Electromagnetic Induction

The relationship between the physics and the small-signal electrical behaviour of transformers

By David W. Knight*

Version 0.10 provisional, 27th May 2012 (The most recent version of this document can be obtained from the author's website: http://www.g3ynh.info/).

dave@camerasunderwater.co.uk

* Ottery St Mary, Devon, England.

Table of Contents

Preface

Author's note:

This document is a work in progress.

An issue under consideration is whether inductance shortfall due to field inhomogeneity should be represented as an operator acting on the number of turns, or as a finite reluctance on the external magnetic path. The turns operator approach has been used in this version and is mathematically consistent for all of the applications considered. The path redefinition approach is another possibility, but needs to be made capable of distinguishing between inductance which fails to materialise (self-inductance shortfall) and inductance which fails to couple (mutual inductance shortfall). It may be that in future revisions, an alternative way of allowing for field inhomogeneity will be discussed.

For the air-cored solenoid, the path length is defined as the length of the coil. This implies that the reluctance of the path for flux loops outside the coil (i.e., the path in series with the path within the coil) is zero. This is strictly correct when the coil is infinitely long and the field is homogeneous. When the coil is not infinitely long, the inductance is reduced because the field becomes inhomogeneous and flux linking is incomplete.

Should this be interpreted as a reduction in the effective number of turns, or as a finite reluctance in the segment of path outside the coil (the gap)? In the former case, the correction parameter might also be regarded as a shunt reluctance. In the latter case the correction is a series reluctance, and Nagaoka's coeff. becomes part of A_L . The latter may be more in keeping with standard practice; but was previously rejected by this author on the basis that it appears to allow for flux shortcuts (incomplete linkage, which is a form of shunting or bypassing) by increasing the effective path length. It also makes no clear distinction between inductance which fails to materialise and inductance which fails to couple to other inductances.

Glossary

Quantity	Symbol	Definition	Units
Poynting vector	$\boldsymbol{P}$	$= \boldsymbol{E} \times \boldsymbol{H}$	[Watts / metre²]
Electric field	$\boldsymbol{E}$		[Volts / metre]
Magnetic field	$\boldsymbol{H}$		[Ampere turns / metre] , [A turns / m]
Field strength	H	$= \mathrm{F} / \ell$	[A turns / m]
Magnetomotive force, MMF, Ampere turns	F	$= \tilde{\mathrm{N}}\,\mathrm{I}$	[Ampere turns] , [A turns]
Instantaneous current	I		[Amperes] , [A]
Flux density (magnetic induction)	$\boldsymbol{B}$ B	$= \mu\boldsymbol{H}$ $= \mu H = \Phi/\mathrm{A}$	[Tesla] , [T] , [gauss $\times 10^4$], [Wb/m²]
Permeability Complex permeability	μ $\boldsymbol{\mu}$	$= \mu_0\ \mu_r$ $= \mu_0\ \boldsymbol{\mu_r}$	[Henrys / metre] , [H/m] [Webers/Ampere metre] , [Wb/A m]
Vacuum permeability	μ_0	$= 4\pi \times 10^{-7}$	[Henrys / metre] , [H/m]

Relative permeability	μ_r $\boldsymbol{\mu_r}$		
Magnetic flux	Φ	$= A\,B$ $= \tilde{N}\, I\, A_L$	[Webers] , [Wb] , [maxwells $\times 10^8$]
Flux linkage	Λ	$= \tilde{N}\, \Phi$ $= \tilde{N}^2 A_L\, I$	[Weber turns] , [Wb.turns]
Effective No. of turns	Ñ	$= N \sqrt{k_H}$	[turns]
Number of turns	N		[turns]
Linkage effy.	$\sqrt{k_H}$		
Current sheet linkage effy.	$\sqrt{k_L}$	√(Nagaoka's coeff.)	
Path area	A		[m²]
Path length	ℓ		[m]
Self inductance	L	$= \Lambda / I$ $= \tilde{N}^2 \mu A / \ell$	[Wb turns / A] , [Henrys] , [H]
Inductance factor (magnetic conductance)	A_L	$= \mu A / \ell$	[Wb / A turns] , [H / turns²]
Reluctance	S	$= 1/A_L$	A turns / Wb
Coupling coefficient	k	$= \sqrt{(k_1 k_2)}$ $= M / \sqrt{(L_1 L_2)}$	
Voltage shortfall factor	k_1 k_2	$= \Phi_{12} / \Phi_1$ $= \Phi_{21} / \Phi_2$	
Mutual inductance	M	$= \Lambda_{12} / I_1$ $= k \sqrt{(L_1 L_2)}$	[H]
Inductive reactance	X_L	$= 2\pi f\, L$	[Ohms] , [Ω]
Mutual reactance	X_M	$= 2\pi f\, M$	[Ohms] , [Ω]

cgs magnetic units	**Symbol**	**Unit**	**Conversion**
Field strength	*H*	Ørsted, Oe	1 Oe = 1000 / 4π A turns/m
Flux	Φ	maxwell (line)	1 maxwell = 10^{-8} Wb
Flux density (magnetic induction)	*B*	gauss	1 gauss = 10^{-4} T 1 Tesla = 10 000 gauss

1. Introduction

When engaged in the business of designing electrical circuits, is possible to come to an understanding of inductors without giving much thought to the magnetic fields which give rise to their properties. Such neglect of fundamentals is no longer possible when we come to the subject of transformers however (or electrical machines for that matter, but they are not the subject of this monograph). For an ability to analyse circuits involving transformers of any type (not just ideal ones), or to quantify the effects of spurious magnetic coupling, it becomes necessary to relate the physics of fields to the mathematical conventions of circuit analysis. The principles involved are fairly straightforward; but, for reasons which warrant careful consideration by anyone attempting to teach the subject, the reconciliation appears to be an area of some difficulty.

A rigorous approach to the subject of electricity involves starting with Maxwell's equations and working out everything from there. Obviously, there is room for a more accessible theory when dealing with lumped-element circuits, and the mathematics of phasors fulfils that requirement. There is a caveat however; which is that the definitions and conventions of phasor analysis need to be consistent with the underlying electromagnetic phenomena. That this issue receives little consideration can be confirmed by examining the various elementary textbooks which purport to cover the subject of magnetic induction. A knowledgeable reader who cares to follow the discussions on offer will soon discover inconsistencies (and may experience a disconcerting reminder of what it felt like to be a beginner).

One problem lies with the definition of flux linkage, which falls victim to casual amendment when texts originally written using cgs (centimetre, gram, second) units are updated to use SI (metre, kilogram, second) units. In particular, in searching through a number of sources, it was found that none of the SI publications mentioned field non-uniformity corrections (Nagaoka's coefficient, etc.). This omission precludes the giving of self-consistent definitions for magnetic flux, inductance and reluctance (reciprocal inductance factor), and so confronts the reader with symbols which have the strange habit of changing their meanings according to context.

Notwithstanding the questionable definition of basic quantities, there is also a strong tradition of confusion over phase relationships and algebraic signs. This problem arises for two reasons. Firstly: flux linkage is habitually transformed as a magnitude, but is used to derive quantities which require a sign[1]. Secondly: the concept of back EMF (which arises from use of the compensation theorem) is not applicable to closed conservative systems such as circuit models. Such vector inconsistencies can be seen in diagrams which are obviously incorrectly labelled; and result in the need for guesswork when attempting to deduce the phase relationships which occur in inductively-coupled systems.

Then there are issues which do not relate to mathematical error, but are deleterious to understanding nonetheless. The main one is that there is no consensus regarding whether 'turns' belong in the system of magnetic units. The upshot is that some authors delete 'turns' wherever it occurs, presumably to divert students from the vice of including numerical factors in dimensional analysis. It seems probable however, that readers *will* be able to grasp the idea that turns do not have the fundamental status of metres, kilograms and seconds; and that people may be prepared to risk perdition in the interests of clarifying (say) the difference between Webers and Weber turns.

A critique of the way in which magnetic induction is typically presented could go on to considerable length, but we will confine ourselves here to one last cause for complaint. That is the tendency to assume that there is no such thing as a transformer having active networks connected to more than one winding. The disease usually begins with a statement to the effect that: "a transformer has a primary winding and one or more secondary windings"; despite the fact that the transformer itself has no opinion on such matters. The designation of windings comes purely from external circuit considerations, and imposes restrictions on the subsequent analysis. Hence, for a

1 **A polarity for flux-linkage**, J. Fiennes and C. R. de Souza, International Journal of Electrical Engineering Education, ISSN: 0020-7209, Volume 38 Issue 3, July 2001, p256-259.

mastery of transformers which displays ambition beyond the ability to design mains power supplies, the theory needs to emphasise the essentially bi-directional nature of inductive coupling. This is not to say that there is any objection to the use of the terms 'primary' and 'secondary'; but that we can only say: 'when a transformer is wired conventionally (i.e., with ports chosen to have DC continuity), and provided that there is only one port connected to an active network; the winding connected to the active network may be called the primary, and all of the others can be called secondaries'. If there are active networks connected to more than one winding, then the windings are neither primary nor secondary, and the transformer is called a 'hybrid'.

The list of issues given above will perhaps go some way towards explaining how transformers come to be regarded as incomprehensible beyond the 'ideal case with lumped parasitics' model. The challenge, of course, is to cover the subject in a manner which addresses those issues. This warrants a somewhat oblique approach, not least because it must convince those who have been taught things differently. In particular, in addition to giving the essential information, it seems necessary to demonstrate the following points:

- The definition of inductance in documents using the SI system requires flux linkages to be counted in a manner which is equivalent to the method used by Maxwell and Faraday.
- Circuit analysis parameters derived from field vectors are pseudoscalar, not scalar; i.e., they change sign when the direction of the parent vector is reversed.
- Contrary to the theory given in certain radio textbooks; voltage-shortfall reciprocity is not an inherent property of coupled inductors; i.e., a 1:1 transformer can have different voltage ratios when the roles of primary and secondary are swapped.
- The extraordinary range or technical possibilities offered by the transformer can be appreciated by thinking of it as a device which allows the impedance of one inductor to be modified (almost arbitrarily) by the time-varying magnetic field produced by another inductor.

Some calculus, of course, is unavoidable in dealing with a subject such as this; but the transition from fields to phasors, with pseudoscalars intact, is nowhere near as arduous as an approach which tries to resolve mathematical inconsistencies by making use of dubious physical arguments. Before we move on to the details however, we will address the most glaring oversight in nearly every discussion of magnetically-coupled devices, which is that the information will be incomprehensible to anyone who does not have a Maxwellian understanding of electricity.

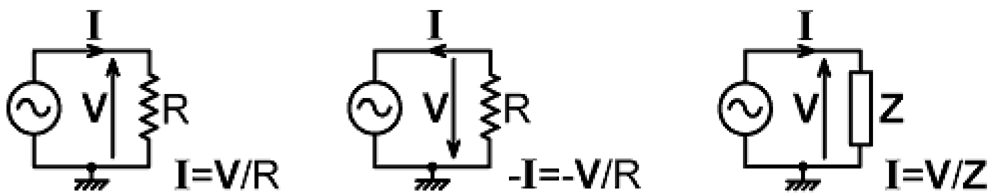
2. Maxwellian Electricity

The phasor approach to circuit analysis can be seen as a set of techniques which allows the empirical laws of DC electricity to be adapted to deal with AC electricity. Phasor analysis manages to be both straightforward and extraordinarily useful, but it has the notable disadvantage that it can encourage bad habits of thought. In particular, it allows the conception that electricity is a fluid which travels through conductive materials; a physical picture which might have seemed promising in the early days of electrical discovery, but which has subsequently turned out to be wrong.

The idea that electrical energy is carried by electrons flowing through wires cannot even account for DC electricity. The reason has to do with the fact that electrons drift through conductors in DC circuits with an average velocity of the order of a few mm/sec, and electrons have very little mass. This means that the transfer of momentum due to electron movements is insufficient to account for the power which a generator delivers to a load. Phasor theory does not concern itself with such issues; but even when strict energy-accounting principles are not applied, accepted techniques of circuit analysis remain inconsistent with electron transport concepts.

The obvious way in which phasor analysis departs from the primitive DC theory is that it involves drawing current direction arrows which point through capacitors. Such arrows serve to give algebraic signs to the circuit parameters, just as they do in DC circuit analysis; but they can no longer be imagined to represent migrating mass. Thus, since AC theory is more general than DC theory (and includes DC theory as a special case), we must conclude that an 'electric current' is not as envisaged when the term was coined.

Power in electrical circuits is usually regarded as a scalar quantity, i.e., it is represented by a positive real number and cannot be made negative by swapping the generator terminals. It still carries a sense of direction however, this being the thermodynamic principle that, on average, energy flows from a source to a sink. Power is, of course, a product of voltage and current; and so it follows that the voltages and currents used in circuit analysis must be defined in such a way that power flows from generator to load. This rule dictates the interpretation of the arrow conventions used in the diagrams below:



In the left-hand diagram, it is assumed that electric potential increases on moving from a reference point to some other part of the circuit. In this case, the reference is taken to be the ground potential; but the choice is, in principle, arbitrary. Hence the voltage arrow shows the assumed direction of increase from zero to a finite potential, as defined purely for the purpose of circuit analysis. Once that definition has been made, the direction of the current is automatically fixed according to the convention that power is positive when flowing into a positive resistance.

When a generator is connected to a pure resistance, the voltage is in phase with the current. In that case, since both phasors point in the same direction and the choice of reference phase is arbitrary; the imaginary parts of both the voltage and the current can be set to zero (this is how AC theory maps onto DC theory). Effectively, the phasors have dropped a dimension and turned into one-dimensional objects; but note that *they have not turned into scalars*. They belong instead to a special class of vectors called *pseudoscalars*, which are distinct from scalars in that they change sign under the operation of 180° rotation. This special property is illustrated in the middle diagram, where a redefinition of the voltage via phase reversal has forced a reversal in the definition of the current direction. The outcome, of course, is that power remains positive, the direction of energy flow being dictated not by the chosen signs of the voltage and current, but by the sign of the resistance.

The right-hand diagram represents the more general situation in which a generator applies a voltage to an impedance. Now, although power will be transferred provided that the impedance has a resistive component; we also have the theoretical possibility that the impedance may be purely reactive, in which case the net power transfer will be zero. We nevertheless adopt the relative sign conventions for voltage and current on the basis that power will flow into the impedance if the impedance has a positive resistive component. The choice remains correct because the mechanism which decides how power will flow is built into the definition of impedance itself. That mechanism, of course, depends on the properties of the 90° rotation operator **j**.

If the load impedance is purely reactive we have:

$\mathbf{Z} = \mathbf{j}X$

where the reactance, X, is (by convention) positive if inductive and negative if capacitive. Thus:

$\mathbf{I} = \mathbf{V}/(\mathbf{j}X) = -\mathbf{j}\mathbf{V}/X$

A pure reactance rotates the phase of **I** by ±90° relative to **V**. The effect of this rotation is that (depending on the instantaneous signs of the voltage and current) sometimes power flows from generator to load, and sometimes it flows from load to generator; and the net power transfer averaged over a whole cycle of the AC waveform is zero.

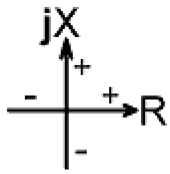
In general, the average power delivered to an impedance is given by the scalar (dot) product of voltage and current, i.e.:

$P = \mathbf{V}\bullet\mathbf{I} = |\mathbf{V}|\ |\mathbf{I}|\ \text{Cos}\varphi$

where Cosφ is known as the *power-factor* and, for an impedance $\mathbf{Z} = R + \mathbf{j}X$

$\varphi = \text{Arctan}(X/R)$

The dot product, according to generally accepted convention, only ever gives a positive result. Hence its use may now seem slightly paradoxical given that we have admitted that instantaneous power can flow the wrong way. We could try to wriggle out of that problem by saying that the formula only relates to average power, but actually, it has to be admitted that the scalar product is not quite as scalar as it might seem. It *will* give a negative answer if the power-factor is negative.



Recall that the impedance of a passive network can only ever lie in the right-hand side of the Z-plane. No true passive network can contain a negative resistance because energy produced from nothing violates the classical principle of causality. There is however, nothing to prevent energy from flowing out of an active network; and so we could, should we happen to be feeling sufficiently misanthropic, devise a theory of circuits using negative resistances instead of generators.

In fact, negative resistances do crop-up in circuit analysis from time to time. They occur when a network 'unexpectedly' turns out to be active. This can happen, for example, when modelling antenna systems with multiple feed-points, the negative resistance at one input being due to energy originating from another input. The same can occur in any set of mutually coupled two-terminal elements (such as a transformer); but in that case, the situation is not unexpected, and we can usually analyse the problem more sensibly by redefining auxiliary energy sources as generators.

Still, it has to be said that negative resistance is not forbidden. It occurs when a two-terminal network defined as an impedance has energy flowing into it from outside the system under consideration (provided, of course, that the energy input exceeds the amount required to cancel the true resistance of the network). The resulting impedance then lies in the *left-hand side* of the Z-plane, and the power in relation to that impedance is *negative*. Thus it might seem that the **V I** scalar product rule for power has let us down in terms of generality; but in fact, there is an interpretation which would be incorrect for general n-dimensional vector theory, but which must be valid in two-dimensional phasor theory because it resolves the power-flow paradox.

The dot product of general vector theory is obtained using the acute angle (θ say) between two vectors lying in n-space. There is also however an obtuse angle (180 - θ) which suggests an alternative solution but is always ignored by convention. Now recall that:

$\text{Cos}(180 - \theta) = -\text{Cos}\theta$

The obtuse-angle dot product is the negative of the ordinary dot product. Hence, in phasor theory, we must conclude that the **V I** dot product should be taken using the actual phase angle φ of the impedance **Z**=**V**/**I**. Then, if |φ| is greater than 90°, as will occur if the resistive component of **Z** is negative, the power 'flowing into' **Z** will be negative as required.

Thus it transpires that power, in phasor theory, although normally considered to be positive, is in fact pseudoscalar. It is not a true vector, but neither is it a magnitude, and it can become negative in some circumstances. Armed with that information, we can now return to the question: "what is an electric current?"; but we cannot give a satisfying answer just yet. The best we can say so far is that 'current is that which has to be defined in conjunction with voltage in order to account for electrical power'. It transpires that, from within the horizons of phasor theory, we do not know what electricity is. That means that the phasor theory, and the DC theory it supersedes, are both incomplete.

The reason why phasor theory cannot answer fundamental questions, is that it is a projection or 'degenerate form' of a higher-dimensional theory. Just what has been lost in projection can be deduced by asking the question: "where exactly on a circuit diagram are the spatial dimensions and coordinates of the components?" It is easy to forget that circuit theory is purely topological. It assumes that all spatial dimensions tend to zero, which is why it is only valid when the wavelength at the analysis frequency is much greater than the length of any energy transmission path.

The parent theory which does answer the question "what is electricity?" is, of course, James Clerk Maxwell's electromagnetic theory. This was presented in early form in 1864, and in a more mature form in 1873. To a 19th Century audience expecting some kind of confirmation of the 'fluid in transit' idea, the change in perception was radical, and for some, impossible to digest. Maxwell established that electricity is an invisible form of light. The difference between it and visible light is determined only by the frequency; and the empirical rules of electrical circuit building turn out to be those which persuade light of very long wavelength to cling to the *outside surfaces* of electrical conductors.

Maxwell's extraordinary contribution to physics was to postulate the existence of a type of electric current (which he called "*displacement current*") which needs no charges in order to flow. It is often said that he made this deduction by noticing that the then known laws of electricity were in violation of the 'principle of local conservation of charge'; but that is a modern re-interpretation. A more plausible route to the discovery lies in the fact that Maxwell was pioneer and advocate of a technique known as *dimensional analysis*. He had, at his disposal, fairly good measurements of the *permeability of free space*, μ_0 (obtained from the force between current-carrying conductors) and the *permittivity of free space*, ε_0 (obtained from the force between charged bodies). What he is likely to have noticed at an early stage therefore, is that the reciprocal of the geometric mean of μ_0 and ε_0 has dimensions of velocity, and that the velocity so obtained is the same as the speed of light, i.e.:

$$c = 1 / \sqrt{\mu_0 \varepsilon_0}$$

Hence it is conceivable that Maxwell realised first that electrical energy is light; and then set about the task of reconciling the known electrical laws with that conviction. On a historical point incidentally, notice that the unification of electricity and light is not the same as the hypothesis that light is electromagnetic radiation. Maxwell attributed the latter to Faraday, who had suggested it in 1846 as a result of studies of the effect of a magnetic field on the polarisation of light passing through rods of diamagnetic glass.

Maxwell produced a complete set of equations which describe the strengths and directions of lines of electric and magnetic force at any point in relation to an electrical system, i.e., a description of the electromagnetic field. With the inclusion of the displacement current to maintain conservation of charge; it then becomes possible to delete all terms relating to physical matter and still have energy present. This energy exists by virtue of continuous exchange between the electric and magnetic fields, exactly according to the ordinary laws of electricity; i.e., a changing magnetic field produces an electric field, and a changing electric field produces a magnetic field; the two fields being represented as vectors which point at right angles. In the absence of matter however, the cyclical variation of the electric field is always exactly in phase with the variation of the magnetic field (from the viewpoint of a stationary observer), and to keep the energy exchange process in balance, the energy must propagate in a direction at right angles to both fields at velocity c.

On the matter of this phase relationship incidentally, it may be helpful to identify the source of a common misconception. In the last years of his life, Maxwell had only one student of note; and that was John Ambrose Fleming, now controversially associated with the invention of the thermionic diode. In 1908, Fleming published his 'Elementary Manual of Radio Telegraphy', in which he described the propagating wave as having the maximum of one field corresponding to the zero of the other[2]. This blunder, from an influential writer who had sat at the feet of the great Maxwell, is an enduring source of confusion. The offending passage was corrected in the second edition of the

2 **The genesis of the thermionic valve**. G W O Howe, IEE Lecture Nov. 1954.

book, but the damage was done and the error found its way into other books to muddle understanding for generations. The reality is that the energy travels at the speed of light and so, if we could observe the fields from the viewpoint of the energy, they would indeed be 90° out of phase. To an ordinary observer however, the electric and magnetic components of a radiation field are perfectly in phase, as is required to make the impedance of free-space resistive (i.e., Fleming's picture of the electromagnetic wave was not consistent with the concept of radiation resistance).

The important inference is that propagating energy will be found to have its electric field in phase with its magnetic field; and the velocity is a constant independent of the physical frame of reference. This is the 'relativistic' property of Maxwell's equations, which eventually led Einstein to the Special Theory of Relativity. For our purposes however, we simply note that this is analogous to the requirement that the voltage must be in phase with the current at any point in a circuit where energy dissipation is the only process occurring. The analogy becomes even clearer when we note that the units of the electric field ***E*** are *Volts per metre*, and the fundamental units of the magnetic field ***H*** are *Amperes per metre*. In a degenerate version of electromagnetism, the spatial dimensions are all replaced by unity; so that the electric field is replaced by voltage, and the magnetic field is replaced by current. The product of voltage and current has units of *energy per unit of time*, or power. The product of ***E*** and ***H*** has units of *power per unit area*, which is a measure of *illumination*.

Hence, an electrical circuit is a structure which causes electromagnetic energy of very long wavelength to follow wires and converge upon resistances. Another part of the story we already know, of course, is that if the structure becomes large in relation to wavelength, then some of the electricity escapes and propagates off into space. The question we have yet to answer however is: "if the free electrons in the wires do not carry the energy, then what exactly do they do?" The answer is that they modify the refractive index of space in the vicinity of conductors, causing energy in transit to be steered into the region just outside, and to some extent slightly inside, the conducting surface.

A problem which we may have now, is that electricity is only the second most badly taught subject in pre-university physics. The worst taught subject is that of optical refraction, which is generally explained in terms of light slowing down. The speed of light in vacuum is the same for all observers, and since matter consists mainly of empty space, there is no other conceivable medium. The apparent velocity can be less than or greater than c however, as a result of interference effects. Light is actually steered or 'refracted' by partially cancelling it in one place and augmenting it in another. The steering fields for electrical conduction are mainly provided by the free electrons in the conductors. What happens is that the electrons oscillate in response to the alternating electric field of a passing electromagnetic wave. This constitutes absorption of energy, but a collection of free charges will re-radiate most of that energy almost immediately, albeit with a slight shift of phase. The process of absorption and near-simultaneous re-emission is called 'scattering'. Once an EM wave enters a region rich in scattering objects, it very quickly loses its identity and is replaced by the sum or 'superposition' of all of the scattered waves. It is the superposition of incident and scattered waves which describes the way in which energy flows through the system; a resultant wave (which can be thought of as a contiguous vector path traced through the overall field) having the ability to negotiate bends, and appearing to travel at a velocity (called the *wave velocity* or the *phase velocity*) which differs from c. Not all of the energy is scattered back incidentally. Some of it is absorbed because the moving electrons can donate energy to other processes occurring in the body of the conductor; i.e., there is some loss due to 'resistance', but that is the price we pay for the steering service.

So now we have a picture of electricity which is very different from the 'magical fluid' idea which prevails in the popular imagination. We need to make a circuit in order to obtain energy from a generator, not because matter must flow from one terminal to the other, but because energy transport requires the combination of an electric field and a magnetic field. In the case of household electricity; the utility company provides the electric field, and the consumer provides the magnetic field by closing the circuit. A further possible manipulation is then to correct the power factor, so that the maximum amount of energy flows to the desired destination. It is also possible to send energy back into to distribution system, by using a local generator to shift the load impedance into the left-hand side of the Z-plane.

Through the logical contortions involved in the process of discovery, a scientific culture is bound to develop many terms which all turn out to mean 'light'. Thus we have: 'gamma rays', 'X-rays', 'radio', 'electromagnetic energy', 'electricity'; and several others. The distinction between electricity and light should be regarded primarily as a technological matter; i.e., the nomenclature changes at the shifting boundary at which it becomes difficult to manufacture structures which are small in comparison to wavelength. The scientific importance of electricity is that it leads to the understanding that light can be decomposed into electric and magnetic fields, which have a quasi-independent existence in the sense that they can be separately detected and manipulated. The technology of electricity is that which allows the relative phases of the two fields from a given energy source to be controlled, so that energy can be made to flow in some desired manner.

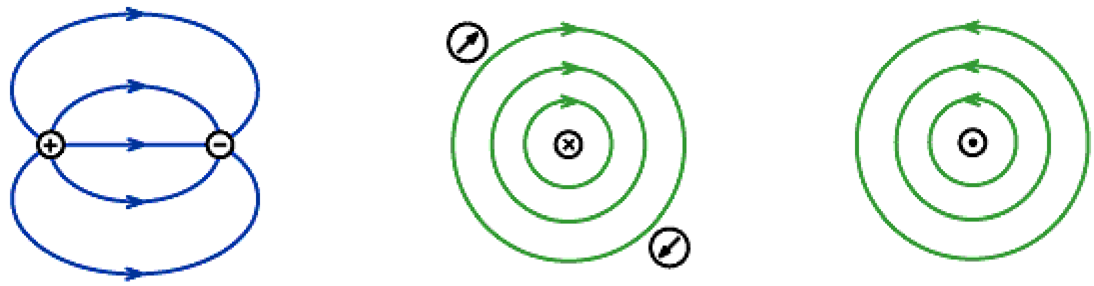
3. Fields

The 'electrons moving through wires' conception of electricity causes magnetism to be regarded as a peripheral phenomenon; which means that an understanding of magnetism usually requires a simultaneous re-appraisal of everything which has previously been learned about electricity. Hence the need to make the point that, if voltage is to be regarded as electric potential, then current must be regarded as magnetic potential; and the corresponding fields have equal status in the transmission of energy. The next step therefore is to visualise electricity in terms of its fields.

In physics, a field is defined as 'a physical quantity which can take on different values at different points in space and time. The field itself is independent of the system of co-ordinates (it stays the same regardless of how we choose to define a point within it) and its value at a given point may be described by a scalar, a vector, or, in general, a tensor (which is a mathematical object having an arbitrary number of variable attributes, not just magnitude and direction).

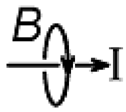
If the medium is homogeneous and isotropic (i.e., on average, the same throughout and the same in all directions), the electric and magnetic fields can be treated as vectors; i.e., they can be defined by stating the magnitude and direction at any given point. It is usual to conceptualise vector fields using the analogy of fluid motion; i.e., each point in the field is assigned an arrow indicating the direction of flow, and a number indicating the flow rate. Thus we find ourselves using the old-fashioned terms "electric flux" and "magnetic flux"; although it must be stressed that an electric or magnetic field, on its own, does not represent a transportation process. Instead, the ***E*** and ***H*** fields are represented as vectors so that their interaction can be represented by an operation called 'vector multiplication', which produces a new vector at right angles to the original two. It is this vector product (or 'cross product') which represents a true flux, it being the illumination field with units of energy per unit time per unit area. Notice, incidentally, that this implies that the ***E*** and ***H*** fields can be treated together as a tensor field; which is known as the *Maxwell bivector* or *electromagnetic tensor*. For many purposes however, it is sufficient to deal with the ***E*** and ***H*** fields separately; albeit while keeping the notion of energy flow in mind.

The electric and magnetic fields can be visualised using Faraday's "lines of force"; as demonstrated by the diagrams below:



The left-hand diagram is a two-dimensional representation of the electric field as it might exist between two charged particles or spheres, or between two conductors seen in cross-section. There is, of course, a line for any arbitrarily chosen point in the field, but the drawing shows only a few of them to give the idea. The important property is that the field lines emerge almost perpendicular to a conducting surface and follow curved curved paths between points of high and low potential. Note, incidentally, that although the field lines are often called 'lines of force', the field does not have the Newtonian dimensions of force (mass × acceleration). The lines represent force only in the loose sense, that a positively charged particle in the field will be repelled by the (+) electrode and attracted to the (-) electrode and will be accelerated in the direction of the field vector.

