Hall Effect Devices

By W. J. GRUBBS

(Manuscript received October 21, 1958)

A wealth of devices which depend on the Hall effect for their operation have been proposed in the last decade. This paper gives the results of a survey of these devices. Original work in this field is included in those sections which describe the circulator, one-piece gyrator, switch, frequency spectrum analyzer, phase discriminator and digital-to-analog encoder. Semiconductor materials are discussed in terms of what type of material is most desirable and how currently available materials limit the usefulness of Hall effect devices.

I. INTRODUCTION

If a magnetic field is applied perpendicular to a current flow in any conductor, the moving charges (which constitute the current) are deflected sidewise and build up a potential difference between the two sides of the conductor. The creation of this transverse electric field (perpendicular to both the magnetic field and the original current flow) is called the Hall effect. During recent years, interest in this effect has increased tremendously. Before semiconductors and their capabilities were understood, the Hall effect in solids was little more than a laboratory curiosity. Now it is not only an important tool in metallurgy and semiconductor device development, but it has been the mode of operation of many proposed devices. This article describes how 20 or so of these devices operate. In each case, the major advantages or disadvantages are mentioned, but no attempt is made actually to determine the usefulness of the device.

The devices to be discussed have been arbitrarily divided into two groups: devices which use a constant magnetic field and devices in which a signal or an oscillator produces at least a part of the magnetic field. Such a division is not entirely arbitrary, because the first group inherently has a very high limit on the operating frequency and the second group has a considerably lower limit. The devices are listed on the following page.

Constant magnetic field

Gyrator
Isolator
Negative-resistance amplifier
Circulator

Signal-produced magnetic field

Switch
Transducer
Magnetic field meter
Electrical compass
Magnetic field variation meter
Ammeter
Wattmeter
Amplifier
Modulator (and nondrift dc amplifier)
Demodulator
Frequency spectrum analyzer
Phase discriminator
Digital-to-analog encoder
Analog multiplier

II. CONSTANT MAGNETIC FIELD DEVICES

2.1 *Gyrator*

The gyrator has received more attention[1,2,3,4,5] than has any other Hall effect device — undoubtedly because it was the first nonreciprocal circuit element to which the electrical world was exposed. Casimir[6] was the first to point out in an English publication that, in a conducting solid, if $R_{ik}(H)$ is the transfer impedance from terminal pair i to pair k in the presence of an orthogonal magnetic field H, then

$$R_{ik}(H) \neq R_{ki}(H),$$

but

$$R_{ik}(H) = R_{ki}(-H). \tag{1}$$

This is simply a statement of the fact that the presence of the magnetic field causes the reciprocity theorem to be violated. With $H = 0$,

$$R_{ik} = R_{ki}\,.$$

Casimir credited Meixner[7] with being the first to prove (1).

McMillan[8] pointed out that many transducers (such as the crystal or condenser types) are reciprocal, but that electrodynamic or magnetic transducers are antireciprocal. An antireciprocal four-pole is a transducer in which the transfer impedance from left-to-right is opposite in sign and equal in magnitude to the transfer from right-to-left. McMillan suggested that by combining the two kinds of transducers (reciprocal and antireciprocal) it should be possible to produce a nonreciprocal transducer — one in which the two transfers have unequal magnitude.

It may be noted that McMillan's antireciprocal four-pole corresponds to at least one definition of the gyrator and his nonreciprocal four-pole bears a resemblance to an isolator. The gyrator may be defined as a four-pole in which the two transfer impedances are equal in magnitude but have phase angles which differ by 180° (i.e., a gyrator is an antireciprocal four-pole).

In a letter to the editor, McMillan[9] suggested the Hall effect as one

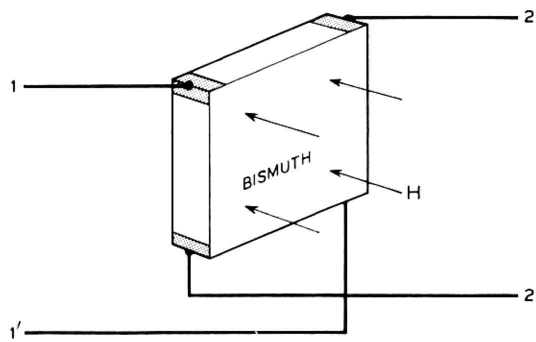

Fig. 1 — McMillan's antireciprocal four-pole (proposed).

possible means of achieving a nonreciprocal electrical system. His scheme was to place a slab of bismuth with contacts in a magnetic field, as shown in Fig. 1. He selected bismuth because it exhibits a very large Hall effect.

