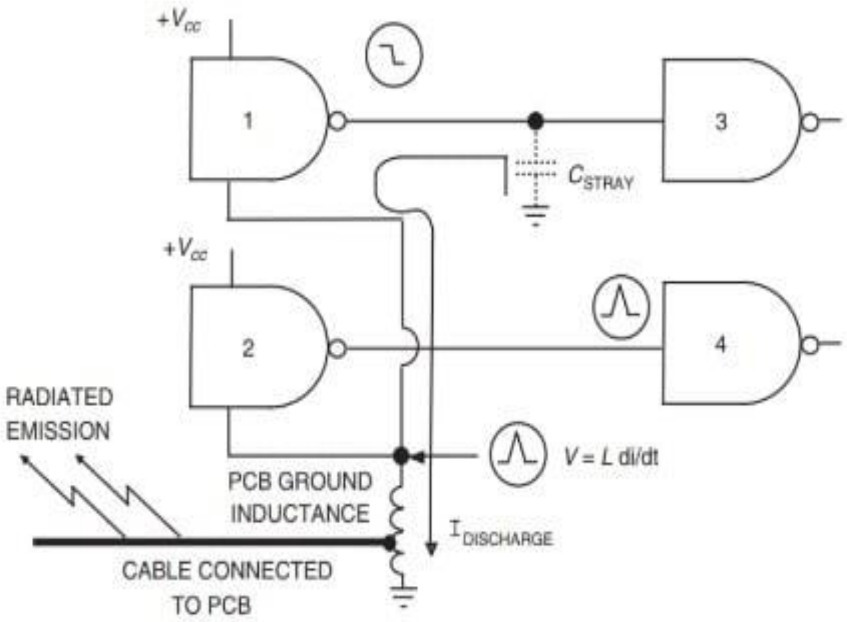
- As shown in Figure Below, the 3dB bandwidth of a low-pass filter is part of the frequency response that lies within 3 dB of the transfer function magnitude at DC (in this figure, the magnitude at DC is 0 dB and it drops to -3 dB at the far end of the transfer function bandwidth).



Bandwidth and Rise Time

Rise Time: An Important Time Domain Parameter

The rise time of a digital signal is a very important time-domain parameter. For example, rise time can directly affect the ground bounce of a PCB. This is illustrated in Figure below



Knee Frequency

Knee Frequency

Knee Frequency

Knee frequency is an estimate of the highest frequency content of the signal. It depends upon the rise time of the signal. If the rise time is smaller, the highest frequency content will have higher frequency. If t denotes the 10% to 90% rise time of a signal, then an estimate for the highest frequency content of the signal or the knee frequency is given by

$$f_{knee} = \frac{0.35}{\tau} \qquad \textbf{1-1}$$

where,

f_{knee} = highest frequency content of the signal

τ = 10% to 90% rise time of the signal

Knee Frequency

Knee Frequency for 2-level Signals

- The digital signal bandwidth depends on the rise time of the signal, this motivates defining some metric for cutting off the bandwidth at some limit. A limit that is often used with two-level signals is the knee frequency, which is defined as shown below:

$$f_{knee} \approx \frac{0.35}{t_{rise}}$$

- The knee frequency for a 2-level square wave with finite edge rate sets a definite limit on the digital signal bandwidth that is useful in many applications.

- The integral of the power spectrum from the repetition rate up to the knee frequency will contain ~75% of the signal power. Note that this is an arbitrary metric, but it is an easy way to set a limit on digital signal bandwidth that everyone can accept and understand.

SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Module 3: Time and Frequency Domains

The Time Domain

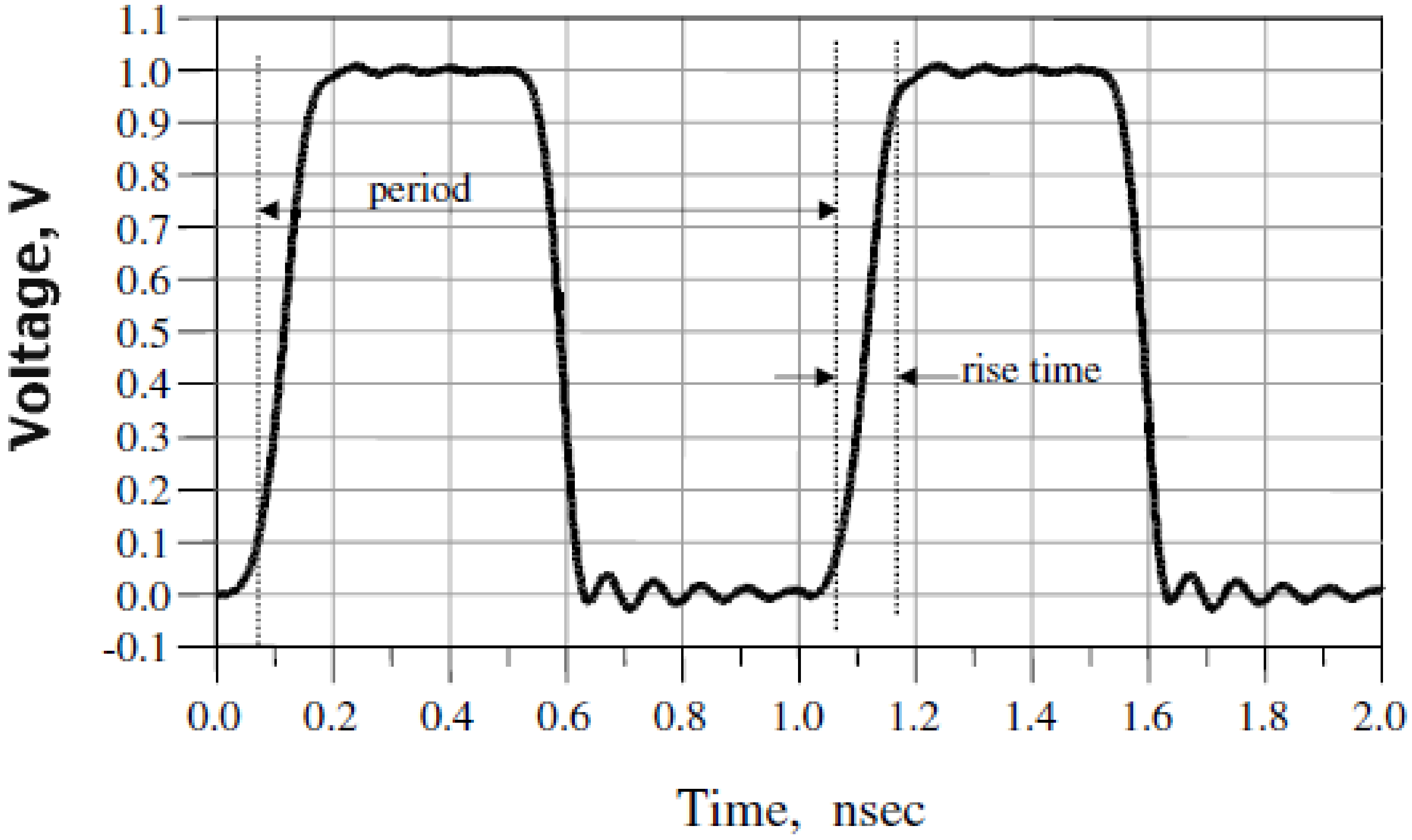
The Time Domain

TIP The time domain is the real world. It is the only domain that actually exists.

The time domain is the world of our experiences and is the domain in which high-speed digital products perform. When evaluating the behavior of a digital product, we typically do the analysis in the time domain because that's where performance is ultimately measured.

The Time Domain

For example, two important properties of a clock waveform are clock period and rise time.



The Time Domain

The clock period is the time interval to repeat one clock cycle, usually measured in nanoseconds (nsec). The clock frequency, F_{clock}, or how many cycles per second the clock goes through, is the inverse of the clock period, T_{clock}.

$$F_{clock} = \frac{1}{T_{clock}}$$

where:

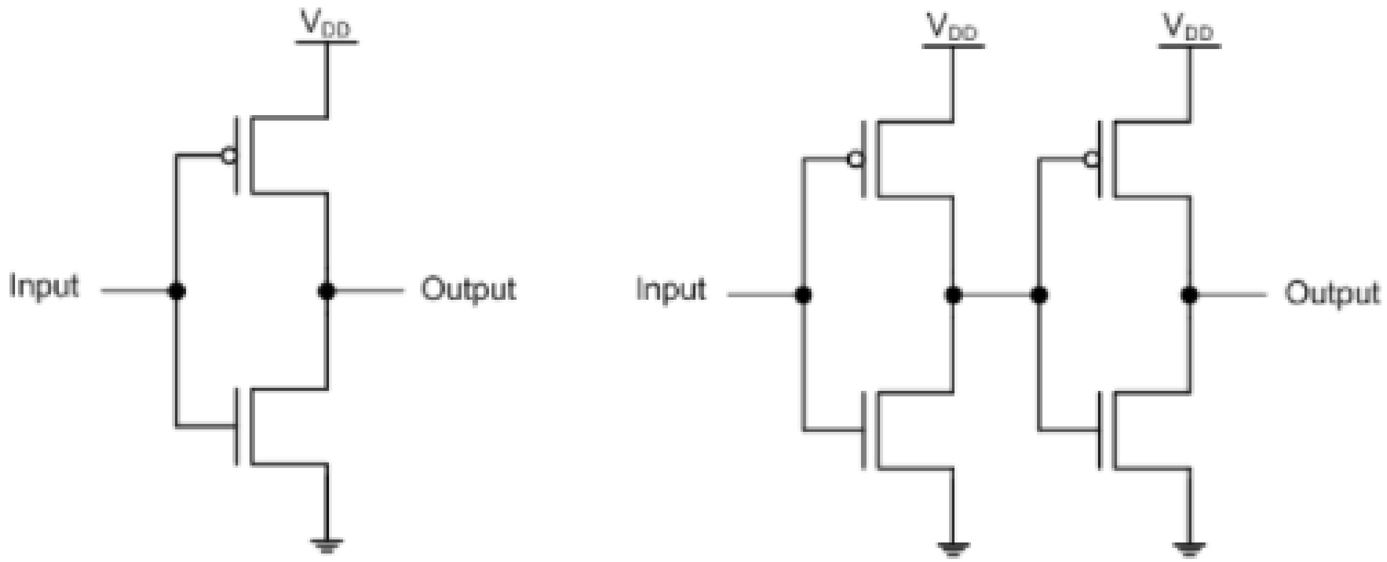
F_{clock} = the clock frequency, in GHz

T_{clock} = the clock period, in nsec

The Time Domain

- Rise Time
  - Time for signal to transition from low to high
  - Two popular definitions
    - Time taken to transition from 10% to 90% of final value
    - Time taken to transition from 20% to 80% of final value
- Fall times are similar defined
  - Usually shorter than rise time depending on logic family
  - Due to design of typical CMOS output driver

Typical CMOS Digital Output Driver



inverted output *non-inverted output*

Only one transistor is on at any one time, depending on whether the output is a low or a high

The Time Domain

- When the driver switches from low to high (rising edge)
  - n transistor turns off, p transistor turns on
- When the driver switches from high to low (falling edge)
  - p transistor turns off, n transistor turns on
- n transistor can turn on faster than p transistor
  - Falling edge shorter than rising edge
  - SI problems more likely to occur on falling edge than rising edge
  - Rising and falling edges can be matched by n channel transistor larger than p channel

Sine Waves in the Frequency Domain

The Frequency Domain

TIP The most important quality of the frequency domain is that it is not real. It is a mathematical construct. The only reality is the time domain. The frequency domain is a mathematical world where very specific rules are followed.

The most important rule in the frequency domain is that the only kind of waveforms that exist are sine waves. Sine waves are the language of the frequency domain.

There are other domains that use other special functions. For example, the JPEG picture-compression algorithm takes advantage of special waveforms that are called wavelets. The wavelet transform takes the space domain, with a lot of x-y amplitude information content, and translates it into a different mathematical description that is able to use less than 10% of the memory to describe the same information. It is an approximation, but a very good one.

The Frequency Domain

It's common for engineers to think that we use sine waves in the frequency domain because we can build any time-domain waveform from combinations of sine waves. This is a very important property of sine waves. However, there are many other waveforms with this property. It is not a property that is unique to sine waves.

In fact, there are four properties that make sine waves very useful for describing any other waveform. These properties are as follows:

The Frequency Domain

1. Any waveform in the time domain can be completely and uniquely described by combinations of sine wave.
2. Any two sine waves with different frequencies are orthogonal to each other. If you multiply them together and integrate over all time, they integrate to zero. This means you can separate each component from every other.
3. They are well defined mathematically.
4. They have a value everywhere with no infinities and they have derivatives that have no infinities anywhere. This means they can be used to describe real world waveforms, since there are no infinities in the real world.

The Frequency Domain

All of these properties are vitally important, but are not unique to sine waves. There is a whole class of functions called *orthonormal functions*, or sometimes called *eigenfunctions* or *basis functions*, which could be used to describe any time-domain waveform. Other orthonormal functions are Hermite Polynomials, Legendre Polynomials, Laguerre Polynomials, and Bessel Functions.

Why did we choose sine waves as our functions in the frequency domain? What's so special about sine waves? The real answer is that by using sine waves, some problems related to the electrical effects of interconnects will be easier to understand and solve using sine waves. If we switch to the frequency domain and use sine-wave descriptions, we can sometimes get to an answer faster than staying in the time domain.

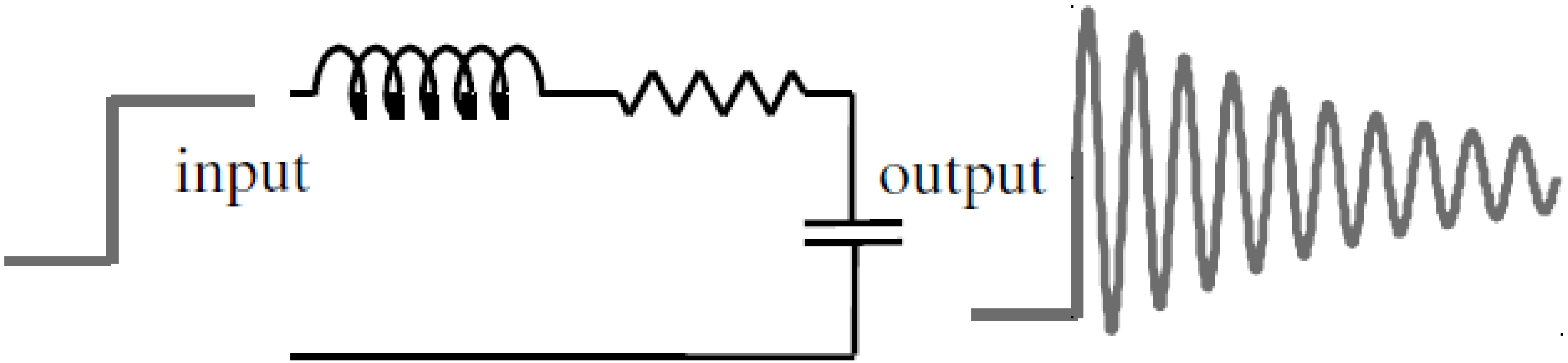
The Frequency Domain

TIP After all, if the time domain is the real world, we would never leave it unless the frequency domain provides a faster route to an acceptable answer.

Sine waves can sometimes provide a faster path to an acceptable answer because of the types of electrical problems we often encounter in signal integrity. If we look at the circuits that describe interconnects, we find that they will often include combinations of resistors (R), inductors (L), and capacitors (C). These elements in a circuit can be described by a second-order linear differential equation. The solution to this type of differential equation is a sine wave. In these circuits, the naturally occurring waveforms will be combinations of the waveforms that are solutions to the differential equation.

The Frequency Domain

We find that in the real world, if we build circuits that contain Rs, Ls, and Cs and send any arbitrary waveform in, more often than not, we get waveforms out that look like sine waves and can more simply be described by a combination of a few sine waves.



Time-domain behavior of a fast edge interacting with an ideal RLC circuit. Sine waves are naturally occurring when digital signals interact with interconnects, which can often be described as combinations of ideal RLC circuit elements.

Shorter Time to a Solution in the Frequency Domain

In some situations, if we use the naturally occurring sine waves in the frequency domain rather than in the time domain, we may arrive at a simpler description to a problem and get to a solution faster.

It is important to keep in mind that there is fundamentally no new information in the frequency domain. The time- and the frequency-domain descriptions of the same waveforms will each have exactly the same information content.

However, some problems are easier to understand and describe in the frequency domain than in the time domain. For example, the concept of bandwidth is intrinsically a frequency-domain idea. We use this term to describe the most significant sine-wave frequency components associated with a signal, a measurement, a model, or an interconnect.

Shorter Time to a Solution in the Frequency Domain

Impedance is defined in both the time and the frequency domain. However, it is far easier to understand, to use, and to apply the concepts of impedance in the frequency domain. We need to understand impedance in both domains, but we will often get to an answer faster by solving an impedance problem in the frequency domain first.

Looking at the impedance of the power and ground distribution in the frequency domain will allow a simpler explanation and solution to rail-collapse problems. As we shall see, the design goal for the power-distribution system is to keep its impedance below a target value from direct current (DC) up to the bandwidth of the typical signals.

Shorter Time to a Solution in the Frequency Domain

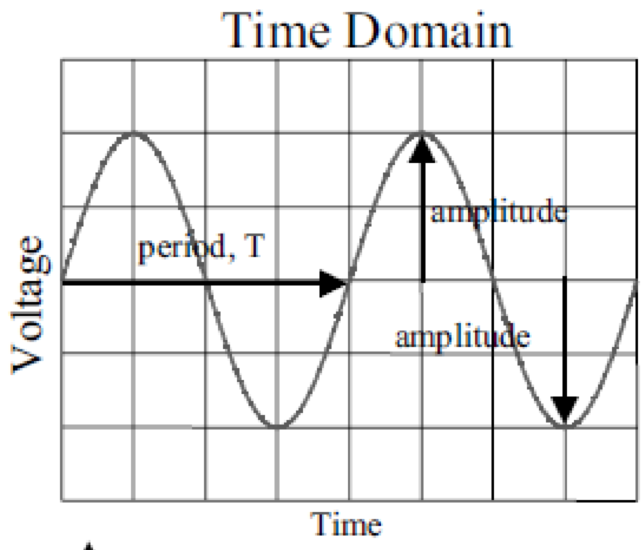
When dealing with EMI issues, both the FCC specifications and the methods of measuring the electromagnetic compliance of a product are more easily performed in the frequency domain.

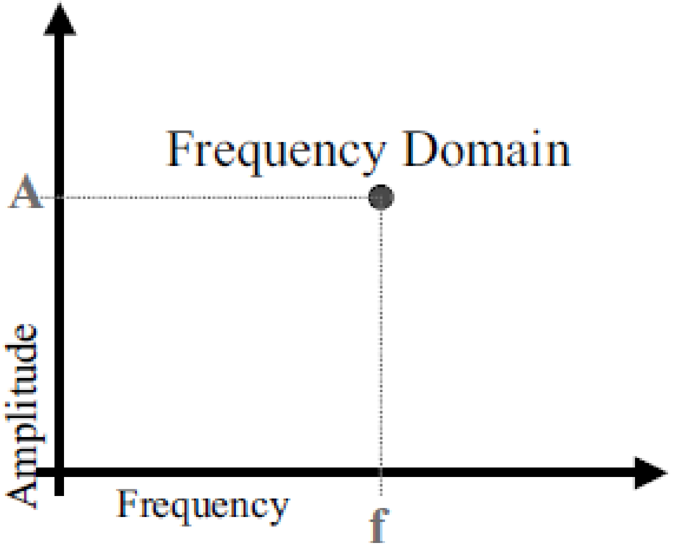
Many of the effects related to lossy transmission lines are more easily analyzed, measured, and simulated by using the frequency domain. The series resistance of a transmission line increases with the square root of frequency, and the shunt AC leakage current in the dielectric increases linearly with frequency. The transient (time-domain) performance of lossy transmission lines is often more easily obtained by first transforming the signal into the frequency domain, looking at how the transmission line affects each frequency component separately, and then transforming the sine-wave components back to the time domain.

Sine Wave Features

The following three terms fully describe a sine wave:

- Frequency
- Amplitude
- Phase





Top: Description of a sine wave in the time domain. It is composed of over one thousand voltage-versus-time data points. Bottom: Description of a sine wave in the frequency domain. Only three terms define a sine wave, which is a single point in the frequency domain.

Sine Wave Features

The sine-wave frequency and the angular frequency are related by:

$$\omega = 2\pi \times f$$

where:

ω = angular frequency, in radians/sec

π = constant, 3.14159...

f = sine-wave frequency, in Hz

Fourier Transform

The Fourier Transform

The starting place for using the frequency domain is being able to convert a waveform from the time domain into a waveform in the frequency domain. We do this with the Fourier Transform. There are three types of Fourier Transforms:

- Fourier Integral (FI)
- Discrete Fourier Transform (DFT)
- Fast Fourier Transform (FFT)

The Fourier Transform

The Fourier Integral (FI) is a mathematical technique of transforming an ideal mathematical expression in the time domain into a description in the frequency domain. For example, if the entire waveform in the time domain were just a short pulse, and nothing else, the Fourier Integral would be used to transform to the frequency domain.

This is done with an integral over all time from – infinity to + infinity. The result is a frequency-domain function that is also continuous from 0 to + infinity frequencies. There is a value for the amplitude at every continuous frequency value in this range.

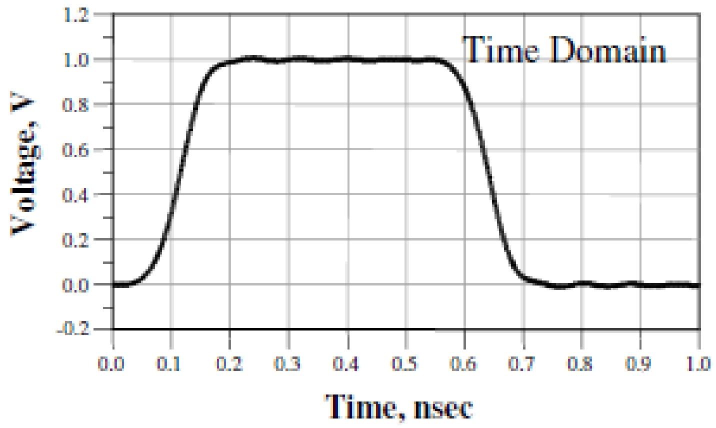
The Fourier Transform: Fourier Integral (FI)

The Fourier Integral (FI) is a mathematical technique of transforming an ideal mathematical expression in the time domain into a description in the frequency domain. For example, if the entire waveform in the time domain were just a short pulse, and nothing else, the Fourier Integral would be used to transform to the frequency domain.

