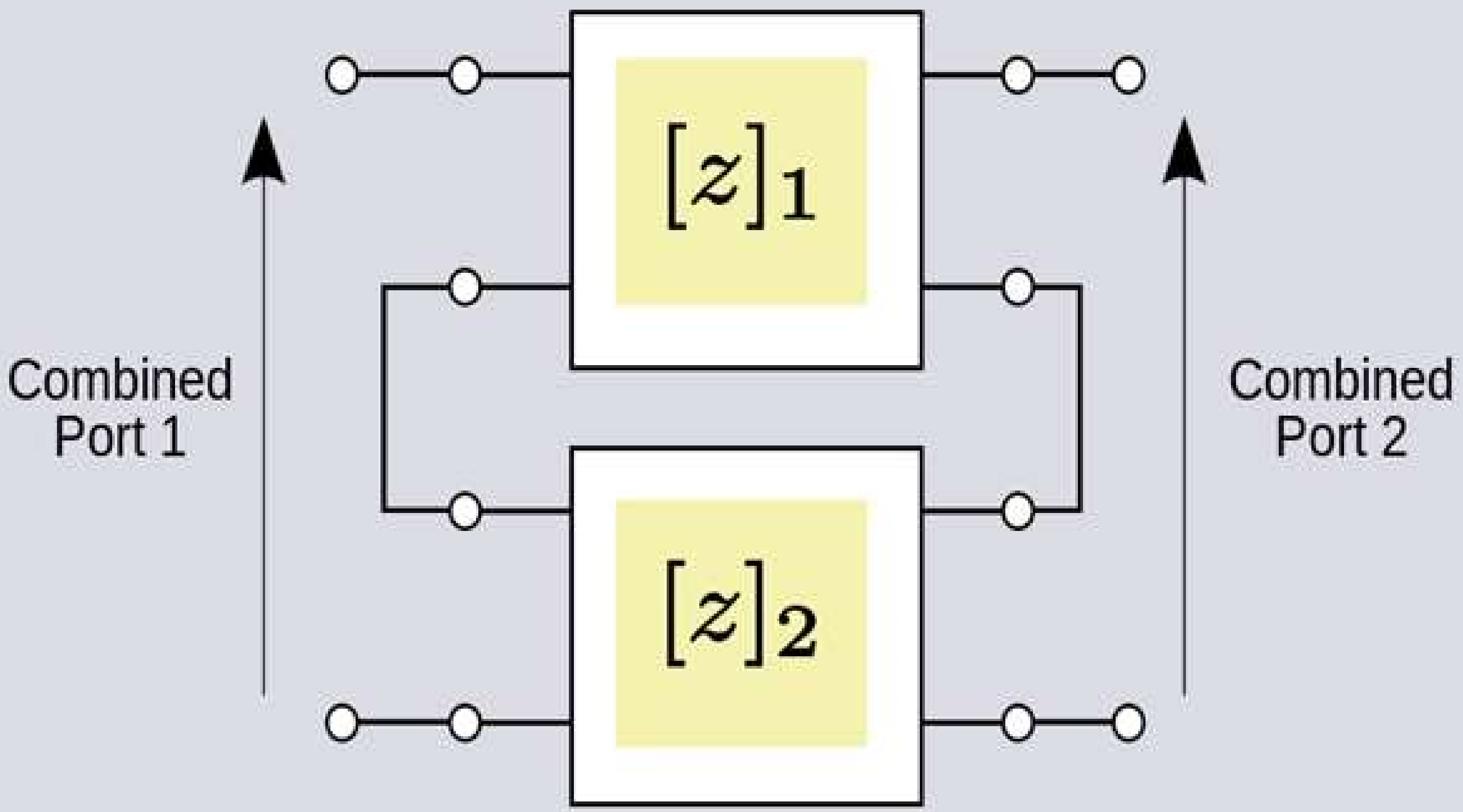
Concepts, Elements and Components of

Network Analysis



Sachiko Woodson

Uriel Matthews

First Edition, 2012

ISBN 978-81-323-1344-1

Published by:
College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Introduction to Network Analysis

A network, in the context of electronics, is a collection of interconnected components. **Network analysis** is the process of finding the voltages across, and the currents through, every component in the network. There are a number of different techniques for achieving this. However, for the most part, they assume that the components of the network are all linear. The methods described here are only applicable to *linear* network analysis except where explicitly stated.

Definitions

Component
A device with two or more terminals into which, or out of which, charge may flow.

Node
A point at which terminals of more than two components are joined. A conductor with a substantially zero resistance is considered to be a node for the purpose of analysis.

Branch
The component(s) joining two nodes.

Mesh
A group of branches within a network joined so as to form a complete loop.

Port
Two terminals where the current into one is identical to the current out of the other.

Circuit
A current from one terminal of a generator, through load component(s) and back into the other terminal. A circuit is, in this sense, a one-port network and is a trivial case to analyse. If there is any connection to any other circuits then a non-trivial network has been formed and at least two ports must exist. Often, "circuit" and "network" are used interchangeably, but many analysts reserve "network" to mean an idealised model consisting of ideal components.

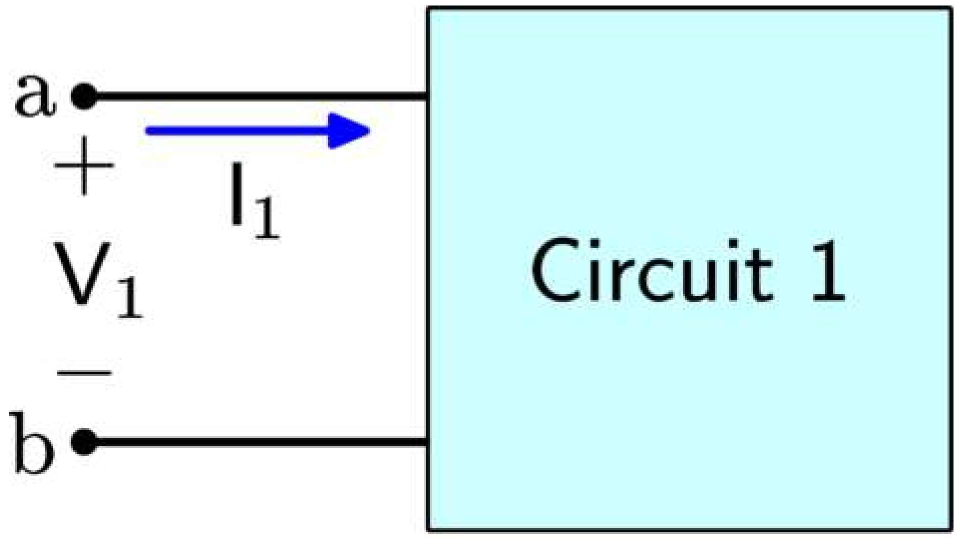
Transfer function

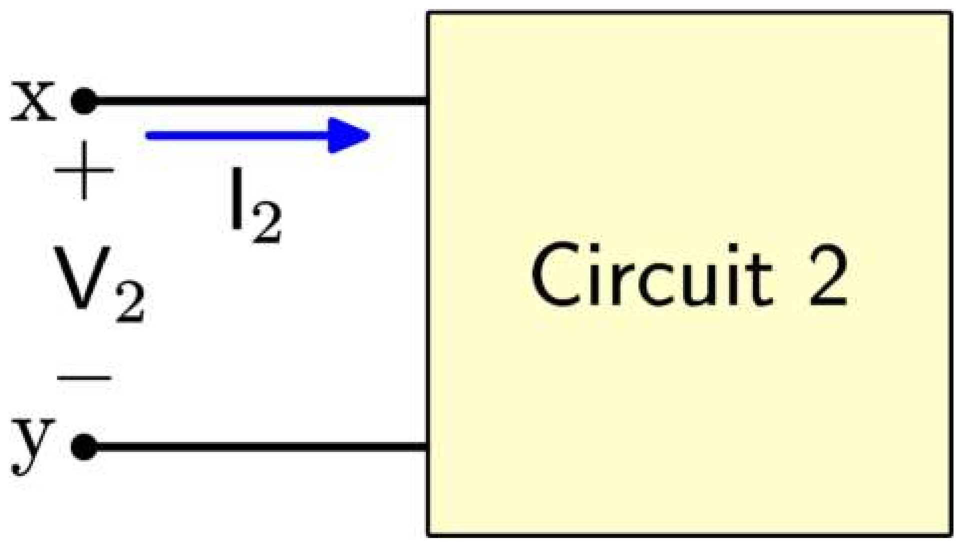
The relationship of the currents and/or voltages between two ports. Most often, an input port and an output port are discussed and the transfer function is described as gain or attenuation.

Component transfer function

For a two-terminal component (i.e. one-port component), the current and voltage are taken as the input and output and the transfer function will have units of impedance or admittance (it is usually a matter of arbitrary convenience whether voltage or current is considered the input). A three (or more) terminal component effectively has two (or more) ports and the transfer function cannot be expressed as a single impedance. The usual approach is to express the transfer function as a matrix of parameters. These parameters can be impedances, but there is a large number of other approaches.

Equivalent circuits





A useful procedure in network analysis is to simplify the network by reducing the number of components. This can be done by replacing the actual components with other notional components that have the same effect. A particular technique might directly reduce the number of components, for instance by combining impedances in series. On the other hand it might merely change the form in to one in which the components can be reduced in a later operation. For instance, one might transform a voltage generator into a current generator using Norton's theorem in order to be able to later combine the internal resistance of the generator with a parallel impedance load.

A resistive circuit is a circuit containing only resistors, ideal current sources, and ideal voltage sources. If the sources are constant (DC) sources, the result is a DC circuit. The analysis of a circuit refers to the process of solving for the voltages and currents present in the circuit. The solution principles outlined here also apply to phasor analysis of AC circuits.

Two circuits are said to be **equivalent** with respect to a pair of terminals if the voltage across the terminals and current through the terminals for one network have the same relationship as the voltage and current at the terminals of the other network.

If $V_2 = V_1$ implies $I_2 = I_1$ for all (real) values of V_1, then with respect to terminals ab and xy, circuit 1 and circuit 2 are equivalent.

The above is a sufficient definition for a one-port network. For more than one port, then it must be defined that the currents and voltages between all pairs of corresponding ports must bear the same relationship. For instance, star and delta networks are effectively three port networks and hence require three simultaneous equations to fully specify their equivalence.

Impedances in series and in parallel

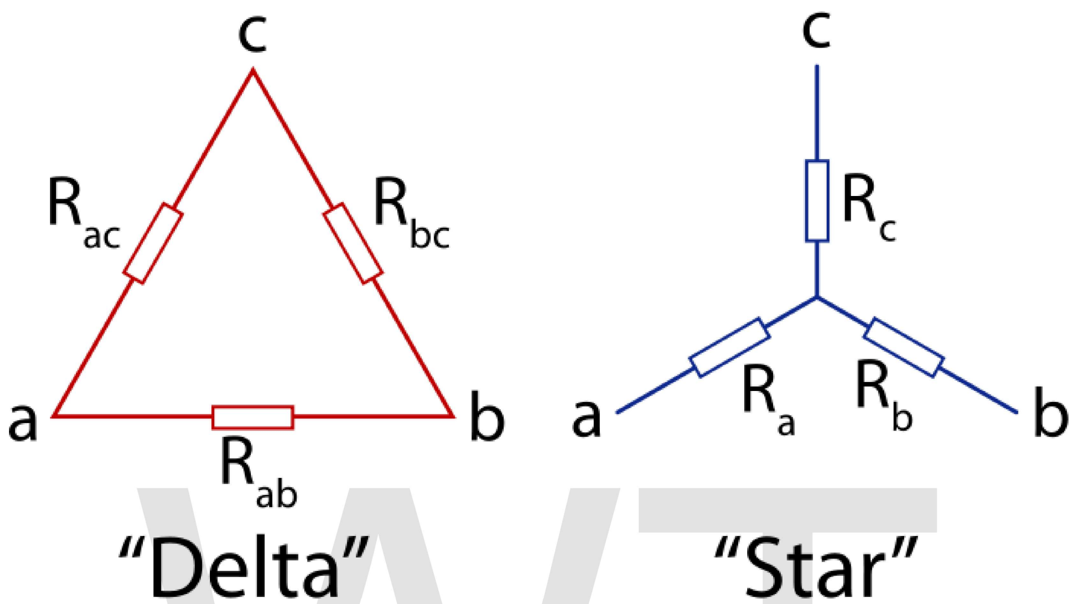
Any two terminal network of impedances can eventually be reduced to a single impedance by successive applications of impendances in series or impendances in parallel.

Impedances in series: $Z_{\mathrm{eq}} = Z_1 + Z_2 + \cdots + Z_n.$

Impedances in parallel: $\frac{1}{Z_{\mathrm{eq}}} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}.$

The above simplified for only two impedances in parallel: $Z_{\mathrm{eq}} = \frac{Z_1 Z_2}{Z_1 + Z_2}.$

Delta-wye transformation



A network of impedances with more than two terminals cannot be reduced to a single impedance equivalent circuit. An n-terminal network can, at best, be reduced to n impedances. For a three terminal network, the three impedances can be expressed as a three node delta (Δ) network or a four node star (Y) network. These two networks are equivalent and the transformations between them are given below. A general network with an arbitrary number of terminals cannot be reduced to the minimum number of impedances using only series and parallel combinations. In general, Y-Δ and Δ-Y transformations must also be used. It can be shown that this is sufficient to find the minimal network for any arbitrary network with successive applications of series, parallel, Y-Δ and Δ-Y; no more complex transformations are required.

For equivalence, the impedances between any pair of terminals must be the same for both networks, resulting in a set of three simultaneous equations. The equations below are expressed as resistances but apply equally to the general case with impedances.

Delta-to-star transformation equations

$$R_a = \frac{R_{\mathrm{ac}}R_{\mathrm{ab}}}{R_{\mathrm{ac}} + R_{\mathrm{ab}} + R_{\mathrm{bc}}}$$
$$R_b = \frac{R_{\mathrm{ab}}R_{\mathrm{bc}}}{R_{\mathrm{ac}} + R_{\mathrm{ab}} + R_{\mathrm{bc}}}$$
$$R_c = \frac{R_{\mathrm{bc}}R_{\mathrm{ac}}}{R_{\mathrm{ac}} + R_{\mathrm{ab}} + R_{\mathrm{bc}}}$$

Star-to-delta transformation equations

$$R_{\mathrm{ac}} = \frac{R_aR_b + R_bR_c + R_cR_a}{R_b}$$
$$R_{\mathrm{ab}} = \frac{R_aR_b + R_bR_c + R_cR_a}{R_c}$$
$$R_{\mathrm{bc}} = \frac{R_aR_b + R_bR_c + R_cR_a}{R_a}$$

General form of network node elimination

The star-to-delta and series-resistor transformations are special cases of the general resistor network node elimination algorithm. Any node connected by N resistors (R_1 .. R_N) to nodes **1** .. **N** can be replaced by $\binom{N}{2}$ resistors interconnecting the remaining N nodes. The resistance between any two nodes x and y is given by:

$$R_{\mathrm{xy}} = R_xR_y\sum_{i=1}^{N}\frac{1}{R_i}$$

For a star-to-delta (N = 3) this reduces to:

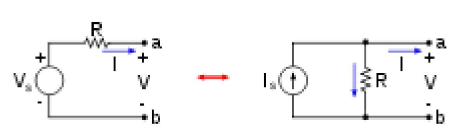
$$R_{\mathrm{ab}} = R_aR_b(\frac{1}{R_a}+\frac{1}{R_b}+\frac{1}{R_c}) = \frac{R_aR_b(R_aR_b+R_aR_c+R_bR_c)}{R_aR_bR_c} = \frac{R_aR_b+R_bR_c+R_cR_a}{R_c}$$

For a series reduction (N = 2) this reduces to:

$$R_{\mathrm{ab}} = R_aR_b(\frac{1}{R_a}+\frac{1}{R_b}) = \frac{R_aR_b(R_a+R_b)}{R_aR_b} = R_a+R_b$$

For a dangling resistor (N = 1) it results in the elimination of the resistor because $\binom{1}{2} = 0$.

Source transformation



A generator with an internal impedance (ie non-ideal generator) can be represented as either an ideal voltage generator or an ideal current generator plus the impedance. These two forms are equivalent and the transformations are given below. If the two networks are equivalent with respect to terminals ab, then V and I must be identical for both networks. Thus,

$$V_s = RI_s \text{or} \; I_s = \frac{V_s}{R}$$

- Norton's theorem states that any two-terminal network can be reduced to an ideal current generator and a parallel impedance.
- Thévenin's theorem states that any two-terminal network can be reduced to an ideal voltage generator plus a series impedance.

Simple networks

Some very simple networks can be analysed without the need to apply the more systematic approaches.

Voltage division of series components

Consider n impedances that are connected in **series**. The voltage V_i across any impedance Z_i is

$$V_i = Z_i I = \left(\frac{Z_i}{Z_1 + Z_2 + \cdots + Z_n}\right) V$$

Current division of parallel components

Consider n impedances that are connected in **parallel**. The current I_i through any impedance Z_i is

$$I_i = \left(\frac{\left(\frac{1}{Z_i}\right)}{\left(\frac{1}{Z_1}\right) + \left(\frac{1}{Z_2}\right) + \cdots + \left(\frac{1}{Z_n}\right)}\right) I$$

for $i = 1,2,...,n$.

Special case: Current division of two parallel components

$$I_1 = \left(\frac{Z_2}{Z_1 + Z_2}\right) I$$

$$I_2 = \left(\frac{Z_1}{Z_1 + Z_2}\right) I$$

Nodal analysis

1. Label all **nodes** in the circuit. Arbitrarily select any node as reference.

2. Define a voltage variable from every remaining node to the reference. These voltage variables must be defined as voltage rises with respect to the reference node.

3. Write a KCL equation for every node except the reference.

4. Solve the resulting system of equations.

Mesh analysis

Mesh — a loop that does not contain an inner loop.

1. Count the number of "window panes" in the circuit. Assign a mesh current to each window pane.

2. Write a KVL equation for every mesh whose current is unknown.

3. Solve the resulting equations

Superposition

In this method, the effect of each generator in turn is calculated. All the generators other than the one being considered are removed; either short-circuited in the case of voltage generators, or open circuited in the case of current generators. The total current through, or the total voltage across, a particular branch is then calculated by summing all the individual currents or voltages.

There is an underlying assumption to this method that the total current or voltage is a linear superposition of its parts. The method cannot, therefore, be used if non-linear components are present. Note that mesh analysis and node analysis also implicitly use superposition so these too, are only applicable to linear circuits.

Choice of method

Choice of method is to some extent a matter of taste. If the network is particularly simple or only a specific current or voltage is required then ad-hoc application of some simple equivalent circuits may yield the answer without recourse to the more systematic methods.

- Superposition is possibly the most conceptually simple method but rapidly leads to a large number of equations and messy impedance combinations as the network becomes larger.
- Nodal analysis: The number of voltage variables, and hence simultaneous equations to solve, equals the number of nodes minus one. Every voltage source connected to the reference node reduces the number of unknowns (and equations) by one. Nodal analysis is thus best for voltage sources.
- Mesh analysis: The number of current variables, and hence simultaneous equations to solve, equals the number of meshes. Every current source in a mesh reduces the number of unknowns by one. Mesh analysis is thus best for current sources. Mesh analysis, however, cannot be used with networks which cannot be drawn as a planar network, that is, with no crossing components.

Transfer function

A transfer function expresses the relationship between an input and an output of a network. For resistive networks, this will always be a simple real number or an expression which boils down to a real number. Resistive networks are represented by a system of simultaneous algebraic equations. However in the general case of linear networks, the network is represented by a system of simultaneous linear differential equations. In network analysis, rather than use the differential equations directly, it is usual practice to carry out a Laplace transform on them first and then express the result in terms of the Laplace parameter s, which in general is complex. This is described as working in the s-domain. Working with the equations directly would be described as working in the time (or t) domain because the results would be expressed as time varying quantities. The Laplace transform is the mathematical method of transforming between the s-domain and the t-domain.

This approach is standard in control theory and is useful for determining stability of a system, for instance, in an amplifier with feedback.

Two terminal component transfer functions

For two terminal components the transfer function, otherwise called the constitutive equation, is the relationship between the current input to the device and the resulting voltage across it. The transfer function, Z(s), will thus have units of impedance - ohms. For the three passive components found in electrical networks, the transfer functions are;

Resistor $Z(s) = R$

Inductor $Z(s) = sL$

Capacitor $Z(s) = \frac{1}{sC}$

For a network to which only steady ac signals are applied, s is replaced with $j\omega$ and the more familiar values from ac network theory result.

Resistor $Z(j\omega) = R$
Inductor $Z(j\omega) = j\omega L$
Capacitor $Z(j\omega) = \frac{1}{j\omega C}$

Finally, for a network to which only steady dc is applied, s is replaced with zero and dc network theory applies.

Resistor $Z = R$
Inductor $Z = 0$
Capacitor $Z = \infty$

Two port network transfer function

Transfer functions, in general, in control theory are given the symbol H(s). Most commonly in electronics, transfer function is defined as the ratio of output voltage to input voltage and given the symbol A(s), or more commonly (because analysis is invariably done in terms of sine wave response), A(jω), so that;

$$A(j\omega) = \frac{V_o}{V_i}$$

The A standing for attenuation, or amplification, depending on context. In general, this will be a complex function of $j\omega$, which can be derived from an analysis of the impedances in the network and their individual transfer functions. Sometimes the analyst is only interested in the magnitude of the gain and not the phase angle. In this case the complex numbers can be eliminated from the transfer function and it might then be written as;

$$A(\omega) = \left|\frac{V_o}{V_i}\right|$$

Two port parameters

The concept of a two-port network can be useful in network analysis as a black box approach to analysis. The behaviour of the two-port network in a larger network can be entirely characterised without necessarily stating anything about the internal structure. However, to do this it is necessary to have more information than just the A(jω) described above. It can be shown that four such parameters are required to fully characterise the two-port network. These could be the forward transfer function, the input impedance, the reverse transfer function (ie, the voltage appearing at the input when a voltage is applied to the output) and the output impedance. There are many others, one of these expresses all four parameters as impedances. It is usual to express the four parameters as a matrix;

$$\begin{bmatrix} V_1 \\ V_0 \end{bmatrix} = \begin{bmatrix} z(j\omega)_{11} & z(j\omega)_{12} \\ z(j\omega)_{21} & z(j\omega)_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_0 \end{bmatrix}$$

The matrix may be abbreviated to a representative element;

$[z(j\omega)]$ or just $[z]$

These concepts are capable of being extended to networks of more than two ports. However, this is rarely done in reality as in many practical cases ports are considered either purely input or purely output. If reverse direction transfer functions are ignored, a multi-port network can always be decomposed into a number of two-port networks.

Distributed components

Where a network is composed of discrete components, analysis using two-port networks is a matter of choice, not essential. The network can always alternatively be analysed in terms of its individual component transfer functions. However, if a network contains distributed components, such as in the case of a transmission line, then it is not possible to analyse in terms of individual components since they do not exist. The most common approach to this is to model the line as a two-port network and characterise it using two-port parameters (or something equivalent to them). Another example of this technique is modelling the carriers crossing the base region in a high frequency transistor. The base region has to be modelled as distributed resistance and capacitance rather than lumped components.

Image analysis

Transmission lines and certain types of filter design use the image method to determine their transfer parameters. In this method, the behaviour of an infinitely long cascade connected chain of identical networks is considered. The input and output impedances and the forward and reverse transmission functions are then calculated for this infinitely long chain. Although, the theoretical values so obtained can never be exactly realised in practice, in many cases they serve as a very good approximation for the behaviour of a finite chain as long as it is not too short.

Non-linear networks

Most electronic designs are, in reality, non-linear. There is very little that does not include some semiconductor devices. These are invariably non-linear, the transfer function of an ideal semiconductor pn junction is given by the very non-linear relationship;

$$i = I_o(e^{\frac{v}{V_T}} - 1)$$

where;

- i and v are the instantaneous current and voltage.
- I_o is an arbitrary parameter called the reverse leakage current whose value depends on the construction of the device.
- V_T is a parameter proportional to temperature called the thermal voltage and equal to about 25mV at room temperature.

There are many other ways that non-linearity can appear in a network. All methods utilising linear superposition will fail when non-linear components are present. There are several options for dealing with non-linearity depending on the type of circuit and the information the analyst wishes to obtain.

Constitutive equations

The diode equation above is an example of a constitutive equation of the general form,

$$f(v, i) = 0$$

This can be thought of as a non-linear resistor. The corresponding constitutive equations for non-linear inductors and capacitors are respectively;

$$f(v, \varphi) = 0$$
$$f(v, q) = 0$$

where f is any arbitrary function, φ is the stored magnetic flux and q is the stored charge.

Existence, uniqueness and stability

An important consideration in non-linear analysis is the question of uniqueness. For a network composed of linear components there will always be one, and only one, unique solution for a given set of boundary conditions. This is not always the case in non-linear circuits. For instance, a linear resistor with a fixed voltage applied to it has only one solution for the current through it. On the other hand, the non-linear tunnel diode has up to three solutions for the current for a given voltage. That is, a particular solution for the current through the diode is not unique, there may be others, equally valid. In some cases there may not be a solution at all: the question of existence of solutions must be considered.

Another important consideration is the question of stability. A particular solution may exist, but it may not be stable, rapidly departing from that point at the slightest stimulation. It can be shown that a network that is absolutely stable for all conditions must have one, and only one, solution for each set of conditions.

Methods

Boolean analysis of switching networks

A switching device is one where the non-linearity is utilised to produce two opposite states. CMOS devices in digital circuits, for instance, have their output connected to either the positive or the negative supply rail and are never found at anything in between except during a transient period when the device is actually switching. Here the non-linearity is designed to be extreme, and the analyst can actually take advantage of that fact. These kinds of networks can be analysed using Boolean algebra by assigning the two states ("on"/"off", "positive"/"negative" or whatever states are being used) to the boolean constants "0" and "1".

The transients are ignored in this analysis, along with any slight discrepancy between the actual state of the device and the nominal state assigned to a boolean value. For instance, boolean "1" may be assigned to the state of +5V. The output of the device may actually be +4.5V but the analyst still considers this to be boolean "1". Device manufacturers will usually specify a range of values in their data sheets that are to be considered undefined (ie the result will be unpredictable).

The transients are not entirely uninteresting to the analyst. The maximum rate of switching is determined by the speed of transition from one state to the other. Happily for the analyst, for many devices most of the transition occurs in the linear portion of the devices transfer function and linear analysis can be applied to obtain at least an approximate answer.

It is mathematically possible to derive boolean algebras which have more than two states. There is not too much use found for these in electronics, although three-state devices are passingly common.

Separation of bias and signal analyses

This technique is used where the operation of the circuit is to be essentially linear, but the devices used to implement it are non-linear. A transistor amplifier is an example of this kind of network. The essence of this technique is to separate the analysis in to two parts. Firstly, the dc biases are analysed using some non-linear method. This establishes the quiescent operating point of the circuit. Secondly, the small signal characteristics of the circuit are analysed using linear network analysis. Examples of methods that can be used for both these stages are given below.

Graphical method of dc analysis

In a great many circuit designs, the dc bias is fed to a non-linear component via a resistor (or possibly a network of resistors). Since resistors are linear components, it is particularly easy to determine the quiescent operating point of the non-linear device from a graph of its transfer function. The method is as follows: from linear network analysis the output transfer function (that is output voltage against output current) is calculated for the network of resistor(s) and the generator driving them. This will be a straight line and can readily be superimposed on the transfer function plot of the non-linear device. The point where the lines cross is the quiescent operating point.

Perhaps the easiest practical method is to calculate the (linear) network open circuit voltage and short circuit current and plot these on the transfer function of the non-linear device. The straight line joining these two point is the transfer function of the network.

In reality, the designer of the circuit would proceed in the reverse direction to that described. Starting from a plot provided in the manufacturers data sheet for the non-linear device, the designer would choose the desired operating point and then calculate the linear component values required to achieve it.

It is still possible to use this method if the device being biased has its bias fed through another device which is itself non-linear - a diode for instance. In this case however, the plot of the network transfer function onto the device being biased would no longer be a straight line and is consequently more tedious to do.