Tellegen[4] referred to McMillan's work and went on to give the gyrator its name,* propose in general terms possible means of making it and suggest many ways in which it would be helpful in electrical network synthesis.

Hogan[10] used the Faraday effect to make a gyrator at microwave frequencies. Mason *et al.*[2] reported making gyrators which use the Hall effect for their operation, and Wick[5] showed that the minimum possible power loss of a Hall effect gyrator is 7.66 db.

The type of Hall effect gyrator which has been studied most thoroughly

* The gyrator is the electrical analog of a mechanical gyroscope.

consists of a square slab of semiconductor with an ohmic contact in the middle of each edge face and a constant magnetic field H perpendicular to the plane of the slab (see Fig. 2). There are no junctions and the reader may readily convince himself that such a device is antireciprocal.

If the Hall angle is small, it is given by

$$\theta_H = \tan^{-1} (\mu_H H \times 10^{-8}), \tag{2}$$

where μ_H is the Hall mobility of the semiconductor's majority carrier in cm²/volt-second and H is the magnetic field intensity in oersteds. (This approximation is valid as long as the product of the majority carrier mobility and density is much greater than the same product for minority carriers.) It can be shown (again, for small θ_H) that, in the Hall effect gyrator with the output leads open-circuited, the output voltage is

$$V_{\text{out}} = V_{\text{in}}(\mu_H H \times 10^{-8}). \tag{3}$$

An n-type germanium gyrator has a loss of about 14 db with $H = 17{,}500$. If a material with higher mobility were used, the same loss would be observed with a smaller field or a smaller loss with the same field. Hogan[10] refers to a vacuum tube Hall effect gyrator proposed by R. O. Grisdale which had four electrodes that could both emit and collect electrons. Such a device could have very high electron mobility. It was built and found to have a loss of "about 7 db". Apparently $\tan \theta_H$ became so very high that the minimum possible loss was approached. Of course, this is more complicated than most users prefer, so the device was not developed for practical use.

It is worth mentioning that a semiconductor Hall effect gyrator operates at direct current and, theoretically, at any frequency up to the dielectric relaxation frequency of the material used.

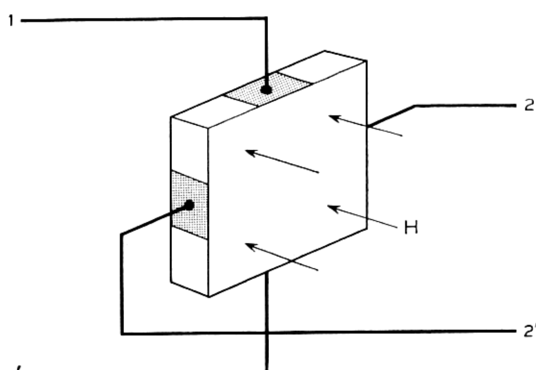

Fig. 2 — Hall effect gyrator (built).

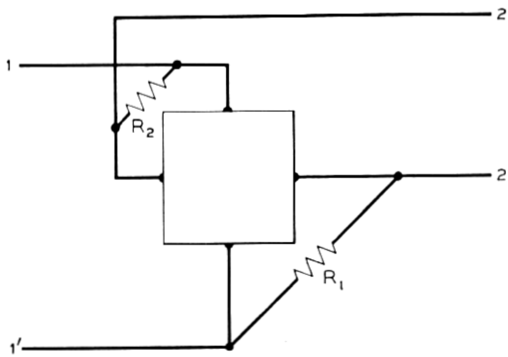

Fig. 3 — Hall effect isolator made from gyrator.

An ideal gyrator is also an ideal impedance inverter, but the loss inherent in the Hall effect gyrator makes it useless as an inverter. This loss is of such a nature that it could be effectively removed by placing negative resistance of the proper value in series with the gyrator. This lossless or ideal gyrator should then find application in network synthesis.

The gyrator has been discussed here in some detail not only because of its historical interest but, primarily, because it has been the basis of the major portion of the devices to be described in the remainder of this article.

