



PAPER • OPEN ACCESS

Clarifying concepts and gaining a deeper understanding of ideal transformers

To cite this article: Chiu-King Ng 2018 *Eur. J. Phys.* **39** 035205

View the [article online](#) for updates and enhancements.

You may also like

- [Notes on a Method of Testing Bars of Magnet Steel](#)
N W McLachlan
- [Magnetic flux in stacks of superconducting tapes of different architecture](#)
I M Harca, K Machaj, L Tomkow et al.
- [Modelling higher trapped fields by pulsed field magnetisation of composite bulk \$MgB_2\$ superconducting rings](#)
V Cientianni, M D Ainslie, H Fujishiro et al.

Clarifying concepts and gaining a deeper understanding of ideal transformers

Chiu-King Ng 

C C C Yenching College (retired), Hong Kong SAR, People's Republic of China

E-mail: ngchiuking@phy.hk

Received 22 September 2017, revised 22 January 2018

Accepted for publication 31 January 2018

Published 9 April 2018



CrossMark

Abstract

Even in ideal transformers, the input and output powers are never exactly equal, thereby causing the familiar ratio between the primary and secondary currents, namely $I_p \cdot I_s = N_s \cdot N_p$, to be slightly incorrect. In this paper, we explain why this is so and derive the correct ratio, as well as clarifying the related prevailing concepts. We conclude that a theory of an ideal transformer without a magnetising current is deficient and self-contradictory. Further, methods to locate the two black (phase) dots in a transformer symbol are elucidated. This paper is suitable for those who are pursuing a deeper understanding on this subject after learning the basics from the literature.

Keywords: ideal transformer, electromagnetic induction, mutual induction, magnetising current, dot convention transformer, equivalent circuit transformer

(Some figures may appear in colour only in the online journal)

1. Introduction

The transformer is conventionally considered to be a minor part of a general physics curriculum. A thousand-page college physics text devotes not more than three pages to the discussion of an ideal transformer. Other than doing simple calculations, students are usually quite confused by it or even misunderstand some key concepts. Sadly, many texts and web resources commit one or two basic conceptual errors on this subject. For example, some



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

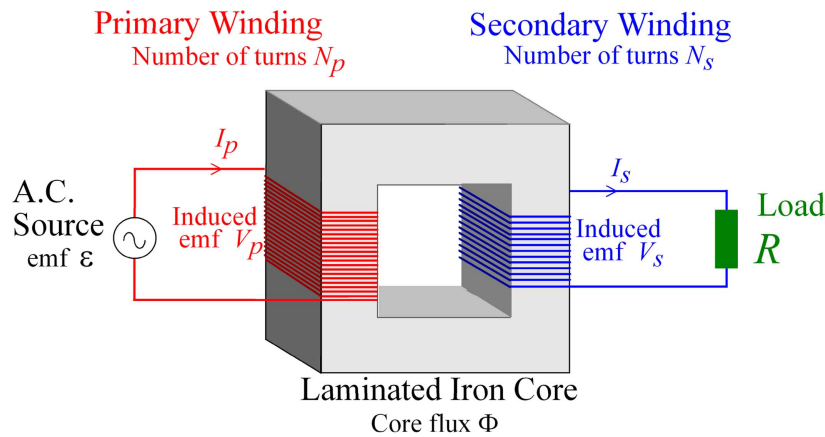


Figure 1. Schematic diagram of a single-phase transformer.

muddle up electromotive force with potential difference, thus writing an expression such as $I_p V_p = I_s V_s$ for input power equalling output power (see section 4).

In this paper, we attempt to provide clarifications, where needed, on the prevailing theory. Our results, which are derived from first principles as far as possible, reveal how important the so-called magnetising or exciting current is. In fact, this current is essential and indispensable, but it is discussed in only some of the commonly used texts [8, 9]. Some authors omit it, possibly because it is difficult for college students. Nonetheless, we try to present it in a way that is easier to grasp.

College-level knowledge of electromagnetism is a basic prerequisite to an understanding of this paper. Section 2 provides the basic knowledge, section 3 shows the incompleteness of the prevailing theory, and section 4 contains the core mathematical derivations. Most of the concepts are elaborated in sections 2–5 and 8. In section 6, we introduce and explain methods to position the two black (phase) dots in a transformer symbol. In section 7, an equivalent circuit of an ideal transformer is put forth.

2. The basics

A transformer is an electromagnetically coupled device consisting of, in a single-phase device, two unconnected windings wrapped around a laminated soft iron core. What is meant by an ‘ideal transformer’? The definitions and properties found in the literature are not completely the same [1–3]. We hope our results are as general as possible, so our analysis is based on as few assumptions as possible. Here, we define an ideal transformer as one which is absolutely free from any energy loss with only the following two conditions satisfied. (1) Both windings are purely inductive, i.e. they have *zero internal resistance*. The connecting wires have zero resistance too. (2) Both windings are perfectly coupled such that the same magnetic flux links them, i.e. there is *no flux leakage*. The transformer that serves as the model of our discussion is shown schematically in figure 1. One winding, called the primary, is connected to a current-independent alternating power source, forming the input circuit. The other winding, called the secondary, is connected to a load of pure resistance R , forming the output circuit.

2.1. Basic equations

Let us denote N_p and N_s as the number of turns, V_p and V_s as the induced electromotive forces, I_p and I_s as the currents, and Φ_p and Φ_s as the magnetic fluxes, where the subscripts ‘p’ and ‘s’ refer to the primary and secondary windings, respectively. Since we assume that the same flux interlinks the two windings, we define a common (mutual) flux $\Phi = \Phi_p = \Phi_s$.

Throughout this paper, ‘magnetic field’ refers to the $\mathbf{B}$ -field (magnetic induction), and the magnetic flux through a coil is defined as $\Phi = \mathbf{B} \cdot \mathbf{A}$, where $\mathbf{A}$ is the area vector of the coil (see section 5).