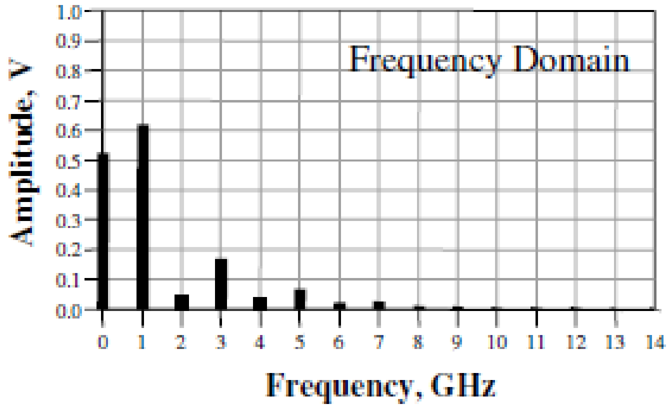
This is done with an integral over all time from – infinity to + infinity. The result is a frequency-domain function that is also continuous from 0 to + infinity frequencies. There is a value for the amplitude at every continuous frequency value in this range.

The Fourier Transform: Discrete Fourier Transform (DFT)

For real-world waveforms, the time-domain waveform is actually composed of a series of discrete points, measured over a finite time, T. For example, a clock waveform may be a signal from 0 v to 1 v and have a period of 1 nsec and a repeat frequency of 1 GHz. To represent one cycle of the clock, there might be as many as 1000 discrete data points, taken at 1-psec intervals. An example of a 1-GHz clock wave in the time domain is shown.



To transform this waveform into the frequency domain, the Discrete Fourier Transform (DFT) would be used. The basic assumption is that the original time-domain waveform is periodic and repeats every T seconds. Rather than integrals, just summations are used so any arbitrary set of data can be converted to the frequency domain using simple numerical techniques.



One cycle of a 1-GHz clock signal in the time domain (top) and frequency domain (bottom).

ORIONPLEX | ORIONTRAIN

The Fourier Transform: Fast Fourier Transform (FFT)

Finally, there is the Fast Fourier Transform (FFT). It is exactly the same as a DFT, except that the actual algorithm used to calculate the amplitude values at each frequency point uses a trick of very fast matrix algebra. This trick works only if the number of time-domain data points is a power of two, for example 256 points, or 512 points, or 1024 points. The result is a DFT, only calculated 100–10,000 times faster than the general DFT algorithm, depending on the number of voltage points.

In general, it is common in the industry to use all three terms, FI, DFT, and FFT, synonymously. We now know there is a difference between them, but they have the same purpose—to translate a time-domain waveform into its frequency-domain spectrum.

Spectral Content of Digital Signals

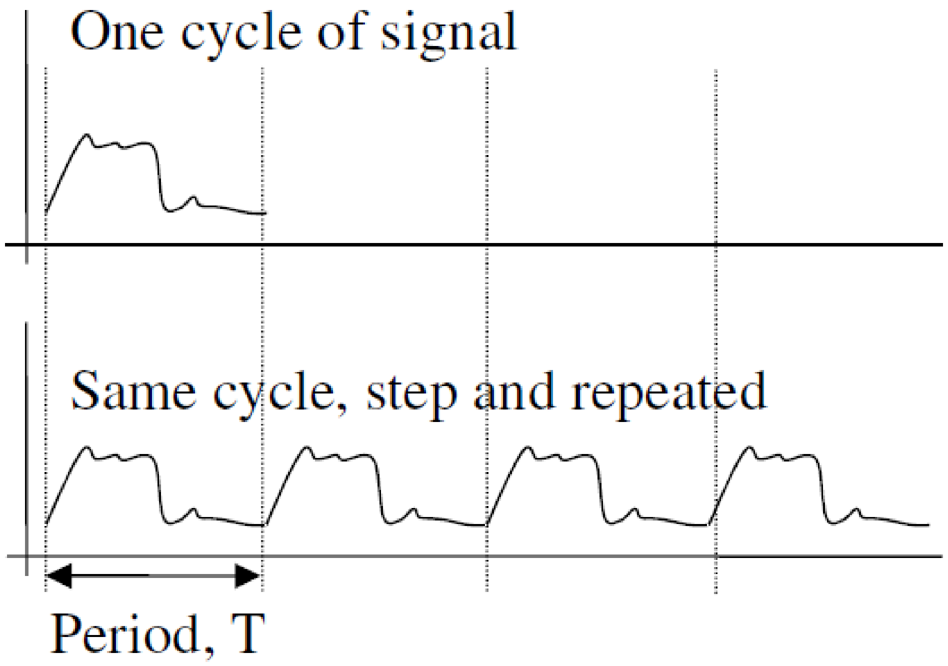
The Spectrum of a Repetitive Signal

In practice, the DFT or FFT is used to translate a real waveform from the time domain to the frequency domain. It is possible to take a DFT of any arbitrary, measured waveform. A key requirement of the waveform is that it be repetitive. We usually designate the repeat frequency of the time-domain waveform with the capital letter F.

For example, an ideal square wave might go from 0 v to 1 v, with a repeat time of 1 nsec and a 50% duty cycle. As an ideal square wave, the rise time to transition from 0 v to 1 v is precisely 0 sec. The repeat frequency would be 1/1 nsec = 1 GHz.

The Spectrum of a Repetitive Signal

If a signal in the time domain is some arbitrary waveform over a time interval from t = 0 to t = T, it may not look repetitive. However, it can be turned into a repetitive signal by just repeating the interval every T seconds. In this case, the repeat frequency would be F = 1/T. Any arbitrary waveform can be made repetitive and the DFT used to convert it to the frequency domain.



Any arbitrary waveform can be made to look repetitive. A DFT can be performed only on a repetitive waveform.

The Spectrum of a Repetitive Signal

For a DFT, only certain frequency values exist in the spectrum. These values are determined by the choice of the time interval or the repeat frequency. When using an automated DFT tool, such as in SPICE, it is recommended to choose a value for the period equal to the clock period. This will simplify the interpretation of the results.

The only sine-wave frequency values that will exist in the spectrum will be multiples of the repeat frequency. If the clock frequency is 1 GHz, for example, the DFT will only have sine wave components at 1 GHz, 2 GHz, 3 GHz, etc.

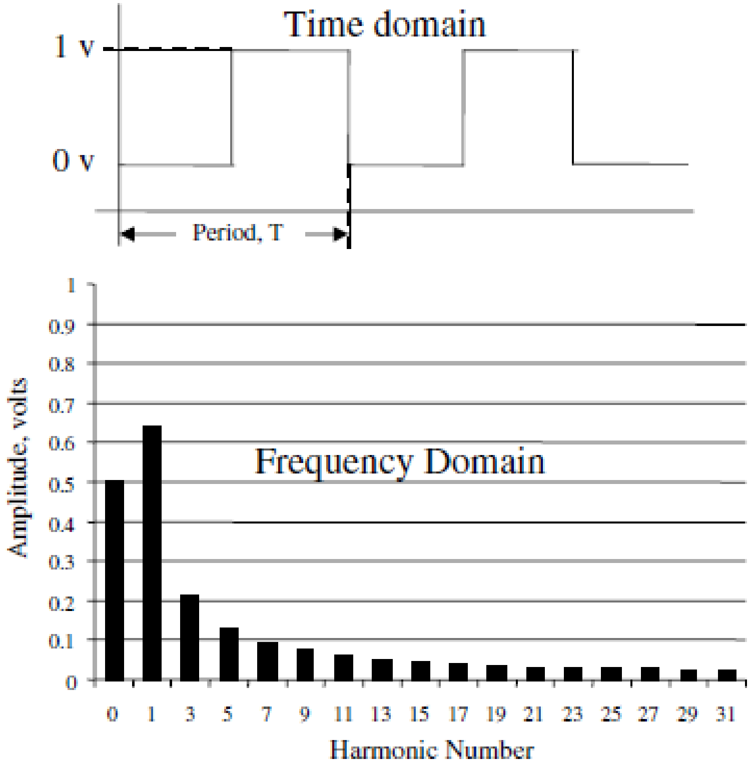
The Spectrum of a Repetitive Signal

The first sine-wave frequency is called the first harmonic. The second sine-wave frequency is called the second harmonic, and so on. Each harmonic will have a different amplitude and phase associated with it. The collection of all the harmonics and their amplitudes is called the spectrum.

The actual amplitudes of each harmonic will be determined by the values calculated by the DFT. Every specific waveform will have its own spectrum.

The Spectrum of an Ideal Square Wave

An ideal square wave has a zero rise time, by definition. It is not a real waveform; it is an approximation to the real world. However, useful insight can be gained by looking at the spectrum of an ideal square wave and using this to evaluate real waveforms later. An ideal square wave has a 50% duty cycle, is symmetrical, and has a peak voltage of 1 v.



Time and frequency domain views of an ideal square wave.

The Spectrum of an Ideal Square Wave

If the ideal square-wave repeat frequency is 1 GHz, the sine-wave frequency values in its spectrum will be multiples of 1 GHz. We expect to see components at f = 1 GHz, 2 GHz, 3 GHz, and so on. But what are the amplitudes of each sine wave? The only way to determine this is to perform a DFT on the ideal square wave. Luckily, it is possible to calculate the DFT exactly for this special case of an ideal square wave. The result is relatively simple.

The Spectrum of an Ideal Square Wave

The amplitudes of all the even harmonics (e.g., 2 GHz, 4 GHz, 6 GHz) are all zero. It is only odd harmonics that have values. The amplitudes, A_n, of the odd harmonics are given by:

$$A_n = \frac{2}{\pi \times n}$$

where:

A_n = the amplitude of the n^{th} harmonic

π = the constant, 3.14159...

n = the harmonic number, only odd allowed

The Spectrum of an Ideal Square Wave

For example, an ideal square wave with 50% duty-cycle and 0 v to 1 v transition has a first harmonic amplitude of 0.63 v. The amplitude of the third harmonic is 0.21 v. We can even calculate the amplitude of the 1001st harmonic. It is 0.00063 v. It is important to note that the amplitudes of higher sine-wave-frequency components decrease with 1/f.

There is one other special frequency value, 0 Hz. Since sine waves are all centered about zero, any combination of sine waves can only describe waveforms in the time domain that are centered about zero. To allow a DC offset, or a nonzero average value, the DC component is stored in the zero-frequency value. This is sometimes called the zeroth harmonic. Its amplitude is equal to the average value of the signal. In the case of the 50% duty-cycle square wave, the zeroth harmonic is 0.5 v.

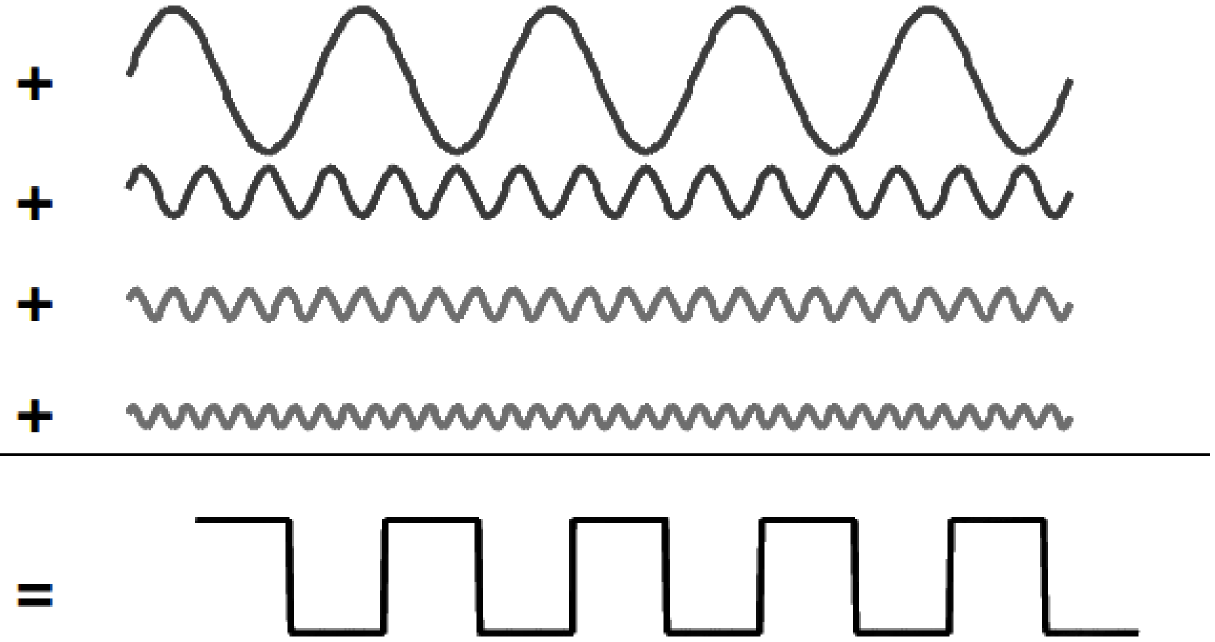
The Spectrum of an Ideal Square Wave

To summarize:

- The collection of sine-wave-frequency components and their amplitudes is called the spectrum. Each component is called a harmonic.
- The zeroth harmonic is the DC value.
- For the special case of a 50% duty-cycle ideal square wave, the even harmonics have an amplitude of zero.
- The amplitude of any harmonic can be calculated as $2/(\pi \times n)$.

From the Frequency Domain to the Time Domain

The spectrum, in the frequency domain, represents all the sine-wave-frequency amplitudes of the time-domain waveform. If we have a spectrum and want to look at the time-domain waveform, we simply take each frequency component, convert it into its time-domain sine wave, then add it to all the rest. This process is called the Inverse Fourier Transform.



Convert the frequency-domain spectrum into the time-domain waveform by adding up each sine-wave component.

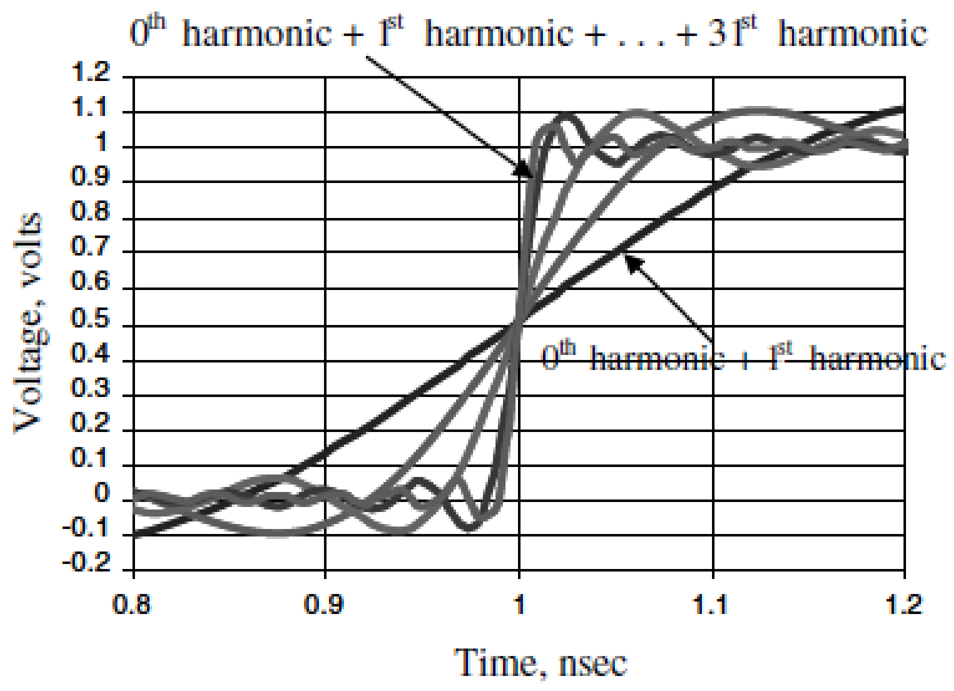
Effect of Bandwidth on Rise Time

The term *bandwidth* is used for the highest sine-wave-frequency component that is significant in the spectrum. This is the highest sine-wave frequency we need to include to adequately approximate the important features of the time-domain waveform. All frequency components of higher frequency than the bandwidth can be ignored. In particular, as we will see, the bandwidth we choose will have a direct effect on the shortest rise time of the signal we are able to describe in the time domain.

The term *bandwidth* historically is used in the rf world to refer to the range of frequencies in a signal. In rf applications, a carrier frequency is typically modulated with some amplitude or phase pattern. The spectrum of frequency components in the signal falls within a band. The range of frequencies in the rf signal is called the bandwidth. Typical rf signals might have a carrier frequency of 1.8 GHz with a bandwidth about this frequency of 100 MHz. The bandwidth of an rf signal defines how dense different communications channels can fit.

Effect of Bandwidth on Rise Time

When we created a time-domain waveform from just the zeroth, the first, and the third harmonics included, the bandwidth of the resulting waveform was just up to the third harmonic, or 3 GHz in this case. By design, the highest sine-wave-frequency component in this waveform is 3 GHz. The amplitude of all other sine-wave components in this time-domain waveform is exactly 0.



The time domain waveform created by adding together the zeroth harmonic and first harmonic, then the third harmonic and then up to the seventh harmonic, then up to the nineteenth harmonic, and then all harmonics up to the thirty-first harmonic, for a 1-GHz ideal square wave.

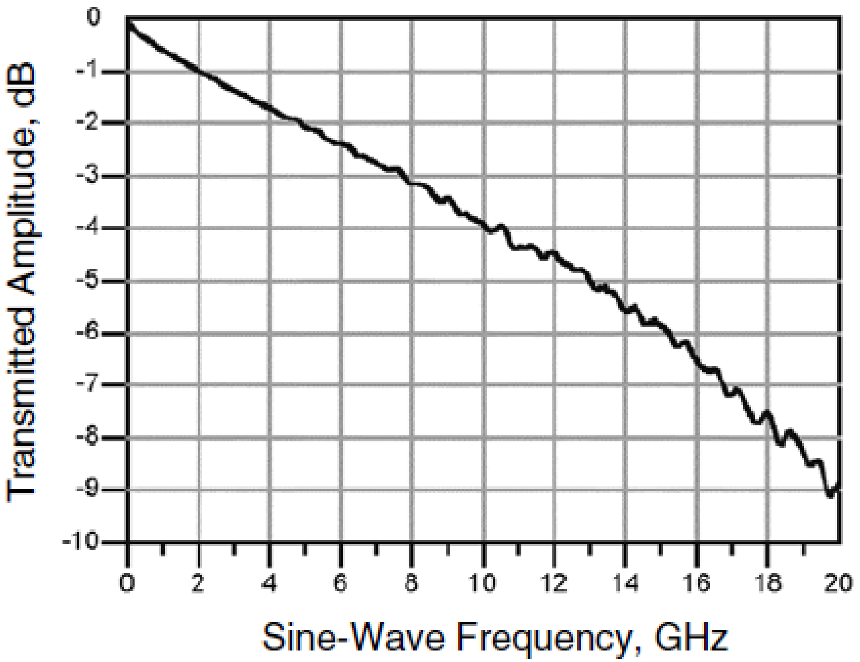
Effect of Bandwidth on Rise Time

In each case, we created a waveform with a higher bandwidth, using the ideal-square-wave's spectrum as the starting place. And, in each case, the higher-bandwidth waveform also had a shorter 10–90 rise time. The higher the bandwidth, the shorter the rise time and the more closely the waveform approximates an ideal square wave. Likewise, if we do something to a short rise-time signal to decrease its bandwidth (i.e., eliminate high-frequency components), its rise time will increase.

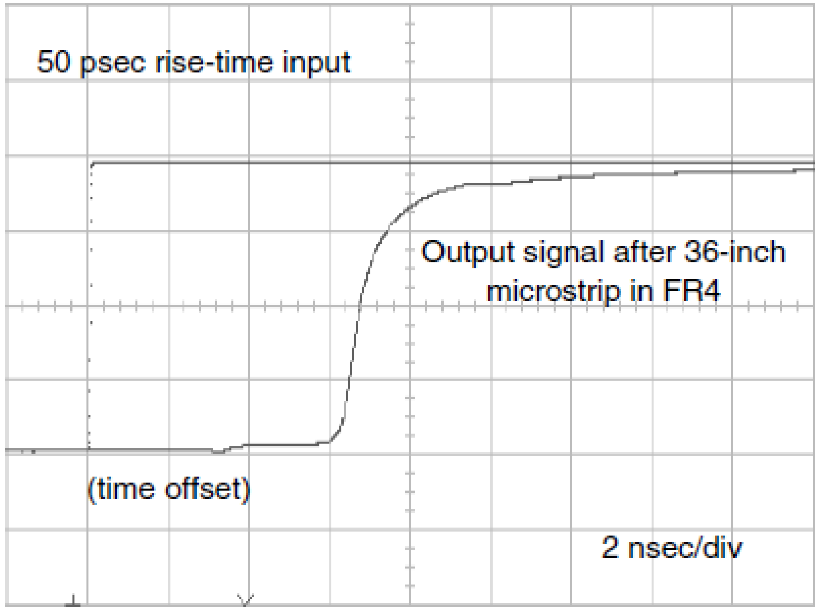
For example, it is initially difficult to evaluate the time-domain response of a signal propagating down a lossy transmission line in FR4. As we will see, there are two loss mechanisms: conductor loss and dielectric loss. If each of these processes were to attenuate low-frequency components the same as they do high-frequency components, there would simply be less signal at the far end, but the pattern of the spectrum would look the same coming out as it does going in. There would be no impact on the rise time of the waveform.

Effect of Bandwidth on Rise Time

However, both conductor loss and dielectric loss will attenuate the higher-frequency components more than the low-frequency components. By the time the signal has traveled through even four inches of trace, the high-frequency components, above about 8 GHz, can have lost more than 50% of their power, leaving the low-frequency terms less affected.



The measured attenuation through a 4-inch length of 50-Ohm transmission line in FR4 showing the higher attenuation at higher frequencies.



The measured input and transmitted signal through a 36-inch 50-Ohm transmission line in FR4, showing the rise time to have degraded from 50 psec to more than 1.5 nsec.

Effect of Bandwidth on Rise Time

This preferential attenuation of higher frequencies has the impact of decreasing the bandwidth of a signal that would propagate through the interconnect.

The rise time has been increased from 50 psec to nearly 1.5 nsec, due to the higher attenuation of the high-frequency components. Thirty-six inches is a typical length for a trace that travels over two 6-inch-long daughter cards and 24 inches of backplane. This rise-time degradation is the chief limitation to the use of FR4 laminate in high-speed serial links above 1 GHz.

TIP In general, a shorter rise-time waveform in the time domain will have a higher bandwidth in the frequency domain. If something is done to the spectrum to decrease the bandwidth of a waveform, the rise time of the waveform will be increased.