2.2 *Isolator*

An isolator[2,3,5] is a nonreciprocal four-pole in which one of the two transfer impedances is zero. Thus it transmits signals in one direction only — say, from terminal pair 1 to pair 2. Such a device can be made from a Hall effect gyrator by the addition of two shunting resistors (see Fig. 3). Assuming the gyrator is symmetrical, its two self impedances may be called z_s and its transfer impedances z_{T1} and z_{T2}. Furthermore,

$$z_{T1} = -z_{T2} = z_T > 0.$$

Then it can be shown that the parallel resistance values R_1 and R_2 must be such that

$$R_1 + R_2 = \frac{z_s^2 + z_T^2}{z_T}$$

in order to obtain an isolator.

Since $z_T/z_s = \tan \theta_H$, it may be seen that, with R_1 and R_2 fixed in value, to maintain isolation one must maintain a constant H, a constant μ_H and, therefore, a constant temperature. Also, if the resistance of the semiconductor contacts is current-sensitive, isolation might be possible only in a limited range of signals. The forward loss of such an isolator is just slightly less than that of the gyrator from which it is made. The minimum possible forward loss is 6 db, and the loss in experimental germanium isolators has been approximately 14 db. Thus, to transmit 1 mw of power, one must apply 26 mw to the input because 25 mw will be dissipated in the isolator, most of it in the germanium. This is another source of current sensitivity — changing the current level changes the amount of power dissipated in the isolator, changing its temperature and therefore z_s, z_T/z_s, R_1 and R_2. The higher the reverse loss is (i.e., the better the balance) the more sensitive the isolator will be to all such variations. The reverse loss in germanium isolators has been made 75 db, and this loss is quite sensitive to small variations.

When indium antimonide (which has a much higher μ_H) was used, the forward loss was reduced to about 7.5 db, even with a smaller H. With carefully adjusted shunt resistance values, a reverse loss of "the order of 100 db" was obtained. Notice that the forward loss was very nearly the minimum possible loss. Unfortunately, InSb has a quite small energy gap and consequently a very high np product. This fact, in conjunction with its very high mobility, inevitably gives rise to a very low resistivity. Thus, the InSb isolator had an impedance of around 1 ohm. Germanium isolators can easily be made with impedance levels ranging from 10 to 1000 ohms.

Typically, μ_H in InSb is a much stronger function of temperature than is μ_H in germanium. Thus, this low-loss isolator has some serious drawbacks.

There is another way of making a Hall effect isolator, and it has been suggested that this second method should remove the temperature sensitivity. Unfortunately, this is not the case. This form of the isolator

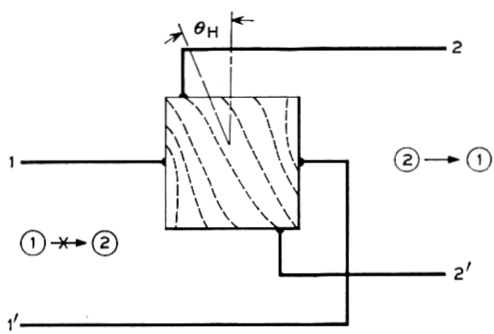

Fig. 4 — Hall effect isolator (or skew gyrator).

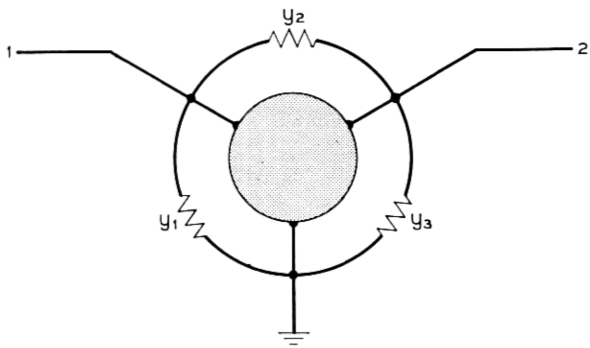

Fig. 5 — Negative resistance amplifier.

uses no resistors and has been called a "skew gyrator," in which the sample is square but the leads are not symmetrically spaced. Fig. 4 shows the sample with the dotted equipotential lines which would occur with a signal applied to side 1 if the magnetic field were the proper value. Obviously, no voltage appears between leads 2 and 2′. If a signal is applied to side 2, there will be an output at side 1. This is stated symbolically as

$$1 \nrightarrow 2 \rightarrow 1.$$

This type of isolator is still sensitive to variations in H, μ_H and temperature (and, therefore, current level). This sensitivity arises from the need for a constant θ_H to maintain the slope of the equipotential lines. Of course, if the semiconductor has the proper doping level, so that μ_H is independent of temperature near room temperature, this device could be quite stable. For that matter, most of these devices could be stable with such a constant mobility material.

