Chapter 1

Hall-Effect Physics

Conceptually, a demonstration of the Hall effect is simple to set up and is illustrated in Figure 1-1. Figure 1-1a shows a thin plate of conductive material, such as copper, that is carrying a current (I), in this case supplied by a battery. One can position a pair of probes connected to a voltmeter opposite each other along the sides of this plate such that the measured voltage is zero.

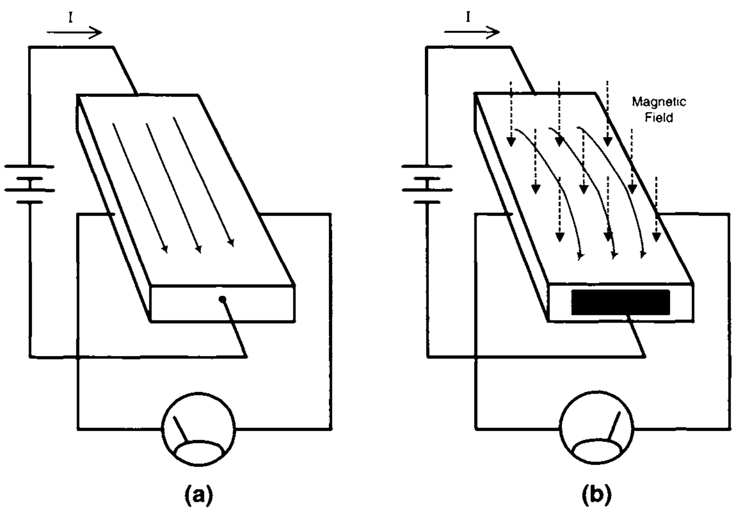

Figure 1-1: The Hall effect in a conductive sheet.

When a magnetic field is applied to the plate so that it is at right angles to the current flow, as shown in Figure 1-1b, a small voltage appears across the plate, which can be measured by the probes. If you reverse the direction (polarity) of the magnetic field, the polarity of this induced voltage will also reverse. This phenomenon is called the Hall effect, named after Edwin Hall.

What made the Hall effect a surprising discovery for its time (1879) is that it occurs under steady-state conditions, meaning that the voltage across the plate persists even when the current and magnetic field are constant over time. When a magnetic field varies with time, voltages are established by the mechanism of induction, and induction was well understood in the late 19th century. Observing a short voltage pulse across the plate when a magnet was brought up to it, and another one when the magnetic field was removed, would not have surprised a physicist of that era. The continuous behavior of the Hall-effect, however, presented a genuinely new phenomenon.

Under most conditions the Hall-effect voltage in metals is extremely small and difficult to measure and is not something that would likely have been discovered by accident. The initial observation that led to discovery of the Hall effect occurred in the 1820s, when Andre A. Ampere discovered that current-carrying wires experienced mechanical force when placed in a magnetic field (Figure 1-2). Hall's question was whether it was the wires or the current in the wires that was experiencing the force. Hall reasoned that if the force was acting on the current itself, it should crowd the current to one side of the wire. In addition to producing a force, this crowding of the current should also cause a slight, but measurable, voltage across the wire.

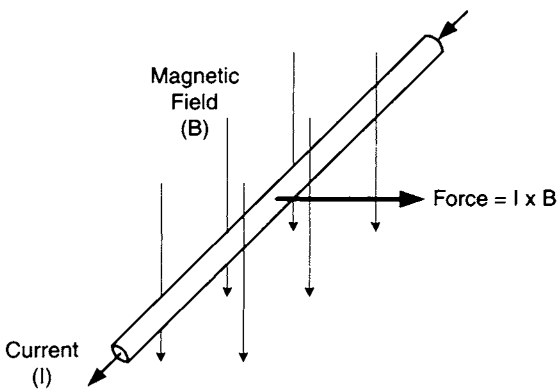

Figure 1-2: A magnetic field exerts mechanical force on a current-carrying wire.

Hall's hypothesis was substantially correct; current flowing down a wire in a magnetic field does slightly crowd to one side, as illustrated in Figure 1-1b, the degree of crowding being highly exaggerated. This phenomenon would occur whether or not the current consists of large numbers of discrete particles, as is now known, or whether it is a continuous fluid, as was commonly believed in Hall's time.

