11 A Short Introduction to Electromagnetism

v1.9 – June 2021

Prerequisite knowledge required: Newtonian Physics, Electric and Magnetic Fields

11.1 Introduction

The word “electromagnetism” contains elements of the words “electric” and “magnetic” pushed together. From this, you might assume that there is some relationship between electric and magnetic fields. You would be right. The subject of electromagnetism is all about these relationships and their consequences.

In terms of analogue electronics, the most important consequences of this relationship are that Kirchhoff’s voltage law doesn’t work with changing magnetic fields going through a circuit, and that this same effect can be harnessed to design useful components such as inductors and transformers, as well as pumps, motors and the dynamos used to generate electricity.

I’ll start this note with a brief reminder of the key results from electric field and magnetic field theory with particular relevance to electronic components (in particular capacitors and inductors), and then go on to describe the experiments which suggested that there was a relationship between electric and magnetic fields, and the theories which account for these relationships. The consequences of these theories is a vast subject, so I’ll just mention a few of the key results here.

11.2 Electric and magnetic fields: a quick summary

The important results (see the chapter on “Electric and Magnetic Fields” for more details) are:

- The electric field strength (*E*) at a point in space is the ratio of the force (*F*) on a small positive charge at that point in space to the charge (*q*) on that small charge:

$$E = \frac{F}{q} \tag{11.1}$$

- The force on a positive charge acts to repel it from another positive charge (like charges repel, unlike charges attract).
- The electric field strength due to charge of q_1 Coulombs a distance *r* away is given by:

$$E = \frac{q_1}{4\pi\varepsilon_0 r^2} \tag{11.2}$$

- The electric potential at a point in space is the amount of energy it takes to move a unit charge from an infinite distance to that point. The electric potential a distance *r* from a charge q_1 Coulombs is given by:

$$V = \frac{q_1}{4\pi\varepsilon_0 r} \tag{11.3}$$

- The potential difference between two points A and B is the amount of energy required to move a unit of charge from A to B.
- The energy stored in an electric field in free space per unit volume is:

$$\text{Energy per unit volume } \left(J/m^3\right) = \frac{\varepsilon_0 E^2}{2} \qquad (11.4)$$

- The energy stored in a magnetic field in free space per unit volume is:

$$Energy/m^3 = \frac{B^2}{2\mu_0} \qquad (11.5)$$

OK, on with the story…

11.3 Ørsted and the link between electricity and magnetism

In 1820, Hans-Christian Ørsted noticed that switching a current on and off caused a compass needle to deflect[1]. This was surprising: at the time it wasn't widely known that there was any connection between electricity and magnetism.

Once the discovery of a connection was made, Ørsted investigated further, and after a series of experiments deduced that a current flowing through a wire produces a circular magnetic field around it.

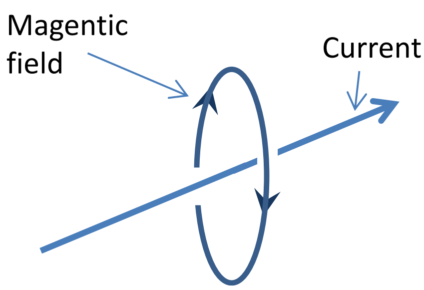

Figure 11.1 Direction of magnetic field around a wire carrying a current

The magnetic field at a point is produced in a direction perpendicular to both the direction the current is flowing in, and the line between the small element of wire and the point. (The direction of the field is given by the "right-hand screw rule": make a "thumbs-up" gesture with your right hand, and point the thumb in the direction of the current flow. The magnetic field produced occurs in the direction of the fingers.)

It was later scientists who quantified this relationship in a famous result known as the Biot-Savart equation. This states any small length of wire δL which carries a current produces a magnetic field around it of magnitude:

[1] It is possible that he was attempting to investigate earlier reports of a connection between electricity and magnetism made by the Italian Gian Domenico Romagnosi. However, the name Ørsted is more often associated with the discovery of the connection between electricity and magnetism.

$$\delta H = \frac{\mu_0}{4\pi}\frac{I}{r^2}\sin\theta\,\delta L \qquad (11.6)$$

where *I* is the current flowing in the wire, *r* is the distance from the wire to the point at which the magnetic field is measured, θ is the angle between the wire and the direction to where the magnetic field is being measured (see Figure 11.2) and μ_0 is the permeability of free space.