Bandwidth and Rise Time

For the special case of a re-created square wave with only some of the higher harmonics included, the bandwidth is inversely related to the rise time. We can fit a straight-line approximation through the points and find the relationship between bandwidth and rise time as:

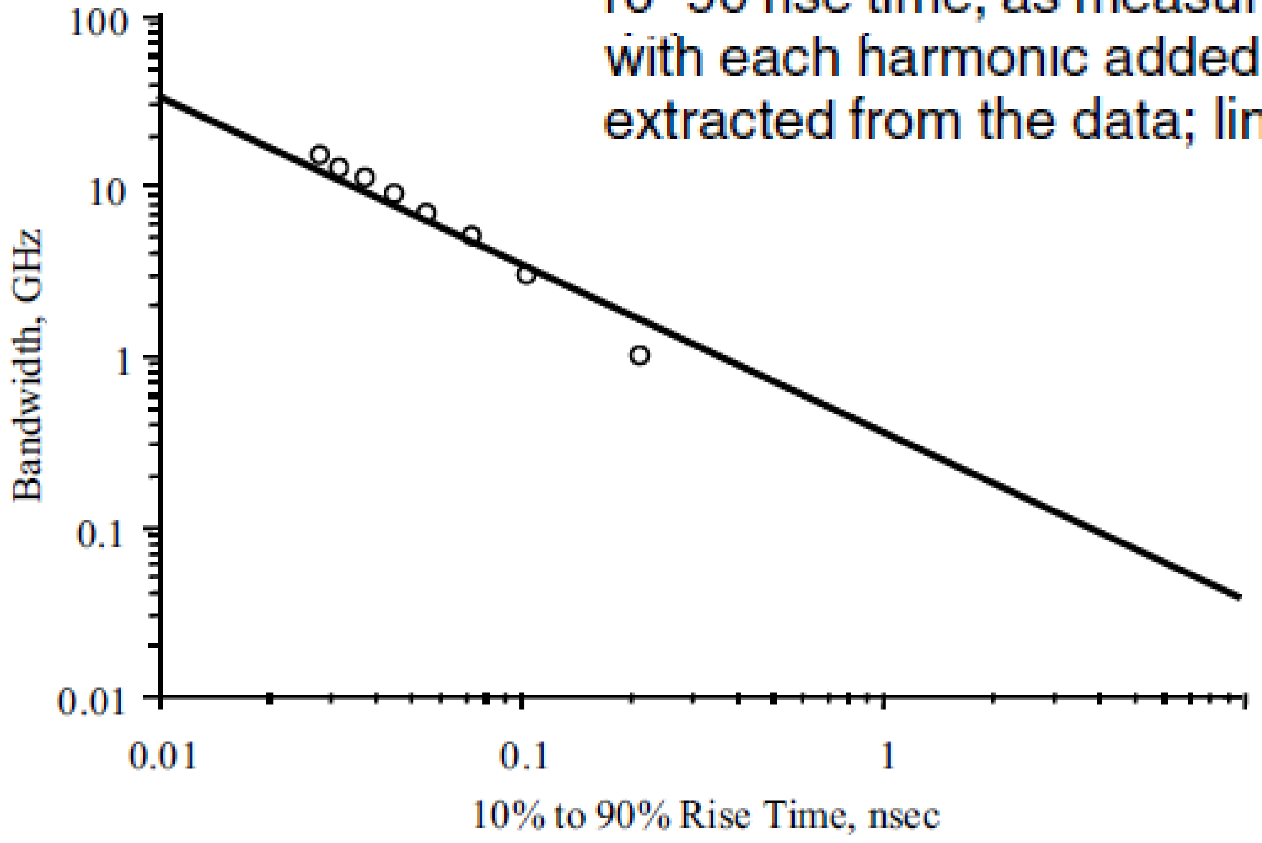
$$BW = \frac{0.35}{RT}$$

where:

BW = the bandwidth, in GHz

RT = the 10–90 rise time, in nsec

Bandwidth and Rise Time



Empirical relationship between the bandwidth of a signal and its 10–90 rise time, as measured from a re-created ideal square wave with each harmonic added one at a time. Circles are the values extracted from the data; line is the approximation of BW = 0.35/rise time.

What Does *Significant* Mean?

If we take an ideal-square-wave clock signal with clock frequency of 1 GHz, its first harmonic will be a 1-GHz sine-wave frequency. If we were to include 100% of every component up to the twenty-first harmonic, the bandwidth would be 21 GHz and the resulting rise time of the re-created signal would be 0.35/21 GHz = 0.0167 nsec or 16.7 psec.

For a real time-domain waveform, the spectral components will almost always drop off in frequency faster than those of an ideal square wave of the same repeat frequency. The question of significance is really about the frequency at which amplitudes of the higher harmonics become small compared to the corresponding amplitudes of an ideal square wave.

What Does *Significant* Mean?

By "small," we usually mean when the power in the component is less than 50% of the power in an ideal square wave's amplitude. A drop of 50% in power is the same as a drop to 70% in amplitude. This is really the definition of significant. Significant is when the amplitude is still above 70% of an ideal square wave's amplitude of the same harmonic.

TIP For any real waveform that has a finite rise time, *significant* refers to the point at which its harmonics are still more than 70% of the amplitude of an equivalent repeat-frequency ideal square wave's.

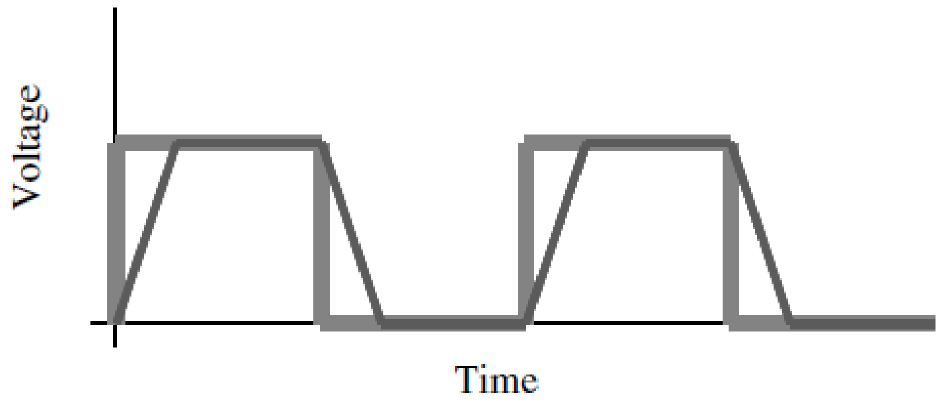
What Does *Significant* Mean?

In a slightly different view, we can define *significant* as the frequency at which the harmonic components of the real waveform begin to drop off faster than 1/f. The frequency at which this happens is sometimes referred to as the *knee frequency*. The harmonic amplitudes of an ideal square wave will initially drop off similarly as 1/f. The frequency at which the harmonic amplitudes of a real waveform begin to significantly deviate from an ideal square wave is the knee frequency.

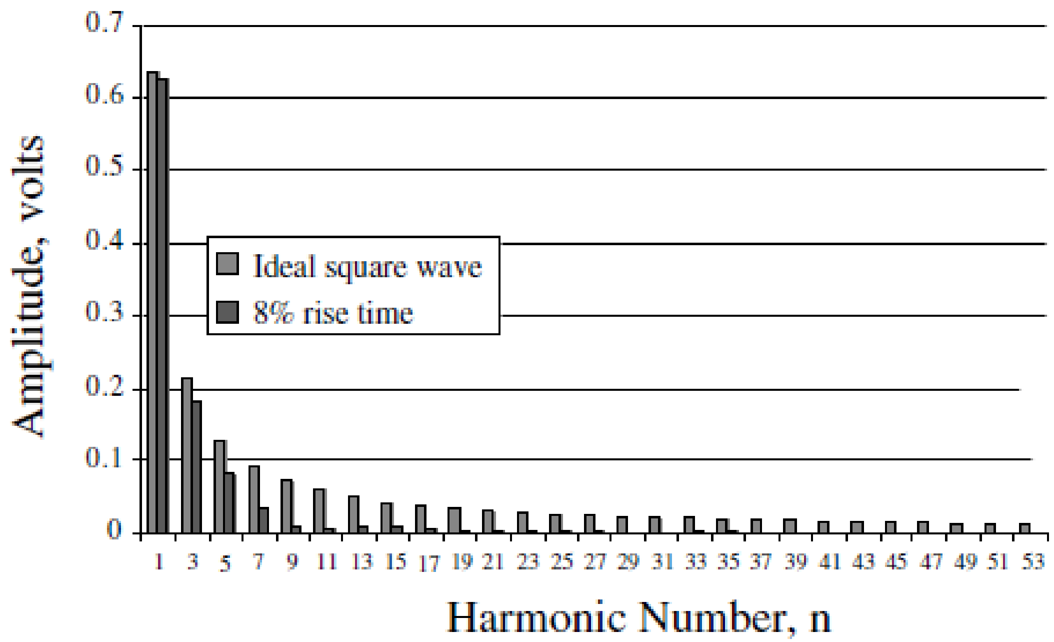
To evaluate the bandwidth of a time-domain waveform, we are really asking what is the highest frequency component that is just barely above 70% of the same harmonic of an equivalent ideal square wave. When the harmonic amplitudes of the real waveform are significantly less than an ideal square wave's, these lower amplitude harmonics will not contribute significantly to decreasing the rise time and we can ignore them.

What Does *Significant* Mean?

For example, we can compare, in the time-domain waveform, two clock waves with a repeat frequency of 1 GHz: an ideal square wave and an ideal trapezoidal waveform, which is a non-ideal square wave with a long rise time. In this example, the 10–90 rise time is about 0.08 nsec, which is a rise time of about 8% of the period, typical of many clock waveforms.



Time domain waveforms of 1-GHz repeat frequency: an ideal square wave and an ideal trapezoidal wave with 0.08-nsec rise time.



Frequency-domain spectra of these waveforms showing the drop-off of the trapezoidal wave's higher harmonics, compared to the square wave's.

What Does *Significant* Mean?

If we compare the frequency components of these waveforms, at what frequency will the trapezoid's spectrum start to differ significantly from the ideal square wave's? We would expect the trapezoid's higher frequency components to begin to become insignificant at about 0.35/0.08 nsec = about 5 GHz. This is the fifth harmonic. After all, we could create a non-ideal square wave with this rise time if we were to take the ideal-square-wave spectrum and drop all components above the fifth harmonic, as we saw earlier.

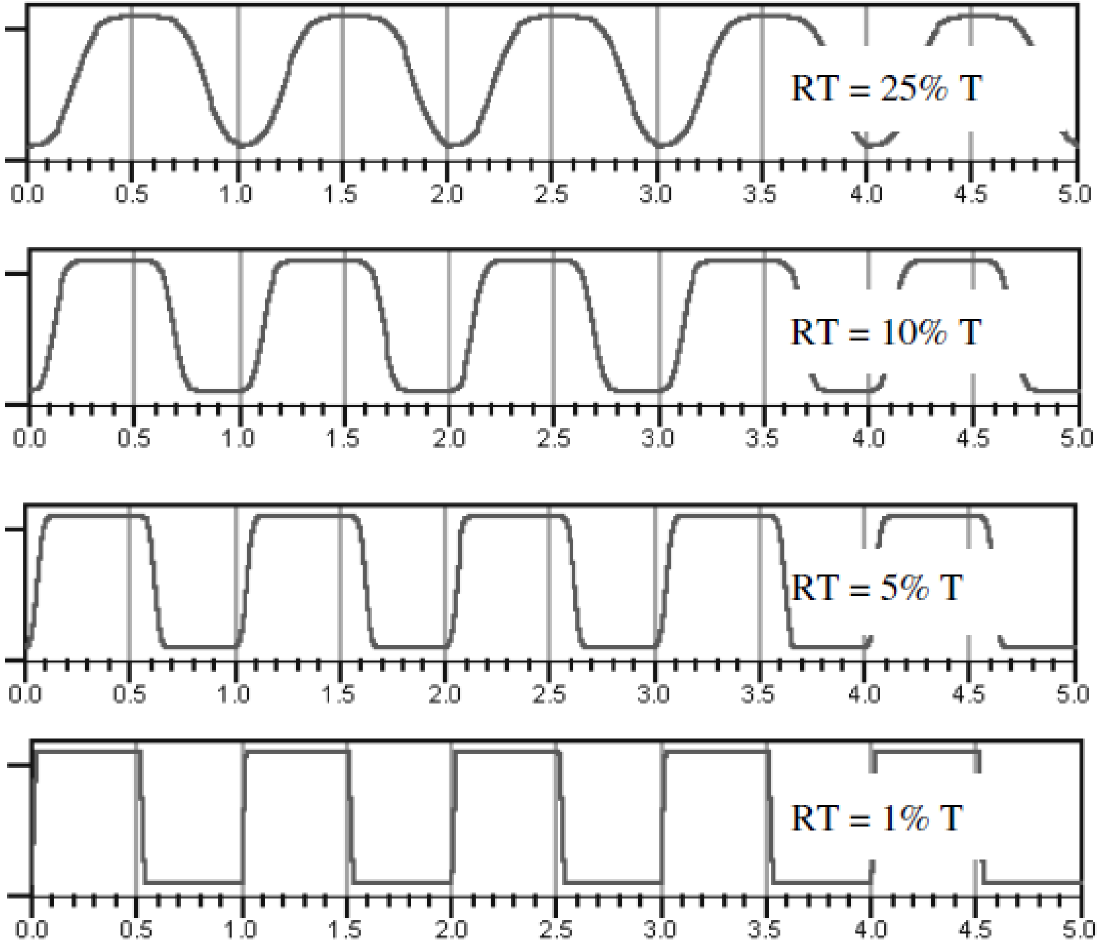
When we look at the actual spectrum of the trapezoid compared to the square wave, we see that the first and third harmonics are about the same for each. The trapezoid's fifth harmonic is about 70% of the square wave's, which is still a large fraction. However, the trapezoid's seventh harmonic is only about 30% of the ideal square wave's.

What Does *Significant* Mean?

The bandwidth of any waveform is always the highest sine-wave-frequency component in its spectrum that is comparable in magnitude to a corresponding ideal square wave. We can find out the bandwidth of any waveform by using a DFT to calculate its spectrum and compare it to an ideal square wave. We identify the frequency component of the waveform that is less than 70% of the ideal square wave, or we can use the rule of thumb developed earlier, that the BW is 0.35/rise time.

Bandwidth and Clock Frequency

As we have seen, bandwidth relates to the rise time of a signal. It is possible to have two different waveforms, with exactly the same clock frequency but different rise times and different bandwidths. Just knowing the clock frequency cannot tell us what the bandwidth is.



Four different waveforms, each with exactly the same 1-GHz clock frequency. Each of them has a different rise time, as a fraction of the period, and hence different bandwidths.

Bandwidth and Clock Frequency

Sometimes, we don't always know the rise time of a signal but need an idea of its bandwidth anyway. Using a simplifying assumption, we can estimate the bandwidth of a clock wave from just its clock frequency. Still, it is important to keep in mind that it is not the clock frequency that determines the bandwidth, it is the rise time. If all we know about the waveform is the clock frequency, we can't know the bandwidth for sure; we can only guess.

To evaluate the bandwidth of a signal from just its clock frequency, we have to make a very important assumption. We need to estimate what a typical rise time might be for a clock wave.

Bandwidth and Clock Frequency

How is the rise time related to the clock period in a real clock waveform? In principle, the only relationship is that the rise time must be less than 50% of the period. Other than this, there is no restriction, and the rise time can be any arbitrary fraction of the period. It could be 25% of the period, as in cases where the clock frequency is pushing the limits of the device technology, such as in 1-GHz clocks. It could be 10% of the period, which is typical of many microprocessor-based products. It could be 5% of the period, which is found in high-end FPGAs driving external low-clock-frequency memory buses. It could even be 1% if the board-level bus is a legacy system.

Bandwidth and Clock Frequency

If we don't know what fraction of the period the rise time is, a reasonable generalization is that the rise time is 7% of the clock period. This approximates many typical microprocessor-based boards and ASICs driving board-level buses. From this, we can estimate the bandwidth of the clock waveform.

It should be kept in mind that this assumption of the rise time being 7% of the period is a bit aggressive. Most systems are probably closer to 10%, so we are assuming a rise time slightly shorter than might typically be found. Likewise, if we are underestimating the rise time, we will be overestimating the bandwidth, which is safer than underestimating it.

Bandwidth and Clock Frequency

If the rise time is 7% of the period, then the period is 1/0.07 or 15 times the rise time. We have an approximation for the bandwidth as 0.35/rise time. We can relate the clock frequency to the clock period, because they are each the inverse of the other. Replacing the clock period for the clock frequency results in the final relationship; the bandwidth is five times the clock frequency:

$$BW_{clock} = 5 \times F_{clock}$$

where:

BW_{clock} = the approximate bandwidth of the clock, in GHz

F_{clock} = the clock repeat frequency, in GHz

Bandwidth and Clock Frequency

For example, if the clock frequency is 100 MHz, the bandwidth of the signal is about 500 MHz. If the clock frequency is 1 GHz, the bandwidth of the signal is about 5 GHz.

This is a generalization and an approximation, based on the assumption that the rise time is 7% of the clock period. Given this assumption, it is a very powerful rule of thumb, which can give an estimate of bandwidth with very little effort. It says that the highest sine-wave-frequency component in a clock wave is typically the fifth harmonic!

It's obvious, but bears repeating, that we always want to use the rise time to evaluate the bandwidth. Unfortunately, we do not always have the luxury of knowing the rise time for a waveform. And yet, we need an answer *now*!

TIP Sometimes getting an OK answer is often more important than getting a BETTER answer LATE.

Bandwidth of a Measurement

So far, we have been using the term *bandwidth* to refer to signals, or clock waveforms. We have said that the bandwidth is the highest significant sine-wave-frequency component in the waveform's spectrum. And, for signals, we said *significant* was based on comparing the amplitude of the signal's harmonic to the amplitude of an equivalent repeat frequency ideal square wave's.

The bandwidth of a measurement is the highest sine-wave-frequency component that has significant accuracy. When the measurement is done in the frequency domain, using an impedance analyzer or a network analyzer, the bandwidth of the measurement is very easy to determine. It is simply the highest sine-wave frequency in the measurement.

Bandwidth of a Model

TIP When we refer to the bandwidth of a model, we are referring to the highest sine-wave-frequency component where the model will accurately predict the actual behavior of the structure it is representing. There are a few tricks that can be used to determine this, but in general, only a comparison to a measurement will give a confident measure of a model's bandwidth.

The simplest starting equivalent circuit model to represent a wire bond is an inductor. Up to what bandwidth might this be a good model? The only way to really tell is to compare a measurement with the prediction of this model. Of course, it will be different for different wire bonds.

Bandwidth of a Model

As an example, we take the case of a very long wire bond, 300 mils long, connecting two pads over a return-path plane 10 mils below.

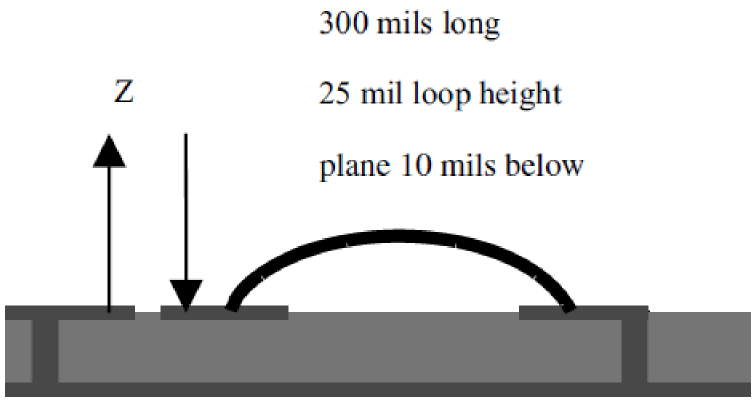
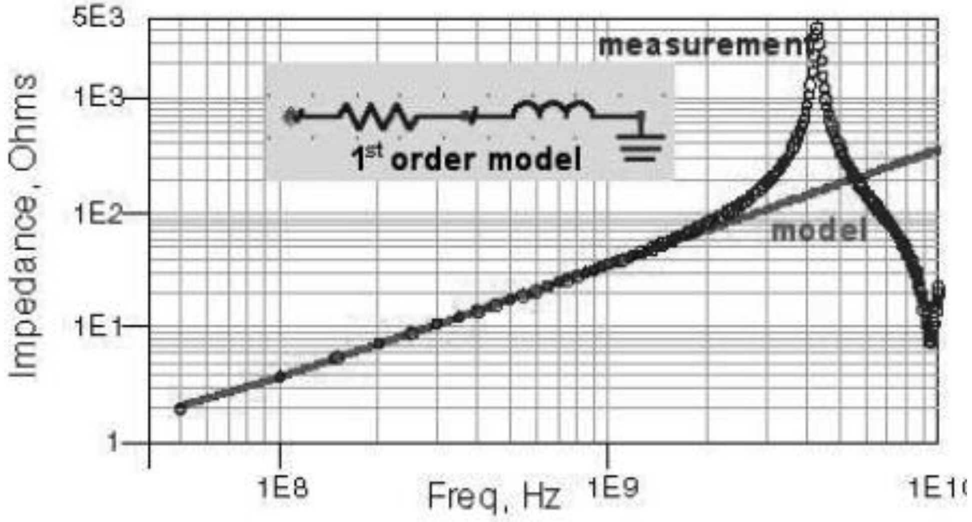
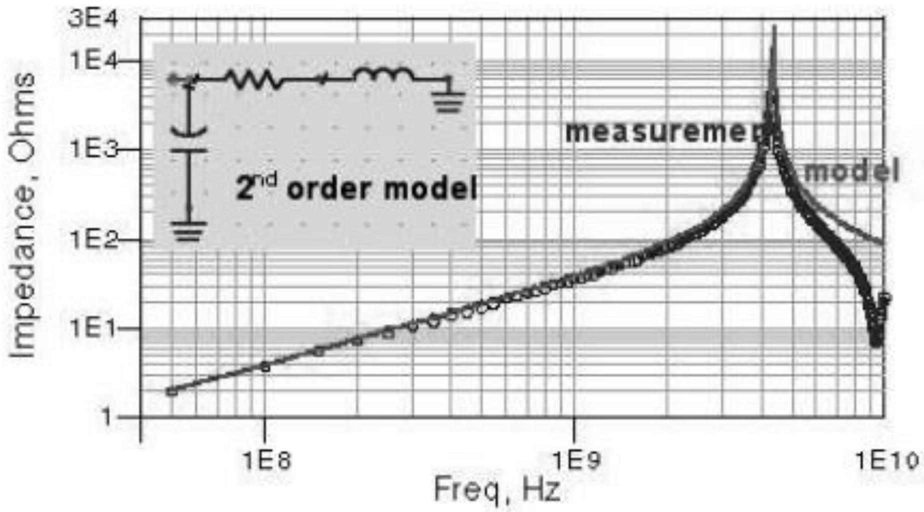


Diagram of a wire-bond loop between two pads, with a return path about 10 mils beneath the wire bond.

Bandwidth of a Model

A simple starting circuit model is a single ideal inductor and ideal resistor in series





Top: Comparison of the measured impedance and the simulation based on the first-order model. The agreement is good up to a bandwidth of about 2 GHz. Bottom: Comparison of the measured impedance and the simulation based on the second-order model. The agreement is good up to a bandwidth of about 4 GHz. The bandwidth of the measurement is 10 GHz, measured with a GigaTest Labs Probe Station.

Bandwidth of a Model

The best values for the L and R give a prediction for the impedance that closely matches the measured impedance up to 2 GHz. The bandwidth of this simple model is 2 GHz.

We could confidently use this simple model to predict performance of this physical structure in applications that had signal bandwidths of 2 GHz. It is surprising that for a wire bond this long, the simplest model, that of a constant ideal inductor and resistor, works so well up to 2 GHz. This is probably higher than the useful bandwidth of the wire bond, but the model is still accurate up to this high a frequency.