The physics of the transformer involves two basic laws. One is Faraday’s law of induction: the electromotive force (emf) induced across the ends of a coil of N turns is $V = -N d\Phi/dt$, and the other is Kirchhoff’s second law: in a closed loop of a circuit, the sum of the emfs is equal to the sum of potential differences, i.e. $\sum \text{emfs} = \sum \text{pds}$.

When looping around the input circuit, one will encounter the power source (emf ε) and the primary winding (emf V_p), so there are two emfs (ε and V_p) but no potential difference (the windings have zero resistance) in the input circuit. Hence,

$$\varepsilon + V_p = 0, \quad (1)$$

where

$$V_p = -N_p \frac{d\Phi}{dt}. \quad (2)$$

In the output loop, there is an induced emf in the secondary winding

$$V_s = -N_s \frac{d\Phi}{dt}, \quad (3)$$

and a potential difference across the load

$$V_s = I_s R. \quad (4)$$

Equation (1) can be written as $\varepsilon = -V_p$, signifying that the primary induced emf V_p is always equal and opposite to the source emf ε . When the transformer is operating, an AC current is drawn from the source to pass through the primary winding. In consequence, the magnetic flux through the primary winding produced by this current varies with time; an induced emf V_p across the primary winding is thus produced to counteract and balance the source emf ε (this cause–effect relationship is shown diagrammatically in figure 5).

One may doubt how a current could appear in a circuit that does not have a net emf to drive the current. Firstly, the appearance of this current does not violate Ohm’s law $V = Ir$, since when the net emf (V) and resistance (r) are both zero, in principle, current I can be any value. Secondly, the eventual possible I is that it can fulfill ‘ $\varepsilon = -V_p$ ’ after the processes just mentioned. Indeed, one objective of this paper is to answer the question ‘what is the primary current?’ This question is crucial to understanding a transformer. The primary current is derived in section 4.3.

In addition, if $I_p dt$ is multiplied to both sides of $\varepsilon = -V_p$, where dt is an infinitely short time interval, we get

$$\varepsilon I_p dt = -V_p I_p dt. \quad (5)$$

The minus sign in the last equation helps us to understand the instantaneous transfer of electrical energy between the source and the primary winding: the source *delivers* energy, and at the same time the primary winding *consumes* energy, and vice-versa.

2.2. Familiar ratios

Dividing equation (2) by (3), we obtain the voltage ratio

$$\frac{V_p}{V_s} = \frac{N_p}{N_s}. \quad (6)$$

In the ideal transformer, input power = power output, so

$$V_p I_p = V_s I_s, \quad (7)$$

where I_p is the primary current. Hence, $I_p/I_s = V_s/V_p$. In view of equation (6), we obtain the current ratio

$$\frac{I_p}{I_s} = \frac{N_s}{N_p}. \quad (8)$$

These two ratios, equations (6) and (8), are widely known to students, but one of them is problematic.

3. Something is wrong

Transformers work on alternating currents. Suppose now that the secondary current I_s is momentarily zero. At this moment, $I_p = 0$ (by the current ratio), and $V_s = 0$ (by $V_s = I_s R$). In addition, we have $d\Phi/dt = 0$ (by $V_s = -N_s d\Phi/dt$), and the core flux Φ , which is produced by the currents in the two windings, is zero as well because $I_p = I_s = 0$.

The satisfaction of $\Phi = 0$ and $d\Phi/dt = 0$ leads to the fact that, an infinitely short time later, Φ is still zero, regardless of what the source voltage may change to. There are two possibilities. The first one is that the two currents, I_p and I_s , continue to be zero. If so, the argument goes back to its starting point $I_s = 0$, and then repeats the deductions again. The argument loops infinitely, reaching an absurd conclusion that Φ is zero at all times. Currents cannot always be zero, so this never occurs. The other possibility is that the two currents are nonzero, but their individual magnetic fluxes cancel each other out. However, the current ratio asserts that the primary and secondary currents are in direct proportion: when one changes, the other changes by the same factor. It is known that magnetic flux is proportional to current, so if the primary and secondary fluxes are equal and opposite at a certain moment, they must always be related in this manner. Unfortunately, we come to the same conclusion: core flux Φ keeps vanishing although the transformer keeps running.

Of course, this is nonsense because a properly functioning transformer must rely on a time-changing magnetic flux. There must be something wrong.

4. Solving an ideal transformer

4.1. Powers

What is wrong? The fact is that, even if the transformer is ideal, the instantaneous input and output powers are not exactly equal. A not-so-correct current ratio has thus been derived. The fault is rooted in two errors.

Error 1: The input power is not $V_p I_p$, it should be εI_p , since the source of emf ε is actually that which delivers electrical energy to the transformer. The voltage $V_p = -N_p d\Phi_p/dt$ is an emf, therefore, by definition, the power $V_p I_p$ is positive (negative) when the primary gives (takes) electrical energy. As long as the primary receives electrical energy from

the source, $V_p I_p$ is negative, so the power $V_p I_p$ can only be understood as the negative of the input power. Some texts and online resources are vague or careless in this respect [3–5]. However, changing the power relationship to

$$\varepsilon I_p = V_s I_s \quad (9)$$

does not help much in averting the ‘nonsense conclusion’ in the previous section.

Error 2: The critical problem in equation (9) is that it misses a term corresponding to *the temporal fluctuation of the total core magnetic energy*. This term is necessary because the energy associated with the time-changing core flux must also be derived from the AC source. For this reason, the power balance must include such a term, for example,

$$\varepsilon I_p = V_s I_s + P_m, \quad (10)$$

where P_m is the rate of change of the total magnetic energy in the iron core.

