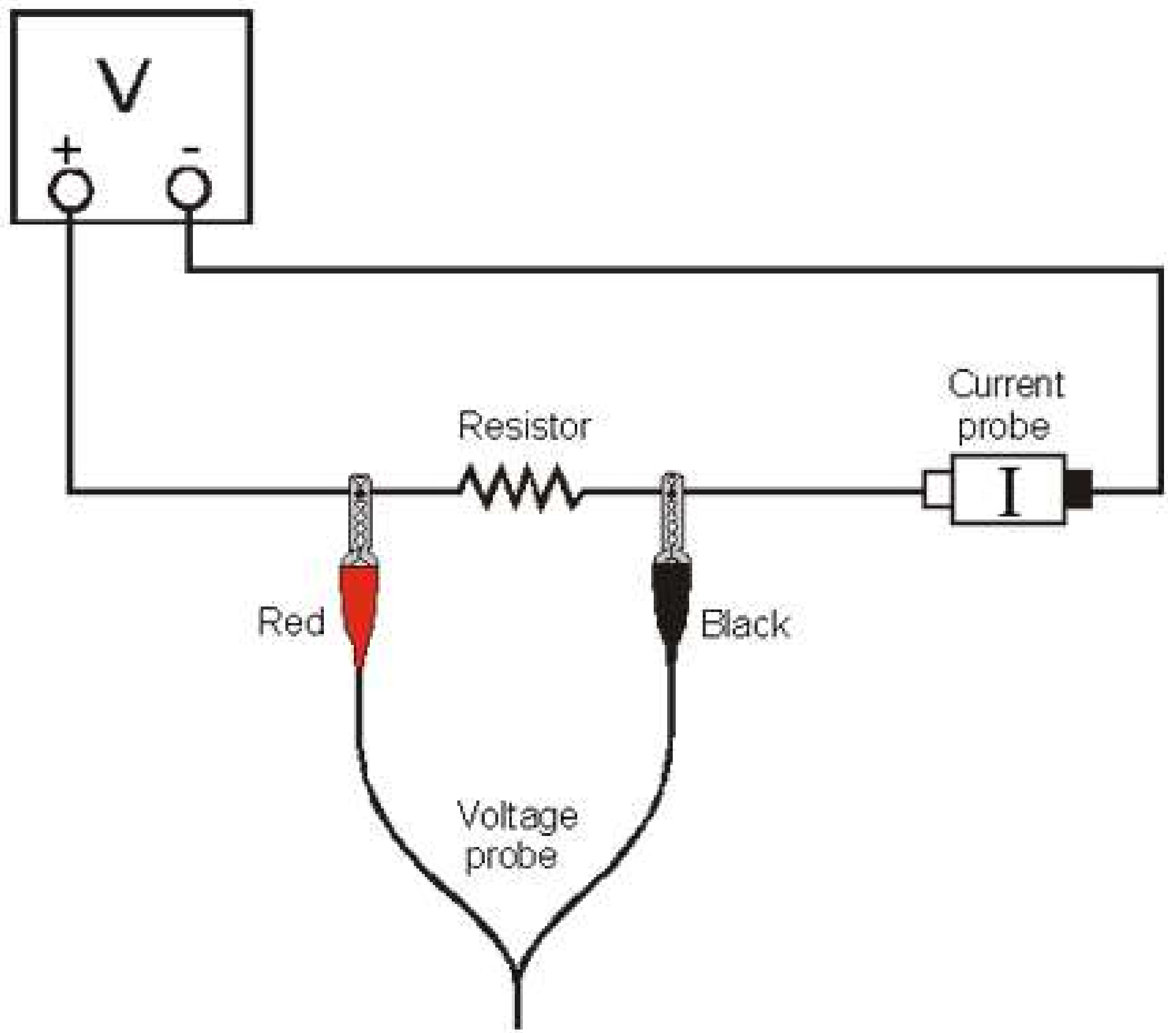
Laws and Theorems of Electronics



Lillie Conyers
Xuan Herrera

First Edition, 2012

ISBN 978-81-323-1355-7

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

Table of Contents

Chapter 1

Foster's Reactance Theorem

Foster's reactance theorem is an important theorem in the fields of electrical network analysis and synthesis. The theorem states that the reactance of a passive, lossless two-terminal (one-port) network always monotonically increases with frequency. The proof of the theorem was first presented by Ronald Martin Foster in 1924.

Explanation

Reactance is the imaginary part of the complex electrical impedance. The specification that the network must be passive and lossless implies that there are no resistors (lossless), or amplifiers or energy sources (passive) in the network. The network consequently must consist entirely of inductors and capacitors and the impedance will be purely an imaginary number with zero real part. Other than that, the theorem is quite general, in particular, it applies to distributed element circuits although Foster formulated it in terms of discrete inductors and capacitors. Foster's theorem applies equally to the admittance of a network, that is the susceptance (imaginary part of admittance) of a passive, lossless one-port monotonically increases with frequency. This result may seem counterintuitive since admittance is the reciprocal of impedance, but is easily proved. If an impedance,

$$Z = iX$$

where,
Z is impedance
X is reactance
i is the imaginary unit

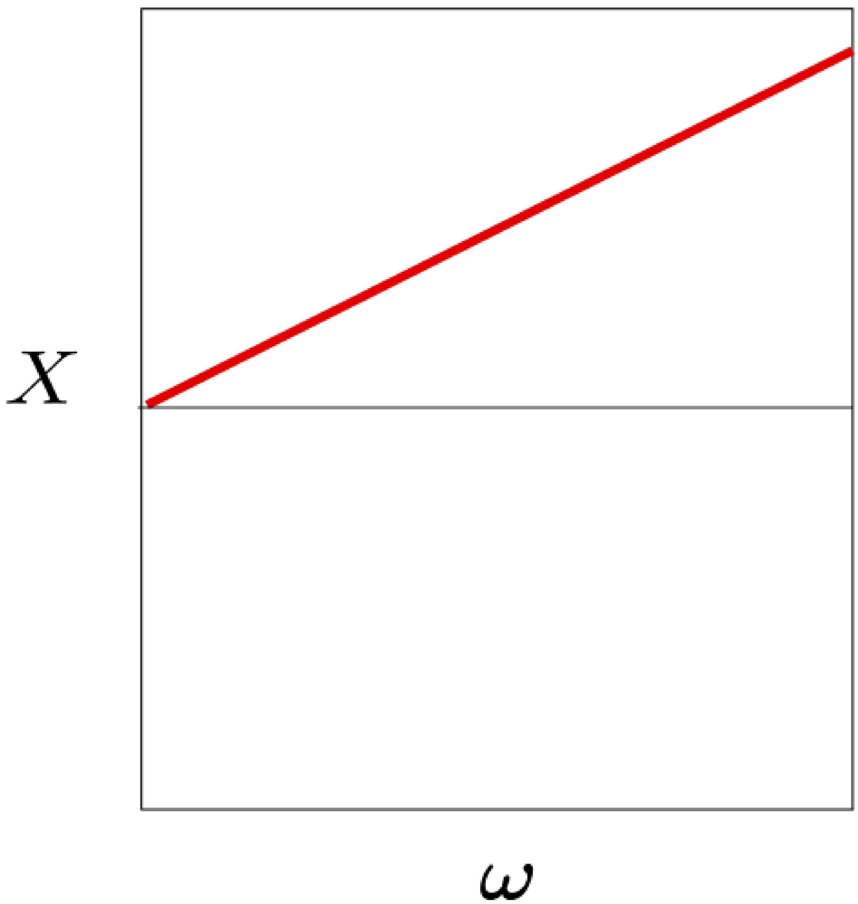
then the admittance is given by

$$Y = \frac{1}{iX} = -i\frac{1}{X} = iB$$

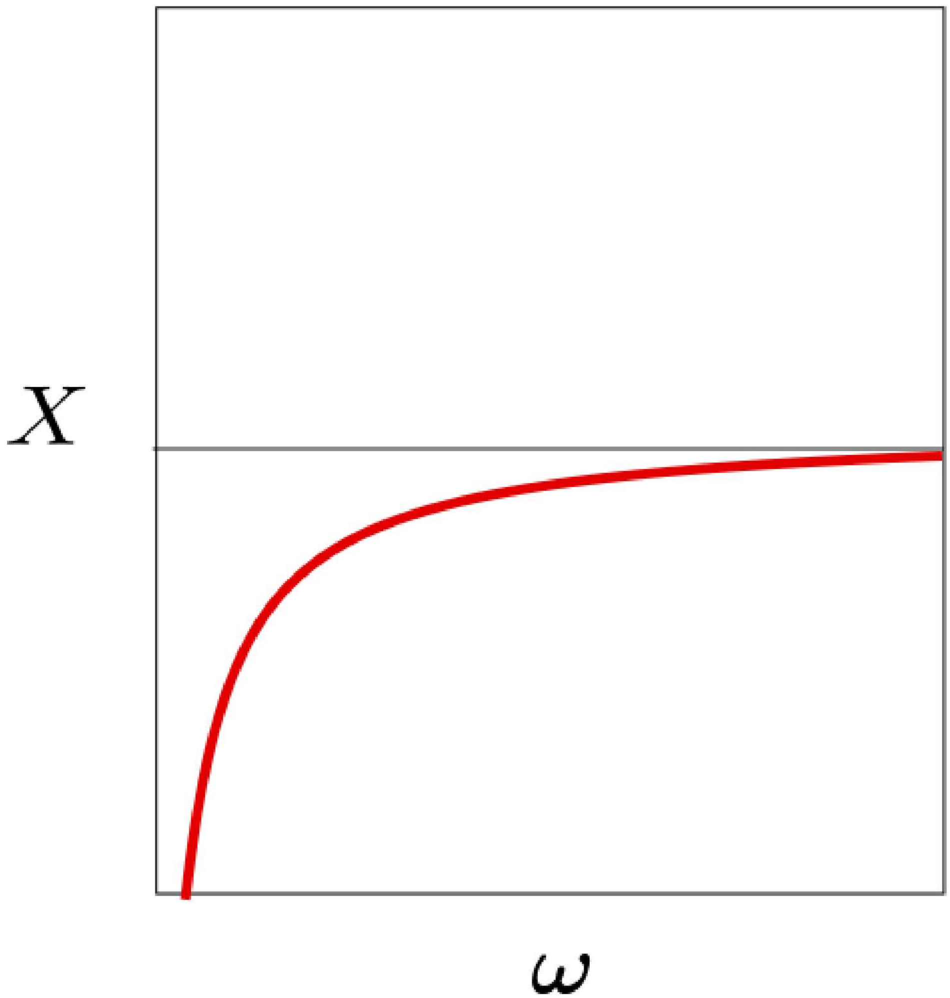
where,
Y is admittance
B is susceptance

If X is monotonically increasing with frequency then 1/X must be monotonically decreasing. −1/X must consequently be monotonically increasing and hence it is proved that B is increasing also. It is often the case in network theory that a principle or procedure apply equally to impedance or admittance as they do here. It is convenient in these circumstances to use the concept of immittance which can mean either impedance or admittance. The mathematics are carried out without stating which it is or specifying units until it is desired to calculate a specific example. Foster's theorem can thus be stated in a more general form as,

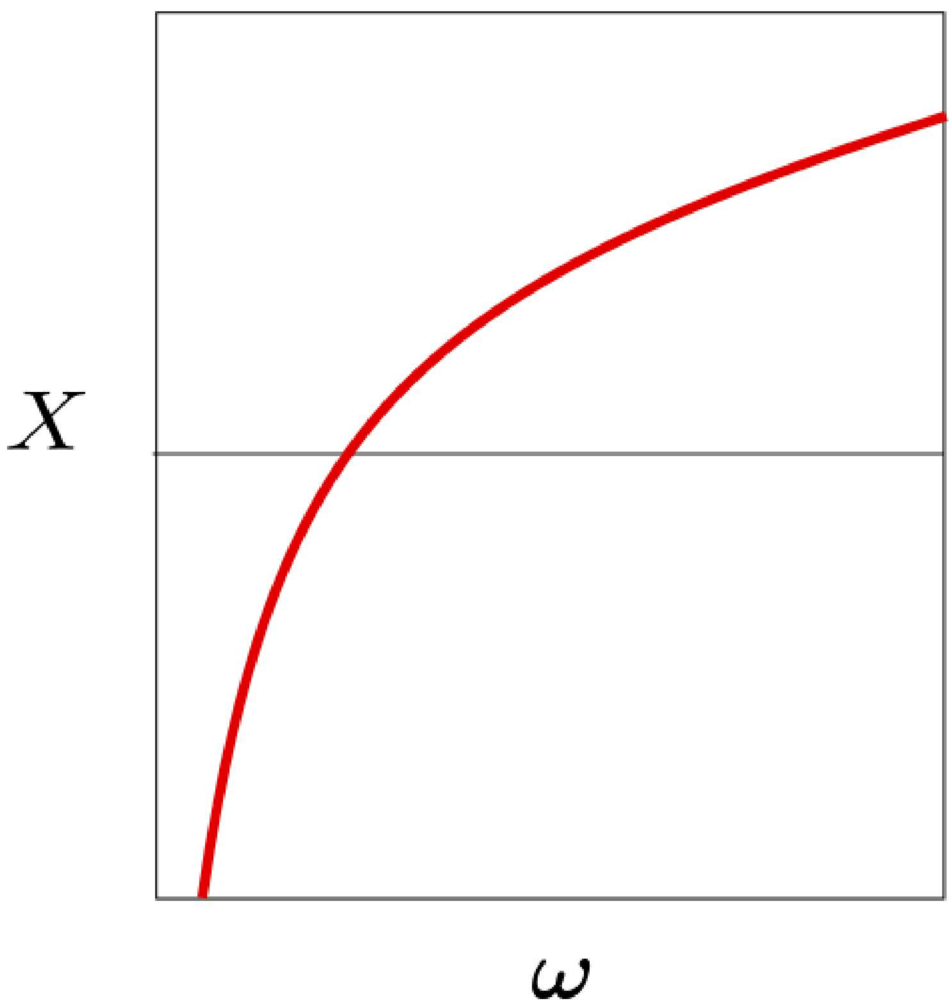
Examples



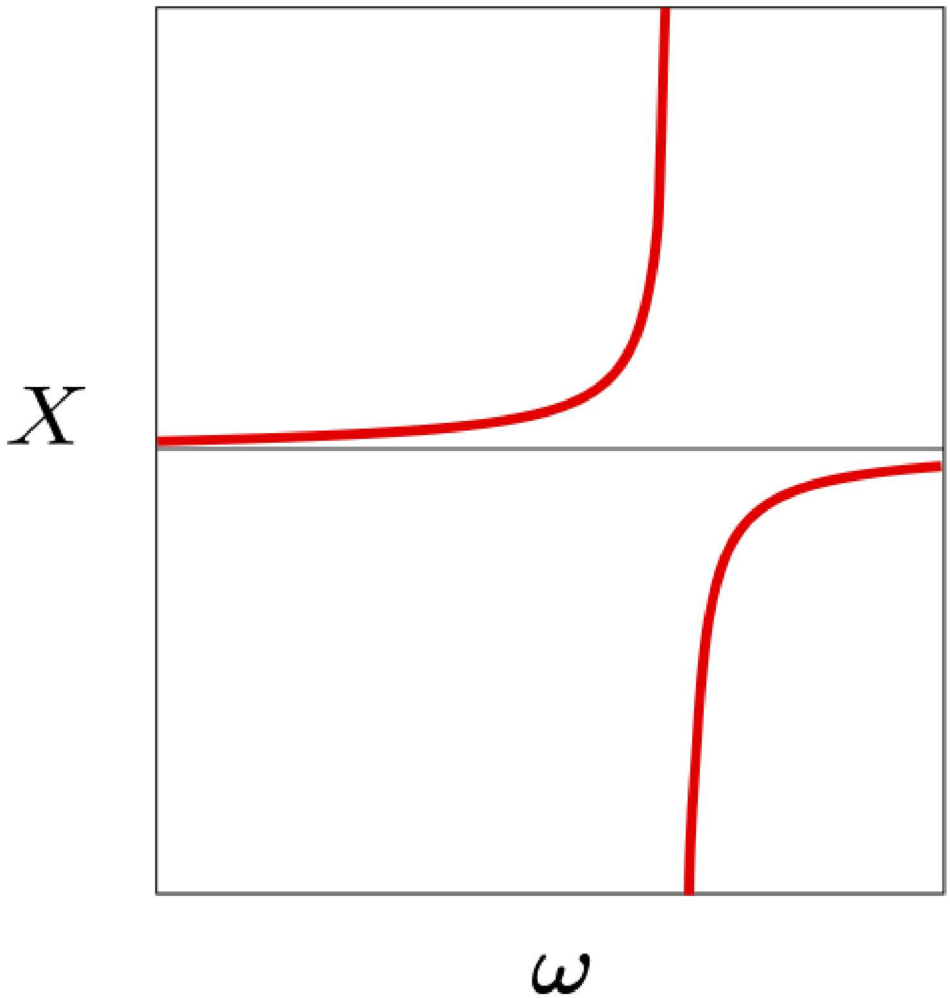
Plot of the reactance of an inductor against frequency



Plot of the reactance of a capacitor against frequency



Plot of the reactance of a series LC circuit against frequency



Plot of the reactance of a parallel LC circuit against frequency

The following examples illustrate this theorem in a number of simple circuits.

Inductor

The impedance of an inductor is given by,

$$Z = i\omega L$$

L is inductance

ω is angular frequency

so the reactance is,

$$X = \omega L$$

which by inspection can be seen to be monotonically (and linearly) increasing with frequency.

Capacitor

The impedance of a capacitor is given by,

$$Z = \frac{1}{i\omega C}$$

C is capacitance

so the reactance is,

$$X = -\frac{1}{\omega C}$$

which again is monotonically increasing with frequency. The impedance function of the capacitor is identical to the admittance function of the inductor and vice versa. It is a general result that the dual of any immittance function that obeys Foster's theorem will also follow Foster's theorem.

Series resonant circuit

A series LC circuit has an impedance that is the sum of the impedances of an inductor and capacitor,

$$Z = i\omega L + \frac{1}{i\omega C} = i\left(\omega L - \frac{1}{\omega C}\right)$$

At low frequencies the reactance is dominated by the capacitor and so is large and negative. This monotonically increases towards zero (the magnitude of the capacitor reactance is becoming smaller). The reactance passes through zero at the point where the magnitudes of the capacitor and inductor reactances are equal (the resonant frequency) and then continues to monotonically increase as the inductor reactance becomes progressively dominant.

Parallel resonant circuit

A parallel LC circuit is the dual of the series circuit and hence its admittance function is the same form as the impedance function of the series circuit,

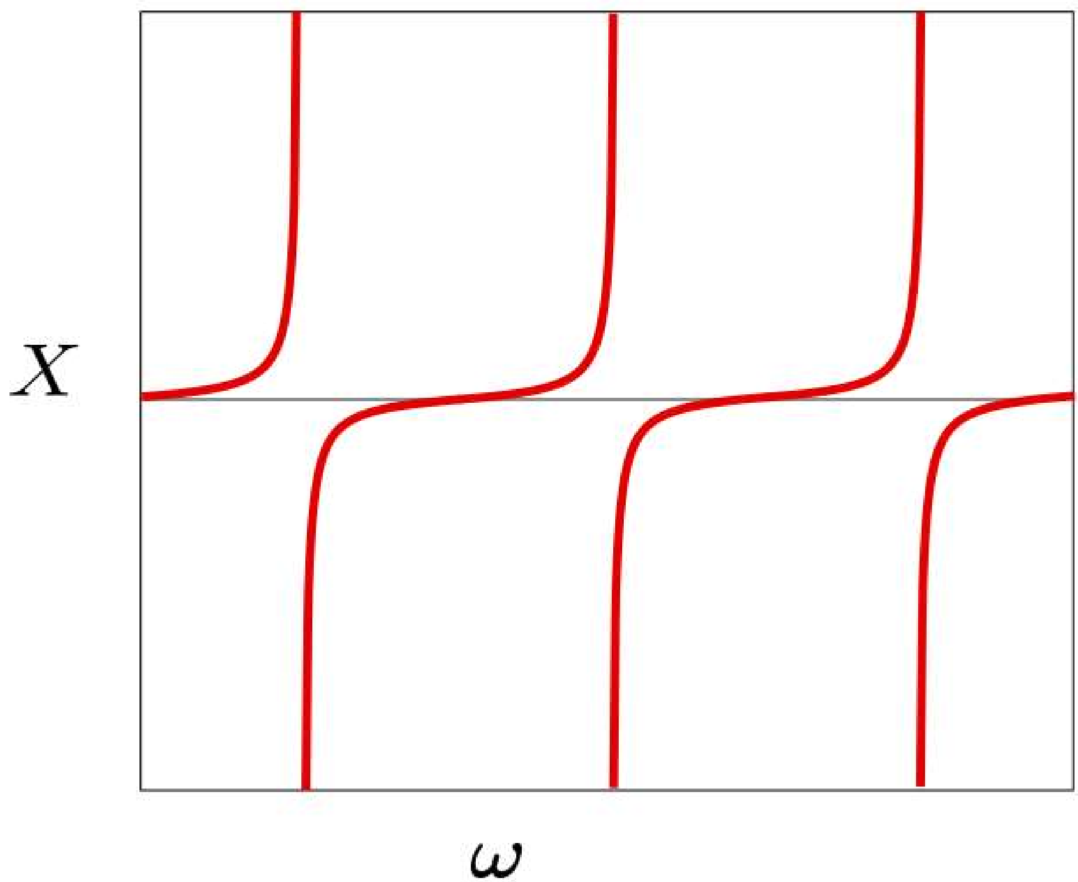
$$Y = i\omega C + \frac{1}{i\omega L}$$

The impedance function is,

$$Z = i\left(\frac{\omega L}{1-\omega^2 LC}\right)$$

At low frequencies the reactance is dominated by the inductor and is small and positive. This monotonically increases towards a pole at the anti-resonant frequency where the susceptance of the inductor and capacitor are equal and opposite and cancel. Past the pole the reactance is large and negative and increasing towards zero where it is dominated by the capacitance.

Poles and zeroes



Plot of the reactance of Foster's first form of canonical driving point impedance showing the pattern of alternating poles and zeroes. Three anti-resonators are required to realise this impedance function.

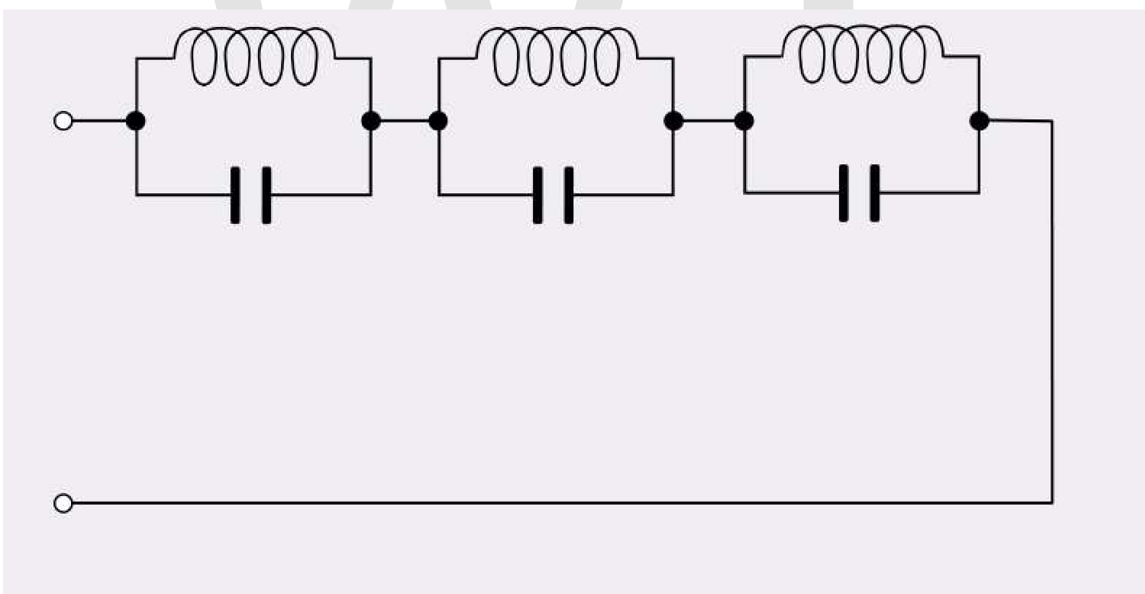
A consequence of Foster's theorem is that the poles and zeroes of any passive immittance function must alternate with increasing frequency. After passing through a pole the function will be negative and is obliged to pass through zero before reaching the next pole if it is to be monotonically increasing.

Not all networks obey Foster's theorem, those that do may be referred to as Foster networks and those that do not are non-Foster networks. A circuit containing an amplifier may well not be a Foster network. In particular, it is possible to simulate negative capacitors and inductors with negative impedance converter circuits. These circuits will have an immittance function of frequency with a negative slope.

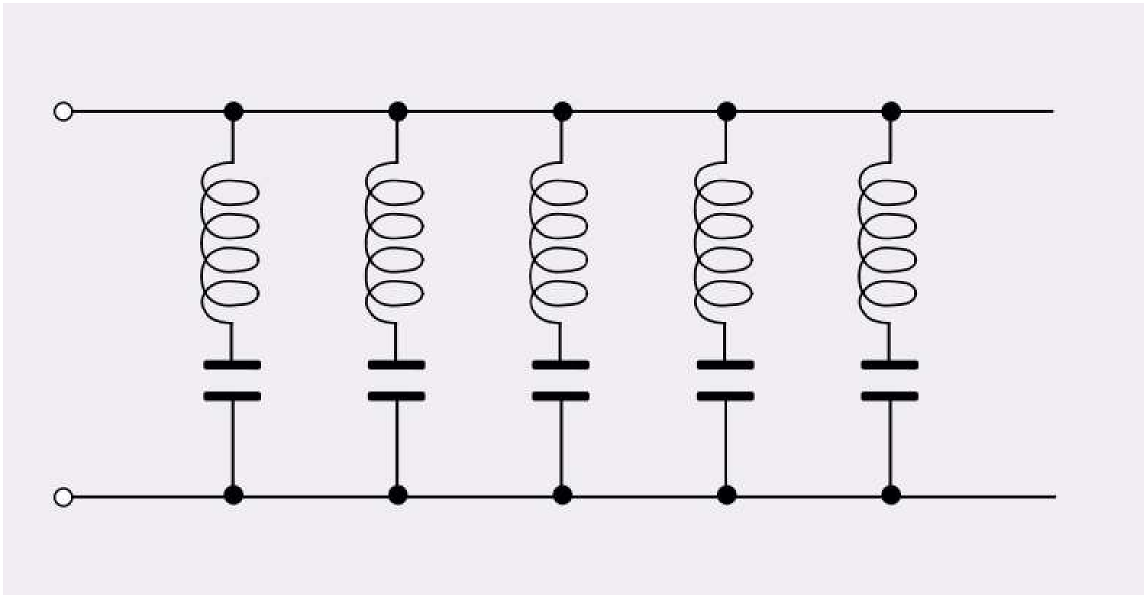
With the addition of a scaling factor, the poles and zeroes of an immittance function completely determine the frequency characteristics of a Foster network. Two Foster networks that have identical poles and zeroes will be equivalent circuits in the sense that their immittance functions will be identical.

Another consequence of Foster's theorem is that the plot of a Foster immittance function on a Smith chart must always travel around the chart in a clockwise direction with increasing frequency.

Realisation



Foster's first form of canonical driving point impedance realisation. If the polynomial function has a pole at $\omega=0$ one of the LC sections will reduce to a single capacitor. If the polynomial function has a pole at $\omega=\infty$ one of the LC sections will reduce to a single inductor. If both poles are present then two sections reduce to a series LC circuit.



Foster's second form of canonical driving point impedance realisation. If the polynomial function has a zero at ω=0 one of the LC sections will reduce to a single capacitor. If the polynomial function has a zero at ω=∞ one of the LC sections will reduce to a single inductor. If both zeroes are present then two sections reduce to a parallel LC circuit.

A one-port passive immittance consisting of discrete elements (that is, not a distributed element circuit) is described as rational in that in can be represented as a rational function of s,

$$Z(s) = \frac{P(s)}{Q(s)}$$

where,
$Z(s)$ is immittance
$P(s),\ Q(s)$ are polynomials with real, positive coefficiencts
s is the Laplace operator, which can be replaced with $i\omega$ when dealing with steady-state AC signals.

This is sometimes referred to as the driving point impedance because it is the impedance at the place in the network at which the external circuit is connected and "drives" it with a signal. Foster in his paper describes how such a lossless rational function may be realised in two ways. Foster's first form consists of a number of series connected parallel LC circuits. Foster's second form of driving point impedance consists of a number of parallel connected series LC circuits. The realisation of the driving point impedance is by no means unique. Foster's realisation has the advantage that the poles and/or zeroes are directly associated with a particular resonant circuit, but there are many other realisations. Perhaps the most well known is Cauer's ladder realisation from filter design.

History

The theorem was developed at American Telephone & Telegraph as part of ongoing investigations into improved filters for telephone multiplexing applications. This work was commercially important, large sums of money could be saved by increasing the number of telephone conversations that could be carried on one line. The theorem was first published by Campbell in 1922 but without a proof. Great use was immediately made of the theorem in filter design, it appears prominently, along with a proof, in Zobel's landmark paper of 1923 which summarised the state of the art of filter design at that time. Foster published his paper the following year which includes his canonical realisation forms.

Cauer in Germany grasped the importance of Foster's work and used it as the foundation of network synthesis. Amongst Cauer's many innovations was to extend Foster's work to all 2-element-kind networks after discovering an isomorphism between them. Cauer was interested in finding the conditions for realisability of a rational one-port network from its polynomial function and the reverse problem of which networks were equivalent, that is, had the same polynomial function. Both of these were important problems in network theory and filter design.

Chapter 2

Maximum Power Transfer Theorem

In electrical engineering, the **maximum power transfer theorem** states that, to obtain maximum external power from a source with a finite internal resistance, the resistance of the load must be equal to the resistance of the source as viewed from the output terminals. Moritz von Jacobi published the maximum power (transfer) theorem around 1840, which is also referred to as "Jacobi's law".

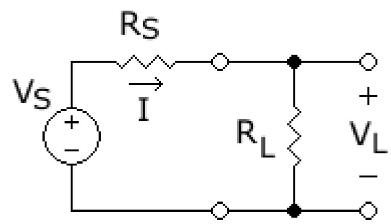
The theorem results in maximum power transfer, and not maximum efficiency. If the resistance of the load is made larger than the resistance of the source, then efficiency is higher, since a higher percentage of the source power is transferred to the load, but the magnitude of the load power is lower since the total circuit resistance goes up.

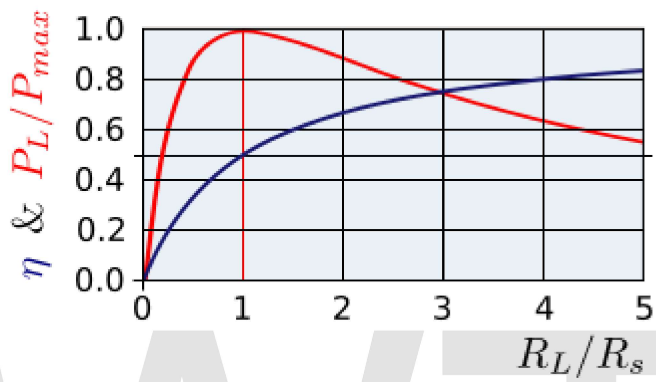
If the load resistance is smaller than the source resistance, then most of the power ends up being dissipated in the source, and although the total power dissipated is higher, due to a lower total resistance, it turns out that the amount dissipated in the load is reduced.

The theorem can be extended to AC circuits that include reactance, and states that maximum power transfer occurs when the load impedance is equal to the complex conjugate of the source impedance.

Maximizing power transfer versus power efficiency

The theorem was originally misunderstood (notably by Joule) to imply that a system consisting of an electric motor driven by a battery could not be more than 50% efficient since, when the impedances were matched, the power lost as heat in the battery would always be equal to the power delivered to the motor. In 1880 this assumption was shown to be false by either Edison or his colleague Francis Robbins Upton, who realized that maximum efficiency was not the same as maximum power transfer. To achieve maximum efficiency, the resistance of the source (whether a battery or a dynamo) could be made close to zero. Using this new understanding, they obtained an efficiency of about 90%, and proved that the electric motor was a practical alternative to the heat engine.





The condition of maximum power transfer does not result in maximum efficiency. If we define the efficiency η as the ratio of power dissipated by the load to power developed by the source, then it is straightforward to calculate from the above circuit diagram that

$$\eta = \frac{R_{\mathrm{load}}}{R_{\mathrm{load}} + R_{\mathrm{source}}} = \frac{1}{1 + \frac{R_{\mathrm{source}}}{R_{\mathrm{load}}}}$$

Consider three particular cases:

- If $R_{\mathrm{load}} = R_{\mathrm{source}}$, then $\eta = 0.5$.
- If $R_{\mathrm{load}} = \infty$ or $R_{\mathrm{source}} = 0$, then $\eta = 1$.
- If $R_{\mathrm{load}} = 0$, then $\eta = 0$.

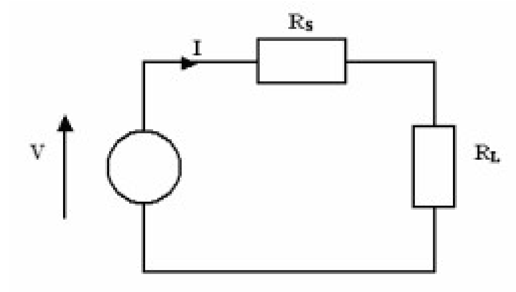
The efficiency is only 50% when maximum power transfer is achieved, but approaches 100% as the load resistance approaches infinity, though the total power level tends towards zero. Efficiency also approaches 100% if the source resistance can be made close to zero. When the load resistance is zero, all the power is consumed inside the source (the power dissipated in a short circuit is zero) so the efficiency is zero.

Note that for a time varying voltage and current, the above applies only to the resistive component of the power, not the reactive component. As shown below, a proper pure reactive match can theoretically achieve 100% power transfer AND 100% efficiency simultaneously in a perfect resistance free reactive circuit. In an imperfect reactive circuit (i.e. one with both reactance and resistance), the efficiency will be the power dissipated in the load divided by the power provided by the source. When the resistance is a small fraction of the reactance, the efficiency will then be very high even at the maximum power transfer point. This is a key reason why power grids use AC versus DC power, and why radio frequency (RF) circuits can operate with relatively high efficiency.

Impedance matching

A related concept is reflectionless impedance matching. In radio, transmission lines, and other electronics, there is often a requirement to match the source impedance (such as a transmitter) to the load impedance (such as an antenna) to avoid reflections in the transmission line.