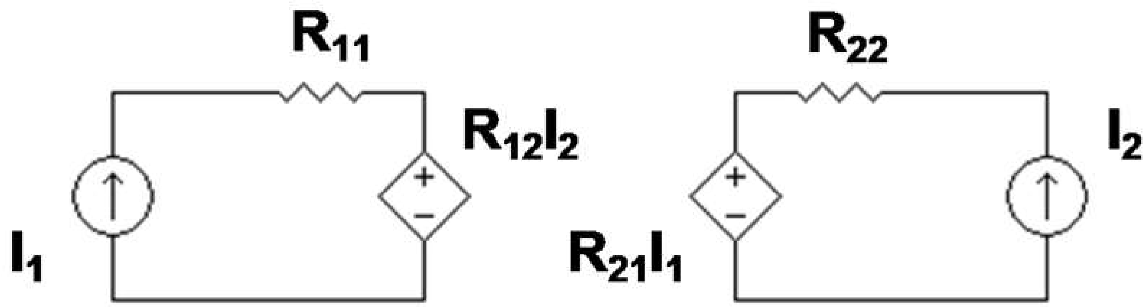
Small signal equivalent circuit

This method can be used where the deviation of the input and output signals in a network stay within a substantially linear portion of the non-linear devices transfer function, or else are so small that the curve of the transfer function can be considered linear. Under a set of these specific conditions, the non-linear device can be represented by an equivalent linear network. It must be remembered that this equivalent circuit is entirely notional and only valid for the small signal deviations. It is entirely inapplicable to the dc biasing of the device.

For a simple two-terminal device, the small signal equivalent circuit may be no more than two components. A resistance equal to the slope of the v/i curve at the operating point (called the dynamic resistance), and tangent to the curve. A generator, because this tangent will not, in general, pass through the origin. With more terminals, more complicated equivalent circuits are required.

A popular form of specifying the small signal equivalent circuit amongst transistor manufacturers is to use the two-port network parameters known as [h] parameters. These are a matrix of four parameters as with the [z] parameters but in the case of the [h] parameters they are a hybrid mixture of impedances, admittances, current gains and voltage gains. In this model the three terminal transistor is considered to be a two port network, one of its terminals being common to both ports. The [h] parameters are quite different depending on which terminal is chosen as the common one. The most important parameter for transistors is usually the forward current gain, h_{21}, in the common emitter configuration. This is designated h_{fe} on data sheets.

The small signal equivalent circuit in terms of two-port parameters leads to the concept of dependent generators. That is, the value of a voltage or current generator depends linearly on a voltage or current elsewhere in the circuit. For instance the [z] parameter model leads to dependent voltage generators as shown in this diagram;



[z] parameter equivalent circuit showing dependent voltage generators

There will always be dependent generators in a two-port parameter equivalent circuit. This applies to the [h] parameters as well as to the [z] and any other kind. These dependencies must be preserved when developing the equations in a larger linear network analysis.

Piecewise linear method

In this method, the transfer function of the non-linear device is broken up into regions. Each of these regions is approximated by a straight line. Thus, the transfer function will be linear up to a particular point where there will be a discontinuity. Past this point the transfer function will again be linear but with a different slope.

A well known application of this method is the approximation of the transfer function of a pn junction diode. The actual transfer function of an ideal diode has been given at the top of this (non-linear) section. However, this formula is rarely used in network analysis, a piecewise approximation being used instead. It can be seen that the diode current rapidly diminishes to $-I_o$ as the voltage falls. This current, for most purposes, is so small it can be ignored. With increasing voltage, the current increases exponentially. The diode is modelled as an open circuit up to the knee of the exponential curve, then past this point as a resistor equal to the bulk resistance of the semiconducting material.

The commonly accepted values for the transition point voltage are 0.7V for silicon devices and 0.3V for germanium devices. An even simpler model of the diode, sometimes used in switching applications, is short circuit for forward voltages and open circuit for reverse voltages.

The model of a forward biased pn junction having an approximately constant 0.7V is also a much used approximation for transistor base-emitter junction voltage in amplifier design.

The piecewise method is similar to the small signal method in that linear network analysis techniques can only be applied if the signal stays within certain bounds. If the signal crosses a discontinuity point then the model is no longer valid for linear analysis purposes. The model does have the advantage over small signal however, in that it is equally applicable to signal and dc bias. These can therefore both be analysed in the same operations and will be linearly superimposable.

Time-varying components

In linear analysis, the components of the network are assumed to be unchanging, but in some circuits this does not apply, such as sweep oscillators, voltage controlled amplifiers, and variable equalisers. In many circumstances the change in component value is periodic. A non-linear component excited with a periodic signal, for instance, can be represented as periodically varying *linear* component. Sidney Darlington disclosed a method of analysing such periodic time varying circuits. He developed canonical circuit forms which are analogous to the canonical forms of Ronald Foster and Wilhelm Cauer used for analysing linear circuits.

Chapter 2

Equivalent Impedance Transforms

An **equivalent impedance** is an equivalent circuit of an electrical network of impedance elements which presents the same impedance between all pairs of terminals as did the given network. Here we, describes mathematical transformations between some passive, linear impedance networks commonly found in electronic circuits.

There are a number of very well known and often used equivalent circuits in linear network analysis. These include resistors in series, resistors in parallel and the extension to series and parallel circuits for capacitors, inductors and general impedances. Also well known are the Norton and Thévenin equivalent current generator and voltage generator circuits respectively, as is the Y-Δ transform. None of these are discussed in detail here; the individual linked articles should be consulted.

The number of equivalent circuits that a linear network can be transformed into is unbounded. Even in the most trivial cases this can be seen to be true, for instance, by asking how many different combinations of resistors in parallel are equivalent to a given combined resistor. Wilhelm Cauer found a transformation that could generate all possible equivalents of a given rational, passive, linear one-port, or in other words, any given two-terminal impedance. Transformations of 4-terminal, especially 2-port, networks are also commonly found and transformations of yet more complex networks are possible.

The vast scale of the topic of equivalent circuits is underscored in a story told by Sidney Darlington. According to Darlington, a large number of equivalent circuits were found by Ronald Foster, following his and George Campbell's 1920 paper on non-dissipative four-ports. In the course of this work they looked at the ways four ports could be interconnected with ideal transformers. They found a number of combinations which might have practical applications and asked the AT&T patent department to have them patented. The patent department replied that it was pointless just patenting some of the circuits if a competitor could use an equivalent circuit to get around the patent; they should patent all of them or not bother. Foster therefore set to work calculating every last one of them. He arrived at an enormous total of 83,539 equivalents. This was too many to patent, so instead the information was released into the public domain in order to prevent any of AT&T's competitors from patenting them in the future.

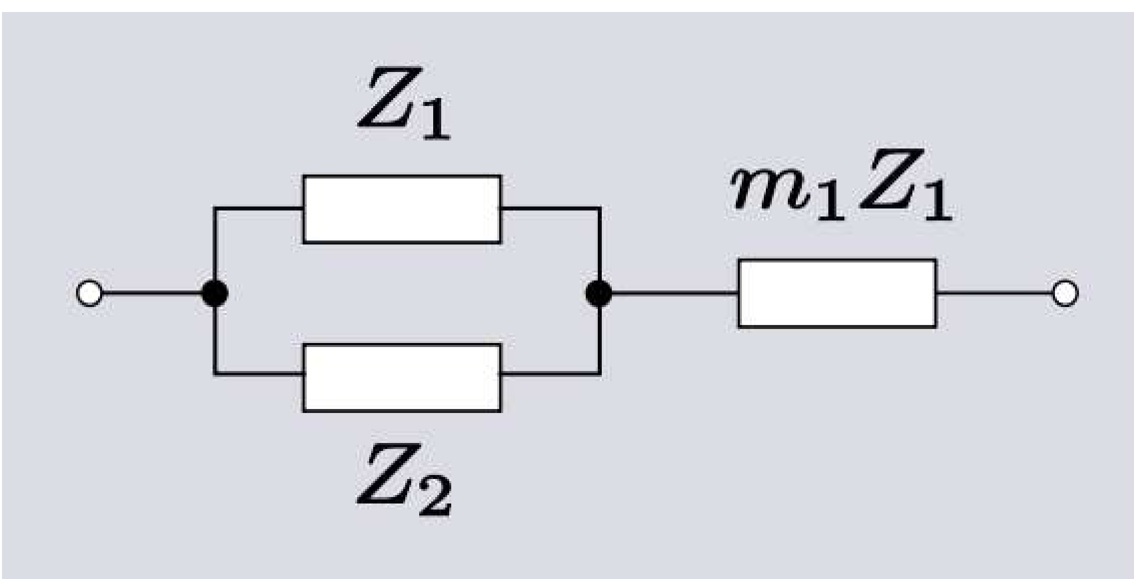
2-terminal, 2-element-kind networks

A single impedance has two terminals to connect to the outside world, hence can be described as a 2-terminal, or a one-port, network. Despite the simple description, there is no limit to the number of meshes, and hence complexity and number of elements, that the impedance network may have. 2-element-kind networks are common in circuit design; filters, for instance, are often LC-kind networks and printed circuit designers favour RC-kind networks because inductors are less easy to manufacture. Transformations are simpler and easier to find than for 3-element-kind networks. One-element-kind networks can be thought of as a special case of two-element-kind. It is possible to use the transformations in this section on a certain few 3-element-kind networks by substituting a network of elements for element Z_n. However, this is limited to a maximum of two impedances being substituted; the remainder will not be a free choice. All the transformation equations given in this section are due to Otto Zobel.

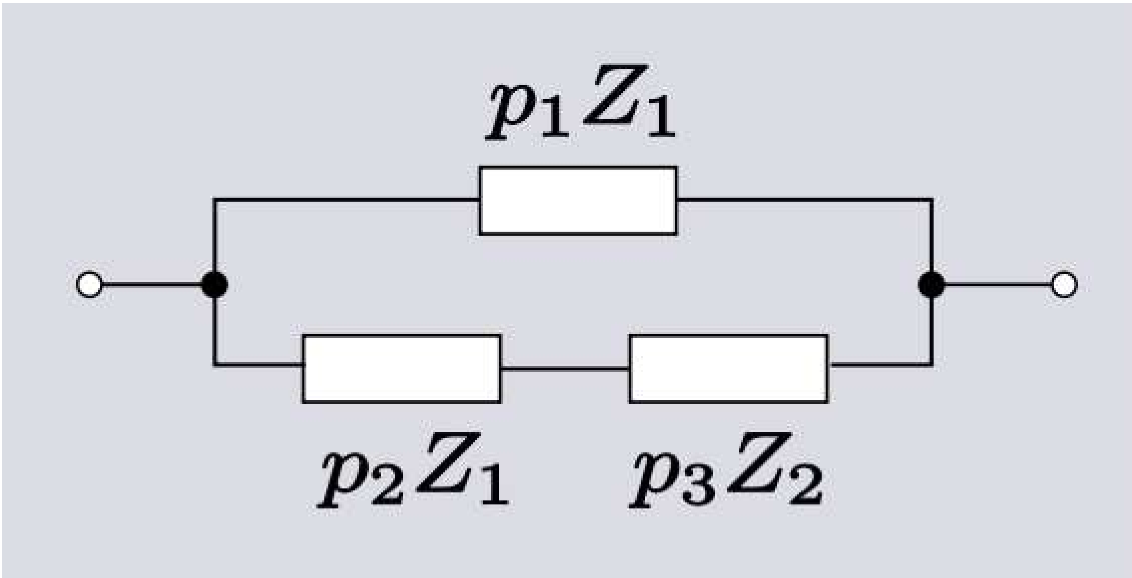
3-element networks

One-element networks are trivial and two-element, two-terminal networks are either two elements in series or two elements in parallel, also trivial. The smallest number of elements that is non-trivial is three, and there are two 2-element-kind non-trivial transformations possible, one being both the reverse transformation and the topological dual, of the other.

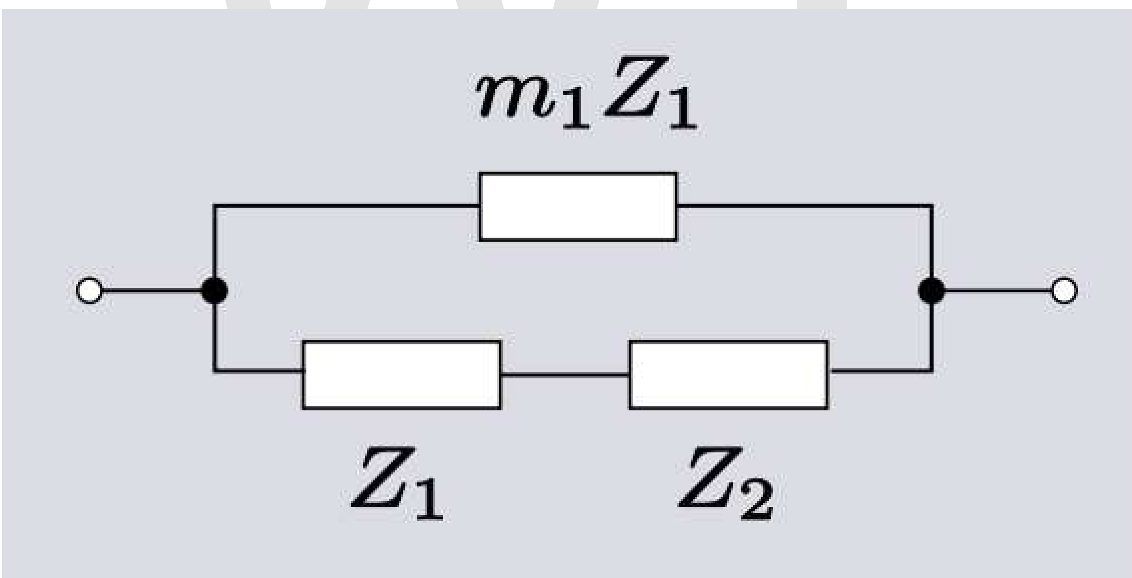
Description
Network
Transform equations
Transformed network
Transform 1.1
Transform 1.2 is the reverse of this transform.



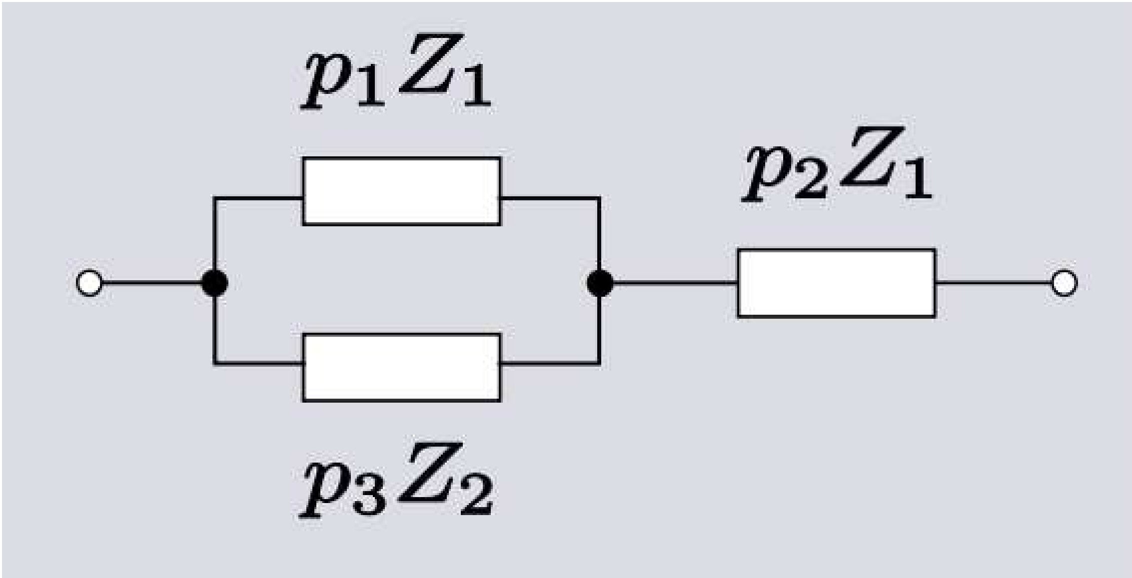
$$p_1 = 1 + m_1\ ,$$
$$p_2 = m_1(1 + m_1)\ ,$$
$$p_3 = (1 + m_1)^2\ .$$



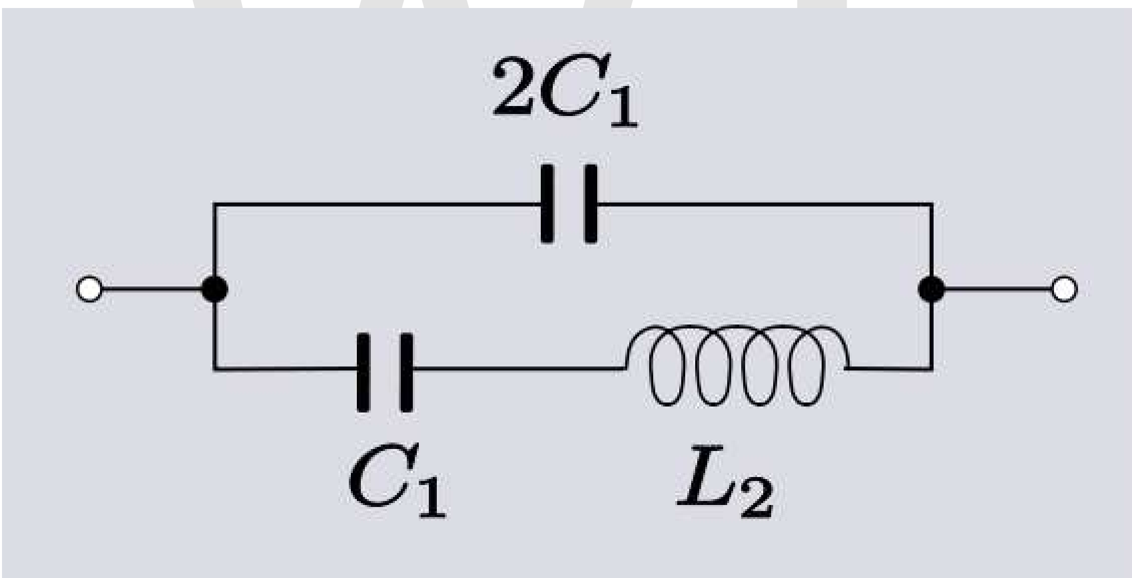
Transform 1.2
The reverse transform, and topological dual, of Transform 1.1.



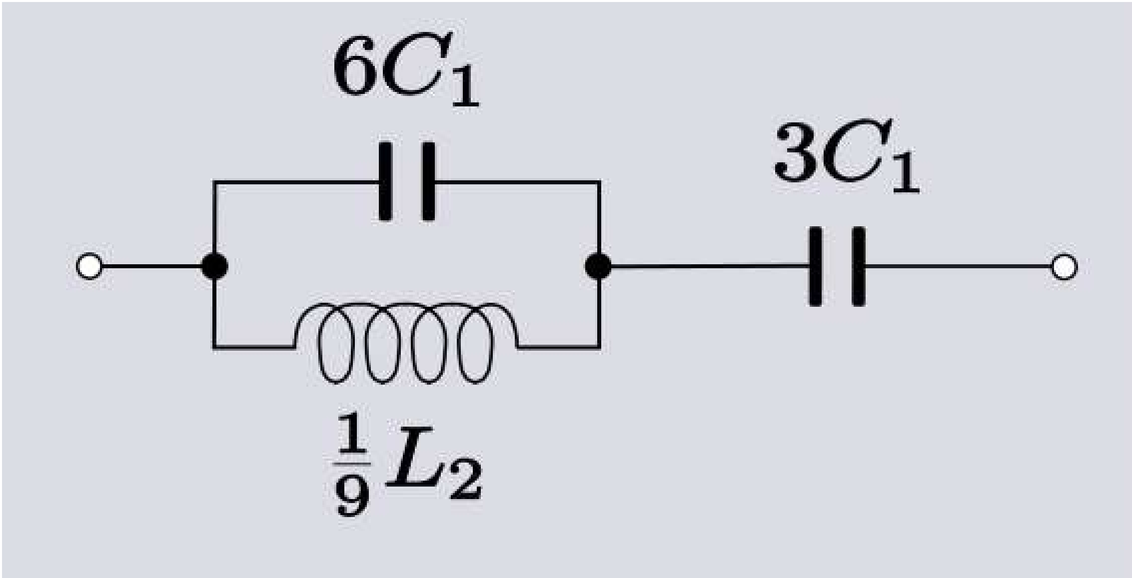
$$p_1 = \frac{{m_1}^2}{1 + m_1}\ ,$$
$$p_2 = \frac{m_1}{1 + m_1}\ ,$$
$$p_3 = \left(\frac{m_1}{1 + m_1}\right)^2\ .$$



Example 1.
An example of Transform 1.2. The reduced size of the inductor has practical advantages.



$m_1 = 0.5$,
$p_1 = \frac{1}{6}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{9}$.

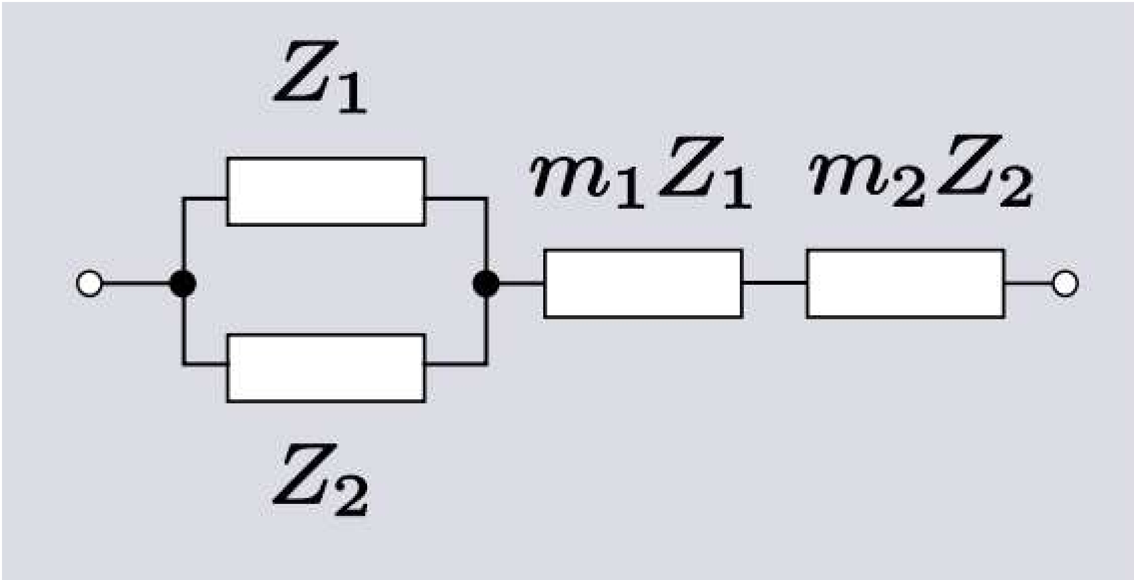


4-element networks

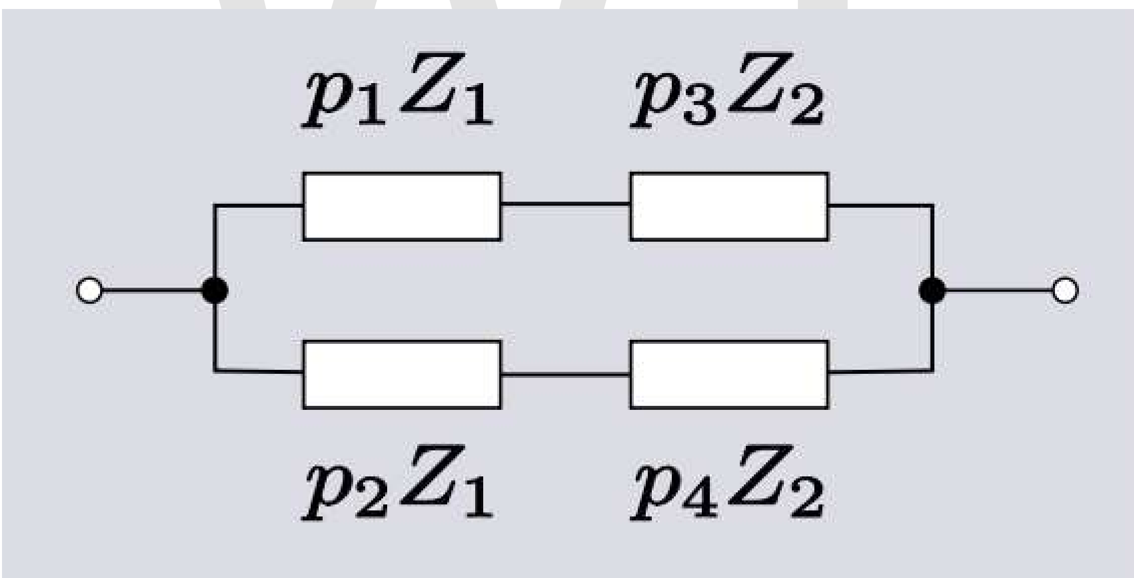
There are four non-trivial 4-element transformations for 2-element-kind networks. Two of these are the reverse transformations of the other two and two are the dual of a different two. Further transformations are possible in the special case of Z_2 being made the same element kind as Z_1, that is, when the network is reduced to one-element-kind. The number of possible networks continues to grow as the number of elements is increased. For all entries in the following table it is defined:

$$q_1 := 1 + m_1 + m_2,$$
$$q_2 := \sqrt{q_1{}^2 - 4m_1m_2},$$
$$q_3 := \frac{(1+m_1)(1+m_2)}{(m_1-m_2)^2},$$
$$q_4 := \frac{q_2 - q_1 + 2m_2}{2q_2},$$
$$q_5 := \frac{q_2 + q_1 - 2m_2}{2q_2}.$$

Description
Network
Transform equations
Transformed network
Transform 2.1
Transform 2.2 is the reverse of this transform. Transform 2.3 is the topological dual of this transform.

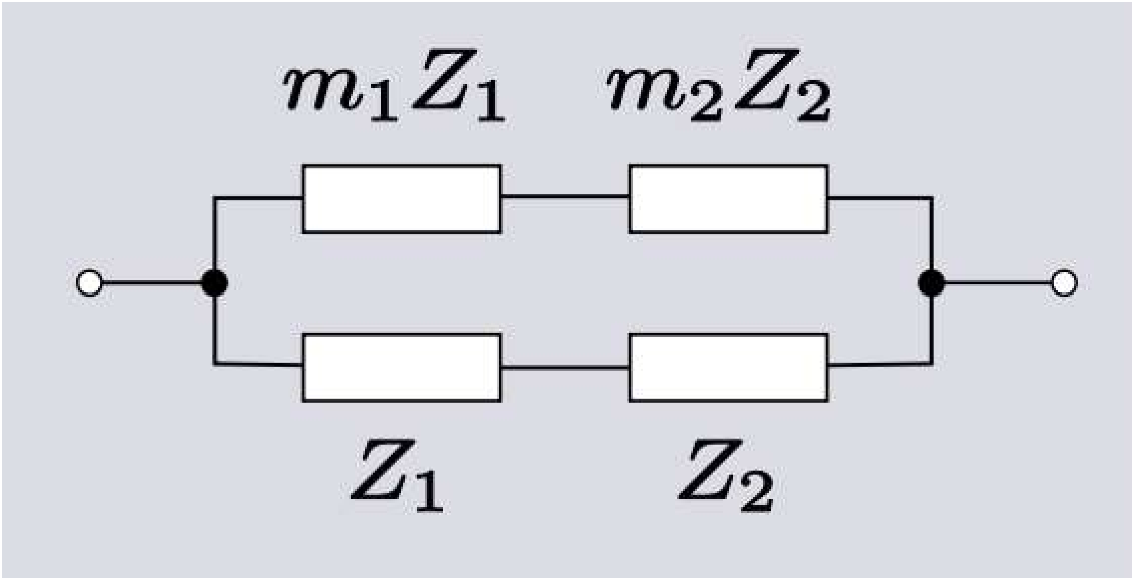


$p_1 = \frac{q_1 + q_2}{2q_5}$, $p_2 = \frac{q_1 - q_2}{2q_4}$, $p_3 = \frac{m_2}{q_5}$, $p_4 = \frac{m_2}{q_4}$.