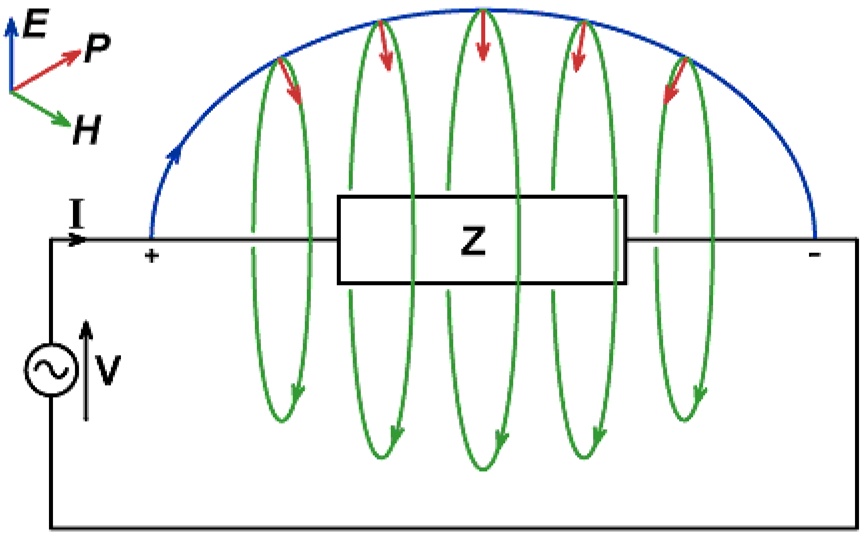
The middle diagram depicts the magnetic field around a wire when the instantaneous current of moving charges (which by convention flows from + to -) is flowing away from the observer. The right hand diagram shows the field when charges are moving towards the observer. The direction of charge migration is indicated by a cross or a dot drawn inside the conductor, the cross representing the tail fins of an arrow moving away, and the dot representing the point of an arrow approaching. The convention that the magnetic field lines rotate in a clockwise direction when positive charges are moving away is known as 'Maxwell's corkscrew rule', or alternatively as 'the right-hand rule' of electromagnetism. Electrons, of course, move contrary to the arrow convention, but nothing was known of their existence in Maxwell's lifetime. Once again, the dimensions of the field [Amperes per metre] do not correspond to Newtonian force, but the direction of the 'magnetic flux' indicates the sense of the force which will be exerted on a compass needle. The North-seeking pole of a magnet will be repelled by field lines coming towards it and attracted by the field lines going away from it. Hence, if the current is large enough to overcome the Earth's magnetic field, a compass placed near the wire will point in the direction of the field vector.



The diagram below shows how the electric and magnetic fields combine to make electromagnetic energy. In this case we consider only a single electric field line (and all of the others are left to the imagination). The key at the top left gives the direction of the vector cross product:

$\boldsymbol{P} = \boldsymbol{E} \times \boldsymbol{H}$

where $\boldsymbol{P}$ is a field representing the energy per unit time flowing through an infinitesimal area, and is known as the *Poynting Vector*. Notice incidentally, that the terms 'field' and 'vector' tend to be used interchangeably. Strictly, a vector is a list of numbers (e.g., magnitude in each co-ordinate direction), whereas a vector field is a list of equations which give a vector when the co-ordinates of a point are put in.



The presence of an impedance in the circuit gives rise to a potential difference, which causes electric field lines to curve from one side of the impedance to the other. The diagram shows an instant when both ***E*** and ***H*** are positive according to the convention of the vector cross product, in which case energy flows into the impedance. If the impedance is purely reactive, the energy will be stored. If the impedance is purely resistive (a black object in optical terms), then the energy will be converted into heat or transduced into useful work. If the impedance represents the terminals of another network, then the energy will be transported away to a place where its fate can also be determined using fields.

The basic rules of phasor theory now follow from Maxwell's discovery that ***E*** is in phase with ***H*** when energy is propagating towards an absorbing boundary. Consider first, what happens when the impedance is a pure resistance. In that case, **V** is in phase with **I**, and so ***E*** is in phase with ***H***. Hence ***E*** and ***H*** change sign together, and ***P*** always points towards the impedance. When the impedance is a pure reactance however, **V** and **I** are exactly 90° out of phase. This means that ***E*** and ***H*** have the same sign for exactly 50% of a generator cycle, and opposite signs for the other 50% of the cycle. Hence the direction of ***P*** reverses cyclically, and the average power flowing into the impedance is exactly the same as the average power flowing out. Intermediates between pure resistance and pure reactance correspond, of course, to the relative amounts of time that ***P*** spends pointing towards or away from the impedance.

On the issue of where the electrical energy is located relative to the connecting wires, note the earlier comment that electric field lines emerge almost perpendicular to a conducting surface. They only bend away from the perpendicular when there is a potential difference parallel to the surface. With the magnetic field lines encircling the wire, this means that the Poynting vector lies parallel to the wire provided that there is no potential difference on moving along the wire. In fact, a detailed analysis shows that the Poynting vector, considered as the average of a large number of microscopic energy transfer processes, is tilted very slightly towards a conductor on the outside (due to the small voltage drop caused by the internal resistance), and points almost directly inwards once under the surface. This means that the electrical energy is located almost completely outside the wire, and the small amount which flows inwards only does so to be absorbed.

So that is how electricity is visualised using fields; but it has to be said that there are subtleties. One peculiar consequence; on which we depend when solving routine electromagnetic problems, but which also relates to the fundamental rules of physics; is known as the ***principle of continuity of energy*** (not to be confused with the principle of *conservation of energy*). Maxwell's equations imply that energy does not simply disappear from one part of the Universe and reappear in another; it flows through space in a definable way and maintains its integrity (i.e., it is conserved in transit[3]).

3 **Is the Universe leaking energy?** Tamara M Davis, Scientific American, July 2010 p20-27.

This was deduced independently and almost simultaneously in 1884 by John Henry Poynting (after whom the Poynting vector is named) and Oliver Heaviside[4]. The underlying reason for continuity was, of course, not understood at the time; but nowadays we associate electromagnetic energy with photons, i.e., de-localised particles which can be emitted or absorbed, but not modified.

The principle of continuity dictates that electric and magnetic fields from different sources do not combine to produce electromagnetic radiation. If they did, the Universe would turn in to a fireball; but then again, perhaps it has already done that, and has subsequently inflated to attain a density at which such interference no longer occurs. The practical significance of energy continuity however, is that it allows us to consider different energy transport processes separately insofar as a system is linear (i.e., does not convert energy at one frequency into energy at other frequencies). Hence, when we model the fields around an electrical system, we know that there are other fields present (such as the Earth's magnetic field), but they can usually be ignored. Thus the fields imagined or depicted relate only to the process under consideration; and this allows us to tackle complicated problems by breaking them down into separate parts. Without continuity of energy, we would not be able to understand the Universe (quite apart from the inconvenience of being blasted apart while trying to do so).

4. Magnetic dipoles

It was allowed to pass without comment in the previous section, but it should have been obvious, that the spatial characteristics of electric and magnetic fields are very different. An electric field line may start from a point (a positive charge say) and terminate on another point (a negative charge say). It transpires however that magnetic field lines cannot do that, which means that they can only be envisaged as as continuous loops. This peculiar principle is embodied in Maxwell's equations as the statement that 'the divergence of the magnetic field is zero', i.e., it doesn't spread out from or converge onto points. The reason is that there are no magnetic charges, i.e., there are no tiny North poles and South poles swarming around on their own. The last statement warrants some qualification however, there being considerable lore attached to 'magnetic monopoles'.

The set of relations we nowadays call "Maxwell's equations" is actually due to Oliver Heaviside [see Nahin, cited above]. Heaviside reduced an original twenty equations to the four which relate the electric and magnetic fields, but he disliked the asymmetry which allowed electric fields to diverge while magnetic fields could not. Therefore he added a term to allow for the possibility of magnetic charges (which he called 'magnetons'), even though it could always be set to zero in the problems he studied. Later on, this idea was followed-up by Paul Dirac, who went on to formulate a relativistic version of quantum mechanics (and thereby predicted the existence of anti-matter) and showed that the existence of magnetic monopoles could explain the quantisation of electric charge (i.e., that charge is not infinitely divisible, but comes in discrete amounts). Monopoles figure in some of the various attempts to formulate a grand unification theory of physics, i.e., they might have existed shortly after the Big Bang; but they do not necessarily exist now, and have so-far never been isolated experimentally (monopole-like behaviour is seen in some exotic magnetic materials[5], but but such particles are composite rather than fundamental).

In the absence of charges on which to land, magnetic field lines can only form loops. Thus, when we imagine the field lines coming out of the North pole of a magnet and looping around to the South pole, we *don't* imagine that there are monopoles crowded at the poles like the charges on a capacitor plate. The act of cutting a bar-magnet in two produces two new magnets, rather than separate poles, and so we must conclude that the lines entering at one pole travel through the body

4 **Oliver Heaviside**, Paul J Nahin. 2nd edition. John Hopkins University Press 2002. ISBN 0-8018-6909-9. Maxwell's Equations: Ch 6, p85-88 and Note 24. Ch 7, p128.

5 **Monopole Position**, John Matson. Scientific American, Nov. 2009. p16. A sighting, of sorts, of separate North and South magnetic poles in spin ices.

of the magnet and emerge at the other.

Which brings us to the subject of magnetic materials and the origins of their properties. Magnetic materials can be classified as either diamagnetic ($\mu \leq \mu_0$) or paramagnetic ($\mu > \mu_0$). The 'ferromagnetism' exhibited by some substances corresponds to the ability to store very large amounts of magnetic energy, and is a giant form of paramagnetism. The ferromagnetic materials can also be further divided into 'hard' and 'soft', the former becoming permanently polarised during exposure to a strong magnetic field, an the latter losing its magnetism upon removal. Hard magnetic materials are, of course, used to make permanent magnets.

Recall that dielectric materials, which lack free charges, can be polarised by an electric field. Polarisation depends on the existence or creation of electric dipoles, which correspond to molecules or crystalline unit cells which have asymmetric internal charge distribution. Similarly, there are polarisable magnetic dipoles within materials; the difference being that there is no metaphorical knife sharp enough to allow us to cut the smallest magnetic dipole in half. The reason is that magnetic dipoles are not associated with pairs of monopoles, but with the spins and orbital behaviour of the charged sub-atomic particles themselves.

A magnetic field of material origin is always associated with an electric current, and the spin of a charged particle can be taken to constitute a current. Normally we say that protons and electrons "have spin", rather than simply saying that they spin; because a fundamental particle is a small package of electromagnetic energy, and it is unrealistic to envisage it spinning in the same way as a macroscopic object. A charged particle is essentially a quantity of energy ($E=mc^2$), trapped in some manner involving a non-divergent energy flow, and having a patent monopolar electric field, and a patent dipolar magnetic field.

Spinning charges can be imagined as tiny bar magnets; or better still, since they are able to change their orientations, as tiny compass needles. The difference is that, while a compass needle can point in any direction to align itself with an externally applied field, the orientations of microscopic dipoles are quantised, i.e., limited in terms of possible energy of orientation relative to an external field. The ordinary charges (protons and electrons) can have spin quantum numbers of $\pm\frac{1}{2}$, which means that they can absorb or emit energy to align their dipoles with an external field, or against it, but there are no intermediate states.

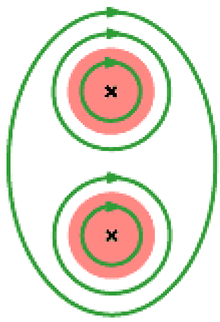
Diamagnetism is associated with paired spins, and paramagnetism is associated with un-paired spins. Spin pairing can be understood by considering what happens when two bar-magnets are placed side by side, with North poles adjacent to South poles. The fields from the two magnets form closed loops, and the combination ceases to interact with external fields and becomes diamagnetic (i.e., a collection of such paired magnets is, on average, slightly less permeable than free space). Hence, among the vast numbers of spinning sub-atomic particles in materials, most are paired, and paramagnetism arises from the relatively small number of un-paired particles which can exist in some types of material.

The materials of primary interest for the construction of inductive devices are, of course, ferromagnetic. Ferromagnetism occurs when the spins of un-paired electrons become aligned over relatively long distances, like strings of bar magnets attached North to South. These regions of constructive alignment are called 'magnetic domains', and are capable of internal re-arrangement to change the orientation of the overall magnetic dipole. In a soft ferromagnetic material, in the absence of an externally applied field, the domains are randomly orientated and there is no overall magnetism. When a field is applied however, the domains are progressively forced into alignment and energy is stored, to be returned when the external field decreases in strength and the material relaxes. Hence, the presence of ferromagnetic material vastly increases the magnetic energy which can be stored in a given volume of space, and the macroscopic average relative permeability of the volume is consequently much greater than 1.

In a hard ferromagnetic material, there exists an energy barrier with respect to domain realignment. Hence, if a field strong enough to overcome the barrier is applied, energy is absorbed, but not all of it is returned when the field is reduced. The domains become permanently aligned to some extent over very long range, and so the material obtains a permanent magnetic field. Hard materials are, of course, not suitable for ordinary inductor cores, because they behave in a highly non-linear fashion. It is important to be aware however, that no ferromagnetic material is completely soft; there will always be some residual magnetism after removal of an external field, and so coils with magnetic cores are not perfect linear devices.

5. Flux linkage

When the term 'linkage' is used in the context of magnetism, the metaphor is meant to be interpreted in the strong sense; i.e., a link is a loop interlocked with another loop, as in a chain. Particularly, the term 'flux linkage' refers to the fact that the magnetic flux lines which form around a current-carrying conductor can also loop around other conductors, this being part of the explanation for the phenomenon of electromagnetic induction. Hence, flux linkage is involved in the process whereby electrical energy supplied to one current loop can be dissipated in a resistance in series with a separate loop, the resulting structure, called a transformer-coupled network, being analogous to a chain. Before we consider inter-circuit linkages however, we need to consider the linkages which occur in a single circuit, particularly when that circuit is coiled into a set of overlapping loops all carrying the same conduction current.



A flux linkage can be considered to exist whenever a loop of magnetic flux completely surrounds a current. The diagram on the right represents the flux surrounding a pair of wires (seen in cross section) which are both carrying the same current (i.e., such as might occur in a two-turn coil). Notice first, that if there is any appreciable depth of current penetration into the body of the conductor (i.e., skin depth), then there can be lines of flux below the conductor surface. These lines, which are associated with internal inductance, do not encircle all of the current, and so contribute less than a whole flux linkage. On the outside, at or very close to the surface, there are lines which encircle the current only once and so correspond to single flux linkages. At greater distances however, the field is best considered as a superposition of the lines from both (or, in general, many) turns, and it takes the form of loops which enclose more than one turn. A flux loop which girdles the current N times is said to contribute N flux linkages.

The reason why we evoke the concept of flux linkage, is that it gives us a simple way of defining inductance. To understand it however, we must use Faraday's way of thinking about magnetic energy storage. Faraday imagined a conductor to be surrounded by a flux of lines (sometimes also called tubes) each having equal weight. In other words, every one of his lines contributed a unit of flux (later called a maxwell); so that regions where the field is strong have many lines, and regions where the field is weak have few lines. Also, lines are created by putting energy into the field, and disappear when energy is removed. Taken in conjunction with the observation that the inductance of a coil is proportional to N^2 (where N is the number of turns); it transpires that the increase in the amount of magnetic energy which can be stored, for a given current, when a conducting loop is coiled, can be determined by counting the flux linkages. Thus, according to Faraday's interpretation, coiling increases inductance because it allows field lines to link the current more than once. Hence the definition:

Inductance = flux linkages per unit of current

Notice that we have not bothered to include a constant of proportionality. The units are to be defined so that such a constant will not be needed.

Nowadays, given wider familiarity with vectors; we are a lot happier to think of magnetic flux in an abstract way. It is sufficient to assign a magnitude and a direction to describe a point in a vector field; but the magnitude which describes the intensity of the magnetic flux can nevertheless be envisaged as the number of maxwells (i.e., lines) enclosed by an infinitesimal area perpendicular to the flux. Hence there is a perfect correspondence between Faraday's and the modern conceptions of flux density (i.e., flux per unit area), provided that we define the relationship correctly. Herein lies a difficulty however, which is that Faraday and Maxwell saw the determination of inductance as a counting problem, whereas later writers saw it as a field integration problem. There is no philosophical difference, but neglect of detail can result in a set of field parameter definitions which are only approximate. We should note that the phrase 'approximate definition' is an oxymoron.

A typical attempt at defining inductance is based on the idea that the flux links the current once in a single loop circuit, and that it links the current N times in an N-turn coil. As we have seen however, neither proposition is strictly true. The flux associated with internal inductance contributes less than 1 flux linkage per line; and of the flux associated with external inductance, not all of it will link with all of the turns. Thus, except in the special case of a single turn loop at very high frequency (when the skin depth has gone to zero), the assumption of maximum possible linkage will lead to error. One way to solve that problem is to define a linkage efficiency parameter, i.e., a correction factor (<1) by which the number of turns can be multiplied in order to ensure that the linkages have been counted correctly. We will start by calling the number of turns so corrected the *effective number of turns*, and give it the symbol Ñ. Thus:

$0 < \tilde{N} < N$

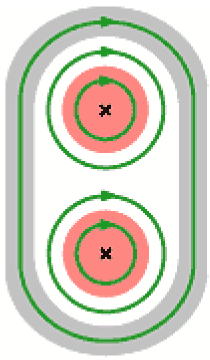
The correction parameter is then Ñ/N, which, for reasons which will shortly become clear, can also be written as $\sqrt{k_H}$, i.e.:

$\tilde{N} = N \sqrt{k_H}$

k_H is a field inhomogeneity factor which, in general, depends on the geometry of the system and the properties of the materials involved. For the hypothetical special case of a current-sheet solenoid (which has no internal inductance):

$k_H = k_L$

where k_L is known as 'Nagaoka's coefficient' and is readily calculated using procedures described elsewhere[6].



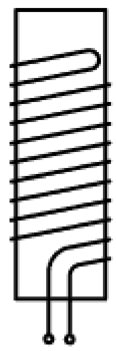
One way in which to maximise linkage efficiency is to provide a magnetic path of low reluctance; i.e., a path of high permeability, short length, and reasonably large cross section. This can be achieved by (say) winding wire onto a ferrite bead. There will still be internal inductance, and there will still be external flux which only links the current once; and in a coil of many turns, there will also be flux which links some but not all of the turns; but the amount of energy which can be stored in the magnetic material will be so great, in comparison to that which can be stored in the wire and the surrounding medium, that it will make other contributions almost negligible. That is why the inductance of a coil wound on a closed transformer core is almost entirely dependent on the number of turns and the core inductance factor (A_L). The 'reluctance' of a transformer core, incidentally, is defined as S=1/A_L. Reluctance in a magnetic circuit is analogous to resistance in an electrical circuit, and so A_L is analogous to conductance.

Notice that the possible range of Ñ (effective turns) was stated above to lie between 0 and N. We might expect a lower limit of 1 (or very slightly less, due to the partial linking associated with internal inductance), but the linking can be considerably less than 1 when the flux from one part of the loop cancels the flux in another part of the loop; i.e.; linking can be constructive or destructive. In a single turn loop, linkage is maximised by maximising the loop area; i.e., by getting a particular point in the conductor to be as far away as possible from any other point which is not carrying current in exactly the same spatial direction. Hence we need a loop shape criterion when defining linkage; and so, unless stated otherwise, it is assumed that the external flux of a single turn loop links the current once (i.e., maximally) when the area is at its maximum; i.e., when the loop is

6 See: Solenoids. D W Knight. Still in HTML form at time of writing.

circular.

In a multi-turn coil, the effective number of turns (and hence the number of flux linkages) can be reduced by changing the winding direction at some point during the making of the coil. This is often done for the purpose of minimising inductance; i.e., to cancel linkages so that Ñ is as close to zero as is possible. This is (for example) the reason for the hairpin bifilar pattern used in some types of wirewound resistor (see right). It follows that if we start counting upwards when winding a coil, and then change direction and start counting downwards; the effective number of turns will eventually become negative (at least, notionally). The definition of a particular winding direction as positive or negative is not completely arbitrary, because it affects the relative directions of the flux lines on the inside and outside of the coil. Hence the handedness of a helix (i.e., left-handed or right-handed) plays a part in determining the orientation of the overall magnetic field. The usual choice however, is to define the effective turns number as positive (i.e., to treat it as a magnitude) and use Maxwell's corkscrew rule to determine the field polarity. This has the effect of fixing the polarity of the overall flux solely according to the sign of the current; i.e.; since inductance is defined as flux linkage per unit current (and is by convention positive), then flux linkage is positive when current is positive.



6. Magnetic induction

The term 'electromotive force' or 'EMF' is sometimes used when referring to a voltage produced by the action of a generator; i.e., an EMF is a voltage associated with a source of energy. Similarly, 'magnetomotive' force or 'MMF', is proportional to the current supplied by a generator; but, due to the phenomenon of flux linkage, MMF is only equal to current when the circuit is composed of a single (i.e., non-overlapping) circular loop.

In electrical systems, power and energy, and also the illumination fields 'power per unit area' and 'energy per unit area' are always proportional to the square of an EMF, the square of an MMF, or to a product of EMF and MMF. This is evident from the elementary formulae: $P=V^2/R$, $P=I^2R$, $P=IV$, $E=(½)CV^2$, $E=(½)LI^2$, and so on. Thus, both EMF and MMF have units which are proportional to the square-root of energy. We can also observe that the energy stored in an inductor is proportional to the inductance, and inductance is proportional to N^2; from which it follows that, if MMF is a function of N, then it must be directly proportional to N.

The act of winding a current loop into a coil must be deemed to increase the MMF, because it increases the magnetic contribution to stored energy without the need for an increase in current. Thus, given that we have deduced the proportionality from dimensional requirements, MMF has units of *Ampere turns*. (and is often referred to as 'Ampere turns'). It is also typically defined as the current multiplied by the number of turns; but that last step is questionable. Flux linkage is a logically-consistent explanation for the increase in MMF which results from coiling, and we know that there will always be flux which cannot link with every turn. Hence we will define MMF as:

$F = \tilde{N} I$

where Ñ is the effective number of turns, as discussed earlier. Now notice that the symbol for current (I) has been written un-bold, whereas in AC circuit analysis we would normally write it in bold script, to indicate that it is a phasor. The reason is that we are presently engaged in the business of relating circuit parameters to the actual magnetic field; and the field strength is strictly proportional to the instantaneous current. It is also the case that, for the purpose of obtaining a complete description of the field, the current has only magnitude and one of two possible directions; i.e., by reversing it, we reverse the direction of all of the flux loops (and thereby reverse the poles in the case of an electromagnet), but we cannot otherwise alter the relative distribution of the field. Hence the current, as defined for the purpose of mapping from electromagnetics to circuit theory, is pseudoscalar, just as it is in DC circuit analysis. We can convert it back into a phasor later, because by then we will have defined all of our new magnetic parameters and the generalisation will not affect them; but for now, we must consider only instantaneous current (or DC).

The units of electric field strength are [Volts / metre]. Similarly, the units of magnetic field strength are [Ampere turns / metre]. Hence the intensity at some point in the magnetic field is given by a vector (***H***) which has a magnitude proportional to the MMF, divided by some distance over which the MMF is considered to act. That distance is known as the *path length*, and is given the symbol ℓ (sometimes with subscripts, depending on the context). The path length in a single-loop circuit is the average length of a flux loop, which is difficult to calculate; but for solenoids and closed transformer cores, the situation is more straightforward. In the case of a solenoid, it is the length (or height) of the cylinder taken from centre-to centre of the connecting wires. In the case of a transformer core, assuming that the field outside the core is negligible, it is the average length of a flux loop in the core body, and is given as the ℓ_e value (effective path length) in the manufacturer's datasheet.

The amount of energy which can be stored in the magnetic field around a current loop is proportional to the average permeability of the magnetic path. Hence, according to Faraday's way of thinking; for a given current, the number of field lines is multiplied by the permeability. Hence we can define a new field which is more fundamental and informative than ***H***, i.e.;

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

where the vector ***B*** is called the ***magnetic induction*** or ***flux density***, i.e., it is the number of Faraday lines per unit area at a given point in the field. It should be obvious, that the inductance of a loop is dependent on ***B***, rather than on ***H*** alone.

To define the complete spatial distribution of the magnetic field, we must solve Maxwell's equations for the system. For the derivation of circuit analysis parameters however, it is not necessary to do that. Instead, we can define an average flux density as the total flux divided by the total path area. Notice that this new quantity is still a field vector. By defining it over the total area however, we imply that its direction is the average direction of all of the field lines in the bundle enclosed by the area. This average direction is always perpendicular to the plane of the loop. Therefore the total flux has only two possible directions (depending on the direction of the current), i.e., it is pseudoscalar, and so we give average flux density the symbol *B* (un-bold, in italic script, not to be confused with susceptance, B). Hence:

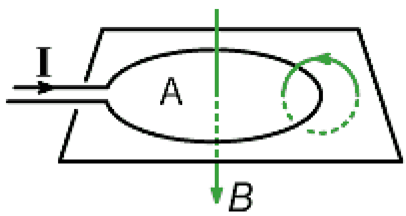
$B = \mu H = \Phi / \mathrm{A}$ [Tesla]

where

$H = \mathrm{F} / \ell$ [Amp. turns / metre]

is the average field strength on the path, and Φ is the total magnetic flux associated with a single loop. Notice that Φ is also pseudoscalar (as is *H*); i.e., it can be positive or negative, and so it is *definitely not* a magnitude. The SI unit of flux is the Weber [Wb], which is equivalent to 10^8 lines (or maxwells). The unit of flux density (magnetic induction) is the Tesla [T], or Wb/m², which is equivalent to 10^4 gauss. The gauss is still commonly used in scientific papers, transformer-core datasheets, etc., and its relationship to the Tesla should therefore be memorised.

When a coil is wound on a closed-circuit high-permeability transformer core, the energy stored in the core is so great in comparison to that stored elsewhere, that the path area is, to a good first-order approximation, the same as the average cross-sectional area of the magnetic channel (i.e., the $\mathrm{A_e}$ value given in the core's datasheet). For an un-cored loop or coil, the path area is perhaps less-immediately obvious, but the definition is not difficult to understand.



Consider a planar (i.e., flat) current loop. However we determine the total flux, the process is equivalent to counting the lines, and each line must be counted only once. Now, noting that each line encircles the current, the number can be determined by counting in the plane of the loop, either on the inside, or on the outside (but not both). Counting the flux on the outside is somewhat problematic, because the field extends (at least notionally) to infinity. Hence the integration is carried out over the area enclosed by the loop; and that area is consequently identified as the path area.

While the logic behind the definition is straightforward however, exact determination of the area is not so easy. The problem is that practical conductors have finite (and often substantial) thickness, which means that the boundary established by the conduction current is necessarily diffuse. The process of incorporating known physical behaviour into a model by defining what happens at various points in a co-ordinate system is known as "establishing the boundary conditions". In this case, we want to count all of the flux lines exactly once, and so the relevant boundary condition is established by defining a perimeter line such that the conduction current flowing outside this line is exactly equal to the current flowing on the inside. Widespread practice is then to assume that the boundary lies at the conductor centre-line; but this is not an accurate choice for several reasons: Firstly, the conduction path on the inside of a loop is shorter than on the outside (i.e., the resistance is less), and so the current will tend to concentrate on the inside. Secondly, there will be a redistribution of the current at high frequencies; especially in multi-turn coils, due to the proximity effect. Lastly, there will be displacement currents and interfacial effects (i.e., it is not possible to make an electrical connection without disturbing the magnetic field); and so, what looks initially like a simple problem turns out to be rather complicated.

Due to the difficulties in establishing the path area, magnetic problems are usually addressed by starting with simplified models. The simplification in question is that of assuming that conductors are filamentary, i.e., of negligible thickness in comparison to other dimensions. Note that a hypothetical filament has no internal inductance, because it is too small to have flux loops inside it. Recall also, that the conductor in a current-sheet solenoid has width, but no radial thickness, and so is filamentary when the solenoid is viewed end-on. The filamentary current model will provide a first (and usually fairly accurate) approximation for the behaviour of a practical magnetic device, and can be further refined by the inclusion of internal inductance and various other correction terms.

Although the process of determining the path area may involve approximation in practice, it is nevertheless rigorously defined as the area bounded by the median line of the conduction current density; i.e., a line chosen so that the integral of the current density on the inside is equal to the integral of the current density on the outside[7]. Hence we may take the final step in defining the inductance of a multi-turn coil, by multiplying the total flux through a single turn by the effective number of turns, to obtain a parameter called the 'number of flux linkages' (or 'flux linkage' for short). To that we assign the symbol Λ (Capital "Lambda"), i.e.:

$\Lambda = \tilde{N}\,\Phi$ [Weber turns]

Notice that Λ is pseudoscalar, because Φ is pseudoscalar; i,e., all quantities derived from the field are bi-directional vectors.

Inductance, the absolute measure of the ability of an electrical device to store energy in the magnetic field, was identified earlier as the number of flux linkages per unit of current. Hence:

$L = \Lambda / I$ [Weber turns / Amp.]

The composite unit is known as the Henry, in honour of Joseph Henry (1797 - 1878) who studied magnetic induction at the same time as Faraday, and made comparable scientific contributions.

The inductance of a coil or current loop is, by convention, positive. Negative inductance can occur in circuit analysis, but only in situations involving inductance subtraction; which means that negative inductance can be synthesised, but does not physically exist. Thus, in the relationship given above, it can be seen that flux linkage must be allowed to be either positive or negative, because instantaneous current can be either positive or negative. This freedom is required for reasons of mathematical consistency; i.e., failure to allow it constitutes the ludicrous proposition that inductance changes sign when the current is reversed. It also allows us to determine the overall direction of the dipolar field around a coil by using Maxwell's corkscrew rule.

---

7 LF effective radius of a single-layer solenoid. D W Knight [available from g3ynh.info]

The general formula for the inductance of a coil can now be obtained by working backwards through the preceding discussion. Thus:

$L = \Lambda / I$	Flux linkage, $\Lambda = \tilde{N}\,\Phi$
$= \tilde{N}\,\Phi / I$	Flux through the path, $\Phi = A\,B$
$= \tilde{N}\,A\,B / I$	Flux density (magnetic induction), $B = \mu\,H$
$= \tilde{N}\,A\,\mu\,H / I$	Field strength, $H = F / \ell$
$= \tilde{N}\,A\,\mu\,F / (\ell\,I)$	Magnetomotive force (MMF), $F = \tilde{N}\,I$
$= \tilde{N}\,A\,\mu\,\tilde{N}\,I / (\ell\,I)$	
$= \mu\,\tilde{N}^2\,A / \ell$	Effective turns, $\tilde{N} = N\sqrt{k_H}$

Hence:

$$L = \mu\,N^2\,k_H\,A / \ell \qquad \text{[Henrys]}$$

For a coil wound on a closed transformer core, where (to a reasonably good approximation) $k_H \rightarrow 1$, this becomes:

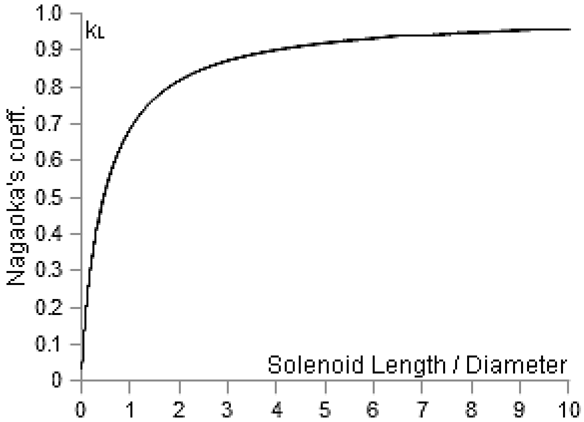
$L = \mu\,N^2\,A_e / \ell_e$

(Where A_e and ℓ_e are the effective path area and effective path length, as given in the manufacturer's datasheet). The quantity:

$A_L = \mu\,A_e / \ell_e$ [Henrys / turn²]

is, of course, the core inductance factor or magnetic conductance (and its reciprocal is the core reluctance).