Calculus-based proof for purely resistive circuits



In the diagram opposite, power is being transferred from the source, with voltage V and fixed source resistance R_S, to a load with resistance R_L, resulting in a current I. By Ohm's law, I is simply the source voltage divided by the total circuit resistance:

$$I = \frac{V}{R_\mathrm{S} + R_\mathrm{L}}$$

The power P_L dissipated in the load is the square of the current multiplied by the resistance:

$$P_\mathrm{L} = I^2 R_\mathrm{L} = \left(\frac{V}{R_\mathrm{S} + R_\mathrm{L}}\right)^2 R_\mathrm{L} = \frac{V^2}{R_\mathrm{S}^2/R_\mathrm{L} + 2R_\mathrm{S} + R_\mathrm{L}}$$

The value of R_L for which this expression is a maximum could be calculated by differentiating it, but it is easier to calculate the value of R_L for which the denominator

$$R_\mathrm{S}^2/R_\mathrm{L} + 2R_\mathrm{S} + R_\mathrm{L}$$

is a minimum. The result will be the same in either case. Differentiating the denominator with respect to R_L:

$$\frac{d}{dR_{\mathrm{L}}}\left(R_{\mathrm{S}}^2/R_{\mathrm{L}}+2R_{\mathrm{S}}+R_{\mathrm{L}}\right)=-R_{\mathrm{S}}^2/R_{\mathrm{L}}^2+1$$

For a maximum or minimum, the first derivative is zero, so

$$R_{\mathrm{S}}^2/R_{\mathrm{L}}^2=1$$

or

$$R_{\mathrm{L}}=\pm R_{\mathrm{S}}$$

In practical resistive circuits, $R_{\mathrm{S}}$and $R_{\mathrm{L}}$are both positive, so the positive sign in the above is the correct solution. To find out whether this solution is a minimum or a maximum, the denominator expression is differentiated again:

$$\frac{d^2}{dR_{\mathrm{L}}^2}\left(R_{\mathrm{S}}^2/R_{\mathrm{L}}+2R_{\mathrm{S}}+R_{\mathrm{L}}\right)=2R_{\mathrm{S}}^2/R_{\mathrm{L}}^3$$

This is always positive for positive values of R_{S} and R_{L}, showing that the denominator is a minimum, and the power is therefore a maximum, when

$$R_{\mathrm{S}}=R_{\mathrm{L}}$$

A note of caution is in order here. This last statement, as written, implies to many people that for a given load, the source resistance must be set equal to the load resistance for maximum power transfer. However, this equation only applies if the source resistance cannot be adjusted, e.g., with antennas. For any given load resistance a source resistance of zero is the way to transfer maximum power to the load. As an example, a 100 volt source with an internal resistance of 10 ohms connected to a 10 ohm load will deliver 250 watts to that load. Make the source resistance zero ohms and the load power jumps to 1000 watts.

In reactive circuits

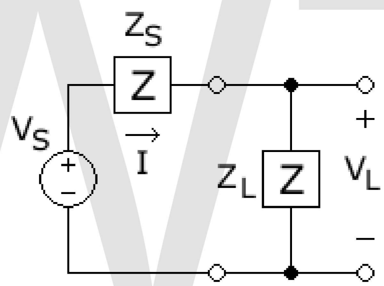
The theorem also applies where the source and/or load are not totally resistive. This invokes a refinement of the maximum power theorem, which says that any reactive components of source and load should be of equal magnitude but opposite phase. This means that the source and load impedances should be complex conjugates of each other. In the case of purely resistive circuits, the two concepts are identical. However, physically realizable sources and loads are not usually totally resistive, having some inductive or capacitive components, and so practical applications of this theorem, under the name of complex conjugate impedance matching, do, in fact, exist.

If the source is totally inductive (capacitive), then a totally capacitive (inductive) load, in the absence of resistive losses, would receive 100% of the energy from the source but send it back after a quarter cycle. The resultant circuit is nothing other than a resonant LC circuit in which the energy continues to oscillate to and fro. This is called reactive power. Power factor correction (where an inductive reactance is used to "balance out" a capacitive one), is essentially the same idea as complex conjugate impedance matching although it is done for entirely different reasons.

For a fixed reactive source, the maximum power theorem maximizes the real power (P) delivered to the load by complex conjugate matching the load to the source.

For a fixed reactive load, power factor correction minimizes the apparent power (S) (and unnecessary current) conducted by the transmission lines, while maintaining the same amount of real power transfer. This is done by adding a reactance to the load to balance out the load's own reactance, changing the reactive load impedance into a resistive load impedance.

Proof



In this diagram, AC power is being transferred from the source, with phasor magnitude voltage $|V_S|$ (peak voltage) and fixed source impedance Z_S, to a load with impedance Z_L, resulting in a phasor magnitude current $|I|$. $|I|$ is simply the source voltage divided by the total circuit impedance:

$$|I| = \frac{|V_\mathrm{S}|}{|Z_\mathrm{S} + Z_\mathrm{L}|}.$$

The average power P_L dissipated in the load is the square of the current multiplied by the resistive portion (the real part) R_L of the load impedance:

$$\begin{aligned} P_\mathrm{L} &= I_\mathrm{rms}^2 R_\mathrm{L} = \frac{1}{2}|I|^2 R_\mathrm{L} = \frac{1}{2}\left(\frac{|V_\mathrm{S}|}{|Z_\mathrm{S} + Z_\mathrm{L}|}\right)^2 R_\mathrm{L} \\ &= \frac{1}{2}\frac{|V_\mathrm{S}|^2 R_\mathrm{L}}{(R_\mathrm{S} + R_\mathrm{L})^2 + (X_\mathrm{S} + X_\mathrm{L})^2}, \end{aligned}$$

where the resistance R_S and reactance X_S are the real and imaginary parts of Z_S, and X_L is the imaginary part of Z_L.

To determine the values of R_L and X_L (since V_S, R_S, and X_S are fixed) for which this expression is a maximum, we first find, for each fixed positive value of R_L, the value of the reactive term X_L for which the denominator

$$(R_\mathrm{S} + R_\mathrm{L})^2 + (X_\mathrm{S} + X_\mathrm{L})^2$$

is a minimum. Since reactances can be negative, this denominator is easily minimized by making

$$X_\mathrm{L} = -X_\mathrm{S}.$$

The power equation is now reduced to:

$$P_\mathrm{L} = \frac{1}{2} \frac{|V_\mathrm{S}|^2 R_\mathrm{L}}{(R_\mathrm{S} + R_\mathrm{L})^2}$$

and it remains to find the value of R_L which maximizes this expression. However, this maximization problem has exactly the same form as in the purely resistive case, and the maximizing condition $R_L = R_S$ can be found in the same way.

The combination of conditions

- $R_\mathrm{L} = R_\mathrm{S}$
- $X_\mathrm{L} = -X_\mathrm{S}$

can be concisely written with a complex conjugate (the *) as:

$$Z_\mathrm{S} = Z_\mathrm{L}^*.$$

Chapter 3

Miller Theorem

Miller theorem refers to the process of creating equivalent circuits. The general circuit theorem asserts that a floating impedance element supplied by two connected in series voltage sources may be split into two grounded elements with corresponding impedances. There is also a dual Miller theorem with regards to impedance supplied by two connected in parallel current sources. The two versions are based on the two Kirchhoff's circuit laws.

Miller theorems are not only pure mathematical expressions. These arrangements explain important circuit phenomena about modifying impedance (Miller effect, virtual ground, bootstrapping, negative impedance, etc.) and help designing and understanding various popular circuits (feedback amplifiers, resistive and time-dependent converters, negative impedance converters, etc.) They are useful in the area of circuit analysis especially for analyzing circuits with feedback and certain transistor amplifiers at high frequencies.

There is a close relationship between Miller theorem and Miller effect: the theorem may be considered as a generalization of the effect and the effect may be thought as a special case of the theorem.

Miller theorem (for voltages)

Definition

Miller theorem establishes that in a linear circuit, if there exists a branch with impedance Z, connecting two nodes with nodal voltages V_1 and V_2, we can replace this branch by two branches connecting the corresponding nodes to ground by impedances respectively $Z/(1 - K)$ and $KZ/(K - 1)$, where $K = V_2/V_1$. Miller theorem may be proved by using the equivalent two-port network technique to replace the two-port to its equivalent and by applying the source absorption theorem. This version of Miller theorem is based on Kirchhoff's voltage law; for that reason, it is named also Miller theorem for voltages.

Explanation

Miller theorem implies that an impedance element is supplied by two arbitrary (not necessarily dependent) voltage sources that are connected in series through the common ground. In practice, one of them acts as a main (independent) voltage source with voltage

V_1 and the other - as an additional (linearly dependent) voltage source with voltage $V_2 = KV_1$. The idea of Miller theorem (modifying circuit impedances seen from the sides of the input and output sources) is revealed below by comparing the two situations - without and with connected an additional voltage source V_2.

If V_2 was zero (there was not a second voltage source or the right end of the element with impedance Z was just grounded), the input current flowing through the element would be determined, according to Ohm's law, only by V_1

$$I_{in0} = \frac{V_1}{Z}$$

and the input impedance of the circuit would be

$$Z_{in0} = \frac{V_1}{I_{in0}} = Z.$$

As a second voltage source is included, the input current depends on both the voltages. According to its polarity, V_2 is subtracted or added from/to V_1; so, the input current decreases/increases

$$I_{in} = \frac{V_1 - V_2}{Z} = \frac{(1-K)}{Z}V_1 = (1-K)I_{in0}$$

and the input impedance of the circuit seen from the side of the input source accordingly increases/decreases

$$Z_{in} = \frac{V_1}{I_{in}} = \frac{Z}{1-K}.$$

So, Miller theorem expresses the fact that **connecting a second voltage source with proportional voltage $V_2 = KV_1$ in series with the input voltage source changes the effective voltage, the current and respectively, the circuit impedance seen from the side of the input source**. Depending on the polarity, V_2 acts as a supplemental voltage source helping or opposing the main voltage source to pass the current through the impedance.

Besides by presenting the combination of the two voltage sources as a new composed voltage source, the theorem may be explained by combining the actual element and the second voltage source into a new virtual element with dynamically modified impedance. From this viewpoint, V_2 is an additional voltage that artificially increases/decreases the voltage drop V_z across the impedance Z thus decreasing/increasing the current. The proportion between the voltages determines the value of the obtained impedance (see the tables below) and gives in total six groups of typical applications.

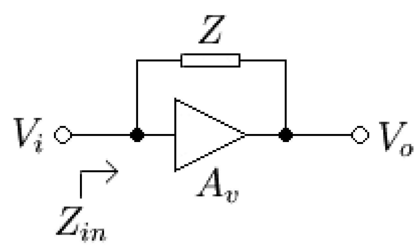
Subtracting V_2 from V_1				
V_2 vs V_1	**$V_2 = 0$**	**$0 < V_2 < V_1$**	**$V_2 = V_1$**	**$V_2 > V_1$**
Impedance	normal	increased	infinite	negative with current inversion

Adding V_2 to V_1				
V_2 vs V_z	**$V_2 = 0$**	**$0 < V_2 < V_z$**	**$V_2 = V_z$**	**$V_2 > V_z$**
Impedance	normal	decreased	zero	negative with voltage inversion

The circuit impedance seen from the side of the output source may be defined by the same reasonings if the voltages V_1 and V_2 are swapped and the coefficient K is replaced by 1/K

$$Z_{in2} = \frac{KZ}{K-1}.$$

Implementation



A typical implementation of Miller theorem based on a single-ended voltage amplifier

Most frequently, Miller theorem may be observed in and implemented by an arrangement consisting of an element with impedance Z connected between the two terminals of a grounded general linear network. Usually, a voltage amplifier with gain of $A_V = K$ serves as such a linear network but also other devices can play this role: a man and a potentiometer in a potentiometric null-balance meter, an electromechanical integrator (servomechanisms using potentiometric feedback sensors), etc. In the amplifier implementation, the input voltage V_i serves as V_1 and the output voltage V_o - as V_2. In many cases, the input voltage source has some internal impedance Z_{int} or an additional input impedance is connected that, in combination with Z, introduces a feedback. Depending on the kind of the amplifier (non-inverting, inverting or differential), the feedback can be positive, negative or mixed.

The Miller amplifier arrangement has two aspects:

- the amplifier may be thought as an additional voltage source converting the actual impedance into a virtual impedance (the amplifier modifies the impedance of the actual element)

- the virtual impedance may be thought as an element connected in parallel to the amplifier input (the virtual impedance modifies the amplifier input impedance).

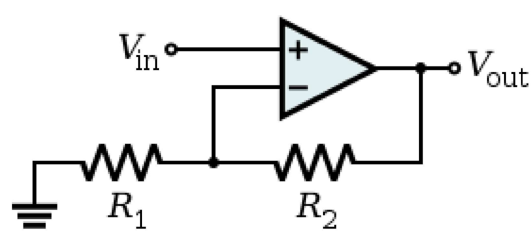
Applications

The introduction of an impedance that connects amplifier input and output ports adds a great deal of complexity in the analysis process. Miller theorem helps reducing the complexity in some circuits particularly with feedback by converting them to simpler equivalent circuits. But Miller theorem is not only an effective tool for creating equivalent circuits; it is also a powerful tool for designing and understanding circuits based on modifying impedance by additional voltage. Depending on the polarity of the output voltage versus the input voltage and the proportion between their magnitudes, there are six groups of typical situations. In some of them, Miller phenomenon appears as desired (bootstrapping) or undesired (Miller effect) unintentional effects; in other cases it is intentionally introduced.

Applications based on subtracting V_2 from V_1

In these applications, the output voltage V_o is inserted with an opposite polarity in respect to the input voltage V_i travelling along the loop (but in respect to ground, the polarities are the same). As a result, the effective voltage across and the current through the impedance decrease; the input impedance increases.

Increased impedance is implemented by a non-inverting amplifier with gain of $0 < A_v < 1$. The (magnitude of) output voltage is less than the input voltage V_i and partially neutralizes it. Examples are imperfect voltage followers (emitter, source, cathode follower, etc.) and amplifiers with series negative feedback (emitter degeneration), whose input impedance is moderately increased.



The op-amp non-inverting amplifier is a typical circuit with series negative feedback based on Miller theorem where the op-amp differential input impedance is apparently increased up to infinite

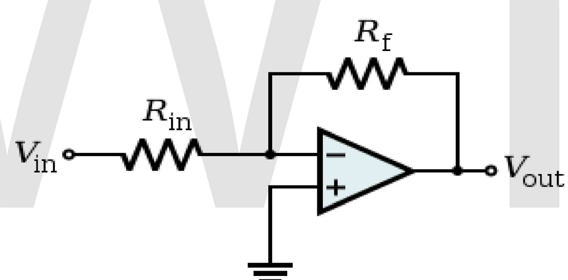
Infinite impedance uses a non-inverting amplifier with $A_v = 1$. The output voltage is equal to the input voltage V_i and completely neutralizes it. Examples are potentiometric null-balance meters and op-amp followers and amplifiers with series negative feedback (op-amp follower and non-inverting amplifier) where the circuit input impedance is enormously increased. This technique is referred to as bootstrapping and is intentionally used in biasing circuits, input guarding circuits, etc.

Negative impedance obtained by current inversion is implemented by a non-inverting amplifier with $A_v > 1$. The current changes its direction as the output voltage is higher than the input voltage. If the input voltage source has some internal impedance Z_{int} or if it is connected through another impedance element, a positive feedback appears. A typical application is the negative impedance converter with current inversion (INIC) that uses both negative and positive feedback (the negative feedback is used to realize a non-inverting amplifier and the positive feedback - to modify the impedance).

Applications based on adding V_2 to V_1

In these applications, the output voltage V_o is inserted with the same polarity in respect to the input voltage V_i travelling along the loop (but in respect to ground, the polarities are opposite). As a result, the effective voltage across and the current through the impedance increase; the input impedance decreases.

Decreased impedance is implemented by an inverting amplifier having some moderate gain, usually $10 < A_v < 1000$. It may be observed as an undesired Miller effect in common-emitter, common-source and common-cathode amplifying stages where effective input capacitance is increased. Frequency compensation for general purpose operational amplifiers and transistor Miller integrator are examples of useful usage of Miller effect.



The op-amp inverting amplifier is a typical circuit with parallel negative feedback based on Miller theorem where the op-amp differential input impedance is apparently decreased up to zero

Zeroed impedance uses an inverting (usually op-amp) amplifier with enormously high gain $A_v \to \infty$. The output voltage is almost equal to the voltage drop V_Z across the impedance and completely neutralizes it. The circuit behaves as a short connection and a virtual ground appears at the input; so, it should not be driven by a constant voltage source. For this purpose, some circuits are driven by a constant current source or by a real voltage source with internal impedance: current-to-voltage converter (transimpedance amplifier), capacitive integrator (named also current integrator or charge amplifier), resistance-to-voltage converter (a resistive sensor connected in the place of the impedance Z). The rest of them have additional impedance connected in series to the input: voltage-to-current converter (transconductance amplifier), inverting amplifier, summing amplifier, inductive integrator, capacitive differentiator, resistive-capacitive integrator, capacitive-resistive differentiator, inductive-resistive differentiator, etc. The inverting integrators from this list are examples of useful and desired applications of Miller effect in its extreme manifestation.

In all these op-amp inverting circuits with parallel negative feedback, the input current is increased to its maximum. It is determined only by the input voltage and the input impedance according to Ohm's law; it does not depend on the impedance Z.

Negative impedance with voltage inversion is implemented by applying both negative and positive feedback to an op-amp amplifier with a differential input. The input voltage source has to have internal impedance $Z_{int} > 0$ or it has to be connected through another impedance element to the input. Under these conditions, the input voltage V_i of the circuit changes its polarity as the output voltage exceeds the voltage drop V_Z across the impedance ($V_i = V_z - V_o < 0$). A typical application is a negative impedance converter with voltage inversion (VNIC). It is interesting that the circuit input voltage has the same polarity as the output voltage although it is applied to the inverting op-amp input; the input source has an opposite polarity to both the circuit input and output voltages.

Generalization of Miller arrangement

The original Miller effect is implemented by capacitive impedance connected between the two nodes. Miller theorem generalizes Miller effect as it implies arbitrary impedance Z connected between the nodes. It is supposed also a constant koefficient K; then the expressions above are valid. But modifying properties of Miller theorem exist even when these requirements are violated and this arrangement can be generalized further by dynamizing the impedance and the coefficient.

Non-linear element. Besides impedance, Miller arrangement can modify the IV characteristic of an arbitrary element. The circuit of a diode log converter is an example of a non-linear virtually zeroed resistance where the logaritmic forward IV curve of a diode is transformed to a vertical straight line overlapping the Y axis.

Not constant coefficient. If the coefficient K varies, some exotic virtual elements can be obtained. A gyrator circuit is an example of such a virtual element where the resistance R_L is modified so that to mimic inductance, capacitance or inversed resistance.

Dual Miller theorem (for currents)

Definition

There is also a dual version of Miller theorem that is based on Kirchhoff's current law (Miller theorem for currents): if there is a branch in a circuit with impedance Z connecting a node, where two currents I_1 and I_2 converge to ground, we can replace this branch by two conducting the referred currents, with impedances respectively equal to $(1 + \alpha)Z$ and $(1 + \alpha)Z/\alpha$, where $\alpha = I_2/I_1$. The dual theorem may be proved by replacing the two-port network by its equivalent and by applying the source absorption theorem.

Explanation

Dual Miller theorem actually expresses the fact that **connecting a second current source producing proportional current $I_2 = KI_1$ in parallel with the main input source and the impedance element changes the current flowing through it, the voltage and accordingly, the circuit impedance seen from the side of the input source**. Depending on the direction, I_2 acts as a supplemental current source helping or opposing the main current source I_1 to create voltage across the impedance. The combination of the actual element and the second current source may be thought as of a new virtual element with dynamically modified impedance.

Implementation

Dual Miller theorem is usually implemented by an arrangement consisting of two voltage sources supplying the grounded impedance Z through floating impedances. The combinations of the voltage sources and belonging impedances form the two current sources - the main and the auxiliary one. As in the case of the main Miller theorem, the second voltage is usually produced by a voltage amplifier. Depending on the kind of the amplifier (inverting, non-inverting or differential) and the gain, the circuit input impedance may be virtually increased, infinite, decreased, zero or negative.

Applications

As the main Miller theorem, besides helping circuit analysis process, the dual version is a powerful tool for designing and understanding circuits based on modifying impedance by additional current. Typical applications are some exotic circuits with negative impedance as load cancellers, capacitance neutralizers, Howland current source and its derivative Deboo integrator. In the last example, the Howland current source consists of an input voltage source V_{IN}, a positive resistor R, a load (the capacitor C acting as impedance Z) and a negative impedance converter INIC ($R_1 = R_2 = R_3 = R$ and the op-amp). The input voltage source and the resistor R constitute an imperfect current source passing current I_R through the load. The INIC acts as a second current source passing "helping" current I_{-R} through the load. As a result, the total current flowing through the load is constant and the circuit impedance seen by the input source is increased. As a comparison, in a load canceller, the INIC passes all the required current through the load; the circuit impedance seen from the side of the input source (the load impedance) is almost infinite.

Chapter 4

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.

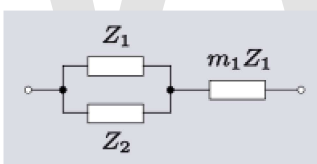
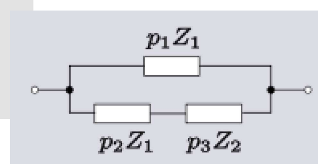
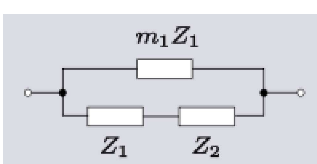
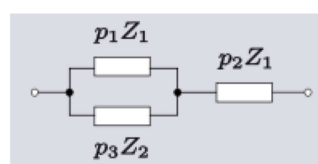
The vast scale of the topic of equivalent circuits is underscored in a story told by Sidney Darlington. According to Darlington, a large number of equivalent circuits were found by Ronald Foster, following his and George Campbell's 1920 paper on non-dissipative four-ports. In the course of this work 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.

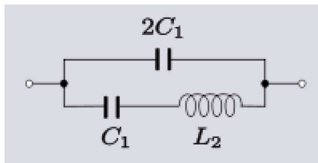
3-element networks

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.		$p_1 = \frac{m_1{}^2}{1 + m_1}$, $p_2 = \frac{m_1}{1 + m_1}$,	

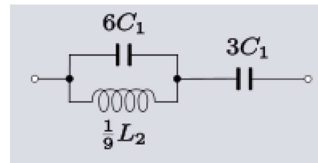
$$p_3 = \left(\frac{m_1}{1+m_1}\right)^2$$

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



$$m_1 = 0.5 \ ,$$

$$p_1 = \tfrac{1}{6} \ , \ p_2 = \tfrac{1}{3} \ ,$$



$$p_3 = \tfrac{1}{9} \ .$$

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:

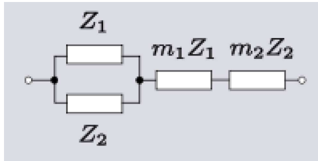
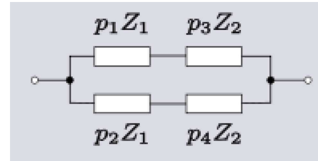
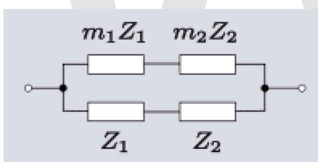
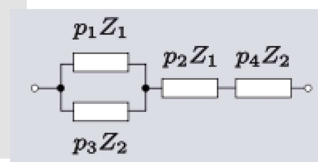
$$q_1 := 1 + m_1 + m_2,$$

$$q_2 := \sqrt{{q_1}^2 - 4m_1m_2},$$

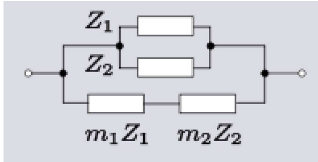
$$q_3 := \frac{(1+m_1)(1+m_2)}{(m_1-m_2)^2},$$

$$q_4 := \frac{q_2 - q_1 + 2m_2}{2q_2},$$

$$q_5 := \frac{q_2 + q_1 - 2m_2}{2q_2}.$$

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.		$p_1 = \dfrac{q_1 + q_2}{2q_5}$, $p_2 = \dfrac{q_1 - q_2}{2q_4}$, $p_3 = \dfrac{m_2}{q_5}$, $p_4 = \dfrac{m_2}{q_4}$.	
Transform 2.2 Transform 2.1 is the reverse of this transform. Transform 2.4 is the topological dual of this transform.		$p_1 = \dfrac{1}{q_3(1 + m_2)}$, $p_2 = \dfrac{m_1}{1 + m_1}$, $p_3 = \dfrac{1}{q_3(1 + m_1)}$, $p_4 = \dfrac{m_2}{1 + m_2}$.	

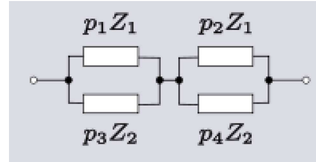
Transform 2.3 Transform 2.4 is the reverse of this transform. Transform 2.1 is the topological dual of this transform.



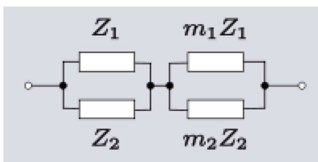
$$p_1 = \frac{q_4(q_1+q_2)}{2m_2} \ ,$$

$$p_2 = \frac{q_5(q_1-q_2)}{2m_2} \ ,$$

$$p_3 = q_4 \ , \ p_4 = q_5 \ .$$



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

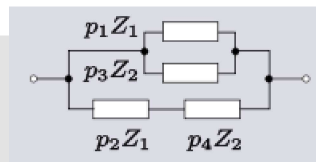


$$p_1 = 1 + m_1 \ ,$$

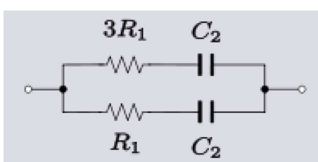
$$p_2 = m_1q_3(1+m_1) \ ,$$

$$p_3 = 1 + m_2 \ ,$$

$$p_4 = m_1q_3(1+m_2) \ .$$



Example 2. An example of Transform 2.2.

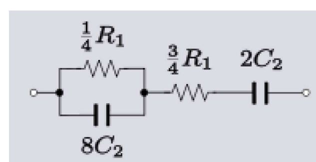


$$m_1 = 3 \ , \ m_2 = 1 \ ,$$

$$q_3 = 2 \ , \ p_1 = \tfrac{1}{4} \ ,$$

$$p_2 = \tfrac{3}{4} \ , \ p_3 = \tfrac{1}{8} \ ,$$

$$p_4 = \tfrac{1}{2} \ .$$



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

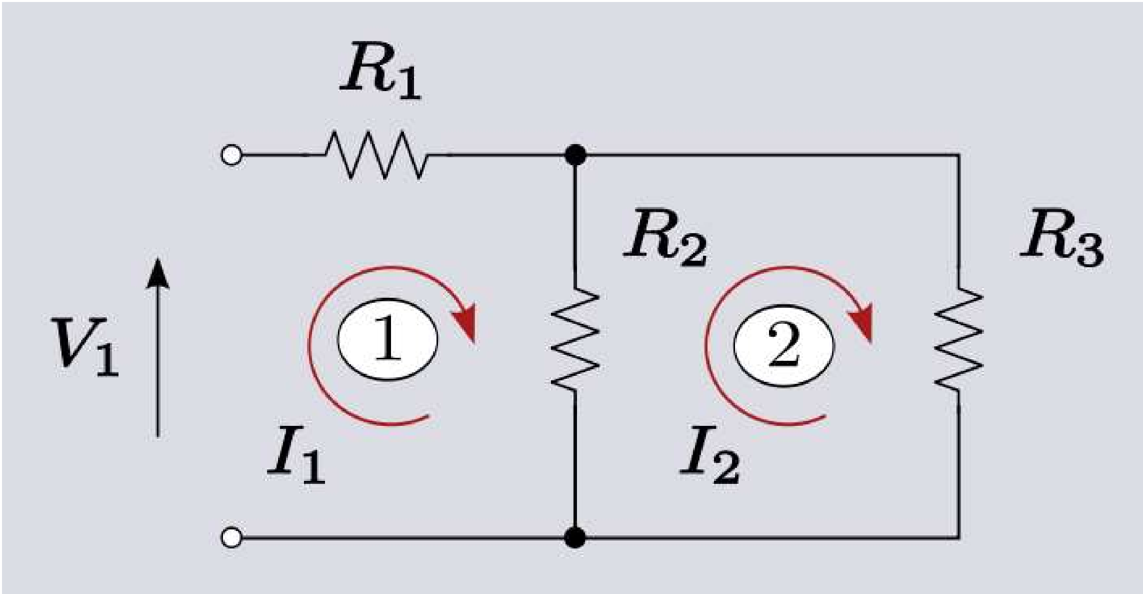


Fig. 1. Simple example of a network of impedances using resistors only for clarity. However, analysis of networks with other impedance elements proceed by the same principles. Two meshes are shown, with numbers in circles. The sum of impedances around each mesh, p, will form the diagonal of the entries of the matrix, Z_{pp}. The impedance of branches shared by two meshes, p and q, will form the entries $-Z_{pq}$. Z_{pq}, $p \neq q$, 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 nxn 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 of here 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 nxn 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;

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

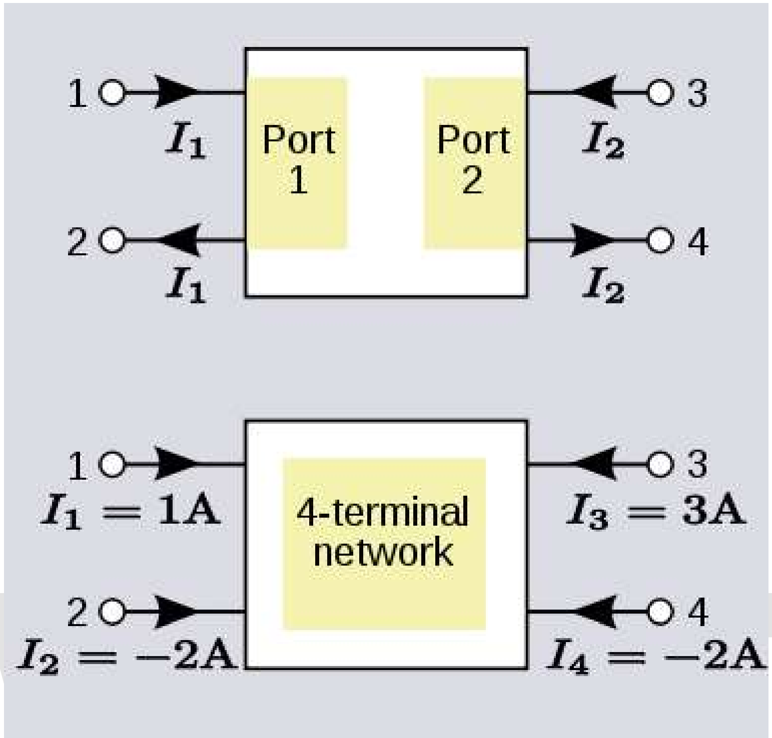
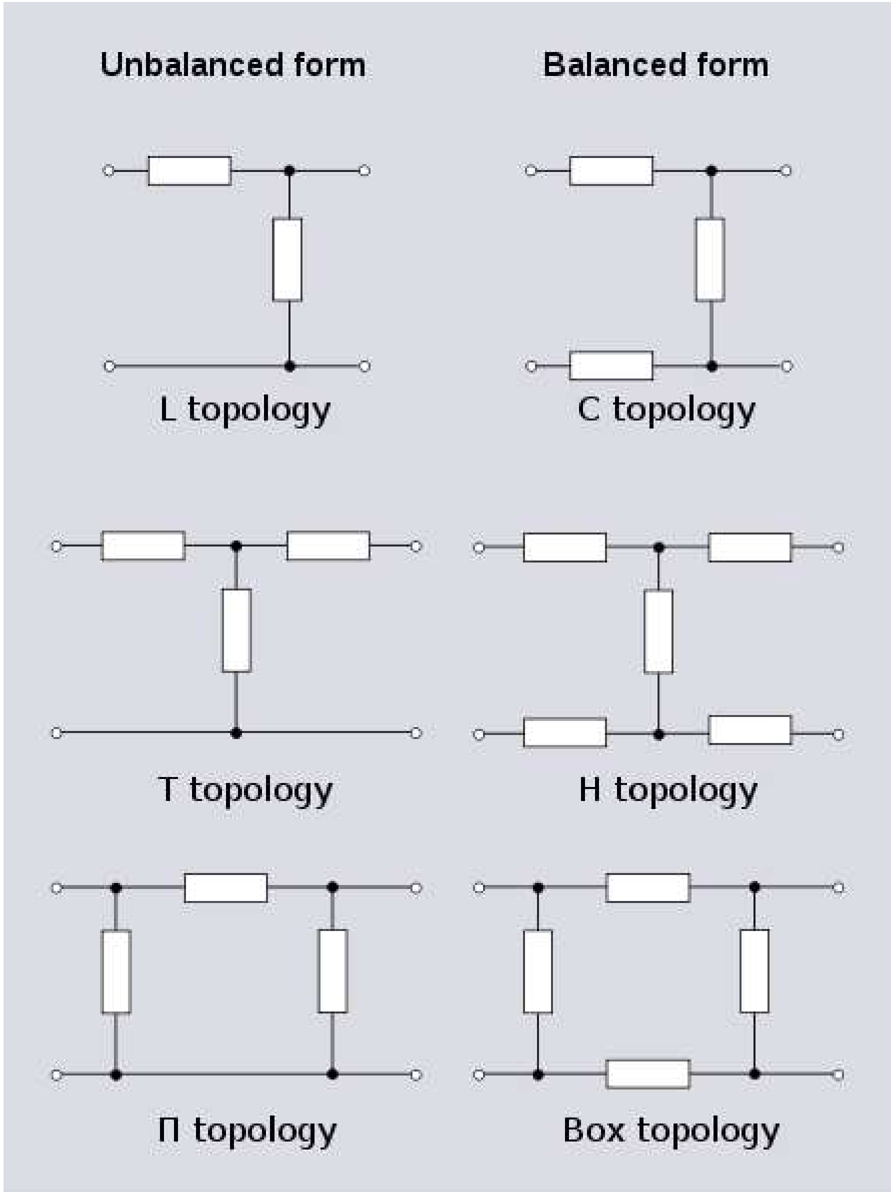


Fig. 2. A 4-terminal network connected by ports (top) has equal and opposite currents in each pair of terminals. The bottom network does not meet the port condition and cannot be treated as a 2-port. It could, however, be treated as an unbalanced 3-port by splitting one of the terminals into three common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

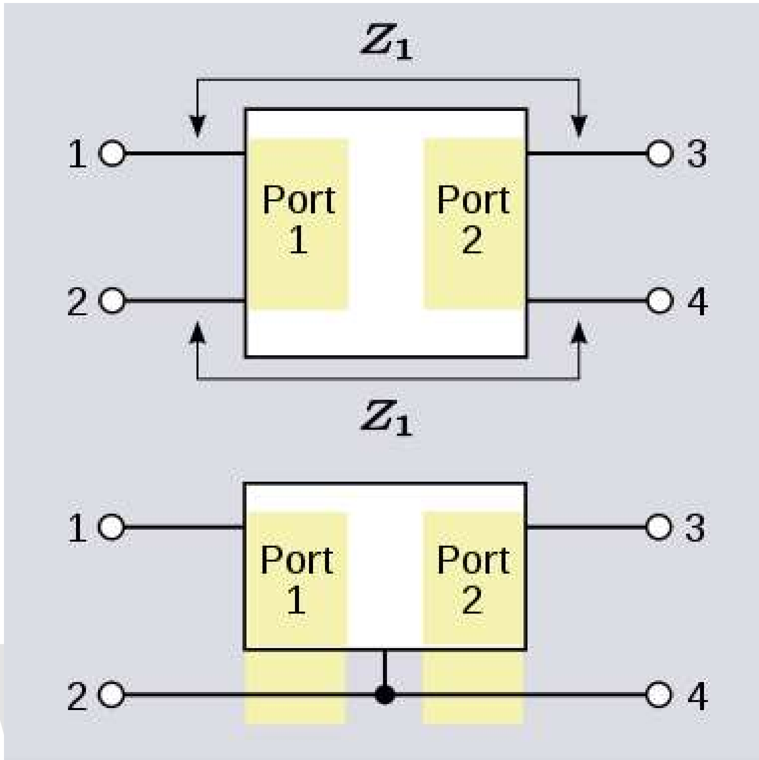


Fig. 3. To be balanced, a network must have the same impedance in each "leg" of the circuit.