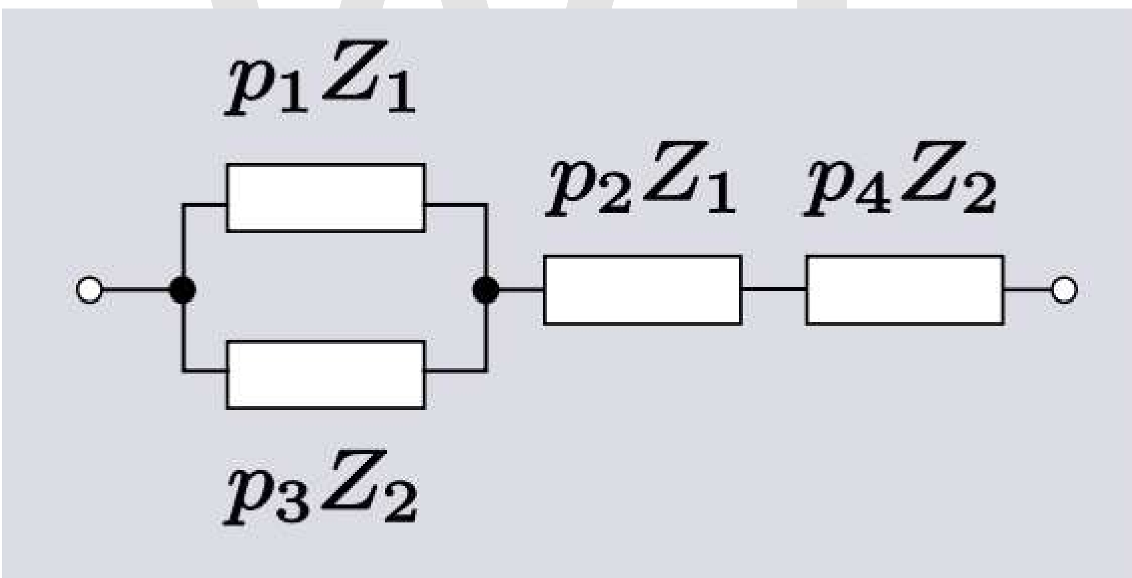


Transform 2.2

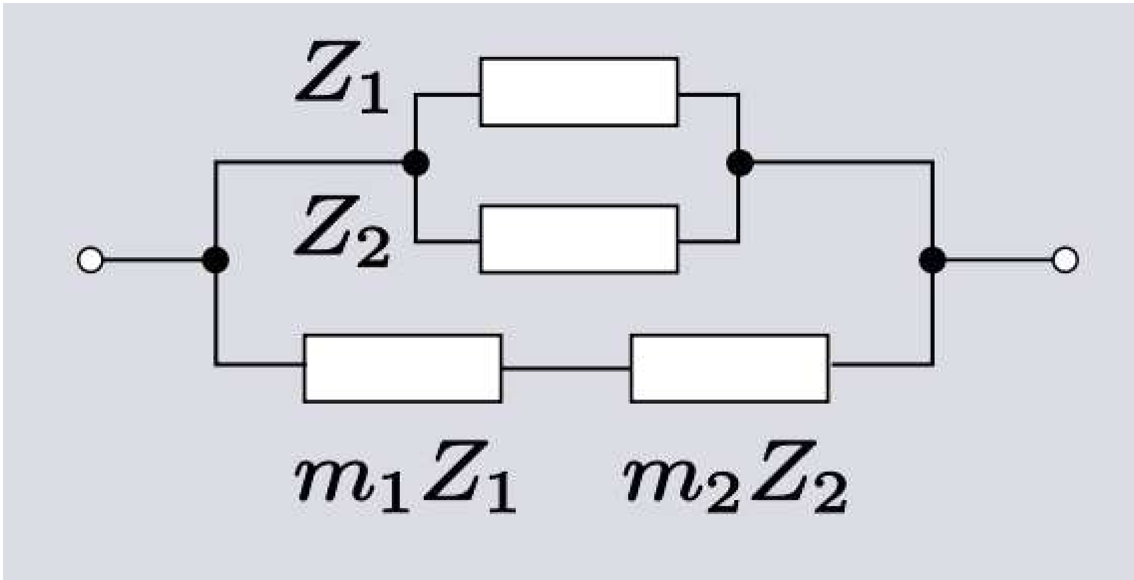
Transform 2.1 is the reverse of this transform. Transform 2.4 is the topological dual of this transform.



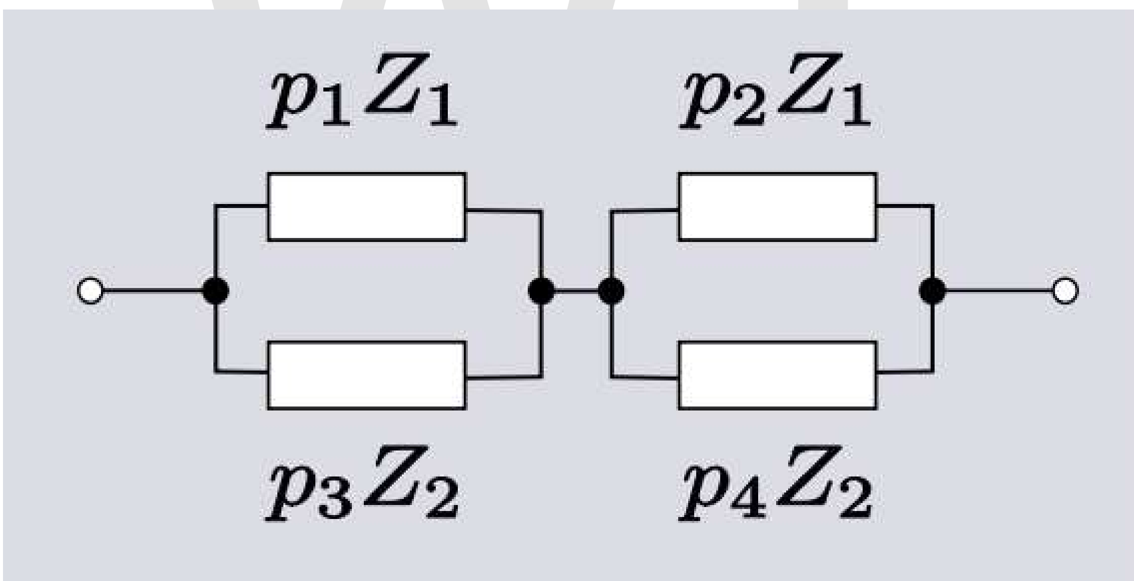
$$p_1 = \frac{1}{q_3(1+m_2)} \; , \; p_2 = \frac{m_1}{1+m_1} \; , \; p_3 = \frac{1}{q_3(1+m_1)} \; , \; p_4 = \frac{m_2}{1+m_2} \; .$$



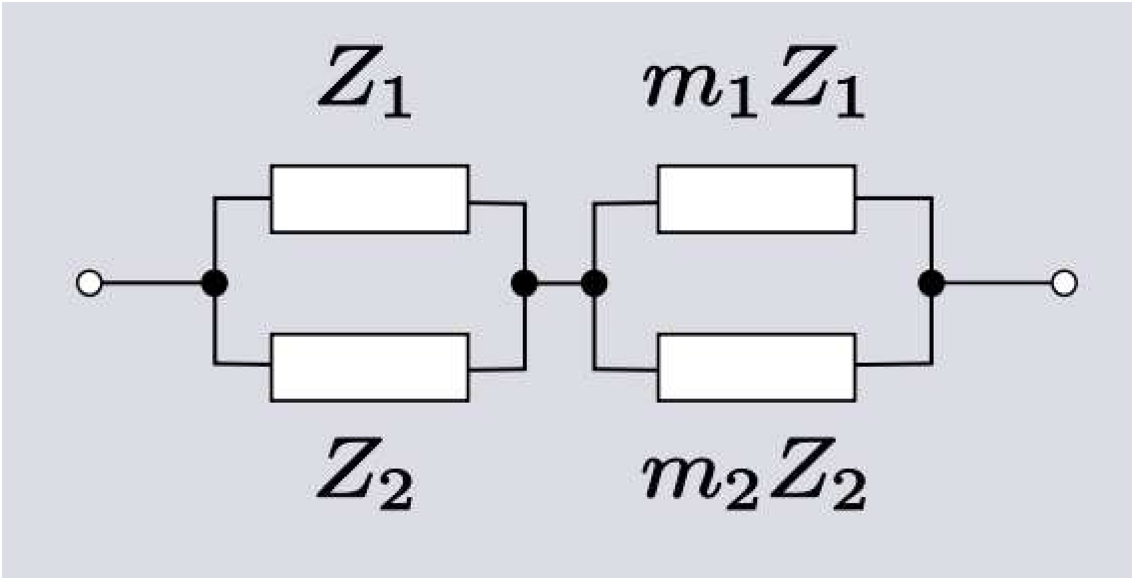
Transform 2.3
Transform 2.4 is the reverse of this transform. Transform 2.1 is the topological dual of this transform.



$p_1 = \frac{q_4(q_1+q_2)}{2m_2}$, $p_2 = \frac{q_5(q_1-q_2)}{2m_2}$, $p_3 = q_4$, $p_4 = q_5$.

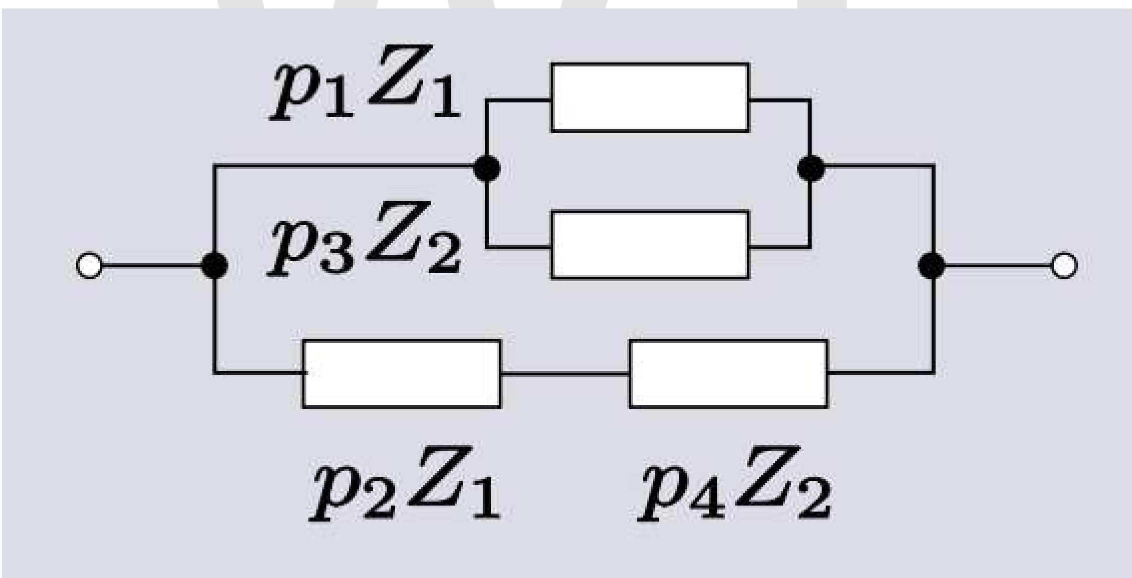


Transform 2.4
Transform 2.3 is the reverse of this transform. Transform 2.2 is the topological dual of this transform.

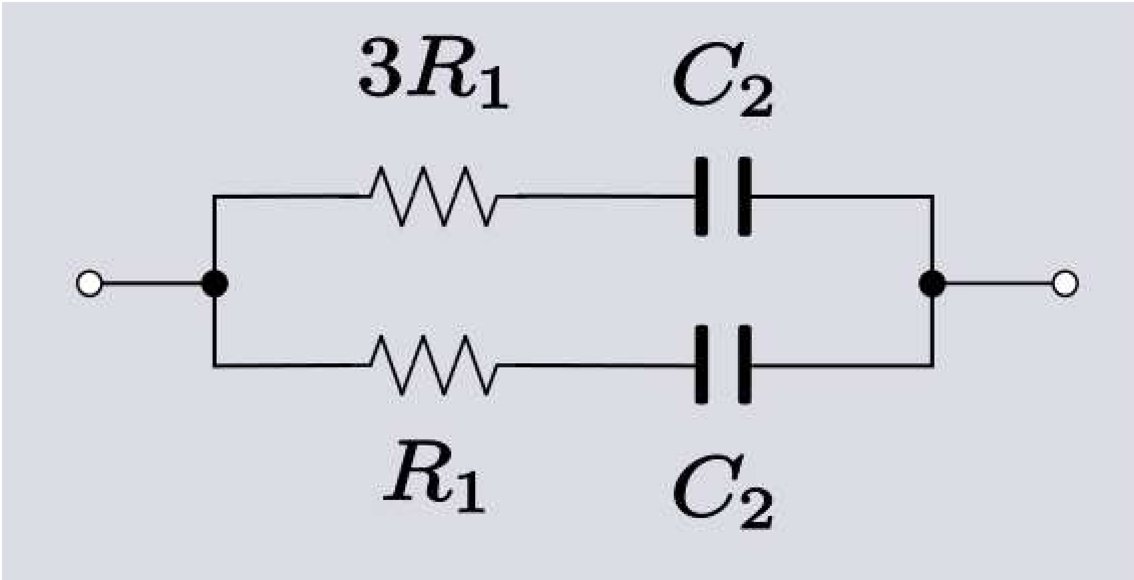


$p_1 = 1 + m_1$, $p_2 = m_1 q_3(1 + m_1)$, $p_3 = 1 + m_2$,

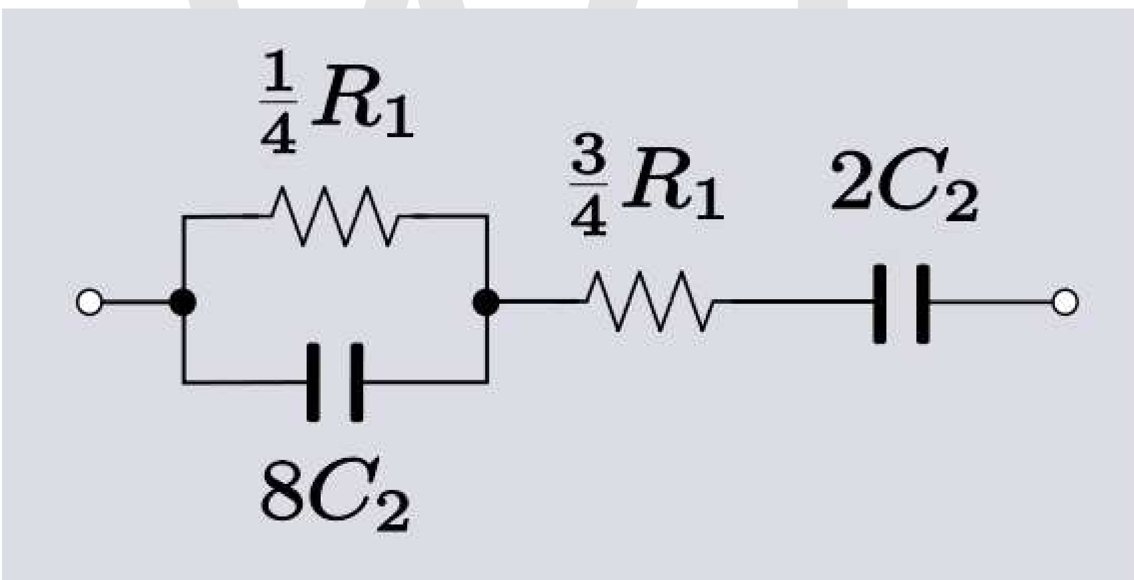
$p_4 = m_1 q_3(1 + m_2)$.



Example 2.
An example of Transform 2.2.



$m_1 = 3$, $m_2 = 1$, $q_3 = 2$, $p_1 = \frac{1}{4}$, $p_2 = \frac{3}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{2}$.



2-terminal, *n*-element, 3-element-kind networks

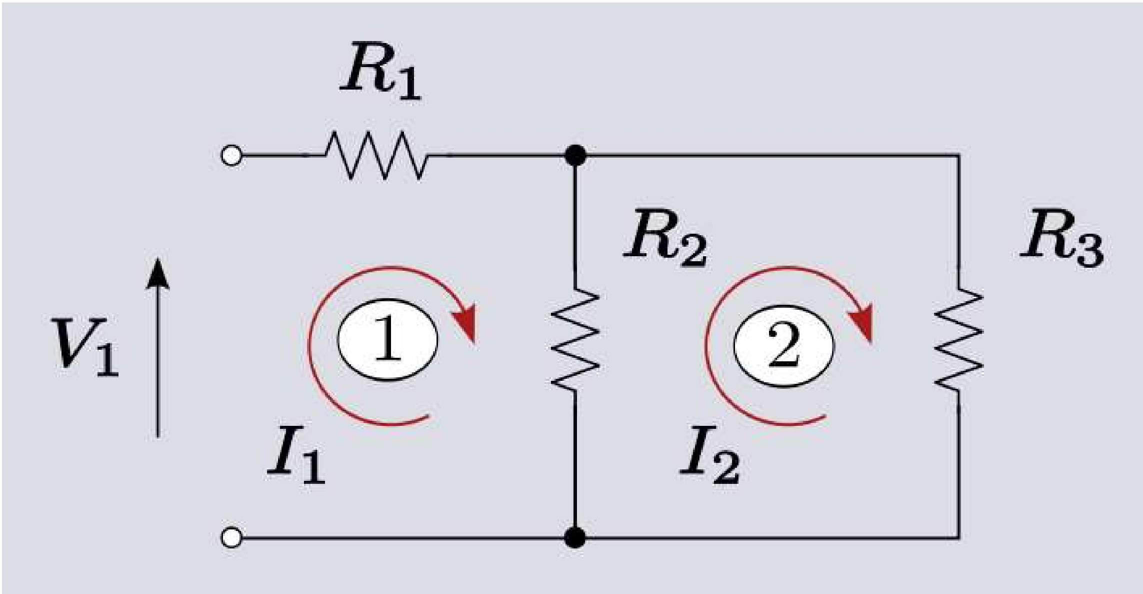


Fig. 1. Simple example of a network of impedances using resistors only for clarity. However, analysis of networks with other impedance elements proceed by the same principles. Two meshes are shown, with numbers in circles. The sum of impedances around each mesh, p, will form the diagonal of the entries of the matrix, Z_{pp}. The impedance of branches shared by two meshes, p and q, will form the entries -Z_{pq}. Z_{pq}, p≠q, will always have a minus sign provided that the convention of loop currents are defined in the same (conventionally counter-clockwise) direction and the mesh contains no ideal transformers or mutual inductors.

Simple networks with just a few elements can be dealt with by formulating the network equations "by hand" with the application of simple network theorems such as Kirchhoff's laws. Equivalence is proved between two networks by directly comparing the two sets of equations and equating coefficients. For large networks more powerful techniques are required. A common approach is to start by expressing the network of impedances as a matrix. This approach is only good for rational networks. Any network that includes distributed elements, such as a transmission line, cannot be represented by a finite matrix. Generally, an *n*-mesh network requires an *n*x*n* matrix to represent it. For instance the matrix for a 3-mesh network might look like;

$$[\mathbf{Z}] = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix}$$

The entries of the matrix are chosen so that the matrix forms a system of linear equations in the mesh voltages and currents (as defined for mesh analysis);

$$[\mathbf{V}] = [\mathbf{Z}][\mathbf{I}]$$

The example diagram in Figure 1, for instance, can be represented as an impedance matrix by;

$$[\mathbf{Z}] = \begin{bmatrix} R_1 + R_2 & -R_2 \\ -R_2 & R_2 + R_3 \end{bmatrix}$$

and the associated system of linear equations are,

$$\begin{bmatrix} V_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 + R_2 & -R_2 \\ -R_2 & R_2 + R_3 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$

In the most general case, each branch, Z_{p}, of the network may be made up of three elements so that,

$$Z_{\mathrm{p}} = sL_{\mathrm{p}} + R_{\mathrm{p}} + \frac{1}{sC_{\mathrm{p}}}$$

where *L*, *R* and *C* represent inductance, resistance, and capacitance respectively and *s* is the complex frequency operator $s=\sigma+i\omega$.

This is the conventional way of representing a general impedance but for the purposes it is mathematically more convenient to deal with elastance, *D*, the inverse of capacitance, *C*. In those terms the general branch impedance can be represented by,

$$sZ_{\mathrm{p}} = s^2 L_{\mathrm{p}} + sR_{\mathrm{p}} + D_{\mathrm{p}}$$

Likewise, each entry of the impedance matrix can consist of the sum of three elements. Consequently, the matrix can be decomposed into three *n*x*n* matrices, one for each of the three element kinds;

$$s[\mathbf{Z}] = s^2[\mathbf{L}] + s[\mathbf{R}] + [\mathbf{D}]$$

It is desired that the matrix [**Z**] represent an impedance, $Z(s)$. For this purpose, the loop of one of the meshes is cut and $Z(s)$ is the impedance measured between the points so cut. It is conventional to assume the external connection port is in mesh 1, and is therefore connected across matrix entry Z_{11}, although it would be perfectly possible to formulate this with connections to any desired nodes. In the following discussion $Z(s)$ taken across Z_{11} is assumed. $Z(s)$ may be calculated from [**Z**] by;

$$Z(s) = \frac{|\mathbf{Z}|}{z_{11}}$$

where z_{11} is the complement of Z_{11} and |**Z**| is the determinant of [**Z**].

For the example network above;

$$|\mathbf{Z}| = (R_1 + R_2)(R_2 + R_3) - {R_2}^2 = R_1R_2 + R_1R_3 + R_2R_3 \ ,$$
$$z_{11} = Z_{22} = R_2 + R_3 \text{ ,and,}$$
$$Z(s) = R_1 + \frac{R_2R_3}{R_2 + R_3} \ .$$

This result is easily verified to be correct by the more direct method of resistors in series and parallel. However, such methods rapidly become tedious and cumbersome with the growth of the size and complexity of the network under analysis.

The entries of [**R**], [**L**] and [**D**] cannot be set arbitrarily. For [**Z**] to be able to realise the impedance *Z*(*s*) then [**R**],[**L**] and [**D**] must all be positive-definite matrices. Even then, the realisation of *Z*(*s*) will, in general, contain ideal transformers within the network. Finding only those transforms that do not require mutual inductances or ideal transformers is a more difficult task. Similarly, if starting from the "other end" and specifying an expression for *Z*(*s*), this again cannot be done arbitrarily. To be realisable as a rational impedance, *Z*(*s*) must be positive-real. The positive-real (PR) condition is both necessary and sufficient but there may be practical reasons for rejecting some topologies.

A general impedance transform for finding equivalent rational one-ports from a given instance of [**Z**] is due to Wilhelm Cauer. The group of real affine transformations,

$$[\mathbf{Z}'] = [\mathbf{T}]^T [\mathbf{Z}][\mathbf{T}]$$

where,

$$[\mathbf{T}] = \begin{bmatrix} 1 & 0 \cdots 0 \\ T_{21} & T_{22} \cdots T_{2n} \\ \cdot & \cdots \\ T_{n1} & T_{n2} \cdots T_{nn} \end{bmatrix}$$

is invariant in *Z*(*s*). That is, all the transformed networks are equivalents according to the definition given here. If the *Z*(*s*) for the initial given matrix is realisable, that is, it meets the PR condition, then all the transformed networks produced by this transformation will also meet the PR condition.

3 and 4-terminal networks

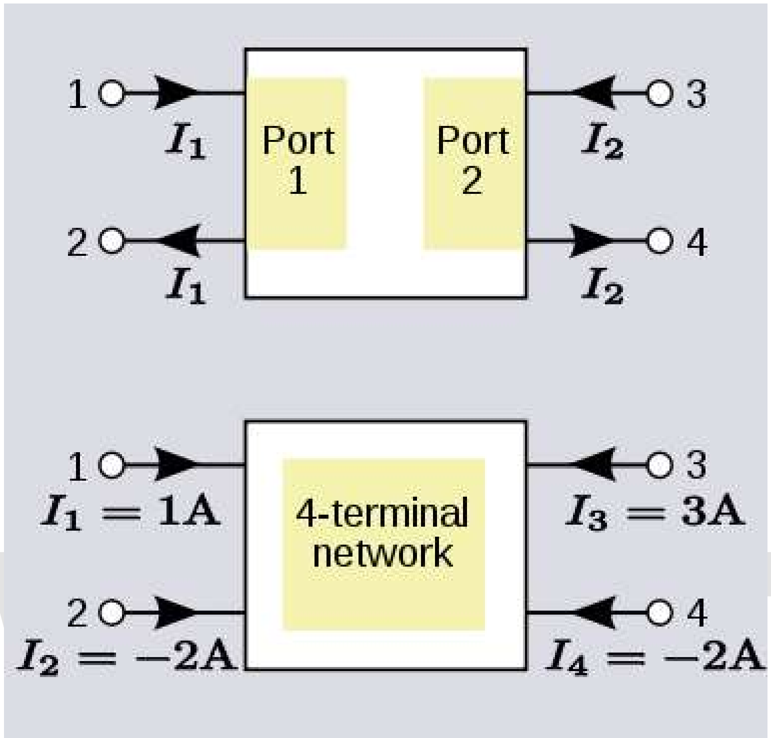
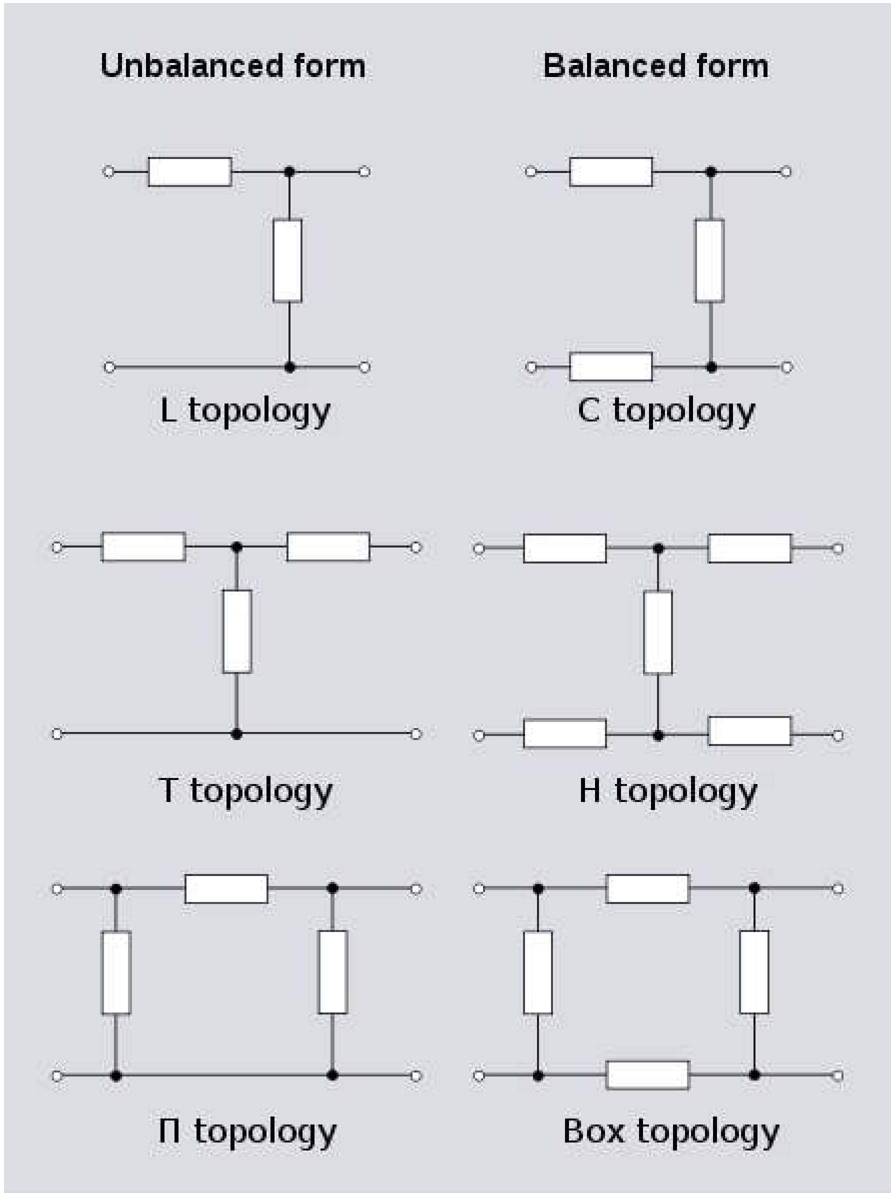


Fig. 2. A 4-terminal network connected by ports (top) has equal and opposite currents in each pair of terminals. The bottom network does not meet the port condition and cannot be treated as a 2-port. It could, however, be treated as an unbalanced 3-port by splitting one of the terminals into three common terminals shared between the ports.

When discussing 4-terminal networks, network analysis often proceeds in terms of 2-port networks, which covers a vast array of practically useful circuits. "2-port", in essence, refers to the way the network has been connected to the outside world: that the terminals have been connected in pairs to a source or load. It is possible to take exactly the same network and connect it to external circuitry in such a way that it is no longer behaving as a 2-port. This idea is demonstrated in Figure 2.



Equivalent unbalanced and balanced networks. The impedance of the series elements in the balanced version is half the corresponding impedance of the unbalanced version.