Although some magnetic energy resides in the iron core, it is not dissipated via eddy currents, hysteresis loss, stray fields, etc, since the transformer is assumed to be ideal. The magnetic energy that is more than the average will return to the source at a later time. Therefore, we have

$$\langle P_m \rangle = 0, \quad (11)$$

and

$$\langle \varepsilon I_p \rangle = \langle V_s I_s \rangle, \quad (12)$$

where $\langle P_m \rangle$ is the power P_m averaged over a cycle, etc. Some texts state that the input and output powers are equal in their root-mean-square or averaged values, rather than being instantaneous, but little or inadequate explanation is provided alongside [6, 7].

4.2. Core flux energy

It is known that when a current I passes through an inductor of inductance L , the magnetic energy stored in L is $E = LI^2/2$ [12]. Furthermore, the inductance L is defined as $LI = N\Phi$, hence $E = N^2\Phi^2/2L$. Without exception, the magnetic energy stored in a transformer core must have the form $\eta\Phi^2/2$, where η is a constant depending on (N_p, L_p) or/and (N_s, L_s) . Later, we will ascertain what this constant actually is. The term P_m in equation (10) is the rate of change of the stored magnetic energy, so

$$P_m = \frac{\eta}{2} \frac{d\Phi^2}{dt} = \eta\Phi \frac{d\Phi}{dt}. \quad (13)$$

On the other hand, equations (1) and (2) allow us to express

$$\frac{d\Phi}{dt} = \frac{\varepsilon}{N_p}. \quad (14)$$

Before proceeding, we assume, without loss of generality, that the source emf ε varies with time t in the form

$$\varepsilon = \varepsilon_o \cos(\omega t), \quad (15)$$

where ε_o is the amplitude and ω is the angular frequency. By integrating equation (14) w.r.t. t with ε given in the last equation, we obtain

$$\Phi = \frac{\varepsilon_o}{\omega N_p} \sin(\omega t), \quad (16)$$

in which the integration constant is zero since Φ must vanish when the amplitude ε_o is zero. Substituting equations (14) and (16) into (13) and then P_m into equation (10), and after rearranging the terms, we get

$$\varepsilon \left[I_p - \left(\frac{\eta}{\omega N_p^2} \right) \varepsilon_o \sin(\omega t) \right] = V_s I_s. \quad (17)$$

Obviously, the current inside the square brackets in the last equation is identical to the primary current in equation (9). We denote this current as $I_{p,L}$ and the current proportional to $\sin(\omega t)$ as $I_{p,M}$, i.e.

$$I_{p,L} = I_p - I_{p,M}, \quad (18)$$

where

$$\varepsilon I_{p,L} = V_s I_s, \quad (19)$$

and

$$I_{p,M} = \left(\frac{\eta}{\omega N_p^2} \right) \varepsilon_o \sin(\omega t). \quad (20)$$

4.3. Currents and core flux

Equation (19) is written as $I_{p,L}/I_s = V_s/\varepsilon$ and is simplified by using $\varepsilon = -V_p$ and the voltage ratio, and we derive the *correct current ratio*,

$$I_{p,L} : I_s = -N_s : N_p. \quad (21)$$

It must be emphasised that this negative current ratio is consistent with the whole theory (see section 5). Among the college-level physics texts surveyed by the author, only one derives a similar negative current ratio [8].

The output is connected to resistor R , so $I_s = V_s/R = V_p(N_s/N_p)/R$ (by the voltage ratio). Because $V_p = -\varepsilon$, I_s is derived to be

$$I_s = -\frac{1}{R} \left(\frac{N_s}{N_p} \right) \varepsilon. \quad (22)$$

Next, $I_{p,L}$ is found by putting the last equation into equation (21),

$$I_{p,L} = \frac{1}{R} \left(\frac{N_s}{N_p} \right)^2 \varepsilon. \quad (23)$$

From the last two equations, we see that I_s and $I_{p,L}$ are both zero when the resistor R is set to infinity (open output circuit); in this case the output circuit is effectively removed and the transformer is reduced to a pure inductor, in which the primary current should lag behind the applied emf ε by $\pi/2$ with peak values (labelled with subscript 'o') related by $I_{po} = \varepsilon_o/X_p$, where X_p is the reactance of the primary winding [11]. By comparing equation (20) with (15), $I_{p,M}$ does lag behind ε by $\pi/2$ (identity $\sin(\omega t) = \cos(\omega t - \pi/2)$). Further, the amplitude of $I_{p,M}$ should be ε_o/X_p , so the factor $\eta/\omega N_p^2$ in equation (20) is now identified as $1/X_p$. Therefore, we obtain the final form of $I_{p,M}$,

$$I_{p,M} = \frac{\varepsilon_o}{X_p} \cos\left(\omega t - \frac{\pi}{2}\right). \quad (24)$$

Combining equations (16) and (24), we find

$$\Phi = \frac{L_p I_{p,M}}{N_p}, \quad (25)$$

where $L_p = X_p/\omega$ is the inductance of the primary winding. The significance of equation (25) lies in its assertion that the core flux Φ is *solely* produced by $I_{p,M}$.