1.1 A Quantitative Examination

Enough is presently known about both electromagnetics and the properties of various materials to enable one to analyze and design practical magnetic transducers based on the Hall effect. Where the previous section described the Hall effect qualitatively, this section will attempt to provide a more quantitative description of the effect and to relate it to fundamental electromagnetic theory.

In order to understand the Hall effect, one must understand how charged particles, such as electrons, move in response to electric and magnetic fields. The force exerted on a charged particle by an electromagnetic field is described by:

$$\vec{F} = q_0\vec{E} + q_0\vec{v} \times \vec{B}$$ **(Equation 1-1)**

where $\vec{F}$ is the resultant force, $\vec{E}$ is the electric field, $\vec{v}$ is the velocity of the charge, $\vec{B}$ is the magnetic field, and q_0 is the magnitude of the charge. This relationship is commonly referred to as the Lorentz force equation. Note that, except for q_0, all of these variables are vector quantities, meaning that they contain independent x, y, and z components. This equation represents two separate effects: the response of a charge to an electric field, and the response of a moving charge to a magnetic field.

In the case of the electric field, a charge will experience a force in the direction of the field, proportional both to the magnitude of the charge and the strength of the field. This effect is what causes an electric current to flow. Electrons in a conductor are pulled along by the electric field developed by differences in potential (voltage) at different points.

In the case of the magnetic field, a charged particle doesn't experience any force unless it is moving. When it is moving, the force experienced by a charged particle is a function of its charge, the direction in which it is moving, and the orientation of the magnetic field it is moving through. Note that particles with opposite charges will experience force in opposite directions; the signs of all variables are significant. In the simple case where the velocity is at right angles to the magnetic field, the force exerted is at right angles to both the velocity and the magnetic field. The cross-product operator ($\times$) describes this relationship exactly. Expanded out, the force in each axis (x,y,z) is related to the velocity and magnetic field components in the various axes by:

$$F_x = q_0\left(v_y B_z - v_z B_y\right)$$
$$F_y = q_0\left(v_z B_z - v_z B_x\right)$$
$$F_z = q_0\left(v_x B_y - v_y B_x\right)$$ **(Equation 1-2)**

The forces a moving charge experiences in a magnetic field cause it to move in curved paths, as depicted in Figure 1-3. Depending on the relationship of the velocity to the magnetic field, the motion can be in circular or helical patterns.

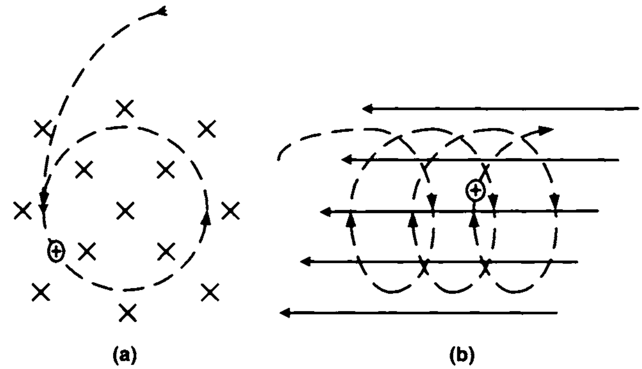

Figure 1-3: Magnetic fields cause charged particles to move in circular (a) or helical (b) paths.

In the case of charge carriers moving through a Hall transducer, the charge carrier velocity is substantially in one direction along the length of the device, as shown in Figure 1-4, and the sense electrodes are connected along a perpendicular axis across the width. By constraining the carrier velocity to the x axis ($vy = 0$, $vz = 0$) and the sensing of charge imbalance to the z axis, we can simplify the above three sets of equations to one:

$$F_z = q_0 v_x B_y \qquad \textbf{(Equation 1-3)}$$

which implies that the Hall-effect transducer will be sensitive only to the y component of the magnetic field. This would lead one to expect that a Hall-effect transducer would be orientation sensitive, and this is indeed the case. Practical devices are sensitive to magnetic field components along a single axis and are substantially insensitive to those components on the two remaining axes. (See Figure 1-4.)