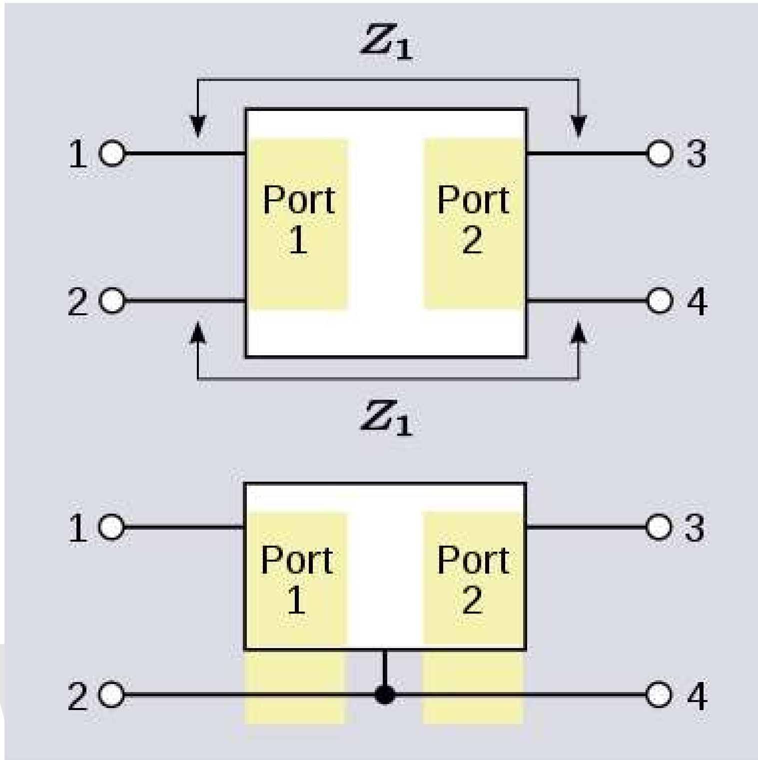


Fig. 3. To be balanced, a network must have the same impedance in each "leg" of the circuit.

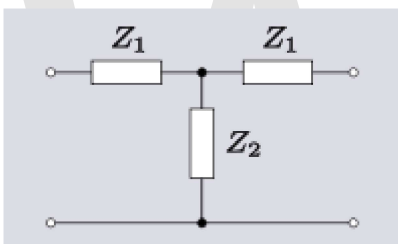
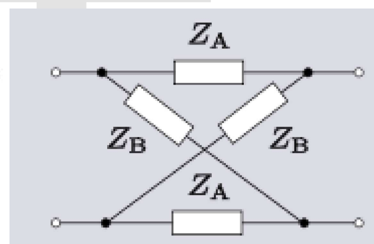
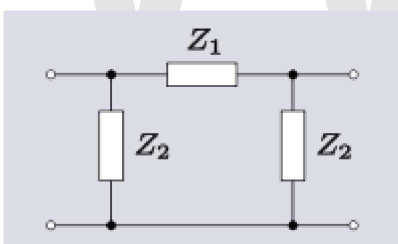
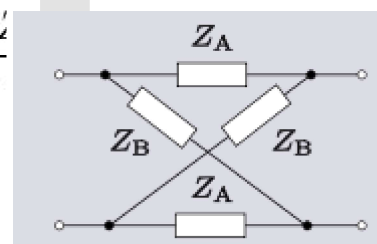
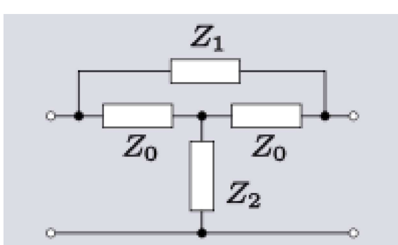
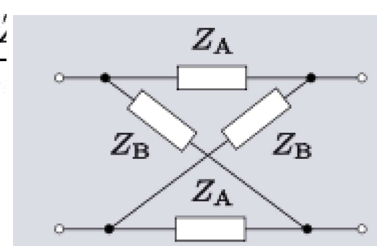
A 3-terminal network can also be used as a 2-port. To achieve this, one of the terminals is connected in common to one terminal of both ports. In other words, one terminal has been split into two terminals and the network has effectively been converted to a 4-terminal network. This topology is known as unbalanced topology and is opposed to balanced topology. Balanced topology requires, referring to Figure 3, that the impedance measured between terminals 1 and 3 is equal to the impedance measured between 2 and 4. This is the pairs of terminals *not* forming ports: the case where the pairs of terminals forming ports have equal impedance is referred to as symmetrical. Strictly speaking, any network that does not meet the balance condition is unbalanced, but the term is most often referring to the 3-terminal topology described above and in Figure 3. Transforming an unbalanced 2-port network into a balanced network is usually quite straightforward: all series connected elements are divided in half with one half being relocated in what was the common branch. Transforming from balanced to unbalanced topology will often be possible with the reverse transformation but there are certain cases of certain topologies which cannot be transformed in this way.

An example of a 3-terminal network transform that is not restricted to 2-ports is the Y-Δ transform. This is a particularly important transform for finding equivalent impedances. Its importance arises from the fact that the total impedance between two terminals cannot be determined solely by calculating series and parallel combinations except for a certain restricted class of network. In the general case additional transformations are required. The Y-Δ transform, its inverse the Δ-Y transform, and the *n*-terminal analogues of these two transforms (star-polygon transforms) represent the minimal additional transforms required to solve the general case. Series and parallel are, in fact, the 2-terminal versions of star and polygon topology. A common simple topology that cannot be solved by series and parallel combinations is the input impedance to a bridge network (except in the special case when the bridge is in balance). The rest of the transforms in this section are all restricted to use with 2-ports only.

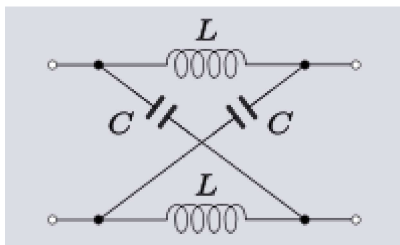
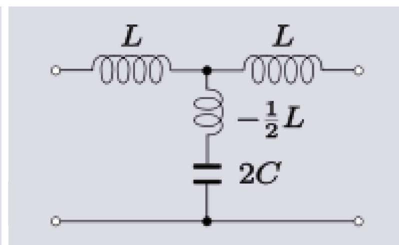
Lattice transforms

Symmetric 2-port networks can be transformed into lattice networks using Bartlett's bisection theorem. The method is limited to symmetric networks but this includes many topologies commonly found in filters, attenuators and equalisers. The lattice topology is intrinsically balanced, there is no unbalanced counterpart to the lattice and it will usually require more components than the transformed network.

Some common networks transformed to lattices (X-networks)

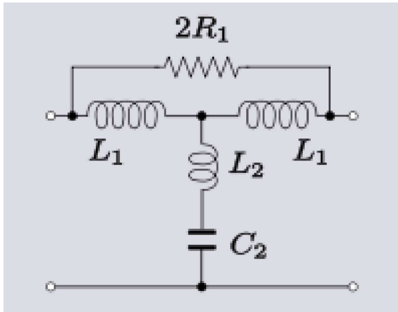
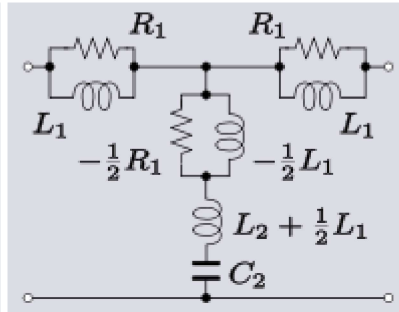
Description	Network	Transform equations	Transformed network
Transform 3.1 Transform of T network to lattice network.		$Z_A = Z_1$, $Z_B = Z_1 +$	
Transform 3.2 Transform of Π network to lattice network.		$Z_A = \frac{Z_1}{Z_1 +}$ $Z_B = Z_2$.	
Transform 3.3 Transform of Bridged-T network to lattice network.		$Z_A = \frac{Z_1}{Z_1 +}$ $Z_B = Z_0 +$	

Reverse transformations from a lattice to an unbalanced topology are not always possible in terms of passive components. For instance, this transform,

Description	Network	Transformed network
Transform 3.4 Transform of a lattice phase equaliser to a T network.		

cannot be realised with passive components because of the negative values arising in the transformed circuit. It can however be realised if mutual inductances and ideal transformers are permitted, for instance, in this circuit. Another possibility is to permit the use of active components which would enable negative impedances to be directly realised as circuit components.

It can sometimes be useful to make such a transformation, not for the purposes of actually building the transformed circuit, but rather, for the purposes of aiding understanding of how the original circuit is working. The following circuit in bridged-T topology is a modification of a mid-series m-derived filter T-section. The circuit is due to Hendrik Bode who claims that the addition of the bridging resistor of a suitable value will cancel the parasitic resistance of the shunt inductor. The action of this circuit is clear if it is transformed into T topology - in this form there is a negative resistance in the shunt branch which can be made to be exactly equal to the positive parasitic resistance of the inductor.

<table>
<tr><th>Description</th><th>Network</th><th>Transformed network</th></tr>
<tr><td>Transform 3.5
Transform of a bridged-T low-pass filter section to a T-section.</td><td>
</td><td>
</td></tr>
</table>

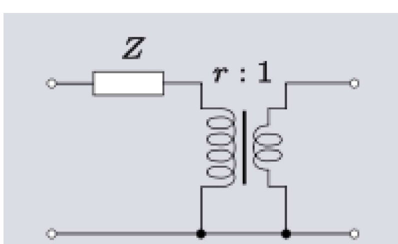
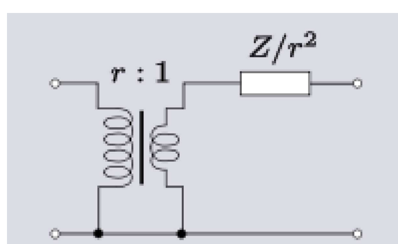
Any symmetrical network can be transformed into any other symmetrical network by the same method, that is, by first transforming into the intermediate lattice form (omitted for clarity from the above example transform) and from the lattice form into the required target form. As with the example, this will generally result in negative elements except in special cases.

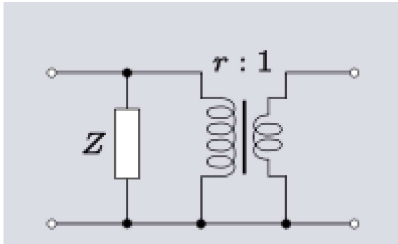
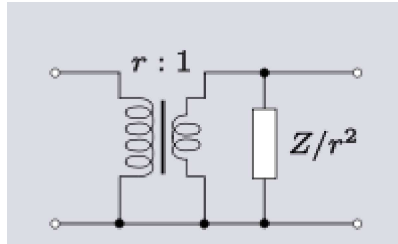
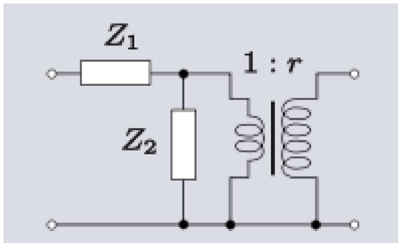
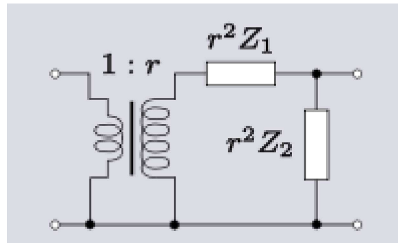
Eliminating resistors

A theorem due to Sidney Darlington states that any PR function $Z(s)$ can be realised as a lossless two-port terminated in a positive resistor R. That is, regardless of how many resistors feature in the matrix [**Z**] representing the impedance network, a transform can be found that will realise the network entirely as an LC-kind network with just one resistor across the output port (which would normally represent the load). No resistors within the network are necessary in order to realise the specified response. Consequently, it is always possible to reduce 3-element-kind 2-port networks to 2-element-kind (LC) 2-port networks provided the output port is terminated in a resistance of the required value.

Eliminating ideal transformers

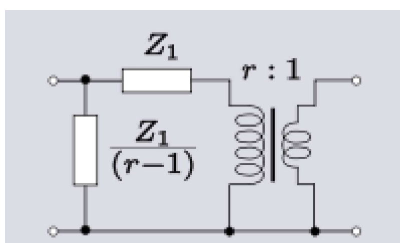
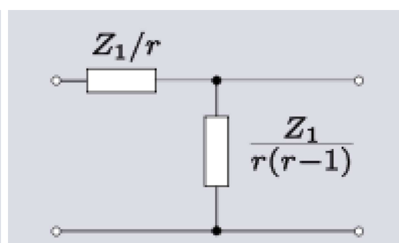
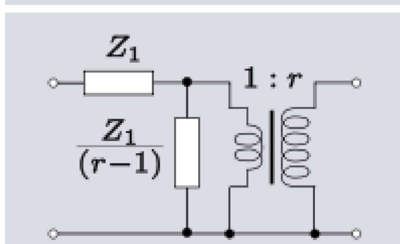
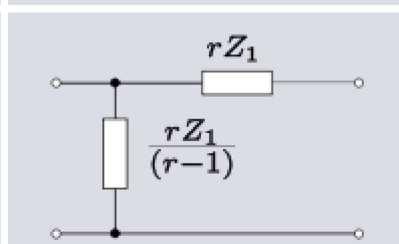
An elementary transformation that can be done with ideal transformers and some other impedance element is to shift the impedance to the other side of the transformer. In all the following transforms, r is the turns ratio of the transformer.

<table>
<tr><th>Description</th><th>Network</th><th>Transformed network</th></tr>
<tr><td>Transform 4.1
Series impedance through a step-down transformer.</td><td>
</td><td>
</td></tr>
</table>

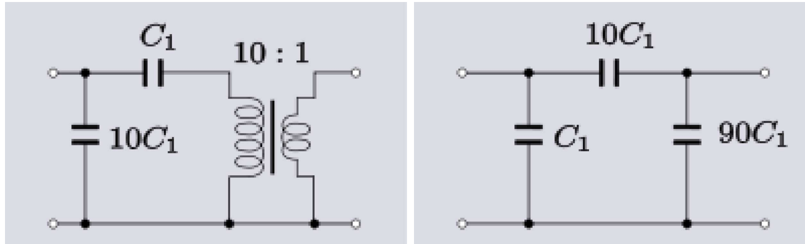
Transform 4.2 Shunt impedance through a step-down transformer.		
Transform 4.3 Shunt and series impedance network through a step-up transformer.		

These transforms do not just apply to single elements; entire networks can be passed through the transformer. In this manner, the transformer can be shifted around the network to a more convenient location.

Darlington gives an equivalent transform that can eliminate an ideal transformer altogether. This technique requires that the transformer is next to (or capable of being moved next to) an "L" network of same-kind impedances. The transform in all variants results in the "L" network facing the opposite way, that is, topologically mirrored.

Description	Network	Transformed network
Transform 5.1 Elimination of a step-down transformer.		
Transform 5.2 Elimination of a step-up transformer.		

Example 3.
Example of transform 5.1.



Example 3 shows the result is a Π-network rather than an L-network. The reason for this is that the shunt element has more capacitance than is required by the transform so some is still left over after applying the transform. If the excess were instead, in the element nearest the transformer, this could be dealt with by first shfting the excess to the other side of the transformer before carrying out the transform.

Chapter 3

Voltage Divider

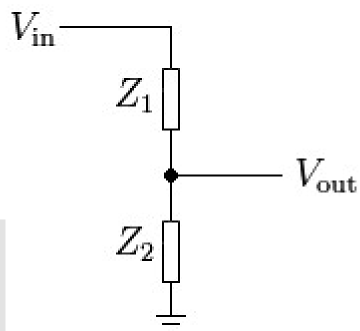


Figure 1: Voltage divider

In electronics, a **voltage divider** (also known as a **potential divider**) is a simple linear circuit that produces an output voltage (V_{out}) that is a fraction of its input voltage (V_{in}). **Voltage division** refers to the partitioning of a voltage among the components of the divider.

The formula governing a voltage divider is similar to that for a current divider, but the ratio describing voltage division places the selected impedance in the numerator, unlike current division where it is the unselected components that enter the numerator.

A simple example of a voltage divider consists of two resistors in series or a potentiometer. It is commonly used to create a reference voltage, and may also be used as a signal attenuator at low frequencies.

General case

A voltage divider referenced to ground is created by connecting two electrical impedances in series, as shown in Figure 1. The input voltage is applied across the series impedances Z_1 and Z_2 and the output is the voltage across Z_2. Z_1 and Z_2 may be composed of any combination of elements such as resistors, inductors and capacitors.

Applying Ohm's Law, the relationship between the input voltage, V_{in}, and the output voltage, V_{out}, can be found:

$$V_{\text{out}} = \frac{Z_2}{Z_1 + Z_2} \cdot V_{\text{in}}$$

Proof:

$$V_{\text{in}} = I \cdot (Z_1 + Z_2)$$
$$V_{\text{out}} = I \cdot Z_2$$
$$I = \frac{V_{\text{in}}}{Z_1 + Z_2}$$
$$V_{\text{out}} = V_{\text{in}} \cdot \frac{Z_2}{Z_1 + Z_2}$$

The transfer function (also known as the divider's **voltage ratio**) of this circuit is simply:

$$H = \frac{V_{out}}{V_{in}} = \frac{Z_2}{Z_1 + Z_2}$$

In general this transfer function is a complex, rational function of frequency.

Resistive divider

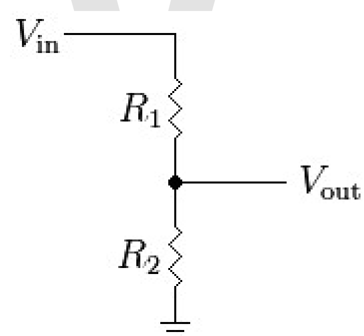


Figure 2: Simple resistive voltage divider

A resistive divider is a special case where both impedances, Z_1 and Z_2, are purely resistive (Figure 2).

Substituting $Z_1 = R_1$ and $Z_2 = R_2$ into the previous expression gives:

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} \cdot V_{\text{in}}$$

As in the general case, R_1 and R_2 may be any combination of series/parallel resistors.

Examples

Resistive divider

As a simple example, if $R_1 = R_2$ then

$$V_{\text{out}} = \frac{1}{2} \cdot V_{\text{in}}$$

As a more specific and/or practical example, if V_{out}=6V and V_{in}=9V (both commonly used voltages), then:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{R_2}{R_1 + R_2} = \frac{6}{9} = \frac{2}{3}$$

and by solving using algebra, R_2 must be twice the value of R_1.

To solve for R1:

$$R_1 = \frac{R_2 \cdot V_{\text{in}}}{V_{\text{out}}} - R_2$$

To solve for R2:

$$R_2 = \frac{R_1}{\left(\frac{V_{\text{in}}}{V_{\text{out}}} - 1\right)}$$

Any ratio between 0 and 1 is possible. That is, using resistors alone it is not possible to either invert the voltage or increase V_{out} above V_{in}.

Low-pass RC filter

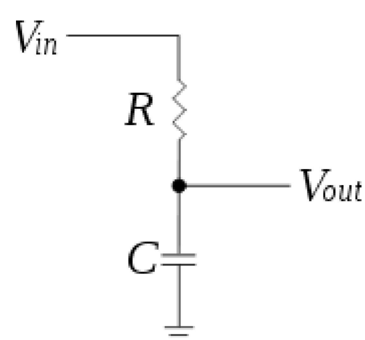


Figure 3: Resistor/capacitor voltage divider

Consider a divider consisting of a resistor and capacitor as shown in Figure 3.

Comparing with the general case, we see $Z_1 = R$ and Z_2 is the impedance of the capacitor, given by

$$Z_2 = -jX_{\mathrm{C}} = \frac{1}{j\omega C},$$

where X_C is the reactance of the capacitor, C is the capacitance of the capacitor, *j* is the imaginary unit, and *ω* (omega) is the radian frequency of the input voltage.

This divider will then have the voltage ratio:

$$\frac{V_{\mathrm{out}}}{V_{\mathrm{in}}} = \frac{Z_2}{Z_1 + Z_2} = \frac{\frac{1}{j\omega C}}{\frac{1}{j\omega C} + R} = \frac{1}{1 + j\omega\, RC}.$$

The product of *τ (tau) = RC* is called the *time constant* of the circuit.

The ratio then depends on frequency, in this case decreasing as frequency increases. This circuit is, in fact, a basic (first-order) lowpass filter. The ratio contains an imaginary number, and actually contains both the amplitude and phase shift information of the filter. To extract just the amplitude ratio, calculate the magnitude of the ratio, that is:

$$\left|\frac{V_{\mathrm{out}}}{V_{\mathrm{in}}}\right| = \frac{1}{\sqrt{1 + (\omega RC)^2}}.$$

Inductive divider

Inductive dividers split DC input according to resistive divider rules above.

Inductive dividers split AC input according to inductance:

$$V_{out} = V_{in} \cdot \frac{L_2}{L_1 + L_2}$$

The above equation is for ideal conditions. In the real world the amount of mutual inductance will alter the results.

Capacitive divider

Capacitive dividers do not pass DC input.

For an AC input a simple capacitive equation is:

$$V_{out} = V_{in} \cdot \frac{C_1}{C_1 + C_2}$$

Capacitive dividers are limited in current by the capacitance of the elements used.

This effect is opposite to resistive division and inductive division.

Loading effect

The voltage output of a voltage divider is not fixed but varies according to the load. To obtain a reasonably stable output voltage the output current should be a small fraction of the input current. The drawback of this is that most of the input current is wasted as heat in the resistors.

The following example describes the effect when a voltage divider is used to drive an amplifier:

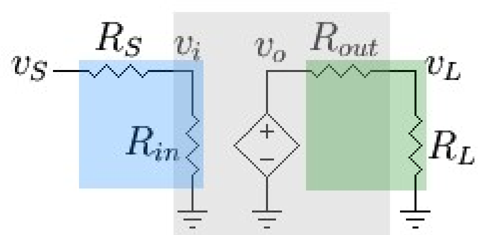


Figure 3: A simple voltage amplifier (gray outline) demonstrating input loading (blue outline) and output loading (green outline)

The gain of an amplifier generally depends on its source and load terminations, so-called **loading effects** that reduce the gain. The analysis of the amplifier itself is conveniently treated separately using idealized drivers and loads, and then supplemented by the use of voltage and current division to include the loading effects of real sources and loads. The choice of idealized driver and idealized load depends upon whether current or voltage is the input/output variable for the amplifier at hand, as described next.

In terms of sources, amplifiers with voltage input (voltage and transconductance amplifiers) typically are characterized using ideal zero-impedance voltage sources. In terms of terminations, amplifiers with voltage output (voltage and transresistance amplifiers) typically are characterized in terms of an open circuit output condition.

Similarly, amplifiers with current input (current and transresistance amplifiers) are characterized using ideal infinite impedance current sources, while amplifiers with current output (current and transconductance amplifiers) are characterized by a short-circuit output condition, As stated above, when any of these amplifiers is driven by a non-ideal source, and/or terminated by a finite, non-zero load, the effective gain is lowered due to the **loading effect** at the input and/or the output. Figure 3 illustrates loading by voltage division at both input and output for a simple voltage amplifier. For any of the four types of amplifier (current, voltage, transconductance or transresistance), these loading effects can be understood as a result of voltage division and/or current division, as described next.

Input loading

A general voltage source can be represented by a Thévenin equivalent circuit with Thévenin series impedance R_S. For a Thévenin driver, the *input* voltage v_i is reduced from v_S by voltage division to a value

$$v_i = v_S \frac{R_{in}}{R_S + R_{in}},$$

where R_{in} is the amplifier input resistance, and the overall gain is reduced below the idealized gain by the same voltage division factor.

In the same manner, the ideal input current for an ideal driver i_i is realized only for an infinite-resistance current driver. For a Norton driver with current i_S and source impedance R_S, the *input* current i_i is reduced from i_S by current division to a value

$$i_i = i_S \frac{R_S}{R_S + R_{in}},$$

where R_{in} is the amplifier input resistance, and the overall gain is reduced below the gain estimated using an ideal driver by the same current division factor.

More generally, complex frequency-dependent impedances can be used instead of the driver and amplifier resistances.

Output loading

For a finite load, R_L an output voltage is reduced by voltage division by the factor $R_L / (R_L + R_{out})$, where R_{out} is the amplifier output resistance. Likewise, as the term *short-circuit* implies, the output current delivered to a load R_L is reduced by current division by the factor $R_{out} / (R_L + R_{out})$. The overall gain is reduced below the gain estimated using an ideal load by the same current division factor.

More generally, complex frequency-dependent impedances can be used instead of the load and amplifier resistances.

Loaded gain - voltage amplifier case

Including both the input and output voltage division factors for the voltage amplifier of Figure 4, the ideal voltage gain A_v realized with an ideal driver and an open-circuit load is reduced to the **loaded gain** A_{loaded}:

$$A_{loaded} = \frac{v_L}{v_S} = \frac{R_{in}}{R_S + R_{in}} \frac{R_L}{R_{out} + R_L} A_v \ .$$

The resistor ratios in the above expression are called the **loading factors**.

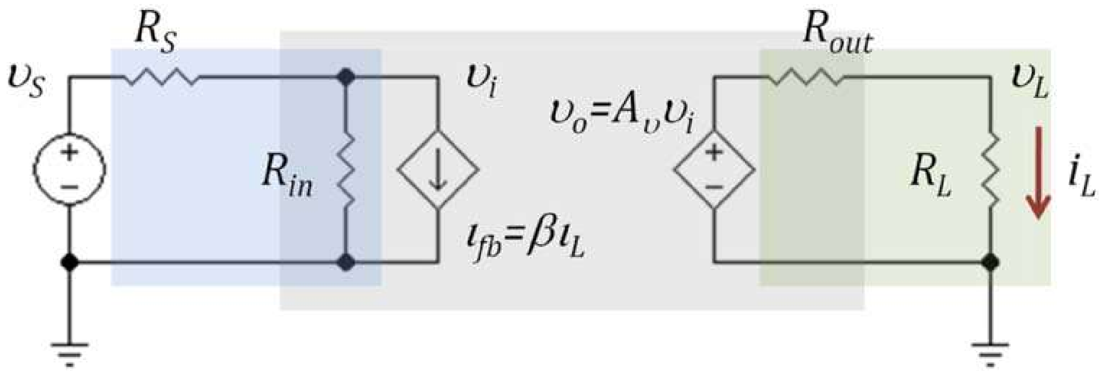


Figure 4: G-parameter voltage amplifier two-port; feedback provided by dependent current-controlled current source of gain β A/A

Unilateral versus bilateral amplifiers

Figure 3 and the associated discussion refers to a unilateral amplifier. In a more general case where the amplifier is represented by a two port, the input resistance of the amplifier depends on its load, and the output resistance on the source impedance. The loading factors in these cases must employ the true amplifier impedances including these bilateral effects. For example, taking the unilateral voltage amplifier of Figure 3, the corresponding bilateral two-port network is shown in Figure 4 based upon g-parameters. Carrying out the analysis for this circuit, the voltage gain with feedback A_{fb} is found to be

$$A_{fb} = \frac{v_L}{v_S} = \frac{A_{loaded}}{1 + \beta (R_S / R_L) A_{loaded}} \ .$$

That is, the ideal current gain A_i is reduced not only by the loading factors, but due to the bilateral nature of the two-port by an additional factor $(1 + \beta\ (R_S / R_L)\ A_{loaded})$, which is typical of negative feedback amplifier circuits. The factor $\beta(R_S / R_L)$ is the voltage feedback provided by the current feedback source of current gain β (A/A). For instance, for an ideal voltage source with $R_S = 0\ \Omega$, the current feedback has no influence, and for $R_L = \infty\ \Omega$, there is zero load current, again disabling the feedback.

Applications

Reference voltage

Voltage dividers are often used to produce stable reference voltages. The term *reference voltage* implies that little or no current is drawn from the divider output node by an attached load. Thus, use of the divider as a reference requires a load device with a high input impedance to avoid **loading** the divider, that is, to avoid disturbing its output voltage. A simple way of avoiding loading (for low power applications) is to simply input the reference voltage into the non-inverting input of an op-amp buffer. Another way is to "neutralize" the load impedance by an equivalent negative impedance (INIC).

Voltage source

While voltage dividers may be used to produce precise reference voltages (that is, when no current is drawn from the reference node), they make poor voltage sources (that is, when current *is* drawn from the reference node). The reason for poor source behavior is that the current drawn by the load passes through resistor R_1, but not through R_2, causing the voltage drop across R_1 to change with the load current, and thereby changing the output voltage.