With the form of ε given in equation (15), the complete primary current, $I_p = I_{p,L} + I_{p,M}$ can be expressed as

$$I_p = \frac{\varepsilon_o X_s}{R X_p} \cos(\omega t) + \frac{\varepsilon_o}{X_p} \cos\left(\omega t - \frac{\pi}{2}\right), \quad (26)$$

where X_s is the reactance of the secondary winding. In the last equation, the relation

$$(N_s : N_p)^2 = X_s : X_p \quad (27)$$

has been used [12]. By comparing the amplitudes of the three currents, $I_{p,L}$, $I_{p,M}$ and I_s , we get

$$(I_{p,L} : I_{p,M} : I_s)_{\text{Amplitude}} = 1 : R/X_s : N_p/N_s. \quad (28)$$

Furthermore, the two components, $I_{p,L}$ and $I_{p,M}$ in equation (26) can be combined mathematically to form

$$I_p = I_{po} \cos(\omega t - \varphi), \quad (29)$$

where

$$\varphi = \tan^{-1}\left(\frac{R}{X_s}\right) \quad (30)$$

is the phase angle by which I_p lags behind ε , and I_{po} is the Pythagorean addition of the amplitudes of $I_{p,L}$ and $I_{p,M}$. The equality of averaged powers $\langle \varepsilon I_p \rangle = \langle V_s I_s \rangle$ (equation (12)) can now be written as $\varepsilon_{\text{rms}}(I_p)_{\text{rms}} \cos \varphi = (V_s)_{\text{rms}}(I_s)_{\text{rms}}$ or $\varepsilon_{\text{rms}}(I_{p,L})_{\text{rms}} = (V_s)_{\text{rms}}(I_s)_{\text{rms}}$, where $\cos \varphi$ is known as the power factor [11] and ‘rms’ stands for the root-mean-square value. A formal mathematical approach to solving the ideal transformer is given in the [appendix](#).

At the end of this section, it is worth taking note of the distinctly different functions of the two primary currents.

- (1) $I_{p,L}$: The ‘L’ in the subscript means delivering energy to the load. It does not contribute to the core flux production (actually its flux is cancelled by that of the secondary current—see section 5). It accounts for the energy transfer between the energy source and load. This current is in phase with the source, so the input power $\varepsilon I_{p,L}$ is non-negative, matching with the output power $I_s^2 R$.
- (2) $I_{p,M}$: The ‘M’ in the subscript means magnetising the core. This current lags behind the source by $\pi/2$ and is responsible for producing the time-changing core flux Φ . It does not matter whether the output circuit is loaded with any R or is even unloaded, the flux Φ must be the same (since $\varepsilon = -V_p = N_p d\Phi/dt$). In other words, the current causing the flux, $I_{p,M}$, is *always present and independent of load R* . Because of the $\pi/2$ phase lag, $\langle P_m \rangle = \langle \varepsilon I_{p,M} \rangle = 0$.

5. The negative sign

The current ratio of $I_{p,L}$ to I_s (equation (21)) is negative—what does it mean? How does this negative sign eventually go to I_s (equation (22)), instead of $I_{p,L}$ (equation (23))? Undoubtedly, this negative sign means the opposite of a direction, but what is this direction? One text states that the negative current ratio implies that the two currents involved are 180° out of phase [8]. Do $I_{p,L}$ and I_s always flow in opposite directions? No, they may flow in the same or opposite directions, depending on how the two windings are coiled around and placed in the core (see figure 4). Actually, this negative sign is related to the *positive sense of rotations* (PSR) of the two windings.

Magnetic flux through a coil is defined as the scalar product of the magnetic field passing through the coil and the area vector of the coil. Whenever an area vector of a coil is chosen, for example, upward or downward if the coil is horizontally placed, subsequently its PSR is determined. The method is the right-hand screw rule: when the right thumb points in the direction of the area vector, the direction the other four fingers then curling around is the PSR of the coil. Indeed, the sign of the final result calculated from Faraday's law of induction is relative to the PSR of the coil [10].

In an ideal transformer, $\Phi_p = \Phi_s$ (the same flux along the core) is assumed, forcing the area vectors of the two windings to be connected and determined altogether as soon as one is chosen. For example, for transformers similar to the one shown in figure 1, the magnetic field directs along the iron core, so satisfying $\Phi_p = \Phi_s$, and the area vectors of the two windings must be opposite. But if the two windings are wound on the same side of the core, their area vectors are parallel.

The negative sign in the current ratio $I_{p,L} : I_s = -N_s : N_p$ is interpreted as follows: one of the two currents flows in the same direction as its PSR, and the other flows in the opposite direction to its PSR.

But it is not up to the primary's PSR to make the choice, it has already been pre-determined in the conventions of physics. In an inductor, we define $LI = N\Phi$, in which the PSR of the coil is chosen to be the same as the flowing direction of the positive current I around the coil (hence both I and Φ always have the same sign) for the sake of defining a positive inductance L . The primary winding of a no-load transformer is essentially an inductor, so the PSR of the primary winding must likewise be preset. Simply put, a positive (negative) current in the primary flows in the same (opposite) direction to the primary's PSR since the primary's PSR is so defined. This is the reason why the negative sign in the current ratio eventually goes to I_s . Unlike the primary's, the secondary's PSR cannot be established unless the relative position of the two windings is known. Nonetheless, using the PSR to determine a direction is somewhat complicated and surely not user-friendly. There is a much better method.

The two PSRs are established, satisfying $\Phi_p = \Phi_s$, so when both winding currents flow along their PSRs, the fluxes produced are of the same sign; put another way, if only one of the two currents does so, as the negative current ratio implies, their fluxes are opposite. Thus, the negative sign in the current ratio suggests an alternative interpretation that seems to be more understandable and utilisable: *in the iron core, the magnetic fluxes (or magnetic fields since they are relative to the same area vector at the same place) produced by the two currents are always opposite*. This can serve as a simple method to find the direction of the secondary current (see section 6.2).