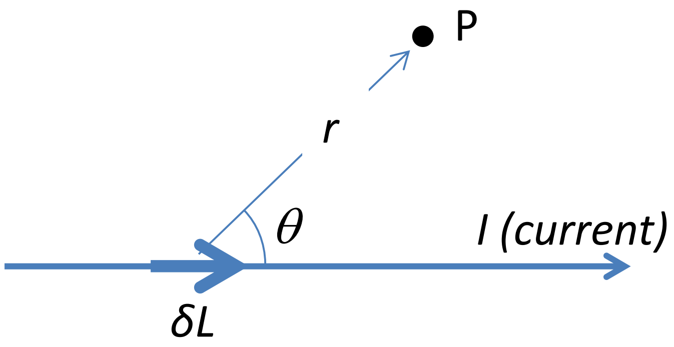

Figure 11.2 Illustration of Biot-Savart equation. In this case the magnetic field at P caused by the current *I* flowing through the wire element *δL* would be coming straight up out of the paper.

11.3.1 B-fields and H-fields

Observant readers with good memories might be wondering why I've used the letter 'H' to represent the magnetic field in equation (11.6) whereas I used the letter 'B' in the chapter about electric and magnetic fields (and in equation (11.5) in this chapter).

Strictly speaking the two are different. Rather confusingly, it's common to refer to both quantities as the "magnetic field", and while there have been several attempts to find a different name for one or the other to avoid the confusion, none of the alternatives has stuck. It's now more common to talk about "the B-field" and "the H-field" when it's important to differentiate between them.

You can perhaps think of the H-field as representing the amount of "magnetic field generating force" that is present in a certain location due to the current flowing in wires. It might not be equal or even linearly related to the total magnetic field that is actually there (the B-field) due to the phenomenon of *magnetisation*.

Warning: this can get a little complicated, mostly due to some rather confusing standard definitions. In free space, if the current is trying to produce a magnetic field, then there's nothing to affect the field produced, and the magnetic field that is actually there (the B-field) is linearly related to the H-field by:

$$B = \mu_0 H \qquad (11.7)$$

Why introduce the constant term μ_0? It's just a matter of convention, and the units which are always used for the B-field (teslas, or T) and for the H-field (ampères per meter, or A/m). μ_0 is known as the permeability[2] of free space, and it's equal to $4\pi\ 10^{-7}$ henrys per meter (H/m). (Henrys are the units of inductance – see later in this chapter for more about inductors).

As mentioned, the B-field is the amount of magnetic field which is actually there. This can be different from the expected $B = \mu_0 H$, since it's possible that there might be some material which makes it harder or easier to produce a magnetic field inside it. Most materials have almost no effect whatsoever, but there are materials which do interact with magnetic fields and which make a significant difference. For the purposes of electronics, the most common material that has a large effect on the B-field present is a piece of *ferromagnetic* material.

You can think of ferromagnetic material as having a lot of small permanent magnetics inside it (called *domains*), all of which can rotate around.

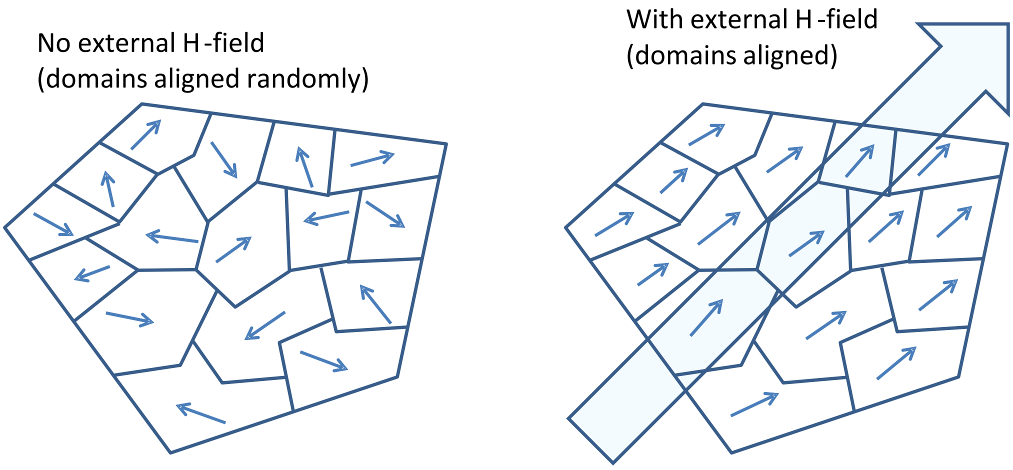

Figure 11.3 Aligning the domains in ferromagnetic material

If you put a bit of ferromagnetic material in a magnetic field, the small permanent magnets in the material will try and align with the external magnetic field. Interestingly, they don't do this like compasses: a compass would tend to move its south pole towards the north pole of the external field, and that would result in a net decrease of the magnetic field inside the material. (There are materials which do this, they are called diamagnetic materials, but they are not the most interesting case for electronics). Ferromagnetic materials have the curious property that they align their domains the other way, resulting in a net increase in the magnetic field in the material: the actual magnetic field present (the B-field) is now made up of the original external magnetic field (the H-field) plus the contribution of all the small magnets in the ferromagnetic material (known as the magnetisation field, or M-field) which act in the same direction.