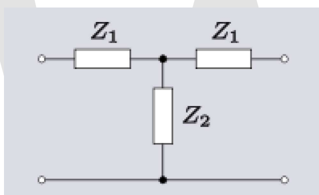
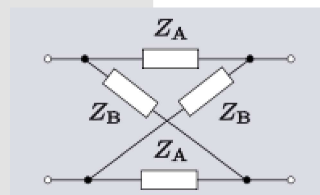
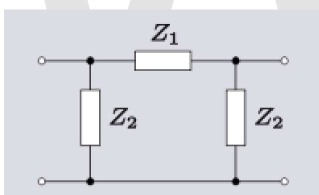
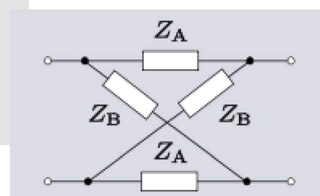
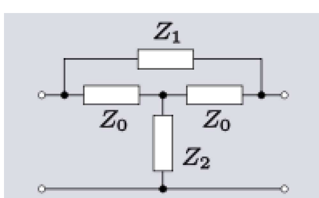
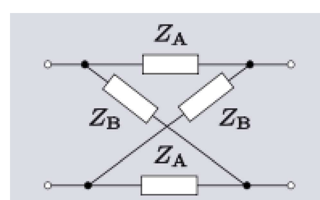
A 3-terminal network can also be used as a 2-port. To achieve this, one of the terminals is connected in common to one terminal of both ports. In other words, one terminal has been split into two terminals and the network has effectively been converted to a 4-terminal network. This topology is known 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.

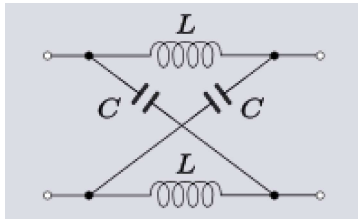
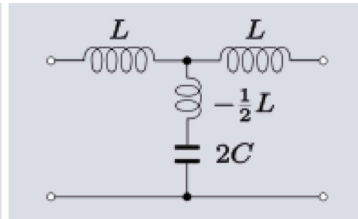
Lattice transforms

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)

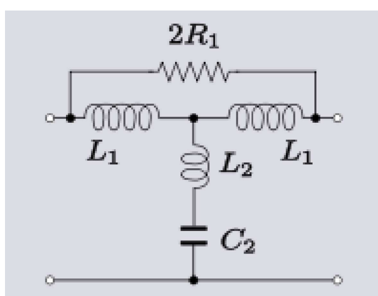
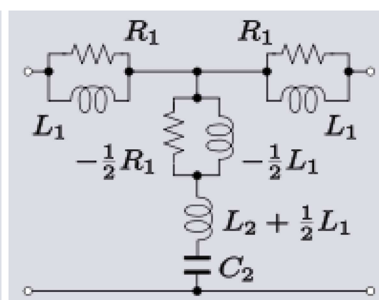
Description	Network	Transform equations	Transformed network
Transform 3.1 Transform of T network to lattice network.		$Z_A = Z_1$, $Z_B = Z_1 + 2Z_2$.	
Transform 3.2 Transform of Π network to lattice network.		$Z_A = \frac{Z_1 Z_2}{Z_1 + 2Z_2}$, $Z_B = Z_2$.	
Transform 3.3 Transform of Bridged-T network to lattice network.		$Z_A = \frac{Z_1 Z_0}{Z_1 + 2Z_0}$, $Z_B = Z_0 + 2Z_2$.	

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.

Description	Network	Transformed network
Transform 3.5 Transform of a bridged-T low-pass filter section to a T-section.		

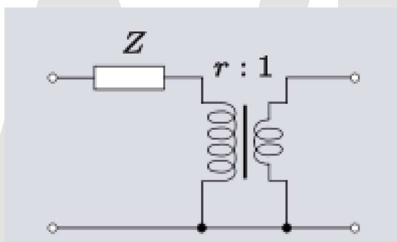
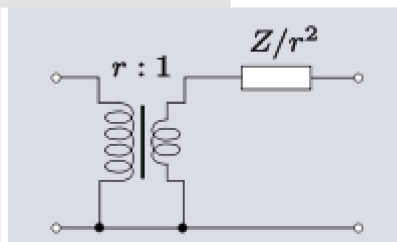
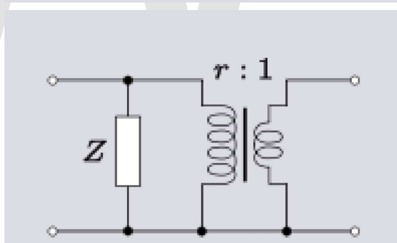
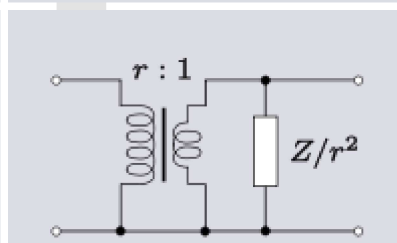
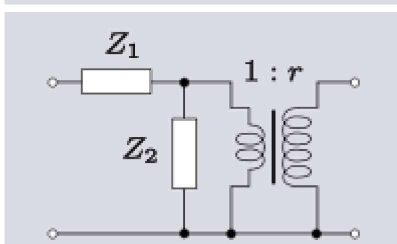
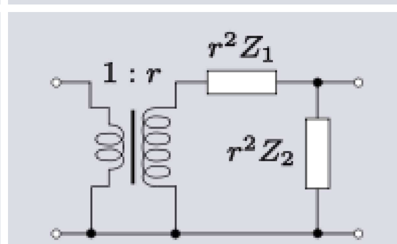
Any symmetrical network can be transformed into any other symmetrical network by the same method, that is, by first transforming into the intermediate lattice form (omitted for clarity from the above example transform) and from the lattice form into the required target form. As with the example, this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

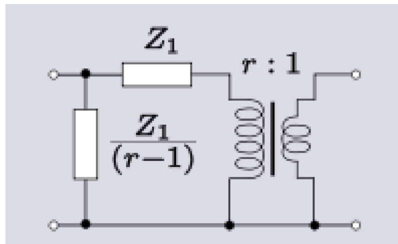
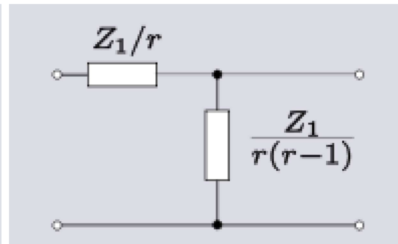
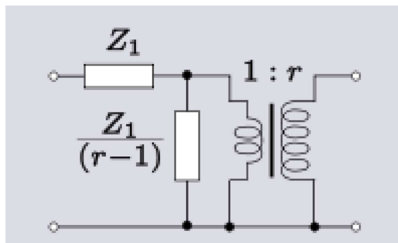
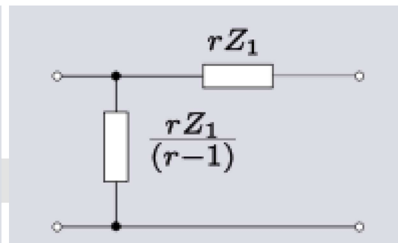
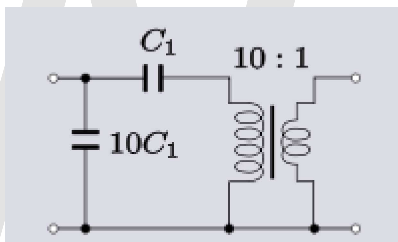
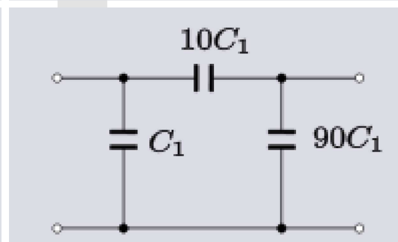
Eliminating ideal transformers

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.

Description	Network	Transformed network
Transform 4.1 Series impedance through a step-down transformer.		
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.	 	

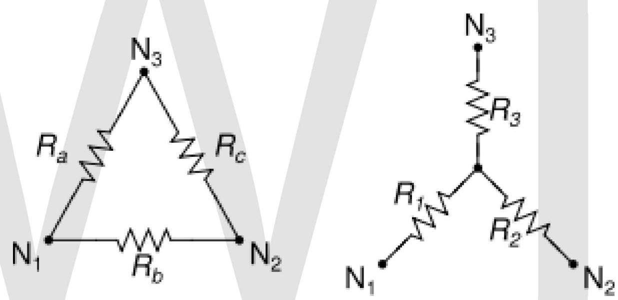
Example 3 shows the result is a Π-network rather than an L-network. The reason for this is that the shunt element has more capacitance than is required by the transform so some is still left over after applying the transform. If the excess were instead, in the element nearest the transformer, this could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

Chapter 5

Y-Δ Transform

The **Y-Δ transform**, also written **Y-delta**, **Wye-delta**, Kennelly's **delta-star transformation**, **star-mesh transformation**, **T-Π** or **T-pi transform**, is a mathematical technique to simplify the analysis of an electrical network. The name derives from the shapes of the circuit diagrams, which look respectively like the letter Y and the Greek capital letter Δ. In the United Kingdom, the wye diagram is sometimes known as a **star**. This circuit transformation theory was published by Arthur Edwin Kennelly in 1899.

Basic Y-Δ transformation



Δ and Y circuits with the labels which are used here

The transformation is used to establish equivalence for networks with three terminals. Where three elements terminate at a common node and none are sources, the node is eliminated by transforming the impedances. For equivalence, the impedance between any pair of terminals must be the same for both networks. The equations given here are valid for complex as well as real impedances.

Equations for the transformation from Δ-load to Y-load 3-phase circuit

The general idea is to compute the impedance R_y at a terminal node of the Y circuit with impedances R', R'' to adjacent nodes in the Δ circuit by

$$R_y = \frac{R'R''}{\sum R_\Delta}$$

where R_Δ are all impedances in the Δ circuit. This yields the specific formulae

$$R_1 = \frac{R_a R_b}{R_a + R_b + R_c},$$

$$R_2 = \frac{R_b R_c}{R_a + R_b + R_c},$$

$$R_3 = \frac{R_a R_c}{R_a + R_b + R_c}.$$

Equations for the transformation from Y-load to Δ-load 3-phase circuit

The general idea is to compute an impedance R_Δ in the Δ circuit by

$$R_\Delta = \frac{R_P}{R_\text{opposite}}$$

where $R_P = R_1R_2 + R_2R_3 + R_3R_1$ is the sum of the products of all pairs of impedances in the Y circuit and $R_{opposite}$ is the impedance of the node in the Y circuit which is opposite the edge with R_Δ. The formula for the individual edges are thus

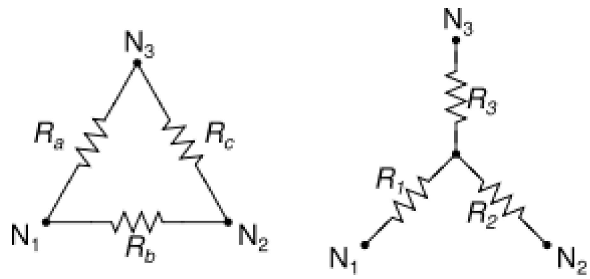
$$R_a = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_2},$$
$$R_b = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_3},$$
$$R_c = \frac{R_1R_2 + R_2R_3 + R_3R_1}{R_1}.$$

Graph theory

In graph theory, the Y-Δ transform means replacing a Y subgraph of a graph with the equivalent Δ subgraph. The transform preserves the number of edges in a graph, but not the number of vertices or the number of cycles. Two graphs are said to be **Y-Δ equivalent** if one can be obtained from the other by a series of Y-Δ transforms in either direction. For example, the Petersen graph family is a Y-Δ equivalence class.

Demonstration

Δ-load to Y-load transformation equations



Δ and Y circuits with the labels that are used here

To relate $\{R_a,R_b,R_c\}$ from Δ to $\{R_1,R_2,R_3\}$ from Y, the impedance between two corresponding nodes is compared. The impedance in either configuration is determined as if one of the nodes is disconnected from the circuit.

The impedance between N_1 and N_2 with N_3 disconnected in Δ:

$$R_\Delta(N_1,N_2) = R_b \parallel (R_a + R_c)$$
$$= \frac{1}{\frac{1}{R_b} + \frac{1}{R_a+R_c}}$$
$$= \frac{R_b(R_a+R_c)}{R_a+R_b+R_c}.$$

To simplify, let R_T be the sum of $\{R_a,R_b,R_c\}$.

$R_T = R_a + R_b + R_c$

Thus,

$$R_\Delta(N_1,N_2) = \frac{R_b(R_a+R_c)}{R_T}$$

The corresponding impedance between N_1 and N_2 in Y is simple:

$R_Y(N_1,N_2) = R_1 + R_2$

hence:

$$R_1 + R_2 = \frac{R_b(R_a + R_c)}{R_T} \quad (1)$$

Repeating for $R(N_2,N_3)$:

$$R_2 + R_3 = \frac{R_c(R_a + R_b)}{R_T} \quad (2)$$

and for $R(N_1,N_3)$:

$$R_1 + R_3 = \frac{R_a(R_b + R_c)}{R_T}. \quad (3)$$

From here, the values of $\{R_1,R_2,R_3\}$ can be determined by linear combination (addition and/or subtraction).

For example, adding (1) and (3), then subtracting (2) yields

$$R_1 + R_2 + R_1 + R_3 - R_2 - R_3 = \frac{R_b(R_a + R_c)}{R_T} + \frac{R_a(R_b + R_c)}{R_T} - \frac{R_c(R_a + R_b)}{R_T}$$

$$2R_1 = \frac{2R_bR_a}{R_T}$$

thus,

$$R_1 = \frac{R_bR_a}{R_T}.$$

where $R_T = R_a + R_b + R_c$

For completeness:

$$R_1 = \frac{R_bR_a}{R_T} \quad (4)$$

$$R_2 = \frac{R_bR_c}{R_T} \quad (5)$$

$$R_3 = \frac{R_a R_c}{R_T} \quad (6)$$

Y-load to Δ-load transformation equations

Let

$R_T = R_a + R_b + R_c$.

We can write the Δ to Y equations as

$$R_1 = \frac{R_a R_b}{R_T} \quad (1)$$

$$R_2 = \frac{R_b R_c}{R_T} \quad (2)$$

$$R_3 = \frac{R_a R_c}{R_T}. \quad (3)$$

Multiplying the pairs of equations yields

$$R_1 R_2 = \frac{R_a R_b^2 R_c}{R_T^2} \quad (4)$$

$$R_1 R_3 = \frac{R_a^2 R_b R_c}{R_T^2} \quad (5)$$

$$R_2 R_3 = \frac{R_a R_b R_c^2}{R_T^2} \quad (6)$$

and the sum of these equations is

$$R_1R_2 + R_1R_3 + R_2R_3 = \frac{R_aR_b^2R_c + R_a^2R_bR_c + R_aR_bR_c^2}{R_T^2} \quad (7)$$

Factor $R_aR_bR_c$ from the right side, leaving R_T in the numerator, canceling with an R_T in the denominator.

$$R_1R_2 + R_1R_3 + R_2R_3 = \frac{(R_aR_bR_c)(R_a + R_b + R_c)}{R_T^2}$$

$$R_1R_2 + R_1R_3 + R_2R_3 = \frac{R_aR_bR_c}{R_T} \quad (8)$$

-Note the similarity between (8) and {(1),(2),(3)}

Divide (8) by (1)

$$\frac{R_1R_2 + R_1R_3 + R_2R_3}{R_1} = \frac{R_aR_bR_c}{R_T}\frac{R_T}{R_aR_b},$$

$$\frac{R_1R_2 + R_1R_3 + R_2R_3}{R_1} = R_c,$$

which is the equation for R_c. Dividing (8) by (2) or (3) (expressions for R_2 or R_3) gives the remaining equations.

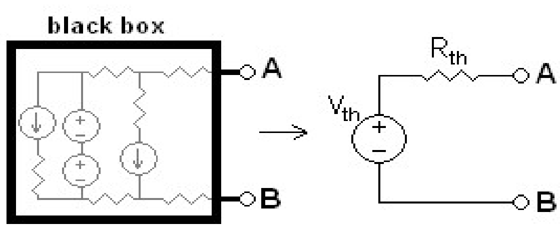
Chapter 6

Thévenin's Theorem and Tellegen's Theorem

Thévenin's theorem

In circuit theory, **Thévenin's theorem** for linear electrical networks states that any combination of voltage sources, current sources, and resistors with two terminals is electrically equivalent to a single voltage source V and a single series resistor R. For single frequency AC systems the theorem can also be applied to general impedances, not just resistors. The theorem was first discovered by German scientist Hermann von Helmholtz in 1853, but was then rediscovered in 1883 by French telegraph engineer Léon Charles Thévenin (1857–1926).

This theorem states that a circuit of voltage sources and resistors can be converted into a **Thévenin equivalent**, which is a simplification technique used in circuit analysis. The Thévenin equivalent can be used as a good model for a power supply or battery (with the resistor representing the internal impedance and the source representing the electromotive force). The circuit consists of an ideal voltage source in series with an ideal resistor.



Any black box containing only voltage sources, current sources, and other resistors can be converted to a Thévenin equivalent circuit, comprising exactly one voltage source and one resistor.

Calculating the Thévenin equivalent

To calculate the equivalent circuit, the resistance and voltage are needed, so two equations are required. These two equations are usually obtained by using the following steps, but any conditions placed on the terminals of the circuit should also work:

1. Calculate the output voltage, V_{AB}, when in open circuit condition (no load resistor—meaning infinite resistance). This is V_{Th}.
2. Calculate the output current, I_{AB}, when the output terminals are short circuited (load resistance is 0). R_{Th} equals V_{Th} divided by this I_{AB}.

The equivalent circuit is a voltage source with voltage V_{Th} in series with a resistance R_{Th}.

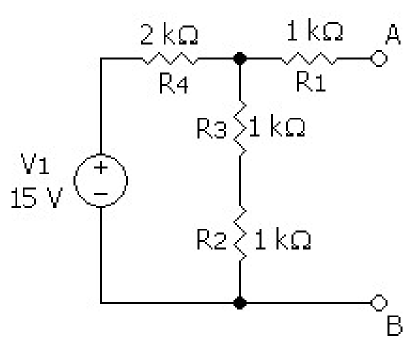
Step 2 could also be thought of as:

2a. Replace voltage sources with short circuits, and current sources with open circuits.
2b. Calculate the resistance between terminals A and B. This is R_{Th}.

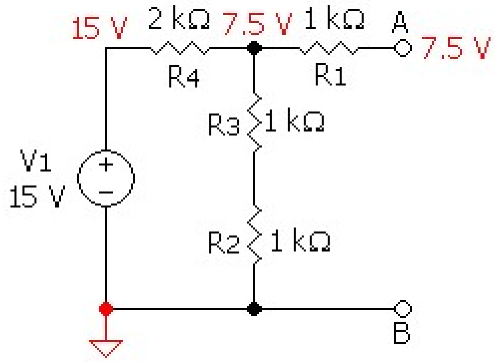
The Thévenin-equivalent voltage is the voltage at the output terminals of the original circuit. When calculating a Thévenin-equivalent voltage, the voltage divider principle is often useful, by declaring one terminal to be V_{out} and the other terminal to be at the ground point.

The Thévenin-equivalent resistance is the resistance measured across points A and B "looking back" into the circuit. It is important to first replace all voltage- and current-sources with their internal resistances. For an ideal voltage source, this means replace the voltage source with a short circuit. For an ideal current source, this means replace the current source with an open circuit. Resistance can then be calculated across the terminals using the formulae for series and parallel circuits. This method is valid only for circuits with independent sources. If there are dependent sources in the circuit, another method must be used such as connecting a test source across A and B and calculating the voltage across or current through the test source.

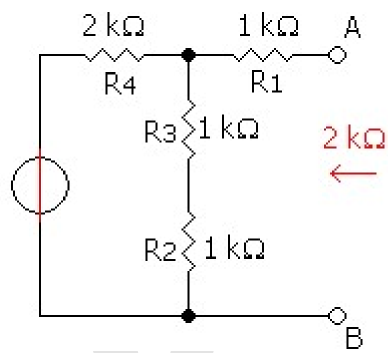
Example



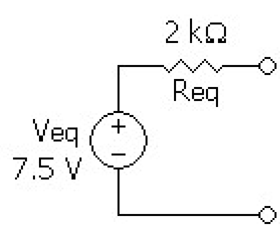
Step 0: The original circuit



Step 1: Calculating the equivalent output voltage



Step 2: Calculating the equivalent resistance



Step 3: The equivalent circuit

In the example, calculating the equivalent voltage:

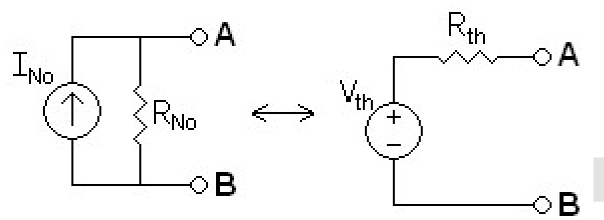
$$V_{\mathrm{Th}} = \frac{R_2 + R_3}{(R_2 + R_3) + R_4} \cdot V_1$$
$$= \frac{1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega}{(1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega) + 2\,\mathrm{k}\Omega} \cdot 15\,\mathrm{V}$$
$$= \frac{1}{2} \cdot 15\,\mathrm{V} = 7.5\,\mathrm{V}$$

(notice that R_1 is not taken into consideration, as above calculations are done in an open circuit condition between A and B, therefore no current flows through this part, which means there is no current through R_1 and therefore no voltage drop along this part)

Calculating equivalent resistance:

$$
\begin{aligned}
R_{\mathrm{Th}} &= R_1 + [(R_2 + R_3) \| R_4)] \\
&= 1\,\mathrm{k}\Omega + [(1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega) \| 2\,\mathrm{k}\Omega)] \\
&= 1\,\mathrm{k}\Omega + \left(\frac{1}{(1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega)} + \frac{1}{(2\,\mathrm{k}\Omega)} \right)^{-1} = 2\,\mathrm{k}\Omega
\end{aligned}
$$

Conversion to a Norton equivalent



A Norton equivalent circuit is related to the Thévenin equivalent by the following:

$$
\begin{aligned}
R_{\mathrm{Th}} &= R_{\mathrm{No}} \\
V_{\mathrm{Th}} &= I_{\mathrm{No}} R_{\mathrm{No}} \\
I_{\mathrm{No}} &= V_{\mathrm{Th}} / R_{\mathrm{Th}}
\end{aligned}
$$

Practical limitations

- Many, if not most circuits are only linear over a certain range of values, thus the Thévenin equivalent is valid only within this linear range and may not be valid outside the range.

- The Thévenin equivalent has an equivalent I-V characteristic only from the point of view of the load.

- The power dissipation of the Thévenin equivalent is not necessarily identical to the power dissipation of the real system. However, the power dissipated by an external resistor between the two output terminals is the same however the internal circuit is represented.

Tellegen's theorem

Tellegen's theorem is one of the most powerful theorems in network theory. Most of the energy distribution theorems and extremum principles in network theory can be derived from it. It was published in 1952 by Bernard Tellegen. Fundamentally, Tellegen's theorem gives a simple relation between magnitudes that satisfy the Kirchhoff's laws of electrical circuit theory.

The Tellegen theorem is applicable to a multitude of network systems. The basic assumptions for the systems are the conservation of flow of extensive quantities (Kirchhoff's current law, KCL) and the uniqueness of the potentials at the network nodes (Kirchhoff's voltage law, KVL). The Tellegen theorem provides a useful tool to analyze complex network systems among them electrical circuits, biological and metabolic networks, pipeline flow networks, and chemical process networks.

The theorem

Consider an arbitrary lumped network whose graph G has b branches and n_t nodes. In an electrical network, the branches are two-terminal components and the nodes are points of interconnection. Suppose that to each branch of the graph we assign arbitrarily a branch potential difference W_k and a branch current F_k for $k = 1, 2, \ldots, b$, and suppose that they are measured with respect to arbitrarily picked associated reference directions. If the branch potential differences $W_1, W_2, \ldots, W_b$ satisfy all the constraints imposed by KVL and if the branch currents $F_1, F_2, \ldots, F_b$ satisfy all the constraints imposed by KCL, then

$$\sum_{k=1}^{b} W_k F_k = 0.$$

Tellegen's theorem is extremely general; it is valid for any lumped network that contains any elements, linear or nonlinear, passive or active, time-varying or time-invariant. The generality is extended when W_k and F_k are linear operations on the set of potential differences and on the set of branch currents (respectively) since linear operations don't affect KVL and KCL. For instance, the linear operation may be the average or the Laplace transform. Another extension is when the set of potential differences W_k is from one network and the set of currents F_k is from an entirely different network, so long as the two networks have the same topology (same incidence matrix). This extension of Tellegen's Theorem leads to many theorems relating to two-port networks.

Definitions

We need to introduce a few necessary network definitions to provide a compact proof.

Incidence matrix: The $n_t \times n_f$ matrix $\mathbf{A}$ is called node-to-branch incidence matrix for the matrix elements a_{ij} being

$$a_{ij} = \begin{cases} 1, & \text{if flow } j \text{ leaves node } i \\ -1, & \text{if flow } j \text{ enters node } i \\ 0, & \text{if flow } j \text{ is not incident with node } i \end{cases}$$

A reference or datum node P_0 is introduced to represent the environment and connected to all dynamic nodes and terminals. The $(n_t - 1) \times n_f$ matrix $\mathbf{A}$, where the row that contains the elements a_{0j} of the reference node P_0 is eliminated, is called reduced incidence matrix.

The **conservation laws (KCL)** in vector-matrix form:

$$\mathbf{AF} = \mathbf{0}$$

The **uniqueness condition for the potentials (KVL)** in vector-matrix form:

$$\mathbf{W} = \mathbf{A^T w}$$

where w_k are the absolute potentials at the nodes to the reference node P_0.

Proof

Using KVL:

$$\mathbf{W^T F} = (\mathbf{A^T w})^\mathbf{T} \mathbf{F} = (\mathbf{w^T A})\mathbf{F} = \mathbf{w^T AF} = \mathbf{0}$$

because $\mathbf{AF} = \mathbf{0}$ by KCL. So:

$$\sum_{k=1}^{b} W_k F_k = \mathbf{W^T F} = 0$$

Applications

Network analogs have been constructed for a wide variety of physical systems, and have proven extremely useful in analyzing their dynamic behavior. The classical application area for network theory and Tellegen's theorem is electrical circuit theory. It is mainly in use to design filters in signal processing applications.

A more recent application of Tellegen's theorem is in the area of chemical and biological processes. The assumptions for electrical circuits (Kirchhoff laws) are generalized for dynamic systems obeying the laws of irreversible thermodynamics. Topology and structure of reaction networks (reaction mechanisms, metabolic networks) can be analyzed using the Tellegen theorem.

Another application of Tellegen's theorem is to determine stability and optimality of complex process systems such as chemical plants or oil production systems. The Tellegen theorem can be formulated for process systems using process nodes, terminals, flow connections and allowing sinks and sources for production or destruction of extensive quantities.

A formulation for Tellegen's theorem of process systems:

$$\sum_{j=1}^{n_P} W_j \frac{\mathrm{d}Z_j}{\mathrm{d}t} = \sum_{k=1}^{n_f} W_k f_k + \sum_{j=1}^{n_P} w_j p_j + \sum_{j=1}^{n_t} w_j t_j, \quad j = 1, \ldots, n_p + n_t$$

where p_j are the production terms, t_j are the terminal connections, and $\frac{\mathrm{d}Z_j}{\mathrm{d}t}$ are the dynamic storage terms for the extensive variables.

Chapter 7

Reciprocity (electromagnetism)

In classical electromagnetism, **reciprocity** refers to a variety of related theorems involving the interchange of time-harmonic electric current densities (sources) and the resulting electromagnetic fields in Maxwell's equations for time-invariant linear media under certain constraints. Reciprocity is closely related to the concept of Hermitian operators from linear algebra, applied to electromagnetism.

Perhaps the most common and general such theorem is **Lorentz reciprocity** (and its various special cases such as **Rayleigh-Carson reciprocity**), named after work by Hendrik Lorentz in 1896 following analogous results regarding sound by Lord Rayleigh and Helmholtz (Potton, 2004). Loosely, it states that **the relationship between an oscillating current and the resulting electric field is unchanged if one interchanges the points where the current is placed and where the field is measured**. For the specific case of an electrical network, it is sometimes phrased as the statement that voltages and currents at different points in the network can be interchanged. More technically, it follows that the mutual impedance of a first circuit due to a second is the same as the mutual impedance of the second circuit due to the first.

There is also an analogous theorem in electrostatics, known as **Green's reciprocity**, relating the interchange of electric potential and electric charge density.

Forms of the reciprocity theorems are used in many electromagnetic applications, such as analyzing electrical networks and antenna systems. For example, reciprocity implies that antennas work equally well as transmitters or receivers, and specifically that an antenna's radiation and receiving patterns are identical. Reciprocity is also a basic lemma that is used to prove other theorems about electromagnetic systems, such as the symmetry of the impedance matrix and scattering matrix, symmetries of Green's functions for use in boundary-element and transfer-matrix computational methods, as well as orthogonality properties of harmonic modes in waveguide systems (as an alternative to proving those properties directly from the symmetries of the eigen-operators).

Lorentz reciprocity

Specifically, suppose that one has a current density $\mathbf{J}_1$ that produces an electric field $\mathbf{E}_1$ and a magnetic field $\mathbf{H}_1$, where all three are periodic functions of time with angular frequency ω, and in particular they have time-dependence exp(− iωt). Suppose that we similarly have a second current $\mathbf{J}_2$ at the same frequency ω which (by itself) produces fields $\mathbf{E}_2$ and $\mathbf{H}_2$. The Lorentz reciprocity theorem then states, under certain simple conditions on the materials of the medium described below, that for an arbitrary surface S enclosing a volume V:

$$\int_V [\mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{E}_1 \cdot \mathbf{J}_2] \, dV = \oint_S [\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1] \cdot \mathbf{dA}.$$

Equivalently, in differential form (by the divergence theorem):

$$\mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{E}_1 \cdot \mathbf{J}_2 = \nabla \cdot [\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1] \, .$$

This general form is commonly simplified for a number of special cases. In particular, one usually assumes that $\mathbf{J}_1$ and $\mathbf{J}_2$ are localized (i.e. have compact support), and that there are no incoming waves from infinitely far away. In this case, if one integrates over all space then the surface-integral terms cancel and one obtains:

$$\int \mathbf{J}_1 \cdot \mathbf{E}_2 \, dV = \int \mathbf{E}_1 \cdot \mathbf{J}_2 \, dV$$

This result (along with the following simplifications) is sometimes called the **Rayleigh-Carson reciprocity theorem**, after Lord Rayleigh's work on sound waves and an extension by John R. Carson (1924; 1930) to applications for radio frequency antennas. Often, one further simplifies this relation by considering point-like dipole sources, in which case the integrals disappear and one simply has the product of the electric field with the corresponding dipole moments of the currents. Or, for wires of negligible thickness, one obtains the applied current in one wire multiplied by the resulting voltage across another and vice versa.

Another special case of the Lorentz reciprocity theorem applies when the volume V entirely contains both of the localized sources (or alternatively if V intersects neither of the sources). In this case:

$$\oint_S (\mathbf{E}_1 \times \mathbf{H}_2) \cdot \mathbf{dA} = \oint_S (\mathbf{E}_2 \times \mathbf{H}_1) \cdot \mathbf{dA}$$

Reciprocity for electrical networks

Above, Lorentz reciprocity was phrased in terms of an externally applied current source and the resulting field. Often, especially for electrical networks, one instead prefers to think of an externally applied voltage and the resulting currents. The Lorentz reciprocity theorem describes this case as well, assuming ohmic materials (i.e. currents that respond linearly to the applied field) with a 3×3 conductivity matrix σ that is required to be symmetric, which is implied by the other conditions below. In order to properly describe this situation, one must carefully distinguish between the externally applied fields (from the driving voltages) and the total fields that result (King, 1963).

More specifically, the $\mathbf{J}$ above only consisted of external "source" terms introduced into Maxwell's equations. We now denote this by $\mathbf{J}^{(e)}$ to distinguish it from the total current produced by both the external source and by the resulting electric fields in the materials. If this external current is in a material with a conductivity σ, then it corresponds to an externally applied electric field $\mathbf{E}^{(e)}$ where, by definition of σ:

$$\mathbf{J}^{(e)} = \sigma \mathbf{E}^{(e)}$$

Moreover, the electric field $\mathbf{E}$ above only consisted of the response to this current, and did not include the "external" field $\mathbf{E}^{(e)}$. Therefore, we now denote the field from before as $\mathbf{E}^{(r)}$, where the total field is given by $\mathbf{E} = \mathbf{E}^{(e)} + \mathbf{E}^{(r)}$.