Bandwidth of a Model

Suppose we wanted a model with an even higher bandwidth that would predict the actual impedance of this real wire bond to higher frequency. We might add the effect of the pad capacitance. Building a new model, a second-order model, and finding the best values for the ideal R, L, and C elements result in a simulated impedance that matches the actual impedance to almost 4 GHz.

Bandwidth of an Interconnect

The bandwidth of an interconnect refers to the highest sine-wave-frequency component that can be transmitted by the interconnect without significant loss. What does *significant* mean? In some applications, a transmitted signal that is within 95% of the incident signal is considered too small to be useful. In other cases, a transmitted signal that is less than 10% of the incident signal is considered usable. In long-distance cable-TV systems, the receivers can use signals that have only 1% of the original power. Obviously, the notion of how much transmitted signal is significant is very dependent on the application and the particular specification. In reality, the bandwidth of an interconnect is the highest sine-wave frequency at which the interconnect still meets the performance specification for the application.

Bandwidth of an Interconnect

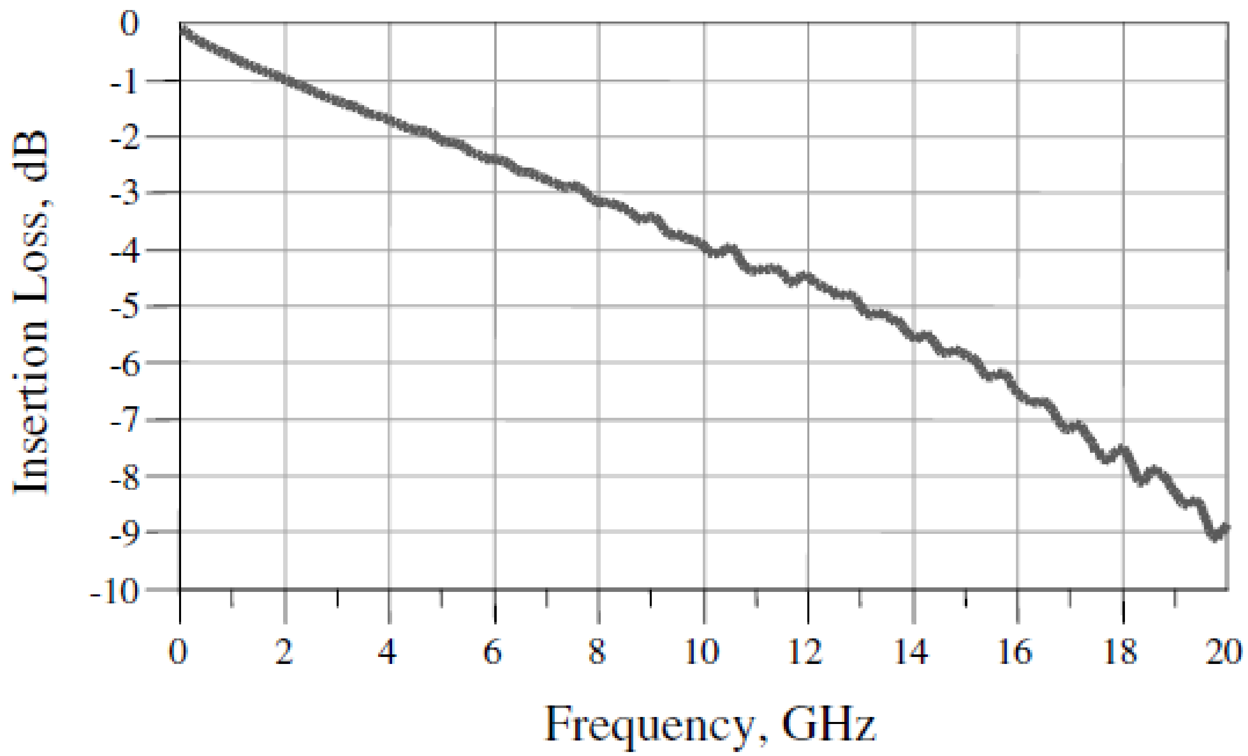
TIP In practice, *significant* means when the transmitted frequency-component amplitude is reduced by –3 dB, which means that its amplitude is reduced to 70% of the incident value. This is often referred to as the 3-dB bandwidth of an interconnect.

The bandwidth of an interconnect can be measured in either the time domain or the frequency domain. In general, we have to be careful interpreting the results if the source impedance is different than the characteristic impedance of the line, due to the complication of multiple reflections.

Measuring the bandwidth of an interconnect in the frequency domain is very straightforward. A network analyzer is used to generate sine waves of various frequencies. It injects the sine waves in the front of the interconnect and measures how much of each sine wave comes out at the far end. It is basically measuring the transfer function of the interconnect, and the interconnect is acting like a filter. This is also sometimes referred to as the *insertion loss* of the interconnect. The interpretation is simple when the interconnect is 50 Ohms, matched to the network analyzer's impedance.

Bandwidth of an Interconnect

For example, the measured transmitted amplitude of sine waves through a 4-inch length of a 50-Ohm transmission line in FR4 is shown below.



Measured transmitted amplitude of different sine-wave signals through a 4-inch-long transmission line made in FR4. The 3 dB bandwidth is seen to be about 8 GHz for this cross section and material properties. Measured with a GigaTest Labs Probe Station.

Bandwidth of an Interconnect

The measurement bandwidth is 20 GHz in this case. The 3-dB bandwidth of the interconnect is seen to be about 8 GHz. This means that if we send in a sine wave at 8 GHz, at least 70% of the amplitude of the 8-GHz sine wave would appear at the far end. More than likely, if the interconnect bandwidth were 8 GHz, nearly 100% of a 1-GHz sine wave would be transmitted to the far end of the same interconnect.

The interpretation of the bandwidth of an interconnect is the approximation that if an ideal square wave were transmitted through this interconnect, each sine-wave component would be transmitted, with those components lower than 8 GHz having roughly the same amplitude coming out as they did going in. But the amplitude of those components above 8 GHz would be reduced to insignificance.

Bandwidth of an Interconnect

A signal that might have a rise time of 1 psec going into the interconnect would have a rise time of 0.35/8 GHz = 0.043 nsec or 43 psec when it came out. The interconnect will degrade the rise time.

TIP The bandwidth of the interconnect is a direct measure of the minimum rise-time signal an interconnect can transmit.

Bandwidth of an Interconnect

If the bandwidth of an interconnect is 1 GHz, the fastest edge it can transmit is 350 psec. This is sometimes referred to as its intrinsic rise time. If a signal with a 350-psec edge enters the interconnect, what will be the rise time coming out? This is a subtle question. The rise time exiting the interconnect can be approximated by:

$$RT^2_{out} = RT^2_{in} + RT^2_{interconnect}$$

where:

RT_{out} = the 10–90 rise time of the output signal

RT_{in} = the 10–90 rise time of the input signal

$RT_{interconnect}$ = the intrinsic 10–90 rise time of the interconnect

Bandwidth of an Interconnect

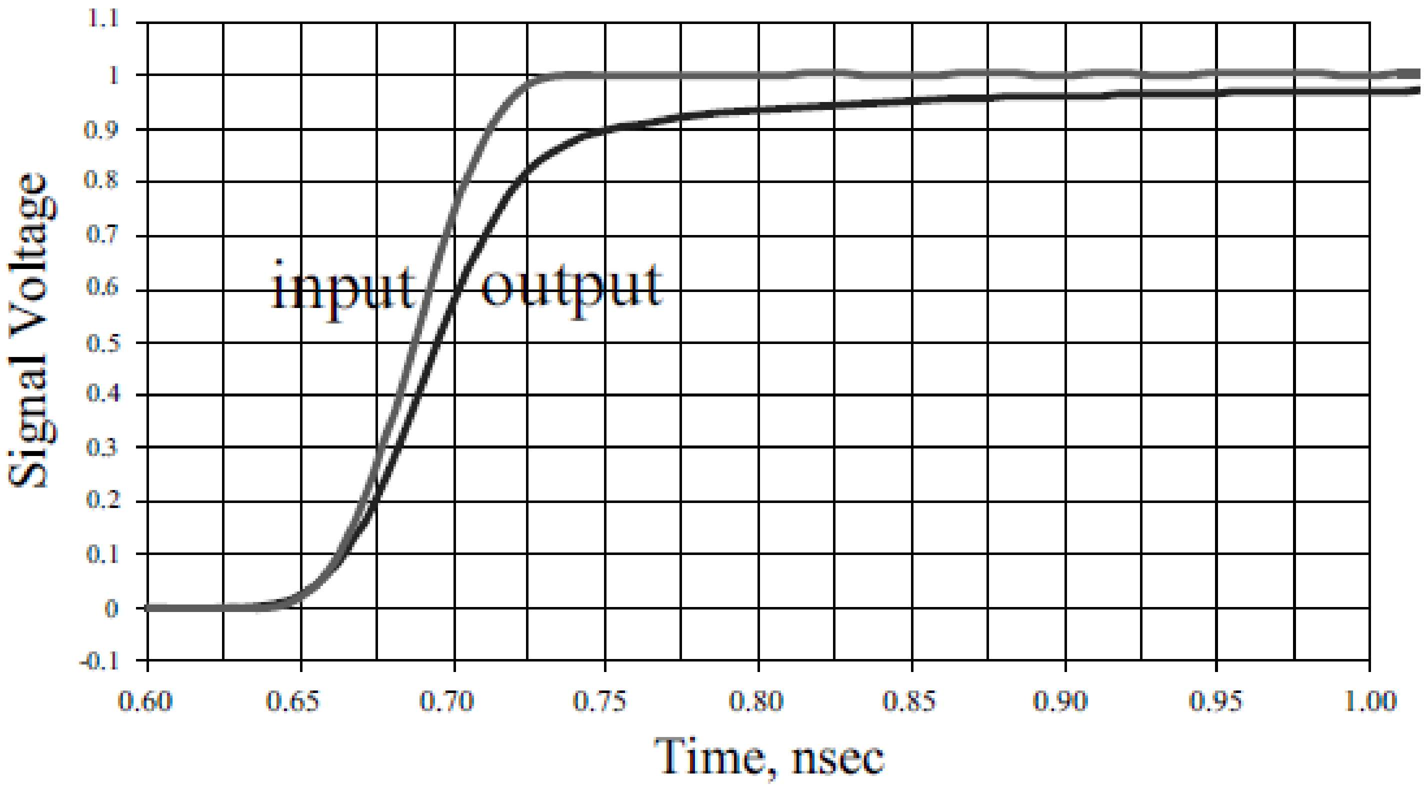
This assumes that both the incident spectra and the response of the interconnect correspond to a Gaussian-shaped rise time.

For example, in the case of this 4-inch-long interconnect, if a signal with a rise time of 50 psec were input, the rise time of the transmitted signal would be:

$$\text{sqrt}(50\ \text{psec}^2 + 43\ \text{psec}^2) = 67\ \text{psec}.$$

This is an increase of about 17 psec in the rise time of the transmitted waveform compared to the incident rise time.

Bandwidth of an Interconnect



Measured input and transmitted signal through a 4-inch long, 50-Ohm transmission line in FR4 showing the rise-time degradation. The input rise time is 50 psec. The predicted output rise time is 67 psec based on the measured bandwidth of the interconnect. Measured with a GigaTest Labs Probe Station.

Bandwidth of an Interconnect

TIP As a simple rule of thumb, in order for the rise time of the signal to be increased by the interconnect less than 10%, the intrinsic rise time of the interconnect should be shorter than 50% of the rise time of the signal.

TIP In the frequency-domain perspective, to support the transmission of a 1-GHz bandwidth signal, we want the bandwidth of the interconnect to be at least twice as high, or 2 GHz.

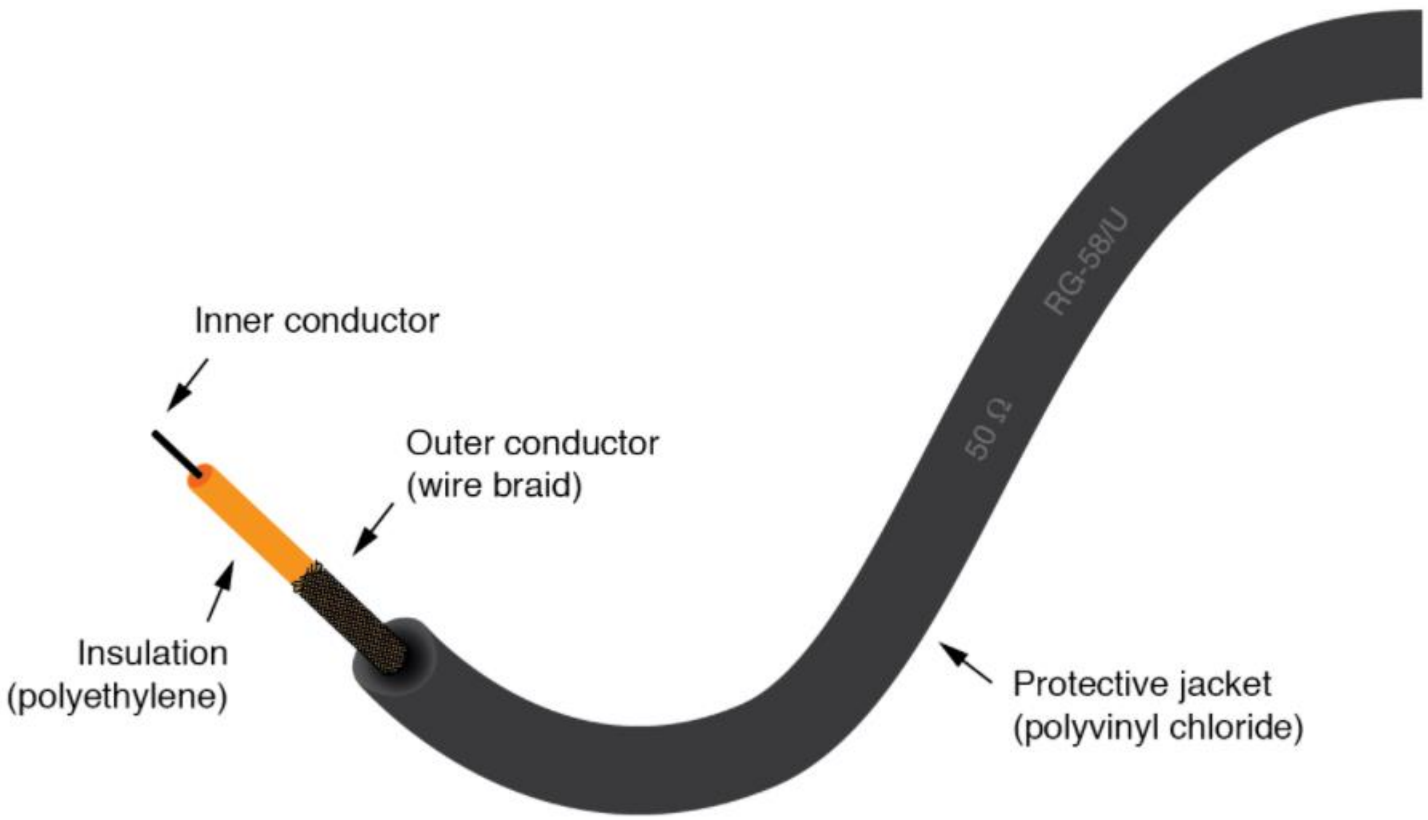
It is important to keep in mind that this is a rule of thumb and it should not be used for design sign-off. It should be used only for a rough estimate or to identify a goal. If the bandwidth of an interconnect is within a factor of two of the bandwidth of the signal, it would probably be important to perform an analysis of how the interconnect affected the entire signal's spectrum.

SIGNAL INTEGRITY & POWER INTEGRITY (SIPI)

Module 4: Transmission Lines

A 50 Ohm Cable?

- The figure below shows a coaxial cable with a label of “50 ohms” printed along its outer sheath:



Coaxial cable construction.

A 50 Ohm Cable?

- Coaxial cable is a two-conductor cable made of
  - A single conductor surrounded by a braided wire jacket
  - A plastic insulating material separating the two
- The outer (braided) conductor completely surrounds the inner (single wire) conductor
  - The two conductors are insulated from each other for the entire length of the cable

A 50 Ohm Cable?

- Commonly used to conduct weak (low-amplitude) voltage signals
  - Excellent ability to shield signals from external interference

- How could two conductors, insulated from each other by a relatively thick layer of plastic, have 50 ohms of resistance between them?

A 50 Ohm Cable?

Measuring the resistance between the outer and inner conductors with my ohmmeter, I found it to be infinite (open-circuit), just as I would have expected from the two insulated conductors.

Measuring each of the two conductors' resistances from one end of the cable to the other indicated nearly zero ohms of resistance: again, exactly what I would have expected from continuous, unbroken lengths of wire.

Nowhere was I able to measure 50 Ω of resistance on this cable, regardless of which points I connected my ohmmeter in between.

A 50 Ohm Cable?

What I didn't understand at that time was the cable's response to high-frequency AC signals and pulses that exhibit fast rise/fall time. Continuous direct current (DC)—such as that used by my ohmmeter to check the cable's resistance—shows the two conductors to be completely insulated from each other, with nearly infinite resistance between the two.

However, due to the effects of capacitance and inductance distributed along the length of the cable, the cable's response to rapidly-changing voltages is such that it acts as a *finite* impedance, drawing current proportional to the applied voltage.

A 50 Ohm Cable?

What we would normally dismiss as being just a pair of wires becomes an important circuit element in the presence of rapidly-changing transients and high-frequency AC signals, with characteristic properties all its own. When expressing such properties, we refer to the wire pair as a *transmission line*.

Many transmission line effects do not appear in significant measure in AC circuits of powerline frequency (50 or 60 Hz), or in continuous DC circuits.

A 50 Ohm Cable?

However, in circuits involving high frequencies and/or extremely long cable lengths, the effects are very significant.

Practical applications of transmission line effects abound in radio-frequency ("RF") communication circuitry, including computer networks, and in low-frequency circuits subject to rapidly-changing voltage transients ("surges") such as lightning strikes on power lines.

Propagation Delay

Circuits and the Speed of Light

Suppose we had a simple one-battery, one-lamp circuit controlled by a switch. When the switch is closed, the lamp immediately lights. When the switch is opened, the lamp immediately darkens: (Figure below)

Lamp appears to immediately respond to switch.

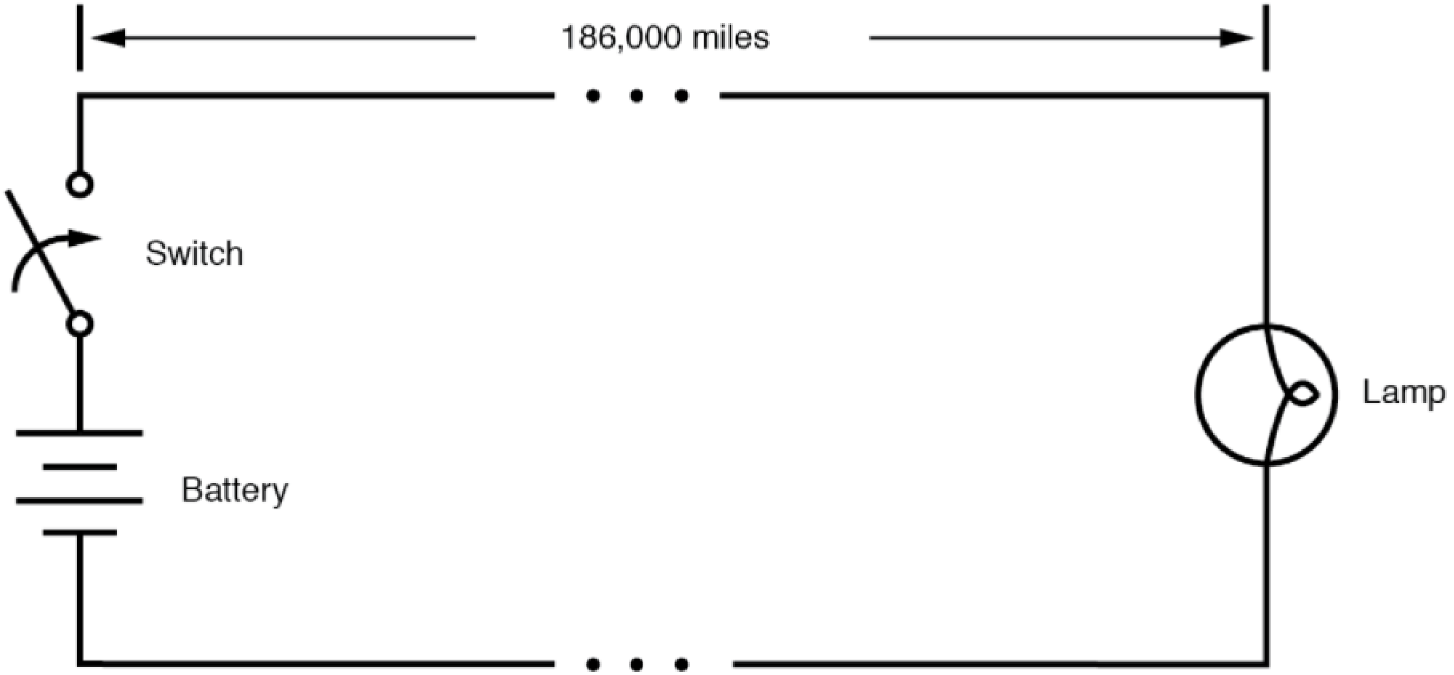
Circuits and the Speed of Light

Actually, an incandescent lamp takes a short time for its filament to warm up and emit light after receiving an electric current of sufficient magnitude to power it, so the effect is not instant. However, what I'd like to focus on is the immediacy of the electric current itself, not the response time of the lamp filament.

For all practical purposes, the effect of switch action is instant at the lamp's location. Although electric charge carriers move through wires very slowly, the overall effect of electric charge carriers pushing against each other happens at the speed of light (approximately 186,000 miles per *second*!).

Circuits and the Speed of Light

What would happen, though, if the wires carrying power to the lamp were 186,000 miles long? Since we know an electric signal has a finite speed (albeit very fast), a set of very long wires should introduce a time delay into the circuit, delaying the switch's action on the lamp: (Figure below)



At the speed of light, lamp responds after 1 second.