Further, the magnetic field inside a solenoid is proportional to the product of the number of turns and the current. The negative current ratio can be rewritten as $N_p I_{p,L} = -N_s I_s$, implying that the magnitudes of these two opposite fields are equal. They add up to exactly

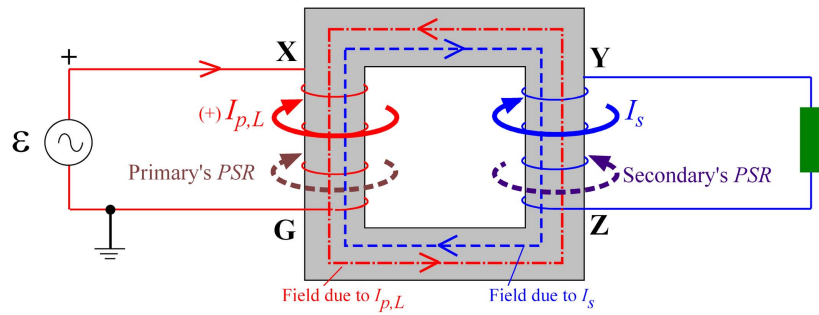


Figure 2. An ideal transformer with an emf ε applied to its primary winding. Using this transformer, we explain the methods used to find the direction of the secondary current and the positions of the black dots.

zero, a scenario that agrees and confirms our earlier findings. The magnetising (load-independent) current $I_{p,M}$ produces an always-adequate core flux Φ while those produced by $I_{p,L}$ and I_s cancel each other out.

6. Black dots

In circuit diagrams, occasionally two black dots are seen next to the terminals of a transformer symbol. One is at the primary, and the other is at the secondary. They are also called phase dots, indicating which input and output terminals are simultaneously positive or negative w.r.t. to the other terminal of the same winding [2]. There are three methods to ascertain where the two black dots should be placed.

Let us consider the transformer shown in figure 2. Suppose, at a certain instant, that the source emf ε is positive, then the terminal X is therefore positive w.r.t. the earthed terminal G. In the figure, the PSRs of the two windings, the magnetic fields produced by the two currents $I_{p,L}$ and I_s at this moment are all shown. With reference to figure 2, the three methods are explained one-by-one in the following. The underlying physics of the first two has been elaborated in the previous section, so only their steps with brief reasons are stated.

6.1. By negative current ratio

$+\varepsilon \rightarrow X$ is positive $\rightarrow$ direction of $+I_{p,L}$ (in phase with ε) $\rightarrow$ primary's PSR (the direction positive primary current flows) $\rightarrow$ secondary's PSR (by 'same flux along core') $\rightarrow$ direction of I_s (by negative current ratio: $I_{p,L}$ now flows along its PSR, so I_s does not) $\rightarrow Y$ is positive (the terminal where current leaves from a seat of emf) $\rightarrow$ place black dots at (X, Y) or (G, Z).

6.2. By flux opposition

$+\varepsilon \rightarrow X$ is positive $\rightarrow$ direction of $+I_{p,L}$ (in phase with ε) $\rightarrow$ magnetic field due to $I_{p,L}$ (by right-hand grip rule) $\rightarrow$ magnetic field due to I_s (must be opposite to that of $I_{p,L}$) $\rightarrow I_s$ (by right-hand grip rule) $\rightarrow Y$ is positive (the terminal where current leaves from a seat of emf) $\rightarrow$ place black dots at (X, Y) or (G, Z).

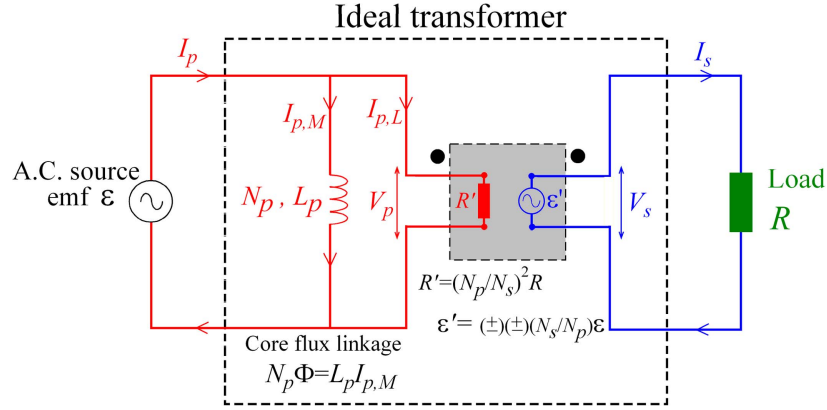


Figure 3. Equivalent circuit of an ideal transformer.

6.3. By positive voltage ratio

The third method is to use the in-phase relationship between V_p and V_s (the positive voltage ratio). This relationship implies their two corresponding electric fields $\mathbf{E}_p$ and $\mathbf{E}_s$, where $\oint \mathbf{E}_p \cdot d\mathbf{x} = V_p$ and $\oint \mathbf{E}_s \cdot d\mathbf{x} = V_s$ point in the same direction relative to their own PSR. That is to say, the induced currents (if any) produced by V_p and V_s would flow in the same direction relative to their own PSR. To avoid confusion, one should be clear that I_s is the induced current of V_s , but neither $I_{p,L}$ nor $I_{p,M}$ is the induced current of V_p . The emf V_p has been cancelled by the source ε , so the induced current of V_p never exists.

To use this method, the two connected PSRs are established by using the condition ‘same flux along core’ first. By taking two assumed currents following these two PSRs, the in-phase terminals, for example, those from which the two currents leave are thus determined.

7. Equivalent circuit

Based on the prevailing models [1, 3], we devise and put forth an equivalent circuit of an ideal transformer. Many major ideas discussed in this paper become obvious with the aid of this equivalent circuit.