For a current-sheet solenoid, $k_H = k_L$, where k_L is the fringing-field or field non-uniformity correction factor known as Nagaoka's coefficient. k_L is always less than 1, because some of the field lines come out through the sides of a solenoid, and so fail to link with all of the turns. $k_L \rightarrow 1$ when a solenoid is very long and thin because the flux lines only spread out towards the ends of a helical coil (to loop around on the outside), and if the ends are far apart, the spreading (non-uniform) region is small in comparison to the major part of the enclosed field.



Before we move on; it is perhaps worth drawing attention to the fact that 'turns' have been included in the system of magnetic units used here. Turns is, of course, just a number, and so has no unit of measurement. Hence some authors do not include turns when stating units, and may even go so far as to claim that the deletion is in aid of mathematical rigour. There is a slight downside; in that the distinctions between current and MMF, and flux and flux linkage, are then lost; not helped by the once widespread practice of using the symbol Φ for both flux and flux linkage; but these privations no doubt help to provide the student with a challenging learning environment.

On a more serious note; those familiar with dimensional analysis will know of it as a technique which allows relationships to be determined by comparing the measurement units of the variables associated with a physical system. The limitation of dimensional analysis is that it cannot determine constants of proportionality, because scale factors (i.e., pure numbers) have no units. The corollary however, is that pseudo-dimensions, such as turns, can have no effect on the validity of dimensional analysis (and may provide valuable additional information). Hence the argument, that the inclusion of numerical parameters is unrigorous, falls down; and the deletion of turns from magnetic units serves no purpose except to spread confusion.

7. Electromagnetic induction

It is important to distinguish between the terms *magnetic induction* and *electromagnetic induction*. Magnetic induction (flux density) is, literally, the amount of magnetism induced (i.e, forced) into a volume of space. Electromagnetic induction, on the other hand, is the propensity for a time-varying magnetic field to create an electric field (and vice versa). The latter phenomenon was discovered by Faraday, in 1831, after he had spent some six years pondering the riddle: 'why is it that a current produces a magnetic field, but a magnetic field does not produce a current'. With hindsight, we can easily identify the barrier to understanding by inserting the adjectives 'direct' in front of 'current, and 'static' in front of 'magnetic field'. Faraday's final experimental apparatus for this investigation was effectively a transformer (although the term was not coined until some years later); i.e., he used a separate coil and a battery to provide the field. The riddle was solved when he noticed that the galvanometer gave tiny kicks in opposite directions as he switched the battery on and off.

Faraday's subsequent interpretation of the induction coil experiment (in terms of flux linkages, of course) became one of the cornerstones of Maxwell's great synthesis of electricity and optics. It was Einstein who gave the full explanation however, by deducing that a magnetic field acquires the character of an electric field when viewed from a moving frame of reference (and vice versa); the degree of mapping from one to the other being proportional to the relative velocity, and becoming absolute at velocity c. This means that there can be no greater relative velocity than c in the world of ordinary matter (but not, as is often assumed, that nothing can travel faster than light[8]), and so the geometry of space-time had to be redefined as a consequence of Maxwell's equations. Einstein's Special Relativity also tells us that there is no distinction between a changing field and a moving field; and so an electric field can be created from a magnetic field in either way. Nowadays, an electromagnetic induction engine of the changing field variety is called a transformer, and a rotating engine of the moving field variety is called an alternator or generator.

Recall that, using Faraday's mental picture of the magnetic field, flux lines (and hence linkages) are created when energy is put into the field, and destroyed when energy is taken out. Hence Faraday's astonishingly simple explanation for what turned out to be the seed for the whole of modern physics (here paraphrased):

'The voltage induced across the ends of a coil is proportional to the rate of change of flux linkages per unit of time' (Faraday's Law of induction).

Note that the word 'proportional' can be replaced by 'equal' if the system of units is chosen correctly; and that was one of the effects of Maxwell's later unification of the ***E*** and ***H*** fields. Hence:

$V = d\Lambda/dt$ [Volts]

This equation has enormous implications; not least, for our purposes, because it lies behind the theory of AC electricity in general, and inductive devices in particular.

8. Self-induction

The definition of inductance was given earlier as:

$L = \Lambda / I$ [Henrys]

This can be re-written:

$\Lambda = L\,I$ [Weber turns]

Now recall that inductance, here seen as the factor of proportionality in the relationship between flux linkages and current, depends only on the physical geometry of the system and the permeability of the magnetic path. Both permeability and geometry can change with field strength, in the latter case (for example) because the force between parallel current-carrying conductors causes coils to contract in length as the current is increased; but these are small effects, and it is not unrigorous to

8 **Quantum Non-Locality and Relativity**, Tim Maudlin, 2nd ed. 2002. ISBN 978-0-6312-3221-6

conceive of an idealised coil, of constant inductance, and then apply corrections when dealing with practical devices. Hence we can differentiate the expression above on the assumption of constant inductance; i.e.:

$\partial\Lambda/\partial t = L\ \partial I/\partial t$

(here we use the partial differential symbol, ∂, to indicate that other dependent variables are being held constant). Now, according to Faraday's law, $\partial\Lambda/\partial t$ is a voltage, and so, assigning a symbol to it, we can write:

$V_b = L\ \partial I/\partial t$

Some readers may be familiar with a different version of this equation, in which the voltage is given as negative. In fact, there is a reason for choosing a sign at this point, but the argument is a subtle one and we will look at it in detail later (it will turn out that the choice made here is the correct one for purposes of circuit theory). For the time being therefore, simply note that both L and t are defined as positive, and so the sign of the voltage is dictated by the derivative (i.e., by the sign of the gradient of the graph of I vs. t). Hence, an increasing current gives a positive voltage, and vice versa.

The expression above tells us that when the current passing through a coil changes, a voltage is induced in the coil itself. This phenomenon is called 'self induction', and was discovered independently by Joseph Henry in 1832 (although it is, of course, a corollary of Faraday's law of 1831). The voltage produced is habitually referred-to as the "back EMF", which is unfortunate because it is not an EMF in the strictest sense. An inductor (not affected by any magnetic field but its own) is a passive electrical device, and no amount of work with a hacksaw will reveal a generator inside it. Hence the self-induced voltage is a reaction force, or potential-difference; and we will refer to it here as the 'back voltage' or 'reaction voltage'. The effect of the reaction voltage is to oppose the change giving rise to it (Lenz's law); which means that when an inductive circuit is connected to a generator of time-varying voltage, there will be a limitation on the current which can flow (this being in addition to the limitation provided by the resistance). By noting that the reaction is in opposition to the driving force, we have cryptically chosen a sign for $L\partial I/\partial t$ (and, as we shall see, it is positive according to the conventions of circuit analysis).

Now suppose that we have an idealised coil of negligible resistance. If we connect it to a battery, the back voltage will prevent the current from rising instantaneously, but the system will eventually reach equilibrium and an enormous constant current will flow. If we connect it to an alternator however, the current will never become constant, and so the back voltage will prevent the current from rising to its DC limit and will remain equal to the instantaneous applied voltage. If the generator produces a sinusoidal output, and the time reference is arbitrary (so that we can assume that V=0 when t=0), the instantaneous driving voltage at time t is given by the expression:

$V(t) = V_p\ \mathrm{Sin}(2\pi f\, t)$

where V(t) should be read as " V, a function of t " (i.e., all parameters are constant, except for t), and V_p is the peak voltage. Hence, using Faraday's law:

$V_b = L\ \partial I/\partial t = V_p\ \mathrm{Sin}(2\pi f\, t)$

i.e.:

$\partial I/\partial t = (V_p / L)\ \mathrm{Sin}(2\pi f\, t)$

The instantaneous current can now be found by integrating both sides of this expression with respect to time, i.e.:

$$\int \frac{\partial I}{\partial t}\, dt = \frac{V_p}{L} \int \mathrm{Sin}(2\pi f\, t)\, dt$$

standard solution $\int \mathrm{Sin}(ax)\, dx = \frac{-1}{a}\ \mathrm{Cos}(ax) + c$

Hence:

$$I = \frac{-V_p}{2\pi f L} \text{Cos}(2\pi f t) + c$$

Note that the constant of integration (c) in the expression above must be zero, because it is obvious that when the peak generator voltage is set to zero, the current will always be zero. Thus Faraday's law has given us the relationship between voltage and current for a pure inductance connected to a sine-wave generator. Notice also the appearance of the quantity 2πfL, which, since the cosine function is dimensionless, has units of [Volts / Amp.] or [Ohms]. Thus we attribute the current limitation to a type of resistance: 'reaction resistance' or 'reactance'; and this quantity is, of course, important enough to warrant its own symbol, X_L.

What we can do now with the expressions for instantaneous voltage and current is to multiply them together to determine the instantaneous power flowing between the generator and the coil. Thus:

$P(t) = -(V_p^2 / X_L) \text{Sin}(2\pi f t) \text{Cos}(2\pi f t)$

This can be simplified using the standard trigonometric identity:

$\text{Sin}(a) \text{Cos}(b) = (½)[\text{Sin}(a+b) + \text{Sin}(a-b)]$

i.e.:

$\text{Sin}(x) \text{Cos}(x) = (½)\text{Sin}(2x)$

Hence:

$P(t) = -(½) (V_p^2 / X_L) \text{Sin}(4\pi f t)$

The quantity V_p^2 is, incidentally, equal to $2V_{rms}^2$ (i.e. V_p is the RMS voltage multiplied by √2), and so:

$P(t) = -(V_{rms}^2 / X_L) \text{Sin}(4\pi f t)$

The instantaneous power is a pure sine wave at twice the excitation frequency; and therefore, as advertised earlier, alternates between positive and negative. The average of a sine wave is, of course, zero; and so the overall work carried out by a generator on an idealised inductance is nil.

Now consider the phase relationship which exists between the instantaneous values of the voltage and current. We have:

$V(t) = V_p \text{Sin}(2\pi f t)$

and

$I(t) = -(V_p /X_L) \text{Cos}(2\pi f t)$

If the time co-ordinate is chosen so that the voltage is a sine wave in the mathematical sense (i.e., V=0 and changing in the positive direction when t=0), then the current is the negative of a cosine wave. There is a fixed 90° phase difference between current and voltage. Notice also that when t=0, -Cos(2πf t)=-1. When the voltage is at zero, but changing in the positive direction; the current is at its most negative, and will take a further ¼ of a cycle before it reaches zero. Thus the current in an inductance lags the applied voltage; a fact learned in kindergarten of course, but here we see how it comes from Faraday's law.

If there is a fixed phase difference between the voltage and current associated with a purely reactive circuit element, then the mathematical relationship can be established without stating the time variations of V and I explicitly. Hence we can dispense with sine and cosine functions, and devise a formal mapping into an algebra which represents the phase difference by means of a unit vector in the +90° direction of a planar co-ordinate system. The choice which puts V 90° ahead of I, assuming that the phase co-ordinate increases in the anticlockwise direction (i.e., adopting the standard trigonometrical convention) is:

$\mathbf{V} / \mathbf{I} = \mathbf{j} X_L$

The unit vector **j** converts the relationship into a two-dimensional vector operation, and so **V** and **I** are now written in bold typeface. **j**, of course, transforms algebraically as √-1; and so we pass from Faraday's law to phasor theory.

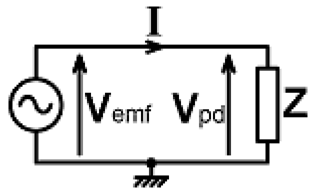
9. Polarity of the induced voltage

In the preceding section, we obtained an expression for the reaction voltage of a passive inductor:
$V_b = L\, \partial I/\partial t$
It was then commented that this expression is frequently written with a minus sign. Such amendment was avoided however, and the AC behaviour of inductors fell out of the algebra without mishap. Now we find ourselves in the strange position of having to justify our failure to make a widely repeated mistake. The confusion results from the misapplication of Lenz's law.

Faraday's law states that a changing magnetic field will induce a voltage across the ends of a conductor suitably disposed within the field. Maxwell's equations have it that a changing magnetic field produces an electric field, but that is the same thing. If the conductor forms part of a closed circuit; then a current will flow, and may be registered by a fast-responding galvanometer. Lenz's law of 1835 tells us that the current is such that the magnetic field *it* produces opposes the change giving rise to it. Lenz's law, as most authors are wont to point out, is actually a manifestation of the principle of conservation of energy; the corollary being that work must be done in order to make the current flow.

The implication of Lenz's law in relation to inductors is that; if a coil or loop of wire is connected to an alternator, the current which flows will induce a voltage which opposes the applied voltage. Now, since the instantaneous value of this reaction force is always in opposition to the applied force; then, according to many textbooks, the expression for it must be negative. Well, perhaps; but here we need to be extremely careful because, if we make a logical error at this point, our attempts to derive the important results of transformer theory will run into trouble. The problem is that a minus sign seems to imply that the reaction voltage is 180° out-of-phase with the driving voltage; but such a view is contrary to the conventions of circuit theory.

The issue can be resolved by considering the distinction between an EMF and a potential difference (PD). Except for a special case known as a 'short circuit', *every* impedance produces a reaction force. We can conceptualise the mechanism by saying that the generator produces a driving or 'electromotive' force, and the load pushes-back by developing a potential difference. Since every action has an equal and opposite reaction (cf. Newton's 3rd law), we may reasonably deduce that:
$\mathbf{V}_{emf} - \mathbf{V}_{pd} = 0$
Thus $\mathbf{V}_{pd}$ gets its minus sign, but that does not not imply that it is out-of-phase with $\mathbf{V}_{emf}$. On the contrary, as should be obvious from the diagram on the right, what we are really saying is that:
$\mathbf{V}_{emf} = \mathbf{V}_{pd}$

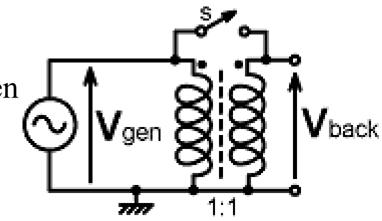


The ambiguity arises because Kirchhoff's second law can be stated in two ways: either, "the algebraic sum of the voltages around a circuit is zero"; or, "the sum of the PDs is equal to the sum of the EMFs". In the first case, PDs are negative with respect to EMFs. When analysing circuits however, we draw arrows, which tell us how to combine the terms in the expressions which describe the circuit behaviour; in which case the minus sign belongs to the analysis, rather than to the term itself. It should all come out the same in the end of course; but it is not unusual to find confusion carried forward into the mathematical working.

The difference between EMF and PD was once emphasised in the teaching of electricity. The matter lost its weight in the latter third of the 20th Century however, there being some doubt about the importance of distinguishing between two quantities which are always equal. It becomes important in dealing with inductance however, as an antidote to the assumption that the reaction force needs to be treated in a special way. There is nothing wrong with Lenz's law. In 1835 it was a brilliant observation; but the principle of conservation of energy is now axiomatic upon all physical investigation, and while Lenz's law is important, it does not need to be specifically applied in this context. What is meant by that is that, if we measure voltages relative to a common reference, the "back EMF" is actually a PD, and is therefore equal in magnitude and phase to the force giving rise to it.

Hence, should we be able to find a way of measuring the back voltage independently, we will find that it is very definitely in phase with the generator voltage. In fact, there is a way of doing that, using an exotic device to which we might apply the snappy title: 'inductor with integral field-probe winding' (although some might be inclined to call it a 'transformer').

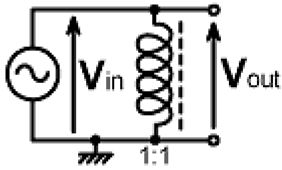
The reason why the reaction force produced by an inductor *is* special, is that the agency giving rise to the force can be intercepted, i.e., back voltage can be observed in a manner which separates it from the driving voltage. If the intention is to measure it accurately, then this can be done by arranging another coil to be affected by the magnetic field in almost exactly the same way as the driven coil. A simple solution is to take two strands of insulated wire and twist them together, then wind several turns of this bifilar pair onto a high-permeability closed-circuit magnetic core, such as a ferrite toroid. One of the windings can then be connected to a signal generator, and the magnitudes and phases of the voltages across the driven and the sensor windings can be compared using a dual-channel oscilloscope. In doing that however, we must keep track of the phasing of the windings, and so we arbitrarily define a lead wire going into one side of the core as the 'start', and the wire coming out of the other side as the 'finish' of the associated winding. We then label the 'starts' with dots on the circuit diagram.



Now, provided that the magnetic coupling between the two coils is very tight, it will be found that closing the switch 's' in the circuit on the right will make practically no difference to the relationship between $\mathbf{V}_{\mathbf{gen}}$ and $\mathbf{V}_{\mathbf{back}}$. In fact, there will be a small difference due to losses and incomplete flux linkage, but not sufficient to be noticeable within the precision achievable from oscilloscope measurements.

So now to the question of why the voltage (strictly, the open-circuit voltage) across the additional winding is equal to the back voltage produced by the inductor. The reason is that the back voltage is due to a changing magnetic field, and the experiment has been constructed in such a way that the fields affecting the two windings are practically the same. Hence, since the two windings in this case are also identical (they have the same number of turns and occupy almost the same volume of space), the voltage across the sensor coil is (practically) equal to the reaction voltage which controls the current flowing from the generator.

We have not learned anything new about magnetism from that; but what we *have* deduced is fundamental. A transformer is a device which allows the reaction voltage produced by one inductor to be sampled by means of another inductor (without necessarily making any direct electrical connection). To that we can add, that the sampling can be in whole or in part, depending on the degree of magnetic coupling; a matter which we will quantify shortly by introducing the concept of mutual inductance. At the risk of getting a little ahead of ourselves, we can also note that all windings can can carry a current. Hence we can use an auxiliary coil to create a time-varying magnetic field which will *modify* the reaction voltage produced by an inductor. The simplest way in which to do that is to connect an impedance to the second winding; in which case, if the load contains a resistive element, then the reaction will be changed in such a way as to allow a net energy flow from the generator to the resistance. Another possibility however, so deleteriously neglected in many textbooks, is to connect an *active* network to the second winding and thereby use a transformer (say) to cancel an unwanted component of a signal, or use a 'neutralisation' winding to cancel the 'self-capacitance' of an inductor.



The more exotic uses of transformers require a clear understanding of the phase relationships between all of the currents and voltages involved. This, as mentioned above, is a general source of confusion; but to begin, we can note that closing the switch in the circuit above converts our bifilar transformer into an inductor wound with two-strand bunch-wire, and we end up with the circuit on the right. What we have now is a 1:1 autotransformer;

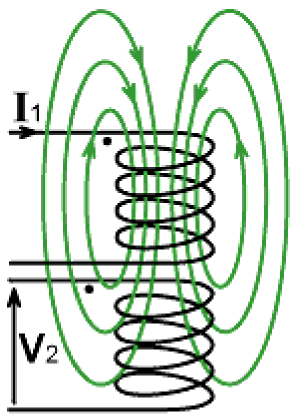
a network of dubious practical merit, but leaving little doubt regarding the relative phases of $\mathbf{V_{in}}$ and $\mathbf{V_{out}}$.

10. Mutual induction

The extraordinary property of Faraday's way of thinking about electromagnetic induction is that it is relativistic. Increasing the current increases the number of flux lines, and reducing it vice versa. Moving a loop through a field of external origin causes flux lines to switch from being on the outside of the loop to being on the inside, and vice versa. Either way, the number of flux linkages changes and a voltage is induced; and there is no distinction between a moving field and a changing field. This strange correspondence fascinated both Maxwell and Einstein, and led to a mathematical formulation which Faraday, by his own confession, did not fully understand. Still Faraday could see what was going on in his mind's eye, and we reject his wonderful metaphor in favour of a more abstract view at our peril. We need a formal and rigorous general approach of course; but there is no substitute for having a good idea of what the answer will be before tackling some complicated mathematical problem.

Here, of course, we focus on the effect of a changing field (transformers), rather than the effect of a moving field (generators); but regardless of how we fix our attention (and given the complicated nature of the three-dimensional problem) it must seem, at first, that we are in theoretical deep water. It transpires however that, by using the concept of flux linkage, the matter of incorporating transformers into circuit theory becomes a straightforward topological problem; i.e., a matter of interconnectedness, like two-terminal network theory itself. By using a topological argument, we can capture the magnetic interaction between a pair of inductors in a single parameter; and although we will still need to use arguments involving fields in order to calculate that parameter from first principles, there will be many situations in which we will be content just to measure it.

The general mechanism whereby a time varying current in one coil induces a voltage in another is illustrated on the right. Here we can see that some, but not all, of the flux links with the pick-up coil; and it should also be obvious that the induced voltage will depend on the proximity of the coils. The law which governs the relationship between inter-coil distance and output voltage is dependent on the geometry of the field, and so will be complicated; but if the system is linear (which it will be to a good approximation in the absence of hard magnetic materials), then the complexity of the field problem cannot add complexity to the circuit-analysis problem. In other words; for any given frequency component in the input current, there will only be one frequency component in the output voltage. This means that, for a given geometry, the relationship between $\mathbf{I}_1$ and $\mathbf{V}_2$ at a particular frequency is governed by a mathematical constant.



Recall that the self-inductance of a coil is defined as the number of flux linkages per unit current. It follows that, in the case where current in one coil produces flux linkages in another; there is a parameter, having units of inductance, which will allow us to calculate the induced voltage. That parameter, not surprisingly, is called the ***mutual inductance***, and is defined as: 'the number of flux linkages in coil 2 per unit of current in coil 1' ; i.e.:

$M = \Lambda_{21} / I_1$ [Henrys]

Notice that the current is written as a pseudoscalar. As before, it is necessary to use instantaneous voltages and currents when relating circuit parameters to field parameters; and the actual phase relationships can be incorporated later.

Now, referring again to the diagram above, consider what will happen if the pick-up coil is rotated through 90° about its centre point in the plane of the diagram. The flux which links with one half of the coil will then be in the opposite sense to that which links the other half; and there will be found a critical point at which the total number of linkages is zero. Individual linkages can be positive or negative. Hence, at the critical point, M will be zero, and there will be no output.

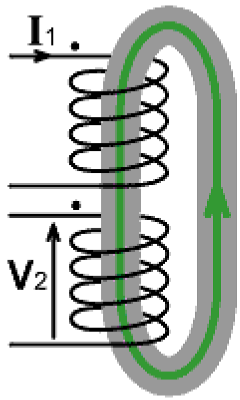
If the pick up coil is rotated through a full 180°, the polarity of the voltage $\mathbf{V}_2$ will be reversed. Notice however, that there will be no intermediate output phases during the course of the rotation; the voltage will simply diminish with constant phase until the 90° point is reached, and then reappear inverted. Hence mutual inductance is pseudoscalar; it can be positive or negative, but it has no intermediate directions because it is conceived as the sum of positive and negative linkages per unit current. A surprising corollary of this observation is that M must be given by a two-valued function; and in view of the parameters from which it has to be derived, it must be related to the geometric mean of the inductances of the two coils. Hence we can deduce the general form:

$M = k\,\surd(L_1 L_2)$ [Henrys]

where k is a dimensionless factor, called the ***coupling coefficient***, which we will shortly derive from the field topology. Notice that, if we take the self-inductances as given, k provides a dimensionless alternative to mutual inductance, and will therefore turn out to be a convenient circuit-analysis parameter.

Before we declare unequivocally that M can be either positive or negative however, we must be mindful of convention. When analysing circuits which have a fixed physical configuration, it is good practice to mark the start of a winding with a dot. Hence, upon 180° rotation of the pick-up coil (or reversal of the winding direction, which amounts to the same thing), the dot jumps from one end of the winding to the other. In that case, if the circuit-analysis arrow adjacent to $\mathbf{V}_2$ is always drawn pointing towards the dot, then M will remain unchanged. Hence M can usually be made positive by considering the phasing of the windings; and this is generally a sensible idea. There is one situation in which M must be allowed to change polarity however, and that is when dealing with variometers; a variometer being a pair of coils with a control shaft allowing one to be rotated relative to the other (see illustration).

Although we will shortly consider the general case in which the coupling between coils is incomplete; it is instructive to begin with the special limiting case, never quite achievable in practice, in which all of the flux from one coil links with the turns of another. This, of course, is the condition which defines an ideal transformer. One way to approximate complete linkage (neglecting internal inductance) is to use twisted bifilar winding (as in the example given in the previous section); but that is only possible when both coils have the same number of turns (and the distributed capacitance between the wires complicates matters at high frequencies). To link the flux (almost) completely between coils having arbitrary numbers of turns, a magnetic circuit of low reluctance (i.e., a transformer core) is required. For circuit-analysis purposes however, we are not concerned with how it is done; provided that the models we end up with can be corrected for the deficiencies of practical devices.

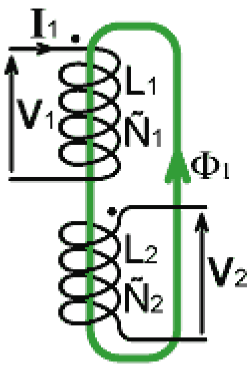


The logical step which will enable us to deduce the input-output relationships (transfer functions) for any system of coupled inductors in the lumped-component limit (i.e., when conductor lengths and the distances between coils are small in comparison to wavelength), lies in the observation that flux linkage is a relationship between the flux through the path and the turns in a particular coil. In other words; flux linkage does not figure directly in the coupling between separate circuits; because the two coils are coupled to the flux, not to each other. Hence the flux on a path through several coils is notionally de-linked from the originating coil and re-linked to the others. When we apply this idea to the simple ideal transformer case, the generalisation will soon become apparent.

Note that, in much of the working to follow, we will assume that all coils are wound with wire which has no resistance. This is permissible in the limit that there are no capacitive currents (i.e., all current follows the wires), because conductor resistance, like internal inductance, is associated with energy-exchange processes in the body of the wire, and is therefore excluded from the magnetic interaction (to a very good approximation at least). Conductor losses can therefore be included in circuit analysis as a separate lumped component in series with the wire. Magnetic losses are a separate matter, and can be handled by defining permeability (and hence inductance) as complex; but for the present purpose we will assume that magnetic media are ideal.

11. Mutual inductance, ideal case

The diagram below gives a topological representation of the coupling which takes place in an ideal transformer; i.e., the spatial distribution of the field is ignored, and only the route taken by the magnetic flux as it threads through the coils is considered. Information about path length and area has not been discarded however, because it is wrapped up in the inductances and turns numbers, i.e., in the parameters L_1, $\tilde{N}_1$, L_2 and $\tilde{N}_2$. Note incidentally; that it might be argued that non-idealities such as linkage inefficiency can be ignored in any discussion of ideal transformers. In the sections to follow however, we will retain the distinction between effective turns and actual turns in order to avoid confusing two separate physical phenomena; these being: *inductance which fails to materialise* (because linking is inefficient); and *inductance which fails to couple with other inductances* (because not all of the flux can loop around the turns of other coils).



The self-induced voltage across the coil L_1 is equal to the instantaneous applied voltage V_1 . It is also equal to the rate of change of flux linkages in L_1 per unit time. Hence:

$V_1 = \partial\Lambda_1 / \partial t = \tilde{N}_1 \, \partial\Phi_1 / \partial t$

Thus the rate of change of flux in the magnetic medium, i.e., the actual flux after de-linking from L_1 is:

$\partial\Phi_1 / \partial t = V_1 / \tilde{N}_1$

The instantaneous voltage induced across the coil L_2 is given by the rate of change of its flux linkages per unit time. Hence:

$V_2 = \partial\Lambda_{21} / \partial t = \tilde{N}_2 \, \partial\Phi_1 / \partial t$

Thus:

$V_2 = V_1 \, \tilde{N}_2 / \tilde{N}_1$

Now recall that:

$\tilde{N} = N \sqrt{k_H}$

For a transformer to be ideal, all of its inductance must be coupled. Therefore it must either be a 1:1 autotransformer, or the coils must have no internal inductance and must be equally subject to any field inhomogeneities. It follows that:

$k_{H1} = k_{H2}$

and so:

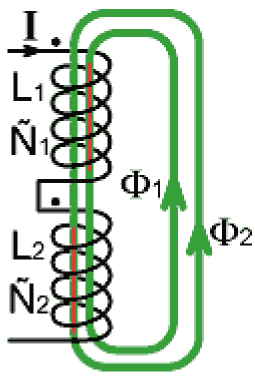
$V_2 = V_1 \, N_2 / N_1$

This, of course, is the classic turns-ratio rule for the ideal transformer; and so offers no surprises, except perhaps for the simplicity of its derivation. A point to note here however is that, although linkage inefficiency is eliminated from the turns *ratio*, it is not eliminated from the process of inductance calculation; so we cannot simply replace every instance of Ñ with N.

Now, mindful that the back voltage is equal to the applied voltage, the inductance L_1 is defined in the relationship:
$V_1 = L_1\, \partial I_1 / \partial t$
Using this as an analogy, we can propose a definition for mutual inductance having the same form; i.e.:
$V_2 = M\, \partial I_1 / \partial t$
According to this definition, the voltage ratio is:
$V_2 / V_1 = M / L_1$ . . . . (**11.1**)
In order to make the next logical step, it is necessary to establish an important point. Mutual inductance is not merely a circuit-analysis parameter; it is an actual physical inductance. It represents an energy-storage mechanism which is accessible to a pair of coils, whereas pure self-inductance is an energy-storage mechanism which is only accessible to a single coil. This means that the mutual inductance cannot change when the roles of the two coils are transposed. Thus we can swap the input and the output; and in terms of the analysis, the only effect will be to swap the subscripts 1 and 2. In this way, V_2 becomes the input voltage, and V_1 becomes the output, and we get:
$V_1 / V_2 = M / L_2$
Thus:
$M / L_1 = L_2 / M$
i.e.:
$M = \sqrt{(L_1 L_2)}$
M is the geometric mean of the coupled inductances. Recall, from the earlier discussion, that the expression for mutual inductance was expected to be of the form:
$M = k \sqrt{(L_1 L_2)}$
Hence, the coupling coefficient, k, for an ideal transformer is 1 (or -1 if we point one of the voltage arrows away from the dot).