We have to modify equation (11.7) to become:

$$B = \mu_0 \left(H + M \right) \qquad (11.8)$$

[2] You can think of permeability as a measure of how easy it is for an H-field to produce a B-field at a certain point in space. It's a function of what is there (in free space, there's nothing there at all).

For ferromagnetic materials M is positive, so the B-field is larger than it would be in free space (indeed this magnification of the magnetic field is the main reason we use ferromagnetic materials in electronics).

In cases where the M-field is proportional to the H-field[3], we can define a relative permeability as:

$$\mu_r = 1 + \frac{M}{H} \tag{11.9}$$

which allows us to rewrite equation (11.8) as:

$$B = \mu_r \mu_0 H \tag{11.10}$$

Almost all materials have a relative permeability that is approximately equal to one; however common ferromagnetic materials (such as iron or nickel) have relative permeabilities measured in the thousands[4], and superconductors, which have relative permeabilities of zero[5].

11.3.2 Magnetic fields in a wire loop

The flux lines of magnetic fields around a long, straight wire occur in loops around the wire. However, what if the wire of interest is not long and straight, but is instead a tightly-packed spiral[6]?

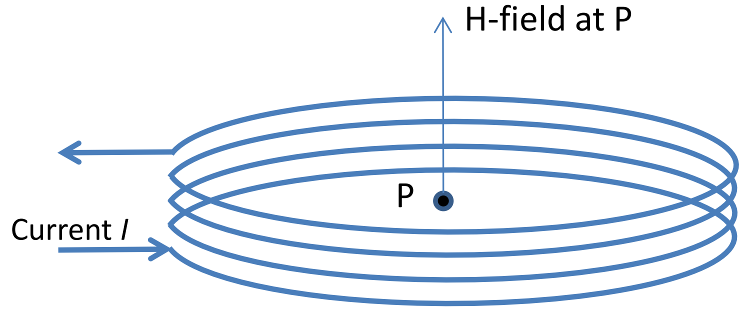

Figure 11.4 Tightly-packet loop of wire with centre at P

We can use the Biot-Savart equation to deduce the magnetic field in the centre of the loop. If there is a current of I amps flowing in the wire, then for each small length of wire δL, the Biot-Savart equation predicts that there will be a small magnetic field produced of magnitude:

---

[3] I should note at this point that this isn't a common situation (at best it's not a very good approximation). Many magnetic materials are non-linear, and the M-field they generate is not proportional to the H-field that is present at the time. Some of them even have memory, and produce an M-field which is a function of what the H-field was some time ago. This makes analysis difficult, which is one of the reasons that electronic components using these materials are difficult to work with in practice.

[4] The exact value is a strong function of how pure the material is.

[5] With a relative permeability of zero, no matter how much H-field is present there will be no B-field at all inside the superconductor.

[6] By "tightly-packed" here, I mean that the radius of the loop is large compared to the height of the loop, so that a point in the centre of the spiral is effectively in the middle of all of the loops of wire in the spiral.

$$\delta H = \frac{\mu_0}{4\pi}\frac{I}{r^2}\delta L \tag{11.11}$$

generated in the exact centre of the loop, acting in the direction up through the loop if the current is flowing anti-clockwise. A moments thought (and/or some tortuous contortions of your hand) should lead you to the conclusion that the same magnetic field will be generated for every small length of wire dL in the loop, and since there is a total length of $2\pi r$ around a loop of radius r, the total magnetic field is the centre of the loop is:

$$H = \frac{\mu_0}{4\pi}\frac{I}{r^2}2\pi r = \frac{\mu_0}{2r}I \tag{11.12}$$

If there are N turns of wire in the spiral, each of them will be contributing the same amount to the total magnetic field, resulting in a field at the centre of the loop of magnitude:

$$H = \frac{\mu_0}{2r}NI \tag{11.13}$$

That's as far as we need to get for now.

11.4 Faraday and electromagnetic induction

Since it had been discovered that electricity (in the form of a current) could produce a magnetic field, it perhaps wasn't too difficult to imagine that perhaps magnetic fields could produce some electric effects. However, it was eleven years later that Michael Faraday famously dropped a magnet through a loop of wire and observed a pulse of current flowing through the wire.