In figure 3, the part inside the dashed line box is the ideal transformer. The primary circuit is made up of a parallel-combination of a pure inductor and a pure resistor. The former has a (presumably large) inductance L_p and number of turns N_p while the latter has a resistance

$$R' = (N_p/N_s)^2 R, \quad (31)$$

which is also known as the equivalent resistance of the load R [9]. The currents passing through these two branches are $I_{p,M}$ and $I_{p,L}$, respectively; so their sum gives the primary current I_p . The equivalent resistor R' obeys Ohm’s law, so the current $I_{p,L}$ is always in phase with the source ε , but the action of the inductor L_p introduces a phase lag of $\pi/2$ in $I_{p,M}$. The input voltage V_p and current $I_{p,L}$ are related by $V_p = I_{p,L} R'$. The connection of the inductor in the circuit is because $I_{p,M}$ solely produces the core flux linkage $N_p \Phi = L_p I_{p,M}$ with a stored magnetic energy $E = N_p^2 \Phi^2 / 2L_p$. Since R' and L_p are connected in parallel, they do not affect each other. Obviously, no matter whether the transformer is loaded with any R or is even unloaded, $I_{p,M}$ is always present and the same.

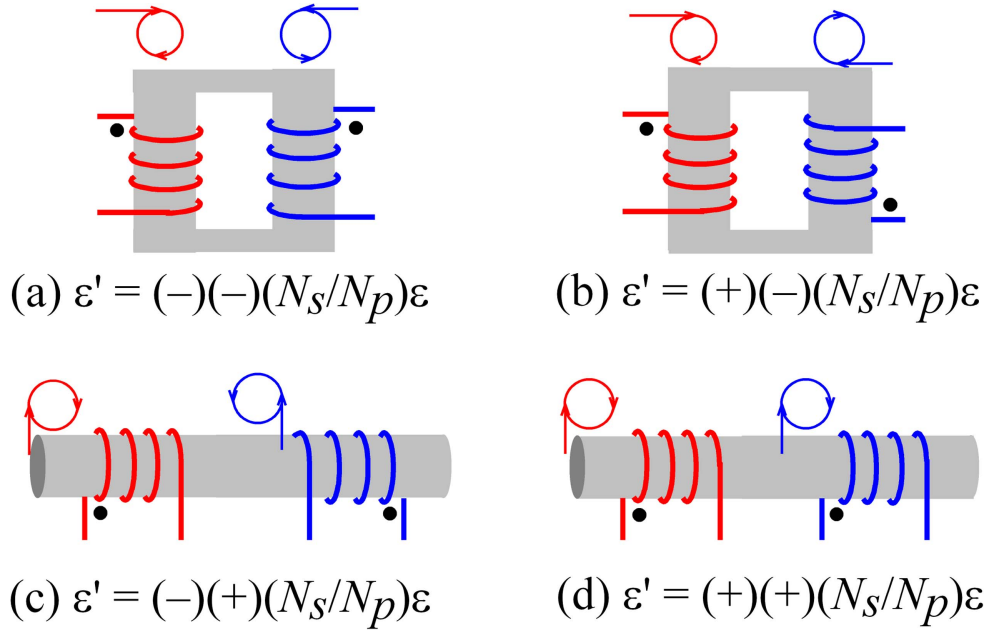


Figure 4. Examples of the determination of the signs in equation (32). The arrows around the circles show the coiling directions of the two windings.

In the equivalent circuit, the secondary side is a separated circuit driven by a source of emf,

$$\varepsilon' = (\pm)(\pm)(N_s/N_p)\varepsilon, \quad (32)$$

where the two signs are determined as follows: (1) the first $(\pm)$ takes the positive (negative) sign when the coils are wound in the same (opposite) sense, and (2) the second $(\pm)$ takes the positive (negative) sign when the coils are wound on the same (opposite) side of the core. Figure 4 provides examples of the four cases. Whenever the overall sign is known, the two black (phase) dots can be placed accordingly, as shown in figure 4. The direction of the output current, which is in phase with ε' , agrees with that found by the methods discussed in section 6. In this equivalent circuit, the PSRs are no longer needed.

When the output circuit is open, R is removed; R' in the input circuit is removed too. Then, $I_{p,L} = I_s = 0$ (in agreement with $N_p I_{p,L} = -N_s I_s$) and $I_p = I_{p,M}$, so the magnetising reactance X_p can be determined experimentally by measuring the values of $(V_p)_{rms}$ and $(I_p)_{rms}$ by AC meters under a no-load condition and using the relation $X_p = (V_p)_{rms}/(I_p)_{rms}$.

The grey area in figure 3 is a magic black box, which somehow absorbs the power used up in R' and creates the same electrical power output in ε' continuously; the values of R' and ε' ensure this occurs all the time.

8. Discussion and summary

First, let us review some of the main findings in this paper. In section 3, by using a recursive argument, we show that a theory formulated on the basis of ' $V_p I_p = V_s I_s$ ' is incomplete and paradoxical (the core flux always vanishes). Unfortunately, this theory is a popular one and is

commonly taught in schools. In section 4, we clarify the concepts: the input power should be εI_p , where ε is the emf of the power source, and, above all, the power relationship should be $\varepsilon I_p = V_s I_s + P_m$, where P_m is the rate of change of the total magnetic energy in the iron core. Starting from this correct power relationship, the ideal transformer is completely solved. The theory thus built depicts a self-contained and comprehensive picture. The main point is that there is always a magnetising current ($I_{p,M}$) flowing in the primary winding to excite the core flux; only this current does this job. Irrespective of what the load R is, the core flux is always present and the same, and is $I_{p,M}$. This explains why, in the argument in section 3, a zero-flux results if $I_{p,M}$ is not taken into account. Another noteworthy result is the negative current ratio $I_{p,L} : I_s = -N_s : N_p$ (equation (21)), where $I_{p,L}$ is the primary current other than $I_{p,M}$. In section 5, we explain how the negative sign of this current ratio can be understood in terms of the PSR of the two windings. Further, this negative current ratio implies that the magnetic fluxes produced by $I_{p,L}$ and I_s cancel each other out, confirming that the core flux is solely produced by $I_{p,M}$. The two primary currents $I_{p,M}$ and $I_{p,L}$ play their own unique roles: the former magnetises the core and the latter gains electrical energy from the source to the load.