In terms of the above equations, if current flows into a load resistance R_L (attached at the output node where the voltage is V_{out}), that load resistance R_L must be considered in parallel with R_2 to determine the voltage at V_{out}. In this case, the voltage at V_{out} is calculated as follows:

$$V_{\text{out}} = \frac{R_2 \| R_{\text{L}}}{R_1 + R_2 \| R_{\text{L}}} V_{\text{in}} = \frac{R_2}{R_1 \left(1 + \frac{R_2}{R_{\text{L}}}\right) + R_2} V_{\text{in}}$$

$$\text{where} \quad R_2 \| R_{\text{L}} \triangleq \left(\frac{1}{R_2} + \frac{1}{R_{\text{L}}}\right)^{-1} = \frac{R_2 R_{\text{L}}}{R_2 + R_{\text{L}}},$$

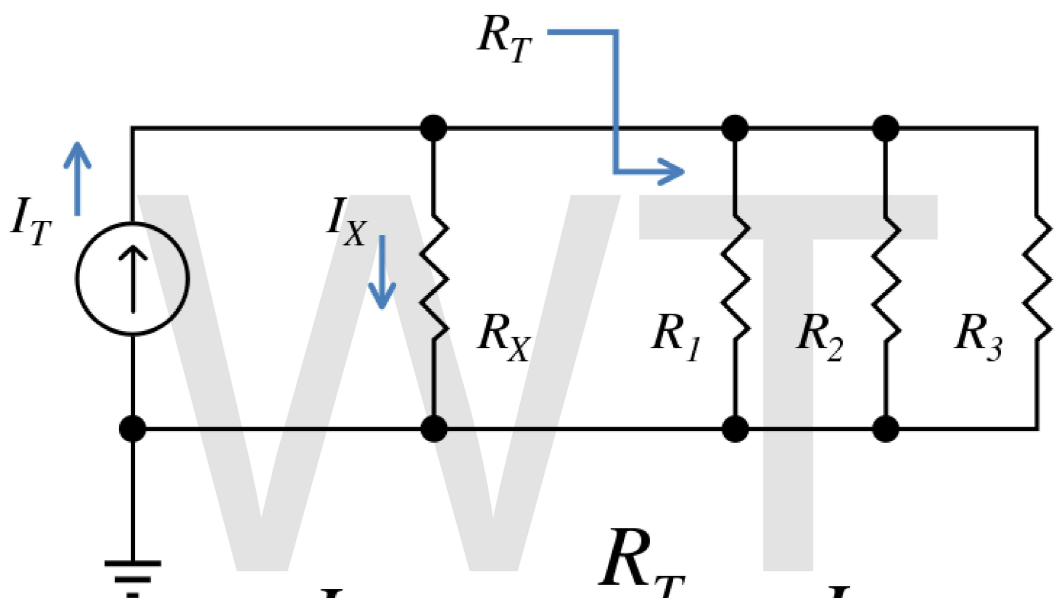
where R_L is a load resistor in parallel with R_2. From this result it is clear that V_{out} is decreased by R_L unless $R_2 // R_L \approx R_2$, that is, unless $R_L >> R_2$.

In other words, for high impedance loads it is possible to use a voltage divider as a voltage source, as long as R_2 has very small value compared to the load. This technique leads to considerable power dissipation in the divider.

A voltage divider is commonly used to set the DC Biasing of a common emitter amplifier, where the current drawn from the divider is the relatively low base current of the transistor.

Chapter 4

Current Divider



$$I_X = \frac{R_T}{R_X + R_T} I_T$$

Figure 1: Schematic of an electrical circuit illustrating current division. Notation $R_{T.}$ refers to the *total* resistance of the circuit to the right of resistor R_X.

In electronics, a **current divider** is a simple linear circuit that produces an output current (I_X) that is a fraction of its input current (I_T). **Current division** refers to the splitting of current between the branches of the divider. The currents in the various branches of such a circuit will always divide in such a way as to minimize the total energy expended. This can be shown by calculus.

The formula describing a current divider is similar in form to that for the voltage divider. However, the ratio describing current division places the impedance of the unconsidered branches in the numerator, unlike voltage division where the considered impedance is in the numerator. This is because in current dividers, total energy expended is minimized, resulting in currents that go through paths of least impedance, therefore the inverse relationship with impedance. On the other hand, voltage divider is used to satisfy Kirchoff's Voltage Law. The voltage around a loop must sum up to zero, so the voltage drops must be divided evenly in a direct relationship with the impedance.

To be specific, if two or more impedances are in parallel, the current that enters the combination will be split between them in inverse proportion to their impedances (according to Ohm's law). It also follows that if the impedances have the same value the current is split equally.

Resistive divider

A general formula for the current I_X in a resistor R_X that is in parallel with a combination of other resistors of total resistance R_T is (see Figure 1):

$$I_X = \frac{R_T}{R_X + R_T} I_T$$

where I_T is the total current entering the combined network of R_X in parallel with R_T. Notice that when R_T is composed of a parallel combination of resistors, say R_1, R_2, ... *etc.*, then the reciprocal of each resistor must be added to find the total resistance R_T:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots\,.$$

General case

Although the resistive divider is most common, the current divider may be made of frequency dependent impedances. In the general case the current I_X is given by:

$$I_X = \frac{Z_T}{Z_X + Z_T} I_T \ ,$$

Using Admittance

Instead of using impedances, the current divider rule can be applied just like the voltage divider rule if admittance (the inverse of impedance) is used.

$$I_X = \frac{Y_X}{Y_{Total}} I_T$$

Take care to note that Y_{Total} is a straightforward addition, not the sum of the inverses inverted (as you would do for a standard parallel resistive network). For Figure 1, the current I_X would be

$$I_X = \frac{Y_X}{Y_{Total}} I_T = \frac{\frac{1}{R_X}}{\frac{1}{R_X} + \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} I_T$$

Example: RC combination

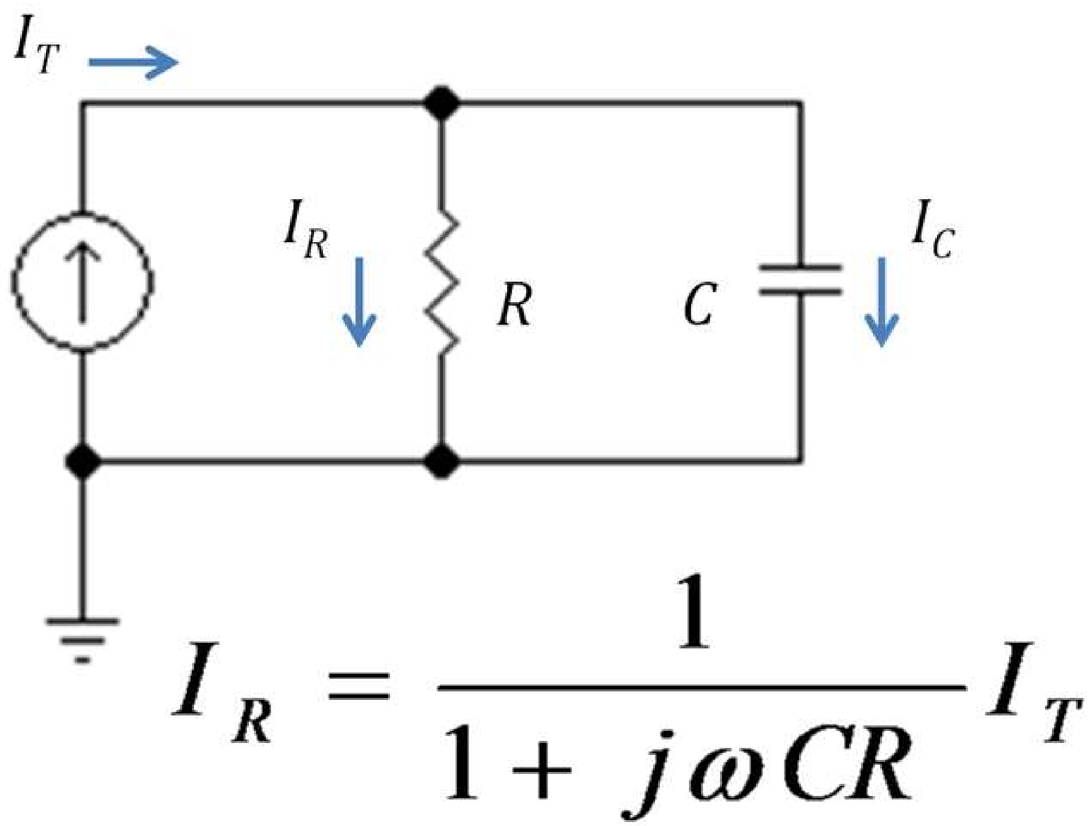


Figure 2: A low pass RC current divider

Figure 2 shows a simple current divider made up of a capacitor and a resistor. Using the formula above, the current in the resistor is given by:

$$\begin{aligned} I_R &= \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} I_T \\ &= \frac{1}{1 + j\omega CR} I_T \ , \end{aligned}$$

where $Z_C = 1/(j\omega C)$ is the impedance of the capacitor.

The product $\tau = CR$ is known as the time constant of the circuit, and the frequency for which ωCR = 1 is called the corner frequency of the circuit. Because the capacitor has zero impedance at high frequencies and infinite impedance at low frequencies, the current in the resistor remains at its DC value I_T for frequencies up to the corner frequency, whereupon it drops toward zero for higher frequencies as the capacitor effectively short-circuits the resistor. In other words, the current divider is a low pass filter for current in the resistor.

Loading effect

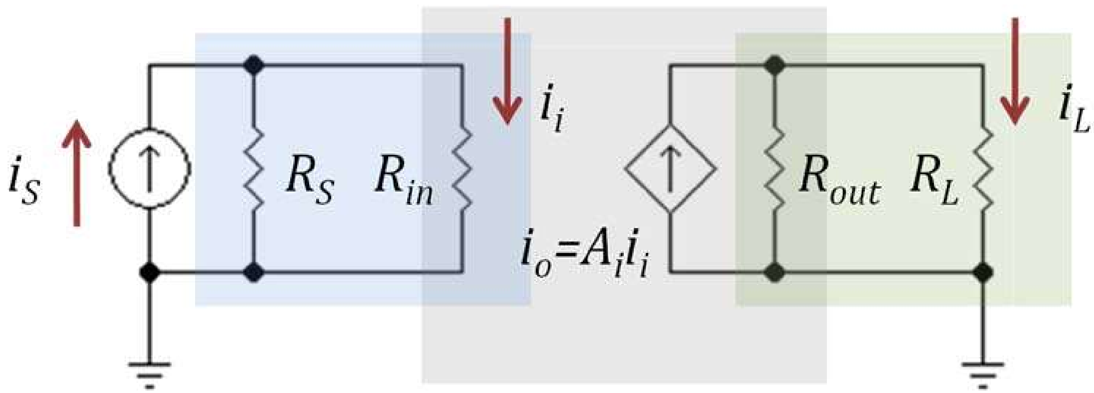


Figure 3: A current amplifier (gray box) driven by a Norton source (i_S, R_S) and with a resistor load R_L. Current divider in blue box at input (R_S,R_{in}) reduces the current gain, as does the current divider in green box at the output (R_{out},R_L)

The gain of an amplifier generally depends on its source and load terminations. Current amplifiers and transconductance amplifiers are characterized by a short-circuit output condition, and current amplifiers and transresistance amplifiers are characterized using ideal infinite impedance current sources. When an amplifier is terminated by a finite, non-zero termination, and/or driven by a non-ideal source, the effective gain is reduced due to the **loading effect** at the output and/or the input, which can be understood in terms of current division.

Figure 3 shows a current amplifier example. The amplifier (gray box) has input resistance R_{in} and output resistance R_{out} and an ideal current gain A_i. With an ideal current driver (infinite Norton resistance) all the source current i_S becomes input current to the amplifier. However, for a Norton driver a current divider is formed at the input that reduces the input current to

$$i_i = \frac{R_S}{R_S + R_{in}} i_S \ ,$$

which clearly is less than i_S. Likewise, for a short circuit at the output, the amplifier delivers an output current $i_o = A_i\, i_i$ to the short-circuit. However, when the load is a non-zero resistor R_L, the current delivered to the load is reduced by current division to the value:

$$i_L = \frac{R_{out}}{R_{out} + R_L} A_i i_i \ .$$

Combining these results, the ideal current gain A_i realized with an ideal driver and a short-circuit load is reduced to the **loaded gain** A_{loaded}:

$$A_{loaded} = \frac{i_L}{i_S} = \frac{R_S}{R_S + R_{in}} \frac{R_{out}}{R_{out} + R_L} A_i \ .$$

The resistor ratios in the above expression are called the **loading factors**.

Unilateral versus bilateral amplifiers

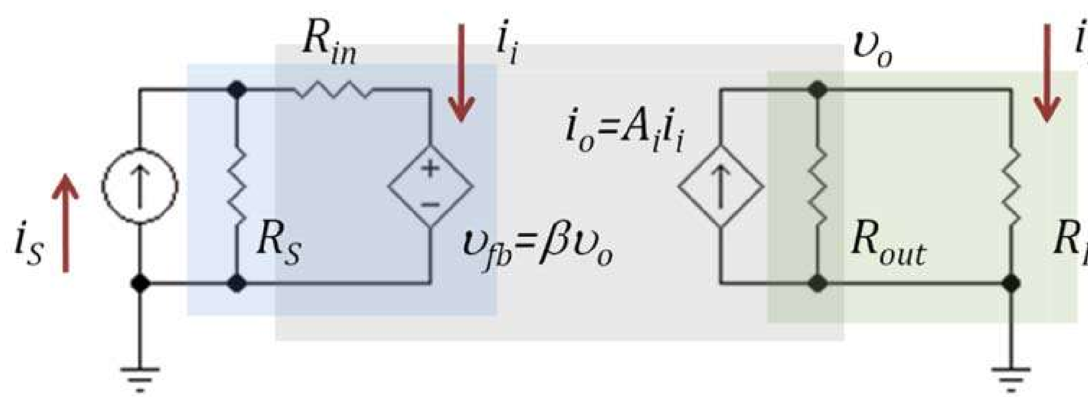


Figure 4: Current amplifier as a bilateral two-port network; feedback through dependent voltage source of gain β V/V

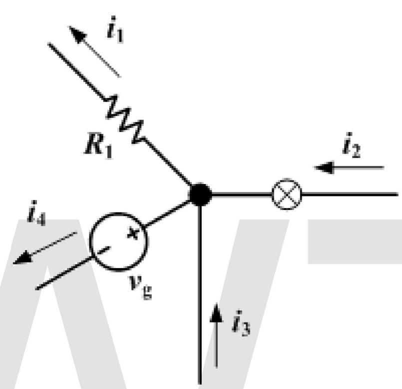
Figure 3 and the associated discussion refers to a unilateral amplifier. In a more general case where the amplifier is represented by a two port, the input resistance of the amplifier depends on its load, and the output resistance on the source impedance. The loading factors in these cases must employ the true amplifier impedances including these bilateral effects. For example, taking the unilateral current amplifier of Figure 3, the corresponding bilateral two-port network is shown in Figure 4 based upon h-parameters. Carrying out the analysis for this circuit, the current gain with feedback A_{fb} is found to be

$$A_{fb} = \frac{i_L}{i_S} = \frac{A_{loaded}}{1 + \beta (R_L / R_S) A_{loaded}} \ .$$

That is, the ideal current gain A_i is reduced not only by the loading factors, but due to the bilateral nature of the two-port by an additional factor $(1 + \beta (R_L / R_S) A_{loaded})$, which is typical of negative feedback amplifier circuits. The factor $\beta (R_L / R_S)$ is the current feedback provided by the voltage feedback source of voltage gain β V/V. For instance, for an ideal current source with $R_S = \infty\ \Omega$, the voltage feedback has no influence, and for $R_L = 0\ \Omega$, there is zero load voltage, again disabling the feedback.

Chapter 5

Nodal Analysis



Kirchhoff's current law is the basis of nodal analysis

In electric circuits analysis, **nodal analysis**, **node-voltage analysis**, or the **branch current method** is a method of determining the voltage (potential difference) between "nodes" (points where elements or branches connect) in an electrical circuit in terms of the branch currents.

In analyzing a circuit using Kirchhoff's circuit laws, one can either do nodal analysis using Kirchhoff's current law (KCL) or mesh analysis using Kirchhoff's voltage law (KVL). Nodal analysis writes an equation at each electrical node, requiring that the branch currents incident at a node must sum to zero. The branch currents are written in terms of the circuit node voltages. As a consequence, each branch constitutive relation must give current as a function of voltage; an admittance representation. For instance, for a resistor, $I_{branch} = V_{branch} * G$, where G (=1/R) is the admittance (conductance) of the resistor.

Nodal analysis is possible when all the circuit elements' branch constitutive relations have an admittance representation. Nodal analysis produces a compact set of equations for the network, which can be solved by hand if small, or can be quickly solved using linear algebra by computer. Because of the compact system of equations, many circuit simulation programs (e.g. SPICE) use nodal analysis as a basis. When elements do not have admittance representations, a more general extension of nodal analysis, modified nodal analysis, can be used.

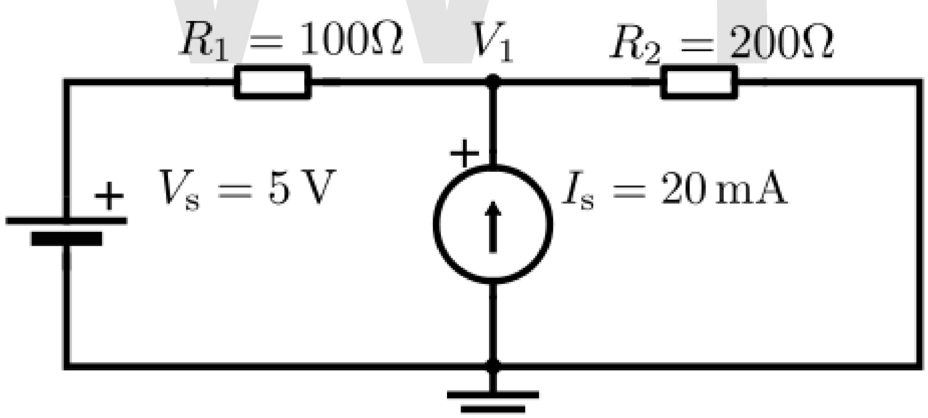
While simple examples of nodal analysis focus on linear elements, more complex nonlinear networks can also be solved with nodal analysis by using Newton's method to turn the nonlinear problem into a sequence of linear problems.

Method

1. Note all connected wire segments in the circuit. These are the *nodes* of nodal analysis.
2. Select one node as the ground reference. The choice does not affect the result and is just a matter of convention. Choosing the node with most connections can simplify the analysis.
3. Assign a variable for each node whose voltage is unknown. If the voltage is already known, it is not necessary to assign a variable.
4. For each unknown voltage, form an equation based on Kirchhoff's current law. Basically, add together all currents leaving from the node and mark the sum equal to zero.
5. If there are voltage sources between two unknown voltages, join the two nodes as a supernode. The currents of the two nodes are combined in a single equation, and a new equation for the voltages is formed.
6. Solve the system of simultaneous equations for each unknown voltage.

Examples

Basic case



Basic example circuit with one unknown voltage, V_1

The only unknown voltage in this circuit is V_1. There are three connections to this node and consequently three currents to consider. The direction of the currents in calculations is chosen to be away from the node.

1. Current through resistor R_1: $(V_1 - V_S) / R_1$
2. Current through resistor R_2: V_1 / R_2

3. Current through current source I_S: $-I_S$

With Kirchhoff's current law, we get:

$$\frac{V_1 - V_S}{R_1} + \frac{V_1}{R_2} - I_S = 0$$

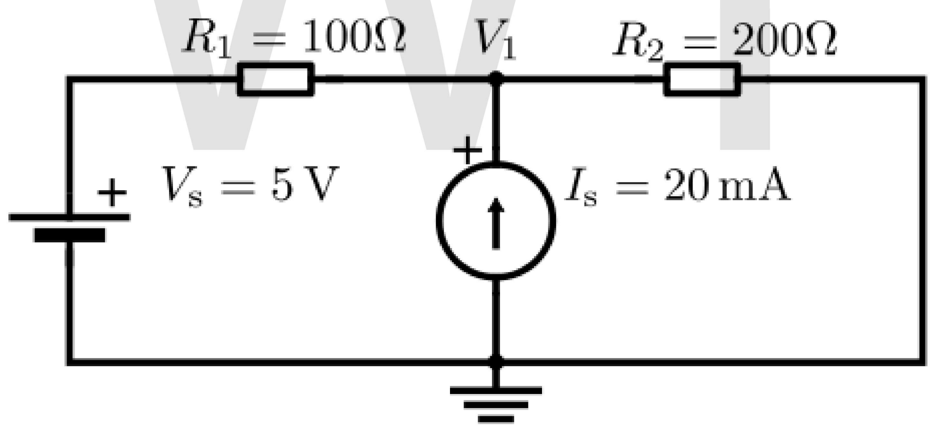
This equation can be solved in respect to V_1:

$$V_1 = \left(\frac{V_S}{R1} + I_S\right) : \left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

Finally, the unknown voltage can be solved by substituting numerical values for the symbols. Any unknown currents are easy to calculate after all the voltages in the circuit are known.

$$V_1 = \left(\frac{5\text{ V}}{100\,\Omega} + 20\text{ mA}\right) : \left(\frac{1}{100\,\Omega} + \frac{1}{200\,\Omega}\right) \approx 4.667\text{ V}$$

Supernodes



In this circuit, V_A is between two unknown voltages, and is therefore a supernode

In this circuit, we initially have two unknown voltages, V_1 and V_2. The voltage at V_3 is already known to be V_B because the other terminal of the voltage source is at ground potential.

The current going through voltage source V_A cannot be directly calculated. Therefore we can not write the current equations for either V_1 or V_2. However, we know that the same current leaving node V_2 must enter node V_1. Even though the nodes can not be individually solved, we know that the combined current of these two nodes is zero. This combining of the two nodes is called the supernode technique, and it requires one additional equation: $V_1 = V_2 + V_A$.

The complete set of equations for this circuit is:

$$\begin{cases} \frac{V_1 - V_B}{R_1} + \frac{V_2 - V_B}{R_2} + \frac{V_2}{R_3} = 0 \\ V_1 = V_2 + V_A \end{cases}$$

By substituting V_1 to the first equation and solving in respect to V2, we get:

$$V_2 = \frac{(R_1 + R_2)R_3 V_B - R_2 R_3 V_A}{(R_1 + R_2)R_3 + R_1 R_2}$$

Chapter 6

Mesh Analysis

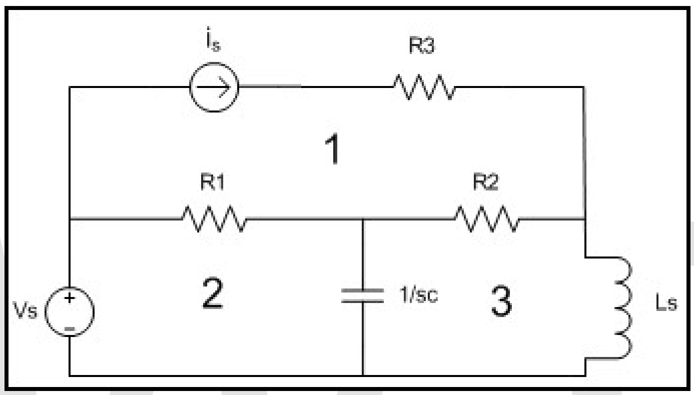


Figure 1: Essential Meshes of the planar circuit labeled 1, 2, and 3. R1, R2, R3, 1/sc, and Ls represent the impedance of the resistors, capacitor, and inductor values in the s-domain. Vs and Is are the values of the voltage source and current source respectively.

Mesh analysis (sometimes referred to as **loop analysis** or **mesh current method**) is a method that is used to solve planar circuits for the voltage and currents at any place in the circuit. Planar circuits are circuits that can be drawn on a plane with no wires overlapping each other. Mesh analysis uses Kirchhoff's voltage law to solve these planar circuits. The advantage of using mesh analysis is that it creates a systematic approach to solving planar circuits and reduces the number of equations needed to solve the circuit for all of the voltages and currents.

Mesh currents and essential meshes

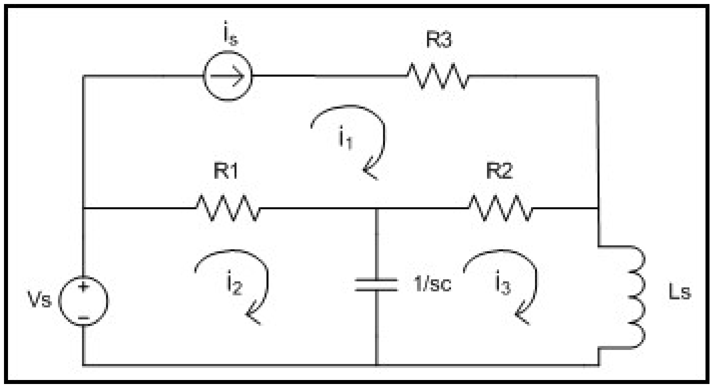


Figure 2: Circuit with Mesh Currents Labeled as i1, i2, and i3. The arrows show the direction of the mesh current.

Mesh analysis works by arbitrarily assigning mesh currents in the essential meshes. An essential mesh is a loop in the circuit that does not contain any other loop. When looking at a circuit schematic, the essential meshes look like a “window pane”. Figure 1 labels the essential meshes with one, two, and three. Once the essential meshes are found, the mesh currents need to be labeled.

A mesh current is a current that loops around the essential mesh. The mesh current might not have a physical meaning but it is used to set up the mesh analysis equations. When assigning the mesh currents it is important to have all the mesh currents loop in the same direction. This will help prevent errors when writing out the equations. The convention is to have all the mesh currents looping in a clockwise direction. Figure 2 shows the same circuit shown before but with the mesh currents labeled.

The reason to use mesh currents instead of just using KCL and KVL to solve a problem is that the mesh currents can account for any unnecessary currents that may be drawn in when using KCL and KVL. Mesh analysis ensures that the least possible number of equations regarding currents is used, greatly simplifying the problem.

Setting up the equations

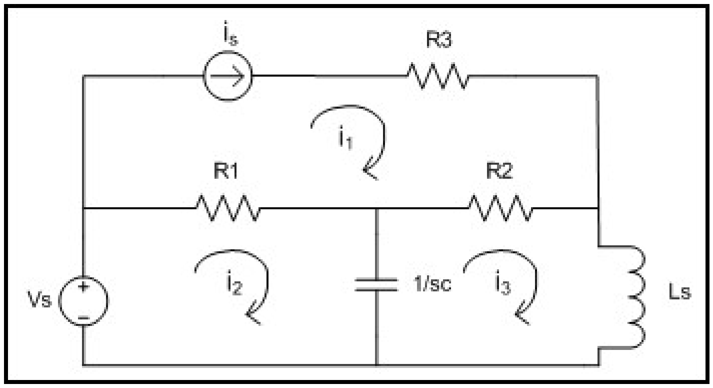


Figure 3: Simple Circuit using Mesh Analysis

After labeling the mesh currents, one only needs to write one equation per mesh in order to solve for all the currents in the circuit. These equations are the sum of the voltage drops in a complete loop of the mesh current. For other than current and voltage sources, the voltage drops will be the impedance of the electronic component multiplied by the mesh current in that loop. It is important to note that if a component exists between two essential meshes, the component's voltage drop will be the impedance of the component times the present mesh current minus the neighboring mesh current (computing the subtraction first).

If a voltage source is present within the mesh loop, the voltage at the source is either added or subtracted depending on if it is a voltage drop or a voltage rise in the direction of the mesh current. For a current source that is not contained between two meshes, the mesh current will take the positive or negative value of the current source depending on if the mesh current is in the same or opposite direction of the current source. The following is the same circuit from above with the equations needed to solve for all the currents in the circuit.

$$\begin{cases} \text{Mesh 1: } i_1 = i_s \\ \text{Mesh 2: } -V_s + R_1(i_2 - i_1) + \frac{1}{sc}(i_2 - i_3) = 0 \\ \text{Mesh 3: } \frac{1}{sc}(i_3 - i_2) + R_2(i_3 - i_1) + Lsi_3 = 0 \end{cases}$$

Once the equations are found, the system of linear equations can be solved by using any technique to solve linear equations.

Special cases

There are two special cases in mesh current: supermesh and dependent sources.

Supermesh

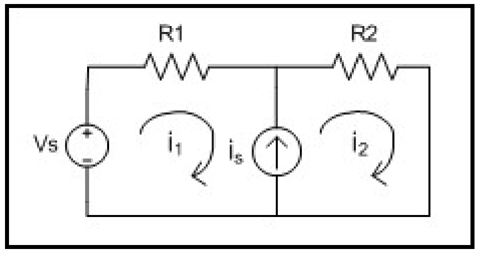


Figure 4: Circuit with a supermesh. Supermesh occurs because the current source is in between the essential meshes.