Although the magnetic field forces the charge carriers to one side of the Hall transducer, this process is self-limiting, because the excess concentration of charges to one side and consequent depletion on the other gives rise to an electric field across the transducer. This field causes the carriers to try to redistribute themselves more evenly. It also gives rise to a voltage that can be measured across the plate. An equilibrium develops where the magnetic force pushing the charge carriers aside is balanced out by the electric force trying to push them back toward the middle

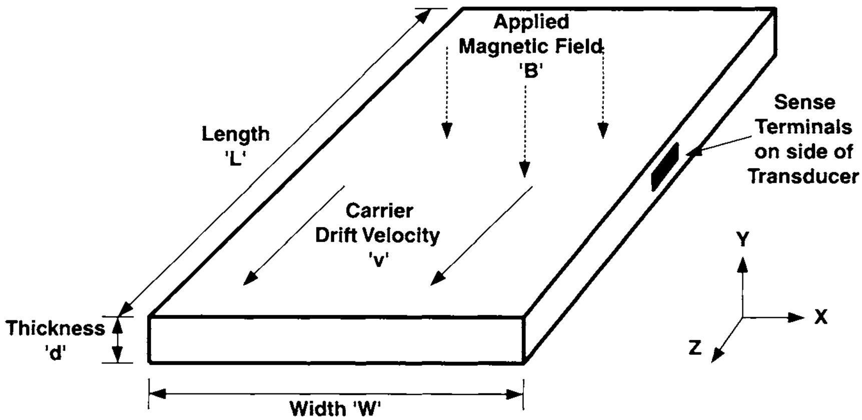

Figure 1-4: Hall-effect transducer showing critical dimensions and reference axis.

$$q_0 E_H + q_0 v \times B = 0 \qquad \textbf{(Equation 1-4)}$$

where E_H is the Hall electric field across the transducer. Solving for E_H yields

$$E_H = -v \times B \qquad \textbf{(Equation 1-5)}$$

which means that the Hall field is solely a function of the velocity of the charge carriers and the strength of the magnetic field. For a transducer with a given width w between sense electrodes, the Hall electric field can be integrated over w, assuming it is uniform, giving us the Hall voltage.

$$V_H = -wvB \qquad \textbf{(Equation 1-6)}$$

The Hall voltage is therefore a linear function of:

a) the charge carrier velocity in the body of the transducer,
b) the applied magnetic field in the "sensitive" axis,
c) the spatial separation of the sense contacts, at right angles to carrier motion.

1.2 Hall Effect in Metals

To estimate the sensitivity of a given Hall transducer, it is necessary to know the average charge carrier velocity. In a metal, conduction electrons are free to move about and do so at random because of their thermal energy. These random "thermal velocities" can be quite high for any given electron, but because the motion is random, the motions of individual electrons average out to a zero net motion, resulting in no current. When an electric field is applied to a conductor, the electrons "drift" in the direction of the applied field, while still performing a fast random walk from their thermal energy. This average rate of motion from an electric field is known as *drift velocity*.

In the case of highly conductive metals, drift velocity can be estimated. The first step is to calculate the density of carriers per unit volume. In the case of a metal such as copper, it can be assumed that every copper atom has one electron in its outer shell that is available for conducting electric current. The volumetric carrier density is therefore the product of the number of atoms per unit of weight and the specific gravity. For the case of copper this can be calculated:

$$N = \frac{N_A}{M_m} D = \frac{6.02 \times 10^{23}\ \text{mol}^{-1}}{63.55\text{g} \cdot \text{mol}^{-1}} \times 8.89\text{g} \cdot \text{cm}^{-3} = 8.42 \times 10^{22}\ \text{cm}^{-3} \quad \textbf{(Equation 1-7)}$$

where: N is the number of carriers per cubic centimeter
N_A is the Avogadro constant (6.02×10^{23} mol^{-1})
M_m is the molar mass of copper (63.55 g $\cdot$ mol^{-1})
D is specific gravity of copper (grams/cm3)

Once one has the carrier density, one can estimate the carrier drift velocity based on current. The unit of current, the ampere (A), is defined as the passage of $\approx 6.2 \times 10^{18}$ charge carriers per second and is equal to $1/q0$. Consider the case of a piece of conductive material with a given cross-sectional area of A. The carrier velocity will be proportional to the current, as twice as much current will push twice as many carriers through per unit time. Assuming that the carrier density is constant and the carriers behave like an incompressible fluid, the velocity will also be inversely proportional to the cross section, a larger cross section meaning lower carrier velocity. The carrier drift velocity can be determined by:

$$v = \frac{I}{q_0 NA} \quad \textbf{(Equation 1-8)}$$

where
v is carrier velocity, cm/sec
I is current in amperes
Q_0 is the charge on an electron (1.60×10^{-19} C)
N is the carrier density, carriers/cm^3
A is the cross section in cm^2