Now, the equation on the left-hand side of the Lorentz reciprocity theorem can be rewritten by moving the σ from the external current term $\mathbf{J}^{(e)}$ to the response field terms $\mathbf{E}^{(r)}$, and also adding and subtracting a $\sigma \mathbf{E}_1^{(e)} \mathbf{E}_2^{(e)}$ term, to obtain the external field multiplied by the total current $\mathbf{J} = \sigma \mathbf{E}$:

$$\int_V \left[\mathbf{J}_1^{(e)} \cdot \mathbf{E}_2^{(r)} - \mathbf{E}_1^{(r)} \cdot \mathbf{J}_2^{(e)}\right] dV = \int_V \left[\sigma \mathbf{E}_1^{(e)} \cdot (\mathbf{E}_2^{(r)} + \mathbf{E}_2^{(e)}) - (\mathbf{E}_1^{(r)} + \mathbf{E}_1^{(e)}) \cdot \sigma \mathbf{E}_2^{(e)}\right] dV$$
$$= \int_V \left[\mathbf{E}_1^{(e)} \cdot \mathbf{J}_2 - \mathbf{J}_1 \cdot \mathbf{E}_2^{(e)}\right] dV$$

For the limit of thin wires, this gives the product of the externally applied voltage (1) multiplied by the resulting total current (2) and vice versa. In particular, the Rayleigh-Carson reciprocity theorem becomes a simple summation:

$$\sum_n V_1^{(n)} I_2^{(n)} = \sum_n V_2^{(n)} I_1^{(n)}$$

where V and I denote the (complex) amplitudes of the AC applied voltages and the resulting currents, respectively, in a set of circuit elements (indexed by n) for two possible sets of voltages V_1 and V_2.

Most commonly, this is simplified further to the case where each system has a single voltage source V, at $V_1^{(1)} = V$ and $V_2^{(2)} = V$. Then the theorem becomes simply $I_1^{(2)} = I_2^{(1)}$: **the current at position (1) from a voltage at (2) is identical to the current at (2) from the same voltage at (1).**

Conditions and proof of Lorentz reciprocity

The Lorentz reciprocity theorem is simply a reflection of the fact that the linear operator $\hat{O}$relating $\mathbf{J}$and $\mathbf{E}$at a fixed frequency (in linear media):

$$\mathbf{J} = \frac{1}{i\omega}\left[\left(\nabla \times \frac{1}{\mu}\nabla\times\right) - \omega^2\varepsilon\right]\mathbf{E} \equiv \hat{O}\mathbf{E}$$

is usually a Hermitian operator under the inner product $(\mathbf{F}, \mathbf{G}) = \int \mathbf{F} \cdot \mathbf{G}\, dV$ for vector fields $\mathbf{F}$and $\mathbf{G}$. (Technically, this unconjugated form is not a true inner product because it is not real-valued for complex-valued fields, but that is not a problem here. In this sense, the operator is not truly Hermitian but is rather complex-symmetric.) This is true whenever the permittivity ε and the magnetic permeability μ, at the given ω, are symmetric 3×3 matrices (symmetric rank-2 tensors) — this includes the common case where they are scalars (for isotropic media), of course. They need not be real—complex values correspond to materials with losses, such as conductors with finite conductivity σ (which is included in ε via $\varepsilon \rightarrow \varepsilon + i\sigma/\omega$)—and because of this the reciprocity theorem does not require time reversal invariance. The condition of symmetric ε and μ matrices is almost always satisfied; see below for an exception.

For any Hermitian operator $\hat{O}$under an inner product (f, g), we have $(f, \hat{O}g) = (\hat{O}f, g)$by definition, and the Rayleigh-Carson reciprocity theorem is merely the vectorial version of this statement for this particular operator $\mathbf{J} = \hat{O}\mathbf{E}$: that is, $(\mathbf{E}_1, \hat{O}\mathbf{E}_2) = (\hat{O}\mathbf{E}_1, \mathbf{E}_2)$. The Hermitian property of the operator here can be derived by integration by parts. For a finite integration volume, the surface terms from this integration by parts yield the more-general surface-integral theorem above. In particular, the key fact is that, for vector fields $\mathbf{F}$and $\mathbf{G}$, integration by parts (or the divergence theorem) over a volume V enclosed by a surface S gives the identity:

$$\int_V \mathbf{F} \cdot (\nabla \times \mathbf{G})\, dV = \int_V (\nabla \times \mathbf{F}) \cdot \mathbf{G}\, dV - \oint_S (\mathbf{F} \times \mathbf{G}) \cdot \mathbf{dA}$$

This identity is then applied twice to $(\mathbf{E}_1, \hat{O}\mathbf{E}_2)$to yield $(\hat{O}\mathbf{E}_1, \mathbf{E}_2)$plus the surface term, giving the Lorentz reciprocity relation.

Surface-term cancellation

The cancellation of the surface terms on the right-hand side of the Lorentz reciprocity theorem, for an integration over all space, is not entirely obvious but can be derived in a number of ways.

The simplest argument would be that the fields goes to zero at infinity for a localized source, but this argument fails in the case of lossless media: in the absence of absorption, radiated fields decay inversely with distance, but the surface area of the integral increases with the square of distance, so the two rates balance one another in the integral.

Instead, it is common (e.g. King, 1963) to assume that the medium is homogeneous and isotropic sufficiently far away. In this case, the radiated field asymptotically takes the form of planewaves propagating radially outward (in the $\hat{\mathbf{r}}$ direction) with $\hat{\mathbf{r}} \cdot \mathbf{E} = 0$ and $\mathbf{H} = \hat{\mathbf{r}} \times \mathbf{E}/Z$ where Z is the impedance $\sqrt{\mu/\epsilon}$ of the surrounding medium. Then it follows that $\mathbf{E}_1 \times \mathbf{H}_2 = \mathbf{E}_1 \times \hat{\mathbf{r}} \times \mathbf{E}_2/Z$, which by a simple vector identity equals $\hat{\mathbf{r}}(\mathbf{E}_1 \cdot \mathbf{E}_2)/Z$. Similarly, $\mathbf{E}_2 \times \mathbf{H}_1 = \hat{\mathbf{r}}(\mathbf{E}_2 \cdot \mathbf{E}_1)/Z$ and the two terms cancel one another.

The above argument shows explicitly why the surface terms can cancel, but lacks generality. Alternatively, one can treat the case of lossless surrounding media by taking the limit as the losses (the imaginary part of ε) go to zero. For any nonzero loss, the fields decay exponentially with distance and the surface integral vanishes, regardless of whether the medium is homogeneous. Since the left-hand side of the Lorentz reciprocity theorem vanishes for integration over all space with any non-zero losses, it must also vanish in the limit as the losses go to zero. (Note that we implicitly assumed the standard boundary condition of zero incoming waves from infinity, because otherwise even an infinitesimal loss would eliminate the incoming waves and the limit would not give the lossless solution.)

Reciprocity and the Green's function

The inverse of the operator $\hat{O}$, i.e. in $\mathbf{E} = \hat{O}^{-1}\mathbf{J}$ (which requires a specification of the boundary conditions at infinity in a lossless system), has the same symmetry as $\hat{O}$ and is essentially a Green's function convolution. So, another perspective on Lorentz reciprocity is that it reflects the fact that convolution with the electromagnetic Green's function is a complex-symmetric (or anti-Hermitian, below) linear operation under the appropriate conditions on ε and μ. More specifically, the Green's function can be written as $G_{nm}(\mathbf{x}', \mathbf{x})$ giving the n-th component of $\mathbf{E}$ at $\mathbf{x}'$ from a point dipole current in the m-th direction at $\mathbf{x}$ (essentially, G gives the matrix elements of $\hat{O}^{-1}$), and Rayleigh-Carson reciprocity is equivalent to the statement that $G_{nm}(\mathbf{x}', \mathbf{x}) = G_{mn}(\mathbf{x}, \mathbf{x}')$. Unlike $\hat{O}$, it is not generally possible to give an explicit formula for the Green's function (except in special cases such as homogeneous media), but it is routinely computed by numerical methods.

Lossless magneto-optic materials

One case in which ε is not a symmetric matrix is for magneto-optic materials, in which case the usual statement of Lorentz reciprocity does not hold. If we allow magneto-optic materials, but restrict ourselves to the situation where material absorption is negligible, then ε and μ are in general 3×3 complex Hermitian matrices. In this case the operator $\nabla \times \frac{1}{\mu}\nabla \times -(\omega^2/c^2)\varepsilon$ is Hermitian under the conjugated inner product $(\mathbf{F}, \mathbf{G}) = \int \mathbf{F}^* \cdot \mathbf{G}\, dV$, and a variant of the reciprocity theorem still holds:

$$-\int_V [\mathbf{J}_1^* \cdot \mathbf{E}_2 + \mathbf{E}_1^* \cdot \mathbf{J}_2]\, dV = \oint_S [\mathbf{E}_1^* \times \mathbf{H}_2 + \mathbf{E}_2 \times \mathbf{H}_1^*] \cdot \mathbf{dA}$$

where the sign changes come from the 1 / iω in the equation above, which makes the operator $\hat{O}$anti-Hermitian (neglecting surface terms). For the special case of $\mathbf{J}_1 = \mathbf{J}_2$, this gives a re-statement of conservation of energy or Poynting's theorem (since here we have assumed lossless materials, unlike above): the time-average rate of work done by the current (given by the real part of $-\mathbf{J}^* \cdot \mathbf{E}$) is equal to the time-average outward flux of power (the integral of the Poynting vector). By the same token, however, the surface terms do not in general vanish if one integrates over all space for this reciprocity variant, so a Rayleigh-Carson form does not hold without additional assumptions.

The fact that magneto-optic materials break Rayleigh-Carson reciprocity is the key to devices such as Faraday isolators and circulators. A current on one side of a Faraday isolator produces a field on the other side but not vice-versa.

Generalization to non-symmetric materials

For a combination of lossy and magneto-optic materials, and in general when the ε and μ tensors are neither symmetric nor Hermitian matrices, one can still obtain a generalized version of Lorentz reciprocity by considering $(\mathbf{J}_1, \mathbf{E}_1)$ and $(\mathbf{J}_2, \mathbf{E}_2)$ to exist in different systems.

In particular, if $(\mathbf{J}_1, \mathbf{E}_1)$ satisfy Maxwell's equations at ω for a system with materials (ε_1, μ_1), and $(\mathbf{J}_2, \mathbf{E}_2)$ satisfy Maxwell's equations at ω for a system with materials $(\varepsilon_1^T, \mu_1^T)$, where T denotes the transpose, then the equation of Lorentz reciprocity holds.

Exceptions to reciprocity

For nonlinear media, no reciprocity theorem generally holds. Reciprocity also does not generally apply for time-varying ("active") media; for example, when ε is modulated in time by some external process. (In both of these cases, the frequency ω is not generally a conserved quantity.)

Feld-Tai reciprocity

A closely related reciprocity theorem was articulated independently by Y. A. Feld and C. T. Tai in 1992 and is known as **Feld-Tai reciprocity** or the **Feld-Tai lemma**. It relates two time-harmonic localized current sources and the resulting magnetic fields:

$$\int \mathbf{J}_1 \cdot \mathbf{H}_2 dV = \int \mathbf{H}_1 \cdot \mathbf{J}_2 dV.$$

However, the Feld-Tai lemma is only valid under much more restrictive conditions than Lorentz reciprocity. It generally requires time-invariant linear media with an isotropic homogeneous impedance, i.e. a constant scalar μ/ε ratio, with the possible exception of regions of perfectly conducting material.

More precisely, Feld-Tai reciprocity requires the Hermitian (or rather, complex-symmetric) symmetry of the electromagnetic operators as above, but also relies on the assumption that the operator relating $\mathbf{E}$ and $i\omega\mathbf{J}$ is a constant scalar multiple of the operator relating $\mathbf{H}$ and $\nabla \times (\mathbf{J}/\varepsilon)$, which is true when ε is a constant scalar multiple of μ (the two operators generally differ by an interchange of ε and μ). As above, one can also construct a more general formulation for integrals over a finite volume.

Green's reciprocity

Whereas the above reciprocity theorems were for oscillating fields, **Green's reciprocity** is an analogous theorem for electrostatics with a fixed distribution of electric charge (Panofsky and Phillips, 1962).

In particular, let φ_1 denote the electric potential resulting from a total charge density ρ_1. The electric potential satisfies Poisson's equation, $-\nabla^2\phi_1 = \rho_1/\varepsilon_0$, where ε_0 is the vacuum permittivity. Similarly, let φ_2 denote the electric potential resulting from a total charge density ρ_2, satisfying $-\nabla^2\phi_2 = \rho_2/\varepsilon_0$. In both cases, we assume that the charge distributions are localized, so that the potentials can be chosen to go to zero at infinity. Then, Green's reciprocity theorem states that, for integrals over all space:

$$\int \rho_1\phi_2 dV = \int \rho_2\phi_1 dV.$$

This theorem is easily proven from Green's second identity. Equivalently, it is the statement that $\int \phi_2(\nabla^2\phi_1)dV = \int \phi_1(\nabla^2\phi_2)dV$, i.e. that ∇^2 is a Hermitian operator (as follows by integrating by parts twice).

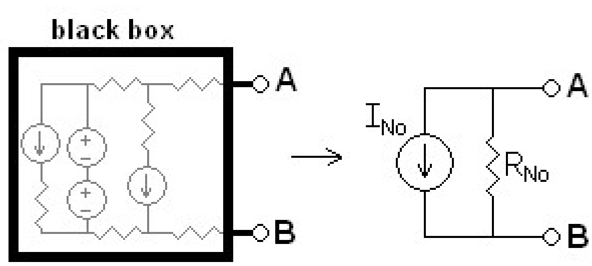
Chapter 8

Norton's Theorem, Extra Element Theorem and Millman's Theorem

Norton's theorem

Norton's theorem for linear electrical networks, known in Europe as the **Mayer–Norton theorem**, states that any collection of voltage sources, current sources, and resistors with two terminals is electrically equivalent to an ideal current source, I, in parallel with a single resistor, R. For single-frequency AC systems the theorem can also be applied to general impedances, not just resistors. The **Norton equivalent** is used to represent any network of linear sources and impedances, at a given frequency. The circuit consists of an ideal current source in parallel with an ideal impedance (or resistor for non-reactive circuits).

Norton's theorem is an extension of Thévenin's theorem and was introduced in 1926 separately by two people: Hause-Siemens researcher Hans Ferdinand Mayer (1895–1980) and Bell Labs engineer Edward Lawry Norton (1898–1983). Only Mayer actually published on this topic, but Norton made known his finding through an internal technical report at Bell Labs.



Any black box containing only voltage sources, current sources, and resistors can be converted to a Norton equivalent circuit.

Calculation of a Norton equivalent circuit

The Norton equivalent circuit is a current source with current I_{No} in parallel with a resistance R_{No}. To find the equivalent,

1. Find the Norton current I_{No}. Calculate the output current, I_{AB}, with a short circuit as the load (meaning 0 resistance between A and B). This is I_{No}.
2. Find the Norton resistance R_{No}. When there are **no dependent sources** (i.e., all current and voltage sources are **independent**), there are two methods of determining the Norton impedance R_{No}.

    - Calculate the output voltage, V_{AB}, when in open circuit condition (i.e., no load resistor — meaning infinite load resistance). R_{No} equals this V_{AB} divided by I_{No}.

    or

    - Replace independent voltage sources with short circuits and independent current sources with open circuits. The total resistance across the output port is the Norton impedance R_{No}.

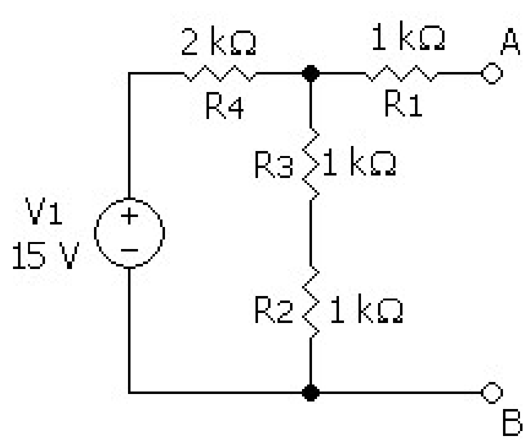
    or

    - Use a given Thevenin resistance: as the two are equal.

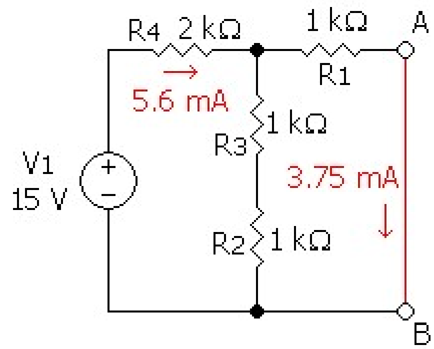
    However, when there are **dependent sources**, the more general method must be used. This method is not shown below in the diagrams.

    - Connect a constant current source at the output terminals of the circuit with a value of 1 Ampere and calculate the voltage at its terminals. The quotient of this voltage divided by the 1 A current is the Norton impedance R_{No}. This method must be used if the circuit contains dependent sources, but it can be used in all cases even when there are no dependent sources.

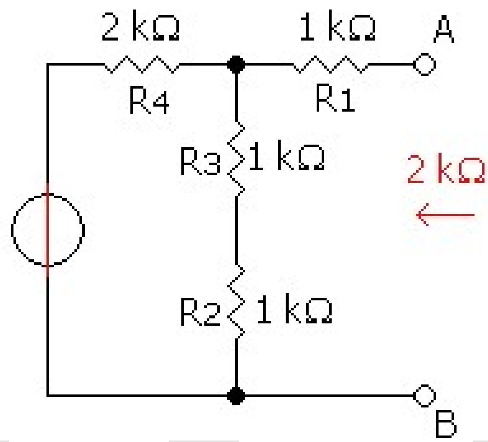
Example of a Norton equivalent circuit



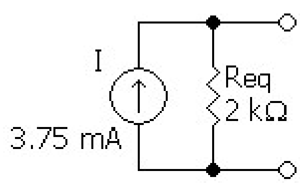
Step 0: The original circuit



Step 1: Calculating the equivalent output current



Step 2: Calculating the equivalent resistance



Step 3: The equivalent circuit

In the example, the total current I_{total} is given by:

$$I_{\text{total}} = \frac{15\text{V}}{2\,\text{k}\Omega + 1\,\text{k}\Omega \| (1\,\text{k}\Omega + 1\,\text{k}\Omega)} = 5.625\text{mA}$$

The current through the load is then, using the current divider rule:

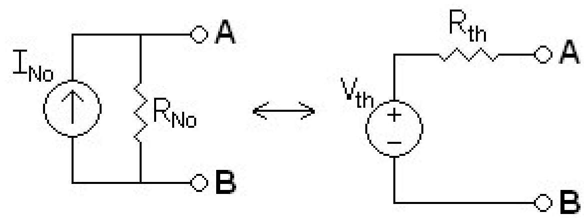
$$I = \frac{1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega}{(1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega)} \cdot I_{\mathrm{total}}$$
$$= 2/3 \cdot 5.625\mathrm{mA} = 3.75\mathrm{mA}$$

And the equivalent resistance looking back into the circuit is:

$$R_{\mathrm{eq}} = 1\,\mathrm{k}\Omega + 2\,\mathrm{k}\Omega \| (1\,\mathrm{k}\Omega + 1\,\mathrm{k}\Omega) = 2\,\mathrm{k}\Omega$$

So the equivalent circuit is a 3.75 mA current source in parallel with a 2 kΩ resistor.

Conversion to a Thévenin equivalent



A Norton equivalent circuit is related to the Thévenin equivalent by the following equations:

$$R_{Th} = R_{No}$$
$$V_{Th} = I_{No} R_{No}$$
$$V_{Th}/R_{Th} = I_{No}$$

Queueing theory

The term "Norton equivalent" is also used in queueing theory for a similar concept. In a reversible queueing system, it is often possible to replace an uninteresting subset of queues by a single (FCFS or PS) queue with an appropriately chosen service rate.

Extra element theorem

The **Extra Element Theorem** (EET) is an analytic technique developed by R.D. Middlebrook for simplifying the process of deriving driving point and transfer functions for linear electronic circuits. Much like Thevenin's theorem, the extra element theorem breaks down one complicated problem into several simpler ones.

Driving point and transfer functions can generally be found using KVL and KCL methods, however several complicated equations may result that offer little insight into the circuit's behavior. Using the extra element theorem, a circuit element (such as a resistor) can be removed from a circuit and the desired driving point or transfer function found. By removing the element that most complicates the circuit (such as an element that creates feedback), the desired function can be easier to obtain. Next two correctional factors must be found and combined with the previously derived function to find the exact expression.

The general form of the extra element theorem is called the N-extra element theorem and allows multiple circuit elements to be removed at once.

Driving point impedances

As a special case, the EET can be used to find the input impedance of a network. For this application the EET can be written as:

$$Z_{in} = Z_{in}^{\infty} \left(\frac{1 + \frac{Z_e^0}{Z}}{1 + \frac{Z_e^{\infty}}{Z}} \right)$$

where

Z is the impedance chosen as the extra element
Z_{in}^{∞} is the input impedance with Z removed (or made infinite)
Z_e^0 is the impedance seen by the extra element Z with the input shorted (or made zero)
Z_e^{∞} is the impedance seen by the extra element Z with the input open (or made infinite)

Computing these three terms may seem like extra effort, but they are often easier to compute than the overall input impedance.

Example

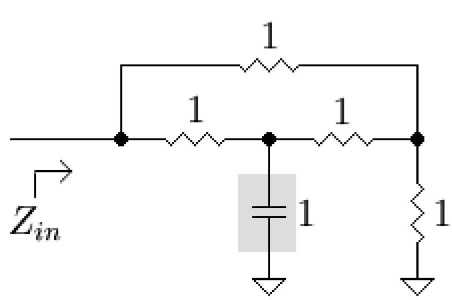


Figure 1: Simple RC circuit to demonstrate the EET. The capacitor (gray shading) is denoted the extra element

Consider the problem of finding Z_{in} for the circuit in Figure 1 using the EET (note all component values are unity for simplicity). If the capacitor (gray shading) is denoted the extra element then

$$Z = \frac{1}{s}$$

Removing this capacitor from the circuit we find

$$Z_{in}^{\infty} = 2\|1 + 1 = \frac{5}{3}$$

Calculating the impedance seen by the capacitor with the input shorted we find

$$Z_e^0 = 1\|(1 + 1\|1) = \frac{3}{5}$$

Calculating the impedance seen by the capacitor with the input open we find

$$Z_e^{\infty} = 2\|1 + 1 = \frac{5}{3}$$

Therefore using the EET, we find

$$Z_{in} = \frac{5}{3}\left(\frac{1 + \frac{3}{5}s}{1 + \frac{5}{3}s}\right)$$

Note that this problem was solved by calculating three simple driving point impedances by inspection.

Feedback amplifiers

The EET is also useful for analyzing single and multi-loop feedback amplifiers. In this case the EET can take the form of the Asymptotic gain model.

Millman's theorem

In electrical engineering, **Millman's theorem** (or the **parallel generator theorem**) is a method to simplify the solution of a circuit. Specifically, Millman's theorem is used to compute the voltage at the ends of a circuit made up of only branches in parallel.

It is named after Jacob Millman, who proved the theorem.

Explanation

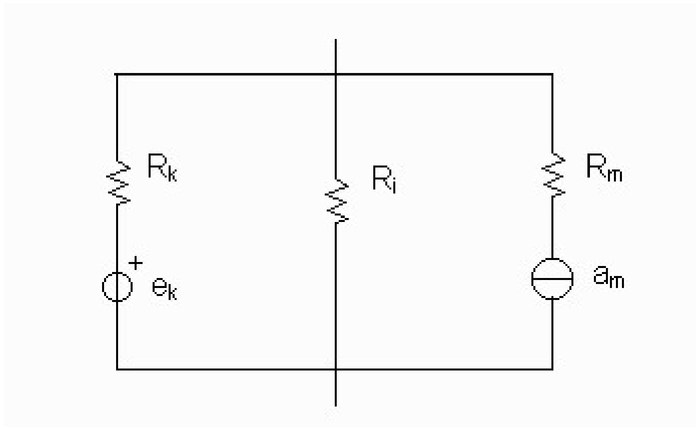


Figure 1

Let e_k be the voltage generators and a_m the current generators.

Let R_i be the resistances on the branches with no generator.

Let R_k be the resistances on the branches with voltage generators.

Let R_m be the resistances on the branches with current generators.

Then Millman states that the voltage at the ends of the circuit is given by:

$$v = \frac{\sum \frac{\pm e_k}{R_k} + \sum \pm a_m}{\sum \frac{1}{R_k} + \sum \frac{1}{R_i}}$$

It can be proved by considering the circuit as a single supernode. Then, according to Ohm and Kirchhoff, "the voltage between the ends of the circuit is equal to the total current entering the supernode divided by the total equivalent conductance of the supernode".

The total current is the sum of the currents flowing in each branch.

The total equivalent conductance of the supernode is the sum of the conductance of each branch, since all the branches are in parallel. When computing the equivalent conductance all the generators have to be switched off, so all voltage generators become short circuits and all current generators become open circuits. That's why the resistances on the branches with current generators do not appear in the expression of the total equivalent conductance.

Chapter 9

Faraday's Law of Induction

Faraday's law of induction is a basic law of electromagnetism relating to the operating principles of transformers, inductors, and many types of electrical motors and generators. The law states that:

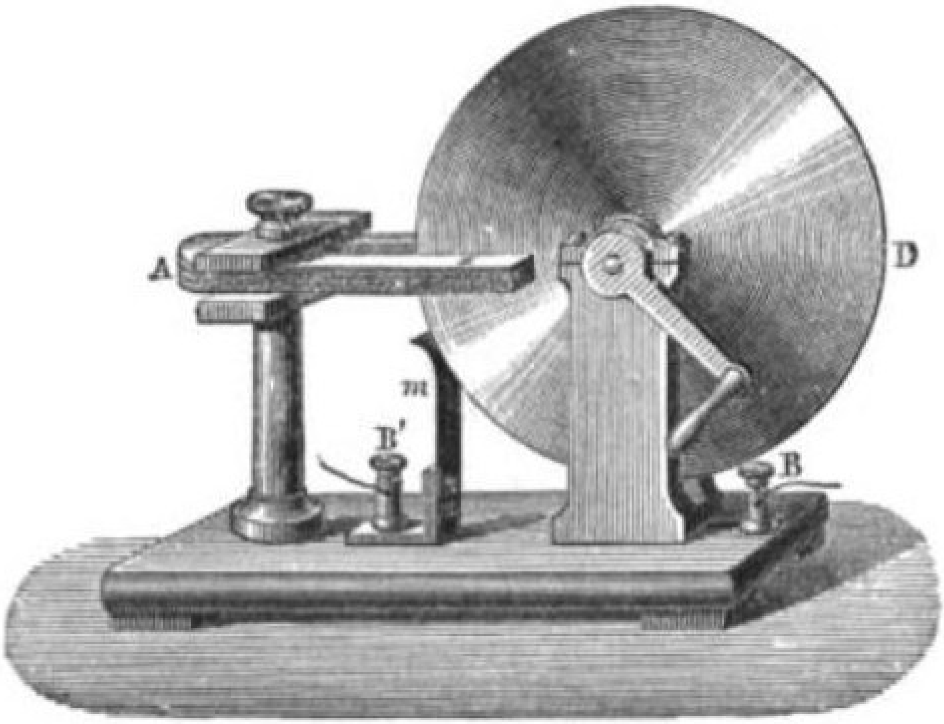
The induced electromotive force (EMF) in any closed circuit is equal to the time rate of change of the magnetic flux through the circuit.

Or alternatively:

The EMF generated is proportional to the rate of change of the magnetic flux.

History

Electromagnetic induction was discovered independently by Michael Faraday and Joseph Henry in 1831; however, Faraday was the first to publish the results of his experiments.

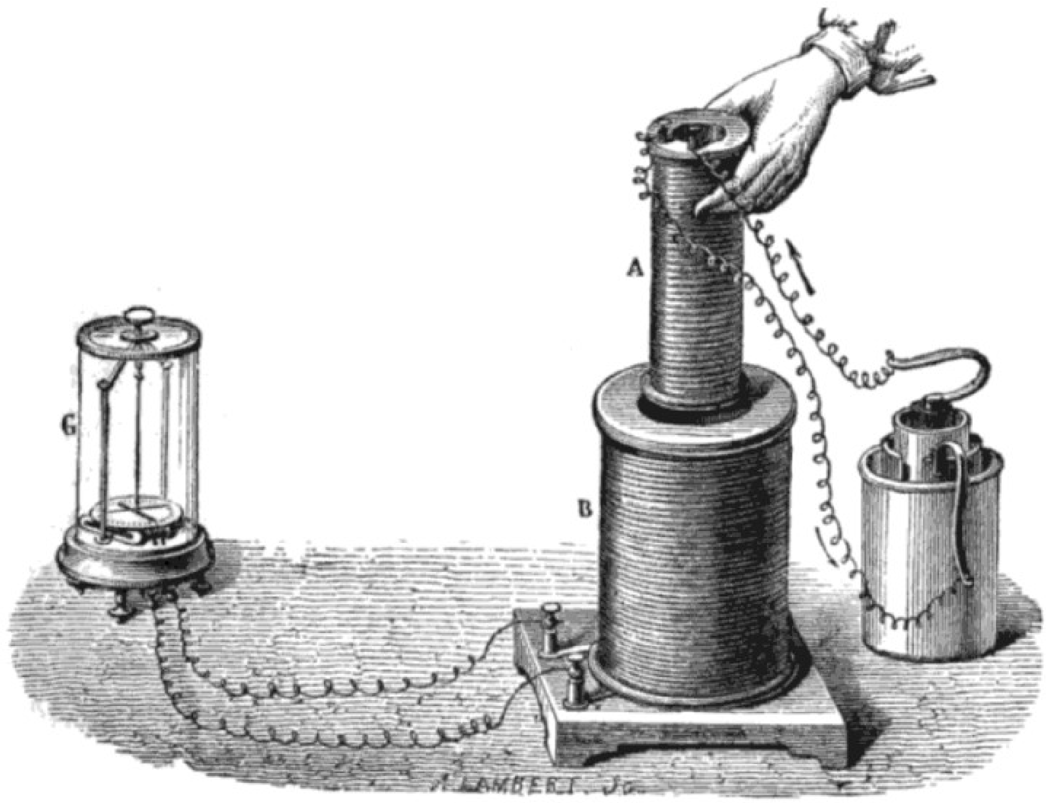


Faraday's disk

In Faraday's first experimental demonstration of electromagnetic induction (August 1831), he wrapped two wires around opposite sides of an iron torus (an arrangement similar to a modern transformer). Based on his assessment of recently-discovered properties of electromagnets, he expected that when current started to flow in one wire, a sort of wave would travel through the ring and cause some electrical effect on the opposite side. He plugged one wire into a galvanometer, and watched it as he connected the other wire to a battery. Indeed, he saw a transient current (which he called a "wave of electricity") when he connected the wire to the battery, and another when he disconnected it. Within two months, Faraday had found several other manifestations of electromagnetic induction. For example, he saw transient currents when he quickly slid a bar magnet in and out of a coil of wires, and he generated a steady (DC) current by rotating a copper disk near a bar magnet with a sliding electrical lead ("Faraday's disk").

Faraday explained electromagnetic induction using a concept he called lines of force. However, scientists at the time widely rejected his theoretical ideas, mainly because they were not formulated mathematically. An exception was Maxwell, who used Faraday's ideas as the basis of his quantitative electromagnetic theory. In Maxwell's papers, the time varying aspect of electromagnetic induction is expressed as a differential equation which Oliver Heaviside referred to as Faraday's law even though it is slightly different in form from the original version of Faraday's law, and doesn't cater for motionally induced EMF. Heaviside's version is the form recognized today in the group of equations known as Maxwell's equations.

Lenz's law, formulated by Heinrich Lenz in 1834, describes "flux through the circuit", and gives the direction of the induced electromotive force and current resulting from electromagnetic induction (elaborated upon in the examples below).



Faraday's experiment showing induction between coils of wire: The liquid battery (right) provides a current which flows through the small coil (A), creating a magnetic field. When the coils are stationary, no current is induced. But when the small coil is moved in or out of the large coil (B), the magnetic flux through the large coil changes, inducing a current which is detected by the galvanometer (G).

Faraday's law as two different phenomena

Some physicists have remarked that Faraday's law is a single equation describing two different phenomena: The **motional EMF** generated by a magnetic force on a moving wire, and the **transformer EMF** generated by an electric force due to a changing magnetic field. James Clerk Maxwell drew attention to this fact in his 1861 paper On Physical Lines of Force. In the latter half of part II of that paper, Maxwell gives a separate physical explanation for each of the two phenomena. A reference to these two aspects of electromagnetic induction is made in some modern textbooks. As Richard Feynman states:

So the "flux rule" that the emf in a circuit is equal to the rate of change of the magnetic flux through the circuit applies whether the flux changes because the field changes or because the circuit moves (or both).... Yet in our explanation of the rule we have used two completely – distinct laws for the two cases $\mathbf{v}\times\mathbf{B}$ for "circuit moves" and $\nabla \times \mathbf{E} = -\partial_t \mathbf{B}$ for "field changes".
We know of no other place in physics where such a simple and accurate general principle requires for its real understanding an analysis in terms of two different phenomena.