As for the shunt-resistor isolator, it is conceivable that resistors could be used that had the proper temperature coefficient for this isolator also to be made stable — perhaps even more stable than the skew gyrator.

2.3 *Negative-Resistance Amplifier*

The negative-resistance amplifier[5,11] uses a slice of semiconductor with three equally spaced edge contacts, a perpendicular magnetic field of constant value H and three negative resistances (see Fig. 5). For the moment, assume the parallel conductances are not in the circuit. The short-circuit admittances may be defined as

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2=0} = y_0, \qquad y_{22} = \left.\frac{i_2}{v_2}\right|_{v_1=0} = y_0,$$

$$y_{12} = \left.\frac{i_1}{v_2}\right|_{v_1=0}, \qquad y_{21} = \left.\frac{i_2}{v_1}\right|_{v_2=0}.$$

A parameter α (an odd function of θ_H) may be such that

$$y_{12} = -\frac{y_0}{2}(1 - \alpha), \qquad y_{21} = -\frac{y_0}{2}(1 + \alpha) \qquad |\alpha| < 1.$$

However, if the parallel admittances are connected and given the values

$$y_1 = y_3 = y_{21},$$

$$y_2 = y_{12},$$

it can be shown that $i_1 = 0$ and $i_2 = \alpha y_0 v_1$. Therefore, with the negative conductances in parallel, the device becomes an isolator which can have gain in the forward direction. One of its advantages is that it permits the construction of negative-resistance amplifiers which are unidirectional, so that higher gain with the same degree of stability is possible with this type of device than with the usual two-terminal negative resistance.

Negative resistance obtained from gas tubes was used in this way to give a forward gain of 6 db and a reverse loss of 46 db.

This type of device was proposed as a high-frequency amplifier because of the inherent insensitivity to frequency of Hall effect gyrators and related devices. Just how good they are at very high frequencies is a question which will be discussed a little later.

The same principle may be applied to the shunt-resistor isolator. Still another form of negative resistance amplifier employing Hall effect will be mentioned in the next section.

2.4 *Circulator*

A circulator is a nonreciprocal n-port device ($n > 2$) in which a signal applied at one port is transmitted only to an adjacent port (e.g., the adjacent clockwise port). No output appears at any other port. A three-port circulator can be made which uses the Hall effect. As one may observe from the skew gyrator, the Hall effect merely enables one to prevent any signal from appearing at one of the outputs. Therefore, if we had four or more ports, a portion of the signal would appear at all ports except one, and the device would not really be a circulator. See Fig. 6 for the symbol of a three-port circulator.

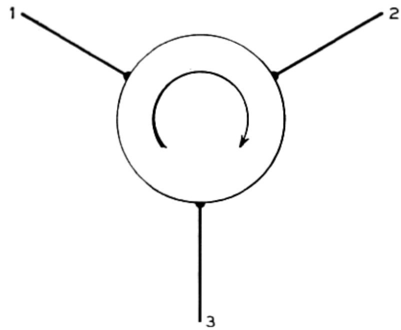

Fig. 6 — General three-port circulator.

The Hall effect circulator[12,13] consists of a circular slab of semiconductor (n-type germanium has been used to date) with six equally spaced edge contacts and a constant magnetic field H applied perpendicular to the plane of the slab. Such a sample is shown in Fig. 7 with equipotential lines dotted in. With the proper value of H, no output appears at 3 but there is an output signal at 2. The results would be similar for an input at 2 or 3 as well. So we may say

$$1 \to 2 \to 3 \to 1,$$

$$3 \nrightarrow 2 \nrightarrow 1 \nrightarrow 3.$$

When a load is attached so that a current may flow at 2 (with the signal applied at 1) this secondary current flow produces an additional Hall electric field which has the net effect of reducing θ_H . To maintain