One surprising result is the drift velocity of carriers in metals. While the electric field that causes the charge carriers to move propagates through a conductor at approximately half the speed of light (300×10^6 m/s), the actual carriers move along at a much more leisurely average pace. To get an idea of the disparity, consider a piece of

#18 gauge copper wire carrying one ampere. This gauge of wire is commonly used for wiring lamps and other household appliances and has a cross section of about 0.0078 cm^2. One ampere is about the amount of current required to light a 100-watt light bulb. Using the previously derived carrier density for copper and substituting into the previous equation gives:

$$v = \frac{1\text{A}}{1.6 \times 10^{-19}\text{C} \cdot 8.42 \times 10^{22}\text{cm}^{-3} \cdot 0.0078\,\text{cm}^2} = 0.009\,\text{cm} \cdot s^{-1} \qquad \textbf{(Equation 1-9)}$$

The carrier drift velocity in the above example is considerably slower than the speed of light; in fact, it is considerably slower than the speed of your average garden snail.

By combining Equations (1-6) and (1-8), we can derive an expression that describes the sensitivity of a Hall transducer as a function of cross-sectional dimensions, current, and carrier density:

$$V_H = \frac{IB}{q_0 N d} \qquad \textbf{(Equation 1-10)}$$

where d is the thickness of the conductor.

Consider the case of a transducer consisting of a piece of copper foil, similar to that shown back in Figure 1-1. Assume the current to be 1 ampere and the thickness to be 25 μm (0.001"). For a magnetic field of 1 tesla (10,000 gauss) the resulting Hall voltage will be:

$$V_H = \frac{1\text{A} \cdot 1T}{1.6 \times 10^{-19}\text{C} \cdot 8.42 \times 10^{28}\text{m}^{-3} \cdot 25 \times 10^{-6}\text{m}} = 3.0 \times 10^{-6}\,\text{V} \qquad \textbf{(Equation 1-11)}$$

Note the conversion of all quantities to SI (meter-kilogram-second) units for consistency in the calculation.

Even for the case of a magnetic field as strong as 10,000 gauss, the voltage resulting from the Hall effect is extremely small. For this reason, it is not usually practical to make Hall-effect transducers with most metals.

1.3 The Hall Effect in Semiconductors

From the previous description of the Hall effect in metals, it can be seen that one means of improvement might be to find materials that do not have as many carriers per unit volume as metals do. A material with a lower carrier density will exhibit the Hall effect more strongly for a given current and depth. Fortunately, semiconductor materials such as silicon, germanium, and gallium-arsenide provide the low carrier densities needed to realize practical transducer elements. In the case of semiconductors, carrier density is usually referred to as *carrier concentration*.

Table 1-1: Intrinsic carrier concentrations at 300°K [Socl85]

Material	Carrier Concentration (cm^{-3})
Copper (est.)	8.4×10^{22}
Silicon	1.4×10^{10}
Germanium	2.1×10^{12}
Gallium-Arsenide	1.1×10^{7}

As can be seen from Table 1-1, these semiconductor materials have carrier concentrations that are orders of magnitude lower than those found in metals. This is because in metals most atoms contribute a conduction electron, whereas the conduction electrons in semiconductors are more tightly held. Electrons in a semiconductor only become available for conduction when they acquire enough thermal energy to reach a conduction state; this makes the carrier concentration highly dependent on temperature.

Semiconductor materials, however, are rarely used in their pure form, but are doped with materials to deliberately raise the carrier concentration to a desired level. Adding a substance like phosphorous, which has five electrons in its outer orbital (and appears in column V of the periodic table) adds electrons as carriers. This results in what is known as an N-type semiconductor. Similarly, one can also add positive charge carriers by doping a semiconductor with column-III materials (three electrons in the outer orbital) such as boron. While this doesn't mean that there are free-floating protons available to carry charge, adding a column-III atom removes an electron from the semiconductor crystal to create a "hole" that moves around and behaves as if it were actually a charge-carrying particle. This type of semiconductor is called a P-type material.