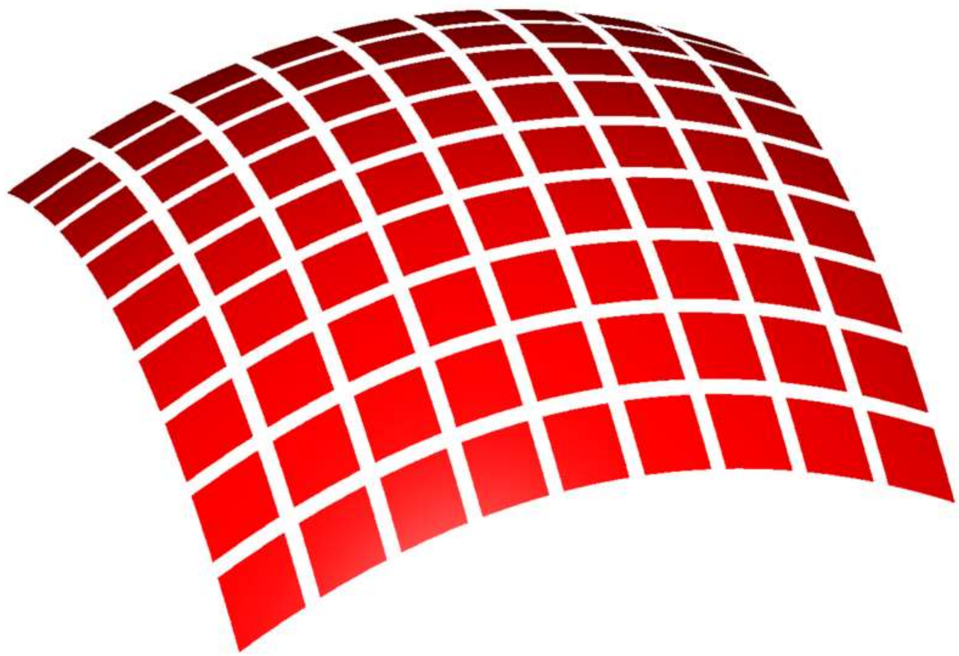
– **Richard P. Feynman** , The Feynman Lectures on Physics

Reflection on this apparent dichotomy was one of the principal paths that led Einstein to develop special relativity:

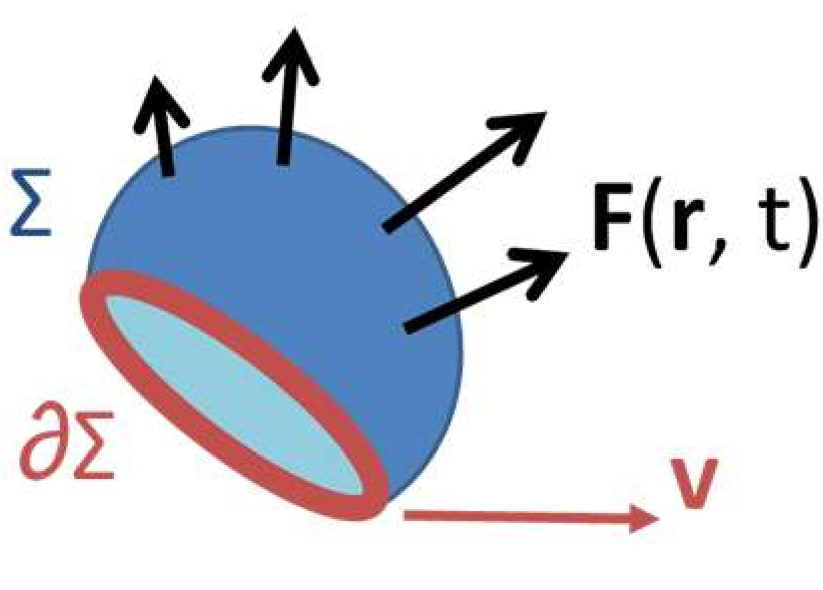
It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

– **Albert Einstein**, On the Electrodynamics of Moving Bodies

Flux through a surface and EMF around a loop



The definition of surface integral relies on splitting the surface Σ into small surface elements. Each element is associated with a vector d**A** of magnitude equal to the area of the element and with direction normal to the element and pointing outward.



e

A vector field **F**(**r**, t) defined throughout space, and a surface Σ bounded by curve ∂Σ moving with velocity **v** over which the field is integrated.

Faraday's law of induction makes use of the magnetic flux Φ_B through a surface Σ, defined by an integral over a surface:

$$\Phi_B = \iint_{\Sigma(t)} \mathbf{B}(\mathbf{r}, t) \cdot d\mathbf{A} \ ,$$

where d**A** is an element of surface area of the moving surface Σ(t), **B** is the magnetic field, and **B**·d**A** is a vector dot product. The surface is considered to have a "mouth" outlined by a closed curve denoted ∂Σ(t). When the flux changes, Faraday's law of induction says that the work $\mathcal{E}$ done (per unit charge) moving a test charge around the closed curve ∂Σ(t), called the electromotive force (EMF), is given by:

$$|\mathcal{E}| = \left| \frac{d\Phi_B}{dt} \right| \ ,$$

where $|\mathcal{E}|$ is the magnitude of the electromotive force (EMF) in volts and Φ_B is the magnetic flux in webers. The direction of the electromotive force is given by Lenz's law.

For a tightly-wound coil of wire, composed of N identical loops, each with the same Φ_B, Faraday's law of induction states that

$$|\mathcal{E}| = N \left| \frac{d\Phi_B}{dt} \right|$$

where N is the number of turns of wire and Φ_B is the magnetic flux in webers through a single loop.

In choosing a path ∂Σ(t) to find EMF, the path must satisfy the basic requirements that (i) it is a closed path, and (ii) the path must capture the relative motion of the parts of the circuit (the origin of the t-dependence in ∂Σ(t)). It is not a requirement that the path follow a line of current flow, but of course the EMF that is found using the flux law will be the EMF around the chosen path. If a current path is not followed, the EMF might not be the EMF driving the current.

Example: Spatially varying Magnetic field

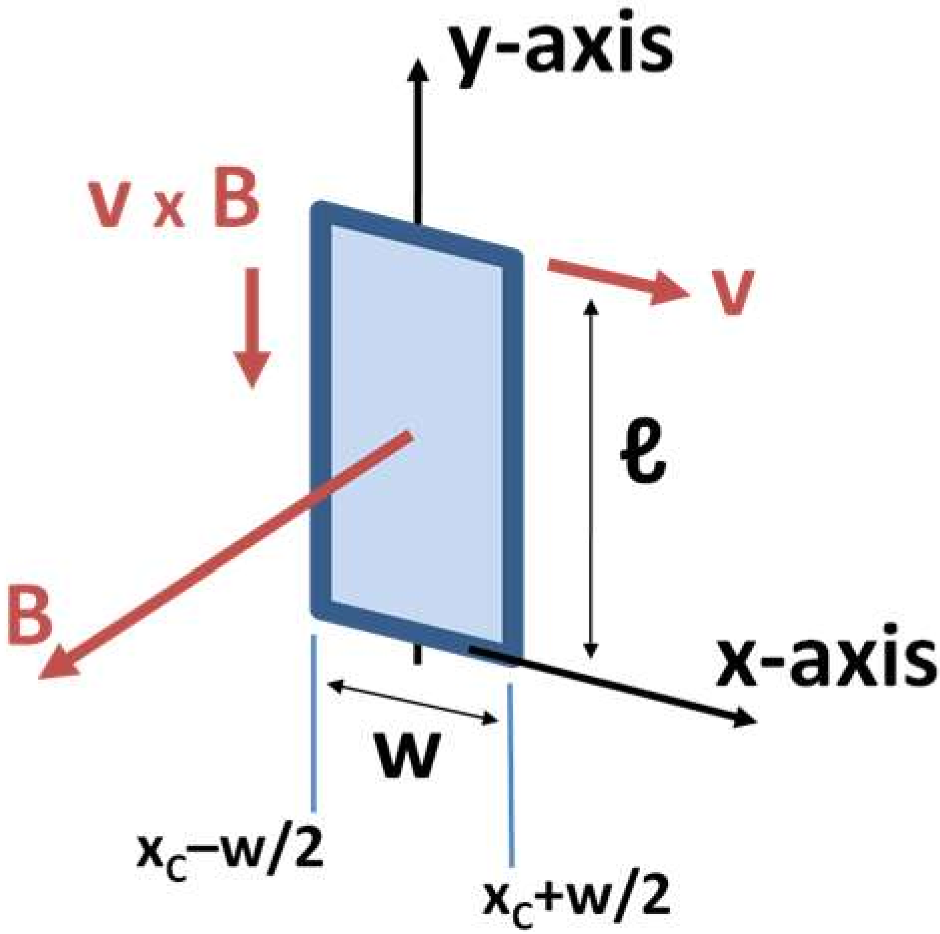


Figure 3: Closed rectangular wire loop moving along x-axis at velocity **v** in magnetic field **B** that varies with position x.

Consider the case in Figure 3 of a closed rectangular loop of wire in the xy-plane translated in the x-direction at velocity v. Thus, the center of the loop at x_C satisfies $v = dx_C / dt$. The loop has length ℓ in the y-direction and width w in the x-direction. A time-independent but spatially varying magnetic field B(x) points in the z-direction. The magnetic field on the left side is $B(x_C - w/2)$, and on the right side is $B(x_C + w/2)$. The electromotive force is to be found by using either the Lorentz force law or equivalently by using Faraday's induction law above.

Lorentz force law method

A charge q in the wire on the left side of the loop experiences a Lorentz force $q\, \mathbf{v} \times B\, \mathbf{k} = -q\, v\, B(x_C - w/2)\, \mathbf{j}$ (**j, k** unit vectors in the y- and z-directions), leading to an EMF (work per unit charge) of $v\, \ell\, B(x_C - w/2)$ along the length of the left side of the loop. On the right side of the loop the same argument shows the EMF to be $v\, \ell\, B(x_C + w/2)$.

The two EMF's oppose each other, both pushing positive charge toward the bottom of the loop. In the case where the **B**-field increases with increase in x, the force on the right side is largest, and the current will be clockwise: using the right-hand rule, the B-field generated by the current opposes the impressed field. The EMF driving the current must increase as we move counterclockwise (opposite to the current). Adding the EMF's in a counterclockwise tour of the loop we find

$$\mathcal{E} = v\ell[B(x_C + w/2) - B(x_C - w/2)] \ .$$

Faraday's law method

At any position of the loop the magnetic flux through the loop is

$$\Phi_B = \pm \int_0^\ell dy \int_{x_C - w/2}^{x_C + w/2} B(x) dx$$
$$= \pm \ell \int_{x_C - w/2}^{x_C + w/2} B(x) dx \ .$$

The sign choice is decided by whether the normal to the surface points in the same direction as **B**, or in the opposite direction. If we take the normal to the surface as pointing in the same direction as the B-field of the induced current, this sign is negative. The time derivative of the flux is then (using the chain rule of differentiation or the general form of Leibniz rule for differentiation of an integral):

$$\frac{d\Phi_B}{dt} = (-)\frac{d}{dx_C}\left[\int_0^\ell dy \ \int_{x_C - w/2}^{x_C + w/2} dx B(x)\right]\frac{dx_C}{dt} \ ,$$
$$= (-)v\ell[B(x_C + w/2) - B(x_C - w/2)] \ ,$$

(where $v = dx_C / dt$ is the rate of motion of the loop in the x-direction) leading to:

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = v\ell[B(x_C + w/2) - B(x_C - w/2)] \ ,$$

as before.

The equivalence of these two approaches is general and, depending on the example, one or the other method may prove more practical.

Example: Moving loop in uniform Magnetic field

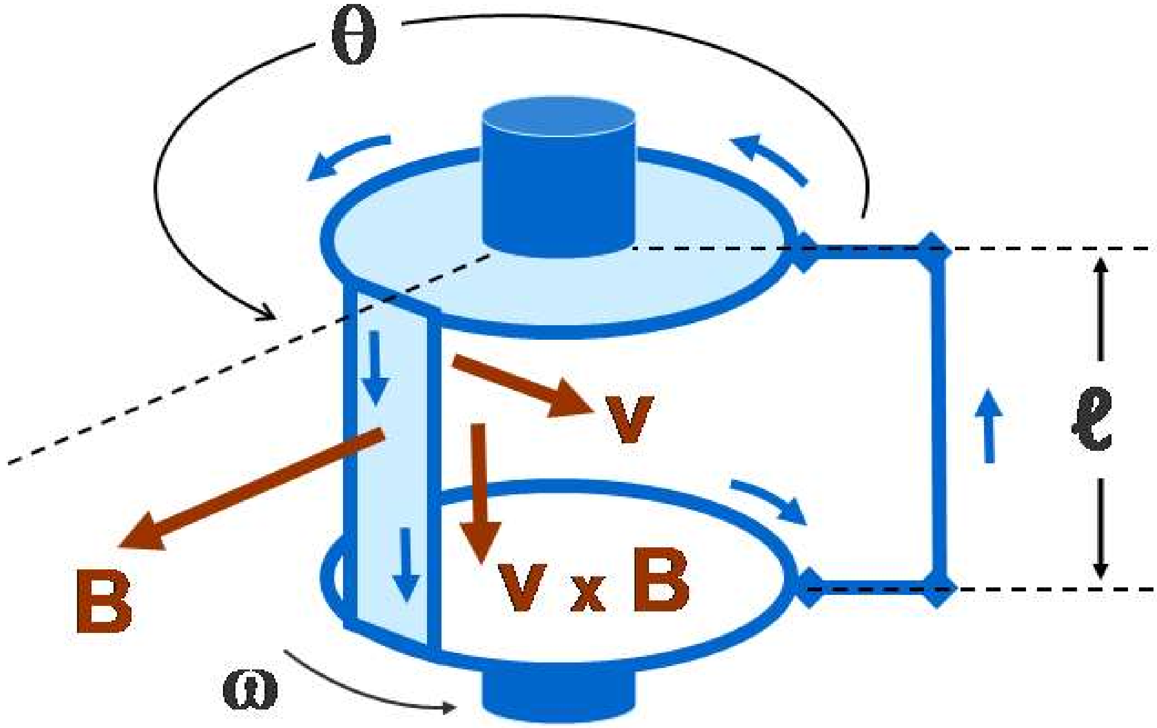


Figure 4: Rectangular wire loop rotating at angular velocity ω in radially outward pointing magnetic field **B** of fixed magnitude. Current is collected by brushes attached to top and bottom discs, which have conducting rims.

Figure 4 shows a spindle formed of two discs with conducting rims and a conducting loop attached vertically between these rims. The entire assembly spins in a magnetic field that points radially outward, but is the same magnitude regardless of its direction. A radially oriented collecting return loop picks up current from the conducting rims. At the location of the collecting return loop, the radial **B**-field lies in the plane of the collecting loop, so the collecting loop contributes no flux to the circuit. The electromotive force is to be found directly and by using Faraday's law above.

Lorentz force law method

In this case the Lorentz force drives the current in the two vertical arms of the moving loop downward, so current flows from the top disc to the bottom disc. In the conducting rims of the discs, the Lorentz force is perpendicular to the rim, so no EMF is generated in the rims, nor in the horizontal portions of the moving loop. Current is transmitted from the bottom rim to the top rim through the external return loop, which is oriented so the **B**-field is in its plane. Thus, the Lorentz force in the return loop is perpendicular to the loop, and no EMF is generated in this return loop. Traversing the current path in the direction opposite to the current flow, work is done against the Lorentz force only in the vertical arms of the moving loop, where

$$F = qBv .$$

where v= velocity of moving charge

Consequently, the EMF is

$$\mathcal{E} = Bv\ell = Br\ell\omega ,$$

where v = velocity of conductor or magnet and l = vertical length of the loop. In this case the velocity is related to the angular rate of rotation by v = r ω, with r = radius of cylinder. Notice that the same work is done on any path that rotates with the loop and connects the upper and lower rim.

Faraday's law method

An intuitively appealing but mistaken approach to using the flux rule would say the flux through the circuit was just $\Phi_B = B\ w\ \ell$, where w = width of the moving loop. This number is time-independent, so the approach predicts incorrectly that no EMF is generated. The flaw in this argument is that it fails to consider the entire current path, which is a closed loop.

To use the flux rule, we have to look at the entire current path, which includes the path through the rims in the top and bottom discs. We can choose an arbitrary closed path through the rims and the rotating loop, and the flux law will find the EMF around the chosen path. Any path that has a segment attached to the rotating loop captures the relative motion of the parts of the circuit.

As an example path, let's traverse the circuit in the direction of rotation in the top disc, and in the direction opposite to the direction of rotation in the bottom disc (shown by arrows in Figure 4). In this case, for the moving loop at an angle θ from the collecting loop, a portion of the cylinder of area A = r ℓ θ is part of the circuit. This area is perpendicular to the **B**-field, and so contributes to the flux an amount:

$$\Phi_B = -Br\theta\ell ,$$

where the sign is negative because the right-hand rule suggests the **B**-field generated by the current loop is opposite in direction to the applied **B** field. As this is the only time-dependent portion of the flux, the flux law predicts an EMF of

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = Br\ell\frac{d\theta}{dt} = Br\ell\omega ,$$

in agreement with the Lorentz force law calculation.

Now let's try a different path. Follow a path traversing the rims via the opposite choice of segments. Then the coupled flux would decrease as θ increased, but the right-hand rule would suggest the current loop added to the applied **B**-field, so the EMF around this path is the same as for the first path. Any mixture of return paths leads to the same result for EMF, so it is actually immaterial which path is followed.

Direct evaluation of the change in flux

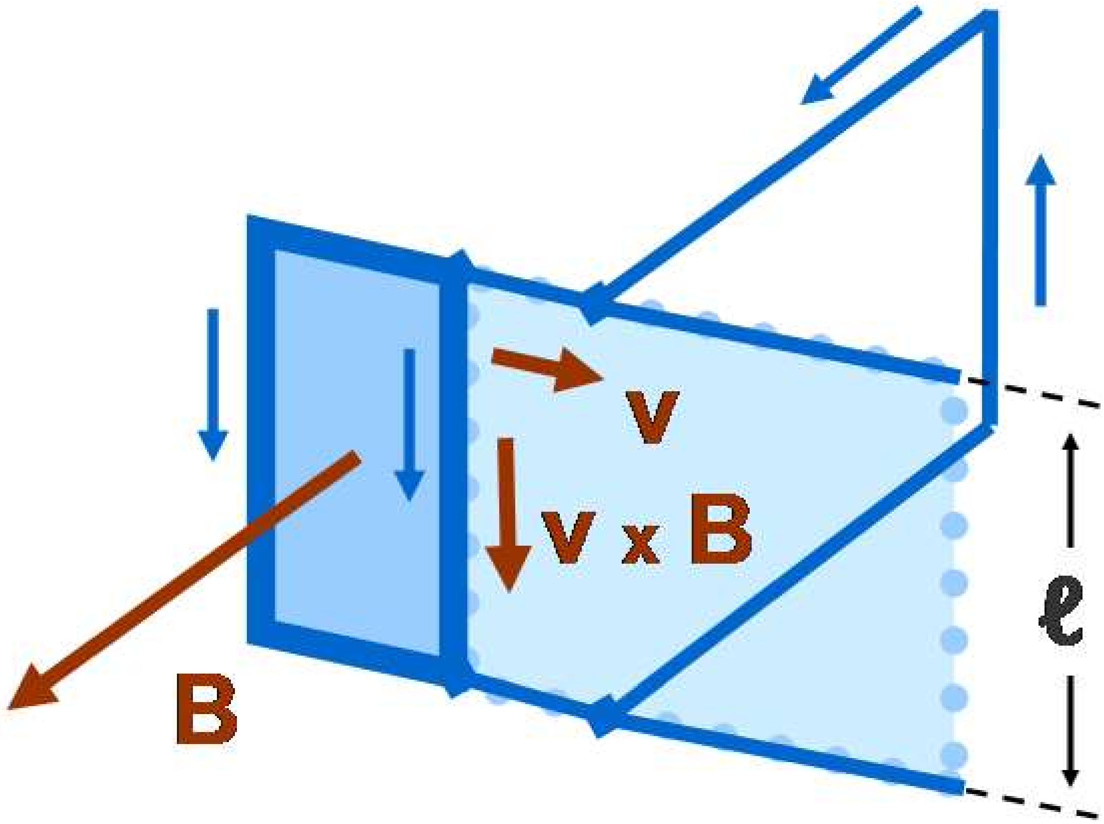


Figure 5: A simplified version of Figure 4. The loop slides with velocity **v** in a stationary, homogeneous **B**-field.

The use of a closed path to find EMF as done above appears to depend upon details of the path geometry. In contrast, the Lorentz-law approach is independent of such restrictions. The following discussion is intended to provide a better understanding of the equivalence of paths and escape the particulars of path selection when using the flux law.

Figure 5 is an idealization of Figure 4 with the cylinder unwrapped onto a plane. The same path-related analysis works, but a simplification is suggested. The time-independent aspects of the circuit cannot affect the time-rate-of-change of flux. For example, at a constant velocity of sliding the loop, the details of current flow through the loop are not time dependent. Instead of concern over details of the closed loop selected to find the EMF, one can focus on the area of **B**-field swept out by the moving loop. This suggestion amounts to finding the rate at which flux is cut by the circuit. That notion provides direct evaluation of the rate of change of flux, without concern over the time-independent details of various path choices around the circuit. Just as with the Lorentz law approach, it is clear that any two paths attached to the sliding loop, but differing in how they cross the loop, produce the same rate-of-change of flux.

In Figure 5 the area swept out in unit time is simply dA / dt = v ℓ, regardless of the details of the selected closed path, so Faraday's law of induction provides the EMF as:

$$\mathcal{E} = \frac{d\Phi_B}{dt} = Bv\ell \ .$$

This path independence of EMF shows that if the sliding loop is replaced by a solid conducting plate, or even some complex warped surface, the analysis is the same: find the flux in the area swept out by the moving portion of the circuit. In a similar way, if the sliding loop in the drum generator of Figure 4 is replaced by a 360° solid conducting cylinder, the swept area calculation is exactly the same as for the case with only a loop. That is, the EMF predicted by Faraday's law is exactly the same for the case with a cylinder with solid conducting walls or, for that matter, a cylinder with a cheese grater for walls. Notice, though, that the current that flows as a result of this EMF will not be the same because the resistance of the circuit determines the current.

The Maxwell-Faraday equation

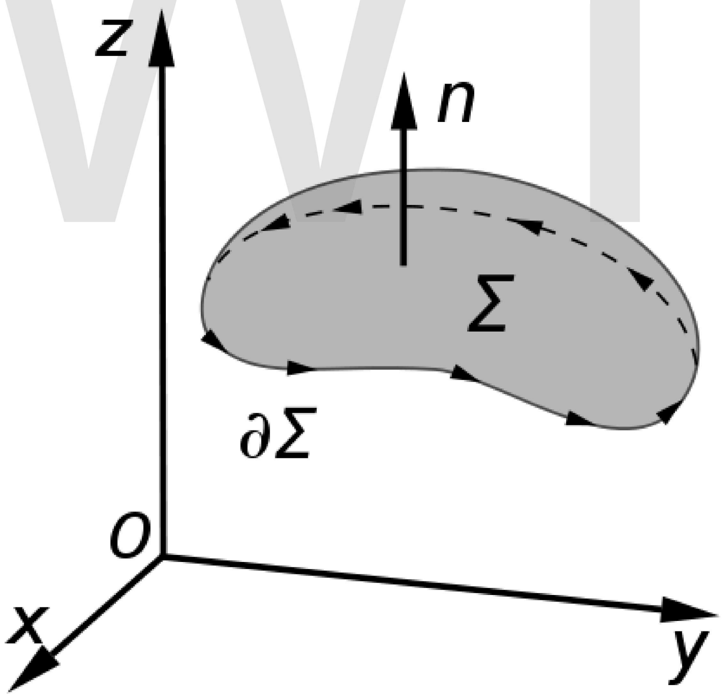


Figure 6: An illustration of Kelvin-Stokes theorem with surface **Σ** its boundary ***∂*Σ** and orientation **n** set by the right-hand rule.

A changing magnetic field creates an electric field; this phenomenon is described by the Maxwell-Faraday equation:

$$\nabla \times \mathbf{E}(\mathbf{r},\ t) = -\frac{\partial \mathbf{B}(\mathbf{r},\ t)}{\partial t}$$

where:

$\nabla \times$denotes curl
E is the electric field
B is the magnetic field

This equation appears in modern sets of Maxwell's equations and is often referred to as Faraday's law. However, because it contains only partial time derivatives, its application is restricted to situations where the test charge is stationary in a time varying magnetic field. It does not account for electromagnetic induction in situations where a charged particle is moving in a magnetic field.

It also can be written in an **integral form** by the Kelvin-Stokes theorem:

$$\oint_{\partial \Sigma} \mathbf{E} \cdot d\boldsymbol{\ell} = - \iint_{\Sigma} \frac{\partial}{\partial t} \mathbf{B} \cdot d\mathbf{A}$$
$$= - \frac{\partial}{\partial t} \iint_{\Sigma} \mathbf{B} \cdot d\mathbf{A}$$

where the movement of the derivative before the integration requires a time-independent surface **Σ** (considered in this context to be part of the interpretation of the partial derivative), and as indicated in Figure 6:

Σ is a surface bounded by the closed contour **∂Σ**; both **Σ** and **∂Σ** are fixed, independent of time
E is the electric field,
d**ℓ** is an infinitesimal vector element of the contour **∂Σ**,
B is the magnetic field.
d**A** is an infinitesimal vector element of surface **Σ** , whose magnitude is the area of an infinitesimal patch of surface, and whose direction is orthogonal to that surface patch.

Both d**ℓ** and d**A** have a sign ambiguity; to get the correct sign, the right-hand rule is used, as explained in the Kelvin-Stokes theorem. For a planar surface Σ, a positive path element d**ℓ** of curve ∂Σ is defined by the right-hand rule as one that points with the fingers of the right hand when the thumb points in the direction of the normal **n** to the surface Σ.

The integral around $\partial\boldsymbol{\Sigma}$ is called a path integral or line integral. The surface integral at the right-hand side of the Maxwell-Faraday equation is the explicit expression for the magnetic flux Φ_B through $\boldsymbol{\Sigma}$. Notice that a nonzero path integral for **E** is different from the behavior of the electric field generated by charges. A charge-generated **E**-field can be expressed as the gradient of a scalar field that is a solution to Poisson's equation, and has a zero path integral.

The integral equation is true for any path $\partial\boldsymbol{\Sigma}$ through space, and any surface $\boldsymbol{\Sigma}$ for which that path is a boundary. Note, however, that $\partial\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}$ are understood not to vary in time in this formula. This integral form cannot treat **motional** EMF because $\boldsymbol{\Sigma}$ is time-independent. Notice as well that this equation makes no reference to EMF $\mathcal{E}$, and indeed cannot do so without introduction of the Lorentz force law to enable a calculation of work.

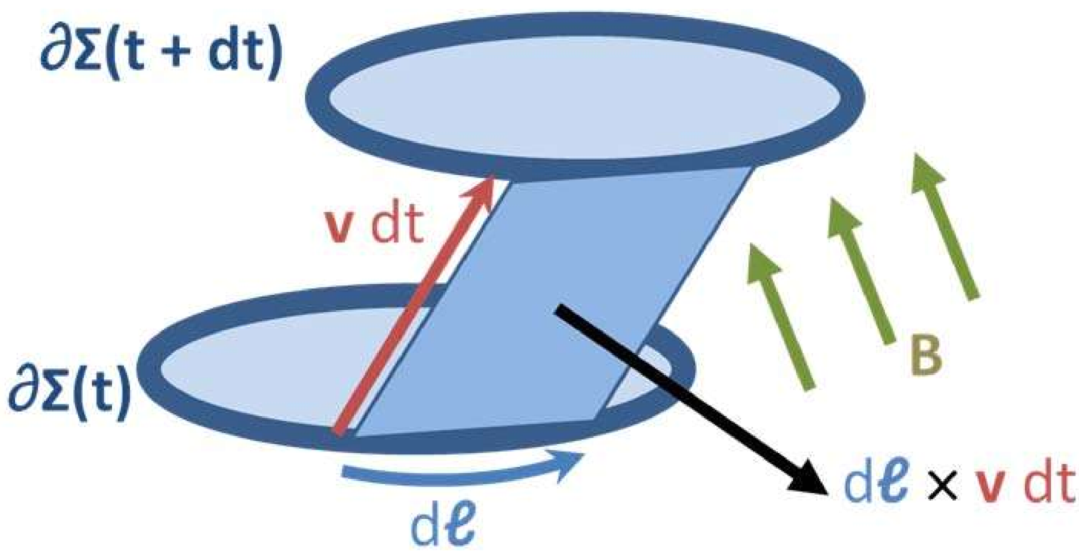


Figure 7: Area swept out by vector element d**ℓ** of curve $\partial\boldsymbol{\Sigma}$ in time dt when moving with velocity **v**.

Using the complete Lorentz force to calculate the EMF,

$$\mathcal{E} = \oint_{\partial\Sigma(t)} (\mathbf{E}(\mathbf{r},\ t) + \mathbf{v} \times \mathbf{B}(\mathbf{r},\ t)) \cdot d\boldsymbol{\ell}\ ,$$

a statement of Faraday's law of induction more general than the integral form of the Maxwell-Faraday equation is:

$$\oint_{\partial\Sigma(t)} (\mathbf{E}(\mathbf{r},\ t) + \mathbf{v} \times \mathbf{B}(\mathbf{r},\ t)) \cdot d\boldsymbol{\ell} = -\frac{d}{dt} \iint_{\Sigma(t)} d\mathbf{A} \cdot \mathbf{B}(\mathbf{r},\ t)\ ,$$

where $\partial\mathbf{\Sigma}(t)$ is the moving closed path bounding the moving surface $\mathbf{\Sigma}(t)$, and **v** is the velocity of movement. See Figure 2. Notice that the ordinary time derivative is used, not a partial time derivative, implying the time variation of $\mathbf{\Sigma}(t)$ must be included in the differentiation. In the integrand the element of the curve $d\boldsymbol{\ell}$ moves with velocity **v**.

Figure 7 provides an interpretation of the magnetic force contribution to the EMF on the left side of the above equation. The area swept out by segment $d\boldsymbol{\ell}$ of curve $\partial\mathbf{\Sigma}$ in time dt when moving with velocity **v** is:

$$d\mathbf{A} = -d\boldsymbol{\ell} \times \boldsymbol{v}dt \ ,$$

so the change in magnetic flux $\Delta\Phi_B$ through the portion of the surface enclosed by $\partial\mathbf{\Sigma}$ in time dt is:

$$\frac{d\Delta\Phi_B}{dt} = -\mathbf{B} \cdot \ d\boldsymbol{\ell} \times \boldsymbol{v} \ = -\mathbf{v} \times \mathbf{B} \cdot \ d\boldsymbol{\ell} \ ,$$

and if we add these $\Delta\Phi_B$-contributions around the loop for all segments $d\boldsymbol{\ell}$, we obtain the magnetic force contribution to Faraday's law. That is, this term is related to motional EMF.

Example: viewpoint of a moving observer

Revisiting the example of Figure 3 in a moving frame of reference brings out the close connection between E- and B-fields, and between motional and induced EMF's. Imagine an observer of the loop moving with the loop. The observer calculates the EMF around the loop using both the Lorentz force law and Faraday's law of induction. Because this observer moves with the loop, the observer sees no movement of the loop, and zero $\mathbf{v} \times \mathbf{B}$. However, because the B-field varies with position x, the moving observer sees a time-varying magnetic field, namely:

$$\mathbf{B} = \mathbf{k}B(x+vt) \ ,$$

where **k** is a unit vector pointing in the z-direction.

Lorentz force law version

The Maxwell-Faraday equation says the moving observer sees an electric field E_y in the y-direction given by:

$$\nabla \times \mathbf{E} = \mathbf{k} \ \frac{dE_y}{dx}$$
$$= -\frac{\partial \mathbf{B}}{\partial t} = -\mathbf{k}\frac{dB(x+vt)}{dt} = -\mathbf{k}\frac{dB}{dx}v \ \ ,$$

Here the chain rule is used:

$$\frac{dB}{dt} = \frac{dB}{d(x+vt)}\frac{d(x+vt)}{dt} = \frac{dB}{dx}v \ .$$

Solving for E_y, to within a constant that contributes nothing to an integral around the loop,

$$E_y(x,\ t) = -B(x+vt)\ v \ .$$

Using the Lorentz force law, which has only an electric field component, the observer finds the EMF around the loop at a time t to be:

$$\mathcal{E} = -\ell[E_y(x_C + w/2,\ t) - E_y(x_C - w/2,\ t)]$$
$$= v\ell[B(x_C + w/2 + vt) - B(x_C - w/2 + vt)] \ ,$$

which is exactly the same result found by the stationary observer, who sees the centroid x_C has advanced to a position $x_C + v\ t$. However, the moving observer obtained the result under the impression that the Lorentz force had only an **electric** component, while the stationary observer thought the force had only a **magnetic** component.

Faraday's law of induction

Using Faraday's law of induction, the observer moving with x_C sees a changing magnetic flux, but the loop does not appear to move: the center of the loop x_C is fixed because the moving observer is moving with the loop. The flux is then:

$$\Phi_B = -\int_0^\ell dy \int_{x_C - w/2}^{x_C + w/2} B(x+vt)dx \ ,$$

where the minus sign comes from the normal to the surface pointing oppositely to the applied B-field. The EMF from Faraday's law of induction is now:

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = \int_0^\ell dy \int_{x_C - w/2}^{x_C + w/2} \frac{d}{dt}B(x+vt)dx$$
$$= \int_0^\ell dy \int_{x_C - w/2}^{x_C + w/2} \frac{d}{dx}B(x+vt)\ v\ dx$$
$$= v\ell\ [B(x_C + w/2 + vt) - B(x_C - w/2 + vt)] \ ,$$

the same result. The time derivative passes through the integration because the limits of integration have no time dependence. Again, the chain rule was used to convert the time derivative to an x-derivative.

The stationary observer thought the EMF was a **motional** EMF, while the moving observer thought it was an **induced** EMF.