12. Windings in series, ideal case

When two windings on the same transformer are placed in series, a single inductance is obtained; but due to the coupling between the coils, the total inductance is not the same as the sum of the two separate inductances. The reason for that, of course, is that the flux from coil 1 links with the turns of coil 2, and vice versa. This situation is represented topologically in the diagram below. The two fluxes are, of course, superposed (and therefore indistinguishable) in the actual field; but it is perfectly legitimate to consider them separately for accounting purposes. Now (noting that both coils carry the same current) we can determine the inductance, by inspection, as the total number of flux linkages per unit current; i.e.:
$L = [\,\Lambda_1 + \Lambda_2 + \Phi_1 \tilde{N}_2 + \Phi_2 \tilde{N}_1\,] / I$



The terms are, in order: the flux linkage in coil 1 due to its own current; the flux linkage in coil 2 likewise; the linkages in coil 2 due to the flux from coil 1; and the linkages in coil 1 due to the flux from coil 2. We can also make the following substitutions:
$\Lambda_1 / I = L_1$, $\Lambda_2 / I = L_2$, $\Phi_1 = \Lambda_1 / \tilde{N}_1$, $\Phi_2 = \Lambda_2 / \tilde{N}_2$
Hence:
$L = L_1 + L_2 + L_1 (\tilde{N}_2 / \tilde{N}_1) + L_2 (\tilde{N}_1 / \tilde{N}_2)$
But, from the previous section, we have:
$\tilde{N}_2 / \tilde{N}_1 = M / L_1 = L_2 / M$
Hence
$L = L_1 + L_2 + 2M$
Now consider what happens when the connections to (say) L_2 are reversed. In that case, the number

of turns $\tilde{N}_2$ becomes negative *and* the direction arrow for the flux Φ_2 is reversed (i.e., Φ_2 becomes negative also). Thus the cross linkages, $\Phi_1\tilde{N}_2$ and $\Phi_2\tilde{N}_1$, both become negative. Hence:
$L = L_1 + L_2 - 2M$
The two fluxes in superposition tend to cancel. The amount of energy which can be stored in the field for a given current is reduced, and so the inductance is reduced. Thus there are two configurations for coupled inductors in series, known as ***series aiding***, and ***series opposition***; and in general, the expression for the total inductance is:
$L = L_1 + L_2 \pm 2M$
One property of this expression which is not immediately obvious is that the total inductance is always positive (assuming that L_1 and L_2 are positive); even when the coils are in opposition and the coil deemed to have negative turns has more turns than the other. This can be seen by considering the requirement for complete inductance cancellation (which is only possible in the ideal case), i.e. L=0 and so:
$L_1 + L_2 = 2M$
Using the substitution: $M=\sqrt{(L_1 L_2)}$, we get:
$L_1 + L_2 = 2 \sqrt{(L_1 L_2)}$
and squaring both sides gives:
$(L_1 + L_2)^2 = L_1^2 + L_2^2 + 2 L_1 L_2 = 4 L_1 L_2$
Hence:
$L_1^2 + L_2^2 = 2 L_1 L_2$
$L_1^2 + L_2^2 - 2 L_1 L_2 = 0$
i.e.:
$(L_1 - L_2)^2 = 0$
Complete cancellation can only occur when $L_1 = L_2$, and so the mutual inductance term cannot be greater than the sum of the separate inductances. This constraint also sets an upper limit on the inductance which can be obtained when the coils are series aiding, i.e., when
$L = L_1 + L_2 + 2\sqrt{(L_1 L_2)}$ *and* $L_1 = L_2$ *then* $L = 4L_1$
What this says is that when the number of turns in a fully-linked coil is doubled, then the inductance is increased by a factor of 4. This rule, of course, is also captured in the expression for L in terms of the transformer-core inductance factor:
$L = \mathbf{A_L} \tilde{N}^2$ where $\mathbf{A_L} = \mu \mathbf{A_e} / \boldsymbol{\ell_e}$
Earlier we gave the inductance for two windings in series as:
$L = L_1 + L_2 + L_1 (\tilde{N}_2 / \tilde{N}_1) + L_2 (\tilde{N}_1 / \tilde{N}_2)$
Putting this in terms of the inductance factor we have:
$L = \mathbf{A_L} [\tilde{N}_1^2 + \tilde{N}_2^2 + \tilde{N}_1^2 (\tilde{N}_2 / \tilde{N}_1) + \tilde{N}_2^2 (\tilde{N}_1 / \tilde{N}_2)]$
i.e.:
$L = \mathbf{A_L} (\tilde{N}_1^2 + \tilde{N}_2^2 + 2\tilde{N}_1\tilde{N}_2)$
which factorises:
$L = \mathbf{A_L} (\tilde{N}_1 + \tilde{N}_2)^2$
This gives a slightly broader interpretation of the inductance factor formula because, in the series opposition case, we can consider the quantity $\tilde{N}_1+\tilde{N}_2$ to represent the effective number of turns as the sum of positive and negative turns; i.e.:
$L = A_L \tilde{N}^2$ where $\tilde{N} = \tilde{N}_1 + \tilde{N}_2$
with $\tilde{N}_1$ (say) positive and representing turns wound clockwise around a flux which recedes when the current is positive; and $\tilde{N}_2$ negative and representing turns wound anticlockwise around the same flux. No matter whether the negative number is greater in magnitude than the positive number however; the inductance is always positive because it is proportional to the square of the total.

13. Mutual inductance, general case

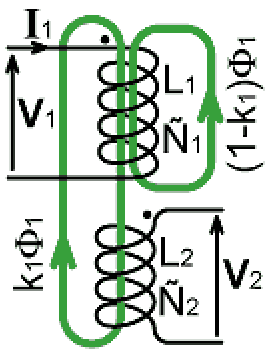
The simple flux-linkage counting principle is straightforwardly extended to the general (i.e., non-ideal) case. Starting with the unloaded transformer, we can say that the driven coil (which we may refer to as the 'primary' when it is the only coil connected to an active network) produces a flux:

$\Phi_1 = \Lambda_1 / \tilde{N}_1$

Of this however, only a proportion, $k_1\Phi_1$ (say), threads through the turns of L_2, and the remainder, $(1\text{-}k_1)\Phi_1$, is private to L_1. Notice here that the communal flux is shown, in the diagram below, as threading through all of the (effective) turns of L_2. What happens in reality is a mixture of complete and partial linkages; but from a topological point-of-view, the only necessary condition is that we count the total number correctly.

Now, as before, we relate the voltage driving L_1 to the rate of change of flux linkages; i.e.:

$V_1 = \partial\Lambda_1 /\partial t = \tilde{N}_1\ \partial\Phi_1 /\partial t$



Thus, bearing in mind that only a proportion ($\times k_1$) of the flux threads L_2, we determine the induced voltage:

$V_2 = \partial\Lambda_{21} /\partial t = \tilde{N}_2\ k_1\ \partial\Phi/\partial t$

Hence:

$V_2 / V_1 = k_1\ \tilde{N}_2 / \tilde{N}_1$

From this, it can be seen that the coefficient k_1 is the factor by which the secondary induced-voltage falls short of that predicted from the turns ratio. Defining mutual inductance as before, we have:

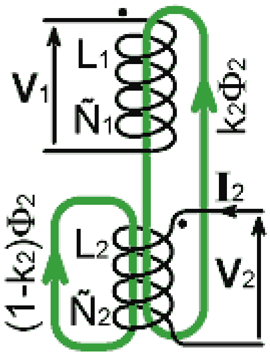
$V_2 = M\ \partial I_1 /\partial t$ and $V_1 = L_1\ \partial I_1 /\partial t$

Hence:

$V_2 / V_1 = M / L_1 = k_1\ \tilde{N}_2 / \tilde{N}_1$

Now, in order to complete the definition of mutual inductance, we swap the roles of the two windings. Thus:

$V_1' / V_2' = M / L_2 = k_2\ \tilde{N}_1 / \tilde{N}_2$



where the voltage-shortfall factor, k_2, which results from using L_2 as primary is, in general, not identical to and rarely exactly the same as k_1 (for which reason the voltages have been primed, to show that their ratio is not the reciprocal of that which appears in the preceding equation). Thus:

$M / (L_1\ k_1) = L_2\ k_2 / M$

and so the general expression for mutual inductance is:

$$M = \pm\sqrt{(L_1\ k_1\ L_2\ k_2)}$$

M can be either positive or negative because it is given by a square-root function; but it is, by convention, taken to be positive when the circuit-analysis currents flow towards the dots. The expression can also be written:

$$M = \pm k \sqrt{(L_1\ L_2)}$$

Where the coupling coefficient, k, is the geometric mean of the two voltage-shortfall factors, i.e.;

$k = \pm\sqrt{(k_1\ k_2)}$

Now notice from the working above that the transformer voltage ratios can be expressed using only the inductances and the coupling coefficient k. This follows because:

$V_2 / V_1 = M / L_1 = k\ [\sqrt{(L_1\ L_2)}] / L_1$

i.e.:

$$V_2 / V_1 = k \sqrt{(L_2 / L_1)}$$

and similarly, when L_2 is used as the primary:

$V_1'/ V_2' = k \sqrt{(L_1 / L_2)}$

This means that the basic properties of a two-coil transformer can be specified using only three parameters, either $\{L_1, L_2, M\}$ or $\{L_1, L_2, k\}$; the latter choice being used for the standard SPICE model and having the convenient property that k is dimensionless and lies between 0 and 1.

Knowledge of the turns ratio is not necessary for circuit-modelling purposes (although it is needed for the physical design). Likewise it is not necessary to specify the individual voltage shortfall factors (k_1 and k_2) explicitly; but note that this latter contraction can lead to a logical trap.

Some influential authors, notably Terman[9] and those who have followed his lead, make the mistake of assuming that because only one k is needed, then, $k = k_1 = k_2$. This implies that the extent to which the flux from one coil fails to couple with another is dictated by the extent to which a flux from the other coil fails to couple with it. Other than by accidental or carefully-engineered coincidence, there is no physical basis on which that can be expected to happen. It is a fallacy, because there will always be some proportion of the inductance of a coil which does not couple with other coils and which is not under the control of other coils (internal inductance for example). The consequence of the logical error is to give the transformer model an inherent property of voltage-shortfall reciprocity which actual devices do not posses. As will be shown later, this is equivalent to assuming that the off-load output voltage of a potential divider remains the same when the input and output terminals are swapped. This mathematical inconsistency renders transformers effectively incomprehensible to those who have failed to notice it.

With regard to the turns-ratio voltage-transformation rule of the ideal transformer, it is often stated that this is no longer meaningful for loosely-coupled transformers. From the working above however, we can see that the relationship has not gone away, it has simply become proportional rather than absolute. Furthermore, the parameters which dictate the constants of proportionality are governed by the physical dimensions and permeabilities of the magnetic paths, and so remain nearly constant when the turns numbers are changed.

Recall that the effective number of turns in a coil is defined as:

$\tilde{N} = N \sqrt{k_\mathbf{H}}$

where, in the case of a current-sheet solenoid, $k_\mathbf{H} = k_\mathbf{L}$ (Nagaoka's coefficient). Hence, for pairs of current-sheet solenoids in proximity:

$V_2 / V_1 = k_1 \tilde{N}_2 / \tilde{N}_1 = k_1 (N_2 / N_1) \sqrt{(k_{\mathbf{L2}} / k_{\mathbf{L1}})}$

and

$V_1 / V_2 = k_2 \tilde{N}_1 / \tilde{N}_2 = k_2 (N_1 / N_2) \sqrt{(k_{\mathbf{L1}} / k_{\mathbf{L2}})}$

$k_\mathbf{L}$ depends only on the length / diameter ratio of the solenoid. It does not change with the number of turns, provided that the winding does not change in length (or diameter) when turns are added or subtracted. Hence, for an air-cored transformer considered in the current-sheet approximation; the induced voltage depends on the turns ratio, the distance between the coils, and the overall shapes of the coils. It does not depend strongly on the turn spacings and wire diameters, provided that the wire diameter is small in comparison to the coil diameter. Even so, the calculation of M (or k) from field considerations is somewhat involved. For those interested in doing so, the subject is covered in detail in F W Grover's monograph[10] and in the (public domain) NBS science paper[11] on which the later book is based.

Note, incidentally, that in the current sheet approximation (i.e., neglecting internal inductance, and other small corrections), k_2 can be made the same as k_1 by symmetry. If the two solenoids are the same in both length and diameter, then the overall field patterns will be practically identical, and the induced voltage shortfall will be the same either-way. Also, in that case, Nagaoka's coefficient will be the same for the two coils, and so we get:

$V_2 / V_1 = k N_2 / N_1$

Thus it is possible to engineer voltage reciprocity to a good first-order approximation; but it is not sensible to assume that such reciprocity is a natural property of transformers.

---

9 **Radio Engineering**. F E Terman. 3rd edn. McGraw-Hill. 1947. p56.

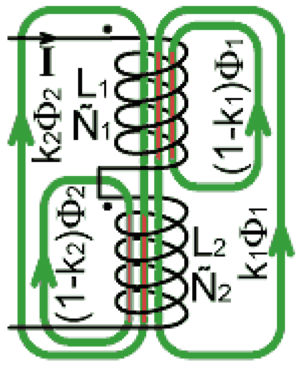
10 **Inductance Calculations: Working Formulas and Tables.** F W Grover, 1946 and 1973. Dover Phoenix Edition 2004. ISBN: 0 486 49577 9.

11 **Formulas and Tables for the Calculation of Mutual and Self Induction**. E B Rosa & F W Grover, 3rd edition 1916, with 1948 corrections. Bureau of Standards scientific paper No. 169. [BS Sci. 169]

14. Coupled inductors in series, general case

Having established what is meant by incomplete coupling, determining the inductance of a series combination of coupled inductors is now straightforward. It is, of course, the total number of linkages per unit current; i.e.:

$L = [\, \Lambda_1 + \Lambda_2 + \Lambda_{21} + \Lambda_{12} \,] / I$



where the last two terms are respectively: the linkages in coil 2 due to the flux from coil 1; and vice versa. Now identifying:

$\Lambda_{21} = \tilde{N}_2\, \Phi_{21}$ where $\Phi_{21} = k_1\, \Phi_1$ and $\Phi_1 = \Lambda_1 / \tilde{N}_1$

and vice versa, we have:

$L = L_1 + L_2 + [\, (\tilde{N}_2 / \tilde{N}_1)\, k_1\, \Lambda_1 + (\tilde{N}_1 / \tilde{N}_2)\, k_2\, \Lambda_2 \,] / I$

i.e.:

$L = L_1 + L_2 + (\tilde{N}_2 / \tilde{N}_1)\, k_1\, L_1 + (\tilde{N}_1 / \tilde{N}_2)\, k_2\, L_2$

but, from the working in the previous section:

$M = (\tilde{N}_2 / \tilde{N}_1)\, k_1\, L_1 = (\tilde{N}_1 / \tilde{N}_2)\, k_2\, L_2$

Hence:

$L = L_1 + L_2 + 2M$

Also, if we swap the connections to one of the coils, we make both the turns in that coil and the flux from that coil negative; causing both of the mutual inductance terms to change sign. Thus, in general:

$L = L_1 + L_2 \pm 2M$

The expression is the same as for the ideal case, the difference being that the magnitude of the mutual inductance is not at its theoretical maximum. An equivalent expression is:

$L = L_1 + L_2 \pm 2k\, \sqrt{(L_1\, L_2)}$

where

$k = \pm\sqrt{(k_1\, k_2)}$

which gives the ideal case when k=1.

15. Measuring M

One matter which must be understood is that a pair of coupled inductors with a coupling coefficient falling considerably short of unity does not constitute a failed transformer design. On the contrary, there are many situations in which a small value of k is desirable; particularly in the design of bandpass filters. An important application (which we will examine later) lies in the flat-topped bandpass characteristic which can be obtained by using a pair of loosely-coupled LC parallel resonators. This, of course, is the basis of the ubiquitous superheterodyne intermediate-frequency (IF) transformer.

In contrast to the usefulness of loosely coupled transformers, there are fewer obvious applications for series-connected pairs of loosely-coupled inductors. Instead, the utility of the configuration lies in the fact that it can be used to measure the mutual inductance of a pair of coils which are intended for use as a transformer. If the coils are connected series aiding, then the inductance is:

$L_+ = L_1 + L_2 + 2M$

And when the coils are connected in series opposition, the inductance is:

$L_- = L_1 + L_2 - 2M$

Subtracting the latter from the former gives:

$L_+ - L_- = 4M$

If the inductances of the coils are measured separately, k can also be determined; i.e.:

$k = M / \sqrt{(L_1\, L_2)}$

16. Magnetic path analysis

The concept of 'effective turns number', Ñ, was introduced earlier as a logical necessity in ensuring that flux-linkages are counted correctly. It is involved in the definition of inductance twice, once in the determination of MMF, and once in the counting of linkages to the resulting flux. Hence the inductance of a coil is given by an expression:

$L = \tilde{N}^2 \mu A / \ell$

This can be turned into an expression involving actual turns, thus:

$L = N^2 k_H \mu A / \ell$

Where k_H accounts for any inductance shortfall due to flux loops not enclosing all of the turns. From this we obtain the identity of the linkage efficiency factor:

$\sqrt{k_H} = \tilde{N} / N$

We can also aggregate the path geometry and permeability parameters into a single inductance factor; which is defined as:

$A_L = \mu A / \ell$

Hence:

$L = \tilde{N}^2 A_L = N^2 k_H A_L$

The use of inductance factors is familiar practice when dealing with transformer cores; but here we generalise it to any coil, with the proviso that it is strictly the effective number of turns, not the actual number, which is required when determining inductance. The square of the linkage efficiency (i.e., k_H) is associated with the turns, not with the inductance factor; because A_L is a measure of magnetic conductance. In other words, we argue that k_H cannot be made to disappear by factoring it into A_L, because the reluctance of a path is not affected by the efficiency with which the fluxes from the individual turns of a particular coil are injected into it.

Now, having separated linkage efficiency from the properties of the magnetic path; we can create a complete topological model from which the physical identities of the voltage-shortfall factors, the coupling coefficient, and the mutual inductance can be deduced. The key to what follows lies in making use of a relationship which exists between magnetic flux and the generalised inductance factor. Recall that inductance is defined as the number of flux linkages per unit current. Thus we have several alternatives:

$L = \tilde{N}^2 A_L = \Lambda / I = \tilde{N} \Phi / I$

Hence:

$\tilde{N}^2 A_L = \tilde{N} \Phi / I$

i.e.:

$\Phi = \tilde{N} I A_L$	**16.1**

Notice here that ÑI is magnetomotive force (MMF). Hence this is a statement of Ohm's law for magnetic circuits:

Flux = MMF × magnetic conductance = MMF / reluctance

'Magnetic conductance' is the reciprocal of reluctance (i.e., magnetic resistance, *S*), which means that (**16.1**) is exactly analogous to the statement:

$I = V G = V / R$ (Ohm's Law)

(G being the standard symbol for electrical conductance). It follows that we can complete the topological theory of transformers by using a version of admittance theory involving magnetic parameters; except that it is easier than general admittance theory because the behaviour we wish to analyse involves only instantaneous fields; i.e., complex numbers are not involved. Phasors come later, when we use Faraday's law to relate currents to induced voltages, using the mapping outlined at the end of section **8**. The instantaneous voltage induced in a pickup coil is, of course, the rate of change of flux on the path multiplied by the effective number of turns. By differentiating (**16.1**) with respect to time we get:

$\partial\Phi/\partial t = \tilde{N} A_L \partial I/\partial t$

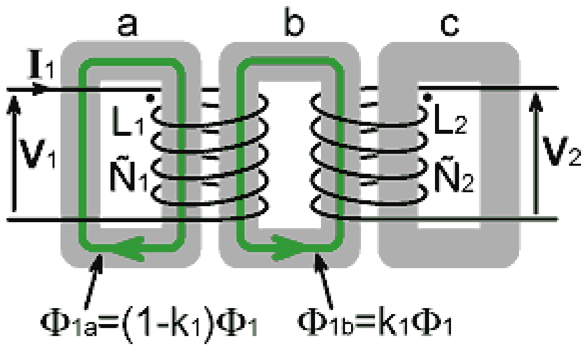
and so we determine induced voltages using expressions of the form:
$V = \tilde{N}\ \partial\Phi/\partial t = \tilde{N}^2 A_L\ \partial I/\partial t = L\ \partial I/\partial t$ (cf. Faraday's Law)
The collection of the physical path parameters into a single coefficient A_L constitutes the mapping from the field to the topological representation; and Faraday's Law will later give us the mapping into circuit theory.

In section **14**, the magnetic flux associated with a pair of coupled inductors was separated into four parts. These were: the flux which is private to coil 1; the flux which is private to coil 2; the flux from coil 1 which links with the turns of coil 2; and the flux from coil 2 which links with the turns of coil 1. In the context of transformer theory, the two private fluxes are known as ***leakage fluxes***; i.e., they are the fluxes which escape from involvement in the coupling. The magnetic system so defined is also meant to account for *all* of the flux; i.e., any flux line which fails to link with the turns of both coils belongs, by definition, to the leakage flux of the coil which produced it. In this way, the four fluxes are notionally divided between three magnetic paths; these being: the communal path and the two leakage paths.

The general situation, as just described, is captured in the topological model shown below. One obvious feature of this model, is that it can also be taken to represent a structure consisting of two coils wound on three transformer cores; the behaviour of the system being considered in the limit where the magnetic flux outside the cores (or paths) is negligible. If we also take the liberty of allowing the term 'ideal transformer' to be used in cases when the coils have $k_H<1$; then we have an ideal transformer made non-ideal by the provision of leakage paths; an arrangement which is exactly analogous to the way in which coupling occurs in practical devices.

An important property of the model is that; if the reluctance is finite on a given path, but elsewhere infinite; then the flux from a particular coil has no choice but to occupy the paths provided for it. Hence, at least notionally, we can calculate the inductance L_1 from the turns and the inductance factors A_{La} and A_{Lb} .

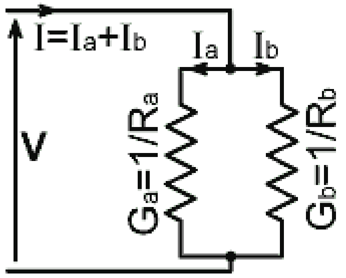
The overall inductance factor which results when a coil is wound around a pair of transformer cores is easily derived from first principles. The applied voltage is equal to the back voltage, which is in turn equal to the rate of change of linkages per unit time. If the flux is linked to every turn, then the total number of linkages is simply $\tilde{N}\Phi$; and since there are two paths, the flux is divided into two parts. Hence, for coil 1 we can write:

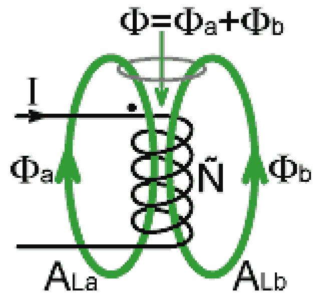


$V_1 = \partial\Lambda_1/\partial t = \tilde{N}_1\ \partial\Phi_1/\partial t$
$\quad = \tilde{N}_1\ \partial(\Phi_{1a} + \Phi_{1b})/\partial t = L_1\ \partial I_1/\partial t$
but earlier we determined the relationship (**16.1**):
$\Phi = \tilde{N}\ I\ A_L$
Thus:
$L_1\ \partial I_1/\partial t = \tilde{N}_1\ \partial(\tilde{N}_1\ I\ A_{La} + \tilde{N}_1\ I\ A_{Lb})/\partial t$
$\quad = \tilde{N}_1^2\ (A_{La} + A_{Lb})\ \partial I_1/\partial t$
i.e.:
$L_1 = \tilde{N}_1^2\ (A_{La} + A_{Lb})$
We can also use an identical method (but with different subscripts) to derive the inductance of coil 2. Thus:
$L_2 = \tilde{N}_2^2\ (A_{Lb} + A_{Lc})$
When a coil is wound around several paths, the overall inductance factor is simply the sum of the individual inductance factors. This can also be deduced directly by analogy with admittance theory; because we know that the overall conductance of a set of parallel conductances is obtained by addition (i.e.; $1/R = 1/R_1 + 1/R_2 + \ldots$).

Admittance theory will take us further than that however. Just as we can calculate the currents in a set of parallel resistors as a proportion of the total, we can calculate the relative fluxes on a set of magnetic paths. Analogous electrical and magnetic systems are depicted in the diagrams below;

and if we bear in mind that flux (Φ) is the counterpart of current; inductance factor (A_L) is the counterpart of conductance (G); and MMF ($\tilde{N}I$) is the counterpart of EMF (V); then it is apparent that we can produce expressions for the partial fluxes Φ_a and Φ_b .





In the electrical case, if $G=G_a+G_b$, and $R=R_a//R_b=R_aR_b/(R_a+R_b)$, then;
$EMF = V = IR = I/G = I/(G_a+G_b)$
$I_a = V/R_a = V\,G_a = I\,G_a / (G_a+G_b)$
Similarly:
$I_b = I\,G_b / (G_a+G_b)$

In the magnetic case, if $A_L=A_{La}+A_{Lb}$, then;
$MMF = \tilde{N}I = \Phi S = \Phi/A_L = \Phi/(A_{La}+A_{Lb})$
Hence:
$\Phi_a = \tilde{N}I\,A_{La} = \Phi\,A_{La}/(A_{La}+A_{Lb})$
and
$\Phi_b = \Phi\,A_{Lb}/(A_{La}+A_{Lb})$

We can now apply this generalised result to the three-path model for a pair of coupled inductors. Using the nomenclature given earlier, for the leakage flux of coil L_1 :
$\Phi_{1a} = (1-k_1)\,\Phi_1 = \Phi_1\,A_{La}/(A_{La}+A_{Lb})$
i.e.;
$1-k_1 = A_{La}/(A_{La}+A_{Lb})$
and for the coupled flux:
$\Phi_{1b} = k_1\,\Phi_1 = \Phi_1\,A_{Lb}/(A_{La}+A_{Lb})$
i.e., the voltage shortfall factor is:
$k_1 = A_{Lb}/(A_{La}+A_{Lb})$ . . . . (**16.2**)
Similarly, for the coil L_2 :
$k_2 = A_{Lb}/(A_{Lb}+A_{Lc})$
and
$1-k_2 = A_{Lc}/(A_{Lb}+A_{Lc})$
Now, recall from section **13**, that the transformer coupling coefficient is the geometric mean of the voltage shortfall factors, i.e.:
$k = \pm\sqrt{(k_1\,k_2)}$
Thus:

$k = \pm\, A_{Lb} / \sqrt{[\,(A_{La}+A_{Lb})\,(A_{Lb}+A_{Lc})\,]}$	**16.3**

Also recall from section **13** that:
$M / L_1 = k_1\,\tilde{N}_2 / \tilde{N}_1$ and $M / L_2 = k_2\,\tilde{N}_1 / \tilde{N}_2$
i.e.:
$M = k_1\,L_1\,\tilde{N}_2 / \tilde{N}_1 = k_2\,L_2\,\tilde{N}_1 / \tilde{N}_2$
but

$L_1 = \tilde{N}_1^2 (A_{La}+A_{Lb})$ and $L_2 = \tilde{N}_2^2 (A_{Lb}+A_{Lc})$
Substituting for L_1 in the expression for M we get:
$M = k_1 (A_{La}+A_{Lb}) \tilde{N}_1^2 \ \tilde{N}_2 / \tilde{N}_1$
and substituting for k_1 gives:
$M = \tilde{N}_1 \tilde{N}_2 (A_{La}+A_{Lb}) A_{Lb} / (A_{La}+A_{Lb})$
i.e.:

$M = \tilde{N}_1 \tilde{N}_2 \ A_{Lb}$	**16.4**

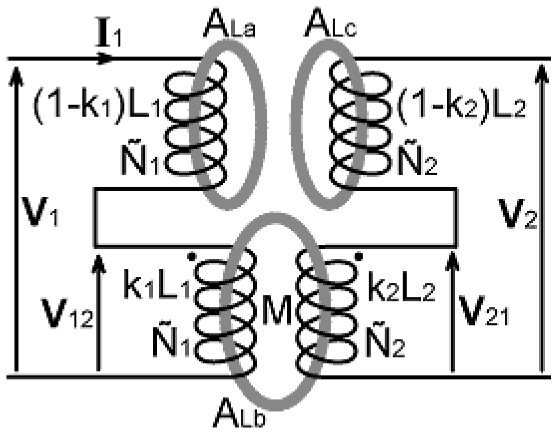
The same result will be obtained if L_2 and k_2 are used in the derivation. Mutual inductance, perhaps unsurprisingly, turns out to be the geometric mean of those parts of the coupled inductances which share a common magnetic path. Note incidentally, that in this formulation (not involving a square-root), the sign of M is dictated by the relative signs of $\tilde{N}_1$ and $\tilde{N}_2$; i.e., if the turns are in opposition, it is negative.

Unfortunately, knowing that M is governed by the inductance factor of the shared path does not simplify the process of calculating it for pairs of loosely coupled coils with non-magnetic cores; the reason being that the difficulty of calculating A_{Lb} is no less than that of calculating M itself. It does however permit quantitative analysis of systems based on high-permeability transformer cores. The three-path topological model also suggests ways of using magnetic cores to make loosely-coupled transformers for bandpass filters, one possible advantage being minimal stray magnetism (i.e., no need for a screening can).

17. Reciprocity failure

Some important general points are raised by the preceding discussion. Most significantly, the magnetic path through an inductor can be arbitrarily divided into sub-paths, and the overall inductance is given by the sum of the inductances on the separate paths. This means that, for circuit analysis purposes, the leakage inductance and the coupled inductance associated with a given coil can be treated as separate elements in series. Furthermore, the leakage inductance can be subdivided into (say) internal inductance and external uncoupled inductance, one possible reason for doing so being that these different parts have different frequency dependences and are described by very different mathematical expressions.