Circuits and the Speed of Light

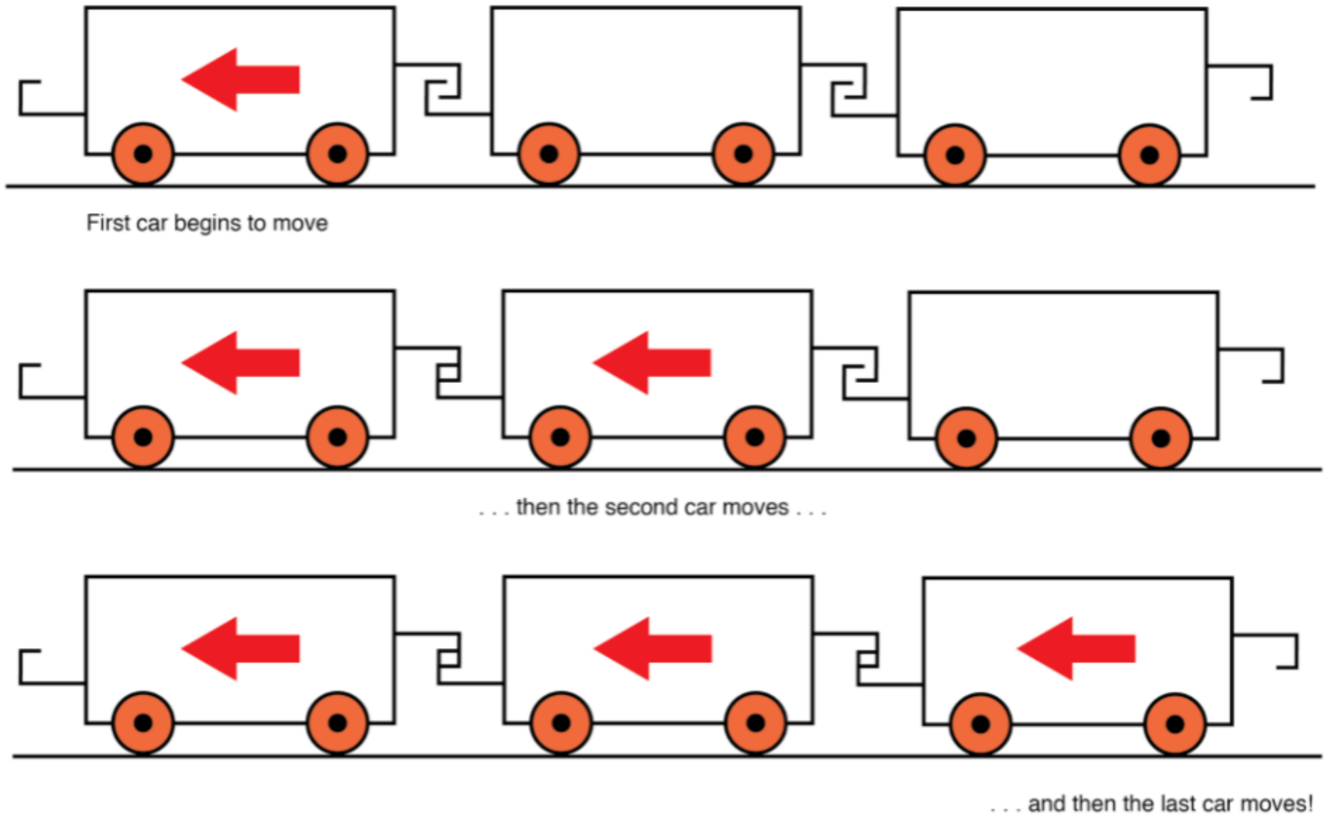
Assuming no warm-up time for the lamp filament, and no resistance along the 372,000 mile length of both wires, the lamp would light up approximately one second after the switch closure.

Although the construction and operation of superconducting wires 372,000 miles in length would pose enormous practical problems, it is theoretically possible, and so this "thought experiment" is valid. When the switch is opened again, the lamp will continue to receive power for one second of time after the switch opens, then it will de-energize.

One way of envisioning this is to imagine the electric charge carriers within a conductor as rail cars in a train: linked together with a small amount of "slack" or "play" in the couplings. When one rail car (an electric charge carrier) begins to move, it pushes on the one ahead of it and pulls on the one behind it, but not before the slack is relieved from the couplings.

Circuits and the Speed of Light

Thus, motion is transferred from car to car (from one electric charge carrier to another) at a maximum velocity limited by the coupling slack, resulting in a much faster transfer of motion from the left end of the train (circuit) to the right end than the actual speed of the cars (electric charge carriers): (Figure below)

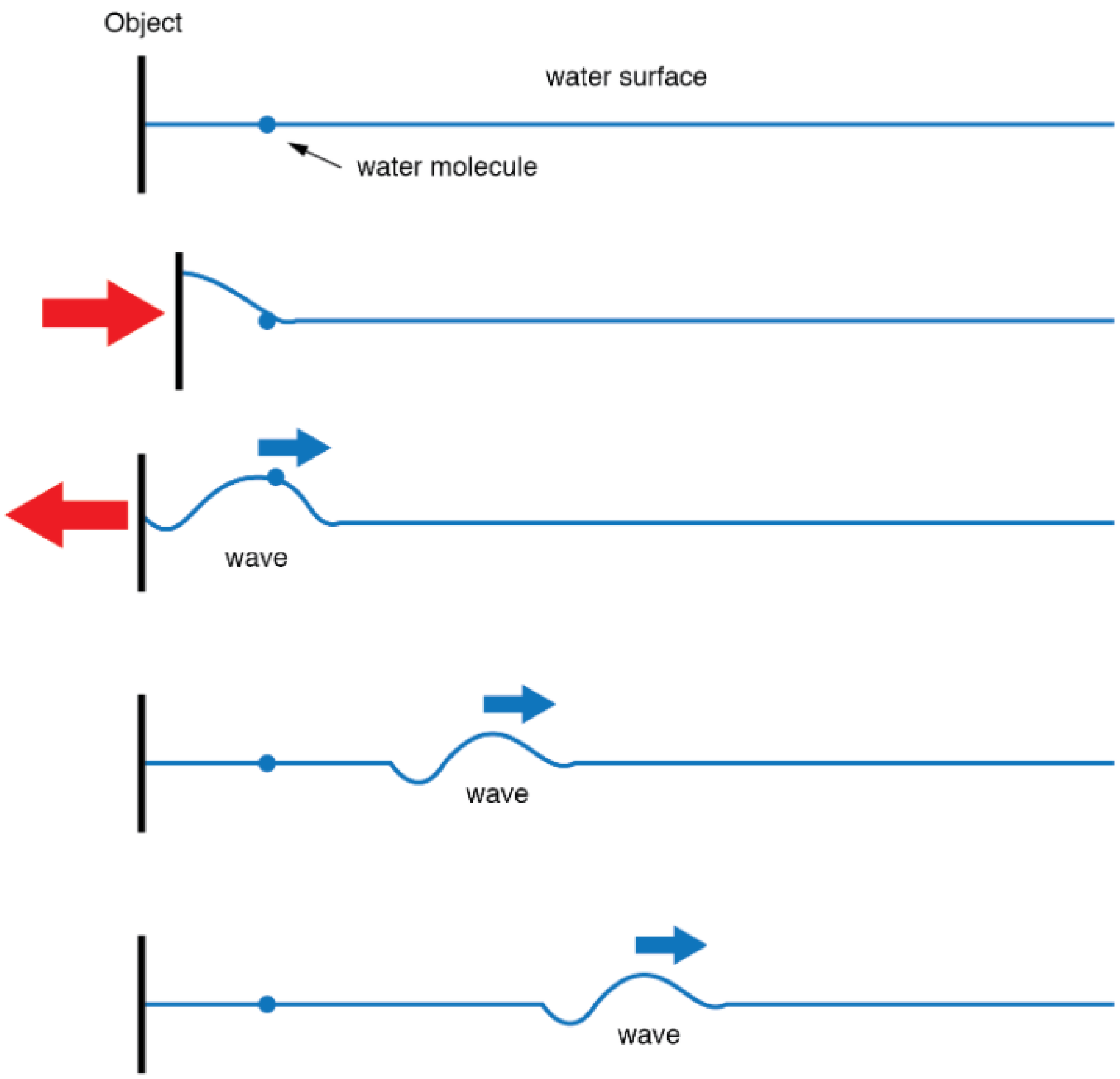


Motion is transmitted successively from one car to next.

Circuits and the Speed of Light

Another analogy, perhaps more fitting for the subject of transmission lines, is that of waves in water. Suppose a flat, wall-shaped object is suddenly moved horizontally along the surface of the water, so as to produce a wave ahead of it.

Circuits and the Speed of Light



Wave motion in water.

Circuits and the Speed of Light

The wave will travel as water molecules bump into each other, transferring wave motion along the water's surface far faster than the water molecules themselves are actually traveling: (Figure below)

Likewise, the electric charge carriers motion "coupling" travels approximately at the speed of light, although the electric charge carriers themselves don't move that quickly. In a very long circuit, this "coupling" speed would become noticeable to a human observer in the form of a short time delay between switch action and lamp action.

Circuits and the Speed of Light

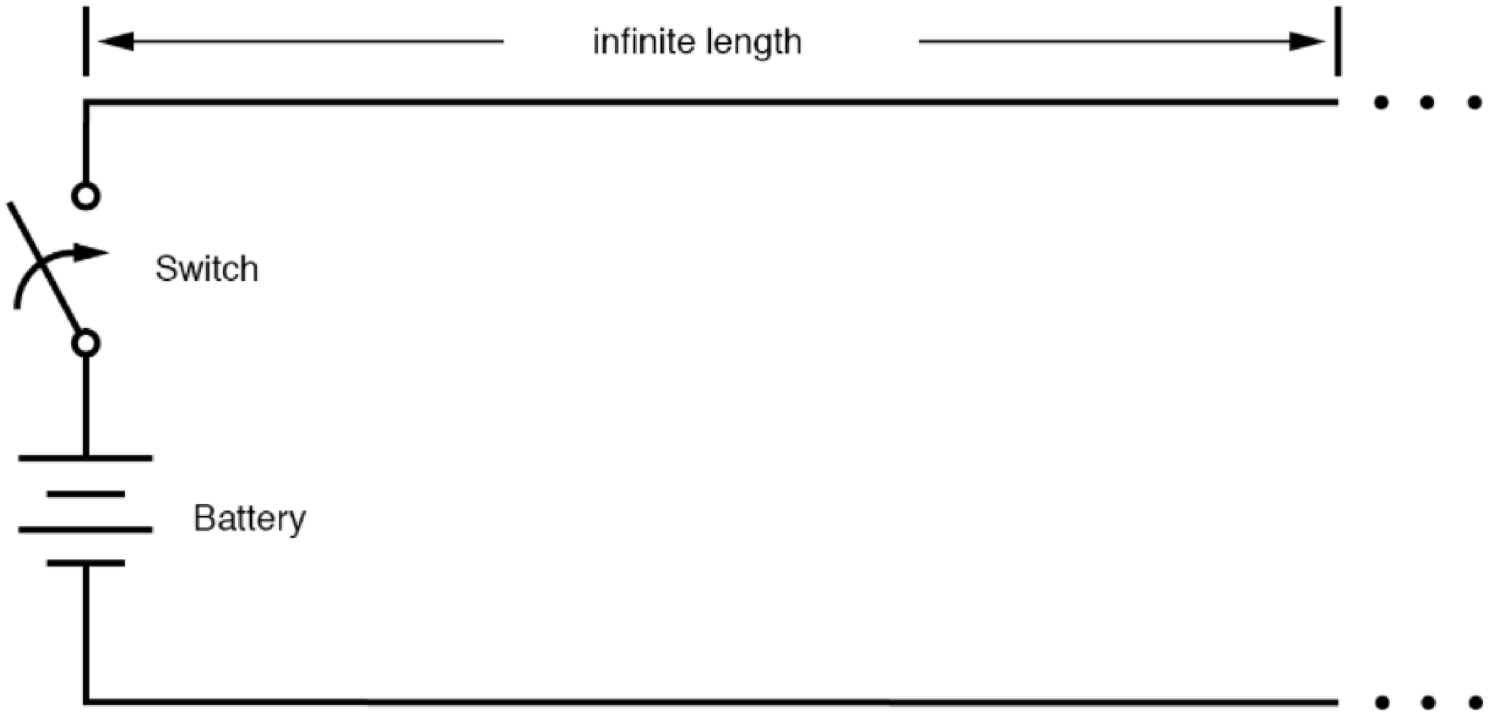
REVIEW:

- In an electric circuit, the electric charge carriers motion “coupling” travels approximately at the speed of light, although the electric charge carriers within the conductors do not travel anywhere near that velocity.

Ideal Transmission Lines

Parallel Wires of Infinite Length

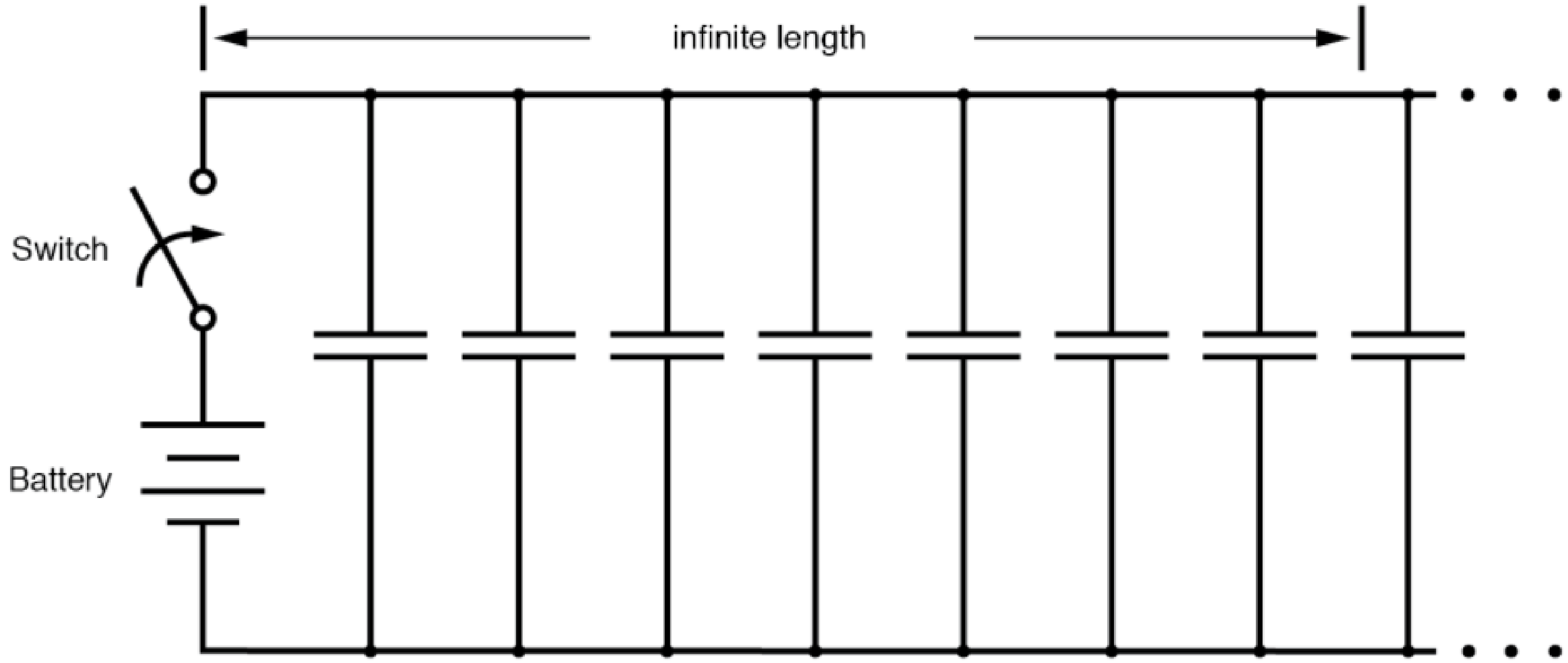
Suppose, though, that we had a set of parallel wires of *infinite* length, with no lamp at the end. What would happen when we close the switch? Being that there is no longer a load at the end of the wires, this circuit is open. Would there be no current at all? (Figure below)



Driving an infinite transmission line.

Parallel Wires of Infinite Length

Despite being able to avoid wire resistance through the use of superconductors in this "thought experiment," we cannot eliminate capacitance along the wires' lengths. *Any* pair of conductors separated by an insulating medium creates capacitance between those conductors: (Figure below)



Equivalent circuit showing stray capacitance between conductors.

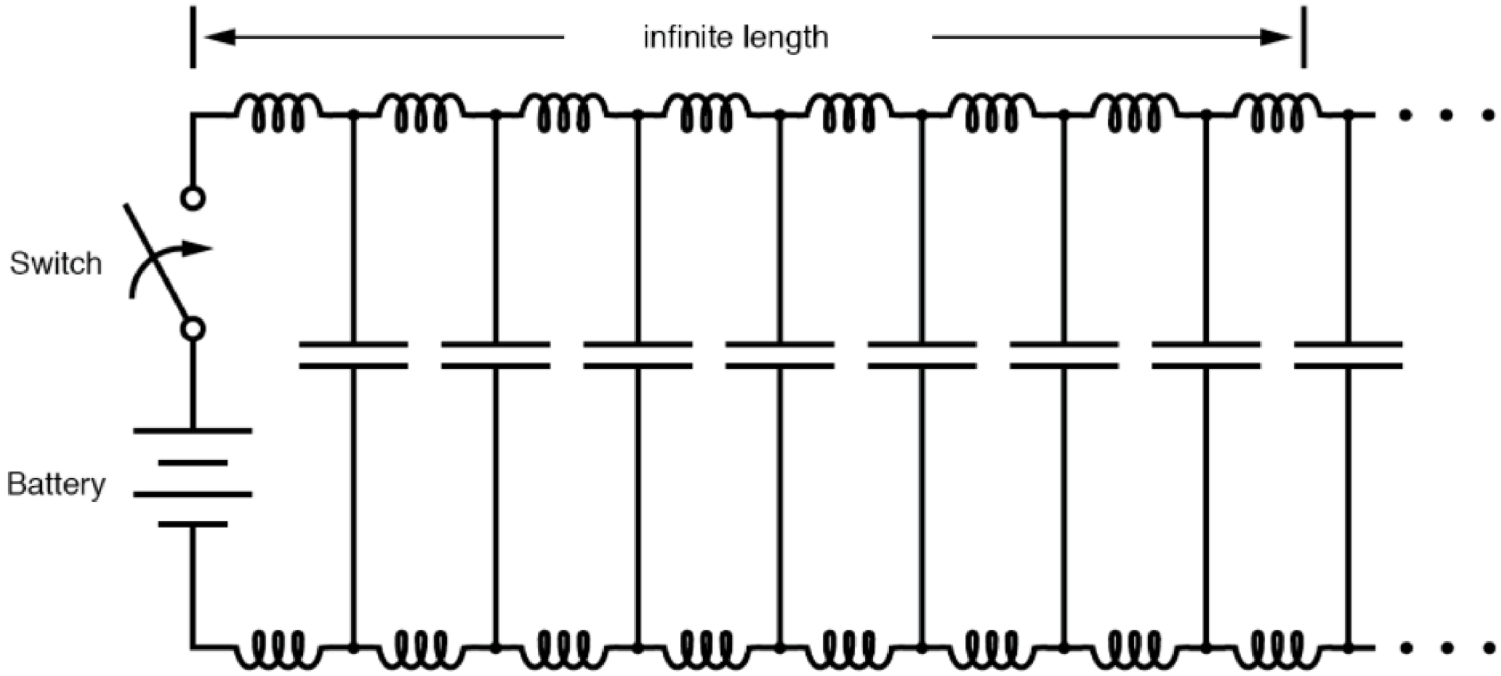
Parallel Wires of Infinite Length

Voltage applied between two conductors creates an electric field between those conductors. Energy is stored in this electric field, and this storage of energy results in an opposition to change in voltage. The reaction of a capacitance against changes in voltage is described by the equation $i = C(de/dt)$, which tells us that current will be drawn proportional to the voltage's rate of change over time. Thus, when the switch is closed, the capacitance between the conductors will react against the sudden voltage increase by charging up and drawing current from the source. According to the equation, an instant rise in applied voltage (as produced by perfect switch closure) gives rise to an infinite charging current.

Capacitance and Inductance

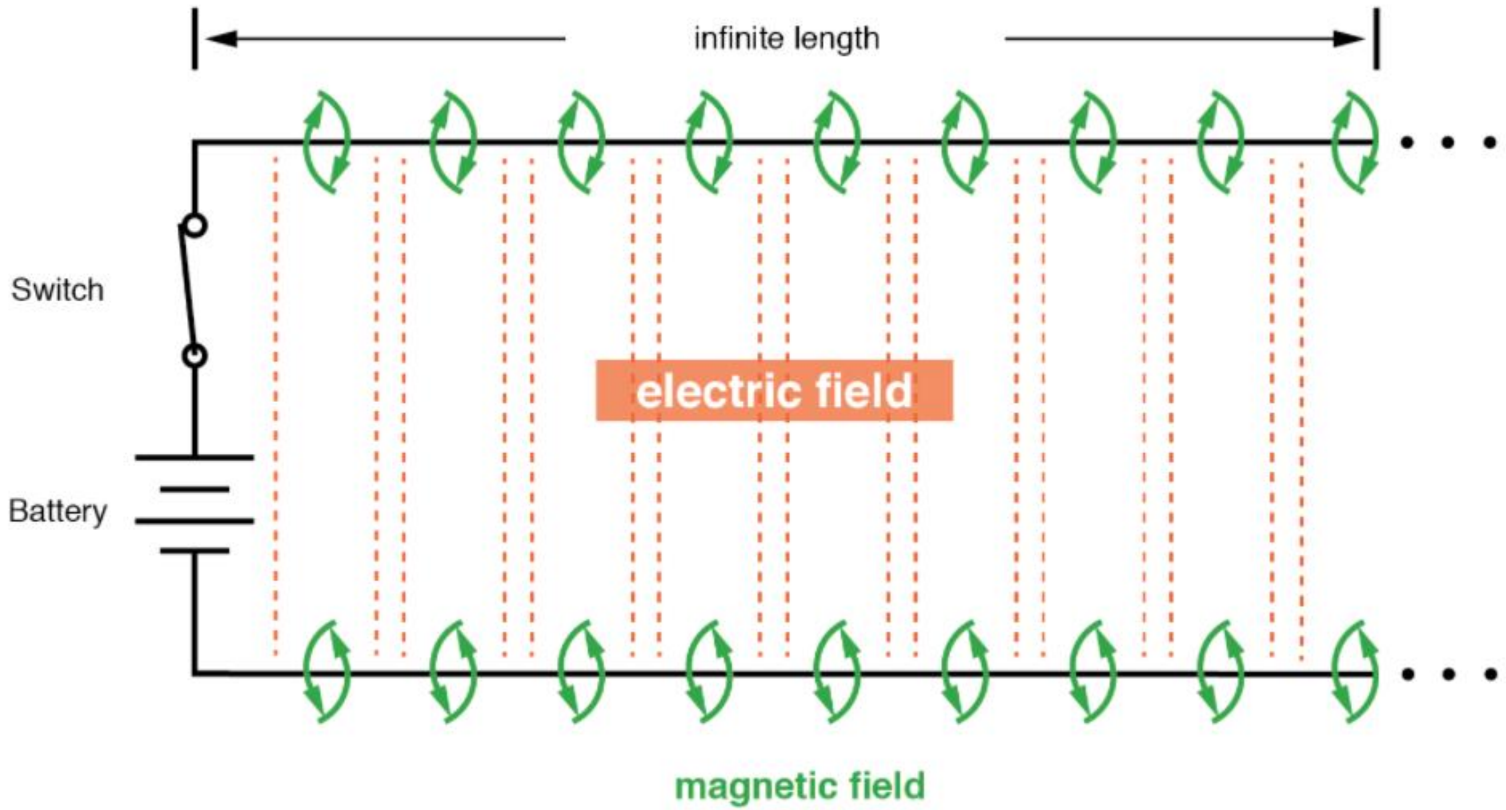
However, the current drawn by a pair of parallel wires will not be infinite, because there exists a series of impedance along the wires due to inductance. (Figure below) Remember that current through *any* conductor develops a magnetic field of proportional magnitude. Energy is stored in this magnetic field, (Figure below) and this storage of energy results in an opposition to change in current. Each wire develops a magnetic field as it carries charging current for the capacitance between the wires, and in so doing drops voltage according to the inductance equation $e = L(di/dt)$. This voltage drop limits the voltage rate-of-change across the distributed capacitance, preventing the current from ever reaching an infinite magnitude:

Capacitance and Inductance



Equivalent circuit showing stray capacitance and inductance.