Texts and online resources are used to supplement a condition in the definition of an ideal transformer, for example, ‘negligible magnetising current is required to set up the flux’ [1], ‘the primary and secondary coils have infinite self-inductances’ [2], or ‘infinitely high core magnetic permeability’ [3]. The inductance of a coil wrapped around an iron core is proportional to the permeability (μ) of the core material [12], so the ‘core of high permeability’ is synonymous with the ‘core of large inductance’. Simply put, all these assumptions try to legitimise the fact that the magnetising current in the theory is ignored. In the following, we focus on discussing whether the magnetising current is truly negligible.

First of all, we need to make clear under what condition(s) the magnetising current $I_{p,M}$ is supposedly negligible. From equation (26), we can see that the amplitude of $I_{p,L}$ is $(\varepsilon_o/R)(X_s/X_p)$, while that of $I_{p,M}$ is ε_o/X_p . When their common denominator X_p is cancelled out, it is obvious the condition $R \ll X_s$ makes $I_{p,L}$ itself too large to outweigh $I_{p,M}$. On the other hand, a larger X_p , the denominator of $I_{p,M}$, will make $I_{p,M}$ itself smaller. When all these factors are considered at once, the most sufficient condition for the validity of $I_{p,M} \ll I_{p,L}$ is $R \ll \text{winding reactance}$, where the winding reactance ($X = \omega L$) refers to the minimum of X_p and X_s .

Does the condition ‘ $R \ll \text{winding reactance}$ ’ imply that the magnetising current is very small and hence justifiably negligible? We have to be very cautious in thinking about this question. Firstly, the condition ‘ $R \ll \text{winding reactance}$ ’ only suggests that $I_{p,M} \ll I_{p,L}$, not necessarily that any effect produced by $I_{p,M}$ is small. Secondly, for $I_{p,M} \ll I_{p,L}$, we can only say $I_{p,M}$ is negligible when both $I_{p,M}$ and $I_{p,L}$ appear and are compared in the same expression. For example, $I_{p,M} \ll I_{p,L}$ justifies the complete primary current $I_p = I_{p,L} + I_{p,M} \approx I_{p,L}$ and the approximate correctness of the famous current ratio $I_p : I_s = N_s : N_p$ (apart from the negative sign). However, the core flux linkage $N_p \Phi = L_p I_{p,M}$ is irrelevant to $I_{p,L}$; it is wholly produced by $I_{p,M}$. Although under the condition $I_{p,M} \ll I_{p,L}$, we do not see any reason why $I_{p,M}$ is negligible here, even if $I_{p,M}$ is assumed to be zero (when L_p is assumed to be infinite) because the core flux linkage is the product of $I_{p,M}$ and L_p . In short, the condition ‘large winding reactance’ or ‘ $R \ll \text{winding reactance}$ ’ does not mean that it is legitimate to neglect $I_{p,M}$ anywhere in the theory.

Nonetheless, the smaller the magnetising current is, the more ‘ideal’ the transformer is. Unlike the always-the-same core flux Φ , the stored magnetic energy in core E diminishes when $I_{p,M}$ is made to be very small since $E = N_p \Phi I_{p,M} / 2$ [12]. A zero (very small) stored magnetic energy entitles the transformer to become more apt to be ‘energy lossless’ because it has no (little) magnetic energy to lose.

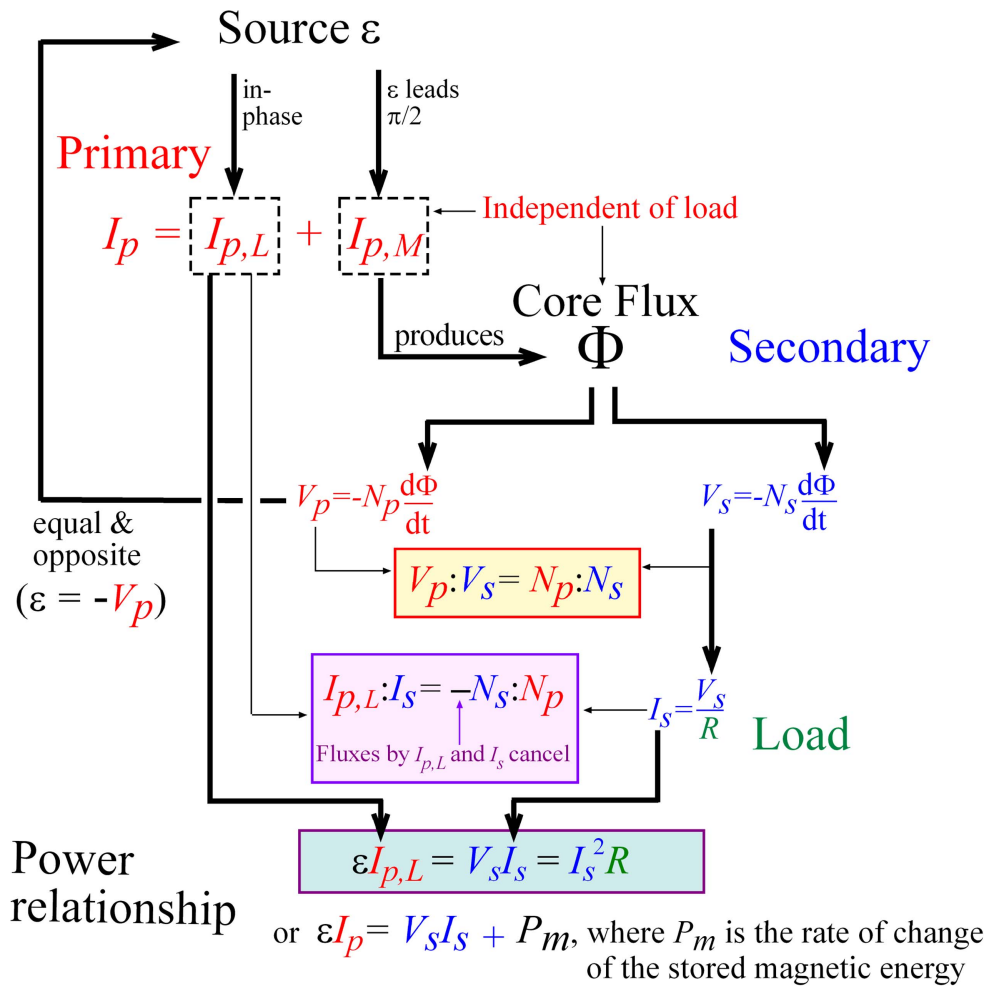


Figure 5. Essential concepts in the physics of an ideal transformer.