As has already been mentioned in section **13**, voltage shortfall reciprocity is not an inherent property of transformers. This can be understood from the working in the previous section, because the three path model can be constructed using (say) toroidal inductor cores, and it is obvious that we are at liberty to choose radically different values for A_{La} and A_{Lc}. The point is perhaps most convincingly made however, if we separate the coupled inductances from the leakage inductances, as the path analysis permits. This is done in the diagram on the right, and results in an electrical model comprising a pair of completely-independent leakage inductances in series with the windings of an ideal transformer. It is for this reason, incidentally, that the ideal transformer is important; not because it is realisable, but because one lies (analytically) at the heart of every practical transformer.



An ideal transformer is completely specified by its A_L value and its turns numbers. Because we have separated the inductances involved into pairs of elements in series however, it might seem that we have lost knowledge of how many turns there are in the coupled inductances; but it transpires that

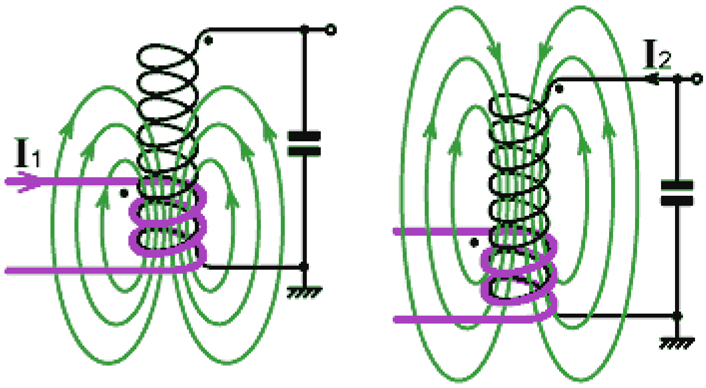
we have not. From an analytical viewpoint, the partial inductances all have the same number of turns as the parent inductance. This curious fact is easily proved as follows:
The two paths accessible to L_1 are A_{La} and A_{Lb} and the coil has $\tilde{N}_1$ turns. Hence:
$L_1 = \tilde{N}_1^2 (A_{La}+A_{Lb})$
The associated leakage inductance is:
$(1\text{-}k_1) L_1 = (1\text{-}k_1) \tilde{N}_1^2 (A_{La}+A_{Lb})$
but, from equation (**16.2**):
$k_1 = A_{Lb}/(A_{La}+A_{Lb})$
Hence:
$(1\text{-}k_1) L_1 = [1\text{-}A_{Lb}/(A_{La}+A_{Lb})] \tilde{N}_1^2 (A_{La}+A_{Lb})$
multiplying out:
$(1\text{-}k_1) L_1 = (A_{La} + A_{Lb} - A_{Lb}) \tilde{N}_1^2$
Hence, the number of turns in the leakage-inductance component of L_1 is $\tilde{N}_1$.
The coupled inductance component of L_1 is:
$k_1 L_1 = \tilde{N}_1^2 (A_{La}+A_{Lb}) A_{Lb} / (A_{La}+A_{Lb})$
Hence, the number of turns in the coupled part of the inductance is also $\tilde{N}_1$. A similar exercise (with different subscripts) will show that the components of L_2 all have $\tilde{N}_2$ turns. The reason is that the separability of a path into sub-paths gives rise to a set of inductances which must be added to obtain the complete inductance; and in circuit-analysis terms, that corresponds to a set of inductances in series. The number of turns around each sub-path however, is always the same as the number of turns in the whole coil. This artefact of the analysis is, from a modelling standpoint, largely irrelevant in the case of leakage inductances, but useful with regard to the coupled inductances. Thus (presuming that we can determine the linkage efficiency factors) we know the turns-ratio, and hence the voltage ratio, of the ideal-transformer part of the model.

Now we can explain the lack of voltage-shortfall reciprocity by treating the series elements as potential dividers. Referring again to the diagram; when the current I_1 is flowing, the voltage (V_{12}) at the input to the ideal transformer is determined by the ratio of the impedances of the leakage inductance and the coupled inductance. Hence:
$V_{12} = k_1 V_1$
The output of the ideal transformer (the induced voltage) is given by the turns ratio, and so:
$V_{21} / V_{12} = \tilde{N}_2 / \tilde{N}_1$
i.e.:
$V_{21} / V_1 = k_1 \tilde{N}_2 / \tilde{N}_1$
But when the transformer is unloaded, the voltage drop across the secondary leakage inductance is zero, and so $V_{21} = V_2$. Hence:
$V_2 / V_1 = k_1 \tilde{N}_2 / \tilde{N}_1$
If we swap the roles of the two windings and repeat the exercise, we get:
$V_1' / V_2' = k_2 \tilde{N}_1 / \tilde{N}_2$
where the voltages have this time been given primes, as a reminder that they are not the same as in the previous case. Voltage reciprocity only occurs when $k_1 = k_2$, and that only occurs when $A_{La} = A_{Lc}$. Note however, that an ideal transformer (considered in isolation) does have perfect reciprocity, because it has no leakage inductances and so, in that case; $A_{La} = A_{Lc} = 0$.

So now, having already laboured the point somewhat; we might wonder why it is that some textbooks assume reciprocity to be true and proceed to analyse networks on that basis. Perhaps the answer lies in the symmetry of the mathematics? But were that correct, it would be possible to take a transformer which does obey the principle of reciprocity and place an arbitrary inductance in series with one of its windings. That inductance would then become part of the leakage inductance, and on adding it to that already existing, we might expect to be able to redefine the network in such a way that reciprocity is once again true. We will leave that as a short exercise for those who are particularly interested in this matter; but be assured that it does not work.

It might be argued however, that the reciprocity defect is not important. Many transformers do have comparable leakage paths on both sides; and as we will see when we come to discuss impedance transformation, it is possible to create equivalent networks by transferring impedances from one side to the other (provided that the external reactances are small in comparison to the coupled reactances). Thus a reciprocal transformer model might prove successful in some circumstances; but there is a great deal of difference between making an approximation as a conscious act and applying a theory which is wrong.

Furthermore, there is an important class of transformers which are highly asymmetric. These are the ubiquitous voltage-magnifier transformers; which find use in such diverse applications as spark generators (ignition coils and Tesla coils), vertical antennas, cathode-ray tube EHT supplies, and the narrow-band filter and oscillator coils used in radio equipment. The basic principle is illustrated in the diagram below, where a coupling-coil of a few turns is wound at the earthed end of a much larger coil. The coupling-coil or 'link-winding' is used to inject energy, so that a large voltage can build-up at the top of the main coil by a process of resonant voltage-magnification. Resonance is either due to to the self-resonance of the large coil, which occurs when the length of the winding wire is an electrical half-wavelength, or it can be brought down in frequency by attaching an external capacitance (which, in the case of a Tesla coil, is often provided by a metal ball at the top). Whatever the details however, the secondary network can be represented as a parallel resonator (in the lumped-element approximation at least).



The asymmetry of the coupling arrangement can be understood by considering the fields. When a current (**I**$_1$) flows in the coupling-coil, it can be seen that most of the lines of magnetic flux enclose turns of the main coil, which means that there will be very little leakage inductance on the primary side. When we consider the flux due to a current (**I**$_2$) flowing in the secondary however, it becomes apparent that the coupling-coil is placed at a location where the field-lines are spreading-out. Hence most of the flux will fail to link with the primary, and the secondary leakage inductance will be large.

The point can also be understood by considering the magnetic paths. If the coupling coil is wound directly over the main coil, then the paths for the two coils will have roughly the same area. The coupling coil will have very little inductance, because it has few turns and its value of Nagaoka's coefficient will be somewhat less than 1; but its path is short and so its A_L value will be relatively large. The main coil, on the other hand, will have a large inductance, because it has many turns and its value of Nagaoka's coefficient will be closer to 1; but its path is long and so its A_L will be relatively small. Hence we have:

$L_1 << L_2$ but $A_{L1} >> A_{L2}$

Now if we let:

$A_{L1} = A_{La} + A_{Lb}$ and $A_{L2} = A_{Lb} + A_{Lc}$

where b is the common path, and a and c are the leakage paths, we have:

$A_{La} + A_{Lb} >> A_{Lb} + A_{Lc}$

i.e.;

$A_{La} >> A_{Lc}$

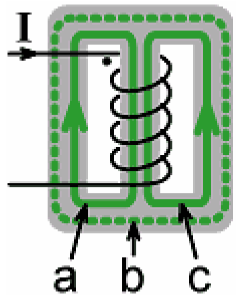
The voltage magnifier is highly asymmetric, but that is an essential feature of the design. If we were to spread the turns of the coupler coil over the whole length of the main coil, then the coupling could be made nearly symmetric, but the Q would be severely degraded by smothering the high-voltage end with lossy insulation and stray capacitive coupling to resistive energy-sinks. In spark coils and transmitter coils, of course, there would also be the problem of insulation breakdown.

Note, incidentally, that the coupling does not have to be via a separate winding. It can also be achieved using an autotransformer connection, i.e., by tapping into the main winding a few turns up from the earthed end. In that case, apart from the lack of DC isolation, the physical behaviour will be similar to that of separate coils; the difference being that there will be no leakage inductance for the coupling winding (because it is physically coincident with the main winding - even the internal inductance will be coupled), and there will be some additional resistive coupling (because the two inductors will share the voltage-drop across the series loss resistance of the common conductor). Thus the autotransformer has the advantage of very tight coupling in one direction, but the possible disadvantage of spurious resistive coupling (which can degrade the out-of-band signal attenuation in narrow-band filter applications).

Notice also that, in the context if intentionally loosely-coupled magnetic systems, the term 'leakage inductance' is not a pejorative. What might be leakage inductance to the transformer analysis, can nevertheless be desired inductance in the overall scheme.

Thus, having established that physically asymmetric transformers are not an oddity to be ignored; it must be observed that there will be serious error if such devices are analysed using the $k = k_1 = k_2$ approach. In order to get the geometric-mean coupling-coefficient (k) right, it will be necessary to shift leakage inductance artificially into the link-winding side. That will give the coupling-coil a fiercely-reactive input impedance, which will not be present in the actual circuit; and the discrepancy can easily be so great as to make the analysis worthless.

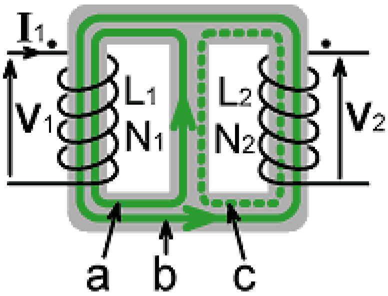
18. Magnetic shunt transformer



Depicted on the right is a coil of wire wound conventionally on a two-hole transformer core. Ignoring the leakage fluxes outside the core material, there are actually three magnetic paths in this system; but since no coil wound on the central pillar can induce a net flux on the path labelled ' b ' , that path can safely be ignored if the core is used according to the manufacturer's instructions. Hence the manufacturer's catalogue gives a nominal (production average) A_L value for the core which is in fact A_{La}+A_{Lc} . Thus (assuming perfect linkage efficiency, i.e., Ñ=N) we calculate the inductance of the coil:

$L = N^2 (A_{La}+A_{Lc})$

There is nothing in the rules of transformer design however, to say that we cannot make use of the outer path; and manufacturers might sell more cores should they care to publish information about it.



In the diagram on the right, a transformer has been constructed by winding coils on the wings of the core, rather than on the central pillar. Now the coupling between the coils is via the undocumented path b, and the fluxes on paths a and c are leakage fluxes. What might not be quite so obvious is that this model is exactly analogous to the three-path system discussed in section **16**. To see why, we need to evoke the electromagnetic principle of continuity, which was introduced in section **3**.

Consider what happens when the two coils are connected in the series-aiding configuration (i.e., when the un-dotted end of L_1 is connected to the dotted end of L_2). That means that current will be flowing into the dotted end of L_2 , and thus, on the left-hand side of that coil, it will be flowing towards the observer. Hence, the resulting flux on path c will be anti-clockwise and, in the central pillar, it will be in opposition to the flux on path a. Transformer cores are usually made to be as nearly symmetric as manufacturing processes will allow. Hence the A_L

values for paths a and c will practically identical. We can therefore make the fluxes on paths a and c equal and opposite in the central pillar by using the same number of turns in both coils; in which case the total magnetic field strength in the central pillar will be zero. We might be tempted to conclude that paths a and c have been rendered inoperative, but that would be incorrect.

The continuity principle states that the fields associated with separate energy-transfer processes (i.e., photons) do not interact, provided that the medium in which wave propagation occurs is linear with respect to electromagnetic stress. Thus (assuming for the moment that the transformer core *is* linear) the fact that there is no *net* flux in the middle pillar does not mean that there is no flux. The two fluxes can perfectly well thread through each other without interaction, and they still exist even when they add up to zero. Hence, in order to understand the electrical behaviour of transformers, we only need to identify the flux paths which are unique, and it does not matter if some of the route is shared by flux from other paths. It follows that paths a, b and c are independent in the regime in which the core material behaves linearly; in which case the analysis of the system shown above can proceed exactly as it did in section **16**.

One gaping hole in the argument so-far, of course, is that no ferromagnetic material is perfectly linear. It turns out however, that non-linearity does not alter the principles of magnetic circuit analysis. The reason is that the divergence of the magnetic field is always zero. No process can cause the existing magnetic lines to split. Thus the topology of the system is unaffected, and non-linearity must be attributed to changes in the overall permeability of the path followed by a particular field line. This implies that $\boldsymbol{\mu}$ for a given region within the material is a non-linear function of the magnetic induction (flux density) B in that region. It also means that $\boldsymbol{\mu}$ is strictly complex, because non-linear processes give rise to losses; i.e., core loss translates into electrical resistance, and this corresponds to a phasor component in the coil impedance which lies at -90° relative to the component due to pure inductive reactance. When large flux-densities are involved (the definition of 'large' being dependent on the material); the upshot is core heating and signal distortion (i.e., the production of harmonics because $\boldsymbol{\mu}$ changes during the course of a wave cycle). None of this however affects the fundamental principles which govern the operation of the device.

So now, we can make sense of the unorthodox transformer depicted above. By winding coils on the wings of the core, we make use of the central pillar as a magnetic shunt. This provides two low-reluctance leakage paths, and thereby drastically reduces the coupling between the two coils. The result is a low-k transformer such as might be used to make double-tuned bandpass filters (IF transformers, etc.). Consequently, the point of interest here is to determine the coupling factor k, but that is now trivial. It is given by equation (**16.3**):

$k = A_{Lb} / \sqrt{[\,(A_{La}+A_{Lb})\,(A_{Lb}+A_{Lc})\,]}$

where, if the core is symmetric, A_{La} and A_{Lc} are the same and can be estimated by dividing the manufacturer's A_L value by 2.

If we presume that the permeability of the material is uniform throughout (a reasonable assumption for the low flux densities involved in signal-processing circuits), then (neglecting leakage paths external to the core) the coupling coefficient is determined by the geometry of the core. Thus manufacturers could publish k-values for two-hole cores used in the shunt configuration. In the absence of that information however, the relevant parameters are easily determined by measurement. It is a matter of winding a few turns of wire around each of the pillars in turn and measuring the inductances. If we call the pillars 'left-hand' (LH), 'middle' (Mid) and 'right-hand' (RH), then we have:

$L_{LH} / N^2 = A_{La} + A_{Lb}$

$L_{Mid} / N^2 = A_{La} + A_{Lc}$

$L_{RH} / N^2 = A_{Lb} + A_{Lc}$

Hence, assuming that the same number of turns was used in each case:

$(L_{LH} + L_{RH} - L_{Mid}) / (2N^2) = A_{Lb}$ . . . . (**18.1**)

To give an idea of the shunting effect which might be obtained in small-signal RF applications, some measurements were made on an Amidon BN-43-202 balun core[12]. Coil impedances were obtained using an admittance bridge operating at a frequency of 10^7 radians/sec (1.591545 MHz), the basic reading accuracy being about ±2.5%.

An admittance bridge gives impedance in parallel form ($Z=R_p//\mathbf{j}X_p$). There is often a tendency to neglect this point; because parallel resistance is very large for high-Q components, and so the series-form reactance is usually practically the same as the parallel-form reactance. This however, is not true for inductors wound on high-permeability RF transformer cores; and so, in this case, a parallel-to-series transformation is required. This involves treating the direct inductance reading from the bridge as apparent inductance (L_p'), converting it to a reactance, then combining it with the parallel resistance to obtain the inductance in series form. Thus we have:

$X_p = 2\pi f\, L_p'$

and

$R_s + \mathbf{j}X_s = R_p // \mathbf{j}X_p = R_p\,\mathbf{j}X_p / (R_p + \mathbf{j}X_p)$

Multiplying numerator and denominator by the complex conjugate of the denominator gives:

$R_s + \mathbf{j}X_s = R_p\,\mathbf{j}X_p (R_p - \mathbf{j}X_p) / (R_p^2 + X_p^2) = (R_p X_p^2 + \mathbf{j}\, R_p^2 X_p) / (R_p^2 + X_p^2)$

Hence, equating imaginaries:

$X_s = R_p^2 X_p / (R_p^2 + X_p^2)$ and $L' = X_s / 2\pi f$

Notice however, that the inductance so extracted is still marked with a prime. This is because there is a contribution to the reactance from the self-capacitance of the coil (C_L), and thus the inductance obtained remains apparent rather than true. It is not possible to determine self-capacitance from a measurement made at a single frequency; but some additional experimentation with the model, using a fairly realistic estimate for C_L for the windings used (about 1pF) indicated that the error due to its neglect is about +1% for the inductance data given below. Furthermore, the object of the exercise is to combine A_L values obtained from three inductance measurements to obtain a coupling coefficient (k); and it can be seen by examining equation (**18.1**) that there is a correlation (because the error is always positive, rather than random) which will tend to cancel the the effect of neglecting C_L.

In the spreadsheet table below, A_L measurements made by winding 3-turn coils on each of the pillars in turn show that the A_{Lb} value for the BN-43-202 core is about 0.3μH/turn², and the coupling factor for a shunt transformer is 0.17 . This is sufficiently low to be useful for making double-tuned bandpass filters; although the lossy nature of the type 43 ferrite might be a counter-indication[13]. To confirm the result, a 3+3 turn shunt transformer was constructed (see illustration) and the coupling was measured using the mutual inductance method described in section **15**. In that case, the correction for parasitics is not so straightforward, but the crude determination given agrees with the A_L method within 2%. The point is thus well established, that the lack of net magnetisation in the central pillar when the coils are connected-series aiding does not diminish the shunting effect.

12 Type 43 ferrite, $\mu_i = 850$, A_L(nom)=2.89μH/turn²; outside dimensions: 13.3 × 7.5 × 14 mm; hole diam: 3.7mm. Datasheet: www.amidoncorp.com/specs/2-34.pdf

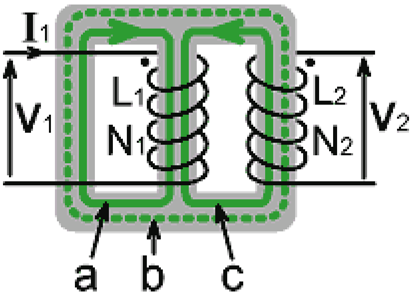
13 A type 61 ferrite version is also available: BN-61-202 ($\mu_i = 125$), requiring more turns for a given inductance, but having lower loss and a greatly reduced temperature coefficient of inductance. The shunt-transformer k is not affected by the material type and will be the same as determined above.

Coupling factor of magnetic shunt transformer
Core is Amidon BN-43-202, AL (nominal)= 2.89 µH/turn^2. f=10^7 radians/sec

	N	Lp /µH	Xp / Ω	Rp / Ω	Rs / Ω	Xs / Ω	Ls /µH		/µH	ALb	k
Central pillar	3	28	280	1050	69.71	261.41	26.14	ALa+ALc	2.90	0.297	0.170
Outer, L1	3	16.8	168	650	40.70	157.48	15.75	ALa+ALb	1.75		
Outer, L2	3	16.9	169	620	42.88	157.31	15.73	ALc+ALb	1.75		
Aiding, L1+L2+2M		39	390	1550	92.29	366.78	36.68				
Opposition, L1+L2-2M		28	280	1060	69.14	261.74	26.17	M	2.63		0.167
									Difference /%		1.624

19. Half turns

It is sometimes stated that when a wire is passed through only one of the holes of a two-hole transformer core, a "half-turn" is obtained. It is also commented, by those who go so far as to make measurements, that the leakage inductances for 'half turns' can be large; and it has been suggested that this is because such turns are somehow "less intimately associated with the core" than full turns.



In the diagram on the right, a two-hole core is depicted having a coil on the central pillar and another on one of the wings. By considering this topology, we can say that the out-of-core leakage inductance for L_2 will be greater than that for L_1 because the sum of the path inductance-factors for L_2 will be less than that for L_1 and so the relative magnetic conductance of the external medium will be somewhat greater. This however, will be a minor effect if the permeability of the core material is large in comparison to that that of air and copper, which is usually the case. Hence L_2 is 'intimate' with the core in any reasonable sense (provided that its turns are wound fairly tightly), and the major reason for its excess leakage inductance is that it can produce a flux on the peripheral magnetic path labelled ' b '. What we can also infer by inspection is that, for a coil wound on the central pillar ("full turns"), alongside or covering L_1, there will be tight coupling between it and L_1 and minimal leakage inductance either-way. For a coil such as L_2 however, assuming negligible induction in the external medium, we can write the relationship between it and L_1 as follows:

Mutual inductance: $M = N_1 N_2 A_{Lc}$

L_1 Coupled inductance: $k_1 L_1 = N_1^2 A_{Lc}$

L_1 Leakage inductance: $(1\text{-}k_1) L_1 = N_1^2 A_{La}$

L_2 Coupled inductance: $k_2 L_2 = N_2^2 A_{Lc}$

L_2 Leakage inductance: $(1\text{-}k_2) L_2 = N_2^2 A_{Lb}$

Voltage shortfall $L_1 \to L_2$: $k_1 = |\mathbf{V}_{2s} / \mathbf{V}_{1p}| = A_{Lc} / (A_{La}+A_{Lc})$

Voltage shortfall $L_2 \to L_1$: $k_2 = |\mathbf{V}_{1s} / \mathbf{V}_{2p}| = A_{Lc} / (A_{Lb}+A_{Lc})$

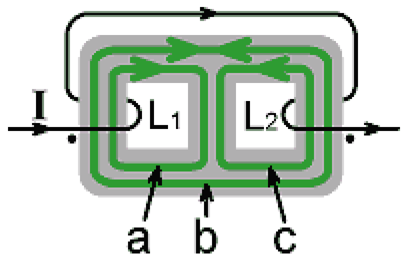
Coupling coefficient: $k = \sqrt{(k_1 k_2)} = A_{Lc} / \sqrt{[(A_{La}+A_{Lc}) (A_{Lb}+A_{Lc})]}$

where the notation: $|\mathbf{V}_{2s} / \mathbf{V}_{1p}|$ means: 'the voltage magnitude ratio when coil 2 is used as secondary and coil 1 is used as primary', etc.. (and it is also the equal to $|\mathbf{V}_{2s}|/|\mathbf{V}_{1p}|$, i.e., the magnitude of a ratio is the same as the ratio of the magnitudes)

Now observe that typical transformer cores are fairly symmetric, in which case $A_{La} = A_{Lc}$, and $k_1 = ½$. This means that, if we take (say) a standard mains transformer and wind some secondary turns around one of the wing pillars, it will require twice as many turns for a given off-load output voltage in comparison to a coil wound on the central pillar. Also this output will droop rather badly on load due to the extra series inductance $N_2^2 A_{Lb}$.

That then is what is meant by "half turns", but note that it is not a rigorous mathematical construct. It can only be applied in the specific context of 'turns ratio in relation to voltage ratio' for transformers with a primary on the central pillar and secondary windings on the wings. Any attempt to generalise the idea by modifying the effective numbers of turns in the coils will fail because the transformers in question are asymmetric and so do not exhibit voltage reciprocity (k_1 is not the same as k_2).

In section **17** it was demonstrated that when a conductor is wound around a magnetic path, the number of turns obtained for analytical purposes (neglecting any linkage inefficiency) is the same as the actual number of turns. That is why it is necessary to be circumspect about (i.e., *avoid*) referring to turns passing through only one hole of a two-hole core as 'half turns'. This poses a conundrum however; which, using the path designations given in the diagram below, can be voiced as follows: If a current loop passes through one hole, it makes a complete turn around (say) paths a and b. If it then passes back through the other hole, it makes a complete turn around paths b and c. Why is it that these two turns in series add up to a single turn around paths a and c, and no turns at all around path b?



It takes a generalisation of the rule for placing phasing dots on transformer windings to solve this problem. It is easy to work out where to put the dots when the core has only one hole, but the solution is far less obvious when separate wires pass through different holes. The rule is this: Designate one end of a wire going into one of the holes as the 'start' of that winding, and place a dot against it. Now, using the corkscrew rule, and assuming a current from a generator flowing towards the dot; determine the direction of rotation of the flux on the path which will be shared with a wire passing through the other hole. In the diagram, the shared path is b , and the flux direction for the turn L_1 is clockwise. Now, considering a wire passing through the other hole (L_2); place a dot against one end of it such that, if a current from a generator were to flow towards that dot, it would produce a flux on the shared path of the same rotation sense as before. In other words; if the two turns were to be connected in series, dotted end to un-dotted end, then the communal fluxes would add.

Having applied the rule, we can now see that when a turn around the middle pillar is considered as a pair of turns around the outer pillars connected in series, then the two inductances are connected in *series opposition*. Hence the total inductance is:

$L = L_1 + L_2 - 2M$

where:

$L_1 = A_{La} + A_{Lb}$, $L_2 = A_{Lb} + A_{Lc}$ and $M = A_{Lb}$

Hence:

$L = A_{La} + A_{Lc}$

A turn around paths a and b, *plus* a turn around paths b and c, *minus* two turns around path b, *is* a turn around paths a and c.

This strange logic is, of course, applied unconsciously when we wind coils on the middle pillar of a two-hole core. It does however raise a point regarding the elimination of a path by flux cancellation. It was observed in the previous discussion of shunt transformers that, when fluxes on separate paths produce cancellation in a shared channel, those two fluxes are nevertheless operational and must be included in the analysis. Here however, we discover two opposing fluxes which can be ignored. The difference is that the cancellation in this case occurs throughout the path, the two fluxes do not give rise to any inductance, and so there is no need to include them in the resulting expressions for inductance.

The foregoing considerations also give us a rule for minimising leakage inductance when constructing tightly-coupled transformers. If a coil is wound on the central pillar of a two-hole core such that, as we look through the holes, the start is (say) on the near side and the finish is on the far side; it is tempting to assume that the coil has a half-integer number of turns. A proper examination however shows that it does not.

Referring again to the diagram above; let us say that we have a coil which passes 10 times through the left-hand hole, but only 9 times through the right-hand hole. This does not constitute a coil of 9½ turns around paths a and c, but rather; a coil of 9 turns around a and c , and one turn around a and b. Alternatively; it can be regarded as 10 turns around path a , 9 turns around c , and one turn around b . This rather complicates the analysis, but there is an even greater downside if the coil is part of a transformer.

If the coil is one of a pair on the central pillar, then the coupling is via paths a and c . Failing to complete the turns of one of the coils allows that coil to produce a net flux on path b, and so gives rise to leakage inductance. This is not particularly important for coils having thousands of turns, such as mains-transformer primaries; but it will degrade the loading characteristics of low-voltage windings, and it will be particularly serious for RF transformers which have few turns.

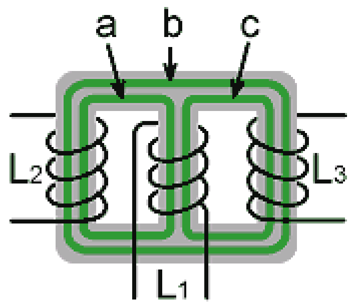
It will be noted, of course, that it is perfectly sensible to construct transformers so that different coils have their terminals on opposite sides of the core. This is good practice, because it separates the electrically-isolated parts of the external circuitry; and for each coil, all of the turns make passes through both holes and so the leakage inductance is minimised.

20: Experimental verification

What has been put forward in the preceding sections is a method for sketching-out the unique magnetic paths associated with a system of coupled inductors and using that information to obtain the complete set of coupling parameters. That does not imply that the topological approach can tell us everything that there is to know about transformers; but it is fair to say that refinements to account for losses, non-linearity, winding capacitance, etc., can later be included without challenge to the basic theory. What is also intended is that the analysis (not to be confused with the problem of estimating the parameters) should be carried out by inspection; i.e., it is a simple procedure, comparable to applying Ohm's law, and should not require constant reference to a list of formulae.

In order to verify the ideas outlined so far and gain familiarity with them; those interested might be amused to carry-out something on the lines of the following practical demonstration. It requires a means for measuring impedance (such as a universal AC bridge), a signal generator, an oscilloscope (or a high-impedance AC voltmeter), a two-hole transformer core and some wire. Here we work at 10^7 radians/sec (1.5915MHz) and use a ferrite core; thereby taking advantage of the fact that radio transformers have few turns and so are easy to construct. The experiment however, can just as well be redesigned to work at audio or power-line frequencies, using (say) a small laminated iron-alloy core.

The procedure is as follows: Wind three coils on the core as shown in the diagram, and take careful note of the number of turns in each coil. For the wing coils L_2 and L_3 , the turns are counted as single passes through the associated hole. For the middle coil L_1 , turns are counted as passes through either of the holes, with the proviso that the number of passes is the same for both holes; in which case, as looking at the diagram, the wires will both come out either on the near side or on the far side. This latter precaution ensures that, by measuring the three inductances individually and solving three simple simultaneous



equations, it is possible extract the A_L values for the magnetic paths a , b , and c (this has already been demonstrated in section **18**).

From the path A_L values and the turns numbers (neglecting linkage inefficiency and out-of-core leakage inductance), all of the device's small-signal inductive parameters can be calculated. These include the mutual inductance and coupling coefficient for any chosen pair of coils, and the voltage-shortfall factors for when a pair of coils is used as a transformer.

A sensible notation is to refer to the three mutual inductances as M_a , M_b and M_c . There is only one mutual inductance per pair of coils, and so it is sufficient to to use a subscript which refers to the shared path; i.e., M_a is the mutual inductance for coils 1 and 2, etc.. The calculated mutual inductances can be verified by measuring the inductance of the corresponding pair of coils in series aiding and series opposition and dividing the difference by 4 (see section **15**). This is a fair test of theories which derive transformer properties from the partitioning of flux between different magnetic paths, but it is not completely convincing because the method of measurement is deduced from a precursor to the overall theory.