For purposes of making Hall transducers, there are several advantages to using doped semiconductor materials. The first is that, because of the low intrinsic carrier concentrations of the pure semiconductors, unless materials can be obtained with part-per-trillion purity levels, the material will be doped anyhow—but it will be unknown with what or to what degree.

The second reason for doping the material is that it allows a choice of the predominant charge carrier. In metals, there is no choice; electrons are the default charge carriers. However, semiconductors there is the choice of either electrons or holes. Since electrons tend to move faster under a given set of conditions than holes, more sensitive Hall transducers can be made using an N-type material in which electrons are the majority carriers than with a P-type material in which current is carried by holes.

The third reason for using doped materials is that, for pure semiconductors, the carrier concentration is a strong function of temperature. The carrier concentration resulting from the addition of dopants is mostly a function of the dopant concentration, which isn't going to change over temperature. By using a high enough concentration of dopant, one can obtain relatively stable carrier concentrations over temperature. Since the Hall voltage is a function of carrier concentration, using highly doped materials results in a more temperature-stable transducer.

In the case of Hall transducers on integrated circuits, there is one more reason for using doped silicon—mainly because that's all that is available. The various silicon layers used in common IC processes are doped with varying levels of N and P materials, depending on their intended function. Layers of pure silicon are not usually available as part of standard IC fabrication processes.

1.4 A Silicon Hall-Effect Transducer

Consider a Hall transducer constructed from N-type silicon that has been doped to a level of 3×10^{15} cm^{-3}. The thickness is 25µm and the current is 1 mA. By substituting the relevant numbers into Equation 1-10, we can calculate the voltage output for a 1-tesla field:

$$V_H = \frac{0.001A \cdot 1\mathrm{T}}{1.6\times10^{-19}C \cdot 3\times10^{21}\ \mathrm{m}^{-3} \cdot 25\times10^{-6}\mathrm{m}} = 0.083\mathrm{V} \qquad \textbf{(Equation 1-12)}$$

The resultant voltage in this case is 83 mV, which is more than 20,000 times the signal of the copper transducer described previously. Equally significant is that the necessary bias current is 1/1000 that used to bias the copper transducer. Millivolt-level output signals and milliamp-level bias currents make for practical sensors.

While one can calculate transducer sensitivity as a function of geometry, doping levels, and bias current, there is one detail we have ignored to this point: the resistance of the transducer. While it is possible to get tremendous sensitivities from thinly doped semiconductor transducers for milliamps of bias current, it may also require hundreds of volts to force that current through the transducer. The resistance of the Hall transducer is a function of the conductivity and the geometry; for a rectangular slab, the resistance can be calculated by:

$$R = \frac{\sigma \cdot l}{w \cdot d} \qquad \textbf{(Equation 1-13)}$$

where

R is resistance in ohms
σ is the resistivity in ohm-cm
l is the length in cm
w is the width in cm
d is the thickness in cm

In the case of metals, σ is a characteristic of the material. In the case of a semiconductor, however, σ is a function of both the doping and a property called carrier mobility. Carrier mobility is a measure of how fast the charge carriers move in response to an electric field, and varies with respect to the type of semiconductor, the dopant concentration level, the carrier type (N or P type), and temperature.

In the case of the silicon Hall-effect transducer described above (d = 25 μm = 0.0025 cm), made from N-type silicon doped to a level of $3 \times 10^{15}\text{cm}^{-3}$, $\sigma \approx 1.7\ \Omega$-cm at room temperature. Let us also assume that the transducer is 0.1 cm long and 0.05 cm wide. The resistance of this transducer is given by:

$$R = \frac{\sigma \cdot l}{w \cdot d} = \frac{1.7\Omega \cdot \text{cm} \times 0.1\,\text{cm}}{0.05\,\text{cm} \times 0.0025\,\text{cm}} = 1360\Omega \qquad \textbf{(Equation 1-14)}$$

With a resistance of 1360Ω, it will take 1.36V to force 1 mA of current through the device. This results in a power dissipation of 1.36 mW, a modest amount of power that can be easily obtained in many electronic systems. The sensitivity and power consumption offered by Hall-effect transducers made from silicon or other semiconductors makes them practical sensing devices.