We are faced with a dilemma. A magnetising current excites the core flux, it is absolutely indispensable; without it, the theory is an intrinsically deficient and self-contradictory zero-flux theory, but a zero magnetising current does make the transformer more apt to be ideal. We advocate that the resolution is either (1) to assume a very small but non-zero magnetising current, or (2) to neglect the magnetising current in the equations but not in the concepts. The first option would be better in general, while the second one would be especially appropriate for introductory courses: the well-known equations ' $V_p/V_s = N_p/N_s$, $I_p/I_s = N_s/N_p$, ...' are taught as usual, but with an additional remark such as 'the core flux is created neither by I_p nor I_s ; it is created by another primary current called the magnetising current ...'. If this notion is told prior to section 3, the conclusion drawn there will no longer be nonsensical. Our advice would be that care should be taken in interpreting an assumption such as 'negligible magnetising current' or 'coils of infinite inductances' in a definition of an ideal transformer.

Finally, to facilitate a quick review, most of the relationships discussed in this paper are shown collectively in figure 5 with arrows showing their causalities.

Acknowledgments

The author would like to thank the anonymous referees for their valuable comments and very useful suggestions regarding the manuscript.

Appendix

The mathematics for solving an ideal transformer is outlined below. Considering equations (1)–(4), and expressing the mutual flux $\Phi_p = \Phi_s = \alpha N_p I_p + \beta N_s I_s$, where α and β are positive constants depending on the geometrical factors of the two windings, we obtain two coupled ODEs:

$$\varepsilon - \alpha N_p^2 \frac{dI_p}{dt} - \beta N_p N_s \frac{dI_s}{dt} = 0, \quad (\text{A1})$$

and

$$-\alpha N_p N_s \frac{dI_p}{dt} - \beta N_s^2 \frac{dI_s}{dt} = I_s R. \quad (\text{A2})$$

Set $N_s = 0$, effectively removing the secondary winding, in which case the transformer becomes a pure inductor. Compared with the equation, $\varepsilon - L_p dI_p/dt = 0$, where L_p is the inductance of the primary coil, the term αN_p^2 is identified as L_p . By symmetry, $\beta N_s^2 = L_s$. Hence, the two ODEs become

$$\varepsilon - L_p \frac{dI_p}{dt} - n L_s \frac{dI_s}{dt} = 0, \quad (\text{A3})$$

and

$$-\left(\frac{L_p}{n}\right) \frac{dI_p}{dt} - L_s \frac{dI_s}{dt} = I_s R, \quad (\text{A4})$$

where $n = N_p/N_s$. Suppose $\varepsilon = \varepsilon_o \exp(-i\omega t)$, $I_p = I_{po} \exp(-i\omega t)$ and $I_s = I_{so} \exp(-i\omega t)$, where $i = \sqrt{-1}$, ε_o is a real amplitude, and I_{po} and I_{so} are complex amplitudes. It is straightforward to solve the unknowns, for example, we get

$$I_{po} = \frac{(X_s + iR)\varepsilon_o}{RX_p}, \quad (\text{A5})$$

where $X_p = \omega L_p$ and $X_s = \omega L_s$. After some algebra, I_p is proven to be exactly the same as equation (26).

ORCID iDs

Chiu-King Ng  <https://orcid.org/0000-0002-1290-1039>

References

- [1] Daniels A R 1976 *Introduction to Electrical Machines* (London: Macmillan) p 43
- [2] Alexander C K and Sadiku M 2013 *Fundamentals of Electric Circuits* 5th edn (New York: McGraw-Hill) p 574
- [3] Wikipedia 2017 Transformers <https://en.wikipedia.org/wiki/Transformer> (Accessed: 7 February 2018)

- [4] Giancoli D C 2005 *Physics: Principles with Applications* 6th edn (Upper Saddle River, NJ: Pearson) pp 595–6
- [5] Serway R A and Vuille C 2017 *College Physics* 11th edn (Boston, MA: Brooks/Cole) pp 701–2
- [6] Knight R D, Jones B and Field S 2015 *College Physics: A Strategic Approach* 3rd edn (Upper Saddle River, NJ: Pearson) pp 844–5
- [7] Cutnell J D and Johnson K W 2012 *Physics* 9th edn (Hoboken, NJ: : Wiley) pp 700–1
- [8] Tipler P A 1999 *Physics for Scientists and Engineers* vol II 4th edn (New York: Freeman) pp 982–3
- [9] Halliday D, Resnick R and Walker J 2014 *Principles of Physics* 10th edn (Hoboken, NJ: : Wiley) pp 838–40
- [10] Jones C 2003 Understanding and using the minus sign in Faraday’s law *Phys. Educ.* **38** 526–30
- [11] Herman S L 2011 *Alternating Current Fundamentals* 8th edn (Clifton Park, NY: Delmar)
- [12] Lorrain P and Corson D 1970 *Electromagnetic Fields and Waves* 2nd edn (New York: Freeman) pp 345–56