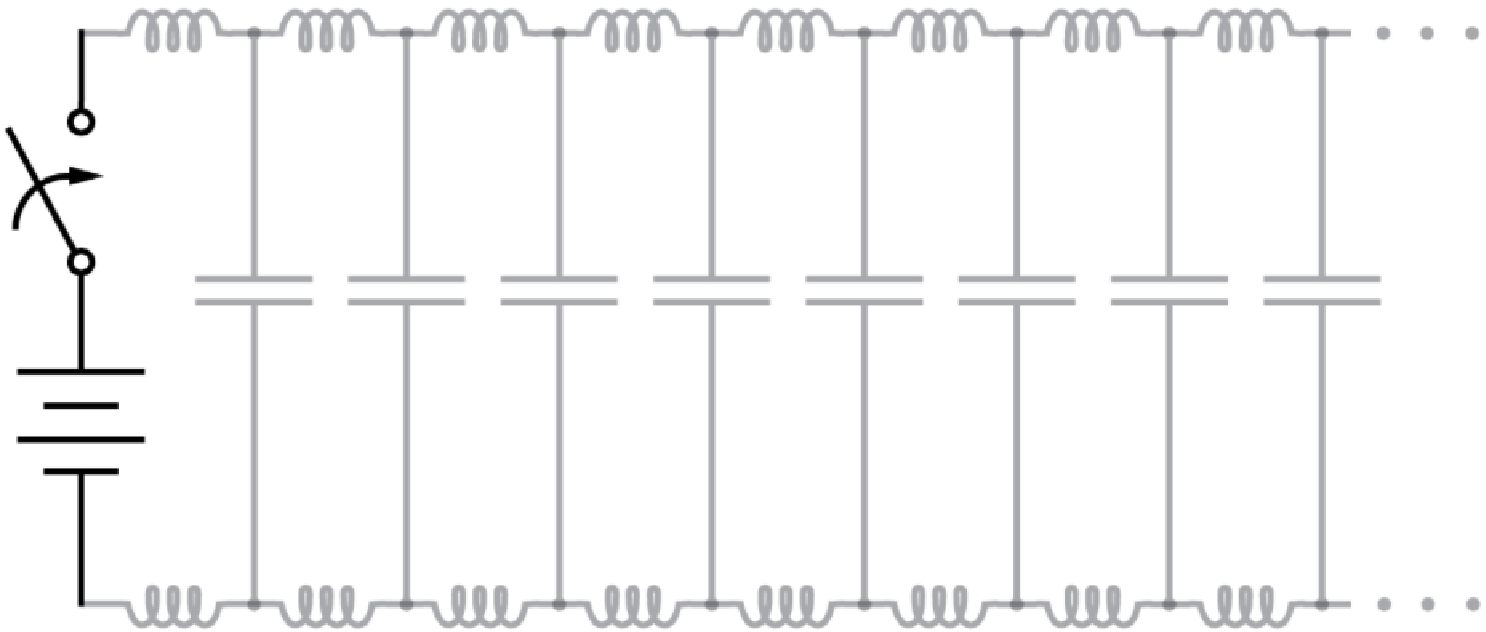
Capacitance and Inductance



Voltage charges capacitance, current charges inductance.

Capacitance and Inductance

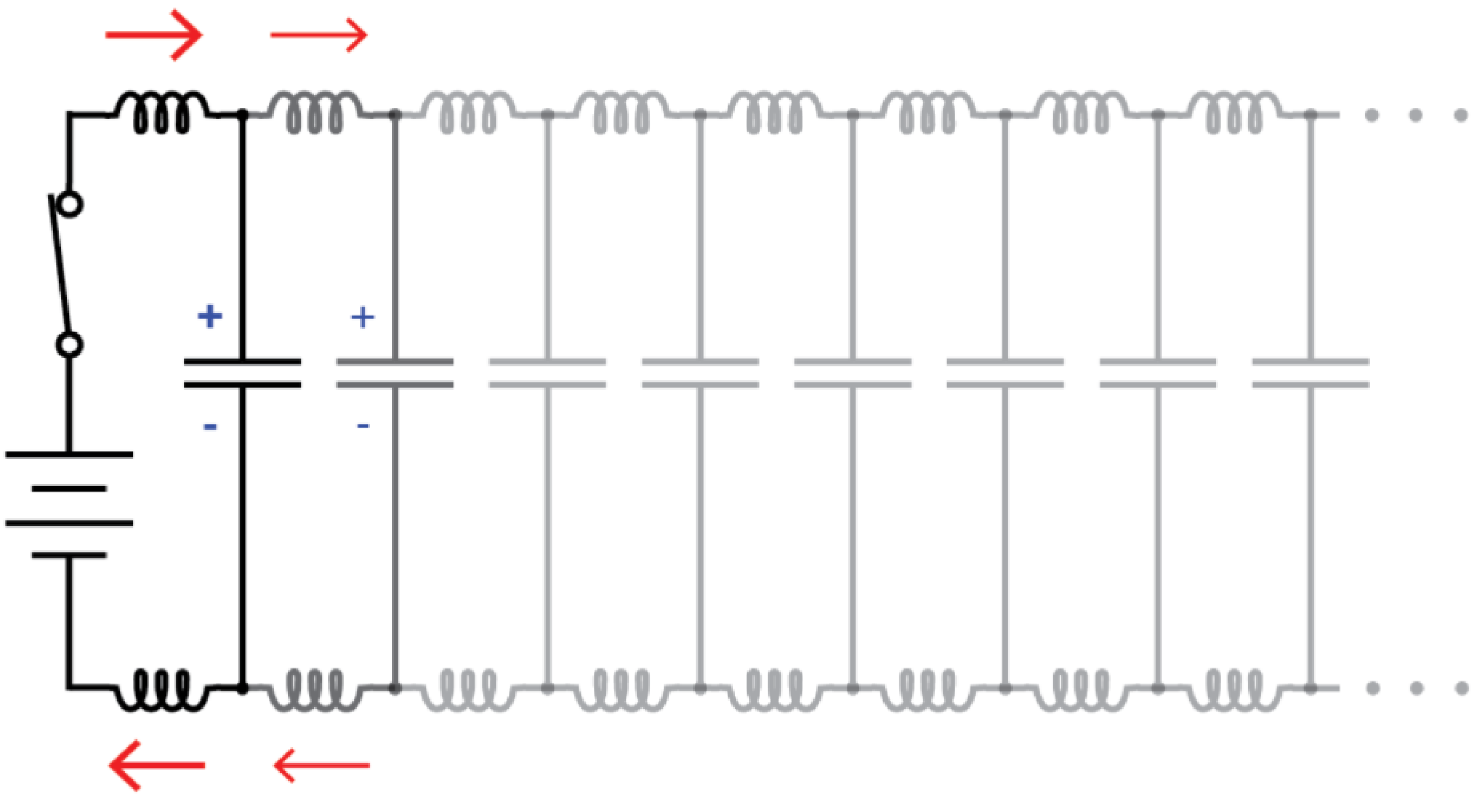
Because the electric charge carriers in the two wires transfer motion to and from each other at nearly the speed of light, the "wave front" of voltage and current change will propagate down the length of the wires at that same velocity, resulting in the distributed capacitance and inductance progressively charging to full voltage and current, respectively, like this:



Uncharged transmission line.

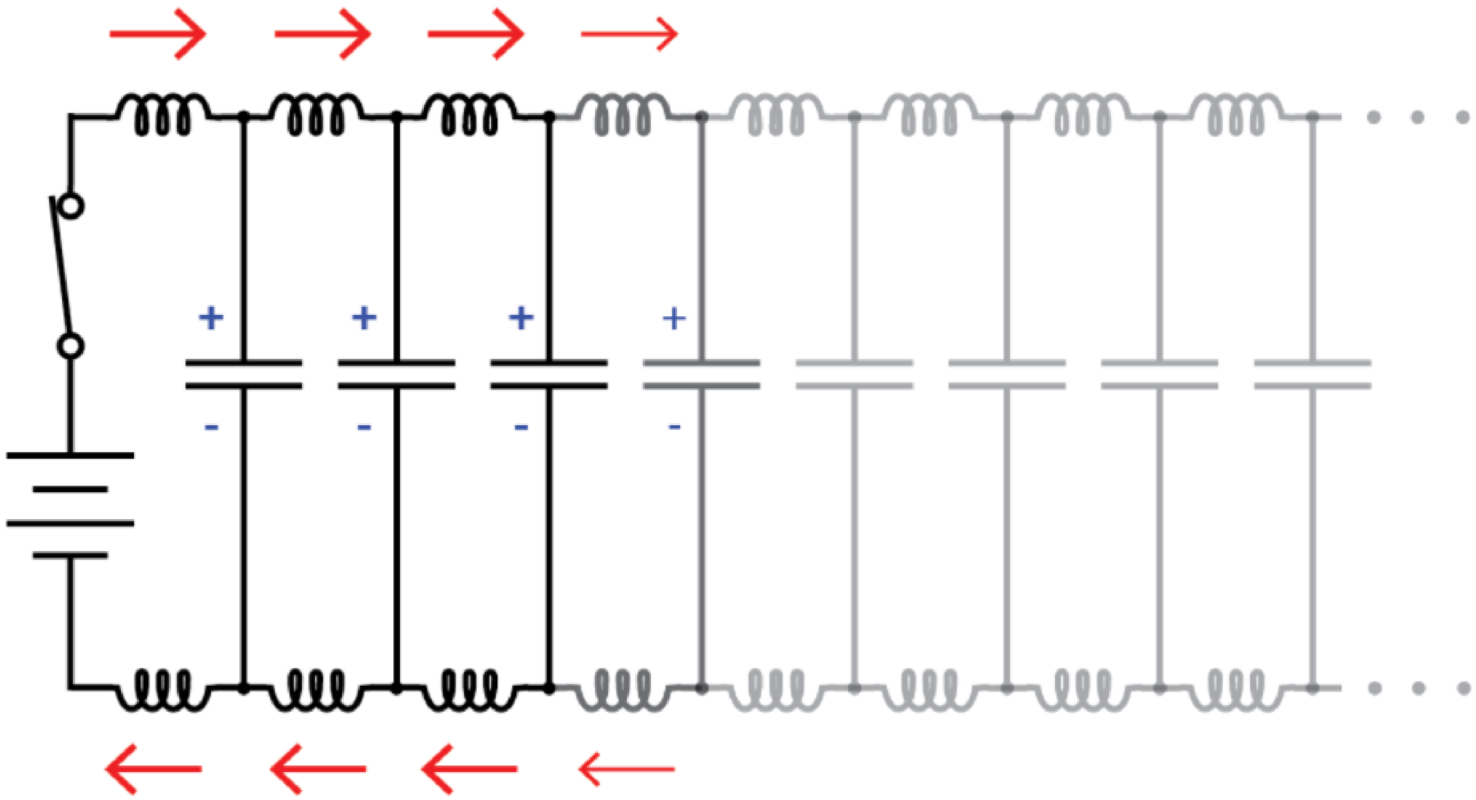
Capacitance and Inductance

Switch closes!



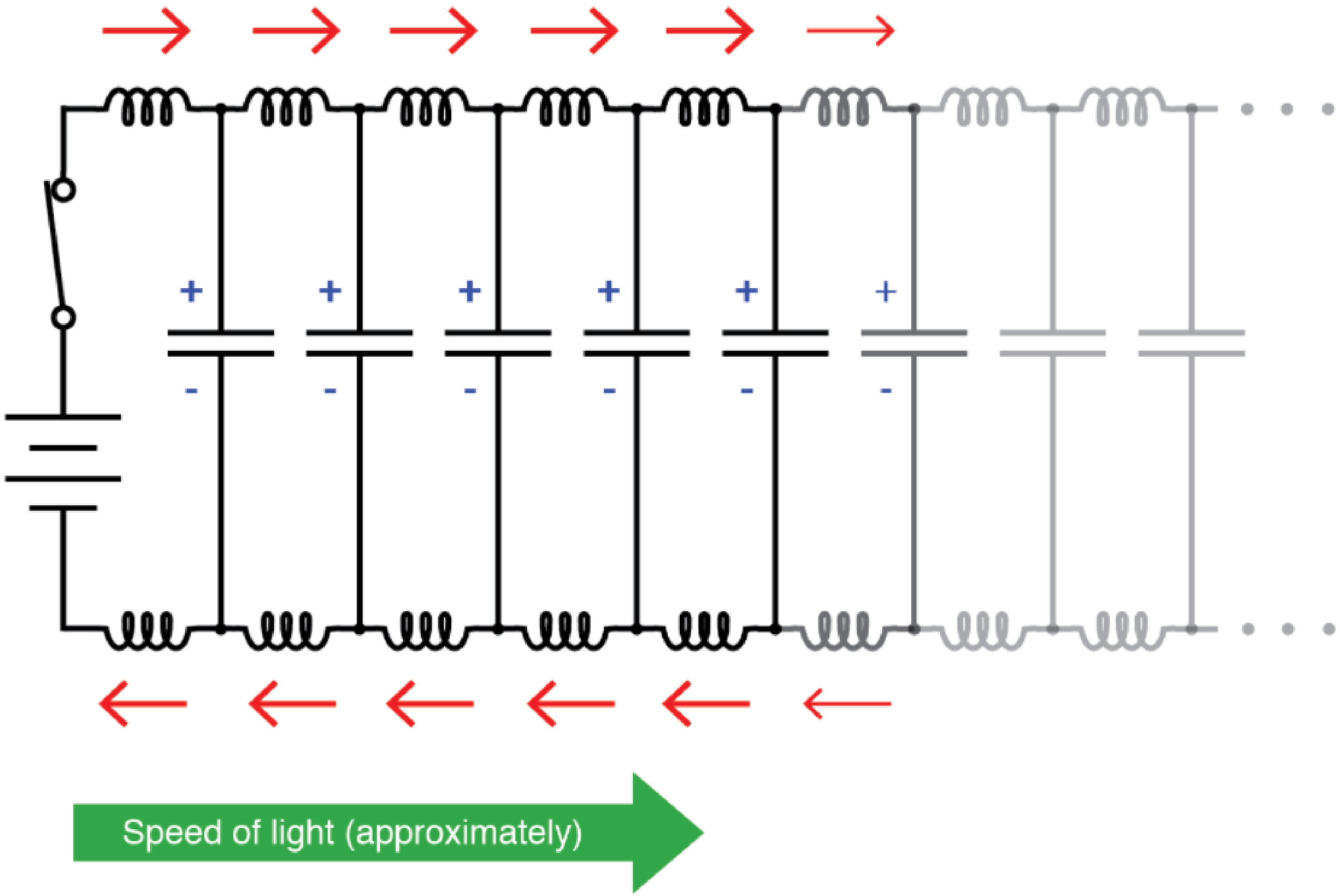
Begin wave propagation.

Capacitance and Inductance



Continue wave propagation.

Capacitance and Inductance



Propagate at speed of light.

The Transmission Line

The end result of these interactions is a constant current of limited magnitude through the battery source. Since the wires are infinitely long, their distributed capacitance will never fully charge to the source voltage, and their distributed inductance will never allow unlimited charging current. In other words, this pair of wires will draw current from the source so long as the switch is closed, behaving as a constant load. No longer are the wires merely conductors of electrical current and carriers of voltage, but now constitute a circuit component in themselves, with unique characteristics. No longer are the two wires merely *a pair of conductors*, but rather a *transmission line*.

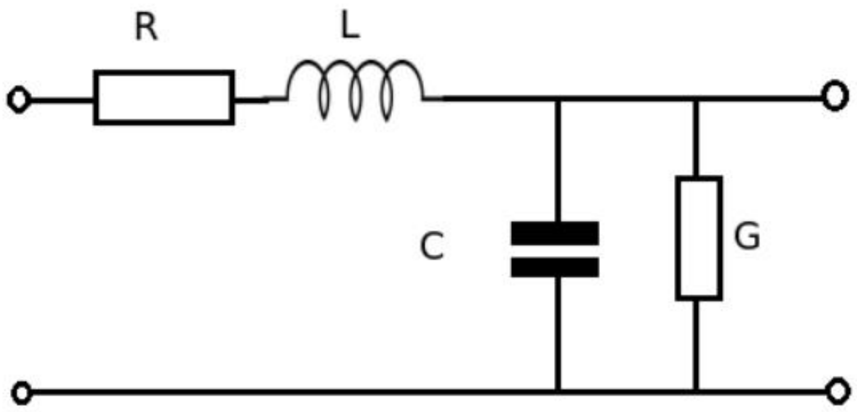
Characteristic Impedance

Characteristic Impedance

As a constant load, the transmission line's response to the applied voltage is resistive rather than reactive, despite being comprised purely of inductance and capacitance (assuming superconducting wires with zero resistance). We can say this because there is no difference from the battery's perspective between a resistor eternally dissipating energy and an infinite transmission line eternally absorbing energy. The impedance (resistance) of this line in ohms is called the *characteristic impedance*, and it is fixed by the geometry of the two conductors. For a parallel-wire line with air insulation, the characteristic impedance may be calculated as such:

Why the Characteristic Impedance is Independent of Signal Frequency

$$Z_o = \sqrt{\frac{R + jwL}{G + jwC}}$$

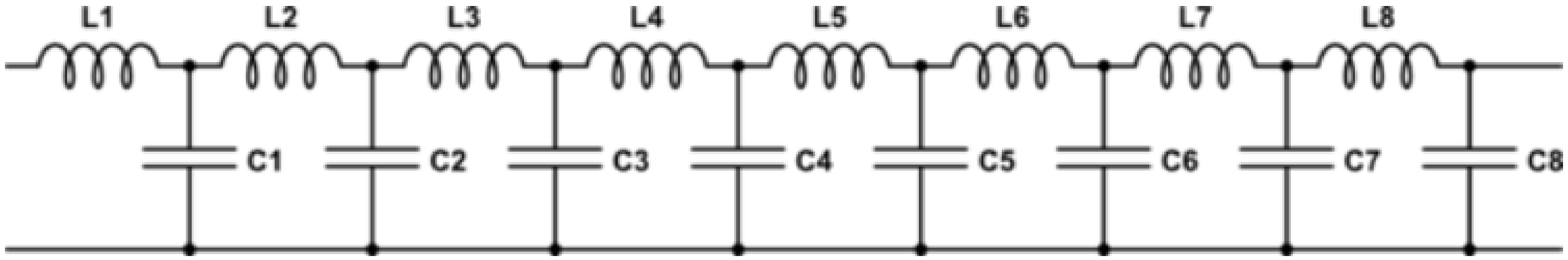


Equivalent circuit per unit length

R is resistance of wire/track
L is the inductance
C is the capacitance
G is the conductance of the dielectric

As R<<jωL and G→0, these values can be ignored and so the expression reduces to $\sqrt{L/C}$ (frequency independent).

Why the Characteristic Impedance is Independent of Cable Length



- Assume that existing line consists of n inductors and n capacitors

- Assume uniform cable with each inductance = L and each capacitance = C
  - Total inductance = nL and total capacitance = nC

- If line becomes longer, number of inductors and capacitors become n + m
  - Total inductance = (n + m)L and total capacitance = (n + m)C

- For both cases, characteristic impedance, $Z_o = \sqrt{\frac{L}{C}}$

Characteristic Impedance

$$Z_0 = \frac{276}{\sqrt{k}} \log \frac{d}{r}$$

Where,

Z_0 = Characteristic impedance of line
d = Distance between conductor centers
r = Conductor radius
k = Relative permittivity of insulation between conductors

Characteristic Impedance

If the transmission line is coaxial in construction, the characteristic impedance follows a different equation:

$$Z_0 = \frac{138}{\sqrt{k}} \log \frac{d_1}{d_2}$$

Where,

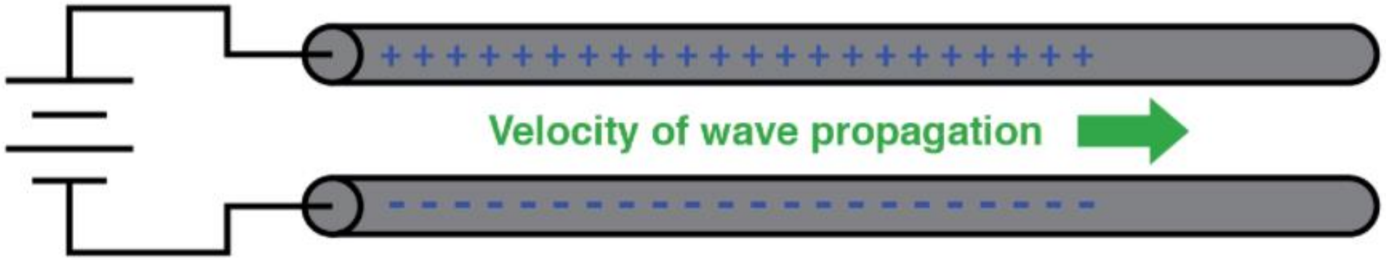
Z_0 = Characteristic impedance of line
d_1 = Inside diameter of outer conductor
d_2 = Outside diameter of inner conductor
k = Relative permittivity of insulation between conductors

Velocity Factor

In both equations, identical units of measurement must be used in both terms of the fraction. If the insulating material is other than air (or a vacuum), both the characteristic impedance and the propagation velocity will be affected. The ratio of a transmission line's true propagation velocity and the speed of light in a vacuum is called the *velocity factor* of that line.

Velocity factor is purely a factor of the insulating material's relative permittivity (otherwise known as its *dielectric constant*), defined as the ratio of a material's electric field permittivity to that of a pure vacuum. The velocity factor of any cable type—coaxial or otherwise—may be calculated quite simply by the following formula:

Velocity Factor



$$\text{Velocity factor} = \frac{v}{c} = \frac{1}{\sqrt{k}}$$

Where,

v = Velocity of wave propagation
c = Velocity of light in a vacuum
k = Relative permittivity of insulation between conductors

Natural Impedance

> Characteristic impedance is also known as *natural impedance*, and it refers to the equivalent resistance of a transmission line if it were infinitely long, owing to distributed capacitance and inductance as the voltage and current "waves" propagate along its length at a propagation velocity equal to some large fraction of light speed.

It can be seen in either of the first two equations that a transmission line's characteristic impedance (Z_0) increases as the conductor spacing increases. If the conductors are moved away from each other, the distributed capacitance will decrease (greater spacing between capacitor "plates"), and the distributed inductance will increase (less cancellation of the two opposing magnetic fields). Less parallel capacitance and more series inductance result in a smaller current drawn by the line for any given amount of applied voltage, which by definition is a greater impedance. Conversely, bringing the two conductors closer together increases the parallel capacitance and decreases the series inductance. Both changes result in a larger current drawn for a given applied voltage, equating to a lesser impedance.