The calculated voltage-shortfall factors can be verified by measuring the the voltage-magnitude ratios. This is easily done using an oscilloscope with a high-impedance (×10) probe. One of the windings is elected primary and connected to the output of a sine-wave signal generator. It may also be helpful to terminate the generator in its preferred load resistance (in parallel with the primary) in some instances, since some oscillators do not take kindly to reactive loads. The 'scope probe is connected across the primary, and the Y-gain and shift controls are adjusted until the waveform just touches the top and bottom rulings of the graticule. The probe and its ground clip are then moved to each of the secondaries in turn, and the new height on the graticule is recorded. Only the ratio of the two voltages is needed (and the ratio of peak-to-peak voltages is the same as the ratio of RMS voltages). The input voltage is not important provided that it is sufficient to give a good clean 'scope trace and not so great as to provoke non-linear behaviour (say, between 200mV and 2V RMS). This method is, incidentally, not particularly precise; but it is a very rigorous test of theory because it is free from assumptions in relation to the behaviour of the device.

From the three initial inductance measurements we obtain:

$A_{La} + A_{Lc} = L_1 / N_1^2$, $A_{La} + A_{Lb} = L_2 / N_2^2$, $A_{Lb} + A_{Lc} = L_3 / N_3^2$

Thus:

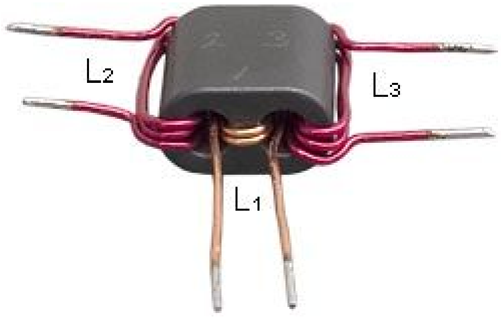
$A_{Lb} = [\, (L_2 / N_2^2) + (L_3 / N_3^2) - (L_1 / N_1^2) \,] / 2$

$A_{La} = (L_2 / N_2^2) - A_{Lb}$ and $A_{Lc} = (L_3 / N_3^2) - A_{Lb}$

For a symmetric core, A_{La} and A_{Lb} will be the same within the experimental uncertainty. $A_{La}+A_{Lc}$ should also be reasonably close to the manufacturer's published A_L value for the core (if known). The determined A_L values allow calculation of the complete set of coupling parameters for the three possible pairs of coils. The relationships are deduced by inspection of the diagram above and are listed in the table below. The theory which leads to these expressions is given in section **16**.

Transformer A (L_1:L_2)	**Transformer B** (L_2:L_3)	**Transformer C** (L_1:L_3)
$M_a = N_1 N_2 A_{La}$	$M_b = N_2 N_3 A_{Lb}$	$M_c = N_1 N_3 A_{Lc}$
$k_{1a} = \lvert \mathbf{V}_{2s} / \mathbf{V}_{1p} \rvert = A_{La} / (A_{La}+A_{Lc})$	$k_{2b} = \lvert \mathbf{V}_{3s} / \mathbf{V}_{2p} \rvert = A_{Lb} / (A_{La}+A_{Lb})$	$k_{1c} = \lvert \mathbf{V}_{3s} / \mathbf{V}_{1p} \rvert = A_{Lc} / (A_{La}+A_{Lc})$
$k_{2a} = \lvert \mathbf{V}_{1s} / \mathbf{V}_{2p} \rvert = A_{La} / (A_{La}+A_{Lb})$	$k_{3b} = \lvert \mathbf{V}_{2s} / \mathbf{V}_{3p} \rvert = A_{Lb} / (A_{Lb}+A_{Lc})$	$k_{3c} = \lvert \mathbf{V}_{1s} / \mathbf{V}_{3p} \rvert = A_{Lc} / (A_{Lb}+A_{Lc})$
$k_a = \sqrt{(k_{1a}\, k_{2a})}$	$k_b = \sqrt{(k_{2b}\, k_{3b})}$	$k_c = \sqrt{(k_{1c}\, k_{3c})}$

$\lvert \mathbf{V}_{2s} / \mathbf{V}_{1p} \rvert$ means 'the voltage magnitude ratio when coil 1 is used as primary and coil 2 is used as secondary', etc.. The voltage-shortfall factor (k_{1a} , etc.) is expected to agree with this ratio in the limit that linkage-inefficiency and out-of-core leakage-inductance are negligible, and on the assumption that the secondary is off load (hence the use of a high-impedance voltmeter).



The spreadsheet analysis given below relates to measurements made on the transformer shown on the right. This is wound on an Amidon BN-43-202 core, the details of which have already been given in section **18**. Impedances were measured in parallel form (±2.5%) and converted into inductances using the parallel-to-series transformation (also described in section **18**). In the spreadsheet; 'obs' means 'observed by measurement not related to the theory under test'; and 'calc' means 'calculated from the path A_L values'.

The uncertainty in the voltage ratio measurements is determined as follows. Relative voltages are recorded in units of centimetres on the oscilloscope graticule. It was estimated that the height setting and reading operations were reproducible to within 0.2cm. Hence the two voltage readings used to determine a ratio each have an uncertainty of ±0.2cm; and those errors were assumed to be random and uncorrelated. Now, if we have:

$y = V_s / V_p$

and the uncertainties in the two voltages are δ_s and δ_p; then, for uncorrelated errors, the overall standard deviation is given by adding the squares of the uncertainty contributions and taking the square root; i.e.:

$\sigma_y = \sqrt{[(\delta_p\, \partial y/\partial V_p)^2 + (\delta_s\, \partial y/\partial V_s)^2]}$

where

$\partial y/\partial V_p = -V_s / V_p^2$ and $\partial y/\partial V_s = 1/V_p$

hence:

$\sigma_y = \sqrt{[(\delta_p V_s / V_p^2)^2 + (\delta_s / V_p)^2]}$

In the spreadsheet, this formula gives uncertainties of about ±0.03 for the ratio measurements. Note that this implies a large proportionate (i.e., percentage) uncertainty when the ratio is small, which means that the experiment is most convincing when the three coils have similar numbers of turns.

Three winding transformer

Amidon BN-43-202, AL (nominal)= 2.89 µH/turn^2. f=10^7 radians/sec

Inductances	N	Lp /µH	Xp / Ω	Rp / Ω	Rs / Ω	Xs / Ω	Ls /µH		/µH		/µH
Middle L1	3	28.1	281.0	1060	69.60	262.5	26.25	ALa+ALc	2.917	ALa	1.456
Outer, L2	3	16.9	169.0	620	42.88	157.3	15.73	ALa+ALb	1.748	ALb	0.292
Outer, L3	3	17	170.0	610	43.96	157.7	15.77	ALc+ALb	1.753	ALc	1.461

Series pairs	Lp /µH	Xp / Ω	Rp / Ω	Rs / Ω	Xs / Ω	Ls /µH		obs	calc	err/%
L1+L2+2Ma	72.5	725.0	3000	165.54	685.0	68.50	Ma /µH	13.123	13.106	0.133
L1+L2-2Ma	17.2	172.0	630	43.70	160.1	16.01	ka	0.646	0.645	0.133
L1+L3+2Mc	73	730.0	3000	167.70	689.2	68.92	Mc /µH	13.207	13.149	0.440
L1+L3-2Mc	17.4	174.0	610	45.90	160.9	16.09	kc	0.649	0.646	0.440
L2+L3+2Mb	39	390.0	1540	92.81	366.5	36.65	Mb /µH	2.626	2.626	0.006
L2+L3-2Mb	28.5	285.0	950	78.44	261.5	26.15	kb	0.167	0.167	0.006

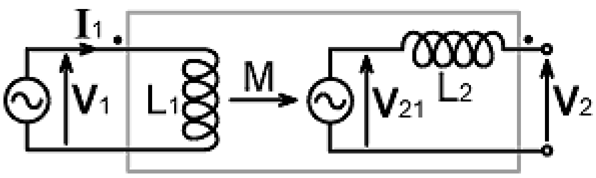
Voltage ratios	Vp/cm	Vs/cm	±	obs	±	calc	err/%
V2s/V1p = k1a	8.0	4.0	0.2	0.50	0.03	0.499	0.17
V1s/V2p = k2a	8.0	6.7	0.2	0.84	0.03	0.833	0.53
V3s/V1p = k1c	8.0	4.0	0.2	0.50	0.03	0.501	-0.17
V1s/V3p = k3c	8.0	6.8	0.2	0.85	0.03	0.834	1.97
V3s/V2p = k2b	8.0	1.3	0.2	0.16	0.03	0.167	-2.64
V2s/V3p = k3b	8.0	1.3	0.2	0.16	0.03	0.166	-2.37

The results agree, within experimental error, with the theory outlined in the preceding sections. They are however, strongly in disagreement with the assumption that voltage-shortfall reciprocity is an inherent property of transformers.

21. Transimpedance

In the discussion up to this point; the object of the exercise has been to infer the electrical properties of systems of coupled inductors from consideration of the associated magnetic fields. That constitutes the mapping from physical fundamentals to circuit theory; but it is now necessary to convert that information into an electrical model. The issue is that a transformer is not a basic circuit element; but rather, it is a 'black box' with ports. The final step in the translation is to represent what lies inside the box as a network of resistances, inductances, capacitances and generators, and to define those elements according to the various physical parameters. In other words; we need to determine the transfer functions which describe the relationships between the ports, and we need to express that information using phasors.

When a transformer is wired conventionally, a port is defined as a pair of terminals leading to one of the coils inside the device. The fundamental circuit element which lies behind each port is therefore, in that particular case (neglecting losses and self capacitance), an inductance. Thus we can start with the concept of a null-transformer, which is simply two inductors in the same box, with no coupling whatsoever between them. This is not a particularly useful construct however; and so we will evoke an arcane process ' M '; which cannot be explained using Kirchhoff's laws, but which is such that, if a generator is connected to one of the ports, then the observable behaviour gives the impression that generators have come into existence behind the other ports. What precious little we can deduce from that, concerning what lies within the box, is shown in the diagram below.



Notice that the generator which has appeared in the secondary network is connected in series with the inductance. This might seem peculiar at first, since we have previously noted that induction causes a voltage to appear across the ends of an inductance; but the arrangement is dictated by the fact that the basic generator-element used in circuit-analysis is a short-circuit when its output is zero. It is obvious that when the external generator is disconnected from the primary port, then the virtual generator in the secondary network gives no output. Thus, if the generator of the induced-voltage were to be connected *across* L_2, then the output-port would be a short-circuit when the primary is disconnected. We know, of course, that when the primary is open-circuit, what we will see looking into the secondary port is the inductance L_2, and so the arrangement shown is the correct one. It is also consistent with the physics of the transformer because, as we will see when we come to discuss the origin of the winding capacitance; magnetic disturbances cause electromagnetic waves to propagate along the wire of a coil, which means that the coil itself is in series with whatever is causing the disturbance. Furthermore, when the secondary is open-circuit, there is no voltage-drop across L_2, and so the arrangement creates the illusion that the voltage is induced directly across the port if (and only if) there is no secondary current.

The next step in building the model is to find a phasor expression for the induced voltage $\mathbf{V}_{21}$. This requires Faraday's Law, and so threatens to embroil us in the business of mapping time-differential equations onto complex numbers (as discussed at the end of section **8**). It transpires however, that we have already done the necessary groundwork in section **11**, where it was determined that the ratio of the secondary induced voltage to the primary voltage is given by the

expression (cf. equation 11.1):
$V_{21} / V_1 = M / L_1$
The two voltages given here are *instantaneous* voltages rather than phasors, but the fixed relationship implies that they are in phase. It is also the case that when two phasors are in phase, their ratio can be treated as a scalar[14]. This means that we can convert the expression above directly into phasor form by applying the converse principle; i.e., a ratio of two instantaneous voltages or magnitudes can be expressed as a ratio of two phasors of equal phase. Hence:
$\mathbf{V}_{21} / \mathbf{V}_1 = M / L_1$
Thus we dispense with differential equations; but we do not yet have $\mathbf{V}_{21}$ in the form required for general transformer analysis. The issue here is that the transformer exploits the phenomenon of electromagnetic induction in order to convert a magnetic field into an electric field. In other words, it converts MMF into EMF, and so it is fundamentally a module which gives a voltage output for a current input. Hence its principal transfer-function is a voltage divided by a current. Such a ratio has units of Ohms (impedance), and so can be described as a 'transfer-impedance' or *transimpedance*. Thus the transformer is by nature a transimpedance device, and is therefore best analysed by making use of that property. The phasor substitution which can be used to eliminate the input voltage is (by inspection of the diagram above, using Ohm's Law):
$\mathbf{V}_1 = \mathbf{j}X_{\mathbf{L1}} \mathbf{I}_1$
Thus:
$\mathbf{V}_{21} / \mathbf{I}_1 = \mathbf{j}X_{\mathbf{L1}} M / L_1$
but $X_{\mathbf{L1}} = 2\pi f L_1$, hence:

$\mathbf{V}_{21} / \mathbf{I}_1 = \mathbf{j}2\pi f M = \mathbf{j}X_{\mathbf{M}}$	[Ohms]	Intrinsic transimpedance

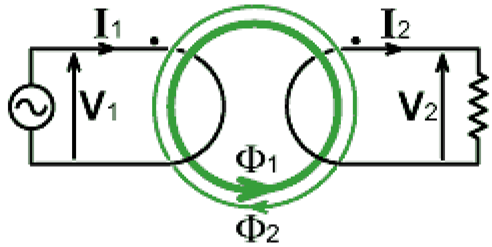
where $X_{\mathbf{M}}$ might reasonably be called the 'mutual reactance'. Notice that $\mathbf{j}X_M$ is referred to here as the *intrinsic* transimpedance. This is because transformers, being natural current-to-voltage converters (with DC isolation) are excellent for making AC and RF ammeters. An ammeter circuit has an overall transimpedance, which is the scale-factor (Volts per Amp) for the voltmeter which displays the result. This is not the same as the intrinsic transimpedance, which governs the fundamental behaviour of the transformer, and so it is necessary to make the distinction.

The basic transimpedance rule given above, incidentally, works for any combination of subscripts, including those which imply self-inductance; i.e., the mutual inductance between a coil and itself is the inductance of the coil. Thus if $\mathbf{V}_{11}$ is the voltage induced in coil 1 due to a current in coil 1, then:
$\mathbf{V}_{11} / \mathbf{I}_1 = \mathbf{j}\, 2\pi f\, L_1$
The voltage induced in port network 2 due to a current in port network 1 is, of course:
$\mathbf{V}_{21} = \mathbf{j}\, X_{\mathbf{M}}\, \mathbf{I}_1$
The sign convention for the relationship is that current is positive when if flows from outside the transformer towards a dot, and voltage is positive when its circuit-analysis arrow points towards a dot. Such induced voltages from any number of currents in any number of ports can simply be added together.

14 **AC Theory**. D W Knight [available from www.g3ynh.info]. Theorem 24.8.

22. Inductive energy transfer

The transimpedance rule can be used to determine the behaviour of any transformer in the lumped-element limit (assuming that losses and parasitic capacitance are considered separately). There is however, a certain subtlety in its application, due to the inherently bi-directional nature of the transformer's ports. This issue is best introduced by considering the way in which energy-transfer occurs in a conventional transformer-coupled system; i.e., in a circuit which has only one energy source, so that one of the coils can be designated primary, and the rest are secondaries.



The diagram on the right represents a pair of inductively-coupled loops with a resistance in the secondary circuit. The primary current induces a voltage $\mathbf{j}X_M\mathbf{I}_1$ in the secondary network, and so closing the circuit causes a current $\mathbf{I}_2$ to flow. Now note that the dot convention sets the polarity of the voltage $\mathbf{V}_2$, as indicated by an arrow. Thus, since the resistance is an energy sink; if the current $\mathbf{I}_2$ is to be considered positive, it must be shown flowing away from the dot. Every current is associated with a magnetic field, and so the secondary current is related to a flux which tends to cancel some of the flux from the primary (i.e., a current flowing away from a dot has the opposite effect of a current flowing towards a dot). Thus we see that the secondary current, by bleeding away some of the energy stored in the field, has the effect of reducing the back-voltage produced by the primary inductance. We can moreover, work out exactly the extent to which the primary back-voltage is reduced by using the transimpedance rule, which enables us to calculate the voltage induced in the primary as a result of the current in the secondary (we will do that in the next section). Also observe, that if the generator connected to the primary is a constant-voltage source; then drawing a current from the secondary will cause an increase in the current flowing into the primary. This is, of course, exactly what happens in practice.

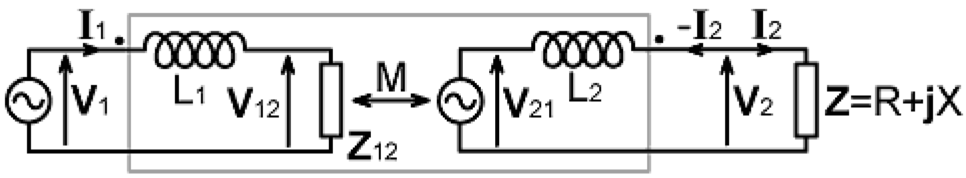
The above explanation however raises an issue of causality. It is tempting to think that the secondary current produces a flux which signals back to the primary to request more current from the generator. Physically, this makes no sense, because it implies that the energy arriving at the sink communicates with the generator in order to request its own existence. This paradox arises because circuit-analysis, albeit mathematically self-consistent, is nevertheless based on pre-Maxwellian notions of electricity. It is the flux which produces the current, and not the other way around. Hence, whatever pragmatic notions we might use in the process of building a circuit model; the flux Φ_2 represents an electromagnetic wave propagating from the source to the sink.

23. Conventional transformer

A curious aspect of coupled-inductor theory is that transformers are described as belonging to categories, such as 'conventional', 'hybrid' or 'transmission line'; and yet these different types are fundamentally the same (at least, when considered in the lumped-element limit). The distinction arises, not in the principle of operation, but in the various ways in which transformers can be used. The reason for that is that an electrical device having multiple bi-directional ports offers a vast range of possibilities; and apart from stating the transimpedance relation, there is little which can be done by way of analysis unless some external circuit conditions are imposed. Subject to that requirement, we will find that most of the ground rules can be established by investigating the conventional transformer (provided that we do not limit our horizons by doing so).

Conventional transformers are, of course, important by virtue of their ubiquity. They are transformers which have DC continuity between the terminals of each port, and which have only one port connected to a generator. Note, incidentally, that autotransformers are included in this category, the matter of electrical isolation making no difference to the underlying magnetic theory.

In a conventional transformer circuit, any net energy flow which occurs is unidirectional. This (if you'll excuse the inversion of causality) allows the voltage which is induced in the primary side, as a result of a secondary current, to be represented as the voltage drop across a virtual impedance ($\mathbf{Z}_{12}$ in the diagram below). Thus, in a basic two-coil transformer, we have two separate electrical networks; with energy from a generator flowing into any resistive component which might be present in $\mathbf{Z}_{12}$, and exactly the same amount of energy emerging from a virtual generator on the secondary side. An implication of that last statement, incidentally, is that the resistive component of $\mathbf{Z}_{12}$ does not include primary-side losses. If loss processes are to be added to the model, it is best to distinguish them from the magnetic coupling by placing separate resistances in series or parallel with the ports.



Now notice that there are two secondary currents marked on the diagram above ($\mathbf{I}_2$ and $-\mathbf{I}_2$). For the purpose of analysing what goes on in the load network $\mathbf{Z}$, we need to know $\mathbf{V}_2$ and $\mathbf{I}_2$. For the purpose of working out induced voltages however, we need to define currents as flowing towards the dots on the ports, and so the current which determines the primary induced voltage is $-\mathbf{I}_2$.
Hence:
$\mathbf{V}_{12} = \mathbf{j}X_\mathbf{M} (-\mathbf{I}_2)$
but we can also define this as the voltage drop across the 'induced' impedance $\mathbf{Z}_{12}$; i.e.:
$\mathbf{V}_{12} = \mathbf{I}_1 \mathbf{Z}_{12}$
Hence:
$\mathbf{Z}_{12} = -\mathbf{j}X_\mathbf{M} \mathbf{I}_2 / \mathbf{I}_1$ . . . . (**23.1**)
This statement must, of needs, appear in any physics-based discussion of transformer modelling, and is sometimes accompanied by 'deep insights' into subject, often with bad effect. Thus some readers may have encountered the assertion that, since $\mathbf{I}_2/\mathbf{I}_1$ is dimensionless, and the phase is $-\mathbf{j}$, then $\mathbf{Z}_{12}$ represents a pure capacitive reactance. Firstly note that $\mathbf{I}_2$ and $\mathbf{I}_1$ are phasors and so the overall phase is only exactly $-\mathbf{j}$ when the two currents are perfectly in phase. Secondly, those two currents can only ever be brought exactly into phase by eliminating all resistances from the secondary network (or by connecting an additional generator to cancel them out). Thirdly, even when the currents are in phase, the frequency law of $\mathbf{Z}_{12}$ is that of an inductance, complete with the property that its reactance goes to zero when $f \to 0$. Hence the $-\mathbf{j}X_\mathbf{M}$ factor in $\mathbf{Z}_{12}$ represents a negative inductance, -M , which we have previously interpreted physically as inductance cancellation (*not* capacitance). What happens in reality is that that the generator pumps energy into the magnetic field surrounding L_1, but the secondary network taps into that energy store, via M, thereby modifying the impedance of the primary network. The actual phase of $\mathbf{Z}_{12}$ depends strongly on the secondary leakage inductance in combination with the reactance of the secondary load impedance; but if the leakage inductance is small, and $\mathbf{Z}$ is purely resistive, then $\mathbf{Z}_{12}$ cancels most (but not all) of the reactance of L_1 and replaces it with resistance.

The impedance $\mathbf{Z}_{12}$ is, of course, not fully defined as a network until we eliminate the current ratio. That can be done by substituting for $\mathbf{I}_2$. Inspection of the secondary network, as represented above, gives:
$\mathbf{I}_2 = \mathbf{V}_{21} / (\mathbf{j}X_{L2} + \mathbf{Z})$
but from the transimpedance rule:
$\mathbf{V}_{21} = \mathbf{j}X_\mathbf{M} \mathbf{I}_1$
Hence:
$\mathbf{I}_2 = \mathbf{j}X_\mathbf{M} \mathbf{I}_1 / (\mathbf{j}X_{L2} + \mathbf{Z})$ . . . . (**23.2**)
substituting this into (**23.1**) gives:

$\mathbf{Z}_{12} = -jX_M\, jX_M\, \mathbf{I}_1 \,/\, [\, (jX_{L2} + \mathbf{Z})\, \mathbf{I}_1\,]$
i.e.;

$\mathbf{Z}_{12} = X_M^2 \,/\, (jX_{L2} + \mathbf{Z})$	**23.3**

This completes the definition of the equivalent network for the lossless, lumped-element transformer. It also demonstrates a special property of transformer theory, which is that the equivalent network depends on the external circuit component **Z**. This does not have particularly profound implications in the case of the conventional transformer, but note that the expression above remains valid when the resistive component of **Z** is *negative*. This gives us the basis for dealing with transformers which have generators connected to more than one port, i.e., **Z** can be allowed to represent an impedance and a generator connected in series. If **Z** is such an impedance+generator hybrid, then so is $\mathbf{Z}_{12}$, and so is the 'impedance' looking into port 1. This is the probable root of the term 'hybrid transformer'; the point being that when there are multiple external energy sources, it becomes necessary to give the virtual 'impedances' in the equivalent network the freedom to occupy the left-hand side of the Z-plane.

24. Loaded voltage ratio

The general expression for the voltage ratio of a transformer (neglecting the effect of losses) can now be obtained. By inspection of the diagram of the preceding section we have:
$\mathbf{V}_2 = \mathbf{I}_2\, \mathbf{Z}$
Which, by substituting for $\mathbf{I}_2$ using (**23.2**), gives:
$\mathbf{V}_2 = jX_M\, \mathbf{I}_1\, \mathbf{Z} \,/\, (\, jX_{L2} + \mathbf{Z})$
Also, by inspection:
$\mathbf{V}_1 = \mathbf{I}_1\, (\, jX_{L1} + \mathbf{Z}_{12}\,)$
Which, by substituting for $\mathbf{Z}_{12}$ using (**23.3**), gives:
$\mathbf{V}_1 = \mathbf{I}_1\, [\, jX_{L1} + X_M^2 \,/\, (jX_{L2} + \mathbf{Z})\,]$
Hence:

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{jX_M\,\mathbf{Z}}{(\, jX_{L2} + \mathbf{Z})\,[\, jX_{L1} + X_M^2 / (jX_{L2} + \mathbf{Z})\,]} = \frac{jX_M\,\mathbf{Z}}{jX_{L1}\,(\, jX_{L2} + \mathbf{Z}) + X_M^2}$$

i.e.:

$\mathbf{V}_2 / \mathbf{V}_1 = jX_M\, \mathbf{Z} \,/\, [\, jX_{L1}\, \mathbf{Z} + X_M^2 - X_{L1}\, X_{L2}\,]$	**24.1**

The first thing to notice about this phasor ratio is that, when $|\mathbf{Z}|\to\infty$, the denominator is overwhelmed by the term $jX_{L1}\mathbf{Z}$ and the other terms can be neglected. Thus $\mathbf{V}_2/\mathbf{V}_1 \rightarrow X_M/X_{L1}$; i.e.:
When $|\mathbf{Z}|\to\infty$, $\mathbf{V}_2 / \mathbf{V}_1 \rightarrow M / L_1$
We have seen this special limiting case before as equation (**11.1**). Neglecting the effect of self-capacitance, the two voltages fall exactly into phase when the transformer is unloaded.

Now recall that, for an ideal transformer, $M=\sqrt{(L_1\, L_2)}$. In that case;
$X_M^2 = X_{L1}\, X_{L2}$
Substituting this into (**24.1**) gives:
$\mathbf{V}_2 / \mathbf{V}_1 = M / L_1$
but, for an ideal transformer, both coils have the same inductance factor. Thus: $M=\tilde{N}_1\tilde{N}_2\, A_L$ and $L_1=\tilde{N}_1^2 A_L$. Hence:
$\mathbf{V}_2 / \mathbf{V}_1 = \tilde{N}_2 / \tilde{N}_1$
For an ideal transformer of the analytical variety, $\mathbf{V}_2$ is always in phase with $\mathbf{V}_1$, and the voltage ratio is given by the effective turns ratio.

The general voltage ratio can also be expressed using the coupling coefficient k instead of M. This alternative form can be obtained by using the substitution $M = k\sqrt{(L_1 L_2)}$ in equation (**24.1**). Notice particularly the pair of terms $X_M^2 - X_{L1}X_{L2}$ in the denominator, where:

$X_M^2 = k^2 X_{L1} X_{L2}$

and so:

$X_M^2 - X_{L1} X_{L2} = (k^2-1) X_{L1} X_{L2}$

Hence:

$\mathbf{V}_2 / \mathbf{V}_1 = \mathbf{j}X_M \mathbf{Z} / [\ \mathbf{j}X_{L1} \mathbf{Z} + (k^2-1) X_{L1} X_{L2}\ ]$

Forcibly factorising $\mathbf{j}X_{L1}$ from the denominator gives:

$\mathbf{V}_2 / \mathbf{V}_1 = (M / L_1)\ \mathbf{Z} / [\ \mathbf{Z} + (k^2-1) X_{L2} / \mathbf{j}\ ]$

i.e.:

$\mathbf{V}_2 / \mathbf{V}_1 = (M / L_1)\ \mathbf{Z} / [\ \mathbf{Z} - \mathbf{j}\ (k^2-1) X_{L2}\ ]$

Now recall from section **13** that:

$M / L_1 = k \sqrt{(L_2 / L_1)}$

Thus, also noticing that $(k^2-1) = -(1-k^2)$:

$\mathbf{V}_2 / \mathbf{V}_1 = k\ [\sqrt{(L_2 / L_1)}]\ \mathbf{Z} / [\ \mathbf{Z} + \mathbf{j}(1-k^2) X_{L2}\ ]$	**24.2**

This is a surprisingly compact expression, which obviously reverts to the ideal case when $k \to 1$.

25. Magnetising current

For a closely-coupled transformer, we expect the current ratio to be approximately the inverse of the voltage ratio. The general expression for the current ratio is given by rearrangement of (**23.2**):

$\mathbf{I}_2 / \mathbf{I}_1 = \mathbf{j}X_M / (\mathbf{j}X_{L2} + \mathbf{Z})$

Inversion gives:

$\mathbf{I}_1 / \mathbf{I}_2 = (L_2 / M) - \mathbf{jZ} / X_M$	**25.1**

Now recall that $M = k\sqrt{(L_1 L_2)}$, and so:

$\mathbf{I}_1 / \mathbf{I}_2 = [\ (1/k)\sqrt{(L_2 / L_1)}\ ] - \mathbf{jZ} / X_M$

For an ideal transformer, $k \to 1$ and because there are no leakage paths; $\sqrt{(L_2 / L_1)} \to \tilde{N}_2 / \tilde{N}_1$.

Hence:

When $k \to 1$, $\mathbf{I}_1 / \mathbf{I}_2 \to [\ \tilde{N}_2 / \tilde{N}_1\ ] - \mathbf{jZ} / X_M$

This limit shows that, even when coupling between the windings is complete, the inverse current-ratio is never exactly equal to the turns ratio. Equality can only be achieved when $|\mathbf{Z}| \to 0$ (the secondary is short circuit and there is no series loss resistance) or when $X_M \to \infty$; and neither of those extremes is physically realisable. The reason for the discrepancy is that there is always a primary current even when the secondary is disconnected. That standing current is known as the ***magnetising current***, and is given by the expression:

$\mathbf{I}_{1o} = \mathbf{V}_1 / (\mathbf{j}\ X_{L1})$

Which is, of course, simply the current through L_1 when $|\mathbf{Z}_{12}| \to 0$. Proof is as follows:

Rearrangement of equation (**25.1**) gives:

$\mathbf{I}_1 = \mathbf{I}_2 (L_2 / M) - \mathbf{j}\ \mathbf{I}_2 \mathbf{Z} / X_M$

but $\mathbf{I}_2 \mathbf{Z} = \mathbf{V}_2$. Therefore:

$\mathbf{I}_1 = \mathbf{I}_2 (L_2 / M) - \mathbf{j}\mathbf{V}_2 / X_M$

A substitution for $\mathbf{V}_2$ is given by equation (**24.1**):

$\mathbf{V}_2 = \mathbf{j}\mathbf{V}_1 X_M \mathbf{Z} / [\ \mathbf{j}X_{L1} \mathbf{Z} + X_M^2 - X_{L1} X_{L2}\ ]$

Hence:

$\mathbf{I}_1 = \mathbf{I}_2 (L_2 / M) - \mathbf{j}(1/X_M)\ \mathbf{j}\mathbf{V}_1 X_M \mathbf{Z} / [\ \mathbf{j}X_{L1} \mathbf{Z} + X_M^2 - X_{L1} X_{L2}\ ]$

i.e.;

$\mathbf{I}_1 = \mathbf{I}_2 (L_2 / M) + \mathbf{V}_1 \mathbf{Z} / [\ \mathbf{j}X_{L1} \mathbf{Z} + X_M^2 - X_{L1} X_{L2}\ ]$	**25.2**

Now, to find the magnetising current, let $|\mathbf{Z}|\rightarrow\infty$. In that case, $\mathbf{I_2}\rightarrow 0$ and the denominator of the second term is overtaken by the quantity $\mathbf{j}X_{L1}\mathbf{Z}$. Thus:
When $|\mathbf{Z}|\rightarrow\infty$, $\mathbf{I_1}\rightarrow \mathbf{V_1} / (\mathbf{j}\, X_{L1})$
This shows, more than anything else, that the model is mathematically self-consistent.