A supermesh occurs when a current source is contained between two essential meshes. To handle the supermesh, first treat the circuit as if the current source is not there. This leads to one equation that incorporates two mesh currents. Once this equation is formed, an equation is needed that relates the two mesh currents with the current source. This will be an equation where the current source is equal to one of the mesh currents minus the other. The following is a simple example of dealing with a supermesh.

$$\begin{cases} \text{Mesh 1, 2:} \quad -V_s + R_1 i_1 + R_2 i_2 = 0 \\ \text{Current source: } i_s = i_2 - i_1 \end{cases}$$

Dependent sources

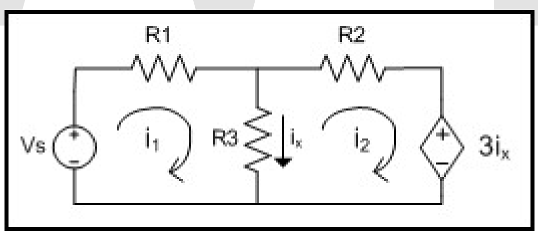


Figure 5: Circuit with dependent source. i_x is the current that the dependent voltage source depends on.

A dependent source is a current source or voltage source that depends on the voltage or current on another element in the circuit. When a dependent source is contained within an essential mesh, the dependent source should be treated like a normal source. After the mesh equation is formed, a dependent source equation is needed. This equation is generally called a constraint equation. This is an equation that relates the dependent source's variable to the voltage or current that the source depends on in the circuit. The following is a simple example of a dependent source.

$$\begin{cases} \text{Mesh 1: } -V_s + R_1 i_1 + R_3(i_1 - i_2) = 0 \\ \text{Mesh 2: } R_2 i_2 + 3i_x + R_3(i_2 - i_1) = 0 \\ \text{Dependent variable: } i_x = i_1 - i_2 \end{cases}$$

Chapter 7

Superposition Theorem & Transfer Function

Superposition theorem

The **superposition theorem** for electrical circuits states that the response (Voltage or Current) in any branch of a bilateral linear circuit having more than one independent source equals the algebraic sum of the responses caused by each independent source acting alone, while all other independent sources are replaced by their internal impedances.

To ascertain the contribution of each individual source, all of the other sources first must be "turned off" (set to zero) by:

1. Replacing all other independent voltage sources with a short circuit (thereby eliminating difference of potential. i.e. V=0, internal impedance of ideal voltage source is ZERO (short circuit)).
2. Replacing all other independent current sources with an open circuit (thereby eliminating current. i.e. I=0, internal impedance of ideal current source is infinite (open circuit).

This procedure is followed for each source in turn, then the resultant responses are added to determine the true operation of the circuit. The resultant circuit operation is the superposition of the various voltage and current sources.

The superposition theorem is very important in circuit analysis. It is used in converting any circuit into its Norton equivalent or Thevenin equivalent.

Applicable to linear networks (time varying or time invariant) consisting of independent sources, linear dependent sources, linear passive elements Resistors, Inductors, Capacitors and linear transformers.

Transfer function

A **transfer function** (also known as the **system function** or **network function**) is a mathematical representation, in terms of spatial or temporal frequency, of the relation between the input and output of a linear time-invariant system. With optical imaging devices, for example, it is the Fourier transform of the point spread function (hence a function of spatial frequency) i.e. the intensity distribution caused by a point object in the field of view.

Explanation

The transfer functions are commonly used in the analysis of single-input single-output filters, for instance. It is mainly used in signal processing, communication theory, and control theory. The term is often used exclusively to refer to linear, time-invariant systems (LTI), as covered here. Most real systems have non-linear input/output characteristics, but many systems, when operated within nominal parameters (not "over-driven") have behavior that is close enough to linear that LTI system theory is an acceptable representation of the input/output behavior.

In its simplest form for continuous-time input signal $x(t)$ and output $y(t)$, the transfer function is the linear mapping of the Laplace transform of the input, $X(s)$, to the output $Y(s)$:

$$Y(s) = H(s)\,X(s)$$

or

$$H(s) = \frac{Y(s)}{X(s)} = \frac{\mathcal{L}\left\{y(t)\right\}}{\mathcal{L}\left\{x(t)\right\}}$$

where $H(s)$ is the transfer function of the LTI system.

In discrete-time systems, the function is similarly written as $H(z) = \frac{Y(z)}{X(z)}$ and is often referred to as the pulse-transfer function.

Direct derivation from differential equations

Consider a linear differential equation with constant coefficients

$$L[u] = \frac{d^n u}{dt^n} + a_1\frac{d^{n-1}u}{dt^{n-1}} + \dots + a_{n-1}\frac{du}{dt} + a_n u = r(t)$$

where u and r are suitably smooth functions of t, and L is the operator defined on the relevant function space, that transforms u into r. That kind of equation can be used to constrain the output function u in terms of the *forcing* function r. The transfer function, written as an operator $F[r] = u$, is the right inverse of L, since $L[F[r]] = r$.

Solutions of the *homogeneous* equation $L[u] = 0$ can be found by trying $u = e^{\lambda t}$. That substitution yields the *characteristic polynomial*

$$p_L(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n$$

The inhomogeneous case can be easily solved if the input function r is also of the form $r(t) = e^{st}$. In that case, by substituting $u = H(s)e^{st}$ one finds that $L[H(s)e^{st}] = e^{st}$ if and only if

$$H(s) = \frac{1}{p_L(s)}, \qquad p_L(s) \neq 0.$$

Taking that as the definition of the *transfer function* requires to carefully disambiguate between complex vs. real values, is traditionally influenced by the interpretation of *abs(H(s))* as the gain and *-atan(H(s))* as the phase lag.

Signal processing

Let $x(t)$ be the input to a general linear time-invariant system, and $y(t)$ be the output, and the bilateral Laplace transform of $x(t)$ and $y(t)$ be

$$X(s) = \mathcal{L}\left\{x(t)\right\} \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} x(t)e^{-st}\,dt$$
$$Y(s) = \mathcal{L}\left\{y(t)\right\} \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} y(t)e^{-st}\,dt.$$

Then the output is related to the input by the **transfer function** $H(s)$ as

$$Y(s) = H(s)X(s)$$

and the transfer function itself is therefore

$$H(s) = \frac{Y(s)}{X(s)}.$$

In particular, if a complex harmonic signal with a sinusoidal component with amplitude $|X|$, angular frequency ω and phase $\arg(X)$

$x(t) = Xe^{j\omega t} = |X| e^{j(\omega t + \arg(X))}$
where $X = |X| e^{j\arg(X)}$

is input to a linear time-invariant system, then the corresponding component in the output is:

$$y(t) = Ye^{j\omega t} = |Y|e^{j(\omega t+\arg(Y))}$$
and $Y = |Y| e^{j\arg(Y)}$.

Note that, in a linear time-invariant system, the input frequency ω has not changed, only the amplitude and the phase angle of the sinusoid has been changed by the system. The frequency response $H(j\omega)$ describes this change for every frequency ω in terms of *gain*:

$$G(\omega) = \frac{|Y|}{|X|} = |H(j\omega)|$$

and *phase shift*:

$$\phi(\omega) = \arg(Y) - \arg(X) = \arg(H(j\omega)).$$

The phase delay (i.e., the frequency-dependent amount of delay introduced to the sinusoid by the transfer function) is:

$$\tau_\phi(\omega) = -\tfrac{\phi(\omega)}{\omega}.$$

The group delay (i.e., the frequency-dependent amount of delay introduced to the envelope of the sinusoid by the transfer function) is found by computing the derivative of the phase shift with respect to angular frequency ω,

$$\tau_g(\omega) = -\tfrac{d\phi(\omega)}{d\omega}.$$

The transfer function can also be shown using the Fourier transform which is only a special case of the bilateral Laplace transform for the case where $s = j\omega$.

Common transfer function families

While any LTI system can be described by some transfer function or another, there are certain "families" of special transfer functions that are commonly used. Typical infinite impulse response filters are designed to implement one of these special transfer functions.

Some common transfer function families and their particular characteristics are:

- Butterworth filter – maximally flat in passband and stopband for the given order
- Chebyshev filter (Type I) – maximally flat in stopband, sharper cutoff than Butterworth of same order
- Chebyshev filter (Type II) – maximally flat in passband, sharper cutoff than Butterworth of same order
- Bessel filter – best pulse response for a given order because it has no group delay ripple
- Elliptic filter – sharpest cutoff (narrowest transition between pass band and stop band) for the given order
- Optimum "L" filter
- Gaussian filter – minimum group delay; gives no overshoot to a step function.
- Hourglass filter
- Raised-cosine filter

Control engineering

In control engineering and control theory the transfer function is derived using the Laplace transform.

The transfer function was the primary tool used in classical control engineering. However, it has proven to be unwieldy for the analysis of multiple-input multiple-output (MIMO) systems, and has been largely supplanted by state space representations for such systems. In spite of this, a transfer matrix can be always obtained for any linear system, in order to analyze its dynamics and other properties: each element of a transfer matrix is a transfer function relating a particular input variable to an output variable.

Chapter 8

Electrical Resistance

The **electrical resistance** of an object measures its opposition to the passage of an electric current. An object of uniform cross section has a resistance proportional to its resistivity and length and inversely proportional to its cross-sectional area.

Discovered by Georg Ohm in 1827, electrical resistance shares some conceptual parallels with the mechanical notion of friction. The SI unit of electrical resistance is the ohm (Ω). Resistance's reciprocal quantity is electrical conductance measured in siemens.

The resistance of an object can be defined as the ratio of voltage to current:

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

For a wide variety of materials and conditions, the electrical resistance R is constant for a given temperature; it does not depend on the amount of current through or the potential difference (voltage) across the object. Such materials are called Ohmic materials. For objects made of ohmic materials the definition of the resistance, with R being a constant for that resistor, is known as Ohm's law.

In the case of a nonlinear conductor (not obeying Ohm's law), this ratio can change as current or voltage changes; the inverse slope of a chord to an I–V curve is sometimes referred to as a "chordal resistance" or "static resistance".

Conductors and resistors

A 65-kΩ resistor, as identified by its electronic color code (blue–green–black). An ohmmeter could be used to verify this value.

Objects such as wires that are designed to have low resistance so that they transfer current with the least loss of electrical energy are called conductors. Objects that are designed to have a specific resistance so that they can dissipate electrical energy or otherwise modify how a circuit behaves are called resistors. Conductors are made of highly conductive materials such as metals, in particular copper and aluminum. Resistors, on the other hand, are made of a wide variety of materials depending on factors such as the desired resistance, amount of energy that it needs to dissipate, precision, and cost.

DC resistance

The resistance of a given resistor or conductor grows with the length of conductor and decreases for larger cross-sectional area. The resistance R of a conductor of uniform cross section, therefore, can be computed as

$$R = \rho \frac{\ell}{A},$$

where ℓ is the length of the conductor, measured in metres [m], A is the cross-sectional area of the conductor measured in square metres [m²], and ρ (Greek: rho) is the electrical resistivity (also called *specific electrical resistance*) of the material, measured in ohm-metres (Ω m). Resistivity is a measure of the material's ability to oppose electric current.

For practical reasons, any connections to a real conductor will almost certainly mean the current density is not totally uniform. However, this formula still provides a good approximation for long thin conductors such as wires.

AC resistance

If a wire conducts high-frequency alternating current, then the effective cross sectional area of the wire is reduced because of the skin effect. If several conductors are together, then due to proximity effect, the effective resistance of each is higher than if that conductor were alone. These effects are so small for low frequency of ordinary household AC that they should ordinarily be treated as if it were DC resistance.

Measuring resistance

An instrument for measuring resistance is called an ohmmeter. Simple ohmmeters cannot measure low resistances accurately because the resistance of their measuring leads causes a voltage drop that interferes with the measurement, so more accurate devices use four-terminal sensing.

Causes of resistance

In metals

A metal consists of a lattice of atoms, each with a shell of electrons. This is also known as a positive ionic lattice. The outer electrons are free to dissociate from their parent atoms and travel through the lattice, creating a 'sea' of electrons, making the metal a conductor. When an electrical potential difference (a voltage) is applied across the metal, the electrons drift from one end of the conductor to the other under the influence of the electric field.

Near room temperatures, the thermal motion of ions is the primary source of scattering of electrons (due to destructive interference of free electron waves on non-correlating potentials of ions), and is thus the prime cause of metal resistance. Imperfections of lattice also contribute into resistance, although their contribution in pure metals is negligible.

The larger the cross-sectional area of the conductor, the more electrons are available to carry the current, so the lower the resistance. The longer the conductor, the more scattering events occur in each electron's path through the material, so the higher the resistance. Different materials also affect the resistance.

In semiconductors and insulators

In metals, the Fermi level lies in the conduction band giving rise to free conduction electrons. However, in semiconductors the position of the Fermi level is within the band gap, approximately half-way between the conduction band minimum and valence band maximum for intrinsic (undoped) semiconductors. This means that at 0 kelvins, there are no free conduction electrons and the resistance is infinite. However, the resistance will continue to decrease as the charge carrier density in the conduction band increases. In extrinsic (doped) semiconductors, dopant atoms increase the majority charge carrier concentration by donating electrons to the conduction band or accepting holes in the valence band. For both types of donor or acceptor atoms, increasing the dopant density leads to a reduction in the resistance. Highly doped semiconductors hence behave metallic. At very high temperatures, the contribution of thermally generated carriers will dominate over the contribution from dopant atoms and the resistance will decrease exponentially with temperature.

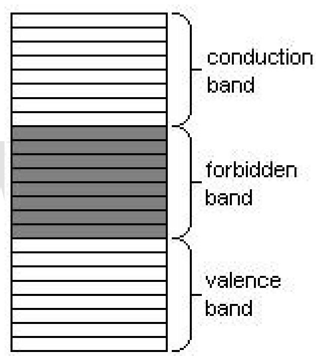
In ionic liquids/electrolytes

In electrolytes, electrical conduction happens not by band electrons or holes, but by full atomic species (ions) traveling, each carrying an electrical charge. The resistivity of ionic liquids varies tremendously by the concentration - while distilled water is almost an insulator, salt water is a very efficient electrical conductor. In biological membranes, currents are carried by ionic salts. Small holes in the membranes, called ion channels, are selective to specific ions and determine the membrane resistance.

Resistivity of various materials

Material	**Resistivity**, ρ ohm-metre
Metals	10^{-8}
Semiconductors	variable
Electrolytes	variable
Insulators	10^{16}
Superconductors	0 (exactly)

Band theory simplified



Electron energy levels in an insulator

Quantum mechanics states that the energy of an electron in an atom cannot be any arbitrary value. Rather, there are fixed energy levels which the electrons can occupy, and values in between these levels are impossible. The energy levels are grouped into two bands: the **valence band** and the **conduction band** (the latter is generally above the former). Electrons in the conduction band may move freely throughout the substance in the presence of an electrical field.

In insulators and semiconductors, the atoms in the substance influence each other so that between the valence band and the conduction band there exists a forbidden band of energy levels, which the electrons cannot occupy. In order for a current to flow, a relatively large amount of energy must be furnished to an electron for it to leap across this forbidden gap and into the conduction band. Thus, even large voltages can yield relatively small currents.

Differential resistance

When the current–voltage dependence is not linear, **differential resistance**, **incremental resistance** or **slope resistance** is defined as the slope of the *V-I* graph at a particular point, thus:

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

This quantity is sometimes called simply *resistance*, although the two definitions are equivalent only for an ohmic component such as an ideal resistor. For example, a diode is a circuit element for which the resistance depends on the applied voltage or current.

If the *V-I* graph is not monotonic (i.e. it has a peak or a trough), the differential resistance will be negative for some values of voltage and current. This property is often known as *negative resistance*, although it is more correctly called *negative differential resistance*, since the absolute resistance *V*/*I* is still positive. An example of such an element is the tunnel diode.

Differential resistance is only useful to compare a nonlinear device with a linear source/load in some small interval; for example if it is necessary to evaluate a zener diode's voltage stability under different current values.

Temperature dependence

Near room temperature, the electric resistance of a typical metal increases linearly with rising temperature, while the electrical resistance of a typical semiconductor decreases with rising temperature. The amount of that change in resistance can be calculated using the temperature coefficient of resistivity of the material using the following formula:

$$R = R_0[\alpha(T - T_0) + 1]$$

where T is its temperature, T_0 is a reference temperature (usually room temperature), R_0 is the resistance at T_0, and α is the percentage change in resistivity per unit temperature. The constant α depends only on the material being considered. The relationship stated is actually only an approximate one, the true physics being somewhat non-linear, or looking at it another way, α itself varies with temperature. For this reason it is usual to specify the temperature that α was measured at with a suffix, such as α_{15} and the relationship only holds in a range of temperatures around the reference.

At lower temperatures (less than the Debye temperature), the resistance of a metal decreases as T^5 due to the electrons scattering off of phonons. At even lower temperatures, the dominant scattering mechanism for electrons is other electrons, and the resistance decreases as T^2. At some point, the impurities in the metal will dominate the behavior of the electrical resistance which causes it to saturate to a constant value. Matthiessen's Rule (first formulated by Augustus Matthiessen in the 1860s; the equation below gives its modern form) says that all of these different behaviors can be summed up to get the total resistance as a function of temperature,

$$R = R_{\text{imp}} + aT^2 + bT^5 + cT$$

where R_{imp} is the temperature independent electrical resistivity due to impurities, and *a*, *b*, and *c* are coefficients which depend upon the metal's properties. This rule can be seen as the motivation to Heike Kamerlingh Onnes's experiments that led in 1911 to discovery of superconductivity.

Intrinsic semiconductors become better conductors as the temperature increases; the electrons are bumped to the conduction energy band by thermal energy, where they flow freely and in doing so leave behind holes in the valence band which also flow freely. The electric resistance of a typical intrinsic (non doped) semiconductor decreases exponentially with the temperature:

$$R = R_0 e^{-aT}$$

Extrinsic (doped) semiconductors have a far more complicated temperature profile. As temperature increases starting from absolute zero they first decrease steeply in resistance as the carriers leave the donors or acceptors. After most of the donors or acceptors have lost their carriers the resistance starts to increase again slightly due to the reducing mobility of carriers (much as in a metal). At higher temperatures it will behave like intrinsic semiconductors as the carriers from the donors/acceptors become insignificant compared to the thermally generated carriers.

The electric resistance of electrolytes and insulators is highly nonlinear, and case by case dependent, therefore no generalized equations are given.

Strain dependence

Just as the resistance of a conductor depends upon temperature, the resistance of a conductor depends upon strain. By placing a conductor under tension (a form of stress that leads to strain in the form of stretching of the conductor), the length of the section of conductor under tension increases and its cross-sectional area decreases. Both these effects contribute to increasing the resistance of the strained section of conductor. Under compression (strain in the opposite direction), the resistance of the strained section of conductor decreases.

Chapter 9

Capacitance

In electromagnetism and electronics, **capacitance** is the ability of a body to hold an electrical charge. Capacitance is also a measure of the amount of electrical energy stored (or separated) for a given electric potential. A common form of energy storage device is a parallel-plate capacitor. In a parallel plate capacitor, capacitance is directly proportional to the surface area of the conductor plates and inversely proportional to the separation distance between the plates. If the charges on the plates are $+Q$ and $-Q$, and V gives the voltage between the plates, then the capacitance is given by

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

The SI unit of capacitance is the farad; 1 farad is 1 coulomb per volt.

The energy (measured in joules) stored in a capacitor is equal to the *work* done to charge it. Consider a capacitance C, holding a charge $+q$ on one plate and $-q$ on the other. Moving a small element of charge $\mathrm{d}q$ from one plate to the other against the potential difference $V = q/C$ requires the work $\mathrm{d}W$:

$$\mathrm{d}W = \frac{q}{C}\,\mathrm{d}q$$

where W is the work measured in joules, q is the charge measured in coulombs and C is the capacitance, measured in farads.

The energy stored in a capacitance is found by integrating this equation. Starting with an uncharged capacitance ($q = 0$) and moving charge from one plate to the other until the plates have charge $+Q$ and $-Q$ requires the work W:

$$W_{\text{charging}} = \int_0^Q \frac{q}{C}\,\mathrm{d}q = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}CV^2 = W_{\text{stored}}.$$

Capacitors

The capacitance of the majority of capacitors used in electronic circuits is several orders of magnitude smaller than the farad. The most common subunits of capacitance in use today are the millifarad (mF), microfarad (µF), nanofarad (nF) and picofarad (pF).

The capacitance can be calculated if the geometry of the conductors and the dielectric properties of the insulator between the conductors are known. For example, the capacitance of a *parallel-plate* capacitor constructed of two parallel plates both of area *A* separated by a distance *d* is approximately equal to the following:

$$C = \varepsilon_r \varepsilon_0 \frac{A}{d},$$

where

> *C* is the capacitance;
> *A* is the area of overlap of the two plates;
> ε_r is the relative static permittivity (sometimes called the dielectric constant) of the material between the plates (for a vacuum, $\varepsilon_r = 1$);
> ε_0 is the electric constant ($\varepsilon_0 \approx 8.854 \times 10^{-12}$ F m^{-1}); and
> *d* is the separation between the plates.

Capacitance is proportional to the area of overlap and inversely proportional to the separation between conducting sheets. The closer the sheets are to each other, the greater the capacitance. The equation is a good approximation if *d* is small compared to the other dimensions of the plates so the field in the capacitor over most of its area is uniform, and the so-called *fringing field* around the periphery provides a small contribution. In CGS units the equation has the form:

$$C = \varepsilon_r \frac{A}{4\pi d}$$

where *C* in this case has the units of length.

Combining the SI equation for capacitance with the above equation for the energy stored in a capacitance, for a flat-plate capacitor the energy stored is:

$$W_{stored} = \frac{1}{2} C V^2 = \frac{1}{2} \varepsilon_r \varepsilon_0 \frac{A}{d} V^2.$$

where *W* is the energy, in joules; *C* is the capacitance, in farads; and *V* is the voltage, in volts.

Voltage dependent capacitors

The dielectric constant for a number of very useful dielectrics changes as a function of the applied electrical field, for example ferroelectric materials, so the capacitance for these devices is more complex. For example, in charging such a capacitor the differential increase in voltage with charge is governed by:

$$dQ = C(V)\, dV \,,$$

where the voltage dependence of capacitance, $C(V)$, stems from the field, which in a large area parallel plate device is given by $\varepsilon = V/d$. This field polarizes the dielectric, which polarization, in the case of a ferroelectric, is a nonlinear *S*-shaped function of field, which, in the case of a large area parallel plate device, translates into a capacitance that is a nonlinear function of the voltage causing the field.

Corresponding to the voltage-dependent capacitance, to charge the capacitor to voltage V an integral relation is found:

$$Q = \int_0^V dV\ C(V) \,,$$

which agrees with $Q = CV$ only when C is voltage independent.

By the same token, the energy stored in the capacitor now is given by

$$dW = QdV = \left[\int_0^V\ dV'\ C(V')\right]\ dV \ .$$

Integrating:

$$W = \int_0^V\ dV\ \int_0^V\ dV'\ C(V') = \int_0^V\ dV'\ \int_{V'}^V\ dV\ C(V')$$
$$= \int_0^V\ dV'\,(V - V')\,C(V') \,,$$

where interchange of the order of integration is used.

The nonlinear capacitance of a microscope probe scanned along a ferroelectric surface is used to study the domain structure of ferroelectric materials.

Another example of voltage dependent capacitance occurs in semiconductor devices such as semiconductor diodes, where the voltage dependence stems not from a change in dielectric constant but in a voltage dependence of the spacing between the charges on the two sides of the capacitor.

Frequency dependent capacitors

If a capacitor is driven with a time-varying voltage that changes rapidly enough, then the polarization of the dielectric cannot follow the signal. As an example of the origin of this mechanism, the internal microscopic dipoles contributing to the dielectric constant cannot move instantly, and so as frequency of an applied alternating voltage increases, the dipole response is limited and the dielectric constant diminishes. A changing dielectric constant with frequency is referred to as dielectric dispersion, and is governed by dielectric relaxation processes, such as Debye relaxation. Under transient conditions, the displacement field can be expressed as:

$$\boldsymbol{D}(t) = \varepsilon_0 \int_{-\infty}^{t} dt' \ \varepsilon_r(t-t')\boldsymbol{E}(t') \ ,$$

indicating the lag in response by the time dependence of ε_r, calculated in principle from an underlying microscopic analysis, for example, of the dipole behavior in the dielectric. For example, linear response function. The integral extends over the entire past history up to the present time. A Fourier transform in time then results in:

$$\boldsymbol{D}(\omega) = \varepsilon_0 \varepsilon_r(\omega)\boldsymbol{E}(\omega) \ ,$$

where $\varepsilon_r(\omega)$ is now a complex function, with an imaginary part related to absorption of energy from the field by the medium. The capacitance, being proportional to the dielectric constant, also exhibits this frequency behavior. Fourier transforming Gauss's law with this form for displacement field:

$$I(\omega) = j\omega Q(\omega) = j\omega \oint_{\Sigma} \boldsymbol{D}(\boldsymbol{r},\ \omega) \cdot d\boldsymbol{\Sigma}$$
$$= \left[G(\omega) + j\omega C(\omega)\right] V(\omega) = \frac{V(\omega)}{Z(\omega)} \ ,$$

where j is the imaginary unit, $V(\omega)$ is the voltage component at angular frequency ω, $G(\omega)$ is the *real* part of the current, called the *conductance*, and $C(\omega)$ determines the *imaginary* part of the current and is the *capacitance*. $Z(\omega)$ is the complex impedance.

When a parallel-plate capacitor is filled with a dielectric, the measurement of dielectric properties of the medium is based upon the relation:

$$\varepsilon_r(\omega) = \varepsilon'_r(\omega) - j\varepsilon''_r(\omega) = \frac{1}{j\omega Z(\omega) C_0} = \frac{C(\omega)}{C_0} \ ,$$

where a single *prime* denotes the real part and a double *prime* the imaginary part, $Z(\omega)$ is the complex impedance with the dielectric present, $C(\omega)$ is the so-called *complex* capacitance with the dielectric present, and C_0 is the capacitance without the dielectric.

(Measurement "without the dielectric" in principle means measurement in free space, an unattainable goal inasmuch as even the quantum vacuum is predicted to exhibit nonideal behavior, such as dichroism. For practical purposes, when measurement errors are taken into account, often a measurement in terrestrial vacuum, or simply a calculation of C_0, is sufficiently accurate.)

Using this measurement method, the dielectric constant may exhibit a resonance at certain frequencies corresponding to characteristic response frequencies (excitation energies) of contributors to the dielectric constant. These resonances are the basis for a number of experimental techniques for detecting defects. The *conductance method* measures absorption as a function of frequency. Alternatively, the time response of the capacitance can be used directly, as in *deep-level transient spectroscopy*.

Another example of frequency dependent capacitance occurs with MOS capacitors, where the slow generation of minority carriers means that at high frequencies the capacitance measures only the majority carrier response, while at low frequencies both types of carrier respond.

At optical frequencies, in semiconductors the dielectric constant exhibits structure related to the band structure of the solid. Sophisticated modulation spectroscopy measurement methods based upon modulating the crystal structure by pressure or by other stresses and observing the related changes in absorption or reflection of light have advanced our knowledge of these materials.

Capacitance matrix

The discussion above is limited to the case of two conducting plates, although of arbitrary size and shape. The definition C=Q/V still holds for a single plate given a charge, in which case the field lines produced by that charge terminate as if the plate were at the center of an oppositely charged sphere at infinity.

$C = Q / V$ does not apply when there are more than two charged plates, or when the net charge on the two plates is non-zero. To handle this case, Maxwell introduced his "coefficients of potential". If three plates are given charges Q_1,Q_2,Q_3, then the voltage of plate 1 is given by

$$V_1 = P_{11}Q_1 + P_{12}Q_2 + P_{13}Q_3 ,$$

and similarly for the other voltages. Maxwell showed that the coefficients of potential are symmetric, so that $P_{12} = P_{21}$, etc. Thus the system can be described by a collection of coefficients known as the "Reciprocal Capacitance Matrix"is used, which is defined as:

$$P_{ij} = \frac{V_i}{Q_j}$$

From this, the mutual capacitance C_m between two objects can be defined by solving for the total charge Q and using $C_m = Q / V$.

$$C_m = \frac{V}{(P_{11} + P_{22}) - (P_{12} + P_{21})}$$

Since no actual device holds perfectly equal and opposite charges on each of the two "plates", it is the mutual capacitance that is reported on capacitors. The collection of coefficients $C_{ij} = Q_i / V_j$ is known as the capacitance matrix and also describes the capacitance of the system.