Natural Impedance

Barring any dissipative effects such as dielectric “leakage” and conductor resistance, the characteristic impedance of a transmission line is equal to the square root of the ratio of the line’s inductance per unit length divided by the line’s capacitance per unit length:

$$\boxed{Z_0 = \sqrt{\frac{L}{C}}}$$

Where,

Z_0 = Characteristic impedance of line
L = Inductance per unit length of line
C = Capacitance per unit length of line

Review: Transmission Lines

REVIEW:

- A *transmission line* is a pair of parallel conductors exhibiting certain characteristics due to distributed capacitance and inductance along its length.
- When a voltage is suddenly applied to one end of a transmission line, both a voltage “wave” and a current “wave” propagate along the line at nearly light speed.
- If a DC voltage is applied to one end of an infinitely long transmission line, the line will draw current from the DC source as though it were a constant resistance.
- The *characteristic impedance* (Z_0) of a transmission line is the resistance it would exhibit if it were infinite in length. This is entirely different from leakage resistance of the dielectric separating the two conductors, and the metallic resistance of the wires themselves. Characteristic impedance is purely a function of the capacitance and inductance distributed along the line’s length and would exist even if the dielectric were perfect (infinite parallel resistance) and the wires superconducting (zero series resistance).
- *Velocity factor* is a fractional value relating to a transmission line’s propagation speed to the speed of light in a vacuum. Values range between 0.66 and 0.80 for typical two-wire lines and coaxial cables. For any cable type, it is equal to the reciprocal (1/x) of the square root of the relative permittivity of the cable’s insulation.

Finite Length Transmission Lines

A transmission line of infinite length is an interesting abstraction, but physically impossible. All transmission lines have some finite length, and as such do not behave precisely the same as an infinite line.

If that piece of 50 Ω "RG-58/U" cable measured with an ohmmeter had been infinitely long, I actually would have been able to measure 50 Ω worth of resistance between the inner and outer conductors. But it was not infinite in length, and so it measured as "open" (infinite resistance).

Nonetheless, the characteristic impedance rating of a transmission line is important even when dealing with limited lengths. An older term for characteristic impedance, which I like for its descriptive value, is *surge impedance*.

Finite Length Transmission Lines

If a transient voltage (a "surge") is applied to the end of a transmission line, the line will draw a current proportional to the surge voltage magnitude divided by the line's surge impedance (I=E/Z). This simple, Ohm's Law relationship between current and voltage will hold true for a limited period of time, but not indefinitely.

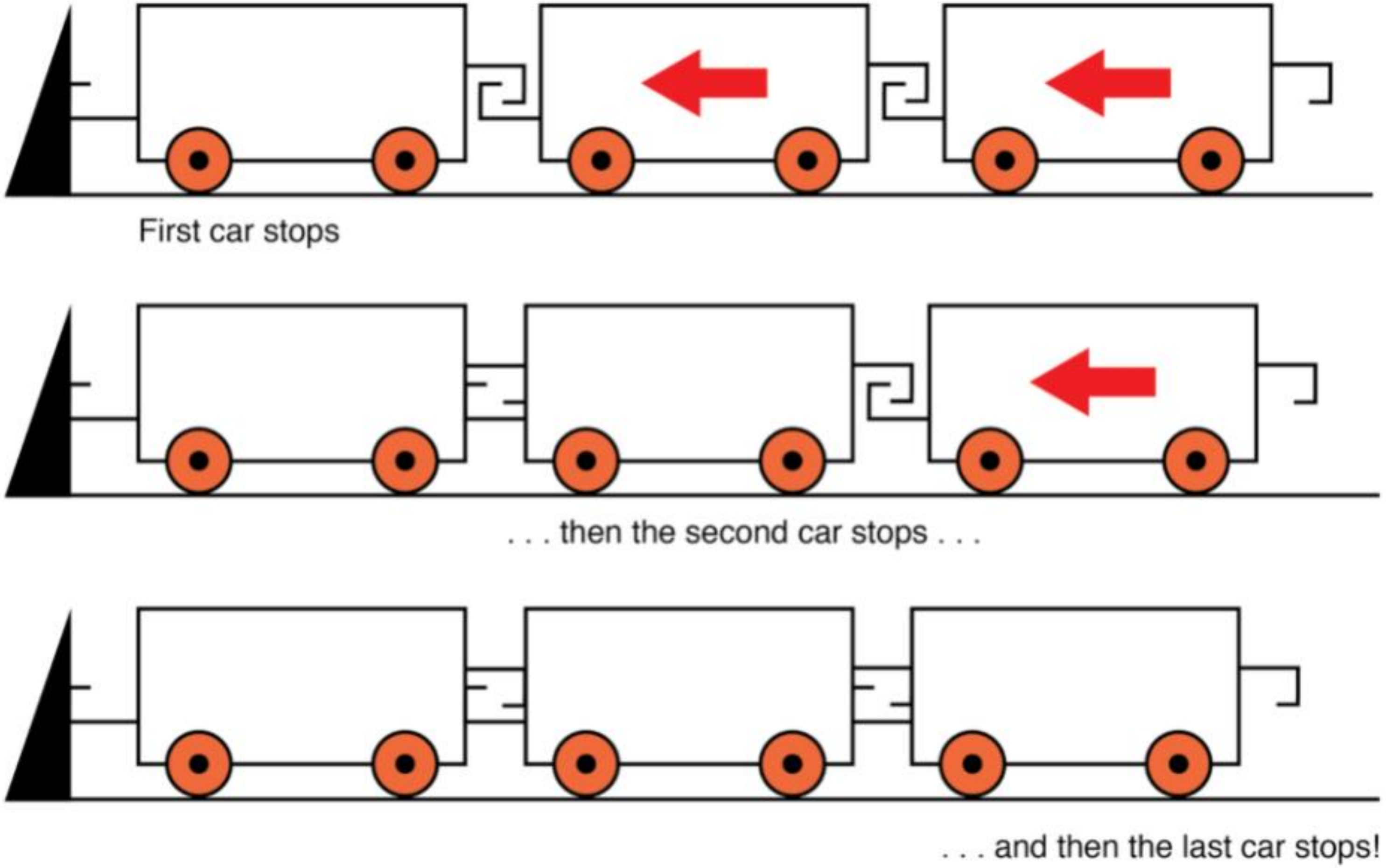
If the end of a transmission line is open-circuited—that is, left unconnected—the current "wave" propagating down the line's length will have to stop at the end, since current cannot flow where there is no continuing path.

This abrupt cessation of current at the line's end causes a "pile-up" to occur along the length of the transmission line, as the electric charge carriers successively find no place to go.

Imagine a train traveling down the track with slack between the railcar couplings: if the lead car suddenly crashes into an immovable barricade, it will come to a stop, causing the one behind it to come to a stop as soon as the first coupling slack is taken up, which causes the next rail car to stop as soon as the next coupling's slack is taken up, and so on until the last rail car stops.

Finite Length Transmission Lines

The train does not come to a halt together, but rather in sequence from first car to last: (Figure below)



Finite Length Transmission Lines

A signal propagating from the source-end of a transmission line to the load-end is called an *incident wave*. The propagation of a signal from load-end to source-end (such as what happened in this example with current encountering the end of an open-circuited transmission line) is called a *reflected wave*.

When this electric charge carrier "pile-up" propagates back to the battery, current at the battery ceases, and the line acts as a simple open circuit.

All this happens very quickly for transmission lines of reasonable length, and so an ohmmeter measurement of the line never reveals the brief time period where the line actually behaves as a resistor.

For a mile-long cable with a velocity factor of 0.66 (signal propagation velocity is 66% of light speed or 122,760 miles per second), it takes only 1/122,760 of a second (8.146 microseconds) for a signal to travel from one end to the other. For the current signal to reach the line's end and "reflect" back to the source, the round-trip time is twice this figure or 16.292 μs.

Significance of Incident and Reflected Waves

High-speed measurement instruments are able to detect this transit time from source to line-end and back to the source again and may be used for the purpose of determining a cable's length.

This technique may also be used for determining the presence *and* location of a break in one or both of the cable's conductors since current will "reflect" off the wire break just as it will off the end of an open-circuited cable.

Instruments designed for such purposes are called *time-domain reflectometers* (TDRs). The basic principle is identical to that of sonar range-finding: generating a sound pulse and measuring the time it takes for the echo to return.

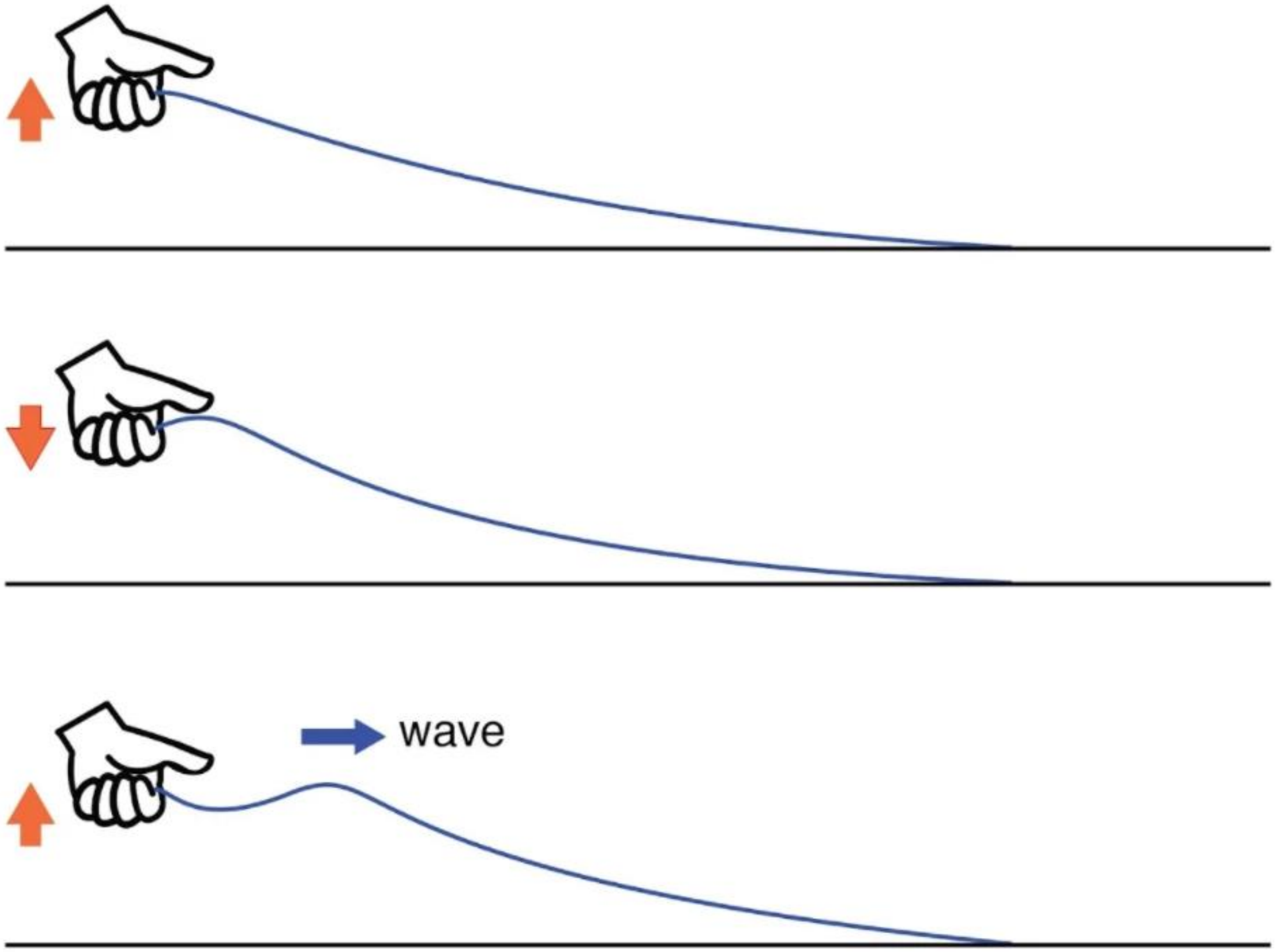
A similar phenomenon takes place if the end of a transmission line is short-circuited: when the voltage wave-front reaches the end of the line, it is reflected back to the source, because voltage cannot exist between two electrically common points.

Significance of Incident and Reflected Waves

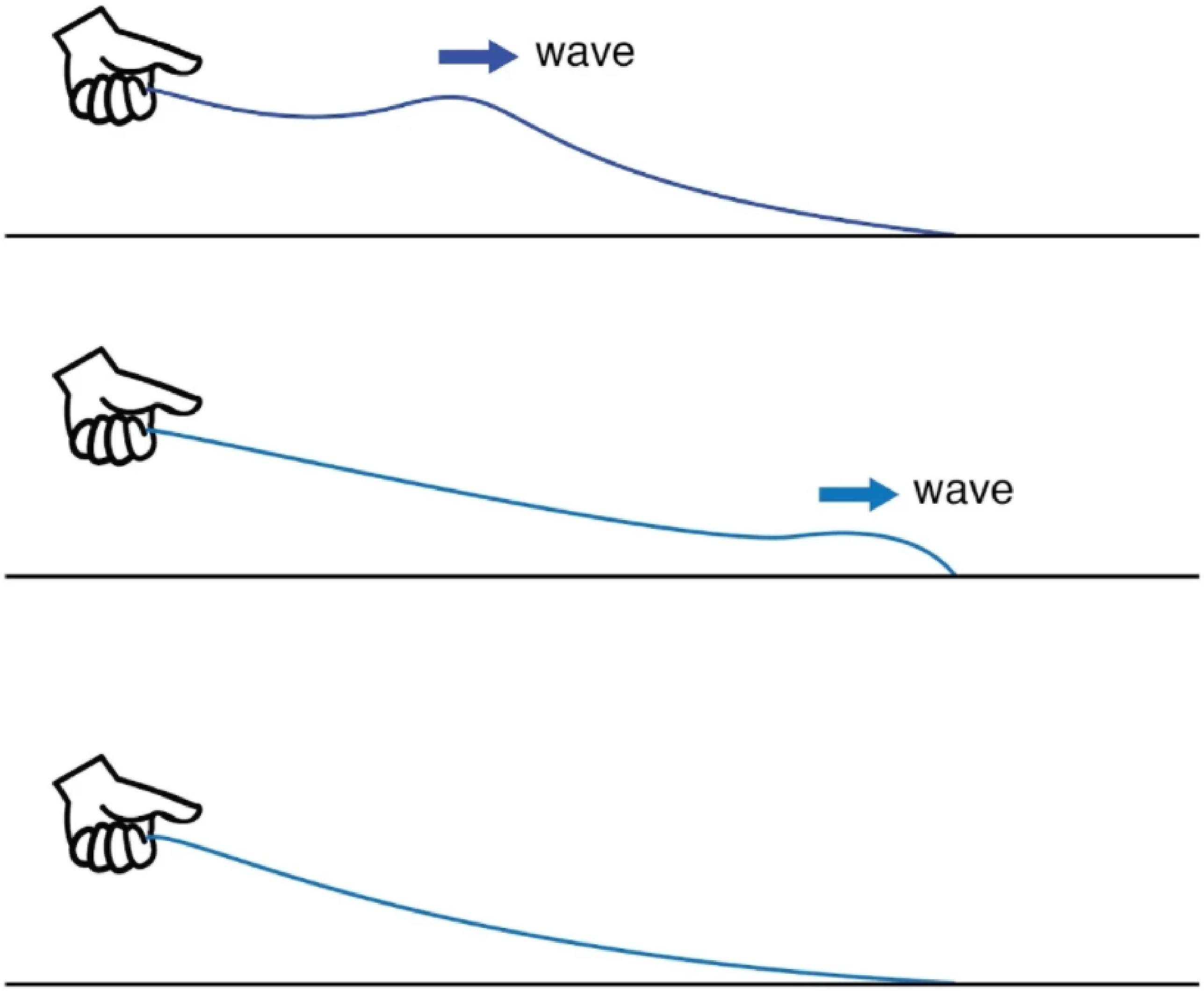
When this reflected wave reaches the source, the source sees the entire transmission line as a short-circuit. Again, this happens as quickly as the signal can propagate round-trip down and up the transmission line at whatever velocity allowed by the dielectric material between the line's conductors.

A simple experiment illustrates the phenomenon of wave reflection in transmission lines. Take a length of rope by one end and "whip" it with a rapid up-and-down motion of the wrist. A wave may be seen traveling down the rope's length until it dissipates entirely due to friction.

Significance of Incident and Reflected Waves



Significance of Incident and Reflected Waves

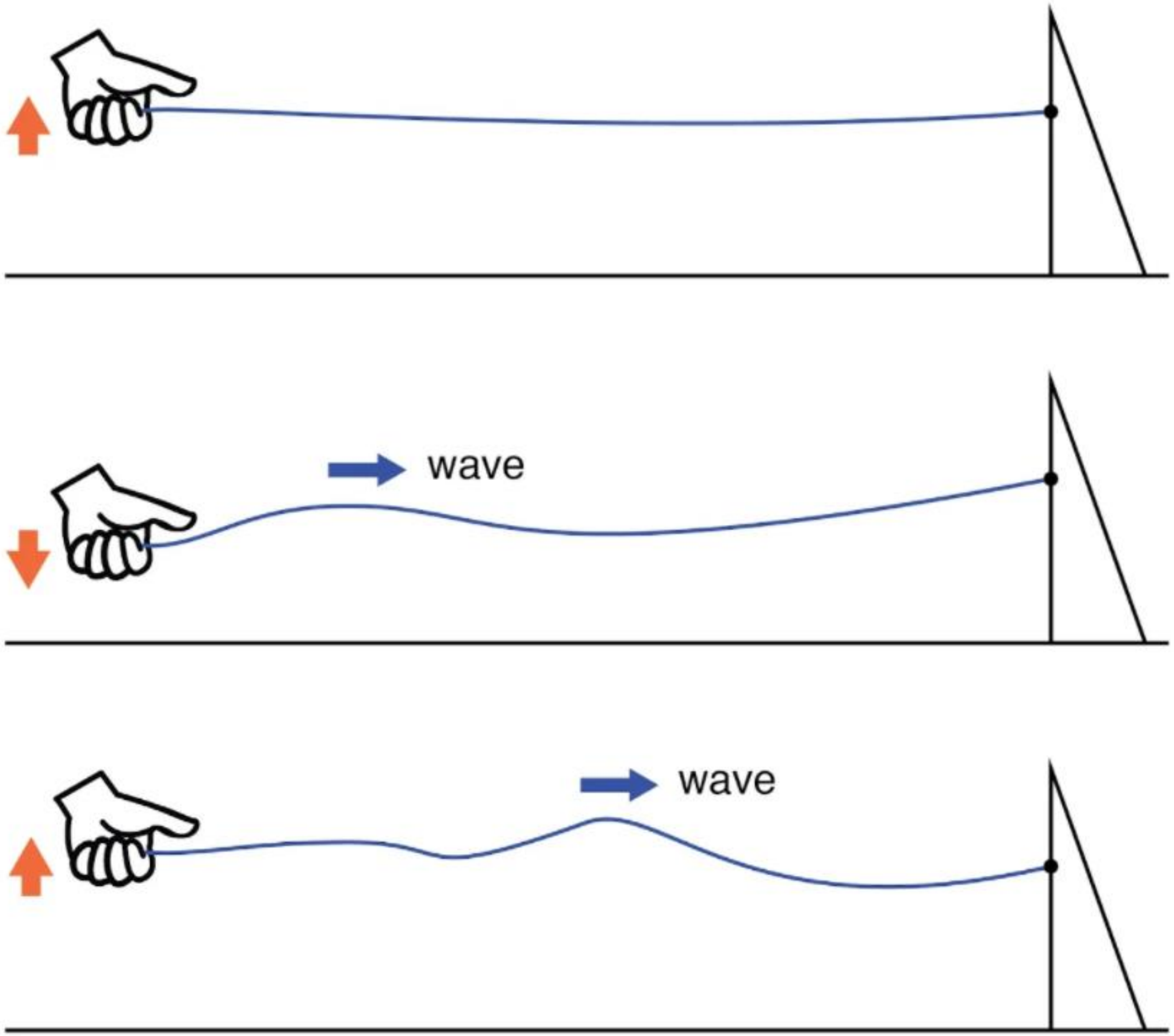


Lossy transmission line.

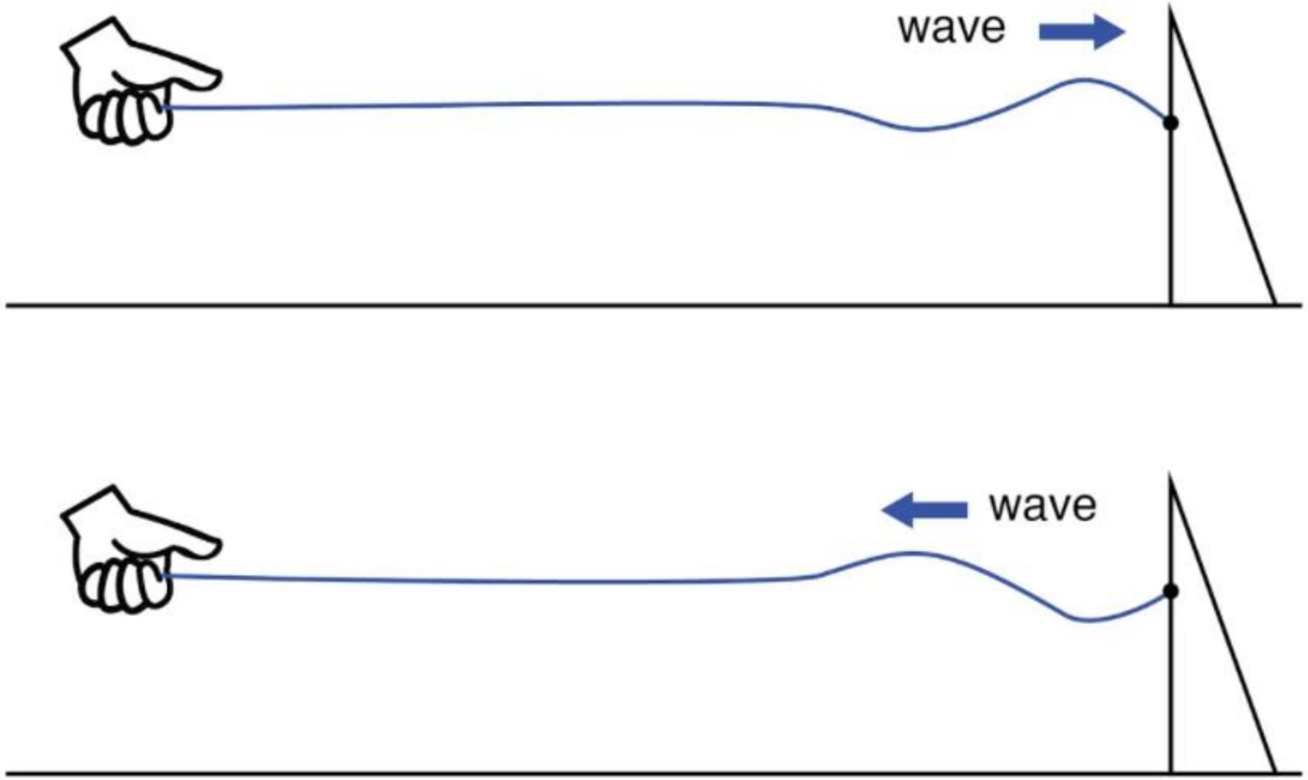
Significance of Incident and Reflected Waves

This is analogous to a long transmission line with internal loss: the signal steadily grows weaker as it propagates down the line's length, never reflecting back to the source. However, if the far end of the rope is secured to a solid object at a point prior to the incident wave's total dissipation, a second wave will be reflected back to your hand: (Figure below)

Significance of Incident and Reflected Waves



Significance of Incident and Reflected Waves



Reflected wave.