Now recall that, for an ideal transformer, $X_M{}^2 = X_{L1}\, X_{L2}$. Hence, in that case, equation (**25.2**) reduces to:

$\mathbf{I_1} = \mathbf{I_2}\, (\tilde{N}_2 / \tilde{N}_1) + \mathbf{V_1} / (\, \mathbf{j}X_{L1}\,)$	**25.3**

The primary current of an ideal transformer is given by the secondary current multiplied by the turns ratio, *plus* the magnetising current.

The current ratio for a closely-coupled transformer will, of course, be approximately same as the inverse turns-ratio in the limit that the magnitude of the working primary current is large relative to the magnetising current. Taking this correspondence to be true is equivalent to the assuming that the primary inductance (and hence the mutual inductance) is infinite. That follows from the limiting relationship given above, since the magnetising current can only go to zero when the input voltage is zero or when $X_{L1}\rightarrow\infty$.

An ideal transformer having infinite inductance (and therefore no magnetising current) is known as a ***perfect transformer***. It has the property that:

$\mathbf{V_2} / \mathbf{V_1} = \mathbf{I_1} / \mathbf{I_2} = \tilde{N}_2 / \tilde{N}_1$	Perfect transformer analytical relation

This (with Ñ=N) is often used, in uncritical situations, as an approximation for actual transformer behaviour; but bear in mind that such an approach has limited validity in RF applications. The reason for the caveat is that, for RF circuit designs which depend on lumped-element behaviour, it is necessary to ensure that the length of the wire used to make an inductor is substantially less than one half-wavelength at the highest frequency of operation. Hence excessive inductance can be problematic, which means that the finite nature of transformer inductance must usually be taken into account (even if only to ensure that it is safe to neglect it). Including the inductance however, is not difficult. Equation (**25.3**) implies that a perfect transformer can be converted into an ideal transformer by connecting an inductance L_1 in parallel with port 1. In later discussion moreover, we will see that exactly the same effect can be had by connecting an inductance L_2 across port 2. In general; a perfect transformer becomes an ideal transformer when an inductance is connected across any one of its ports.

Note incidentally, that some older texts use the term "ideal transformer" to refer to the perfect transformer. Modern usage however seems to favour the convention that ideal transformers have finite inductance.

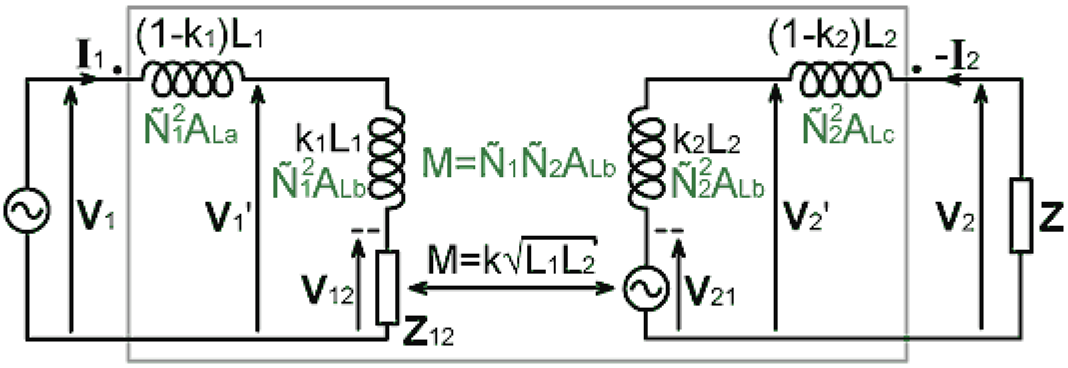
Also note that, in this article, we have allowed that ideal transformers do not have to have perfect flux-linkage; i.e., Ñ can be less than N. It was observed in section **11**, that for a physical realisation of an ideal transformer (insofar as that is possible): $\tilde{N}_2/\tilde{N}_1 = N_2/N_1$. This correspondence however is not valid when an ideal transformer with separate leakage inductances is used for circuit analysis purposes. In the analytical case, the effective turns ratio for the actual transformer becomes the true ratio for the ideal transformer. This creates an auxiliary rule: leave the tildes on the Ns until sure that they are not required. Generally, although we will use Ñ here instead of N (where appropriate) for the sake of mathematical consistency, the difference is only likely to be significant when dealing with transformers which do not have low-reluctance cores. Bear in mind however, that when the tildes are discarded, it is the approximation (not the analytical relation) which is intended; i.e.:

$\mathbf{V_2} / \mathbf{V_1} \approx \mathbf{I_1} / \mathbf{I_2} \approx N_2 / N_1$	Perfect transformer approximation

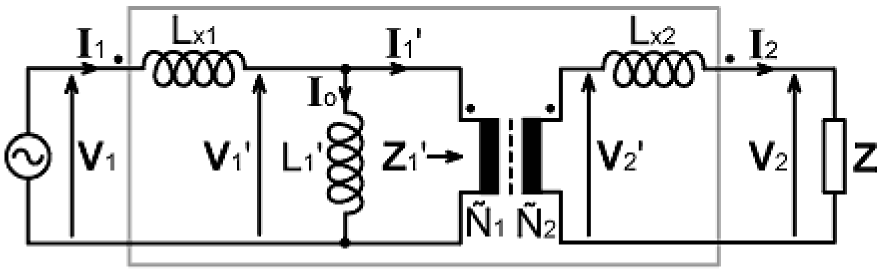
26. Perfect transformer as a circuit-element

This is a good point at which to examine the pros and cons of treating the transformer as a whole versus treating it as an ideal transformer with separate leakage inductances. The latter approach will, of course, be familiar to many as that which can be learned without recourse to electromagnetic fundamentals; but it is perhaps worth demonstrating that avoidance of the underlying concept (i.e., transimpedance) does not necessarily constitute the best line of attack. In general, it is preferable to be aware of the relationship between the two models, and pick whichever is best suited to the problem at hand.

In order to establish the mapping which relates the two approaches, it is necessary to partition the inductances using the voltage shortfall factors (k_1 and k_2). This operation puts us back in touch with the turns numbers and the path inductance factors. The various relationships, established over the course of the discussion from section **13** onwards, are summarised in the diagram below:



The magnetic coupling can be represented analytically using an ideal transformer having turns $\tilde{N}_1$ and $\tilde{N}_2$. Furthermore, as discussed in the preceding section; the ideal transformer can be represented as a perfect transformer with an inductance k_1L_1 in parallel with its primary port to provide the magnetising current. The redefined network is shown below:



Some comments regarding symbols and notation might be of value at this point. Notice that the leakage inductances are are given the subscript ' x ', which should be taken to stand for 'extra'. The high-voltage overwind of a voltage-magnifier transformer is sometimes referred-to as the 'extra coil', and the underlying concept is exactly the same as it is here. Also, using the letter ' l ' for leakage has the problem that it is indistinguishable from the figure ' 1 '.

The symbol used here for the perfect transformer is not reserved for that purpose. Instead, it is a general transformer symbol of the 'invented by draughtsmen for the convenience of draughtsmen' school; but it has the advantage in this context that it gives not the slightest hint that the inter-network coupling has anything to do with inductance. That is appropriate because all of the inductance has been removed and placed in the attached networks.

The perfect transformer becomes a new quasi-fundamental circuit element. Just like the other basic elements (resistance, capacitance, inductance); it cannot be realised in practice other than as an approximation, but it can be made to describe the behaviour practical devices by the inclusion of parasitics (i.e., the inductances shown in the diagram). Its defining properties, using the notation given above, are:

$\tilde{N}_2 / \tilde{N}_1 = \mathbf{V}_2' / \mathbf{V}_1' = \mathbf{I}_1' / \mathbf{I}_2$	**26.1**

When an inductance is connected across any one of the ports, the perfect transformer becomes an ideal transformer. For this discussion, we have elected to connect an inductance L_1' across port 1. Thus we can also avail ourselves of the ideal transformer relationship:

$\tilde{N}_2 / \tilde{N}_1 = \sqrt{(L_2' / L_1')}$ **26.2**

Squaring that gives:

$L_2' = L_1' \tilde{N}_2^2 / \tilde{N}_1^2$

The inductance L_2' does not appear on the diagram. It is the image of L_1' as seen, magnified by the square of the turns ratio, when looking into port 2. This means, of course, that we could account for the magnetising current by removing L_1' and placing an inductance L_2' across port 2 instead.

Recall that in section **24** we obtained the general voltage-transformation rule for the loaded transformer in a few lines of algebra. We will now repeat the exercise using the perfect-transformer approach.

Referring to the diagram above: by inspection of the secondary network; $\mathbf{V}_2$ is obtained from $\mathbf{V}_2'$ via a potential divider comprising $\mathbf{j}X_{Lx2}$ and $\mathbf{Z}$. Thus:

$\mathbf{V}_2 = \mathbf{V}_2' \mathbf{Z} / (\mathbf{j}X_{Lx2} + \mathbf{Z})$

but (**26.1**) gives:

$\mathbf{V}_2' = \mathbf{V}_1' \tilde{N}_2 / \tilde{N}_1$

Hence:

$\mathbf{V}_2 = \mathbf{V}_1' (\tilde{N}_2 / \tilde{N}_1) \mathbf{Z} / (\mathbf{j}X_{Lx2} + \mathbf{Z})$

$\mathbf{V}_1'$ is obtained from $\mathbf{V}_1$ via a potential divider comprising $\mathbf{j}X_{Lx1}$ and $\mathbf{j}X_{L1}'$ in parallel with the impedance looking into port 1. This gives:

$\mathbf{V}_1' = \mathbf{V}_1 (\mathbf{j}X_{L1}' // \mathbf{Z}_1') / (\mathbf{j}X_{Lx1} + \mathbf{j}X_{L1}' // \mathbf{Z}_1')$

Thus the unsimplified voltage ratio is:

$\mathbf{V}_2 / \mathbf{V}_1 = (\tilde{N}_2 / \tilde{N}_1) \mathbf{Z} (\mathbf{j}X_{L1}' // \mathbf{Z}_1') / [(\mathbf{j}X_{Lx2} + \mathbf{Z}) (\mathbf{j}X_{Lx1} + \mathbf{j}X_{L1}' // \mathbf{Z}_1')]$

This expression becomes more tractable when inverted, i.e.;

$\mathbf{V}_1 / \mathbf{V}_2 = (\tilde{N}_1 / \tilde{N}_2) [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 + \mathbf{j}X_{Lx1} / (\mathbf{j}X_{L1}' // \mathbf{Z}_1')]$

Also $1/(\mathbf{a}//\mathbf{b}) = (1/\mathbf{a})+(1/\mathbf{b})$. Therefore:

$\mathbf{V}_1 / \mathbf{V}_2 = (\tilde{N}_1 / \tilde{N}_2) [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 +(\mathbf{j}X_{Lx1} / \mathbf{j}X_{L1}') + \mathbf{j}X_{Lx1} / \mathbf{Z}_1']$

The impedance looking into port 1 is defined as:

$\mathbf{Z}_1' = \mathbf{V}_1' / \mathbf{I}_1'$

but (**26.1**) gives:

$\mathbf{V}_1' = \mathbf{V}_2' \tilde{N}_1 / \tilde{N}_2$ and $\mathbf{I}_1' = \mathbf{I}_2 \tilde{N}_2 / \tilde{N}_1$

Hence:

$\mathbf{Z}_1' = (\tilde{N}_1 / \tilde{N}_2)^2 \mathbf{V}_2' / \mathbf{I}_2 = (\tilde{N}_1 / \tilde{N}_2)^2 (\mathbf{j}X_{Lx2} + \mathbf{Z})$

i.e., the impedance looking into port 1 is the impedance of the secondary network multiplied by the square of the inverse turns ratio. Thus:

$\mathbf{V}_1 / \mathbf{V}_2 = (\tilde{N}_1 / \tilde{N}_2) [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 + (L_{x1} / L_1') + \mathbf{j}X_{Lx1} (\tilde{N}_2 / \tilde{N}_1)^2 / (\mathbf{j}X_{Lx2} + \mathbf{Z})]$

Using (**26.2**) to replace the turns ratios gives:

$\mathbf{V}_1 / \mathbf{V}_2 = [\sqrt{(L_1' / L_2')}] [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 + (L_{x1} / L_1') + \mathbf{j}X_{Lx1} (L_2' / L_1') / (\mathbf{j}X_{Lx2} + \mathbf{Z})]$

Now notice that:

$X_{Lx1} L_2' = 2\pi f L_{x1} L_2' = L_{x1} X_{L2}'$

Therefore:

$\mathbf{V}_1 / \mathbf{V}_2 = [\sqrt{(L_1' / L_2')}] [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 + (L_{x1} / L_1') + \mathbf{j}(L_{x1} / L_1') X_{L2}' / (\mathbf{j}X_{Lx2} + \mathbf{Z})]$

i.e.:

$\mathbf{V}_1 / \mathbf{V}_2 = [\sqrt{(L_1' / L_2')}] [(\mathbf{j}X_{Lx2} + \mathbf{Z}) / \mathbf{Z}] [1 + (L_{x1} / L_1')\{1 + \mathbf{j}X_{L2}' / (\mathbf{j}X_{Lx2} + \mathbf{Z}) \}]$

Multiplying $(\mathbf{j}X_{Lx2} + \mathbf{Z})$ into the right most square bracket gives:

$\mathbf{V}_1 / \mathbf{V}_2 = [\sqrt{(L_1' / L_2')}]\ (1 / \mathbf{Z})\ [\ \mathbf{j}X_{Lx2} + \mathbf{Z} + (\ L_{x1} / L_1')(\ \mathbf{j}X_{Lx2} + \mathbf{Z} + \mathbf{j}X_{L2}'\ )\ ]$
and re-inversion gives the solution:

$\mathbf{V}_2 / \mathbf{V}_1 = [\sqrt{(L_2' / L_1')}]\ \mathbf{Z} / [\ \mathbf{Z} + \mathbf{j}X_{Lx2} + (\ L_{x1} / L_1')(\ \mathbf{Z} + \mathbf{j}X_{Lx2} + \mathbf{j}X_{L2}'\ )\ ]$

Also, we might note that $L_{x2}+L_2' = L_2$, and so alternatively:

$\mathbf{V}_2 / \mathbf{V}_1 = [\sqrt{(L_2' / L_1')}]\ \mathbf{Z} / [\ \mathbf{Z} + \mathbf{j}X_{Lx2} + (\ L_{x1} / L_1')(\ \mathbf{Z} + \mathbf{j}X_{L2}\ )\ ]$	**26.3**

The relationships which give the mapping from the perfect transformer analysis to the transimpedance model are as follows:

$L_{x1} = (1-k_1)\ L_1$	$L_1' = k_1\ L_1$	$L_2' = k_2\ L_2$	$L_{x2} = (1-k_2)\ L_2$	$k = \sqrt{(k_1\ k_2)}$

Using these in (**26.3**) gives:
$\mathbf{V}_2 / \mathbf{V}_1 = [\sqrt{(k_2L_2 / k_1L_1)}]\ \mathbf{Z} / [\ \mathbf{Z} + \mathbf{j}(1-k_2)X_{L2} + \{\ (1-k_1) / k_1\ \}(\ \mathbf{Z} + \mathbf{j}X_{L2}\ )\ ]$
but notice that:
$k / k_1 = \sqrt{(k_1\ k_2 / k_1^{\ 2})} = \sqrt{(k_2 / k_1)}$
Using this as a substitution gives:
$\mathbf{V}_2 / \mathbf{V}_1 = k\ [\sqrt{(L_2 / L_1)}]\ \mathbf{Z} / [\ k_1\ \mathbf{Z} + \mathbf{j}(k_1 - k_1\ k_2)X_{L2} + (1-k_1)(\ \mathbf{Z} + \mathbf{j}X_{L2}\ )\ ]$
Multiplying out the denominator and noting that $k_1\ k_2 = k^2$ gives:
$\mathbf{V}_2 / \mathbf{V}_1 = k\ [\sqrt{(L_2 / L_1)}]\ \mathbf{Z} / [\ k_1\ \mathbf{Z} + \mathbf{j}k_1\ X_{L2} - \mathbf{j}k^2\ X_{L2} + \mathbf{Z} + \mathbf{j}X_{L2} - k_1\ \mathbf{Z} - \mathbf{j}k_1\ X_{L2}\ ]$
i.e.:

$\mathbf{V}_2 / \mathbf{V}_1 = k\ [\sqrt{(\ L_2 / L_1)}]\ \mathbf{Z} / [\ \mathbf{Z} + \mathbf{j}(1-k^2)\ X_{L2}\ ]$

This is the same as equation (**24.2**) and shows that the perfect-transformer-based analysis is exactly equivalent to the transimpedance approach.

Transimpedance is essentially Faraday's Law in phasor form. It captures the fundamental simplicity of transformer coupling; which is, after all, merely a reflection of the fact that inductors in proximity have overlapping magnetic fields. The result is that the magnetic interaction between any pair of coils can be completely specified for circuit-analysis purposes by using three easy-to-measure parameters: $\{L_1, L_2, M\}$; or alternatively: $\{L_1, L_2, k\}$.

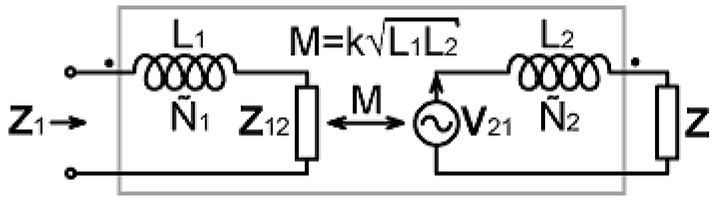
The perfect transformer is a 'black-box' circuit element which, as we have just seen, can be extracted from the underlying theory. It is governed by simple rules (**26.1**), which are easily learned by those who do not have a field-based understanding of electricity; and so it provides a way of sidestepping the conceptual difficulties associated with magnetic theory. That might provide an explanation for its widespread adoption; but as can be seen from the laborious working up to expression (**26.3**), it does not simplify the general analysis of transformer coupled networks, and neither does it produce an elegant expression for the voltage transfer function. Instead, it complicates matters by partitioning the inductances, and thereby increases the number of parameters needed to specify a coupled system. The possible minimal parameter sets are as established in the discussion above:
$\{L_{x1}, L_1', \tilde{N}_2/\tilde{N}_1, L_{x2}\}$ or $\{L_{x1}, \tilde{N}_2/\tilde{N}_1, L_2', L_{x2}\}$ or $\{L_{x1}, L_1', L_2', L_{x2}\}$
from which it can also be seen that, except perhaps for the turns ratio (and then only when $\tilde{N}=N$); the chosen parameters are equal to quantities which are difficult to estimate.

The intention here however, is not to denigrate the perfect transformer as a circuit-analysis construct, but merely to counter the idea that it is the culmination of all of the foregoing theory. It is a valuable addition to the analytical tool-kit, well suited to problems involving nearly-ideal transformer behaviour; but it is not the only available approach, and it can be cumbersome when inappropriately applied.

27. Impedance transformation and referral

One of the major uses of the transformer, particularly in RF power-transmission and signal-processing applications, is that of impedance transformation. This is a very large subject area, and no attempt will be made to do justice to it here; but we must at least establish the basic principles. A significant practical issue is that, in the process of connecting the output of one electrical network to the input of another; it is often found that the preferred load impedance of the driving stage is not the same as the input impedance of the driven stage. A conventional transformer, with its ability to step voltage up and step current down (or vice versa), is therefore a good candidate for all or part of an intervening impedance-matching network.

The generalised circuit-analysis problem is illustrated below. An impedance **Z** is connected to the secondary network; an impedance $\mathbf{Z}_1$ is seen looking into the primary port; and we want to know how the relationship between $\mathbf{Z}_1$ and **Z** is determined by the transformer parameters.



From the diagram we have:

$\mathbf{Z}_1 = \mathbf{j}X_{L1} + \mathbf{Z}_{12}$

The transferred impedance $\mathbf{Z}_{12}$ was given earlier as equation (23.3):

$\mathbf{Z}_{12} = X_M^2 / (\mathbf{j}X_{L2} + \mathbf{Z})$

Hence:

$\mathbf{Z}_1 = \mathbf{j}X_{L1} + X_M^2 / (\mathbf{j}X_{L2} + \mathbf{Z})$

which, using the substitution $M=k\sqrt{(L_1 L_2)}$ becomes:

$\mathbf{Z}_1 = \mathbf{j}X_{L1} + k^2 X_{L1} X_{L2} / (\mathbf{j}X_{L2} + \mathbf{Z})$

Putting both terms onto a common denominator gives:

$$\mathbf{Z}_1 = \frac{\mathbf{j}X_{L1}(\mathbf{j}X_{L2} + \mathbf{Z}) + k^2 X_{L1} X_{L2}}{\mathbf{j}X_{L2} + \mathbf{Z}} = \frac{\mathbf{j}X_{L1}\mathbf{Z} + k^2 X_{L1} X_{L2} - X_{L1} X_{L2}}{\mathbf{j}X_{L2} + \mathbf{Z}}$$

i.e.:

$$\mathbf{Z}_1 = \frac{\mathbf{j}X_{L1}\mathbf{Z} + (k^2-1) X_{L1} X_{L2}}{\mathbf{j}X_{L2} + \mathbf{Z}}$$

This is about as simple as the general expression can be; but note that some textbooks try to simplify it further by multiplying numerator and denominator by $(\mathbf{Z}-\mathbf{j}X_{L2})$. The problem with that is that, because **Z** is complex, $(\mathbf{Z}-\mathbf{j}X_{L2})$ is not the same as the complex conjugate $(\mathbf{Z}+\mathbf{j}X_{L2})^*$. Hence the operation does not have the effect of making the denominator real and so misses the point.

The practical simplification comes when the transformer is tightly coupled. In that case, $k \rightarrow 1$ and we have:

$\mathbf{Z}_1 = \mathbf{j}X_{L1}\mathbf{Z} / (\mathbf{j}X_{L2} + \mathbf{Z})$

This is very nearly an expression for two impedances in parallel, and will become so if we can represent X_{L1} as a factor of X_{L2}. The relationship which does that, in the ideal transformer approximation is, of course:

$L_1 = L_2 N_1^2/N_2^2$

Hence:

$\mathbf{Z}_1 = (N_1^2/N_2^2)\, \mathbf{j}X_{L2}\mathbf{Z} / (\mathbf{j}X_{L2} + \mathbf{Z})$

i.e.: When $k \rightarrow 1$, $\mathbf{Z}_1 \rightarrow (N_1^2/N_2^2)\,(\mathbf{j}X_{L2} // \mathbf{Z})$

Also, we can represent X_{L2} as a factor of X_{L1} using:

$L_2 = L_1 N_2^2/N_1^2$

Hence:

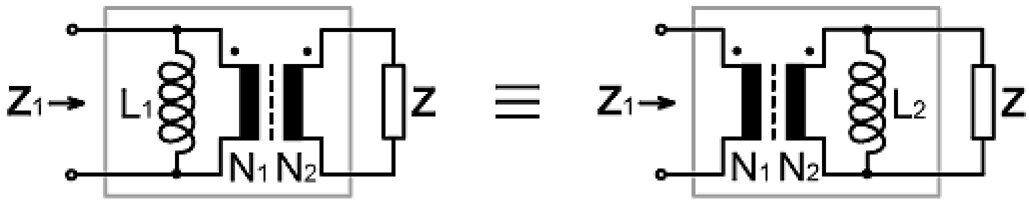
$\mathbf{Z}_1 = \mathbf{j}X_{L1}\mathbf{Z} / [\,(N_2^2/N_1^2)\, \mathbf{j}X_{L1} + \mathbf{Z}]$

This rearranges to:

$\mathbf{Z}_1 = \mathbf{j}X_{L1}\,(N_1^2/N_2^2)\,\mathbf{Z}\,/\,[\,\mathbf{j}X_{L1} + (N_1^2/N_2^2)\,\mathbf{Z}]$
i.e.; $\mathbf{Z}_1 = \mathbf{j}X_{L1} \,//\, (N_1^2/N_2^2)\,\mathbf{Z}$
Thus, we have two equivalent expressions:

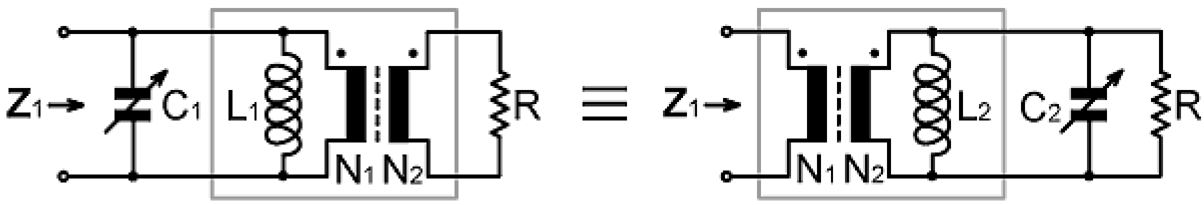
When $k\rightarrow 1$, $\mathbf{Z}_1 \approx \mathbf{j}X_{L1} \,//\, (N_1^2/N_2^2)\,\mathbf{Z} = (N_1^2/N_2^2)\,(\,\mathbf{j}X_{L2} \,//\, \mathbf{Z}\,)$

The meaning of the two alternatives is illustrated by the pair of equivalent perfect-transformer networks shown below:



A closely-coupled transformer transforms an impedance according to the inverse-square of the turns-ratio (assuming that the impedance is connected to port 2, and we define the turns ratio as N_2/N_1); but it always places its own inductance in parallel with the result. The transformation rule also implies that it makes no difference to $\mathbf{Z}_1$ whether the inductance is represented as L_2 across port 2 or as L_1 across port 1. An important corollary is that any parallel impedance component can be transferred from one side of an ideal transformer to the other by multiplying it by the square of the inverse turns ratio (2→1) or the square of the turns ratio (1→2) as appropriate.

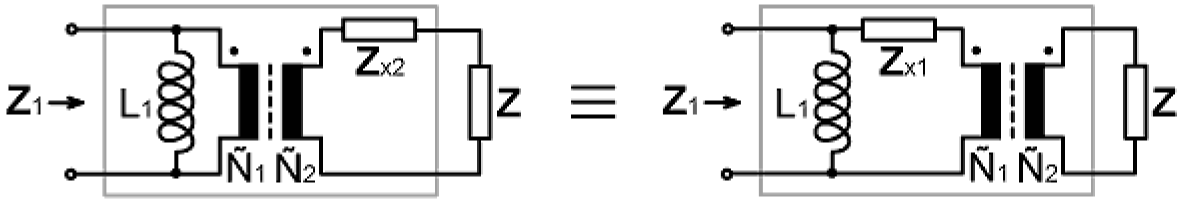
It was mentioned earlier that it is not a good idea to neglect transformer inductance when designing RF circuits. There are however, various ways of ensuring that it has negligible effect. For the inter-stage coupling and output transformers used in broadband radio systems, there is often little recourse but to engineer it out. That cannot usually be done by using large numbers of turns (because the winding wire must be kept short), and so the preferred method is to use cores of very low reluctance (i.e., large A_L). For narrow-band systems however, it is possible to resonate the inductance against a capacitance and so send the combined reactance to infinity. The principle is illustrated below, where we see that there are also two possible placements for the tuning capacitor.



Here we have assumed, for the sake of argument, that the load impedance is a pure resistance. The combined reactance then becomes open circuit when:
$|\,X_{L1} \,//\, X_{C1}\,|\rightarrow\infty$; or when: $|\,X_{L2} \,//\, X_{C2}\,|\rightarrow\infty$
This, in either case, corresponds to the condition $X_L+X_C=0$, which occurs when $f=1/2\pi\sqrt{(LC)}$.
Now recall that $L_2=(N_2/N_1)^2 L_1$. It follows that $X_{C2} = (N_2/N_1)^2\, X_{C1}$. Capacitive reactance is however inversely proportional to capacitance, and so:
$C_2 = (N_1/N_2)^2\, C_1$
Thus, if we have (say) $N_2>N_1$, then we can tune out the inductance by using a small capacitance across port 2 or a large capacitance across port 1. A corollary incidentally, is that the effect of a tuning capacitor can be scaled or modified (say, for bandspreading) by coupling it to a resonator coil either via a separate winding or via a tap connection. A thorough discussion of this technique, for both loosely and tightly coupled transformers, is given by Bob Weaver[15].