Electrical generator

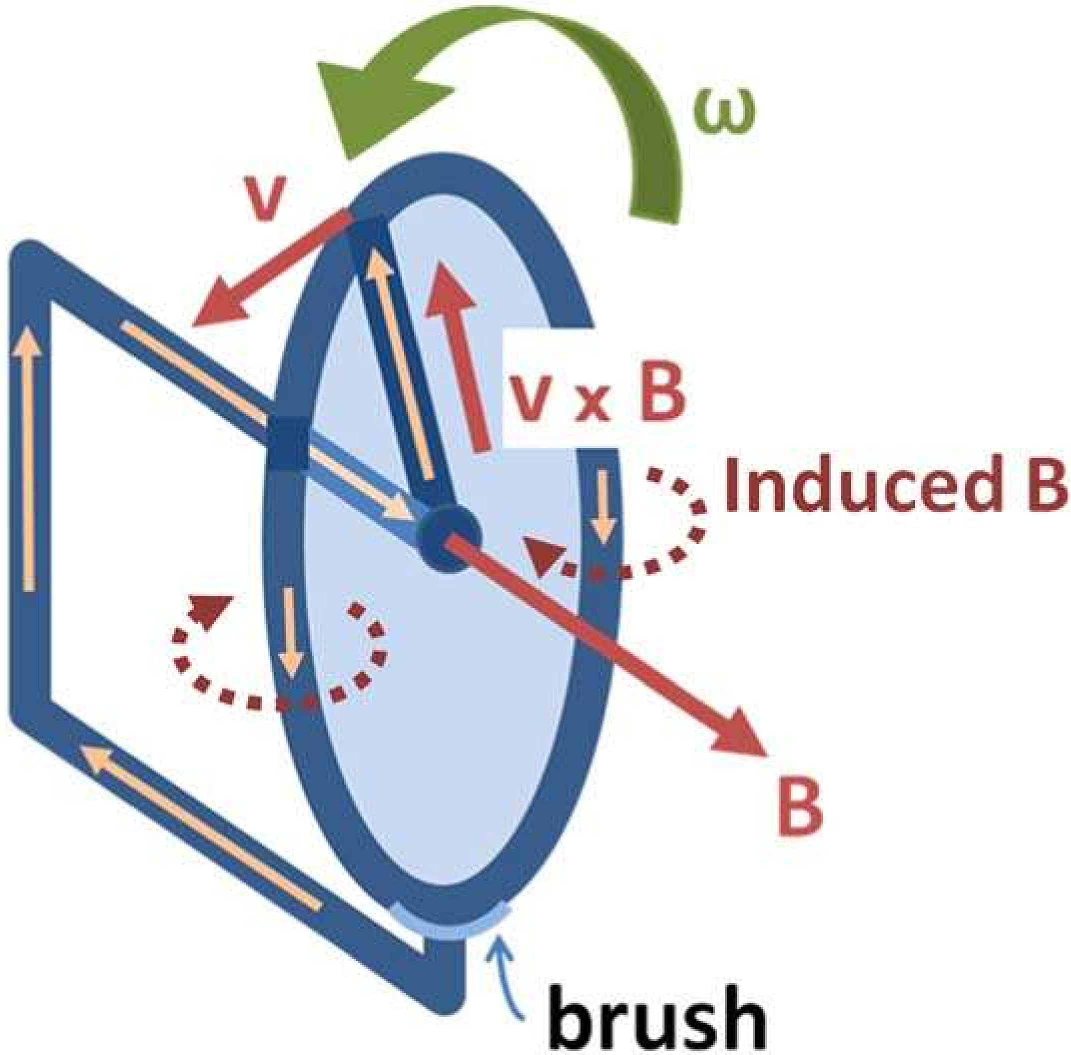


Figure 8: Faraday's disc electric generator. The disc rotates with angular rate ω, sweeping the conducting radius circularly in the static magnetic field **B**. The magnetic Lorentz force **v** × **B** drives the current along the conducting radius to the conducting rim, and from there the circuit completes through the lower brush and the axle supporting the disc. Thus, current is generated from mechanical motion.

The EMF generated by Faraday's law of induction due to relative movement of a circuit and a magnetic field is the phenomenon underlying electrical generators. When a permanent magnet is moved relative to a conductor, or vice versa, an electromotive force is created. If the wire is connected through an electrical load, current will flow, and thus electrical energy is generated, converting the mechanical energy of motion to electrical energy. For example, the drum generator is based upon Figure 4. A different implementation of this idea is the Faraday's disc, shown in simplified form in Figure 8. Note that either the analysis of Figure 5, or direct application of the Lorentz force law, shows that a solid conducting disc works the same way.

In the Faraday's disc example, the disc is rotated in a uniform magnetic field perpendicular to the disc, causing a current to flow in the radial arm due to the Lorentz force. It is interesting to understand how it arises that mechanical work is necessary to drive this current. When the generated current flows through the conducting rim, a magnetic field is generated by this current through Ampere's circuital law (labeled "induced B" in Figure 8). The rim thus becomes an electromagnet that resists rotation of the disc (an example of Lenz's law). On the far side of the figure, the return current flows from the rotating arm through the far side of the rim to the bottom brush. The B-field induced by this return current opposes the applied B-field, tending to decrease the flux through that side of the circuit, opposing the increase in flux due to rotation. On the near side of the figure, the return current flows from the rotating arm through the near side of the rim to the bottom brush. The induced B-field increases the flux on this side of the circuit, opposing the decrease in flux due to rotation. Thus, both sides of the circuit generate an emf opposing the rotation. The energy required to keep the disc moving, despite this reactive force, is exactly equal to the electrical energy generated (plus energy wasted due to friction, Joule heating, and other inefficiencies). This behavior is common to all generators converting mechanical energy to electrical energy.

Although Faraday's law always describes the working of electrical generators, the detailed mechanism can differ in different cases. When the magnet is rotated around a stationary conductor, the changing magnetic field creates an electric field, as described by the Maxwell-Faraday equation, and that electric field pushes the charges through the wire. This case is called an **induced** EMF. On the other hand, when the magnet is stationary and the conductor is rotated, the moving charges experience a magnetic force (as described by the Lorentz force law), and this magnetic force pushes the charges through the wire. This case is called **motional** EMF.

Electrical motor

An **electric motor** converts electrical energy into mechanical energy. Most electric motors operate through interacting magnetic fields and current-carrying conductors to generate force, although a few use electrostatic forces. The reverse process, producing electrical energy from mechanical energy, is accomplished by some type of generator such as an alternator or a dynamo. Many types of electric motors can be run as generators, and vice versa. For example a starter/generator for a gas turbine, or traction motors used on vehicles, often perform both tasks.

Electric motors are found in applications as diverse as industrial fans, blowers and pumps, machine tools, household appliances, power tools, and disk drives. They may be powered by direct current (e.g., a battery powered portable device or motor vehicle), or by alternating current from a central electrical distribution grid. The smallest motors may be found in electric wristwatches. Medium-size motors of highly standardized dimensions and characteristics provide convenient mechanical power for industrial uses. The very largest electric motors are used for propulsion of large ships, and for such purposes as pipeline compressors, with ratings in the millions of watts. Electric motors may be classified by the source of electric power, by their internal construction, by their application, or by the type of motion they give.

The physical principle of production of mechanical force by the interactions of an electric current and a magnetic field was known as early as 1821. Electric motors of increasing efficiency were constructed throughout the 19th century, but commercial exploitation of electric motors on a large scale required efficient electrical generators and electrical distribution networks.

Some devices, such as magnetic solenoids and loudspeakers, although they generate some mechanical power, are not generally referred to as electric motors, and are usually termed actuators and transducers, respectively.

Electric motors

Categorization of electric motors

The classic division of electric motors has been that of Alternating Current (AC) types vs Direct Current (DC) types. This is more a de facto convention, rather than a rigid distinction. For example, many classic DC motors run on AC power, these motors being referred to as universal motors.

Rated output power is also used to categorise motors, those of less than 746 Watts, for example, are often referred to as fractional horsepower motors (FHP) in reference to the old imperial measurement.

The ongoing trend toward electronic control further muddles the distinction, as modern drivers have moved the commutator out of the motor shell. For this new breed of motor, driver circuits are relied upon to generate sinusoidal AC drive currents, or some approximation thereof. The two best examples are: the brushless DC motor and the stepping motor, both being poly-phase AC motors requiring external electronic control, although historically, stepping motors (such as for maritime and naval gyrocompass repeaters) were driven from DC switched by contacts.

Considering all rotating (or linear) electric motors require synchronism between a moving magnetic field and a moving current sheet for average torque production, there is a clearer distinction between an asynchronous motor and synchronous types. An asynchronous motor requires slip between the moving magnetic field and a winding set to induce current in the winding set by mutual inductance; the most ubiquitous example being the common AC induction motor which must slip to generate torque. In the synchronous types, induction (or slip) is not a requisite for magnetic field or current production (e.g. permanent magnet motors, synchronous brush-less wound-rotor doubly-fed electric machine).

DC motors

A DC motor is designed to run on DC electric power. Two examples of pure DC designs are Michael Faraday's homopolar motor (which is uncommon), and the ball bearing motor, which is (so far) a novelty. By far the most common DC motor types are the brushed and brushless types, which use internal and external commutation respectively to periodicaly reverse the current in the rotor windings.

Permanent-magnet motors

A permanent-magnet motor does not have a field winding on the stator frame, instead relying on permanent magnets to provide the magnetic field against which the rotor field interacts to produce torque. Compensating windings in series with the armature may be used on large motors to improve commuation under load. Because this field is fixed, it cannot be adjusted for speed control. Permanent-magent motors are convenient in miniature motors to eliminate the power consumption of the field winding. Most larger

DC motors are of the "dynamo" type, which requires current to flow in field windings to provide the stator magnetic field.

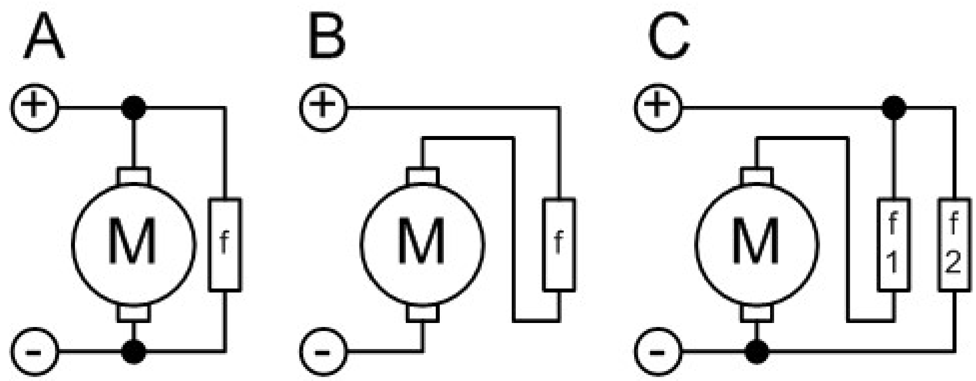
To minimize overall weight and size, miniature permanent-magnet motors may use high energy magnets made with neodymium or other strategic elements.

Brushed DC motors

DC motor design generates an oscillating current in a wound rotor, or armature, with a split ring commutator, and either a wound or permanent magnet stator. A rotor consists of one or more coils of wire wound around a core on a shaft; an electrical power source is connected to the rotor coil through the commutator and its brushes, causing current to flow in it, producing electromagnetism. The commutator causes the current in the coils to be switched as the rotor turns, keeping the magnetic poles of the rotor from ever fully aligning with the magnetic poles of the stator field, so that the rotor never stops (like a compass needle does) but rather keeps rotating indefinitely (as long as power is applied and is sufficient for the motor to overcome the shaft torque load and internal losses due to friction, etc.)

Many of the limitations of the classic commutator DC motor are due to the need for brushes to press against the commutator. This creates friction. Sparks are created by the brushes making and breaking circuits through the rotor coils as the brushes cross the insulating gaps between commutator sections. Depending on the commutator design, this may include the brushes shorting together adjacent sections—and hence coil ends—momentarily while crossing the gaps. Furthermore, the inductance of the rotor coils causes the voltage across each to rise when its circuit is opened, increasing the sparking of the brushes. This sparking limits the maximum speed of the machine, as too-rapid sparking will overheat, erode, or even melt the commutator. The current density per unit area of the brushes, in combination with their resistivity, limits the output of the motor. The making and breaking of electric contact also causes electrical noise, and the sparks additionally cause RFI. Brushes eventually wear out and require replacement, and the commutator itself is subject to wear and maintenance (on larger motors) or replacement (on small motors). The commutator assembly on a large motor is a costly element, requiring precision assembly of many parts. On small motors, the commutator is usually permanently integrated into the rotor, so replacing it usually requires replacing the whole rotor.

Large brushes are desired for a larger brush contact area to maximize motor output, but small brushes are desired for low mass to maximize the speed at which the motor can run without the brushes excessively bouncing and sparking (comparable to the problem of "valve float" in internal combustion engines). (Small brushes are also desirable for lower cost.) Stiffer brush springs can also be used to make brushes of a given mass work at a higher speed, but at the cost of greater friction losses (lower efficiency) and accelerated brush and commutator wear. Therefore, DC motor brush design entails a trade-off between output power, speed, and efficiency/wear.



A: shunt
B: series
C: compound
f = field coil

There are five types of brushed DC motor:

A. DC shunt-wound motor

B. DC series-wound motor

C. DC compound motor (two configurations):

- Cumulative compound
- Differentially compounded

D. Permanent magnet DC motor (not shown)

E. Separately excited (sepex) (not shown).

Brushless DC motors

Some of the problems of the brushed DC motor are eliminated in the brushless design. In this motor, the mechanical "rotating switch" or commutator/brushgear assembly is replaced by an external electronic switch synchronised to the rotor's position. Brushless motors are typically 85–90% efficient or more (higher efficiency for a brushless electric motor of up to 96.5% were reported by researchers at the Tokai University in Japan in 2009), whereas DC motors with brushgear are typically 75–80% efficient.

Midway between ordinary DC motors and stepper motors lies the realm of the brushless DC motor. Built in a fashion very similar to stepper motors, these often use a permanent magnet external rotor, three phases of driving coils, one or more Hall effect sensors to sense the position of the rotor, and the associated drive electronics. The coils are activated, one phase after the other, by the drive electronics as cued by the signals from either Hall effect sensors or from the back EMF (electromotive force) of the undriven coils. In effect, they act as three-phase synchronous motors containing their own variable-frequency drive electronics. A specialized class of brushless DC motor controllers utilize EMF feedback through the main phase connections instead of Hall effect sensors to determine position and velocity. These motors are used extensively in electric radio-controlled vehicles. When configured with the magnets on the outside, these are referred to by modellers as outrunner motors.

Brushless DC motors are commonly used where precise speed control is necessary, as in computer disk drives or in video cassette recorders, the spindles within CD, CD-ROM (etc.) drives, and mechanisms within office products such as fans, laser printers and photocopiers. They have several advantages over conventional motors:

- Compared to AC fans using shaded-pole motors, they are very efficient, running much cooler than the equivalent AC motors. This cool operation leads to much-improved life of the fan's bearings.
- Without a commutator to wear out, the life of a DC brushless motor can be significantly longer compared to a DC motor using brushes and a commutator. Commutation also tends to cause a great deal of electrical and RF noise; without a commutator or brushes, a brushless motor may be used in electrically sensitive devices like audio equipment or computers.
- The same Hall effect sensors that provide the commutation can also provide a convenient tachometer signal for closed-loop control (servo-controlled) applications. In fans, the tachometer signal can be used to derive a "fan OK" signal.
- The motor can be easily synchronized to an internal or external clock, leading to precise speed control.
- Brushless motors have no chance of sparking, unlike brushed motors, making them better suited to environments with volatile chemicals and fuels. Also, sparking generates ozone which can accumulate in poorly ventilated buildings risking harm to occupants' health.
- Brushless motors are usually used in small equipment such as computers and are generally used to get rid of unwanted heat.
- They are also very quiet motors which is an advantage if being used in equipment that is affected by vibrations.

Modern DC brushless motors range in power from a fraction of a watt to many kilowatts. Larger brushless motors up to about 100 kW rating are used in electric vehicles. They also find significant use in high-performance electric model aircraft.

Coreless or ironless DC motors

Nothing in the principle of any of the motors described above requires that the iron (steel) portions of the rotor actually rotate. If the soft magnetic material of the rotor is made in the form of a cylinder, then (except for the effect of hysteresis) torque is exerted only on the windings of the electromagnets. Taking advantage of this fact is the **coreless or ironless DC motor**, a specialized form of a brush or brushless DC motor. Optimized for rapid acceleration, these motors have a rotor that is constructed without any iron core. The rotor can take the form of a winding-filled cylinder, or a self-supporting structure comprising only the magnet wire and the bonding material. The rotor can fit inside the stator magnets; a magnetically soft stationary cylinder inside the rotor provides a return path for the stator magnetic flux. A second arrangement has the rotor winding basket surrounding the stator magnets. In that design, the rotor fits inside a magnetically soft cylinder that can serve as the housing for the motor, and likewise provides a return path for the flux.

Because the rotor is much lighter in weight (mass) than a conventional rotor formed from copper windings on steel laminations, the rotor can accelerate much more rapidly, often achieving a mechanical time constant under 1 ms. This is especially true if the windings use aluminum rather than the heavier copper. But because there is no metal mass in the rotor to act as a heat sink, even small coreless motors must often be cooled by forced air.

Related limited-travel actuators have no core and a bonded coil placed between the poles of high-flux thin permanent magnets. These are the fast head positioners for rigid-disk ("hard disk") drives.

Printed armature or pancake DC motors

A rather unusual motor design the pancake/printed armature motor has the windings shaped as a disc running between arrays of high-flux magnets, arranged in a circle, facing the rotor and forming an axial air gap. This design is commonly known the pancake motor because of its extremely flat profile, although the technology has had many brand names since its inception, such as ServoDisc.

The printed armature (originally formed on a printed circuit board) in a printed armature motor is made from punched copper sheets that are laminated together using advanced composites to form a thin rigid disc. The printed armature has a unique construction, in the brushed motor world, in that it does not have a separate ring commutator. The brushes run directly on the armature surface making the whole design very compact.

An alternative manufacturing method is to use wound copper wire laid flat with a central conventional commutator, in a flower and petal shape. The windings are typically stabilized by being impregnated with electrical epoxy potting systems. These are filled epoxies that have moderate mixed viscosity and a long gel time. They are highlighted by low shrinkage and low exotherm, and are typically UL 1446 recognized as a potting compound for use up to 180°C (Class H) (UL File No. E 210549).

The unique advantage of ironless DC motors is that there is no cogging (vibration caused by attraction between the iron and the magnets) and parasitic eddy currents cannot form in the rotor as it is totally ironless. This can greatly improve efficiency, but variable-speed controllers must use a higher switching rate (>40 kHz) or direct current because of the decreased electromagnetic induction.

These motors were originally invented to drive the capstan(s) of magnetic tape drives, in the burgeoning computer industry. Pancake motors are still widely used in high-performance servo-controlled systems, humanoid robotic systems, industrial automation and medical devices. Due to the variety of constructions now available the technology is used in applications from high temperature military to low cost pump and basic servo applications.

Universal motors

A series-wound motor is referred to as a **universal motor** when it has been designed to operate on either AC or DC power. The ability to operate on AC is because the current in both the field and the armature (and hence the resultant magnetic fields) will alternate (reverse polarity) in synchronism, and hence the resulting mechanical force will occur in a constant direction.

Operating at normal power line frequencies, universal motors are often found in a range rarely larger than one kilowatt (about 1.3 horsepower). Universal motors also form the basis of the traditional railway traction motor in electric railways. In this application, the use of AC to power a motor originally designed to run on DC would lead to efficiency losses due to eddy current heating of their magnetic components, particularly the motor field pole-pieces that, for DC, would have used solid (un-laminated) iron. Although the heating effects are reduced by using laminated pole-pieces, as used for the cores of transformers and by the use of laminations of high permeability electrical steel, one solution available at start of the 20th century was for the motors to be operated from very low frequency AC supplies, with 25 and 16.7 hertz (Hz) operation being common. Because they used universal motors, locomotives using this design were also commonly capable of operating from a third rail powered by DC.

An advantage of the universal motor is that AC supplies may be used on motors which have some characteristics more common in DC motors, specifically high starting torque and very compact design if high running speeds are used. The negative aspect is the maintenance and short life problems caused by the commutator. As a result, such motors are usually used in AC devices such as food mixers and power tools which are used only intermittently, and often have high starting-torque demands. Continuous speed control of a universal motor running on AC is easily obtained by use of a thyristor circuit, while (imprecise) stepped speed control can be accomplished using multiple taps on the field coil. Household blenders that advertise many speeds frequently combine a field coil with several taps and a diode that can be inserted in series with the motor (causing the motor to run on half-wave rectified AC).

Universal motors generally run at high speeds, making them useful for appliances such as blenders, vacuum cleaners, and hair dryers where high RPM operation is desirable. They are also commonly used in portable power tools, such as drills, sanders (both disc and orbital), circular and jig saws, where the motor's characteristics work well. Many vacuum cleaner and weed trimmer motors exceed 10,000 RPM, while Dremel and other similar miniature grinders will often exceed 30,000 RPM.

Universal motors also lend themselves to electronic speed control and, as such, are an ideal choice for domestic washing machines. The motor can be used to agitate the drum (both forwards and in reverse) by switching the field winding with respect to the armature. The motor can also be run up to the high speeds required for the spin cycle.

Motor damage may occur due to overspeeding (running at an RPM in excess of design limits) if the unit is operated with no significant load. On larger motors, sudden loss of load is to be avoided, and the possibility of such an occurrence is incorporated into the motor's protection and control schemes. In some smaller applications, a fan blade attached to the shaft often acts as an artificial load to limit the motor speed to a safe value, as well as a means to circulate cooling airflow over the armature and field windings.

AC motors

In 1882, Nikola Tesla discovered the rotating magnetic field, and pioneered the use of a rotary field of force to operate machines. He exploited the principle to design a unique two-phase induction motor in 1883. In 1885, Galileo Ferraris independently researched the concept. In 1888, Ferraris published his research in a paper to the Royal Academy of Sciences in Turin.

Tesla had suggested that the commutators from a machine could be removed and the device could operate on a rotary field of force. Professor Poeschel, his teacher, stated that would be akin to building a perpetual motion machine. Tesla would later attain U.S. Patent 0,416,194, Electric Motor (December 1889), which resembles the motor seen in many of Tesla's photos. This classic alternating current electro-magnetic motor was an induction motor.

Michail Osipovich Dolivo-Dobrovolsky later invented a three-phase "cage-rotor" in 1890. This type of motor is now used for the vast majority of commercial applications.

An AC motor has two parts. An stationary stator having coils supplied with AC current to produce a rotating magnetic field, and a rotor attached to the output shaft that is given a torque by the rotating field.

AC Motor with sliding rotor

Conical rotor brake motor incorporates the brake as an integral part of the conical sliding rotor. When the motor is at rest, a spring acts on the sliding rotor and forces the brake ring against the brake cap in the motor, holding the rotor stationary. When the motor is energized, its magnetic field generates both an axial and a radial component. The axial component overcomes the spring force, releasing the brake; while the radial component causes the rotor to turn. There is no additional brake control required.

Synchronous electric motor

A synchronous electric motor is an AC motor distinguished by a rotor spinning with coils passing magnets at the same rate as the alternating current and resulting magnetic field which drives it. Another way of saying this is that it has zero slip under usual operating conditions. Contrast this with an induction motor, which must slip to produce torque. A synchronous motor is like an induction motor except the rotor is excited by a DC field. Slip rings and brushes are used to conduct current to rotor. The rotor poles connect to each other and move at the same speed hence the name synchronous motor.

Induction motor

An induction motor is an asynchronous AC motor where power is transferred to the rotor by electromagnetic induction. An induction motor resembles a rotating transformer, because the stator (stationary part) is essentially the primary side of the transformer and the rotor (rotating part) is the secondary side. Polyphase induction motors are widely used in industry.

Induction motors may be further divided into squirrel-cage motors and wound-rotor motors. Squirrel-cage motors have a heavy winding made up of solid bars, usually aluminum or copper, joined by rings at the ends of the rotor. Currents induced into this winding provide the rotor magnetic field. The shape of the rotor bars determines the speed-torque characterstics. At low speeds, the current induced in the squirrel cage is nearly at line frequency and tends to flow in the outer parts of the rotor cage. As the motor accelerates, the slip frequency becomes lower, and more current flows in the interior of the winding. By shaping the bars to change the resistance of the windings portions in the interior and outer parts of the cage, effectively a variable resistance is inserted in the rotor circuit.

In a wound-rotor motor, the rotor winding is made of many turns of insulated wire and is connected to slip rings on the motor shaft. An external resistor or other control devices can be connected in the rotor circuit. Resistors allow control of the motor speed, although significant power is dissipated in the external resistance. A converter can be fed from the rotor circuit and return the slip-freqeuncy power that would otherwise be wasted back into the power system.

The wound-rotor induction motor is used primarily to start a high inertia load or a load that requires a very high starting torque across the full speed range. By correctly selecting the resistors used in the secondary resistance or slip ring starter, the motor is able to produce maximum torque at a relatively low supply current from zero speed to full speed. This type of motor also offers controllable speed.

Motor speed can be changed because the torque curve of the motor is effectively modified by the amount of resistance connected to the rotor circuit. Increasing the value of resistance will move the speed of maximum torque down. If the resistance connected to the rotor is increased beyond the point where the maximum torque occurs at zero speed, the torque will be further reduced.

When used with a load that has a torque curve that increases with speed, the motor will operate at the speed where the torque developed by the motor is equal to the load torque. Reducing the load will cause the motor to speed up, and increasing the load will cause the motor to slow down until the load and motor torque are equal. Operated in this manner, the slip losses are dissipated in the secondary resistors and can be very significant. The speed regulation and net efficiency is also very poor.

Doubly-fed electric motor

Doubly-fed electric motors have two independent multiphase windings on both the stator and the rotor, with at least one of the winding sets electronically controlled for variable speed operation. Doubly-fed electric motors are machines with an effective constant torque speed range that is twice synchronous speed for a given frequency of excitation. This is twice the constant torque speed range as singly-fed electric machines, which have only one active winding set.

A doubly-fed motor allows for a smaller electronic converter but the cost of the rotor winding and slip rings may offset the saving in the power electronics components. Difficulties with controlling speed near synchronous speed limit applications.

Singly-fed electric motor

Most AC motors are singly-fed. Singly-fed electric motors have a single multiphase winding set that is connected to a power supply. Singly-fed electric machines may be either induction or synchronous. The active winding set can be electronically controlled. Singly-fed electric machines have an effective constant torque speed range up to synchronous speed for a given excitation frequency.

Electrical transformer

The EMF predicted by Faraday's law is also responsible for electrical transformers. When the electric current in a loop of wire changes, the changing current creates a changing magnetic field. A second wire in reach of this magnetic field will experience this change in magnetic field as a change in its coupled magnetic flux, a $d\Phi_B / dt$. Therefore, an electromotive force is set up in the second loop called the **induced EMF** or **transformer EMF**. If the two ends of this loop are connected through an electrical load, current will flow.

Magnetic flow meter

Faraday's law is used for measuring the flow of electrically conductive liquids and slurries. Such instruments are called magnetic flow meters. The induced voltage $\mathcal{E}$

generated in the magnetic field B due to a conductive liquid moving at velocity v is thus given by:

$$\mathcal{E} = B\ell v,$$

where ℓ is the distance between electrodes in the magnetic flow meter.

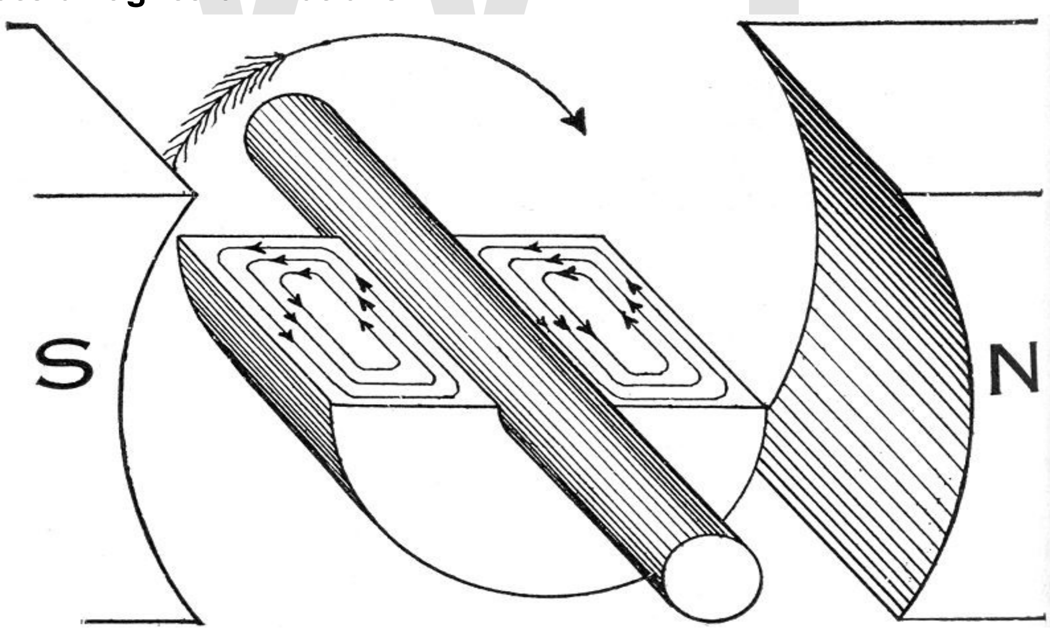
Parasitic induction and waste heating

All metal objects moving in relation to a static magnetic field will experience inductive power flow, as do all stationary metal objects in relation to a moving magnetic field. These power flows are occasionally undesirable, resulting in flowing electric current at very low voltage and heating of the metal.

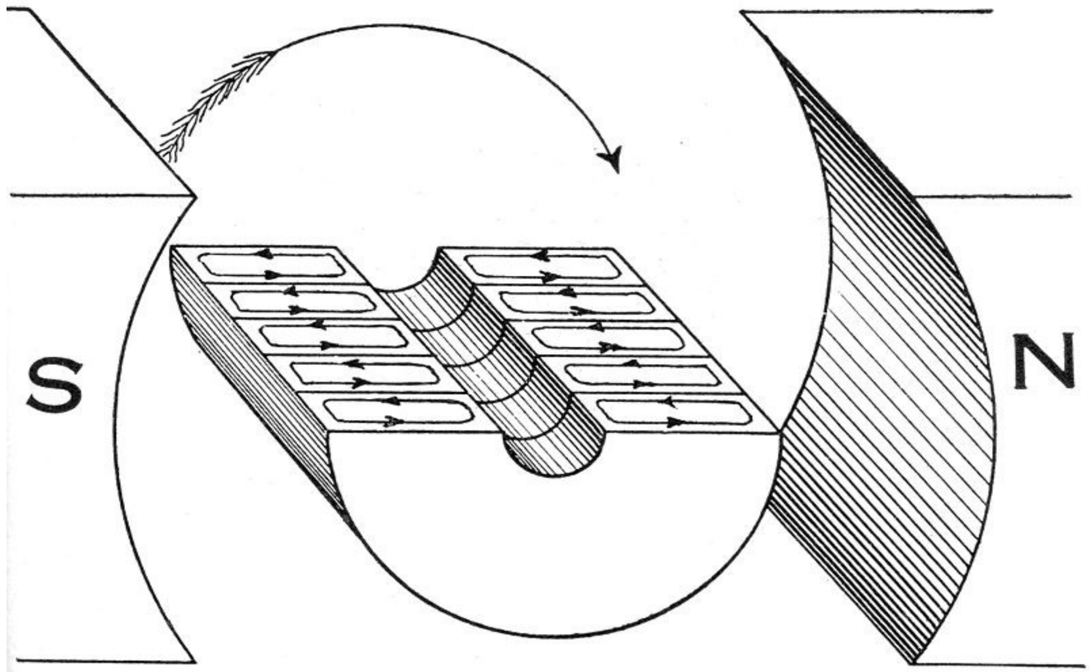
There are a number of methods employed to control these undesirable inductive effects.

- Electromagnets in electric motors, generators, and transformers do not use solid metal, but instead use thin sheets of metal plate, called laminations. These thin plates reduce the parasitic eddy currents, as described below.
- Inductive coils in electronics typically use magnetic cores to minimize parasitic current flow. They are a mixture of metal powder plus a resin binder that can hold any shape. The binder prevents parasitic current flow through the powdered metal.

Electromagnet laminations

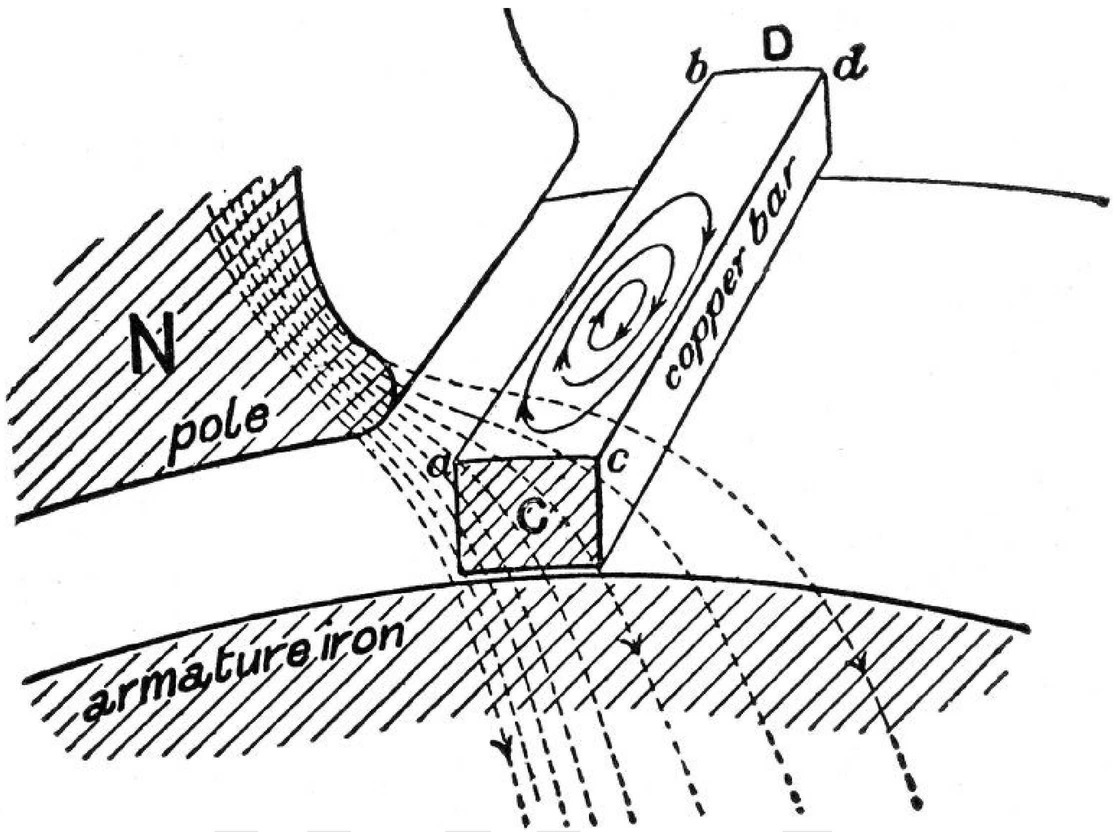


Eddy currents occur when a solid metallic mass is rotated in a magnetic field, because the outer portion of the metal cuts more lines of force than the inner portion, hence the induced electromotive force not being uniform, tends to set up currents between the points of greatest and least potential. Eddy currents consume a considerable amount of energy and often cause a harmful rise in temperature.



Only five laminations or plates are shown in this example, so as to show the subdivision of the eddy currents. In practical use, the number of laminations or punchings ranges from 40 to 66 per inch, and brings the eddy current loss down to about one percent. While the plates can be separated by insulation, the voltage is so low that the natural rust/oxide coating of the plates is enough to prevent current flow across the laminations.

This is a rotor approximately 20mm in diameter from a DC motor used in a CD player. Note the laminations of the electromagnet pole pieces, used to limit parasitic inductive losses.