Self-capacitance

In electrical circuits, the term *capacitance* is usually a shorthand for the *mutual capacitance* between two adjacent conductors, such as the two plates of a capacitor. There also exists a property called *self-capacitance*, which is the amount of electrical charge that must be added to an isolated conductor to raise its *electrical potential* by one volt. The reference point for this potential is a theoretical hollow conducting sphere, of infinite radius, centered on the conductor. Using this method, the self-capacitance of a conducting sphere of radius R is given by:

$$C = 4\pi\varepsilon_0 R$$

Example values of self-capacitance are:

- for the top "plate" of a van de Graaff generator, typically a sphere 20 cm in radius: 20 pF
- the planet Earth: about 709 µF

Elastance

The inverse of capacitance is called elastance. The unit of elastance is the daraf.

Stray capacitance

Any two adjacent conductors can be considered a capacitor, although the capacitance will be small unless the conductors are close together for long. This (often unwanted) effect is termed "stray capacitance". Stray capacitance can allow signals to leak between otherwise isolated circuits (an effect called crosstalk), and it can be a limiting factor for proper functioning of circuits at high frequency.

Stray capacitance is often encountered in amplifier circuits in the form of "feedthrough" capacitance that interconnects the input and output nodes (both defined relative to a common ground). It is often convenient for analytical purposes to replace this capacitance with a combination of one input-to-ground capacitance and one output-to-ground capacitance. (The original configuration — including the input-to-output capacitance — is often referred to as a pi-configuration.) Miller's theorem can be used to effect this replacement. Miller's theorem states that, if the gain ratio of two nodes is 1/K, then an impedance of Z connecting the two nodes can be replaced with a Z/(1-k) impedance between the first node and ground and a KZ/(K-1) impedance between the second node and ground. (Since impedance varies inversely with capacitance, the internode capacitance, C, will be seen to have been replaced by a capacitance of KC from input to ground and a capacitance of (K-1)C/K from output to ground.) When the input-to-output gain is very large, the equivalent input-to-ground impedance is very small while the output-to-ground impedance is essentially equal to the original (input-to-output) impedance.

Capacitance of simple systems

Calculating the capacitance of a system amounts to solving the Laplace equation $\nabla^2\varphi=0$ with a constant potential φ on the surface of the conductors. This is trivial in cases with high symmetry. There is no solution in terms of elementary functions in more complicated cases.

For quasi two-dimensional situations analytic functions may be used to map different geometries to each other.

Capacitance of simple systems

Parallel-plate capacitor

$$\varepsilon A/d$$

A: Area
d: Distance
ε: Permittivity

Coaxial cable

$$\frac{2\pi\varepsilon l}{\ln\left(a_2/a_1\right)}$$

a_1: Inner radius
a_2: Outer radius
l: Length

Pair of parallel wires

$$\frac{\pi\varepsilon l}{\operatorname{arcosh}\left(\frac{d}{2a}\right)}=\frac{\pi\varepsilon l}{\ln\left(\frac{d}{2a}+\sqrt{\frac{d^2}{4a^2}-1}\right)}$$

a: Wire radius
d: Distance, d > 2a
l: Length of pair

Wire parallel to wall

$$\frac{2\pi\varepsilon l}{\operatorname{arcosh}\left(\frac{d}{a}\right)} = \frac{2\pi\varepsilon l}{\ln\left(\frac{d}{a}+\sqrt{\frac{d^2}{a^2}-1}\right)}$$

a: Wire radius
d: Distance, d > a
l: Wire length

Two parallel coplanar strips

$$\varepsilon l \frac{K\left(\sqrt{1-k^2}\right)}{K\left(k\right)}$$

d: Distance
w_1, w_2: Strip width
k_i: $d/(2w_i+d)$

k^2: $k_1 k_2$
K: Elliptic integral
l: Length

Concentric spheres

$$\frac{4\pi\varepsilon a_1 a_2}{a_2 - a_1}$$

a_1: Inner radius
a_2: Outer radius

Two spheres, equal radius

$$2\pi\varepsilon a \sum_{n=1}^{\infty} \frac{\sinh\left(\ln\left(D+\sqrt{D^2-1}\right)\right)}{\sinh\left(n\ln\left(D+\sqrt{D^2-1}\right)\right)}$$
$$= 2\pi\varepsilon a \left\{1+\frac{1}{2D}+\frac{1}{4D^2}+\frac{1}{8D^3}+\frac{1}{8D^4}+\frac{3}{32D^5}+O\left(\frac{1}{D^6}\right)\right\}$$
$$= 2\pi\varepsilon a \left\{\ln 2+\gamma-\frac{1}{2}\ln\left(\frac{d}{a}-2\right)+O\left(\frac{d}{a}-2\right)\right\}$$

a: Radius
d: Distance, d > 2a
D = d/2a
γ: Euler's constant

Sphere in front of wall

$$4\pi\varepsilon a\sum_{n=1}^{\infty}\frac{\sinh\left(\ln\left(D+\sqrt{D^2-1}\right)\right)}{\sinh\left(n\ln\left(D+\sqrt{D^2-1}\right)\right)}$$

a: Radius
d: Distance, d > a
D = d/a

Sphere

$$4\pi\varepsilon a$$

a: Radius

Circular disc

$$8\varepsilon a$$

a: Radius

Thin straight wire, finite length

$$\frac{2\pi\varepsilon l}{\Lambda}\left\{1+\frac{1}{\Lambda}(1-\ln 2)+\frac{1}{\Lambda^2}\left[1+(1-\ln 2)^2-\frac{\pi^2}{12}\right]+O\left(\frac{1}{\Lambda^3}\right)\right\}$$

a: Wire radius
l: Length
Λ: ln(*l*/a)

Chapter 10

Inductance

Inductance is the property of an electrical circuit measuring the induced electric voltage compared to the rate of change of the electric current in the circuit. This property also is called **self inductance** to discriminate it from **mutual inductance**, describing the voltage induced in one electrical circuit by the rate of change of the electric current in another circuit.

The quantitative definition of the self inductance *L* of an electrical circuit in SI units (webers per ampere, known as henries) is

$$v = L\frac{di}{dt},$$

where *v* denotes the voltage in volts and *i* the current in amperes. This is a linear relation between voltage and current akin to Ohm's law, but with an extra time derivate. The simplest solutions of this equation are a constant current with no voltage or a current changing linearilly in time with a constant voltage.

The term 'inductance' was coined by Oliver Heaviside in February 1886. It is customary to use the symbol *L* for inductance, possibly in honour of the physicist Heinrich Lenz. The SI unit of inductance is the **henry** (H), named after American scientist and magnetic researcher Joseph Henry. 1 H = 1 Wb/A.

Inductance is caused by the magnetic field generated by electric currents according to Ampere's law. To add inductance to a circuit, electronic components called inductors are used, typically consisting of coils of wire to concentrate the magnetic field and to collect the induced voltage. This is analogous to adding capacitance to a circuit by adding capacitors. Capacitance is caused by the electric field generated by electric charge according to Gauss's law.

The generalization to the case of *K* electrical circuits with currents i_m and voltages v_m reads

$$v_m = \sum_{n=1}^{K} L_{m,n}\frac{di_n}{dt}.$$

Inductance here is a symmetric matrix. The diagonal coefficients $L_{m,m}$ are called coefficients of self inductance, the off-diagonal elements are called coefficients of mutual inductance. The coefficients of inductance are constant as long as no magnetizable material with nonlinear characteristics is involved. This is a direct consequence of the linearity of Maxwell's equations in the fields and the current density. The coefficients of inductance become functions of the currents in the nonlinear case.

Derivation from Faraday's law of inductance

The inductance equations above are a consequence of Maxwell's equations. There is a straightforward derivation in the important case of electrical circuits consisting of thin wires.

Consider a system of K wire loops, each with one or several wire turns. The flux linkage of loop m is given by

$$N_m\Phi_m = \sum_{n=1}^{K} L_{m,n} i_n.$$

Here N_m denotes the number of turns in loop m, Φ_m the magnetic flux through this loop, and $L_{m,n}$ are some constants. This equation follows from Ampere's law - magnetic fields and fluxes are linear functions of the currents. By Faraday's law of induction we have

$$v_m = N_m \frac{d\Phi_m}{dt} = \sum_{n=1}^{K} L_{m,n} \frac{di_n}{dt},$$

where v_m denotes the voltage induced in circuit m. This agrees with the definition of inductance above if the coefficients $L_{m,n}$ are identified with the coefficients of inductance. Because the total currents $N_n i_n$ contribute to Φ_m it also follows that $L_{m,n}$ is proportional to the product of turns $N_m N_n$.

Inductance and magnetic field energy

Multiplying the equation for v_m above with $i_m dt$ and summing over m gives the energy transferred to the system in the time interval dt,

$$\sum_{m}^{K} i_m v_m dt = \sum_{m,n=1}^{K} i_m L_{m,n} di_n \stackrel{!}{=} \sum_{n=1}^{K} \frac{\partial W\left(i\right)}{\partial i_n} di_n.$$

This must agree with the change of the magnetic field energy W caused by the currents. The integrability condition

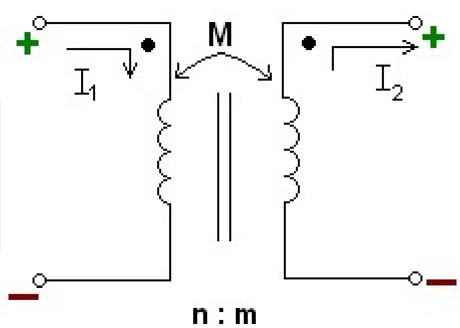
$$\partial^2 W/\partial i_m \partial i_n = \partial^2 W/\partial i_n \partial i_m$$

requires $L_{m,n}=L_{n,m}$. The inductance matrix $L_{m,n}$ thus is symmetric. The integral of the energy transfer is the magnetic field energy as a function of the currents,

$$W\left(i\right) = \frac{1}{2}\sum_{m,n=1}^{K} i_m L_{m,n} i_n.$$

This equation also is a direct consequence of the linearity of Maxwell's equations. It is helpful to associate changing electric currents with a build-up or decrease of magnet field energy. The corresponding energy transfer requires or generates a voltage. A mechanical analogy in the K=1 case with magnetic field energy $(1/2)Li^2$ is a body with mass M, velocity u and kinetic energy $(1/2)Mu^2$. The rate of change of velocity (current) multiplied with mass (inductance) requires or generates a force (an electrical voltage).

Coupled inductors



The circuit diagram representation of mutually coupled inductors. The two vertical lines between the inductors indicate a *solid core* that the wires of the inductor are wrapped around. "n:m" shows the ratio between the number of windings of the left inductor to windings of the right inductor. This picture also shows the dot convention.

Mutual inductance occurs when the change in current in one inductor induces a voltage in another nearby inductor. It is important as the mechanism by which transformers work, but it can also cause unwanted coupling between conductors in a circuit.

The mutual inductance, M, is also a measure of the coupling between two inductors. The mutual inductance by circuit i on circuit j is given by the double integral *Neumann formula*

The mutual inductance also has the relationship:

$$M_{21} = N_1 N_2 P_{21}$$

where

M_{21} is the mutual inductance, and the subscript specifies the relationship of the voltage induced in coil 2 to the current in coil 1.
N_1 is the number of turns in coil 1,
N_2 is the number of turns in coil 2,
P_{21} is the permeance of the space occupied by the flux.

The mutual inductance also has a relationship with the coupling coefficient. The coupling coefficient is always between 1 and 0, and is a convenient way to specify the relationship between a certain orientation of inductor with arbitrary inductance:

$$M = k\sqrt{L_1 L_2}$$

where

k is the *coupling coefficient* and $0 \leq k \leq 1$,
L_1 is the inductance of the first coil, and
L_2 is the inductance of the second coil.

Once the mutual inductance, M, is determined from this factor, it can be used to predict the behavior of a circuit:

$$V_1 = L_1 \frac{dI_1}{dt} - M \frac{dI_2}{dt}$$

where

V_1 is the voltage across the inductor of interest,
L_1 is the inductance of the inductor of interest,
dI_1/dt is the derivative, with respect to time, of the current through the inductor of interest,
dI_2/dt is the derivative, with respect to time, of the current through the inductor that is coupled to the first inductor, and
M is the mutual inductance.

The minus sign arises because of the sense the current I_2 has been defined in the diagram. With both currents defined going into the dots the sign of M will be positive.

When one inductor is closely coupled to another inductor through mutual inductance, such as in a transformer, the voltages, currents, and number of turns can be related in the following way:

$$V_s = \frac{N_s}{N_p} V_p$$

where

V_s is the voltage across the secondary inductor,
V_p is the voltage across the primary inductor (the one connected to a power source),
N_s is the number of turns in the secondary inductor, and
N_p is the number of turns in the primary inductor.

Conversely the current:

$$I_s = \frac{N_p}{N_s} I_p$$

where

I_s is the current through the secondary inductor,
I_p is the current through the primary inductor (the one connected to a power source),
N_s is the number of turns in the secondary inductor, and
N_p is the number of turns in the primary inductor.

Note that the power through one inductor is the same as the power through the other. Also note that these equations don't work if both transformers are forced (with power sources).

When either side of the transformer is a tuned circuit, the amount of mutual inductance between the two windings determines the shape of the frequency response curve. Although no boundaries are defined, this is often referred to as loose-, critical-, and over-coupling. When two tuned circuits are loosely coupled through mutual inductance, the bandwidth will be narrow. As the amount of mutual inductance increases, the bandwidth continues to grow. When the mutual inductance is increased beyond a critical point, the peak in the response curve begins to drop, and the center frequency will be attenuated more strongly than its direct sidebands. This is known as overcoupling.

Calculation techniques

In the most general case, inductance can be calculated from Maxwell's equations. Many important cases can be solved using simplifications. Where high frequency currents are considered, with skin effect, the surface current densities and magnetic field may be obtained by solving the Laplace equation. Where the conductors are thin wires, self inductance still depends on the wire radius and the distribution of the current in the wire. This current distribution is approximately constant (on the surface or in the volume of the wire) for a wire radius much smaller than other length scales.

Mutual inductance

The mutual inductance by a filamentary circuit *i* on a filamentary circuit *j* is given by the double integral *Neumann formula*

$$M_{ij} = \frac{\mu_0}{4\pi} \oint_{C_i} \oint_{C_j} \frac{\mathbf{ds}_i \cdot \mathbf{ds}_j}{|\mathbf{R}_{ij}|}$$

The symbol μ_0 denotes the magnetic constant ($4\pi\times10^{-7}$ H/m), 'C_i *and* C_j *are the curves spanned by the wires,* R_{ij} *is the distance between two points.*

Self-inductance

Formally the self-inductance of a wire loop would be given by the above equation with $i = j$. However, *1/R* now becomes infinite and thus the finite radius *a* along with the distribution of the current in the wire must be taken into account. There remain the contribution from the integral over all points where $|R| \geq a/2$ and a correction term,

$$M_{ii} = L \approx \left(\frac{\mu_0}{4\pi} \oint_{C} \oint_{C'} \frac{\mathbf{ds} \cdot \mathbf{ds}'}{|\mathbf{R}|} \right)_{|\mathbf{R}| \geq a/2} + \frac{\mu_0}{2\pi} l Y$$

Here *a* and *l* denote radius and length of the wire, and *Y* is a constant that depends on the distribution of the current in the wire: $Y = 0$ when the current flows in the surface of the wire (skin effect), $Y = 1/4$ when the current is homogeneous across the wire. This approximation is accurate when the wires are long compared to their cross-sectional dimensions. Here is a derivation of this equation.

Method of images

In some cases different current distributions generate the same magnetic field in some section of space. This fact may be used to relate self inductances (method of images). As an example consider the two systems:

- A wire at distance *d/2* in front of a perfectly conducting wall (which is the return)
- Two parallel wires at distance *d*, with opposite current

The magnetic field of the two systems coincides (in a half space). The magnetic field energy and the inductance of the second system thus are twice as large as that of the first system.

Relation between inductance and capacitance

Inductance per length L' and capacitance per length C' are related to each other in the special case of transmission lines consisting of two parallel perfect conductors of arbitrary but constant cross section,

$$L'C' = \varepsilon\mu.$$

Here ε and μ denote dielectric constant and magnetic permeability of the medium the conductors are embedded in. There is no electric and no magnetic field inside the conductors (complete skin effect, high frequency). Current flows down on one line and returns on the other. The signal propagation speed coincides with the propagation speed of electromagnetic waves in the bulk.

Self-inductance of simple electrical circuits in air

The self-inductance of many types of electrical circuits can be given in closed form. Examples are listed in the table.

Inductance of simple electrical circuits in air

Type	Inductance / μ_0	Comment
Single layer solenoid	$\frac{r^2N^2}{3l}\left\{-8w+4\frac{\sqrt{1+m}}{m}\left(K\left(\frac{m}{1+m}\right)-(1-m)E\left(\frac{m}{1+m}\right)\right)\right\}$ $=\frac{r^2N^2\pi}{l}\left\{1-\frac{8w}{3\pi}+\sum_{n=1}^{\infty}\left(\frac{1\cdot 3...(2n-3)}{2\cdot 4\cdot 6...2n}\right)^2\frac{2n-1}{2n+2}2^{2n+1}(-1)^{n+1}w^{2n}\right\}$ $=\frac{r^2N^2\pi}{l}\left(1-\frac{8w}{3\pi}+\frac{w^2}{2}-\frac{w^4}{4}+\frac{5w^6}{16}-\frac{35w^8}{64}+...\right)$ for w << 1 $=rN^2\left\{\left(1+\frac{1}{32w^2}+O(\frac{1}{w^4})\right)\ln 8w-1/2+\frac{1}{128w^2}+O(\frac{1}{w^4})\right\}$ for w >> 1	*N*: Number of turns *r*: Radius *l*: Length $w = r/l$ $m = 4w^2$ *E*,*K*: Elliptic integrals
Coaxial cable, high frequency	$\frac{l}{2\pi}\ln\frac{a_1}{a}$	a_1: Outer radius a: Inner radius *l*: Length

Circular loop	$r \cdot \left(\ln \frac{8r}{a} - 2 + Y\right)$	r: Loop radius a: Wire radius
Rectangle	$\frac{1}{\pi}\left(b\ln\frac{2b}{a} + d\ln\frac{2d}{a} - (b+d)(2-Y) + 2\sqrt{b^2+d^2} - b\cdot\operatorname{arsinh}\frac{b}{d} - d\cdot\operatorname{arsinh}\frac{d}{b}\right)$	b, d: Border length d >> a, b >> a a: Wire radius
Pair of parallel wires	$\frac{l}{\pi}\left(\ln\frac{d}{a} + Y\right)$	a: Wire radius d: Distance, d ≥ 2a *l*: Length of pair
Pair of parallel wires, high frequency	$\frac{l}{\pi}\operatorname{arcosh}\left(\frac{d}{2a}\right) = \frac{l}{\pi}\ln\left(\frac{d}{2a} + \sqrt{\frac{d^2}{4a^2} - 1}\right)$	a: Wire radius d: Distance, d ≥ 2a *l*: Length of pair
Wire parallel to perfectly conducting wall	$\frac{l}{2\pi}\left(\ln\frac{2d}{a} + Y\right)$	a: Wire radius d: Distance, d ≥ a *l*: Length
Wire parallel to conducting wall,	$\frac{l}{2\pi}\operatorname{arcosh}\left(\frac{d}{a}\right) = \frac{l}{2\pi}\ln\left(\frac{d}{a} + \sqrt{\frac{d^2}{a^2} - 1}\right)$	a: Wire radius d: Distance, d ≥ a *l*: Length

high frequency

The symbol μ_0 denotes the magnetic constant ($4<\pi\times10^{-7}$ H/m). For high frequencies the electric current flows in the conductor surface (skin effect), and depending on the geometry it sometimes is necessary to distinguish low and high frequency inductances. This is the purpose of the constant Y: $Y = 0$ when the current is uniformly distributed over the surface of the wire (skin effect), $Y = 1/4$ when the current is uniformly distributed over the cross section of the wire. In the high frequency case, if conductors approach each other, an additional screening current flows in their surface, and expressions containing Y become invalid. Details for some circuit types are available on another page.

Phasor circuit analysis and impedance

Using phasors, the equivalent impedance of an inductance is given by:

$$Z_L = V/I = jL\omega$$

where

j is the imaginary unit,
L is the inductance,
$\omega = 2\pi f$ is the angular frequency,
f is the frequency and
$L\omega = X_L$ is the inductive reactance.

Nonlinear inductance

Many inductors make use of magnetic materials. These materials over a large enough range exhibit a nonlinear permeability with such effects as saturation. This in-turn makes the resulting inductance a function of the applied current. Faraday's Law still holds but inductance is ambiguous and is different whether you are calculating circuit parameters or magnetic fluxes.

The secant or large-signal inductance is used in flux calculations. It is defined as:

$$L_s(i) \overset{\text{def}}{=} \frac{N\Phi}{i} = \frac{\Lambda}{i}$$

The differential or small-signal inductance, on the other hand, is used in calculating voltage. It is defined as:

$$L_d(i) \stackrel{\text{def}}{=} \frac{d(N\Phi)}{di} = \frac{d\Lambda}{di}$$

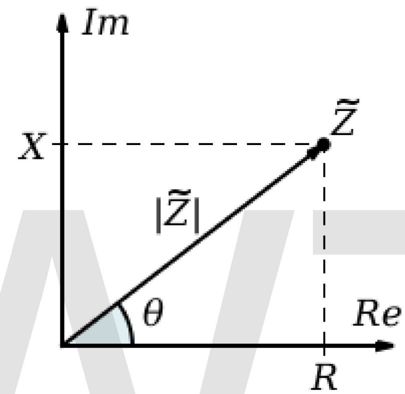
The circuit voltage for a nonlinear inductor is obtained via the differential inductance as shown by Faraday's Law and the chain rule of calculus.

$$v(t) = \frac{d\Lambda}{dt} = \frac{d\Lambda}{di}\frac{di}{dt} = L_d(i)\frac{di}{dt}$$

There are similar definitions for nonlinear mutual inductances.

Chapter 11

Electrical Impedance



A graphical representation of the complex impedance plane

Electrical impedance, or simply **impedance**, describes a measure of opposition to alternating current (AC). Electrical impedance extends the concept of resistance to AC circuits, describing not only the relative amplitudes of the voltage and current, but also the relative phases. When the circuit is driven with direct current (DC) there is no distinction between impedance and resistance; the latter can be thought of as impedance with zero phase angle.

The symbol for impedance is usually Z and it may be represented by writing its magnitude and phase in the form $|Z|\angle\theta$. However, complex number representation is more powerful for circuit analysis purposes. The term *impedance* was coined by Oliver Heaviside in July 1886. Arthur Kennelly was the first to represent impedance with complex numbers in 1893.

Impedance is defined as the frequency domain ratio of the voltage to the current. In other words, it is the voltage–current ratio for a single complex exponential at a particular frequency ω. In general, impedance will be a complex number, with the same units as resistance, for which the SI unit is the ohm. For a sinusoidal current or voltage input, the polar form of the complex impedance relates the amplitude and phase of the voltage and current. In particular,

- The magnitude of the complex impedance is the ratio of the voltage amplitude to the current amplitude.
- The phase of the complex impedance is the phase shift by which the current is ahead of the voltage.

The reciprocal of impedance is admittance (i.e., admittance is the current-to-voltage ratio, and it conventionally carries units of siemens, formerly called mhos).

Complex impedance

Impedance is represented as a complex quantity $Z$and the term *complex impedance* may be used interchangeably; the polar form conveniently captures both magnitude and phase characteristics,

$$Z = |Z|e^{j\theta}$$

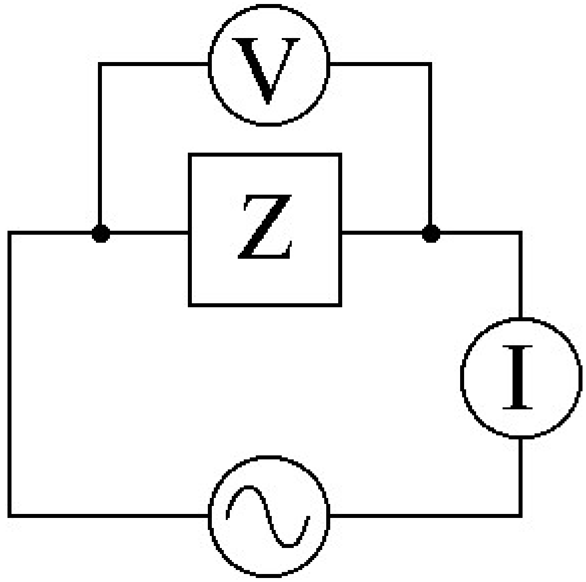
where the magnitude $|Z|$represents the ratio of the voltage difference amplitude to the current amplitude, while the argument θgives the phase difference between voltage and current and $j$is the imaginary unit. In Cartesian form,

$$Z = R + jX$$

where the real part of impedance is the resistance $R$and the imaginary part is the reactance X.

Where it is required to add or subtract impedances the cartesian form is more convenient, but when quantities are multiplied or divided the calculation becomes simpler if the polar form is used. A circuit calculation, such as finding the total impedance of two impedances in parallel, may require conversion between forms several times during the calculation. Conversion between the forms follows the normal conversion rules of complex numbers.

Ohm's law



An AC supply applying a voltage V, across a load Z, driving a current I

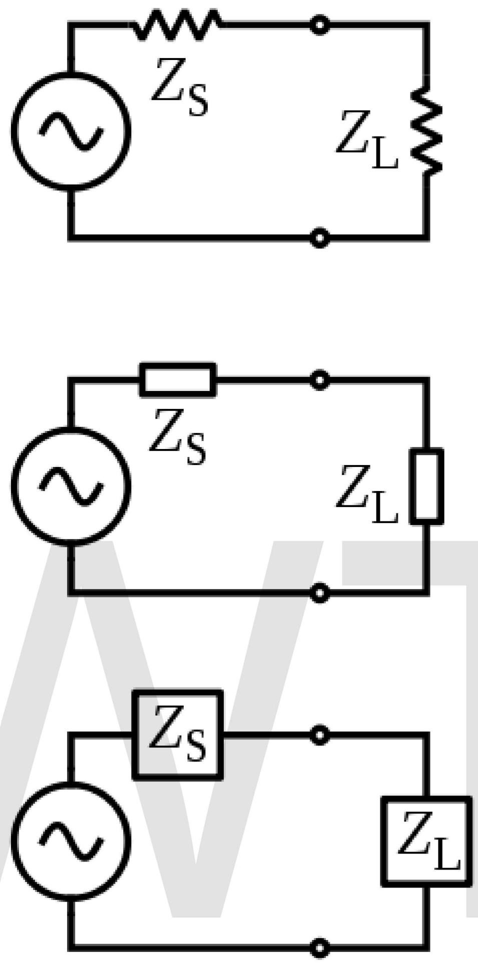
The meaning of electrical impedance can be understood by substituting it into Ohm's law.

$$V = IZ = I|Z|e^{j\theta}$$

The magnitude of the impedance $|Z|$ acts just like resistance, giving the drop in voltage amplitude across an impedance Z for a given current I. The phase factor tells us that the current lags the voltage by a phase of θ (i.e. in the time domain, the current signal is shifted $\frac{\theta}{2\pi}T$ to the right with respect to the voltage signal).

Just as impedance extends Ohm's law to cover AC circuits, other results from DC circuit analysis such as voltage division, current division, Thevenin's theorem, and Norton's theorem, can also be extended to AC circuits by replacing resistance with impedance.

Complex voltage and current



Generalized impedances in a circuit can be drawn with the same symbol as a resistor (US ANSI or DIN Euro) or with a labeled box.

In order to simplify calculations, sinusoidal voltage and current waves are commonly represented as complex-valued functions of time denoted as $V$and I.

$$V = |V|e^{j(\omega t + \phi_V)}$$
$$I = |I|e^{j(\omega t + \phi_I)}$$

Impedance is defined as the ratio of these quantities.

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

Substituting these into Ohm's law we have

$$
\begin{aligned}
|V|e^{j(\omega t+\phi_V)} &= |I|e^{j(\omega t+\phi_I)}|Z|e^{j\theta} \\
&= |I||Z|e^{j(\omega t+\phi_I+\theta)}
\end{aligned}
$$

Noting that this must hold for all *t*, we may equate the magnitudes and phases to obtain

$$
\begin{aligned}
&|V| = |I||Z| \\
&\phi_V = \phi_I + \theta
\end{aligned}
$$

The magnitude equation is the familiar Ohm's law applied to the voltage and current amplitudes, while the second equation defines the phase relationship.