Significance of Incident and Reflected Waves

Usually, the purpose of a transmission line is to convey electrical energy from one point to another.

Even if the signals are intended for information only, and not to power some significant load device, the ideal situation would be for all of the original signal energy to travel from the source to the load, and then be completely absorbed or dissipated by the load for maximum signal-to-noise ratio.

Thus, “loss” along the length of a transmission line is undesirable, as are reflected waves, since reflected energy is energy not delivered to the end device.

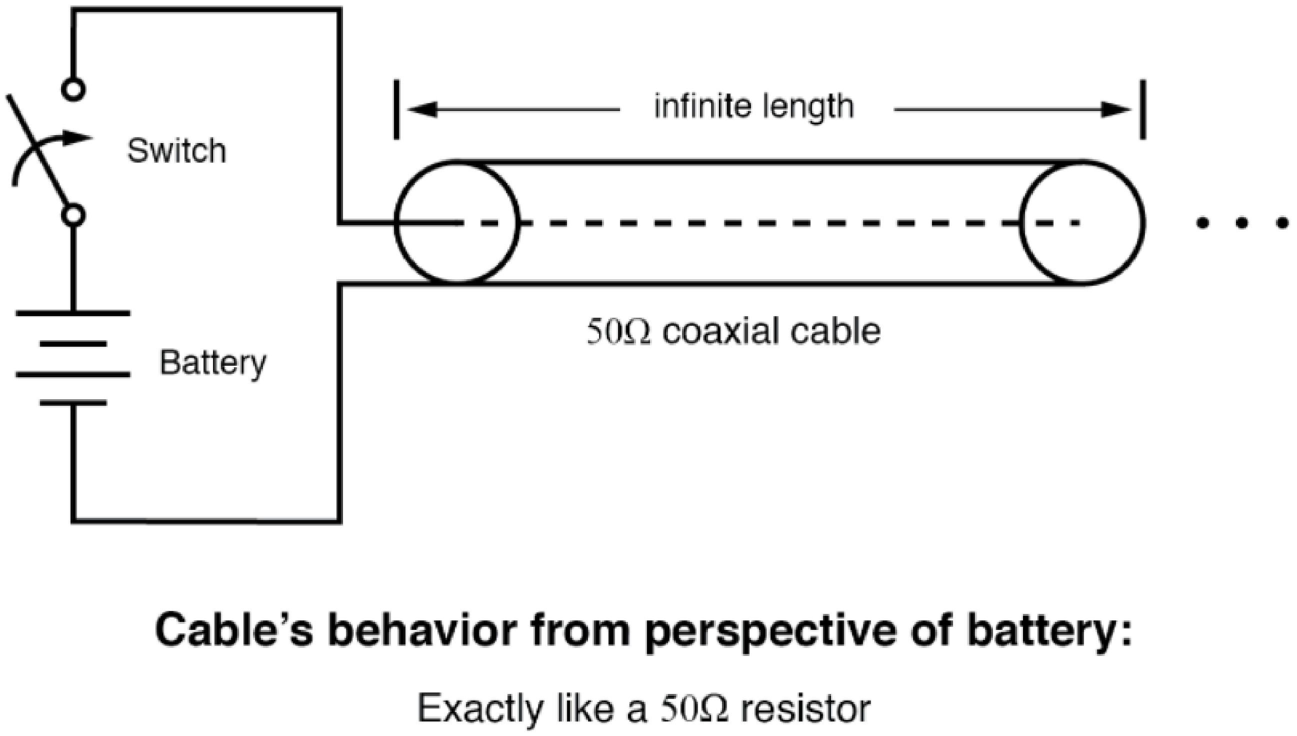
How to Eliminate Reflections in the Transmission Line

Reflections may be eliminated from the transmission line if the load's impedance exactly equals the characteristic ("surge") impedance of the line.

For example, a 50 Ω coaxial cable that is either open-circuited or short-circuited will reflect all of the incident energy back to the source. However, if a 50 Ω resistor is connected at the end of the cable, there will be no reflected energy, all signal energy being dissipated by the resistor.

This makes perfect sense if we return to our hypothetical, infinite-length transmission line example. A transmission line of 50 Ω characteristic impedance and infinite length behaves exactly like a 50 Ω resistance as measured from one end. (Figure below)

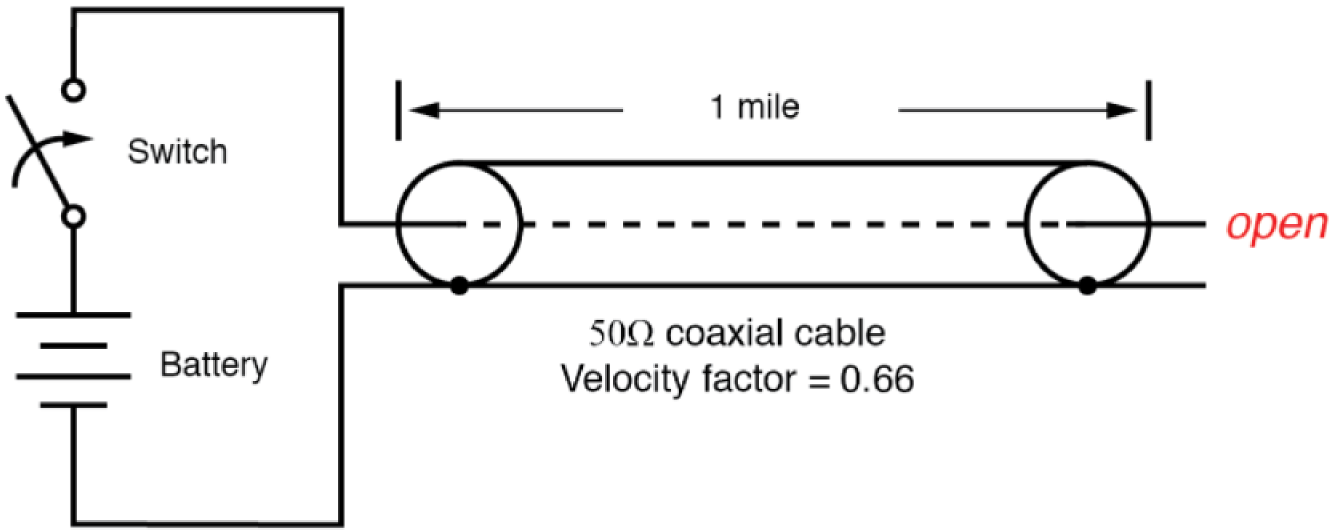
How to Eliminate Reflections in the Transmission Line



Infinite transmission line looks like resistor.

How to Eliminate Reflections in the Transmission Line

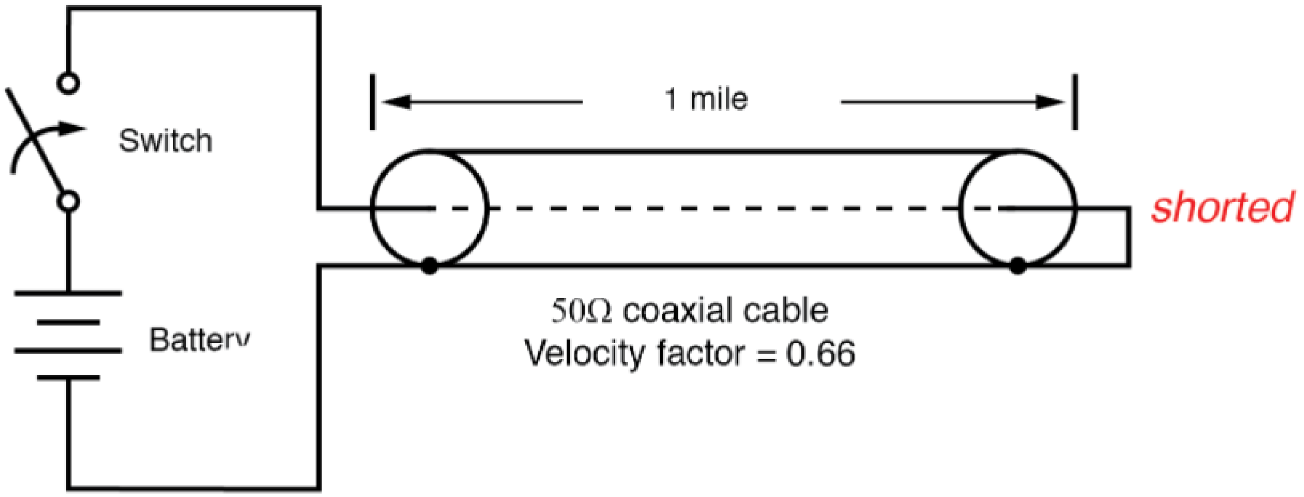
If we cut this line to some finite length, it will behave like a 50 Ω resistor to a constant source of DC voltage for a brief time, but then behave like an open- or a short-circuit, depending on what condition we leave the cut end of the line: open or shorted. (Figure below)



Cable's behavior from perspective of battery:
Like a 50Ω resistor for 16.292 µs,
then like an open (infinite resistance)

One mile transmission.

How to Eliminate Reflections in the Transmission Line



Cable's behavior from perspective of battery:

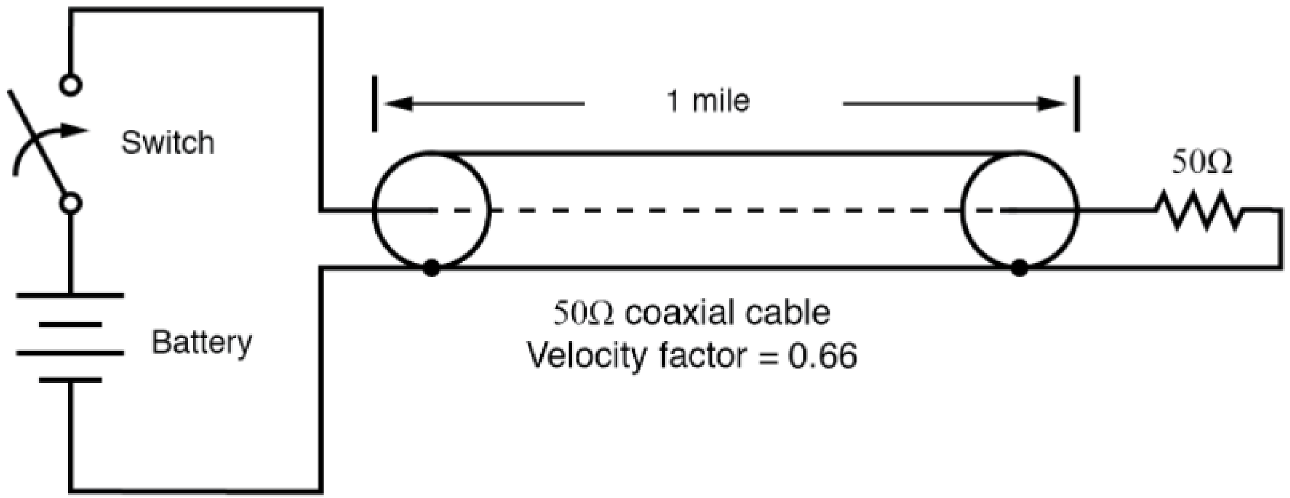
Like a 50Ω resistor for 16.292 μs,

then like a short (zero resistance)

Shorted transmission line.

How to Eliminate Reflections in the Transmission Line

However, if we *terminate* the line with a 50 Ω resistor, the line will once again behave as a 50 Ω resistor, indefinitely: the same as if it were of an infinite length again: (Figure below)



Cable's behavior from perspective of battery:
Exactly like a 50Ω resistor

Line terminated in characteristic impedance.

How to Eliminate Reflections in the Transmission Line

In essence, a terminating resistor matching the natural impedance of the transmission line makes the line "appear" infinitely long from the perspective of the source, because a resistor has the ability to eternally dissipate energy in the same way a transmission line of infinite length is able to eternally absorb energy.

Reflected waves will also manifest if the terminating resistance isn't precisely equal to the characteristic impedance of the transmission line, not just if the line is left unconnected (open) or jumpered (shorted).

Though the energy reflection will not be total with a terminating impedance of a slight mismatch, it will be partial. This happens whether or not the terminating resistance is *greater* or *less* than the line's characteristic impedance.

How to Eliminate Reflections in the Transmission Line

Re-reflections of a reflected wave may also occur at the *source end* of a transmission line if the source's internal impedance (Thevenin equivalent impedance) is not exactly equal to the line's characteristic impedance.

A reflected wave returning back to the source will be dissipated entirely if the source impedance matches the line's, but will be reflected back toward the line end like another incident wave, at least partially, if the source impedance does not match the line.

This type of reflection may be particularly troublesome, as it makes it appear that the source has transmitted another pulse.

Review: How to Eliminate Reflections in the Transmission Line (1/2)

- Characteristic impedance is also known as *surge impedance*, due to the temporarily resistive behavior of any length transmission line.

- A finite-length transmission line will appear to a DC voltage source as a constant resistance for some short time, then as whatever impedance, the line is terminated with. Therefore, an open-ended cable simply reads "open" when measured with an ohmmeter, and "shorted" when its end is short-circuited.

- A transient ("surge") signal applied to one end of an open-ended or short-circuited transmission line will "reflect" off the far end of the line as a secondary wave. A signal traveling on a transmission line from source to load is called an *incident wave*; a signal "bounced" off the end of a transmission line, traveling from load to source, is called a *reflected wave*.

Review: How to Eliminate Reflections in the Transmission Line (2/2)

- Reflected waves will also appear in transmission lines terminated by resistors not precisely matching the characteristic impedance.

- A finite-length transmission line may be made to appear infinite in length if terminated by a resistor of equal value to the line's characteristic impedance. This eliminates all signal reflections.

- A reflected wave may become re-reflected off the source-end of a transmission line if the source's internal impedance does not match the line's characteristic impedance. This re-reflected wave will appear, of course, like another pulse signal transmitted from the source.

“Long” and “Short” Transmission Lines

In DC and low-frequency AC circuits, the characteristic impedance of parallel wires is usually ignored. This includes the use of coaxial cables in instrument circuits, often employed to protect weak voltage signals from being corrupted by induced “noise” caused by stray electric and magnetic fields.

This is due to the relatively short timespans in which reflections take place in the line, as compared to the period of the waveforms or pulses of the significant signals in the circuit.

As we saw in the last section, if a transmission line is connected to a DC voltage source, it will behave as a resistor equal in value to the line’s characteristic impedance only for as long as it takes the incident pulse to reach the end of the line and return as a reflected pulse, back to the source.

“Long” and “Short” Transmission Lines

After that time (a brief 16.292 μs for the mile-long coaxial cable of the last example), the source “sees” only the terminating impedance, whatever that may be.

If the circuit in question handles low-frequency AC power, such short time delays introduced by a transmission line between when the AC source outputs a voltage peak and when the source “sees” that peak loaded by the terminating impedance (round-trip time for the incident wave to reach the line’s end and reflect back to the source) are of little consequence.

Even though we know that signal magnitudes along the line’s length are not equal at any given time due to signal propagation at (nearly) the speed of light, the actual phase difference between start-of-line and end-of-line signals is negligible, because line-length propagations occur within a very small fraction of the AC waveform’s period.

"Long" and "Short" Transmission Lines

For all practical purposes, we can say that voltage along all respective points on a low-frequency, two-conductor line are equal and in-phase with each other at any given point in time.

In these cases, we can say that the transmission lines in question are *electrically short*, because their propagation effects are much quicker than the periods of the conducted signals.

By contrast, an *electrically long* line is one where the propagation time is a large fraction or even a multiple of the signal period. A "long" line is generally considered to be one where the source's signal waveform completes at least a quarter-cycle (90° of "rotation") before the incident signal reaches line's end.

How to Calculate the Wavelength?

To put this into perspective, we need to express the distance traveled by a voltage or current signal along a transmission line in relation to its source frequency. An AC waveform with a frequency of 60 Hz completes one cycle in 16.66 ms.

At light speed (186,000 mile/s), this equates to a distance of 3100 miles that a voltage or current signal will propagate in that time. If the velocity factor of the transmission line is less than 1, the propagation velocity will be less than 186,000 miles per second, and the distance less by the same factor.

But even if we used the coaxial cable’s velocity factor from the last example (0.66), the distance is still a very long 2046 miles! Whatever distance we calculate for a given frequency is called the *wavelength* of the signal.

How to Calculate the Wavelength?

A simple formula for calculating wavelength is as follows:

$$\lambda = \frac{v}{f}$$

Where,

λ = Wavelength

v = Velocity of propagation

f = Frequency of signal

How to Calculate the Wavelength?

The lowercase Greek letter "lambda" (λ) represents wavelength, in whatever unit of length used in the velocity figure (if miles per second, then wavelength in miles; if meters per second, then wavelength in meters).

Velocity of propagation is usually the speed of light when calculating signal wavelength in open air or in a vacuum, but will be less if the transmission line has a velocity factor less than 1.

If a "long" line is considered to be one at least 1/4 wavelength in length, you can see why all connecting lines in the circuits discussed thus far have been assumed "short."

How to Calculate the Wavelength?

For a 60 Hz AC power system, power lines would have to exceed 775 miles in length before the effects of propagation time became significant. Cables connecting an audio amplifier to speakers would have to be over 4.65 miles in length before line reflections would significantly impact a 10 kHz audio signal!

When dealing with radio-frequency systems, though, transmission line length is far from trivial. Consider a 100 MHz radio signal: its wavelength is a mere 9.8202 feet, even at the full propagation velocity of light (186,000 mile/s).

A transmission line carrying this signal would not have to be more than about 2-1/2 feet in length to be considered “long!” With a cable velocity factor of 0.66, this critical length shrinks to 1.62 feet.

What Happens if the Transmission Line is "Short"?

When an electrical source is connected to a load via a "short" transmission line, the load's impedance dominates the circuit. This is to say, when the line is short, its own characteristic impedance is of little consequence to the circuit's behavior.

We see this when testing a coaxial cable with an ohmmeter: the cable reads "open" from center conductor to outer conductor if the cable end is left unterminated.

Though the line acts as a resistor for a very brief period of time after the meter is connected (about 50 Ω for an RG-58/U cable), it immediately thereafter behaves as a simple "open circuit:" the impedance of the line's open end.

What Happens if the Transmission Line is “Short”?

Since the combined response time of an ohmmeter and the human being using it *greatly exceeds* the round-trip propagation time up and down the cable, it is “electrically short” for this application, and we only register the terminating (load) impedance.

It is the extreme speed of the propagated signal that makes us unable to detect the cable’s 50 Ω transient impedance with an ohmmeter.

If we use a coaxial cable to conduct a DC voltage or current to a load, and no component in the circuit is capable of measuring or responding quickly enough to “notice” a reflected wave, the cable is considered “electrically short” and its impedance is irrelevant to circuit function.

What Happens if the Transmission Line is “Short”?

Note how the electrical “shortness” of a cable is relative to the application: in a DC circuit where voltage and current values change slowly, nearly any physical length of cable would be considered “short” from the standpoint of characteristic impedance and reflected waves.

Taking the same length of cable, though, and using it to conduct a high-frequency AC signal could result in a vastly different assessment of that cable’s “shortness!”

When a source is connected to a load via a “long” transmission line, the line’s own characteristic impedance dominates over load impedance in determining circuit behavior. In other words, an electrically “long” line acts as the principal component in the circuit, its own characteristics overshadowing the load’s.

What Happens When the Transmission Line is Electrically "Long"?

With a source connected to one end of the cable and a load to the other, current drawn from the source is a function primarily of the line and not the load. This is increasingly true the longer the transmission line is.

Consider our hypothetical 50 Ω cable of infinite length, surely the ultimate example of a "long" transmission line: no matter what kind of load we connect to one end of this line, the source (connected to the other end) will only see 50 Ω of impedance, because the line's infinite length prevents the signal from *ever reaching* the end where the load is connected.

In this scenario, line impedance exclusively defines circuit behavior, rendering the load completely irrelevant.

How to Minimize the Impact of Transmission Line Length on a Circuit?

The most effective way to minimize the impact of transmission line length on circuit behavior is to match the line's characteristic impedance to the load impedance.

If the load impedance is equal to the line impedance, then *any* signal source connected to the other end of the line will "see" the exact same impedance, and will have the exact same amount of current drawn from it, regardless of line length.

In this condition of perfect impedance matching, line length only affects the amount of time delay from signal departure at the source to signal arrival at the load. However, perfect matching of line and load impedances is not always practical or possible.

The next section discusses the effects of "long" transmission lines, especially when line length happens to match specific fractions or multiples of signal wavelength.

Review: How to Minimize the Impact of Transmission Line Length on a Circuit? (1/2)

- Coaxial cabling is sometimes used in DC and low-frequency AC circuits as well as in high-frequency circuits, for the excellent immunity to induced “noise” that it provides for signals.

- When the period of a transmitted voltage or current signal greatly exceeds the propagation time for a transmission line, the line is considered *electrically short*. Conversely, when the propagation time is a large fraction or multiple of the signal’s period, the line is considered *electrically long*.

- A signal’s *wavelength* is the physical distance it will propagate in the timespan of one period. Wavelength is calculated by the formula $\lambda=v/f$, where “λ” is the wavelength, “v” is the propagation velocity, and “f” is the signal frequency.

Review: How to Minimize the Impact of Transmission Line Length on a Circuit? (2/2)

- A rule-of-thumb for transmission line "shortness" is that the line must be at least 1/4 wavelength before it is considered "long."

- In a circuit with a "short" line, the terminating (load) impedance dominates circuit behavior. The source effectively sees nothing but the load's impedance, barring any resistive losses in the transmission line.

- In a circuit with a "long" line, the line's own characteristic impedance dominates circuit behavior. The ultimate example of this is a transmission line of infinite length: since the signal will *never* reach the load impedance, the source only "sees" the cable's characteristic impedance.

- When a transmission line is terminated by a load precisely matching its impedance, there are no reflected waves and thus no problems with line length.

Training Enquiry

PSDC

admin@psdc.org.my

www.psdc.org.my

04-643 7929

Solution Enquiry

ORIONPLEX

ORIONTRAIN

gary.lee@orionplex.com.my

www.orionplex.com.my

019-4106712

Gary Lee Engineering Solutions Penang

espenang.6712

Thank you