In this illustration, a solid copper bar inductor on a rotating armature is just passing under the tip of the pole piece N of the field magnet. Note the uneven distribution of the lines of force across the bar inductor. The magnetic field is more concentrated and thus stronger on the left edge of the copper bar (a,b) while the field is weaker on the right edge (c,d). Since the two edges of the bar move with the same velocity, this difference in field strength across the bar creates whirls or current eddies within the copper bar.

This is one reason high voltage devices tend to be more efficient than low voltage devices. High voltage devices use many turns of small-gauge wire in motors, generators, and transformers. These many small turns of inductor wire in the electromagnet break up the eddy flows that can form within the large, thick inductors of low voltage, high current devices.

Chapter 10

Gauss's Law

In physics, **Gauss's law**, also known as **Gauss's flux theorem**, is a law relating the distribution of electric charge to the resulting electric field. Gauss's law states that:

The electric flux through any closed surface is proportional to the enclosed electric charge.

The law was formulated by Carl Friedrich Gauss in 1835, but was not published until 1867. It is one of four of Maxwell's equations which form the basis of classical electrodynamics, the other three being Gauss's law for magnetism, Faraday's law of induction, and Ampère's law with Maxwell's correction. Gauss's law can be used to derive Coulomb's law, and vice versa.

Gauss's law may be expressed in its integral form:

$$\oint_S \mathbf{E} \cdot \mathrm{d}\mathbf{A} = \frac{Q}{\varepsilon_0},$$

where the left-hand side of the equation is a surface integral denoting the electric flux through a closed surface S, and the right-hand side of the equation is the total charge enclosed by S divided by the electric constant.

Gauss's law also has a differential form:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

where $\nabla \cdot \mathbf{E}$ is the divergence of the electric field, and ρ is the charge density.

The integral and differential forms are related by the divergence theorem, also called Gauss's theorem. Each of these forms can also be expressed two ways: In terms of a relation between the electric field **E** and the total electric charge, or in terms of the electric displacement field **D** and the free electric charge.

Gauss's law has a close mathematical similarity with a number of laws in other areas of physics, such as Gauss's law for magnetism and Gauss's law for gravity. In fact, any "inverse-square law" can be formulated in a way similar to Gauss's law: For example, Gauss's law itself is essentially equivalent to the inverse-square Coulomb's law, and Gauss's law for gravity is essentially equivalent to the inverse-square Newton's law of gravity.

Gauss's law can be used to demonstrate that all electric fields inside a Faraday cage have an electric charge. Gauss's law is something of an electrical analogue of Ampère's law, which deals with magnetism.

In terms of total charge

Integral form

For a volume V with surface S, Gauss's law states that

$$\Phi_{E,S} = \frac{Q}{\varepsilon_0}$$

where $\Phi_{E,S}$ is the electric flux through S, Q is total charge inside V, and ε_0 is the electric constant. The electric flux is given by a surface integral over S:

$$\oint_S \mathbf{E} \cdot \mathrm{d}\mathbf{A}$$

where **E** is the electric field, d**A** is a vector representing an infinitesimal element of area, and · represents the dot product..

Applying the integral form

If the electric field is known everywhere, Gauss's law makes it quite easy, in principle, to find the distribution of electric charge: The charge in any given region can be deduced by integrating the electric field to find the flux.

However, much more often, it is the reverse problem that needs to be solved: The electric charge distribution is known, and the electric field needs to be computed. This is much more difficult, since if you know the total flux through a given surface, that gives almost no information about the electric field, which (for all you know) could go in and out of the surface in arbitrarily complicated patterns.

An exception is if there is some symmetry in the situation, which mandates that the electric field passes through the surface in a uniform way. Then, if the total flux is known, the field itself can be deduced at every point. Common examples of symmetries which lend themselves to Gauss's law include cylindrical symmetry, planar symmetry, and spherical symmetry.

Differential form

In differential form, Gauss's law states:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

where $\nabla \cdot$ denotes divergence, **E** is the electric field, and ρ is the total electric charge density (including both free and bound charge), and ε_0 is the electric constant. This is mathematically equivalent to the integral form, because of the divergence theorem.

Equivalence of integral and differential forms

The integral and differential forms are mathematically equivalent, by the divergence theorem. Here is the argument more specifically:

The integral form of Gauss's law is:

$$\oint_S \mathbf{E} \cdot \mathrm{d}\mathbf{A} = \frac{Q}{\varepsilon_0}$$

for any closed surface S containing charge Q. By the divergence theorem, this equation is equivalent to:

$$\iiint_V \nabla \cdot \mathbf{E} \ \mathrm{d}V = \frac{Q}{\varepsilon_0}$$

for any volume V containing charge Q. By the relation between charge and charge density, this equation is equivalent to:

$$\iiint_V \nabla \cdot \mathbf{E} \ \mathrm{d}V = \iiint_V \frac{\rho}{\varepsilon_0} \ \mathrm{d}V$$

for any volume V. In order for this equation to be simultaneously true for every possible volume V, it is necessary (and sufficient) for the integrands to be equal everywhere. Therefore, this equation is equivalent to:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}.$$

Thus the integral and differential forms are equivalent.

In terms of free charge

Free versus bound charge

The electric charge that arises in the simplest textbook situations would be classified as "free charge"—for example, the charge which is transferred in static electricity, or the charge on a capacitor plate. In contrast, "bound charge" arises only in the context of dielectric (polarizable) materials. (All materials are polarizable to some extent.) When such materials are placed in an external electric field, the electrons remain bound to their respective atoms, but shift a microscopic distance in response to the field, so that they're more on one side of the atom than the other. All these microscopic displacements add up to give a macroscopic net charge distribution, and this constitutes the "bound charge".

Although microscopically, all charge is fundamentally the same, there are often practical reasons for wanting to treat bound charge differently from free charge. The result is that the more "fundamental" Gauss's law, in terms of **E**, is sometimes put into the equivalent form below, which is in terms of **D** and the free charge only.

Integral form

This formulation of Gauss's law states that, for any volume V in space, with surface S, the following equation holds:

$$\Phi_{D,S} = Q_{\mathrm{free}},$$

where $\Phi_{D,S}$ is the flux of the electric displacement field **D** through S, and Q_{free} is the free charge contained in V. The flux $\Phi_{D,S}$ is defined analogously to the flux $\Phi_{E,S}$ of the electric field **E** through S. Specifically, it is given by the surface integral

$$\Phi_{D,S} = \oint_S \mathbf{D} \cdot \mathrm{d}\mathbf{A}.$$

Differential form

The differential form of Gauss's law, involving free charge only, states:

$$\nabla \cdot \mathbf{D} = \rho_{\mathrm{free}}$$

where $\nabla \cdot \mathbf{D}$ is the divergence of the electric displacement field, and ρ_{free} is the free electric charge density.

The differential form and integral form are mathematically equivalent. The proof primarily involves the divergence theorem.

Equivalence of total and free charge statements

Proof that the formulations of Gauss's law in terms of free charge are equivalent to the formulations involving total charge.

In linear materials

In homogeneous, isotropic, nondispersive, linear materials, there is a nice, simple relationship between **E** and **D**:

$$\varepsilon \mathbf{E} = \mathbf{D}$$

where ε is the permittivity of the material. Under these circumstances, there is yet another pair of equivalent formulations of Gauss's law:

$$\Phi_{E,S} = \frac{Q_{\text{free}}}{\varepsilon}$$
$$\nabla \cdot \mathbf{E} = \frac{\rho_{\text{free}}}{\varepsilon}$$

Relation to Coulomb's law

Deriving Gauss's law from Coulomb's law

Gauss's law can be derived from Coulomb's law, which states that the electric field due to a stationary point charge is:

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\frac{\mathbf{e_r}}{r^2}$$

where

$\mathbf{e_r}$ is the radial unit vector,
r is the radius, |**r**|,
ε_0 is the electric constant,
q is the charge of the particle, which is assumed to be located at the origin.

Using the expression from Coulomb's law, we get the total field at **r** by using an integral to sum the field at **r** due to the infinitesimal charge at each other point **s** in space, to give

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{s})(\mathbf{r}-\mathbf{s})}{|\mathbf{r}-\mathbf{s}|^3} d^3\mathbf{s}$$

where ρ is the charge density. If we take the divergence of both sides of this equation with respect to **r**, and use the known theorem

$$\nabla \cdot \left(\frac{\mathbf{s}}{|\mathbf{s}|^3}\right) = 4\pi\delta(\mathbf{s})$$

where δ(**s**) is the Dirac delta function, the result is

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon_0} \int \rho(\mathbf{s})\ \delta(\mathbf{r}-\mathbf{s})\ d^3\mathbf{s}$$

Using the "sifting property" of the Dirac delta function, we arrive at

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \rho(\mathbf{r})/\epsilon_0$$

which is the differential form of Gauss's law, as desired.

Note that since Coulomb's law only applies to stationary charges, there is no reason to expect Gauss's law to hold for moving charges based on this derivation alone. In fact, Gauss's law does hold for moving charges, and in this respect Gauss's law is more general than Coulomb's law.

Deriving Coulomb's law from Gauss's law

Strictly speaking, Coulomb's law cannot be derived from Gauss's law alone, since Gauss's law does not give any information regarding the curl of **E**. However, Coulomb's law can be proven from Gauss's law if it is assumed, in addition, that the electric field from a point charge is spherically-symmetric (this assumption, like Coulomb's law itself, is exactly true if the charge is stationary, and approximately true if the charge is in motion).

Taking S in the integral form of Gauss's law to be a spherical surface of radius r, centered at the point charge Q, we have

$$\oint_S \mathbf{E} \cdot d\mathbf{A} = Q/\varepsilon_0$$

By the assumption of spherical symmetry, the integrand is a constant which can be taken out of the integral. The result is

$$4\pi r^2 \hat{\mathbf{r}} \cdot \mathbf{E}(\mathbf{r}) = Q/\varepsilon_0$$

where $\hat{\mathbf{r}}$ is a unit vector pointing radially away from the charge. Again by spherical symmetry, **E** points in the radial direction, and so we get

$$\mathbf{E}(\mathbf{r}) = \frac{Q}{4\pi\varepsilon_0} \frac{\hat{\mathbf{r}}}{r^2}$$

which is essentially equivalent to Coulomb's law. Thus the inverse-square law dependence of the electric field in Coulomb's law follows from Gauss's law.

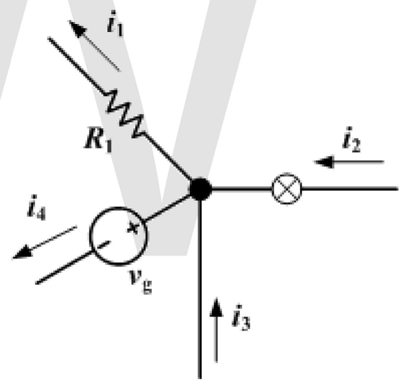
Chapter 11

Kirchhoff's Circuit Laws

Kirchhoff's circuit laws are two equalities that deal with the conservation of charge and energy in electrical circuits, and were first described in 1845 by Gustav Kirchhoff. Widely used in electrical engineering, they are also called Kirchhoff's rules or simply Kirchhoff's laws.

Both circuit rules can be directly derived from Maxwell's equations, but Kirchhoff preceded Maxwell and instead generalized work by Georg Ohm.

Kirchhoff's current law (KCL)



The current entering any junction is equal to the current leaving that junction. $i_1 + i_4 = i_2 + i_3$

This law is also called **Kirchhoff's point rule**, **Kirchhoff's junction rule** (or nodal rule), and **Kirchhoff's first rule**.

The principle of conservation of electric charge implies that:

> At any node (junction) in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node.
> or
> The algebraic sum of currents in a network of conductors meeting at a point is zero. (Assuming that current entering the junction is taken as positive and current leaving the junction is taken as negative).

Recalling that current is a signed (positive or negative) quantity reflecting direction towards or away from a node, this principle can be stated as:

$$\sum_{k=1}^{n} I_k = 0$$

n is the total number of branches with currents flowing towards or away from the node.

This formula is also valid for complex currents:

$$\sum_{k=1}^{n} \tilde{I}_k = 0$$

The law is based on the conservation of charge whereby the charge (measured in coulombs) is the product of the current (in amperes) and the time (which is measured in seconds).

Changing charge density

Physically speaking, the restriction regarding the "capacitor plate" means that Kirchhoff's current law is only valid if the charge density remains constant in the point that it is applied to. This is normally not a problem because of the strength of electrostatic forces: the charge buildup would cause repulsive forces to disperse the charges.

However, a charge build-up can occur in a capacitor, where the charge is typically spread over wide parallel plates, with a physical break in the circuit that prevents the positive and negative charge accumulations over the two plates from coming together and cancelling. In this case, the sum of the currents flowing into one plate of the capacitor is not zero, but rather is equal to the rate of charge accumulation. However, if the displacement current d**D**/dt is included, Kirchhoff's current law once again holds. (This is really only required if one wants to apply the current law to a point on a capacitor plate. In circuit analyses, however, the capacitor as a whole is typically treated as a unit, in which case the ordinary current law holds since exactly the current that enters the capacitor on the one side leaves it on the other side.)

More technically, Kirchhoff's current law can be found by taking the divergence of Ampère's law with Maxwell's correction and combining with Gauss's law, yielding:

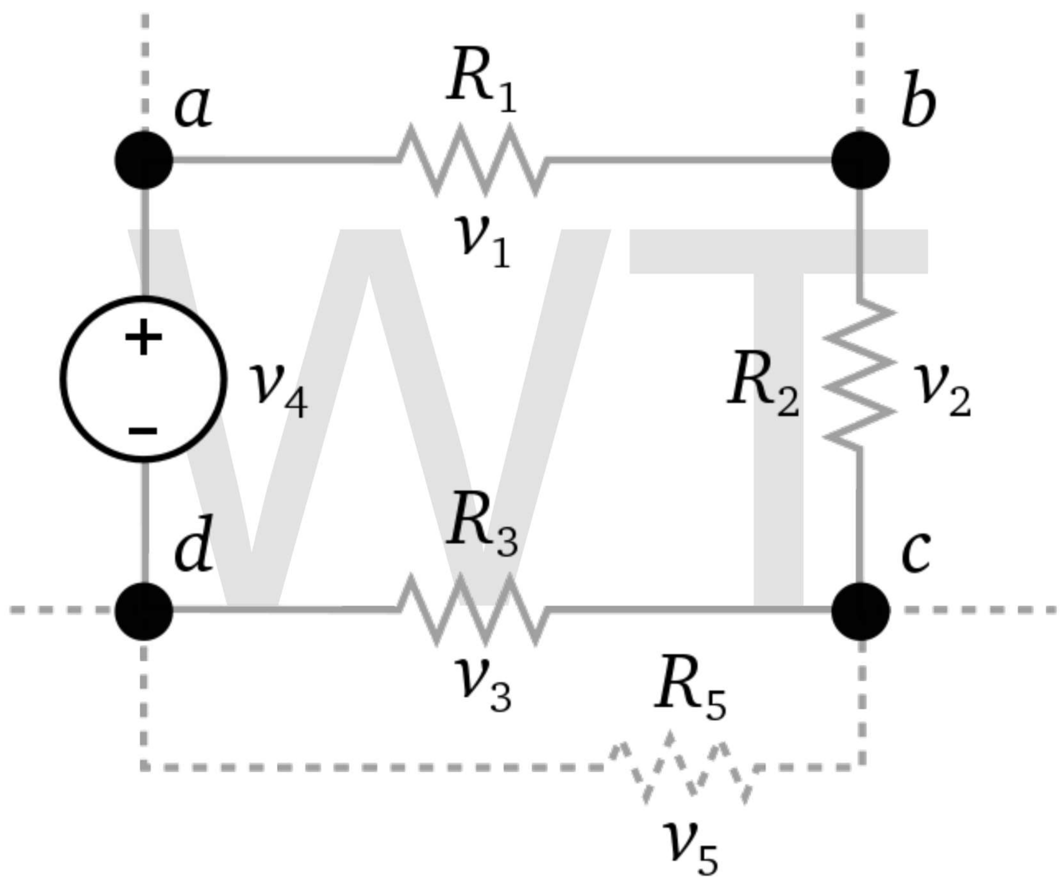
$$\nabla \cdot \mathbf{J} = -\nabla \cdot \frac{\partial \mathbf{D}}{\partial t} = -\frac{\partial \rho}{\partial t}$$

This is simply the charge conservation equation (in integral form, it says that the current flowing out of a closed surface is equal to the rate of loss of charge within the enclosed volume (Divergence theorem). Kirchhoff's current law is equivalent to the statement that the divergence of the current is zero, true for time-invariant ρ, or always true if the displacement current is included with **J**.

Uses

A matrix version of Kirchhoff's current law is the basis of most circuit simulation software, such as SPICE.

Kirchhoff's voltage law (KVL)



The sum of all the voltages around the loop is equal to zero. $v_1 + v_2 + v_3 - v_4 = 0$

This law is also called **Kirchhoff's second law**, **Kirchhoff's loop (or mesh) rule**, and **Kirchhoff's second rule**.

The principle of conservation of energy implies that

> The directed sum of the electrical potential differences (voltage) around any closed circuit is zero.

or
More simply, the sum of the emfs in any closed loop is equivalent to the sum of the potential drops in that loop.
or
The algebraic sum of the products of the resistances of the conductors and the currents in them in a closed loop is equal to the total emf available in that loop.

Similarly to KCL, it can be stated as:

$$\sum_{k=1}^{n} V_k = 0$$

Here, n is the total number of voltages measured. The voltages may also be complex:

$$\sum_{k=1}^{n} \tilde{V}_k = 0$$

This law is based on the conservation of "energy given/taken by potential field" (not including energy taken by dissipation). Given a voltage potential, a charge which has completed a closed loop doesn't gain or lose energy as it has gone back to initial potential level.

This law holds true even when resistance (which causes **dissipation** of energy) is present in a circuit. The validity of this law in this case can be understood if one realizes that a charge in fact doesn't go back to its starting point, due to dissipation of energy. A charge will just terminate at the negative terminal, instead of positive terminal. This means all the energy given by the potential difference has been fully consumed by resistance which in turn loses the energy as heat dissipation.

To summarize, Kirchhoff's voltage law has nothing to do with gain or loss of energy by electronic components (resistors, capacitors, etc). It is a law referring to the potential field generated by voltage sources. In this potential field, regardless of what electronic components are present, the gain or loss in "energy given by the potential field" must be zero when a charge completes a closed loop.

Electric field and electric potential

Kirchhoff's voltage law could be viewed as a consequence of the principle of conservation of energy. Otherwise, it would be possible to build a perpetual motion machine that passed a current in a circle around the circuit.

Considering that electric potential is defined as a line integral over an electric field, Kirchhoff's voltage law can be expressed equivalently as

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = 0,$$

which states that the line integral of the electric field around closed loop C is zero.

In order to return to the more special form, this integral can be "cut in pieces" in order to get the voltage at specific components.

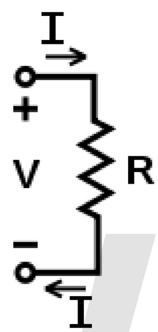
Limitations

This is a simplification of Faraday's law of induction for the special case where there is no fluctuating magnetic field linking the closed loop. Therefore, it practically suffices for explaining circuits containing only resistors and capacitors.

In the presence of a changing magnetic field the electric field is not conservative and it cannot therefore define a pure scalar potential—the line integral of the electric field around the circuit is not zero. This is because energy is being transferred from the magnetic field to the current (or vice versa). In order to "fix" Kirchhoff's voltage law for circuits containing inductors, an effective potential drop, or electromotive force (emf), is associated with each inductance of the circuit, exactly equal to the amount by which the line integral of the electric field is not zero by Faraday's law of induction.

Chapter 12

Ohm's Law



V, I, and R, the parameters of Ohm's law

Ohm's law states that the current through a conductor between two points is directly proportional to the potential difference or voltage across the two points, and inversely proportional to the resistance between them.

The mathematical equation that describes this relationship is:

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

where I is the current through the resistance in units of amperes, V is the potential difference measured across the resistance in units of volts, and R is the resistance of the conductor in units of ohms. More specifically, Ohm's law states that the R in this relation is constant, independent of the current.

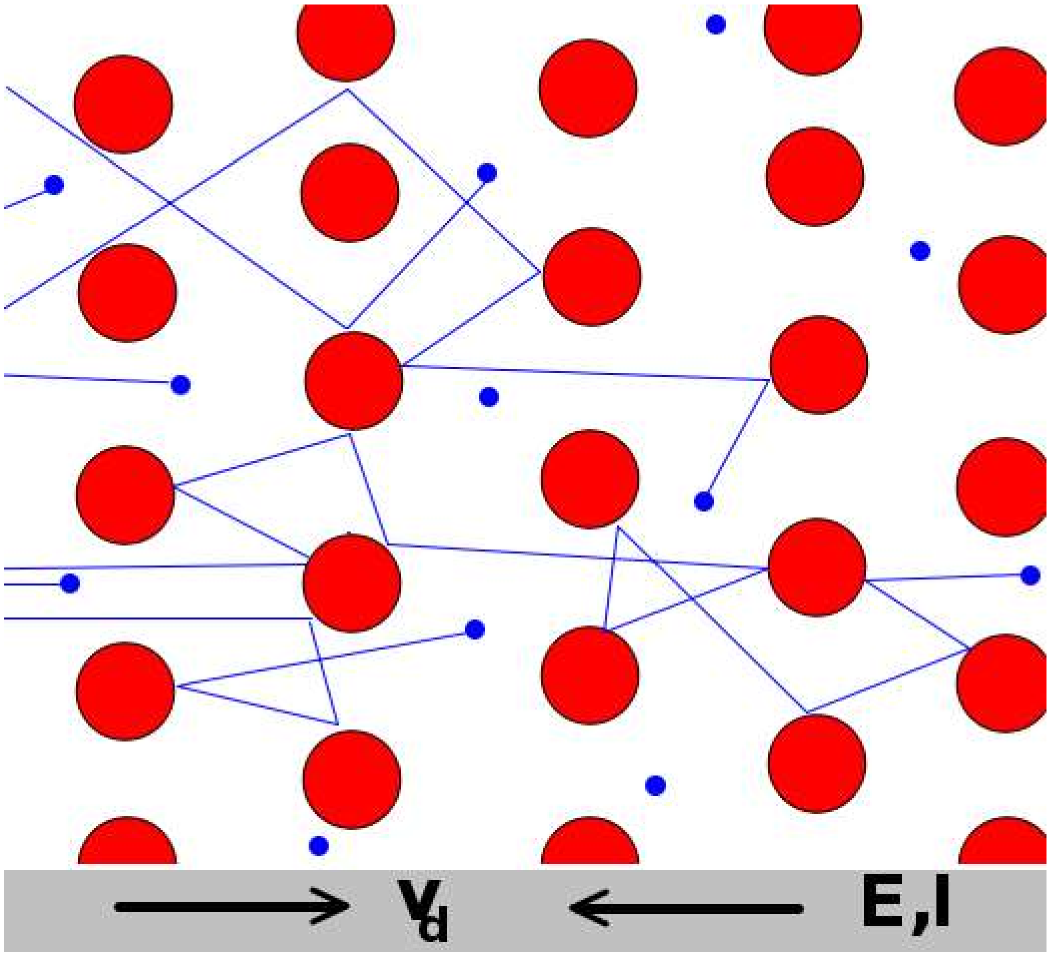
The law was named after the German physicist Georg Ohm, who, in a treatise published in 1827, described measurements of applied voltage and current through simple electrical circuits containing various lengths of wire. He presented a slightly more complex equation than the one above to explain his experimental results. The above equation is the modern form of Ohm's law.

In physics, the term Ohm's law is also used to refer to various generalizations of the law originally formulated by Ohm. The simplest example of this is:

$$J = \sigma E,$$

where **J** is the current density at a given location in a resistive material, **E** is the electric field at that location, and σ is a material dependent parameter called the conductivity. This reformulation of Ohm's law is due to Gustav Kirchhoff.

Microscopic origins of Ohm's law



Drude Model electrons (shown here in blue) constantly bounce between heavier, stationary crystal ions (shown in red).

The dependence of the current density on the applied electric field is essentially quantum mechanical in nature. A qualitative description leading to Ohm's law can be based upon classical mechanics using the Drude model developed by Paul Drude in 1900.

The Drude model treats electrons (or other charge carriers) like pinballs bouncing between the ions that make up the structure of the material. Electrons will be accelerated in the opposite direction to the electric field by the average electric field at their location. With each collision, though, the electron is deflected in a random direction with a velocity that is much larger than the velocity gained by the electric field. The net result is that electrons take a tortuous path due to the collisions, but generally drift in a direction opposing the electric field.

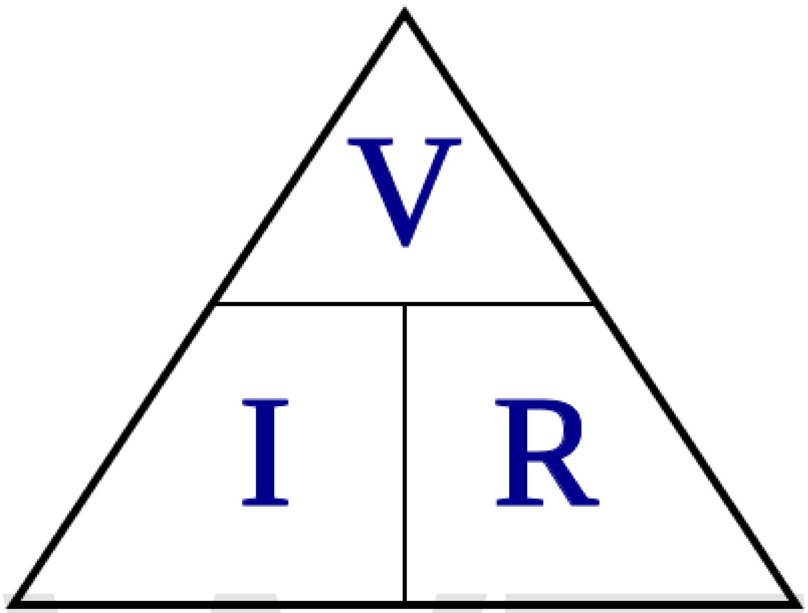
The drift velocity then determines the electric current density and its relationship to **E** and is independent of the collisions. Drude calculated the average drift velocity from $\mathbf{p} = -e\mathbf{E}\tau$ where **p** is the average momentum, −e is the charge of the electron and τ is the average time between the collisions. Since both the momentum and the current density are proportional to the drift velocity, the current density becomes proportional to the applied electric field; this leads to Ohm's law.

Hydraulic analogy

A hydraulic analogy is sometimes used to describe Ohm's Law. Water pressure, measured by pascals (or PSI), is the analog of voltage because establishing a water pressure difference between two points along a (horizontal) pipe causes water to flow. Water flow rate, as in liters per second, is the analog of current, as in coulombs per second. Finally, flow restrictors — such as apertures placed in pipes between points where the water pressure is measured — are the analog of resistors. We say that the rate of water flow through an aperture restrictor is proportional to the difference in water pressure across the restrictor. Similarly, the rate of flow of electrical charge, that is, the electric current, through an electrical resistor is proportional to the difference in voltage measured across the resistor.

Flow and pressure variables can be calculated in fluid flow network with the use of the hydraulic ohm analogy. The method can be applied to both steady and transient flow situations. In the linear laminar flow region, Poiseuille's law describes the hydraulic resistance of a pipe, but in the turbulent flow region the pressure–flow relations become nonlinear.

Circuit analysis



Ohm's law triangle

In circuit analysis, three equivalent expressions of Ohm's law are used interchangeably:

$$I = \frac{V}{R} \quad \text{or} \quad V = IR \quad \text{or} \quad R = \frac{V}{I}$$

Indeed, each is quoted by some sources as the defining relationship of Ohm's law, or all three are quoted, or derived from a proportional form, or even just the two that do not correspond to Ohm's original statement may sometimes be given.

The interchangeability of the equation may be represented by a triangle, where V (voltage) is placed on the top section, the I (current) is placed to the left section, and the R (resistance) is placed to the right. The line that divides the left and right sections indicate multiplication, and the divider between the top and bottom sections indicates division (hence the division bar).

Resistive circuits

Resistors are circuit elements that impede the passage of electric charge in agreement with Ohm's law, and are designed to have a specific resistance value R. In a schematic diagram the resistor is shown as a zig-zag symbol. An element (resistor or conductor) that behaves according to Ohm's law over some operating range is referred to as an ohmic device (or an ohmic resistor) because Ohm's law and a single value for the resistance suffice to describe the behavior of the device over that range.

Ohm's law holds for circuits containing only resistive elements (no capacitances or inductances) for all forms of driving voltage or current, regardless of whether the driving voltage or current is constant (DC) or time-varying such as AC. At any instant of time Ohm's law is valid for such circuits.

Reactive circuits with time-varying signals

When reactive elements such as capacitors, inductors, or transmission lines are involved in a circuit to which AC or time-varying voltage or current is applied, the relationship between voltage and current becomes the solution to a differential equation, so Ohm's law (as defined above) does not directly apply since that form contains only resistances having value R, not complex impedances which may contain capacitance ("C") or inductance ("L").

Equations for time-invariant AC circuits take the same form as Ohm's law, however, the variables are generalized to complex numbers and the current and voltage waveforms are complex exponentials.

In this approach, a voltage or current waveform takes the form Ae^{st}, where t is time, s is a complex parameter, and A is a complex scalar. In any linear time-invariant system, all of the currents and voltages can be expressed with the same s parameter as the input to the system, allowing the time-varying complex exponential term to be canceled out and the system described algebraically in terms of the complex scalars in the current and voltage waveforms.

The complex generalization of resistance is impedance, usually denoted Z; it can be shown that for an inductor,

$$Z = sL$$

and for a capacitor,

$$Z = \frac{1}{sC}.$$

We can now write,

$$\boldsymbol{V} = \boldsymbol{I} \cdot \boldsymbol{Z}$$

where **V** and **I** are the complex scalars in the voltage and current respectively and **Z** is the complex impedance.

This form of Ohm's law, with Z taking the place of R, generalizes the simpler form. When Z is complex, only the real part is responsible for dissipating heat.

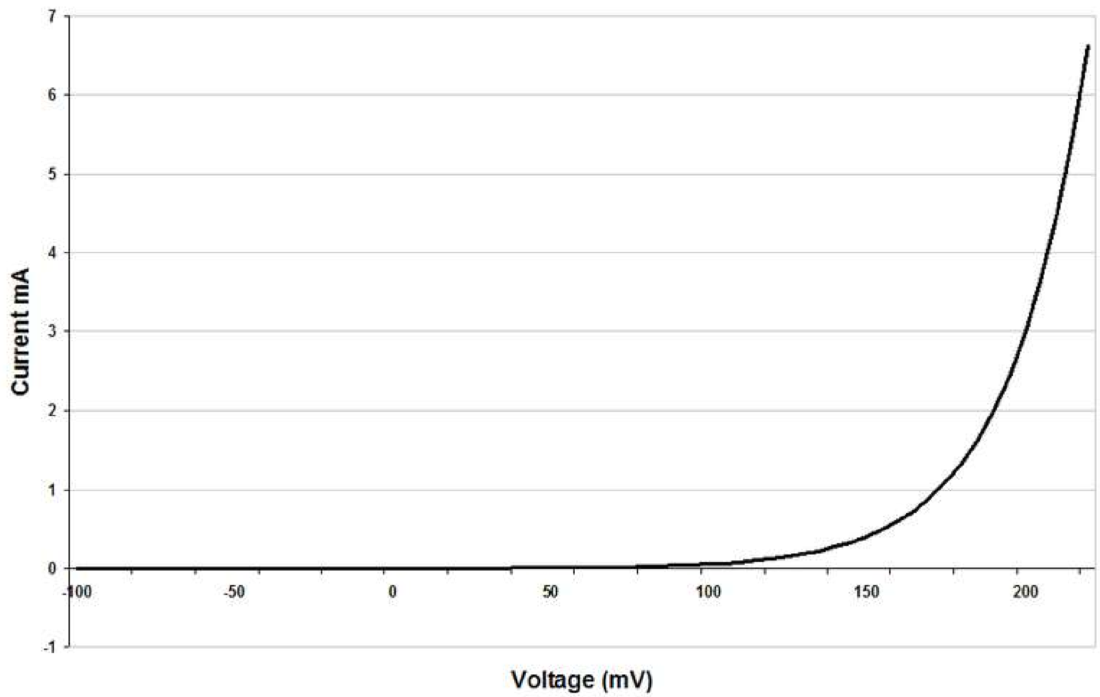
In the general AC circuit, Z varies strongly with the frequency parameter s, and so also will the relationship between voltage and current.

For the common case of a steady sinusoid, the s parameter is taken to be $j\omega$, corresponding to a complex sinusoid $Ae^{j\omega t}$. The real parts of such complex current and voltage waveforms describe the actual sinusoidal currents and voltages in a circuit, which can be in different phases due to the different complex scalars.

Linear approximations

Ohm's law is one of the basic equations used in the analysis of electrical circuits. It applies to both metal conductors and circuit components (resistors) specifically made for this behaviour. Both are ubiquitous in electrical engineering. Materials and components that obey Ohm's law are described as "ohmic" which means they produce the same value for resistance (R = V/I) regardless of the value of V or I which is applied and whether the applied voltage or current is DC (direct current) of either positive or negative polarity or AC (alternating current).

In a true ohmic device, the same value of resistance will be calculated from R = V/I regardless of the value of the applied voltage V. That is, the ratio of V/I is constant, and when current is plotted as a function of voltage the curve is linear (a straight line). If voltage is forced to some value V, then that voltage V divided by measured current I will equal R. Or if the current is forced to some value I, then the measured voltage V divided by that current I is also R. Since the plot of I versus V is a straight line, then it is also true that for any set of two different voltages V_1 and V_2 applied across a given device of resistance R, producing currents $I_1 = V_1/R$ and $I_2 = V_2/R$, that the ratio $(V_1-V_2)/(I_1-I_2)$ is also a constant equal to R. The operator "delta" (Δ) is used to represent a difference in a quantity, so we can write $\Delta V = V_1-V_2$ and $\Delta I = I_1-I_2$. Summarizing, for any truly ohmic device having resistance R, $V/I = \Delta V/\Delta I = R$ for any applied voltage or current or for the difference between any set of applied voltages or currents.



Plot of I–V curve of an ideal p-n junction diode at 1µA reverse leakage current. Failure of the device to follow Ohm's law is clearly shown since the curve is not a straight line.