Validity of complex representation

This representation using complex exponentials may be justified by noting that (by Euler's formula):

$$
\cos(\omega t+\phi) = \frac{1}{2}\Big[e^{j(\omega t+\phi)} + e^{-j(\omega t+\phi)}\Big]
$$

i.e. a real-valued sinusoidal function (which may represent our voltage or current waveform) may be broken into two complex-valued functions. By the principle of superposition, we may analyse the behaviour of the sinusoid on the left-hand side by analysing the behaviour of the two complex terms on the right-hand side. Given the symmetry, we only need to perform the analysis for one right-hand term; the results will be identical for the other. At the end of any calculation, we may return to real-valued sinusoids by further noting that

$$
\cos(\omega t+\phi) = \Re\Big\{e^{j(\omega t+\phi)}\Big\}
$$

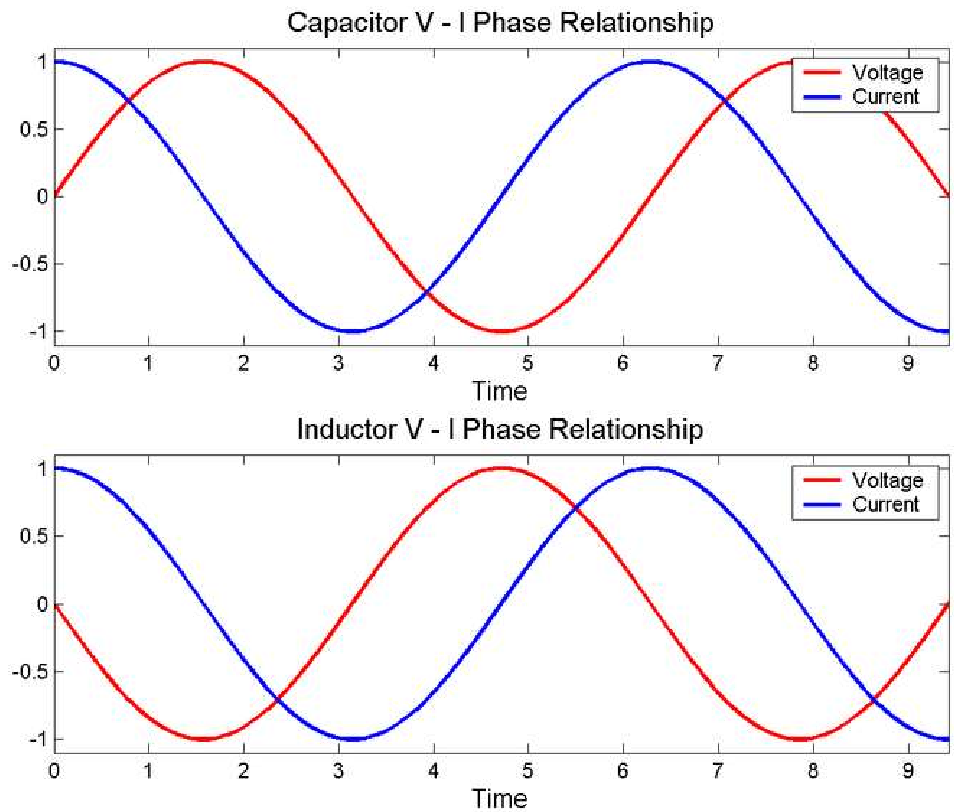
In other words, we simply take the real part of the result.

Phasors

A phasor is a constant complex number, usually expressed in exponential form, representing the complex amplitude (magnitude and phase) of a sinusoidal function of time. Phasors are used by electrical engineers to simplify computations involving sinusoids, where they can often reduce a differential equation problem to an algebraic one.

The impedance of a circuit element can be defined as the ratio of the phasor voltage across the element to the phasor current through the element, as determined by the relative amplitudes and phases of the voltage and current. This is identical to the definition from Ohm's law given above, recognising that the factors of $e^{j\omega t}$ cancel.

Device examples



The phase angles in the equations for the impedance of inductors and capacitors indicate that the voltage across a capacitor *lags* the current through it by a phase of $\pi / 2$, while the voltage across an inductor *leads* the current through it by $\pi / 2$. The identical voltage and current amplitudes tell us that the magnitude of the impedance is equal to one.

The impedance of an ideal resistor is purely real and is referred to as a *resistive impedance*:

$$Z_R = R.$$

Ideal inductors and capacitors have a purely imaginary *reactive impedance*:

$$Z_L = j\omega L,$$
$$Z_C = \frac{1}{j\omega C}.$$

Note the following identities for the imaginary unit and its reciprocal:

$$j = \cos\left(\frac{\pi}{2}\right) + j\sin\left(\frac{\pi}{2}\right) = e^{j\frac{\pi}{2}},$$
$$\frac{1}{j} = -j = \cos\left(-\frac{\pi}{2}\right) + j\sin\left(-\frac{\pi}{2}\right) = e^{j(-\frac{\pi}{2})}.$$

Thus we can rewrite the inductor and capacitor impedance equations in polar form:

$$Z_L = \omega L e^{j\frac{\pi}{2}},$$
$$Z_C = \frac{1}{\omega C} e^{j(-\frac{\pi}{2})}.$$

The magnitude tells us the change in voltage amplitude for a given current amplitude through the impedance, while the exponential factors give the phase relationship.

Deriving the device specific impedances

What follows below is a derivation of impedance for each of the three basic circuit elements, the resistor, the capacitor, and the inductor. Although the idea can be extended to define the relationship between the voltage and current of any arbitrary signal, these derivations will assume sinusoidal signals, since any arbitrary signal can be approximated as a sum of sinusoids through Fourier Analysis.

Resistor

For a resistor, we have the relation:

$$v_{\mathrm{R}}(t) = i_{\mathrm{R}}(t) R.$$

This is simply a statement of Ohm's Law.

Considering the voltage signal to be

$$v_{\mathrm{R}}(t) = V_p \sin(\omega t),$$

it follows that

$$\frac{v_{\mathrm{R}}(t)}{i_{\mathrm{R}}(t)} = \frac{V_p \sin(\omega t)}{I_p \sin(\omega t)} = R.$$

This tells us that the ratio of AC voltage amplitude to AC current amplitude across a resistor is R, and that the AC voltage leads the AC current across a resistor by 0 degrees.

This result is commonly expressed as

$$Z_{\text{resistor}} = R\,.$$

Capacitor

For a capacitor, we have the relation:

$$i_{\text{C}}(t) = C\frac{\mathrm{d}v_{\text{C}}(t)}{\mathrm{d}t}.$$

Considering the voltage signal to be

$$v_{\text{C}}(t) = V_p \sin(\omega t)$$

it follows that

$$\frac{\mathrm{d}v_{\text{C}}(t)}{\mathrm{d}t} = \omega V_p \cos\left(\omega t\right).$$

And thus

$$\frac{v_{\text{C}}\left(t\right)}{i_{\text{C}}\left(t\right)} = \frac{V_p \sin(\omega t)}{\omega V_p C \cos\left(\omega t\right)} = \frac{\sin(\omega t)}{\omega C \sin\left(\omega t + \frac{\pi}{2}\right)}.$$

This tells us that the ratio of AC voltage amplitude to AC current amplitude across a capacitor is $\frac{1}{\omega C}$, and that the AC voltage leads the AC current across a capacitor by -90 degrees (or the AC current leads the AC voltage across a capacitor by 90 degrees).

This result is commonly expressed in polar form, as

$$Z_{\text{capacitor}} = \frac{1}{\omega C} e^{-j\frac{\pi}{2}}$$

or, by applying Euler's formula, as

$$Z_{\text{capacitor}} = -j\frac{1}{\omega C} = \frac{1}{j\omega C}\,.$$

Inductor

For the inductor, we have the relation:

$$v_{\text{L}}(t) = L\frac{\mathrm{d}i_{\text{L}}(t)}{\mathrm{d}t}.$$

This time, considering the current signal to be

$$i_{\mathrm{L}}(t) = I_p \sin(\omega t)\,,$$

it follows that

$$\frac{\mathrm{d}i_{\mathrm{L}}(t)}{\mathrm{d}t} = \omega I_p \cos\left(\omega t\right).$$

And thus

$$\frac{v_{\mathrm{L}}\left(t\right)}{i_{\mathrm{L}}\left(t\right)} = \frac{\omega I_p L \cos(\omega t)}{I_p \sin\left(\omega t\right)} = \frac{\omega L \sin\left(\omega t + \frac{\pi}{2}\right)}{\sin(\omega t)}.$$

This tells us that the ratio of AC voltage amplitude to AC current amplitude across an inductor is ωL, and that the AC voltage leads the AC current across an inductor by 90 degrees.

This result is commonly expressed in polar form, as

$$Z_{\text{inductor}} = \omega L e^{j\frac{\pi}{2}}.$$

Or, more simply, using Euler's formula, as

$$Z_{\text{inductor}} = j\omega L.$$

Generalised s-plane impedance

Impedance defined in terms of $j\omega$ can strictly only be applied to circuits which are energised with a steady-state AC signal. The concept of impedance can be extended to a circuit energised with any arbitrary signal by using complex frequency instead of $j\omega$. Complex frequency is given the symbol s and is, in general, a complex number. Signals are expressed in terms of complex frequency by taking the Laplace transform of the time domain expression of the signal. The impedance of the basic circuit elements in this more general notation is as follows:

Element	Impedance expression
Resistor	R
Inductor	sL
Capacitor	$\frac{1}{sC}$

For a DC circuit this simplifies to $s = 0$. For a steady-state sinusoidal AC signal $s = j\omega$.

Resistance vs reactance

Resistance and reactance together determine the magnitude and phase of the impedance through the following relations:

$$|Z| = \sqrt{ZZ^*} = \sqrt{R^2 + X^2}$$
$$\theta = \arctan\left(\frac{X}{R}\right)$$

In many applications the relative phase of the voltage and current is not critical so only the magnitude of the impedance is significant.

Resistance

Resistance R is the real part of impedance; a device with a purely resistive impedance exhibits no phase shift between the voltage and current.

$$R = |Z| \cos \theta$$

Reactance

Reactance X is the imaginary part of the impedance; a component with a finite reactance induces a phase shift θ between the voltage across it and the current through it.

$$X = |Z| \sin \theta$$

A purely reactive component is distinguished by the fact that the sinusoidal voltage across the component is in quadrature with the sinusoidal current through the component. This implies that the component alternately absorbs energy from the circuit and then returns energy to the circuit. A pure reactance will not dissipate any power.

Capacitive reactance

A capacitor has a purely reactive impedance which is inversely proportional to the signal frequency. A capacitor consists of two conductors separated by an insulator, also known as a dielectric.

At low frequencies a capacitor is open circuit, as no charge flows in the dielectric. A DC voltage applied across a capacitor causes charge to accumulate on one side; the electric field due to the accumulated charge is the source of the opposition to the current. When the potential associated with the charge exactly balances the applied voltage, the current goes to zero.

Driven by an AC supply, a capacitor will only accumulate a limited amount of charge before the potential difference changes sign and the charge dissipates. The higher the frequency, the less charge will accumulate and the smaller the opposition to the current.

Inductive reactance

Inductive reactance $X_L$is proportional to the signal frequency $f$and the inductance L.

$$X_L = \omega L = 2\pi f L$$

An inductor consists of a coiled conductor. Faraday's law of electromagnetic induction gives the back emf $\mathcal{E}$(voltage opposing current) due to a rate-of-change of magnetic flux density Bthrough a current loop.

$$\mathcal{E} = -\frac{d\Phi_B}{dt}$$

For an inductor consisting of a coil with *N* loops this gives.

$$\mathcal{E} = -N\frac{d\Phi_B}{dt}$$

The back-emf is the source of the opposition to current flow. A constant direct current has a zero rate-of-change, and sees an inductor as a short-circuit (it is typically made from a material with a low resistivity). An alternating current has a time-averaged rate-of-change that is proportional to frequency, this causes the increase in inductive reactance with frequency.

Combining impedances

The total impedance of many simple networks of components can be calculated using the rules for combining impedances in series and parallel. The rules are identical to those used for combining resistances, except that the numbers in general will be complex numbers. In the general case however, equivalent impedance transforms in addition to series and parallel will be required.

Series combination

For components connected in series, the current through each circuit element is the same; the total impedance is simply the sum of the component impedances.

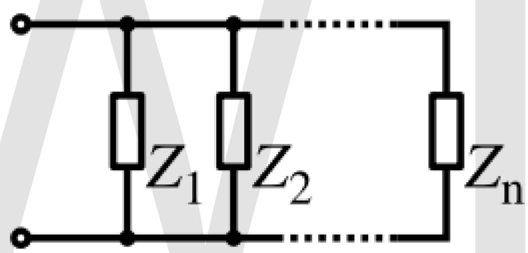
$$Z_{\text{eq}} = Z_1 + Z_2 + \cdots + Z_n$$

Or explicitly in real and imaginary terms:

$$Z_{\text{eq}} = R + jX = (R_1 + R_2 + \cdots + R_n) + j(X_1 + X_2 + \cdots + X_n)$$

Parallel combination

For components connected in parallel, the voltage across each circuit element is the same; the ratio of currents through any two elements is the inverse ratio of their impedances.



Hence the inverse total impedance is the sum of the inverses of the component impedances:

$$\frac{1}{Z_{\text{eq}}} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}$$

or, when n = 2:

$$Z_{\text{eq}} = Z_1 \| Z_2 \stackrel{\text{def}}{=} \frac{Z_1 Z_2}{Z_1 + Z_2}$$

The equivalent impedance Z_{eq}can be calculated in terms of the equivalent resistance $R_{\text{eq}}$and reactance X_{eq}.

$$Z_{\text{eq}} = R_{\text{eq}} + jX_{\text{eq}}$$

$$R_{\rm eq} = \frac{(X_1 R_2 + X_2 R_1)(X_1 + X_2) + (R_1 R_2 - X_1 X_2)(R_1 + R_2)}{(R_1 + R_2)^2 + (X_1 + X_2)^2}$$
$$X_{\rm eq} = \frac{(X_1 R_2 + X_2 R_1)(R_1 + R_2) - (R_1 R_2 - X_1 X_2)(X_1 + X_2)}{(R_1 + R_2)^2 + (X_1 + X_2)^2}$$

Measurement

According to Ohm's law the impedance of a device can be calculated by complex division of the voltage and current. The impedance of the device can be calculated by applying a sinusoidal voltage to the device in series with a resistor, and measuring the voltage across the resistor and across the device. Performing this measurement by sweeping the frequencies of the applied signal provides the impedance phase and magnitude.

Impulse impedance spectroscopy

The use of an impulse response may be used in combination with the fast Fourier transform (FFT) to rapidly measure the electrical impedance of various electrical devices. The technique compares well to other methodologies such as network and impedance analyzers while providing additional versatility in the electrical impedance measurement. The technique is theoretically simple, easy to implement and completed with ordinary laboratory instrumentation for minimal cost.

Variable impedance

In general, neither impedance nor admittance can be time varying as they are defined for complex exponentials for $-\infty < t < +\infty$. If the complex exponential voltage–current ratio changes over time or amplitude, the circuit element cannot be described using the frequency domain. However, many systems (e.g., varicaps that are used in radio tuners) may exhibit non-linear or time-varying voltage–current ratios that appear to be LTI for small signals over small observation windows; hence, they can be roughly described as having a time-varying impedance. That is, this description is an approximation; over large signal swings or observation windows, the voltage–current relationship is non-LTI and cannot be described by impedance.

Chapter 12

Current Source

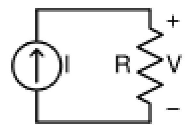
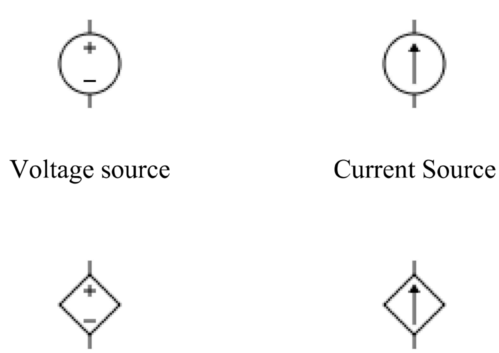


Figure 1: An ideal current source, *I*, driving a resistor, *R*, and creating a voltage *V*

A **current source** is an electrical or electronic device that delivers or absorbs electric current. A current source is the dual of a voltage source. The term constant-current **sink** is sometimes used for sources fed from a negative voltage supply. Figure 1 shows a schematic for an ideal current source driving a resistor load.

Ideal current sources

In circuit theory, an **ideal current source** is a circuit element where the current through it is independent of the voltage across it. It is a mathematical model, which real devices can only approach in performance. If the current through an ideal current source can be specified independently of any other variable in a circuit, it is called an *independent* current source. Conversely, if the current through an ideal current source is determined by some other voltage or current in a circuit, it is called a **dependent** or **controlled current source**. Symbols for these sources are shown in Figure 2.



Controlled Voltage Source Controlled Current Source

Battery of cells Single cell

Figure 2: Source symbols

An independent current source with zero current is identical to an ideal open circuit. For this reason, the internal resistance of an ideal current source is infinite. The voltage across an ideal current source is completely determined by the circuit it is connected to. When connected to a short circuit, there is zero voltage and thus zero power delivered. When connected to a load resistance, the voltage across the source approaches infinity as the load resistance approaches infinity (an open circuit). Thus, an ideal current source, if such a thing existed in reality, could supply unlimited power and so would represent an unlimited source of energy.

No real current source is ideal (no unlimited energy sources exist) and all have a finite internal resistance (none can supply unlimited voltage). However, the internal resistance of a physical current source is effectively modeled in circuit analysis by combining a non-zero resistance in parallel with an ideal current source (the Norton equivalent circuit). The connection of an ideal open circuit to an ideal non-zero current source does not represent any physically realizable system.

Physical current sources

Resistor current source

The simplest current source consists of a voltage source in series with a resistor. The current available from such a source is given by the ratio of the voltage across the voltage source to the resistance of the resistor. For a nearly ideal current source, the value of this resistor should be very large but this implies that, for a specified current, the voltage source must be very large. Thus, efficiency is low (due to power loss in the resistor) and it is usually impractical to construct a 'good' current source this way. Nonetheless, it is often the case that such a circuit will provide adequate performance when the specified current and load resistance are small. For example, a 5 V voltage source in series with a 4.7 kilohm resistor will provide an *approximately* constant current of 1 mA (±5%) to a load resistance in the range of 50 to 450 ohm.

Active current sources

Active current sources have many important applications in electronic circuits. Current sources (current-stable resistors) are often used in place of ohmic resistors in analog integrated circuits to generate a current without causing attenuation at a point in the signal path to which the current source is attached. The collector of a bipolar transistor, the drain of a field effect transistor, or the plate of a vacuum tube naturally behave as current sources (or sinks) when properly connected to an external source of energy (such as a power supply) because the output impedance of these devices is naturally high when used in the current source configuration.

JFET and N-FET current source

A JFET can be made to act as a current source by tying its gate to its source. The current then flowing is the I_{DSS} of the FET. These can be purchased with this connection already made and in this case the devices are called current regulator diodes or constant current diodes or current limiting diodes (CLD). An enhancement mode N channel MOSFET can be used in the circuits listed below.

Simple transistor current source

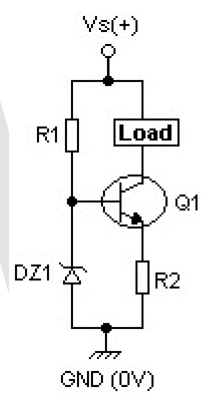


Figure 3: Typical constant current source (CCS)

Figure 3 shows a typical constant current source (CCS). DZ1 is a zener diode which, when reverse biased (as shown in the circuit) has a constant voltage drop across it irrespective of the current flowing through it. Thus, as long as the zener current (I_Z) is above a certain level (called holding current), the voltage across the zener diode (V_Z) will be constant. Resistor R1 supplies the zener current and the base current (I_B) of NPN transistor (Q1). The constant zener voltage is applied across the base of Q1 and emitter resistor R2. The operation of the circuit is as follows:

Voltage across R2 (V_{R2}) is given by V_Z - V_{BE}, where V_{BE} is the base-emitter drop of Q1. The emitter current of Q1 which is also the current through R2 is given by

$$I_{R2}(= I_E) = \frac{V_{R2}}{R2} = \frac{V_Z - V_{BE}}{R2}$$

Since V_Z is constant and V_{BE} is also (approximately) constant for a given temperature, it follows that V_{R2} is constant and hence I_E is also constant. Due to transistor action, emitter current I_E is very nearly equal to the collector current I_C of the transistor (which in turn, is the current through the load). Thus, the load current is constant (neglecting the output resistance of the transistor due to the Early effect) and the circuit operates as a constant current source. As long as the temperature remains constant (or doesn't vary much), the load current will be independent of the supply voltage, R1 and the transistor's gain. R2 allows the load current to be set at any desirable value and is calculated by

$R2 = \frac{V_Z - V_{BE}}{I_{R2}}$ or $R2 = \frac{V_Z - 0.65}{I_{R2}}$, since V_{BE} is typically 0.65 V for a silicon device.

(I_{R2} is also the emitter current and is assumed to be the same as the collector or required load current, provided h_{FE} is sufficiently large). Resistance R_1 at resistor R1 is calculated as

$R_1 = \frac{V_S - V_Z}{I_Z + K \cdot I_B}$ where, K = 1.2 to 2 (so that R_1 is low enough to ensure adequate I_B), $I_B = \frac{I_C (= I_E = I_{R2})}{h_{FE(min)}}$ and $h_{FE(min)}$ is the lowest acceptable current gain for the particular transistor type being used.

A more common current source in integrated circuits is the current mirror.

Simple transistor current source with diode compensation

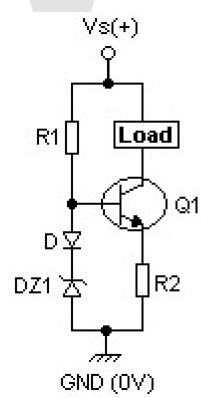


Figure 4: Typical constant current source (CCS) with diode compensation

Temperature changes will change the output current delivered by the circuit of Figure 3 because V_{BE} is sensitive to temperature. Temperature dependence can be compensated using the circuit of Figure 4 that includes a standard diode D (of the same semiconductor material as the transistor) in series with the Zener diode as shown in the image on the left. The diode drop (V_D) tracks the V_{BE} changes due to temperature and thus significantly counteracts temperature dependence of the CCS.

Resistance R_2 is now calculated as

$$R_2 = \frac{V_Z + V_D - V_{BE}}{I_{R2}}$$

Since $V_D = V_{BE} = 0.65$ V,

Therefore, $$R_2 = \frac{V_Z}{I_{R2}}$$

(In practice V_D is never exactly equal to V_{BE} and hence it only suppresses the change in V_{BE} rather than nulling it out.)

and R_1 is calculated as

$$R_1 = \frac{V_S - V_Z - V_D}{I_Z + K \cdot I_B}$$ (the compensating diode's forward voltage drop V_D appears in the equation and is typically 0.65 V for silicon devices.)

This method is most effective for Zener diodes rated at 5.6 V or more. For breakdown diodes of less than 5.6 V, the compensating diode is usually not required because the breakdown mechanism is not as temperature dependent as it is in breakdown diodes above this voltage.

Simple transistor current source with LED

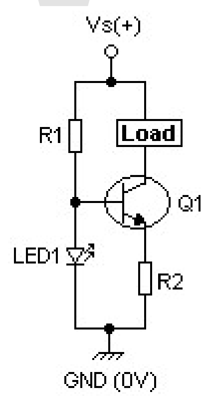


Figure 5: Typical constant current source (CCS) using LED instead of zener

Another method is to replace the Zener diode with a light-emitting diode LED1 as shown in Figure 5. The LED voltage drop (V_D) is now used to derive the constant voltage and also has the additional advantage of tracking (compensating) V_{BE} changes due to temperature. R_2 is calculated as

$$R_2 = \frac{V_D - V_{BE}}{I_{R2}}$$

and R_1 as

$$R_1 = \frac{V_S - V_D}{I_D + K \cdot I_B}$$, where I_D is the LED current.

Feedback

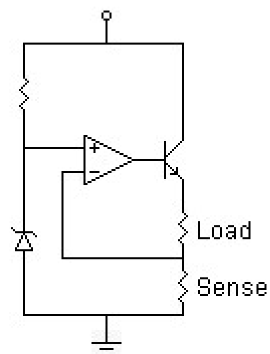


Figure 6: Typical op-amp current source. The transistor is not needed if the required current doesn't exceed the sourcing ability of the op-amp. The current will be the zener voltage divided by the sense resistor.

Another common method is to use feedback to set the current and remove the dependence on the Vbe of the transistor. Figure 6 shows a very common approach using an op amp with the non-inverting input connected to a voltage source (such as the Zener in an above example) and the inverting input connected to the same node as the resistor and emitter of the transistor. This way the generated voltage is across the resistor, rather than both the resistor and transistor.

Current mirror

Feedback is also used in the two-transistor current mirror with emitter degeneration. Feedback is a basic feature in some current mirrors using multiple transistors, such as the Widlar current source and the Wilson current source.

Other practical sources

In the case of opamp circuits sometimes it is desired to inject a precisely known current to the inverting input (as an offset of signal input for instance) and a resistor connected between the source voltage and the inverting input will approximate an ideal current source with value V/R.

Current source made by a voltage regulator

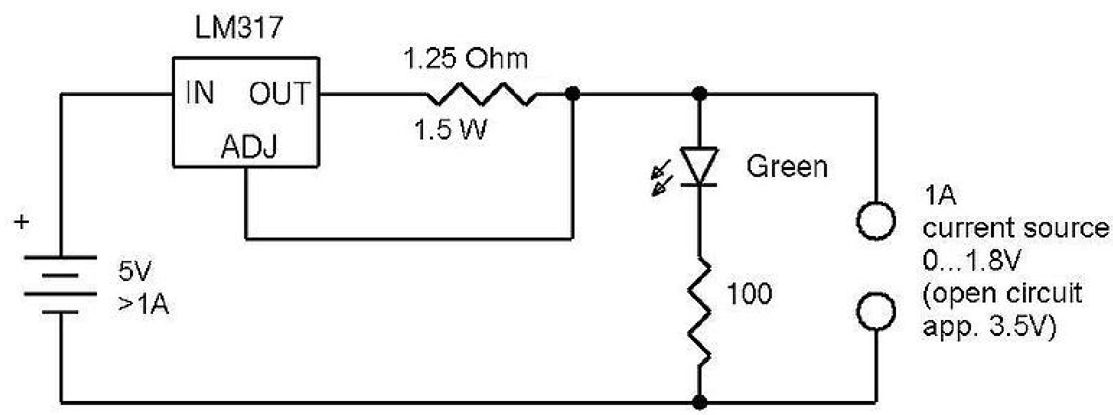


Figure 7: Constant current source using the LM317 voltage regulator

The circuit of Figure 7 using the LM317 voltage regulator is used to present a constant current source. The voltage regulator keeps up a constant voltage drop (1.25 V) across a constant resistor (1.25 Ω); so, a constant current (1 A) flows through the resistor and the load. The LED is on when the voltage across the load exceeds 1.8 V (the indicator circuit introduces some error).

High voltage current sources

A Van de Graaff generator behaves as a current source because of its very high output voltage coupled with its very high output resistance and so it supplies the same few microamperes at any output voltage up to hundreds of thousands of volts (or even tens of megavolts) for large laboratory versions.

Current and voltage source comparison

Most sources of electrical energy (mains electricity, a battery, ...) are best modeled as voltage sources. Such sources provide constant voltage, which means that as long as the amount of current drawn from the source is within the source's capabilities, its output voltage stays constant. An ideal voltage source provides no energy when it is loaded by an open circuit (i.e. an infinite impedance), but approaches infinite power and current when the load resistance approaches zero (a short circuit). Such a theoretical device would have a zero ohm output impedance in series with the source. A real-world voltage source has a very low, but non-zero output impedance: often much less than 1 ohm.

Conversely, a current source provides a constant current, as long as the load connected to the source terminals has sufficiently low impedance. An ideal current source would provide no energy to a short circuit and approach infinite energy and voltage as the load resistance approaches infinity (an open circuit). An *ideal* current source has an infinite output impedance in parallel with the source. A *real-world* current source has a very high, but finite output impedance. In the case of transistor current sources, impedances of a few megohms (at DC) are typical.

An *ideal* current source cannot be connected to an *ideal* open circuit because this would create the paradox of running a constant, non-zero current (from the current source) through an element with a defined zero current (the open circuit). Nor can an *ideal* voltage source be connected to an *ideal* short circuit (R=0), since this would result a similar paradox of finite non zero voltage across an element with defined zero voltage (the short circuit).

Because no ideal sources of either variety exist (all real-world examples have finite and non-zero source impedance), any current source can be considered as a voltage source with the *same* source impedance and vice versa. These concepts are dealt with by Norton's and Thévenin's theorems.