Faraday, like any good scientist, immediately began investigating and quantifying the results. After a whole series of experiments with different magnets and different loops of wire, he deduced his *law of electromagnetic induction*: that an induced electromotive force (similar to a potential difference[7]) occurs in a closed circuit equal to the rate of change of the magnetic flux enclosed by the circuit.

The law can be represented in an equation:

$$\varepsilon = -\frac{d\Phi_B}{dt} \tag{11.14}$$

where ε is the electromotive force (emf) around the circuit (measured in volts), and Φ_B is the magnetic flux through it (measured in volt-seconds, or sometimes Webers (Wb)). Magnetic flux is the product of the field strength (this time the B-field since it's the actual strength of the magnetic field that's important) multiplied by the area A:

[7] I can't call it an "induced potential difference" since strictly speaking it's not a difference in a potential. (Kirchhoff's voltage law tells us that the sum of the potential differences around a closed circuit is zero.) However electromotive force is also measured in volts, and causes current to move through resistors in the same way as a potential difference does. You can think of "electromotive force" as any force which causes charges (in this case electrons) to move in a conductor. This includes forces due to an external electric field (which does imply an electric potential difference), but it also includes forces due to a changing magnetic field (which doesn't).

$$\Phi_B = B \times A \tag{11.15}$$

The negative sign is known as *Lenz's law* and states that an increase in magnetic flux through an area causes a negative EMF around the area's boundary, although to appreciate why the negative sign is here you have to know the convention for giving a direction to an area enclosed by a circumference[8].

For example, consider that tightly-packed coil of wire again:

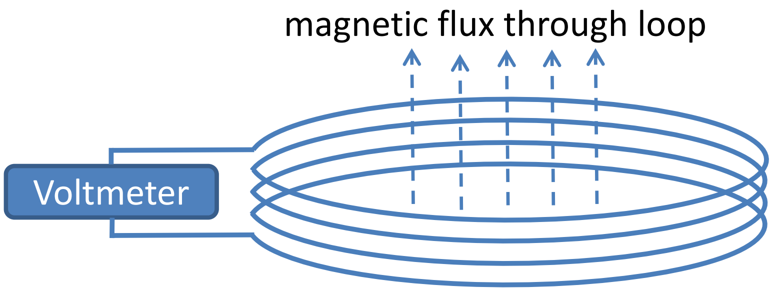

Figure 11.5 Changing the magnetic field through a loop of wire

If the magnetic field through the coil changes, then an electromotive force given by Faraday's equation appears around each loop of the wire. With N turns of wire, the total EMF between one end of the wire and the other is given by:

$$\varepsilon = -NA\frac{dB}{dt} \tag{11.16}$$

and this will cause currents to flow in any circuit containing this coil.

11.5 Inductors

You might have noticed at this point that we have an equation describing the magnetic field produced by a current flowing through a spiral of wire, and another equation describing the effect of that magnetic field producing an electromotive force from one end of the looped wire to the other.

They appear to predict that a changing current through the spiral would produce changing magnetic field within the spiral, which would in turn produce a voltage across the terminals. In some ways this is like a resistance: the current through the component is associated with a voltage across the terminals of the component.

However, there are some key differences: an unchanging current would produce an unchanging magnetic field, which according to equation (11.16) would not produce any voltage at all, since a changing magnetic field is required to produce the voltage. Also, the faster the current is changing, the greater the rate of change of the magnetic field, and the greater the voltage that is generated. We seem to have some sort of resistance here which is a function of frequency: the faster the current is changing, the greater the voltage that appears across the component.

---

[8] Put your right hand in a "thumbs-up" gesture, with the fingers in the chosen direction around the circumference of the area. The direction your thumb is pointing is the direction of the area.

We can derive a formula for what is happening; all we need is some way to relate the magnetic flux Φ_B in Faraday's equation to the magnetic field *B* in the Biot-Savart equation. Unfortunately this requires working out the magnetic field everywhere in the loop (not just in the middle), which makes it a rather difficult calculation. However, the result does indeed produce something of the form:

$$\varepsilon \propto -\frac{dI}{dt} \tag{11.17}$$

which says that the potential induced is proportional to the rate of change of the current, and acts in the opposite direction to the current. The larger the rate of change of current, the larger the voltage produced.