15 http://electronbunker.ca/Bandspreading_3.html

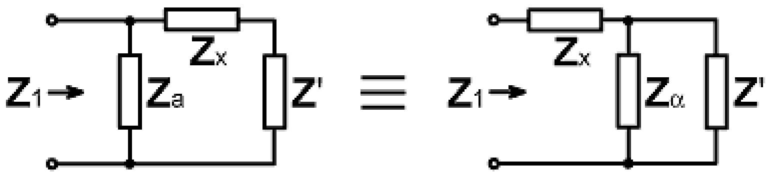
What we have so far established is that a parallel impedance component, whether wanted or parasitic, can be referred from one side of an ideal transformer to the other by multiplying it by the square of the turns ratio (or the inverse of that, depending on the definition). It is also possible to separate the impedance loading a transformer into two or more series-connected elements and refer any of those elements to the other side. The operation is illustrated below:



In the left-hand case: $\mathbf{Z}_1 = \mathbf{j}X_{L1} \,//\, (\tilde{N}_1/\tilde{N}_2)^2\,(\mathbf{Z}_{x2} + \mathbf{Z}) = \mathbf{j}X_{L1} \,//\, [\,(\tilde{N}_1/\tilde{N}_2)^2\,\mathbf{Z}_{x2} + (\tilde{N}_1/\tilde{N}_2)^2\,\mathbf{Z}\,]$
In the right-hand case: $\mathbf{Z}_1 = \mathbf{j}X_{L1} \,//\, [\,\mathbf{Z}_{x1} + (\tilde{N}_1/\tilde{N}_2)^2\,\mathbf{Z}\,]$
Therefore: $\mathbf{Z}_{x1} = (\tilde{N}_1/\tilde{N}_2)^2\,\mathbf{Z}_{x2}$

A common reason for wanting to split an impedance into series-connected elements is, of course, to separate the external network from the transformer parasitics. That being the case, it would be convenient not only to shift the parasitics to the other side of the prefect-transformer element, but also to shift them to the other side of the parallel inductance (L_1 in this representation, but it is to be understood that the referral can go either way). That would permit aggregation of primary and secondary leakage inductances and series losses into a single impedance, to be placed either on the primary or secondary side. The required network transformation is generalised in the diagram below, where $\mathbf{Z}_a$ (which might represent the magnetising inductance L_1) must be replaced by a new impedance $\mathbf{Z}_\alpha$ on the other side of $\mathbf{Z}_x$. The impedance $\mathbf{Z}'$ might represent the load impedance $\mathbf{Z}$ (say) after it has been referred through a perfect transformer.

In the left-hand network:
$\mathbf{Z}_1 = \mathbf{Z}_a \,//\, (\mathbf{Z}_x + \mathbf{Z}')$
In the right-hand network:
$\mathbf{Z}_1 = \mathbf{Z}_x + \mathbf{Z}_\alpha \,//\, \mathbf{Z}'$
Hence:
$\mathbf{Z}_\alpha \,//\, \mathbf{Z}' = \mathbf{Z}_a \,//\, (\mathbf{Z}_x + \mathbf{Z}') - \mathbf{Z}_x$
and taking the reciprocal gives:



$$\frac{1}{\mathbf{Z}_\alpha} + \frac{1}{\mathbf{Z}'} = \frac{1}{\mathbf{Z}_a \,//\, (\mathbf{Z}_x + \mathbf{Z}') - \mathbf{Z}_x} = \frac{1}{\dfrac{\mathbf{Z}_a\,(\mathbf{Z}_x + \mathbf{Z}')}{\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z}'} - \mathbf{Z}_x}$$

The denominator can be put on to a common denominator, which becomes the numerator:

$$\frac{1}{\mathbf{Z}_\alpha} + \frac{1}{\mathbf{Z}'} = \frac{\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z}'}{\mathbf{Z}_a\,(\mathbf{Z}_x + \mathbf{Z}') - (\,\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z}'\,)\,\mathbf{Z}_x} = \frac{\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z}'}{\mathbf{Z}_a\mathbf{Z}_x + \mathbf{Z}_a\mathbf{Z}' - \mathbf{Z}_a\mathbf{Z}_x - \mathbf{Z}_x^2 - \mathbf{Z}_x\mathbf{Z}'}$$

Moving $1/\mathbf{Z}'$ to the right hand side then gives:

$$\frac{1}{\mathbf{Z}_\alpha} = \frac{\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z'}}{\mathbf{Z}_a \mathbf{Z'} - \mathbf{Z}_x^2 - \mathbf{Z}_x \mathbf{Z'}} - \frac{1}{\mathbf{Z'}} = \frac{\mathbf{Z'}(\mathbf{Z}_a + \mathbf{Z}_x + \mathbf{Z'}) - \mathbf{Z}_a \mathbf{Z'} + \mathbf{Z}_x^2 + \mathbf{Z}_x \mathbf{Z'}}{(\mathbf{Z}_a \mathbf{Z'} - \mathbf{Z}_x^2 - \mathbf{Z}_x \mathbf{Z'})\,\mathbf{Z'}}$$

Notice here that the numerator factorises, i.e.: $\mathbf{Z'}^2 + 2\mathbf{Z}_x \mathbf{Z'} + \mathbf{Z}_x^2 = (\mathbf{Z'} + \mathbf{Z}_x)^2$. Therefore, on re-inversion, we get:

$$\mathbf{Z}_\alpha = \frac{\mathbf{Z'}^2 \mathbf{Z}_a - \mathbf{Z'} \mathbf{Z}_x^2 - \mathbf{Z'}^2 \mathbf{Z}_x}{(\mathbf{Z'} + \mathbf{Z}_x)^2} = \frac{\mathbf{Z'}^2 \mathbf{Z}_a - \mathbf{Z'} \mathbf{Z}_x (\mathbf{Z'} + \mathbf{Z}_x)}{(\mathbf{Z'} + \mathbf{Z}_x)^2}$$

Now observe that: $\mathbf{Z'} \mathbf{Z}_x / (\mathbf{Z'} + \mathbf{Z}_x) = \mathbf{Z'} // \mathbf{Z}_x$. Therefore:

$$\boxed{\mathbf{Z}_\alpha = \frac{\mathbf{Z'}^2 \mathbf{Z}_a}{(\mathbf{Z'} + \mathbf{Z}_x)^2} - \mathbf{Z'} // \mathbf{Z}_x}$$

It is obvious that, if $\mathbf{Z}_a$ represents a pure inductance, then the impedance $\mathbf{Z}_\alpha$ which replaces it not a pure inductance. The new parallel impedance will acquire some resistive character if $\mathbf{Z}_x$ or $\mathbf{Z'}$ contains resistance, and its reactance can become capacitive under some circumstances. The importance of the expression however, lies not in its exact form, but in its limiting behaviour. Firstly:

When $|\mathbf{Z'}| \to \infty$, $\mathbf{Z}_\alpha \to \mathbf{Z}_a - \mathbf{Z}_x$; i.e.: $\mathbf{Z}_\alpha + \mathbf{Z}_x \to \mathbf{Z}_a$

This can be deduced directly by inspecting the diagram above. It implies that when a transformer is lightly loaded, a parallel inductance can be referred to the other side of a series inductance and still retain an approximately pure inductive character. The outcome of greatest utility hoverer is that:

When $|\mathbf{Z}_x| \to 0$, $\mathbf{Z}_\alpha \to \mathbf{Z}_a$

If the series impedance elements are of relatively small magnitude, they can be referred unchanged from one side of a parallel impedance to the other. This facility is the basis upon which the ideal transformer, with relatively small leakage inductances and loss resistances, is converted into the standard model for the closely-coupled transformer. It is also the basis on which a short length of transmission line can be represented, in the lumped-element approximation, as a series inductance and a parallel capacitance; a correspondence which will give us insight into the lesser phase-shifting effect commonly known as 'winding capacitance'.

28. Capacitance and propagation delay

For the purposes of lumped-element circuit analysis; the reactance of an inductor is best expressed not as that of a pure inductance, but as that of an inductance in parallel with a small capacitance. In discussion of isolated coils, the negative reactive component is known as the 'self-capacitance', whereas in the context of transformers, it is often referred-to as the 'winding capacitance'. In either case, the parallel capacitance allows for the fact that every coil has a fundamental self-resonance frequency (SRF) at which the reactance goes to infinity. When modelling transformers moreover, it will be found that parallel capacitances, judiciously placed, can help to account for a small phase lag (relative to that of a purely inductive model) which occurs in the voltage output of actual devices.

In coils and transformers having layered windings, a major proportion of the self-capacitance can be attributed to the overlap of the layers. This is simple static capacitance, albeit easier to explain than to calculate; which can be fairly represented as a single capacitance placed across the terminals of an inductor, or as capacitances across the ports and across the perfect-transformer element of a transformer model. In RF practice however, it is normally desirable to minimise stray capacitance and, since the required inductances are usually small, that can be done by winding each coil in a single layer. For single-layer windings, with the obscuring effect of static capacitance largely removed, a somewhat different explanation for the phase-lag arises from the experimental data; and it is one which provides an unexpected corroboration of the theory advanced so far.

The self-resonance mechanism is discussed in detail in a separate article[16], but a brief summary is appropriate here. The SRF of an isolated coil is most easily measured by scattering radiation from it. That can be done, for example, by using a grid-dip oscillator (GDO). If the coil-wire is stiff enough to allow the structure to be self-supporting, then the measurement can be made in the absence of dielectric materials other than air; in which case it will be found that the fundamental SRF occurs when the wavelength of the incident radiation is twice the length of the electrical conductor (i.e., the length of the wire when straightened-out). Thus, if the wire length is ℓ_w, we have:

$\ell_w \approx \lambda_0 / 2$

where $\lambda_0 = c/f$, is the free-space wavelength. The relationship is accurate to within a few % for solenoid coils which have some space between the turns and are made of wire which is thin relative to the overall coil diameter.

It follows that the SRF corresponds to a transmission-line resonance. When all of the parasitic effects are taken away, the fundamental resonance occurs at the frequency at which an electromagnetic wave, reflecting from the impedance-discontinuities which occur at the ends of the wire, arrives back at its starting point in phase with itself. Furthermore, there is not just a single resonance, but a series of overtones; with parallel (high-impedance) resonances occurring when the wire-length is a half-integer number of electrical wavelengths, and series (low-impedance) resonances occurring when the length is a whole-integer number of electrical wavelengths. This is the behaviour of a short-circuited transmission line.

The impedance of a lossless transmission line is given by the expression:

$$\mathbf{Z_{in}} = \frac{R_0\,[\,\mathbf{Z} + \mathbf{j}R_0\,\mathrm{Tan}(2\pi\,\ell_{TL}/\lambda)\,]}{[R_0 + \mathbf{jZ}\,\mathrm{Tan}(2\pi\,\ell_{TL}/\lambda)\,]} \qquad (28.1)$$

where $\mathbf{Z}$ is the terminating impedance, R_0 is the characteristic resistance, ℓ_{TL} is the line length, and $\lambda=v/f$ is the electrical wavelength (v is the phase velocity). If the line is short-circuited by putting $\mathbf{Z}=0$ we have:

$\mathbf{Z_{in}} = \mathbf{j}R_0\,\mathrm{Tan}(2\pi\,\ell_{TL}/\lambda)$

Parallel resonances occur when $\mathrm{Tan}(2\pi\ell_{TL}/\lambda)\rightarrow\infty$, and the fundamental instance corresponds to $\ell_{TL} = \lambda/4$. Thus the SRF of a coil can be represented as the fundamental resonance frequency of a short-circuited quarter-wave line. Notice, incidentally, that this resonance occurs when the length of the winding wire is $\lambda/2$. This is not a paradox, the reason being that the total length of conductor in a $\lambda/4$ line is $\lambda/2$.

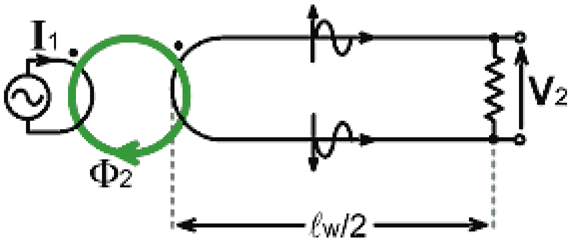
Further information about the transmission-line behaviour of coils can be had by making precise phase measurements on transformers. In another article[17], a method is described whereby the phase of the output voltage of a transformer, relative to the input current, can be measured with a precision of a few milli-degrees. The outcome of experiments is that, after correction for static capacitance and inductance, the graph of phase error vs. frequency is a near-perfect straight line of negative

16 **Self-resonance and self-capacitance of solenoid coils**. D W Knight. [available from www.g3ynh.info]

17 **Evaluation and optimisation of current-transformer bridges**. D W Knight. [www.g3ynh.info]

gradient. A quantity measured in (say) degrees per MHz is a *time*. It needs only to be divided by 360 to convert it into a time in microseconds, and the minus sign indicates that it is a delay. Furthermore, the delay-figure obtained per winding is close to the time it takes for light to propagate a distance equal to half the length of the winding wire.

The foregoing gives us a clear picture of the way in which electromagnetic energy propagates through transformers. To understand it, start by considering the effect of a magnetic disturbance occurring exactly half-way along the wire of a secondary winding (see diagram). The disturbance applies equal and opposite electromagnetic forces to the conductor in the upstream (towards the dot) and downstream directions, resulting in the launch of equal-and opposite waves which propagate towards the terminals. The two waves combine at the terminals to produce a contribution to the output voltage, because they produce equal and opposite displacements of the terminal potentials.

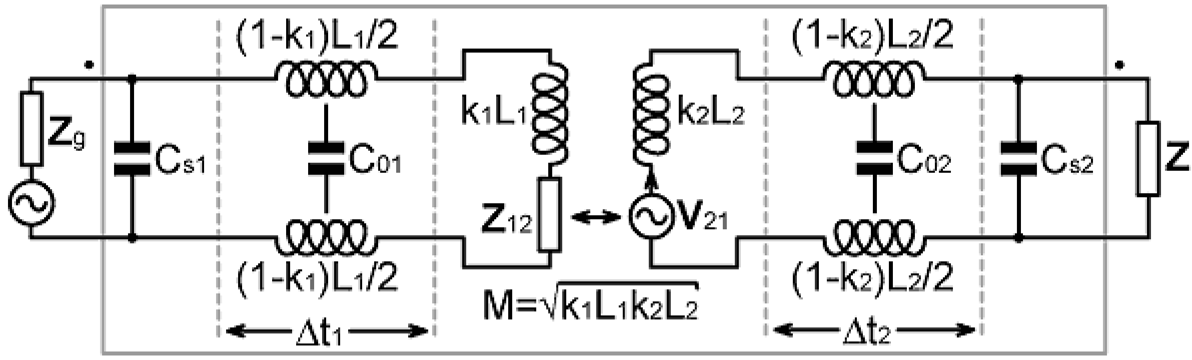


Now consider what happens when a magnetic disturbance occurs at an arbitrary point on the wire. In that case, this distances travelled by the two waves are not equal, and so the potential displacements at the terminals are not equal and opposite. The propagation process however (as will be later inferred from the phase velocity) occurs mainly outside any transformer-core material which may be present, and so takes place in an environment which is reasonably linear with regard to electromagnetic stress. Hence there will be no mixing between components at different frequencies, and we may analyse the behaviour of the system by considering only sine-waves. The sum of two sine-waves is another sine-wave. Hence the displacements at the two terminals will produce a resultant which is a sine-wave, and the phase of that will be the same as that of a wave which has travelled the average of the two distances. Thus the resultant output for a disturbance occurring at any point on the wire will be the same as that arising from a disturbance exactly half-way along the wire.

The consequences of this superposition process are twofold. Firstly; electromagnetic waves emerging from a port of a conventional transformer will appear to have come from a point exactly half-way along the winding wire. Secondly; waves resulting from an infinite number of simultaneous disturbances occurring at points distributed over the length of the wire will add together in phase. The latter corollary constitutes a dynamic (i.e., electromagnetic, rather than purely magnetic) explanation for the phenomenon of flux-linkage; the reason being that the divergence theorem (div$\boldsymbol{B}$=0) implies that flux lines are always continuous and are therefore contours of simultaneity. It also implies that the coupled inductance of the winding will be cancelled from the point-of-view of electromagnetic energy propagating out of (or into) the transformer; which is exactly what was established from circuit-theory considerations in section **23**.

The inductance which is not cancelled is, of course, the leakage inductance. Hence a wave propagating out of a transformer winding does so via a transmission line of total inductance equal to the leakage inductance, and physical length equal to half the length of the wire. The view that the leakage inductance has a definable physical length is, incidentally, corroborated by the path-analysis discussion of section **17**. There it was established that, from an analytical viewpoint, the leakage inductance of a particular winding has the same number of turns as the total inductance. We have tended to regard the turns number as a measure of flux linkage, but it is also a measure of wire length. The leakage inductance is distributed along the winding, not isolated into a lump, and so it retards waves progressively as they travel the wire.

The result of these deliberations is a modified transformer model as shown in the diagram below. This may appear to be somewhat complicated on first appraisal, but we will later resolve it into lumped elements using the impedance-referral techniques discussed in the previous section. Notice first that some stray or static capacitances, C_{s1} and C_{s2} have been placed across the ports. Such capacitances are inevitable in any device capable of sustaining a potential difference across its terminals, which is another way of saying that capacitance results from the absence of a short circuit. Also notice that the generator connected to the primary has been provided with a series impedance $\mathbf{Z_g}$, the point being that the transformer input capacitance can have no effect on circuit behaviour unless the generator has a finite output impedance. The major change however, is the inclusion of two transmission lines, which give rise to time delays Δt_1 and Δt_2.



As energy flows through a transformer, each winding causes a delay equal to the time it takes for an electromagnetic wave to travel half the length of the conductor. Velocity is distance divided by time, and so:
$\Delta t = -(\ell_w/2) / v$
The delay is given a minus sign because it gives rise to a phase lag in the transformer output voltage. The quantity v is the phase velocity (or wave velocity), and is given by:
$v = 1/\sqrt{(\mu\varepsilon)}$
where μ and ε are respectively the absolute permeability and permittivity of the transmission-line environment. Note that μ, written here without subscripts, is *not the same* as the transformer-core permeability; the lack of adornment being purely to minimise clutter for the purposes of this discussion. Absolute permeability and permittivity can be further separated into the free-space and relative quantities, i.e.;
$v = 1/\sqrt{(\mu_0\, \mu_r\, \varepsilon_0\, \varepsilon_r)}$
but $1/\sqrt{(\mu_0\, \varepsilon_0)}$ is the speed of light, c. Hence:
$v = c/\sqrt{(\mu_r\, \varepsilon_r)}$
The quantity $\sqrt{(\mu_r\, \varepsilon_r)}$ is known as the *refractive index* of the medium, and in an optical context is usually given the symbol ' n '. The quantity $1/\sqrt{(\mu_r\, \varepsilon_r)}$ or $1/n$ is known as the *velocity factor*, and is simply the proportion by which the phase velocity differs from the speed of light. Hence:
$v = c / n$

When considering the transfer of energy via a transformer, it is tempting to imagine that it involves the propagation of electromagnetic waves through the transformer core . Were that the case however, we would expect to measure enormous retardation in any device having a ferromagnetic core. This issue can be understood by estimating the refractive index of the core material. For the time-delay study cited earlier (for example), the transformers were based on Micrometals type 61 nickel-zinc ferrite, which has in initial permeability (i.e., relative permeability at low flux densities) of 125 at frequencies of a few MHz. The relative permittivity (dielectric constant) of ferromagnetic materials is more difficult to obtain, but data given by Snelling[18] suggest that it lies in the 10 to 100 range for NiZn ferrites. Thus, if we assume, for the sake of argument, that the core has an ε_r of about 20, we might expect a refractive index in the region of $\sqrt{(125 \times 20)} = 50$, and a velocity factor of $1/50 = 0.02$. In fact, for a single-layer winding of

18 **Soft Ferrites: Properties and Applications**. E C Snelling. 2nd ed. Butterworth. 1988. ISBN 0-408-02760-6. Permittivity of ferrites: p127 - 129.

enamelled-copper magnet-wire on a type 61 toroid, the time-delay corresponded to a refractive index of about 1.2 , i.e., a transmission-line velocity factor of about 0.83. From this we must conclude that the major part of the propagation process takes place in an environment which has a low relative permeability. This is consistent with the earlier observation; that the distributed inductance of the line is inductance which has failed to participate in the magnetic coupling, i.e., it is the leakage inductance. The leakage inductance is composed of internal inductance (flux loops within the body of the conductor) and the inductance due to flux loops which fail to enter the magnetic core. Hence, presuming that what lies outside the core (including the wire) is non-ferromagnetic; the relative permeability for the transmission-line environment will be that which is typical of non-ferromagnetic materials, i.e., it will be within a few parts-per-thousand of unity.

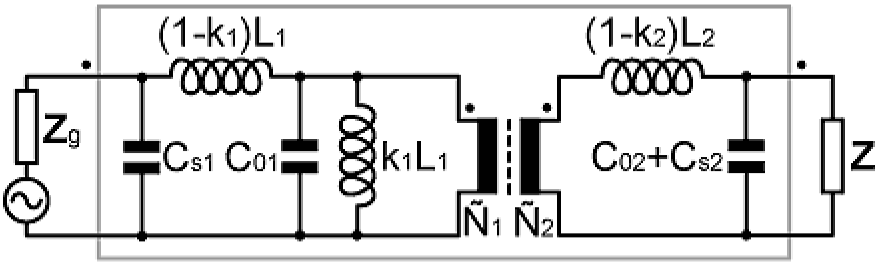
When making coils and transformers, it is necessary to use wire of high conductivity in order to minimise resistive losses. Hence nearly all transformers are wound either with copper wire, or with silver-plated copper wire. Both metals have μ_r close to 1, as do wire insulation materials such as enamel or PTFE. It follows that we can usually assume $\mu_r \approx 1$ when estimating the transmission-line phase velocity, and any deviation from c is almost entirely attributable to the average relative permittivity (dielectric constant) of the environment in the vicinity of the wire surface. Thus we have: $n=\sqrt{(1 \times \varepsilon_r)}$, which gives ε_r=1.44 when n=1.2 . This is reasonable given that the ***E***-field of a travelling-wave in an RF transformer will exist mostly in the air around the wire, but also partly in the wire insulation, and it may penetrate the magnetic core material to a small extent. An additional refractive effect will be that of the coil itself, which scatters radiation in such a way as to make the resultant wave follow a helical path, and must therefore make a contribution to the local permittivity. The latter will be somewhat frequency-dependent (a coil is a dispersive transmission line); and so it is to be expected that the propagation delay will not be exactly constant. This means that there will be some temporal-spreading of pulse-type signals as they pass through the transformer (i.e., different frequency components will have different phase velocities).

The presence of transmission lines leading into and out-of the transformer implies, of course, that there will be impedance transformations in addition to that which is performed by the coupled inductances. Those extra transformations can, in principle, be handled using equation (**28.1**); but in practice it is a difficult problem due to lack of knowledge of the line characteristic-resistance (or 'surge resistance') R_0 . The surge resistance of a lossless line is given by:

$R_0 = \sqrt{(L_0 / C_0)}$

where L_0 is the inductance per unit length, and C_0 is the capacitance per unit length, of the line. The total line inductance is, of course, the leakage inductance of the winding, and finding the length is straightforward; which means that the average distributed inductance is experimentally determinable. The distributed capacitance however, is difficult to separate from stray-capacitance, and is therefore usually unknown. The issue is not completely intractable, but pragmatism dictates that the most straightforward approach to RF transformer design is to keep the length of the winding-wire sufficiently short that the spurious impedance transformation can be modelled using a lumped series inductance and a lumped parallel capacitance. A rough rule-of-thumb for that approximation is that the wire should be no longer than about $\lambda/10$ at the highest frequency of operation, but note that even that can sometimes prove to be excessive. Generally however, most circuits are only required to conform to the model to within a few % , and in situations requiring very high precision, it is usually possible to define a set of parameter adjustments which are capable of absorbing model deficiencies.

What has been established is that the transmission-line impedance-transformation can be modelled by placing empirically-determined shunt capacitances on the load-sides of the lumped leakage inductances, *provided that* the length of the winding wire is substantially less than $\lambda/2$. The textbook standard transformer model relies on this approach; even though most commentators do not mention the length issue or identify what the network actually does. The reason for the neglect of transmission-line effects is that the model is inherited from the utility power-distribution industry, where $\lambda_0/2$ is about 3000Km for 50Hz, and 2500Km for 60Hz. Thus, for mains transformers, it would be unusual to have a conductor-length approaching even $\lambda_0/1000$; and so the lumped-element approach is very accurate. For RF transformers however, it is often difficult to achieve $\ell_w<\lambda/10$ (e.g., only 1m of wire to make a transformer for 30MHz). This implies that model validity is (or should be) an ever-present concern for the RF designer; but the lumped-parameter representation can still give excellent results when judiciously applied. Thus, cautiously, we arrive at the interim lossless network representation shown below.



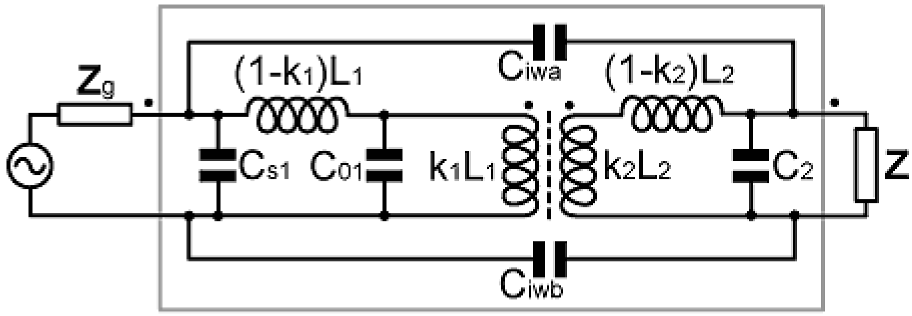
The element sets { $(1\text{-}k_1)L_1$, C_{01} } and { $(1\text{-}k_2)L_2$, C_{02} } are L-networks, which can be adjusted to reproduce the transmission-line impedance transformations over a reasonably broad range of frequencies provided that the winding conductor-lengths are electrically short. The capacitance C_{01} can moreover be referred to the other side of the perfect-transformer element (by dividing it by the square of the turns-ratio), and if the secondary leakage inductance is small, it can then be aggregated with the capacitances across the secondary port. Also, if the magnetising inductance (k_1L_1) is relatively large, the primary leakage inductance can be moved to the secondary side (by multiplying it by the square of the turns ratio). In this way, for transformers having only two windings, the impedance transformations due to both lines can be modelled by a single L-network. A further point worth noting here is the matter of parameter correlation. If the model parameters are to be determined by (say) fitting them to frequency-response data; then it will not be possible to distinguish between the effects of primary and secondary-side parasitics in a closely-coupled two-coil transformer, and in that case it will be necessary to aggregate parameters in order to get the fitting process to converge. Such parameter aggregation has the effect of divorcing the network representation from the underlying physics, but the resulting empirical model still gives an accurate account of the behaviour of actual devices.

The modelling error which results from representing a transmission line as an L-network is, incidentally, usually small; but it can become significant in circuits (such as RF bridges) which require very precise phase-frequency tracking . The issue is that the addition of a parallel capacitance moves an impedance clockwise around a circle of constant conductance[19], whereas a time-delay moves a **V/I** relationship clockwise around a circle centred on the origin of the Z-plane. These two effects are experimentally indistinguishable when the phase-shift is small, but diverge when the phase-shift becomes large. This is, of course, a reiteration of the point that it is desirable to have $\ell_w << \lambda/2$.

On the subject of adding capacitances to the transformer model, there remains one final issue. This is that the separate coils must be placed in close proximity; which means that, in the absence of some kind of electrostatic shielding, there will be stray capacitance between the windings. Such capacitance is known as *inter-winding capacitance,* and can cause pronounced effects in signal-processing circuits. In bandpass filters, for example, it can give rise to poor out-of band signal attenuation; and in RF bridge circuits, it gives rise to phase errors which require correction. Strictly, inter-winding capacitance is distributed in the gap between the coils, but the effect is fairly well

19 **Impedance matching**. D W Knight. [available from www.g3ynh.info]

approximated by adding capacitances as shown below:



29. Faraday shielding

The standard solution to the inter-winding capacitance problem is to provide an electrostatic screen (known as a *Faraday shield*). In mains transformers, this is done by placing a sheet of copper foil between the primary and secondary windings (with insulating layers on both sides); the foil having an insulated overlap, so that it eliminates any line-of-sight from one coil to the other but does not form a shorted turn. When the shield is grounded (via a wire of low inductance) capacitive coupling is almost completely eliminated. The effect of this arrangement depends on the fact that power-line transformers are extremely lossy at high-audio and radio frequencies; and so the shield prevents mains-borne noise from appearing on the power-rail, and it prevents high-frequency signals from being injected into the house wiring.

In RF transformers, Faraday shielding is often effected by making one of the windings from coaxial cable (see illustration). This makes the transmission-line parameters of the shielded winding easier to estimate; but note that it is imperative that the shield is only earthed at one end (it constitutes a shorted winding if earthed at both ends), and it is therefore not completely safe to assume that all of the leakage inductance is confined within the cable (shield and centre-conductor currents are not equal-and-opposite).

It is commonly assumed that the Faraday shield converts all inter-winding capacitance into capacitance to ground. This however is not a proper interpretation of the effect, as can be understood by considering the coaxial-cable screening method. The inner and outer conductors of the cable make the same number of turns around the transformer core, and there is a relatively large capacitance between them. Hence the shield forms an additional transformer winding, with its free-end capacitively coupled to to the other windings (particularly to the one it shields). This, depending on the external circuit configuration, affects the phase-frequency response of the system; and it can also result in extraneous coupling, such as the injection of power-supply noise into the signal chain. Consequently, the shield-winding currents should be considered in the process of determining where to connect the earth wire, or indeed, in deciding whether or not inter-winding capacitance might be preferable.

>>> To be continued . . . .

Topics for discussion:

External reluctance:
Design and Calculation of Induction Heating Coils. R M Baker. Trans. AIEE, Vol. 76, Mar. 1957. p31-40.
"A coil is electrically short when an appreciable portion of its ampere-turns is used to overcome the reluctance of the external flux path".

neutralisation
current transformers

30. Transformer losses
Hysteresis
Eddy current losses
Residual loss
Wire resistance "Copper loss", skin and proximity effects
Dielectric loss

xx. Multi-port transformers.

Current transformers
Overall transimpedance. Sampling of current for bridges or ammeters. Quadrature transformer.

short-circuit secondary. Measuring leakage inductance.
Shorted turns and screening cans.
"The Effective Inductance and Resistance of Screened Coils", A G Bogle, JIEE, 1940, 87, p299-.
Reprinted: Wireless Section, Proc. IEE, Vol 15, iss 45, Sept. 1940, p221-238.
"The short circuited turn", Thomas Roddam, WW March 1957 p114-117
"The short circuited screen" (Letters to the editor), Thomas Roddam, WW June 1957, p274.
"Effect of a Conducting Shield on the Inductance of an Air-Core Solenoid." Ted L Simpson. IEEE Trans. on Magnetics, Vol. 35, No.1, Jan 1999, p508-515.

Coupled resonators
Double-tuned bandpass (IF) transformer
coupled oscillators - a model for QM tunnelling, perturbation matrix

Hybrid transformers
Origin of the term. Compensation theorem. Impedance-generator hybrid
Trunk-line hybrid. Side-tone hybrid (why people shout into cellphones).

Neugebaur-Perrault neutralisation

Transmission-line transformers. Differential and common modes. Impedance transformation.