There are, however, components of electrical circuits which do not obey Ohm's law; that is, their relationship between current and voltage (their I–V curve) is nonlinear. An example is the p-n junction diode (curve at right). As seen in the figure, the current does not increase linearly with applied voltage for a diode. One can determine a value of current (I) for a given value of applied voltage (V) from the curve, but not from Ohm's law, since the value of "resistance" is not constant as a function of applied voltage. Further, the current only increases significantly if the applied voltage is positive, not negative. The ratio V/I for some point along the nonlinear curve is sometimes called the static, or chordal, or DC, resistance , but as seen in the figure the value of total V over total I varies depending on the particular point along the nonlinear curve which is chosen. This means the "DC resistance" V/I at some point on the curve is not the same as what would be determined by applying an AC signal having peak amplitude ΔV volts or ΔI amps centered at that same point along the curve and measuring ΔV/ΔI. However, in some diode applications, the AC signal applied to the device is small and it is possible to analyze the circuit in terms of the dynamic, small-signal, or incremental resistance, defined as the one over the slope of the V–I curve at the average value (DC operating point) of the voltage (that is, one over the derivative of current with respect to voltage). For sufficiently small signals, the dynamic resistance allows the Ohm's law small signal resistance to be calculated as approximately one over the slope of a line drawn tangentially to the V-I curve at the DC operating point.

Temperature effects

Ohm's law has sometimes been stated as, "for a conductor in a given state, the electromotive force is proportional to the current produced." That is, that the resistance, the ratio of the applied electromotive force (or voltage) to the current, "does not vary with the current strength ." The qualifier "in a given state" is usually interpreted as meaning "at a constant temperature," since the resistivity of materials is usually temperature dependent. Because the conduction of current is related to Joule heating of the conducting body, according to Joule's first law, the temperature of a conducting body may change when it carries a current. The dependence of resistance on temperature therefore makes resistance depend upon the current in a typical experimental setup, making the law in this form difficult to directly verify. Maxwell and others worked out several methods to test the law experimentally in 1876, controlling for heating effects.

Relation to heat conductions

Ohm's principle predicts the flow of electrical charge (i.e. current) in electrical conductors when subjected to the influence of voltage differences; Jean-Baptiste-Joseph Fourier's principle predicts the flow of heat in heat conductors when subjected to the influence of temperature differences.

The same equation describes both phenomena, the equation's variables taking on different meanings in the two cases. Specifically, solving a heat conduction (Fourier) problem with temperature (the driving "force") and flux of heat (the rate of flow of the driven "quantity", i.e. heat energy) variables also solves an analogous electrical conduction (Ohm) problem having electric potential (the driving "force") and electric current (the rate of flow of the driven "quantity", i.e. charge) variables.

The basis of Fourier's work was his clear conception and definition of thermal conductivity. He assumed that, all else being the same, the flux of heat is strictly proportional to the gradient of temperature. Although undoubtedly true for small temperature gradients, strictly proportional behavior will be lost when real materials (e.g. ones having a thermal conductivity that is a function of temperature) are subjected to large temperature gradients.

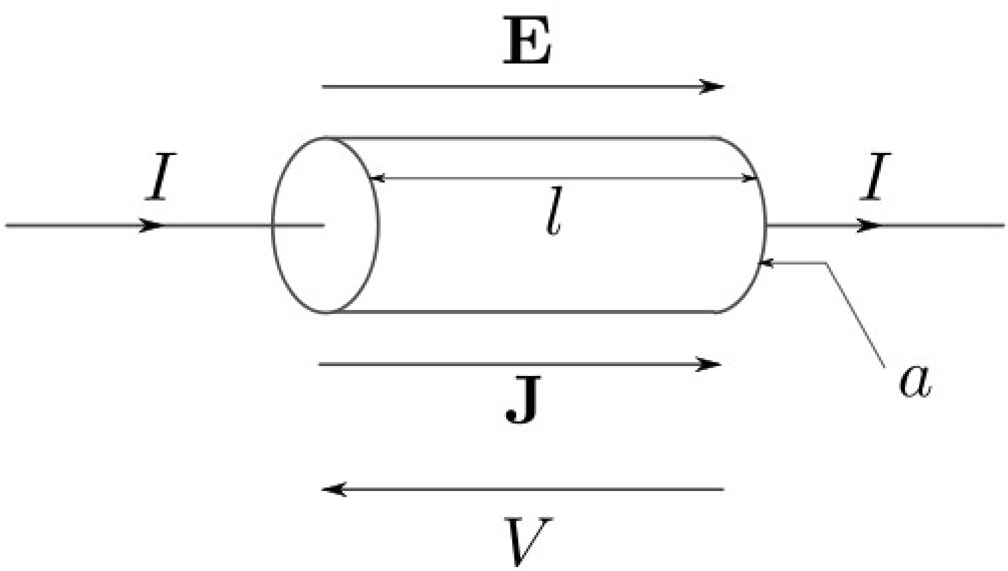
A similar assumption is made in the statement of Ohm's law: other things being alike, the strength of the current at each point is proportional to the gradient of electric potential. The accuracy of the assumption that flow is proportional to the gradient is more readily tested, using modern measurement methods, for the electrical case than for the heat case.

Other versions of Ohm's law

Ohm's law, in the form above, is an extremely useful equation in the field of electrical/electronic engineering because it describes how voltage, current and resistance are interrelated on a "macroscopic" level, that is, commonly, as circuit elements in an electrical circuit. Physicists who study the electrical properties of matter at the microscopic level use a closely related and more general vector equation, sometimes also referred to as Ohm's law, having variables that are closely related to the V, I, and R scalar variables of Ohm's law, but are each functions of position within the conductor. Physicists often use this continuum form of Ohm's Law:

$$\mathbf{E} = \rho \mathbf{J}$$

where "**E**" is the electric field vector with units of volts per meter (analogous to "V" of Ohm's law which has units of volts), "**J**" is the current density vector with units of amperes per unit area (analogous to "I" of Ohm's law which has units of amperes), and "ρ" (Greek "rho") is the resistivity with units of ohm·meters (analogous to "R" of Ohm's law which has units of ohms). The above equation is sometimes written as **J** = σ**E** where "σ" is the conductivity which is the reciprocal of ρ.



Current flowing through a uniform cylindrical conductor (such as a round wire) with a uniform field applied.

The potential difference between two points is defined as:

$$\Delta V = -\int \mathbf{E} \cdot d\mathbf{l}$$

with $d\mathbf{l}$ the element of path along the integration of electric field vector **E**. If the applied **E** field is uniform and oriented along the length of the conductor as shown in the figure, then defining the voltage V in the usual convention of being opposite in direction to the field (see figure), and with the understanding that the voltage V is measured differentially across the length of the conductor allowing us to drop the Δ symbol, the above vector equation reduces to the scalar equation:

$$V = El \quad \text{or} \quad E = \frac{V}{l}$$

Since the **E** field is uniform in the direction of wire length, for a conductor having uniformly consistent resistivity ρ, the current density **J** will also be uniform in any cross-sectional area and oriented in the direction of wire length, so we may write:

$$J = \frac{I}{a}$$

Substituting the above 2 results (for E and J respectively) into the continuum form shown at the beginning of this section:

$$\frac{V}{l} = \frac{I}{a}\rho \qquad \text{or} \qquad V = I\rho\frac{l}{a}$$

The electrical resistance of a uniform conductor is given in terms of resistivity by:

$$R = \rho\frac{l}{a}$$

where l is the length of the conductor in SI units of meters, a is the cross-sectional area (for a round wire $a = \pi r^2$ if r is radius) in units of meters squared, and ρ is the resistivity in units of ohm·meters.

After substitution of R from the above equation into the equation preceding it, the continuum form of Ohm's law for a uniform field (and uniform current density) oriented along the length of the conductor reduces to the more familiar form:

$$V = IR$$

A perfect crystal lattice, with low enough thermal motion and no deviations from periodic structure, would have no resistivity, but a real metal has crystallographic defects, impurities, multiple isotopes, and thermal motion of the atoms. Electrons scatter from all of these, resulting in resistance to their flow.

The more complex generalized forms of Ohm's law are important to condensed matter physics, which studies the properties of matter and, in particular, its electronic structure. In broad terms, they fall under the topic of constitutive equations and the theory of transport coefficients.

Magnetic effects

The continuum form of the equation is only valid in the reference frame of the conducting material. If the material is moving at velocity **v** relative to a magnetic field **B**, a term must be added as follows:

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = \mathbf{J}\rho$$

History

In January 1781, before Georg Ohm's work, Henry Cavendish experimented with Leyden jars and glass tubes of varying diameter and length filled with salt solution. He measured the current by noting how strong a shock he felt as he completed the circuit with his body. Cavendish wrote that the "velocity" (current) varied directly as the "degree of electrification" (voltage). He did not communicate his results to other scientists at the time, and his results were unknown until Maxwell published them in 1879.

Ohm did his work on resistance in the years 1825 and 1826, and published his results in 1827 as the book Die galvanische Kette, mathematisch bearbeitet (The galvanic Circuit investigated mathematically). He drew considerable inspiration from Fourier's work on heat conduction in the theoretical explanation of his work. For experiments, he initially used voltaic piles, but later used a thermocouple as this provided a more stable voltage source in terms of internal resistance and constant potential difference. He used a galvanometer to measure current, and knew that the voltage between the thermocouple terminals was proportional to the junction temperature. He then added test wires of varying length, diameter, and material to complete the circuit. He found that his data could be modeled through the equation

$$x = \frac{a}{b + l},$$

where x was the reading from the galvanometer, l was the length of the test conductor, a depended only on the thermocouple junction temperature, and b was a constant of the entire setup. From this, Ohm determined his law of proportionality and published his results.

Ohm's law was probably the most important of the early quantitative descriptions of the physics of electricity. We consider it almost obvious today. When Ohm first published his work, this was not the case; critics reacted to his treatment of the subject with hostility. They called his work a "web of naked fancies" and the German Minister of Education proclaimed that "a professor who preached such heresies was unworthy to teach science." The prevailing scientific philosophy in Germany at the time asserted that experiments need not be performed to develop an understanding of nature because nature is so well ordered, and that scientific truths may be deduced through reasoning alone. Also, Ohm's brother Martin, a mathematician, was battling the German educational system. These factors hindered the acceptance of Ohm's work, and his work did not become widely accepted until the 1840s. Fortunately, Ohm received recognition for his contributions to science well before he died.

In the 1850s, Ohm's law was known as such, and was widely considered proved, and alternatives such as "Barlow's law" discredited, in terms of real applications to telegraph system design, as discussed by Samuel F. B. Morse in 1855.

While the old term for electrical conductance, the mho (the inverse of the resistance unit ohm), is still used, a new name, the siemens, was adopted in 1971, honoring Ernst Werner von Siemens. The siemens is preferred in formal papers.

In the 1920s, it was discovered that the current through an ideal resistor actually has statistical fluctuations, which depend on temperature, even when voltage and resistance are exactly constant; this fluctuation, now known as Johnson–Nyquist noise, is due to the discrete nature of charge. This thermal effect implies that measurements of current and voltage that are taken over sufficiently short periods of time will yield ratios of V/I that fluctuate from the value of R implied by the time average or ensemble average of the measured current; Ohm's law remains correct for the average current, in the case of ordinary resistive materials.

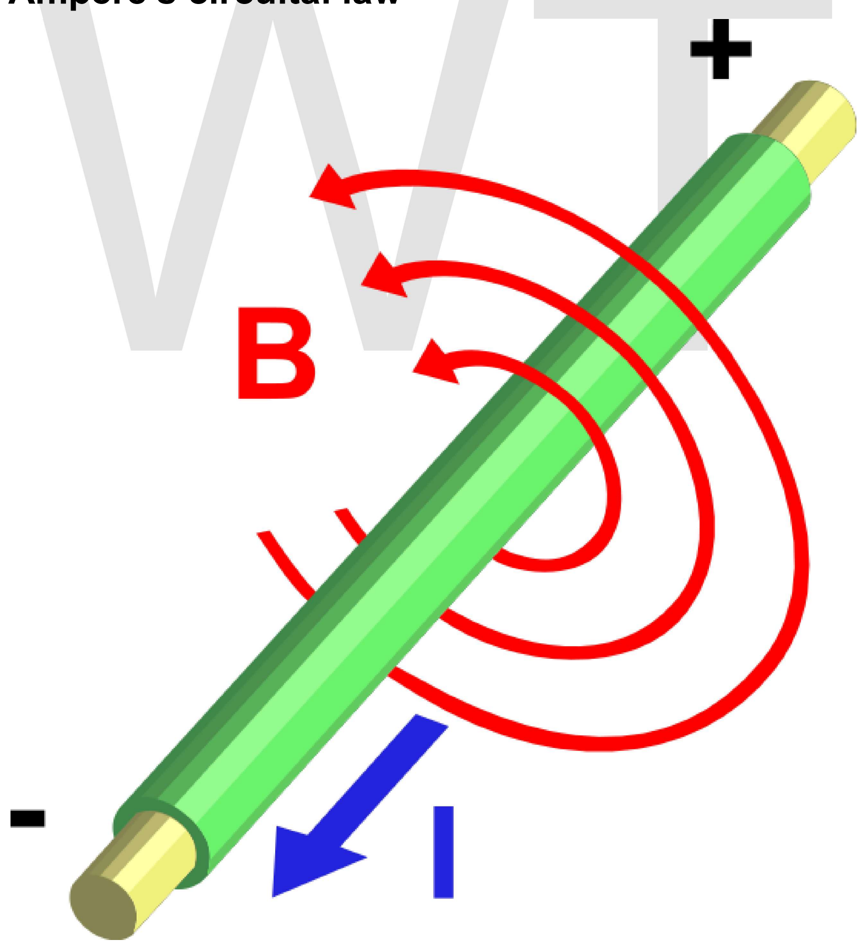
Ohm's work long preceded Maxwell's equations and any understanding of frequency-dependent effects in AC circuits. Modern developments in electromagnetic theory and circuit theory do not contradict Ohm's law when they are evaluated within the appropriate limits.

Chapter 13

Ampère's Circuital Law

In classical electromagnetism, **Ampère's circuital law**, discovered by André-Marie Ampère in 1826, relates the integrated magnetic field around a closed loop to the electric current passing through the loop. James Clerk Maxwell derived it again hydrodynamically in his 1861 paper On Physical Lines of Force and it is now one of the Maxwell equations, which form the basis of classical electromagnetism.

Original Ampère's circuital law



An electric current produces a magnetic field

In its historically original form, Ampère's Circuital Law relates the magnetic field to its electric current source. The law can be written in two forms, the "integral form" and the "differential form". The forms are equivalent, and related by the Kelvin–Stokes theorem. It can also be written in terms of either the **B** or **H** magnetic fields. Again, the two forms are equivalent.

Ampère's circuital law is now known to be a correct law of physics in a magnetostatic situation: The system is static except possibly for continuous steady currents within closed loops. In all other cases the law is incorrect unless Maxwell's correction is included (see below).

Integral form

In SI units (the version in cgs units is in a later section), the "integral form" of the original Ampère's circuital law is:

$$\oint_C \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 \iint_S \mathbf{J} \cdot \mathrm{d}\mathbf{S}$$

$$\oint_C \mathbf{H} \cdot \mathrm{d}\boldsymbol{\ell} = \iint_S \mathbf{J}_\mathrm{f} \cdot \mathrm{d}\mathbf{S}$$

or equivalently,

$$\oint_C \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \mu_0 I_\mathrm{enc}$$

$$\oint_C \mathbf{H} \cdot \mathrm{d}\boldsymbol{\ell} = I_\mathrm{f,enc}$$

where

$\oint_C$ is the closed line integral around the closed curve C;
B is the magnetic B-field in teslas;
H is the magnetic H-field in ampere per metre;
· is the vector dot product;
d**ℓ** is an infinitesimal element (a differential) of the curve C (i.e. a vector with magnitude equal to the length of the infinitesimal line element, and direction given by the tangent to the curve C, see below);
$\iint_S$ denotes an integral over the surface S enclosed by the curve C (see below; the double integral sign is meant simply to denote that the integral is two-dimensional in nature);
μ_0 is the magnetic constant;
$\mathbf{J}_\mathrm{f}$ is the free current density through the surface S enclosed by the curve C (see below);

J is the total current density through the surface S enclosed by the curve C, including both free and bound current (see below);
d**S** is the vector area of an infinitesimal element of surface S (that is, a vector with magnitude equal to the area of the infinitesimal surface element, and direction normal to surface S. The direction of the normal must correspond with the orientation of C by the right hand rule, see below for further discussion);
$I_{f,enc}$ is the net free current that penetrates through the surface S (see below);
I_{enc} is the total net current that penetrates through the surface S, including both free and bound current (see below).

There are a number of ambiguities in the above definitions that warrant elaboration.

First, three of these terms are associated with sign ambiguities: the line integral $\oint_C$ could go around the loop in either direction (clockwise or counterclockwise); the vector area d**S** could point in either of the two directions normal to the surface; and I_{enc} is the net current passing through the surface S, meaning the current passing through in one direction, minus the current in the other direction—but either direction could be chosen as positive. These ambiguities are resolved by the right-hand rule: With the palm of the right-hand toward the area of integration, and the index-finger pointing along the direction of line-integration, the outstretched thumb points in the direction that must be chosen for the vector area d**S**. Also the current passing in the same direction as d**S** must be counted as positive. The right hand grip rule can also be used to determine the signs.

Second, there are infinitely many possible surfaces S that have the curve C as their border. (Imagine a soap film on a wire loop, which can be deformed by blowing gently at it.) Which of those surfaces is to be chosen? If the loop does not lie in a single plane, for example, there is no one obvious choice. The answer is that it does not matter; it can be proven that any surface with boundary C can be chosen.

Differential form

By the Kelvin–Stokes theorem, this equation can also be written in a "differential form". Again, this equation only applies in the case where the electric field is constant in time; see below for the more general form. In SI units, the equation states:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$
$$\nabla \times \mathbf{H} = \mathbf{J_f}$$

where

$\nabla \times$ is the curl operator.

Note on free current versus bound current

The electric current that arises in the simplest textbook situations would be classified as "free current"—for example, the current that passes through a wire or battery. In contrast, "bound current" arises in the context of bulk materials that can be magnetized and/or polarized. (All materials can to some extent.)

When a material is magnetized (for example, by placing it in an external magnetic field), the electrons remain bound to their respective atoms, but behave as if they were orbiting the nucleus in a particular direction, creating a microscopic current. When the currents from all these atoms are put together, they create the same effect as a macroscopic current, circulating perpetually around the magnetized object. This magnetization current $\mathbf{J_M}$ is one contribution to "bound current".

The other source of bound current is bound charge. When an electric field is applied, the positive and negative bound charges can separate over atomic distances in polarizable materials, and when the bound charges move, the polarization changes, creating another contribution to the "bound current", the polarization current $\mathbf{J_P}$.

The total current density **J** due to free and bound charges is then:

$$\mathbf{J} = \mathbf{J_f} + \mathbf{J_M} + \mathbf{J_P} \ ,$$

with $\mathbf{J_f}$ the "free" or "conduction" current density.

All current is fundamentally the same, microscopically. Nevertheless, there are often practical reasons for wanting to treat bound current differently from free current. For example, the bound current usually originates over atomic dimensions, and one may wish to take advantage of a simpler theory intended for larger dimensions. The result is that the more microscopic Ampère's law, expressed in terms of **B** and the microscopic current (which includes free, magnetization and polarization currents), is sometimes put into the equivalent form below in terms of **H** and the free current only.

Shortcomings of the original formulation of Ampère's circuital law

There are two important issues regarding Ampère's law that require closer scrutiny. First, there is an issue regarding the continuity equation for electrical charge. There is a theorem in vector calculus that states the divergence of a curl must always be zero. Hence $\nabla\cdot(\nabla\times\mathbf{B}) = 0$ and so the original Ampère's law implies that $\nabla\cdot\mathbf{J} = 0$. But in general $\nabla\cdot\mathbf{J} = -\partial\rho/\partial t$, which is non-zero for a time-varying charge density. An example occurs in a capacitor circuit where time-varying charge densities exist on the plates.

Second, there is an issue regarding the propagation of electromagnetic waves. For example, in free space, where $\mathbf{J} = 0$, Ampère's law implies that $\nabla\times\mathbf{B} = 0$, but instead $\nabla\times\mathbf{B} = -(1/c^2)\ \partial\mathbf{E}/\partial t$.

To treat these situations, the contribution of displacement current must be added to the current term in Ampère's law.

James Clerk Maxwell conceived of displacement current as a polarization current in the dielectric vortex sea, which he used to model the magnetic field hydrodynamically and mechanically. He added this displacement current to Ampère's circuital law at equation (112) in his 1861 paper On Physical Lines of Force.

Displacement current

In free space, the displacement current is related to the time rate of change of electric field.

In a dielectric the above contribution to displacement current is present too, but a major contribution to the displacement current is related to the polarization of the individual molecules of the dielectric material. Even though charges cannot flow freely in a dielectric, the charges in molecules can move a little under the influence of an electric field. The positive and negative charges in molecules separate under the applied field, causing an increase in the state of polarization, expressed as the polarization density **P**. A changing state of polarization is equivalent to a current.

Both contributions to the displacement current are combined by defining the displacement current as:

$$\mathbf{J_D} = \frac{\partial}{\partial t}\mathbf{D}(\boldsymbol{r},\ t)\ ,$$

where the electric displacement field is defined as:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 \varepsilon_r \mathbf{E}\ ,$$

where ε_0 is the electric constant, ε_r the relative static permittivity, and **P** is the polarization density. Substituting this form for **D** in the expression for displacement current, it has two components:

$$\mathbf{J}_{\mathrm{D}} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t}.$$

The first term on the right hand side is present everywhere, even in a vacuum. It doesn't involve any actual movement of charge, but it nevertheless has an associated magnetic field, as if it were an actual current. Some authors apply the name displacement current to only this contribution.

The second term on the right hand side is the displacement current as originally conceived by Maxwell, associated with the polarization of the individual molecules of the dielectric material.

Maxwell's original explanation for displacement current focused upon the situation that occurs in dielectric media. In the modern post-aether era, the concept has been extended to apply to situations with no material media present, for example, to the vacuum between the plates of a charging vacuum capacitor. The displacement current is justified today because it serves several requirements of an electromagnetic theory: correct prediction of magnetic fields in regions where no free current flows; prediction of wave propagation of electromagnetic fields; and conservation of electric charge in cases where charge density is time-varying.

Extending the original law: the Maxwell–Ampère equation

Next Ampère's equation is extended by including the polarization current, thereby remedying the limited applicability of the original Ampère's circuital law.

Treating free charges separately from bound charges, Ampère's equation including Maxwell's correction in terms of the **H**-field is (the **H**-field is used because it includes the magnetization currents, so $\mathbf{J}_M$ does not appear explicitly, H-field and also Note):

$$\oint_C \mathbf{H} \cdot \mathrm{d}\boldsymbol{\ell} = \iint_S \left(\mathbf{J}_\mathrm{f} + \frac{\partial}{\partial t}\mathbf{D} \right) \cdot \mathrm{d}\mathbf{A}$$

(integral form), where **H** is the magnetic H field (also called "auxiliary magnetic field", "magnetic field intensity", or just "magnetic field", **D** is the electric displacement field, and $\mathbf{J}_\mathrm{f}$ is the enclosed conduction current or free current density. In differential form,

$$\nabla \times \mathbf{H} = \mathbf{J}_\mathrm{f} + \frac{\partial}{\partial t}\mathbf{D} \ .$$

On the other hand, treating all charges on the same footing (disregarding whether they are bound or free charges), the generalized Ampère's equation (also called the Maxwell–Ampère equation) is:

$$\oint_C \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \iint_S \left(\mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t}\mathbf{E} \right) \cdot \mathrm{d}\mathbf{A}$$

in integral form. In differential form,

$$\nabla \times \mathbf{B} = \left(\mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t}\mathbf{E} \right) \ .$$

In both forms **J** includes magnetization current density as well as conduction and polarization current densities. That is, the current density on the right side of the Ampère–Maxwell equation is:

$$\mathbf{J_f} + \mathbf{J_D} + \mathbf{J_M} = \mathbf{J_f} + \mathbf{J_P} + \mathbf{J_M} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \ ,$$

where current density $\mathbf{J}_D$ is the displacement current, and **J** is the current density contribution actually due to movement of charges, both free and bound. Because $\nabla \cdot \mathbf{D} = \rho$, the charge continuity issue with Ampère's original formulation is no longer a problem. Because of the term in $\varepsilon_0 \partial \mathbf{E} / \partial t$, wave propagation in free space now is possible.

With the addition of the displacement current, Maxwell was able to hypothesize (correctly) that light was a form of electromagnetic wave.

Proof of equivalence

Proof that the formulations of Ampère's law in terms of free current are equivalent to the formulations involving total current.

Ampère's law in cgs units

In cgs units, the integral form of the equation, including Maxwell's correction, reads

$$\oint_C \mathbf{B} \cdot \mathrm{d}\boldsymbol{\ell} = \frac{1}{c} \iint_S \left(4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} \right) \cdot \mathrm{d}\mathbf{S}$$

where c is the speed of light.

The differential form of the equation (again, including Maxwell's correction) is

$$\nabla \times \mathbf{B} = \frac{1}{c} \left(4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} \right).$$

Chapter 14

Biot–Savart Law

The **Biot–Savart law** is an equation in electromagnetism that describes the magnetic field **B** generated by an electric current. The vector field **B** depends on the magnitude, direction, length, and proximity of the electric current, and also on a fundamental constant called the magnetic constant. The law is valid in the magnetostatic approximation, and results in a **B** field consistent with both Ampère's circuital law and Gauss's law for magnetism.

Introduction

The Biot–Savart law is used to compute the magnetic field generated by a steady current, i.e. a continual flow of charges, for example through a wire, which is constant in time and in which charge is neither building up nor depleting at any point. The equation in SI units is

$$\mathbf{B} = \int \frac{\mu_0}{4\pi} \frac{I d\mathbf{l} \times \hat{\mathbf{r}}}{|r|^2},$$

or, equivalently,

$$\mathbf{B} = \int \frac{\mu_0}{4\pi} \frac{I d\mathbf{l} \times \mathbf{r}}{|r|^3},$$

where

I is the current,
d**l** is a vector, whose magnitude is the length of the differential element of the wire, and whose direction is the direction of conventional current,
B is the net magnetic field,
μ_0 is the magnetic constant,
$\hat{\mathbf{r}}$ is the displacement unit vector in the direction pointing from the wire element towards the point at which the field is being computed, and
$\mathbf{r} = r\hat{\mathbf{r}}$ is the full displacement vector from the wire element to the point at which the field is being computed.

The symbols in boldface denote vector quantities.

To apply the equation, you choose a point in space at which you want to compute the magnetic field. Holding that point fixed, you integrate over the path of the current(s) to find the total magnetic field at that point. The application of this law implicitly relies on the superposition principle for magnetic fields, i.e. the fact that the magnetic field is a vector sum of the field created by each infinitesimal section of the wire individually.

The formulations given above work well when the current can be approximated as running through an infinitely-narrow wire. If the current has some thickness, the proper formulation of the Biot–Savart law (again in SI units) is:

$$\mathbf{B} = \int \frac{\mu_0}{4\pi} \frac{(\mathbf{J}\, dV) \times \hat{\mathbf{r}}}{r^2},$$ or (equivalently), $$\mathbf{B} = \int \frac{\mu_0}{4\pi} \frac{(\mathbf{J}\, dV) \times \mathbf{r}}{r^3},$$

where dV is the differential element of volume and **J** is the current density vector in that volume.

The Biot–Savart law is fundamental to magnetostatics, playing a similar role to Coulomb's law in electrostatics. When magnetostatics does not apply, the Biot–Savart law should be replaced by Jefimenko's equations.

Forms

General

In the magnetostatic approximation, the magnetic field can be determined if the current density **J** is known:

$$\mathbf{B} = K_m \int \frac{\mathbf{J} \times \hat{\mathbf{r}}}{r^2} dV$$

where:

dV is the differential element of volume.

$K_m = \frac{\mu_0}{4\pi}$ is the magnetic constant

Constant uniform current

In the special case of a constant, uniform current **I**, the magnetic field **B** is

$$\mathbf{B} = K_m I \int \frac{d\mathbf{l} \times \hat{\mathbf{r}}}{r^2}$$

Point charge at constant velocity

In the case of a charged point particle **q** moving at a constant velocity **v**, then Maxwell's equations give the following expression for the electric field and magnetic field:

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0}\frac{1 - v^2/c^2}{(1 - v^2\sin^2\theta/c^2)^{3/2}}\frac{\hat{\mathbf{r}}}{r^2}$$
$$\mathbf{B} = \mathbf{v} \times \frac{1}{c^2}\mathbf{E}$$

where $\hat{\mathbf{r}}$ is the vector pointing from the current (non-retarded) position of the particle to the point at which the field is being measured, and θ is the angle between $\mathbf{v}$ and $\mathbf{r}$.

When $v^2 \ll c^2$, the electric field and magnetic field can be approximated as

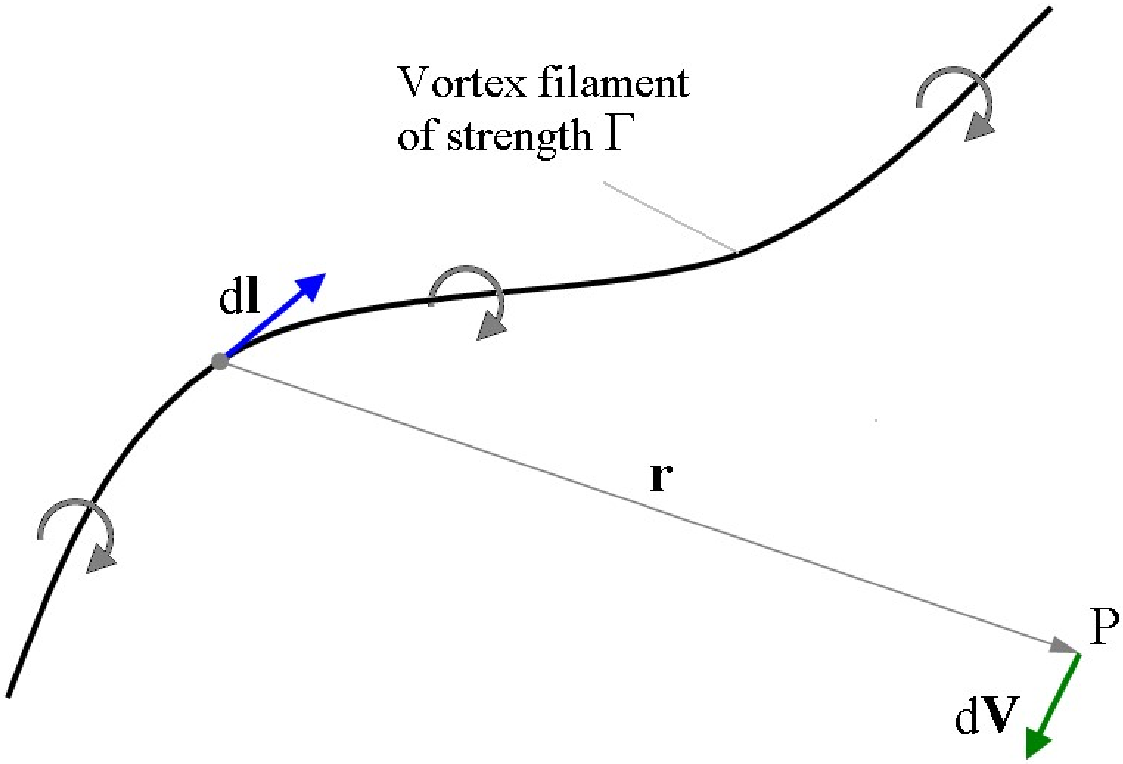
$$\mathbf{E} = \frac{q}{4\pi\epsilon_0}\ \frac{\hat{\mathbf{r}}}{r^2}$$
$$\mathbf{B} = \frac{\mu_0 q\mathbf{v}}{4\pi} \times \frac{\hat{\mathbf{r}}}{r^2}$$

These equations are called the "Biot–Savart law for a point charge" due to its closely analogous form to the "standard" Biot–Savart law given previously. These equations were first derived by Oliver Heaviside in 1888.

Magnetic responses applications

The Biot–Savart law can be used in the calculation of magnetic responses even at the atomic or molecular level, e.g. chemical shieldings or magnetic susceptibilities, provided that the current density can be obtained from a quantum mechanical calculation or theory.

Aerodynamics applications



The figure shows the velocity (dV) induced at a point P by an element of vortex filament (dL) of strength Γ.

The Biot–Savart law is also used in aerodynamic theory to calculate the velocity induced by vortex lines.

In the aerodynamic application, the roles of vorticity and current are reversed as when compared to the magnetic application.

In Maxwell's 1861 paper 'On Physical Lines of Force', magnetic field strength **H** was directly equated with pure vorticity (spin), whereas **B** was a weighted vorticity that was weighted for the density of the vortex sea. Maxwell considered magnetic permeability μ to be a measure of the density of the vortex sea. Hence the relationship,

(1) **Magnetic induction current**

$$\mathbf{B} = \mu \mathbf{H}$$

was essentially a rotational analogy to the linear electric current relationship,

(2) **Electric convection current**

$$\mathbf{J} = \rho \mathbf{v}$$

where **ρ** is electric charge density. **B** was seen as a kind of magnetic current of vortices aligned in their axial planes, with **H** being the circumferential velocity of the vortices.

The electric current equation can be viewed as a convective current of electric charge that involves linear motion. By analogy, the magnetic equation is an inductive current involving spin. There is no linear motion in the inductive current along the direction of the **B** vector. The magnetic inductive current represents lines of force. In particular, it represents lines of inverse square law force.

In aerodynamics the induced air currents are forming solenoidal rings around a vortex axis that is playing the role that electric current plays in magnetism. This puts the air currents of aerodynamics into the equivalent role of the magnetic induction vector **B** in electromagnetism.

In electromagnetism the **B** lines form solenoidal rings around the source electric current, whereas in aerodynamics, the air currents form solenoidal rings around the source vortex axis.

Hence in electromagnetism, the vortex plays the role of 'effect' whereas in aerodynamics, the vortex plays the role of 'cause'. Yet when we look at the **B** lines in isolation, we see exactly the aerodynamic scenario in so much as that **B** is the vortex axis and **H** is the circumferential velocity as in Maxwell's 1861 paper.

For a vortex line of infinite length, the induced velocity at a point is given by

$$v = \frac{\Gamma}{2\pi r}$$

where

Γ is the strength of the vortex
r is the perpendicular distance between the point and the vortex line.

This is a limiting case of the formula for vortex segments of finite length:

$$v = \frac{\Gamma}{4\pi r} \left[\cos A - \cos B\right]$$

where A and B are the (signed) angles between the line and the two ends of the segment.

The Biot–Savart law, Ampère's circuital law, and Gauss's law for magnetism

Here is a demonstration that the magnetic field **B** as computed from the Biot–Savart law will always satisfy Ampere's circuital law and Gauss's law for magnetism.