These turn out to be very useful circuit components: they're called *inductors*, with their value (their *inductance*) given by the constant of proportionality in equation (11.17). So we could write:

$$\varepsilon = -L\frac{dI}{dt} \tag{11.18}$$

where *L* is the inductance is measured in Henrys (H).

Note that the EMF tends to act in a direction to try and limit the rate of change of current: an increasing current will produce a negative EMF which will reduce the voltage difference in the circuit causing the current to flow.

11.5.1 Inductors in circuit theory

Lenz's law can be a little confusing if you are trying to derive the performance of inductors in circuits. Due to the way in which the EMF is defined, it's in the reverse direction to the voltage which appears across the component.

If you have an inductor in a circuit between node A and node B, then the formula to use in circuit analysis is:

$$V_A - V_B = L\frac{dI_{AB}}{dt} \tag{11.19}$$

where V_A is the voltage at node A, V_B is the voltage at node B and I_{AB} is the current in the direction from A to B. This is almost the same as the equivalent equation for resistors:

$$V_A - V_B = R \times I_{AB} \tag{11.20}$$

11.5.2 Transformers

What if there are two coils of wire? If they are placed close together like this:

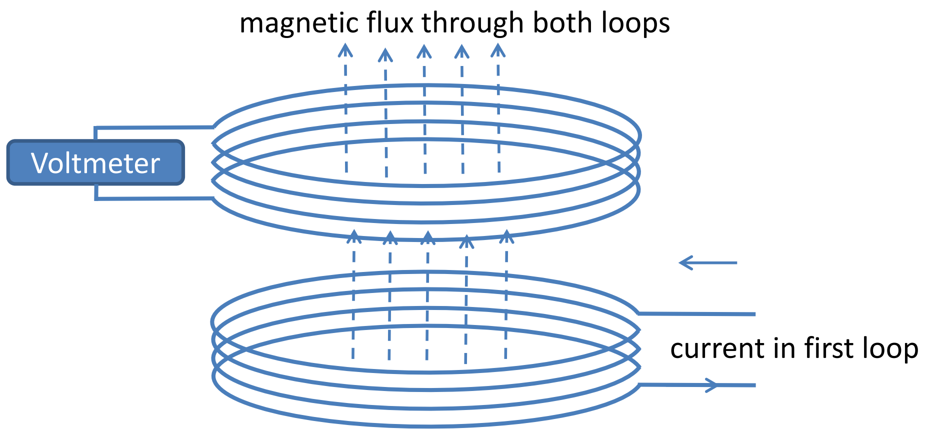

Figure 11.6 Two loops sharing the same magnetic field

then the magnetic field generated by the current in one coil will also pass through the other coil. This suggests that a changing current in one coil will produce an EMF in the other coil as well as in its own coil. This is very useful: this is a way to pass signals (and power) from one circuit to another while having no electrical connection between. It's called a transformer.

Most transformers don't have the same number of windings in both loops. Since the flux generated and the EMF produced are both proportional to the number of windings, this allows transformers to reduce (or increase) the voltage in an AC circuit without loss of power. (At least ideal transformers don't lose power, real transformers will lose some power in the resistance of the wires in the windings and the effect of the non-ideal core causing heat losses in the ferromagnetic core.)

11.6 Maxwell's equations and electromagnetic waves

We've seen above that electric fields can generate magnetic fields, and time-varying magnetic fields can generate electric fields. So in theory, if you had a time-varying electric field, it would produce a time-varying magnetic field around it, and these time-varying magnetic fields would produce time-varying electric fields around them, which would in turn produce time-varying magnetic fields around them, which would in turn produce time-varying electric fields around them…

The whole pattern of magnetic and electric fields would propagate outwards into space from the original source (which can be a source of either electric or magnetic fields). A sensible name for these patterns of fields might be "electromagnetic waves".

They were first predicted by James Clerk Maxwell[9] in 1864, in which he calculated the expected speed of these waves to be 310,740,000 m/s. Due to some very elegant experiments with rotating mirrors, the speed of light was already known at that time, and within experimental error, the two results agreed. This was the first evidence that light was an electromagnetic wave.

11.7 Summary: the most important things to know

- Current produce magnetic fields around themselves.
- Changing the magnetic field lines going through a circuit results in an EMF (an effective voltage) appearing around the circuit, which can cause currents to flow.

---

[9] Although a brilliant mathematical physicist, you might be well-advised to avoid his poetry, in particular the excruciating "Song of the Cub".

- The rate of change of the current in an inductor is linearly proportional to the rate of change of the magnetic field produced in the core, and that in turn is linearly proportional to the EMF produced around the circuit.
    - This leads to the inductor equation: $e(t) = L\, di/dt$.