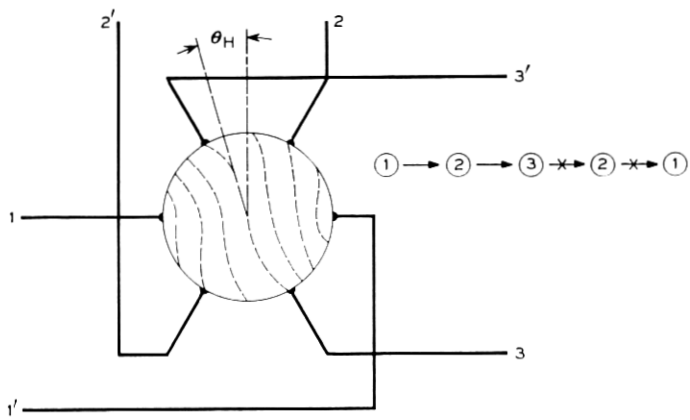

Fig. 7 — Hall effect circulator.

circulation, H must be increased. Thus, it is obvious that the circulator must be terminated in the proper impedances if its circulation property is to be retained. At a particular magnetic field, H_0, the load impedance necessary to "balance" the circulator is the same as the circulator's input impedance, so H_0 is the logical field to use to avoid impedance mismatches.

When n-type germanium is used, $H_0 \approx 14{,}500$ oersteds, and the matching load impedance, Z_L, is a function of the sample's thickness and resistivity, and might vary from 10 to 1000 ohms. If n-type InSb were used, H_0 would be in the order of 1000 oersteds, but (as with the isolator) Z_L would have to be quite low — 1 ohm or even less.

It is almost impossible to make a circulator which has the proper impedance level and is symmetrical so that it will circulate a signal applied at any input. Fortunately there is a means of overcoming this difficulty so that a circulator of the type which is quite easily fabricated may be made to appear symmetrical and to operate at the proper impedance level. All that is involved is placing a network of six resistors in parallel with the circulator (see Fig. 8). Use C_P with $H < H_0$ and C_s with $H > H_0$. Apparently the three-resistor networks shown in Fig. 9 can be used equally effectively but, since C_P and C_s have six variables, they should offer greater flexibility than does C_p' or either form of C_s'.

The parallel networks permit Z_L to be adjusted to any value between about one tenth its normal value and infinity (open-circuit), using $H = H_0$ all the while. One can also use the normal value of Z_L with any H from zero to infinity. Another possible function of the networks is to permit an approximately symmetrical circulator to be used with unequal loads at the three-ports.

The forward loss of a circulator is about 17 db if it is used without parallel networks, and reverse losses of about 65 db have been achieved. With n-type germanium, H must be about 14,500 oersteds. If InSb were

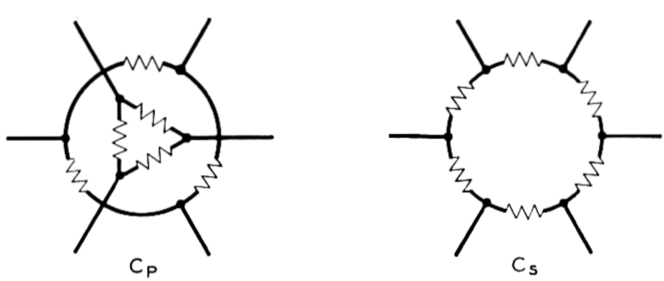

Fig. 8 — Parallel networks for circulator.

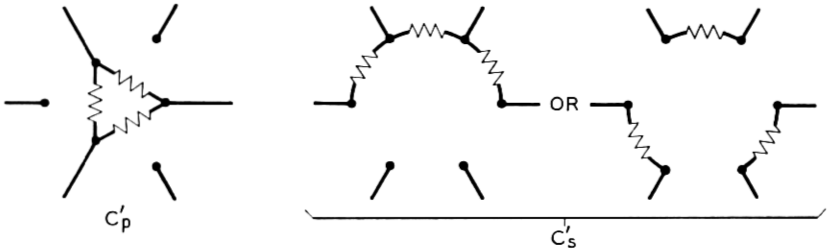

Fig. 9 — Alternative parallel networks.

used, H would be only about 1000. However, the forward loss would still be 17 db, since the loss is determined by θ_H, and θ_H must be the same in any Hall effect circulator.

On the other hand, if one uses network C_s and increases H, the forward loss is reduced. With n-type germanium, it has been reduced to about 15 db. Judging by the low loss achieved with an InSb isolator, one would expect that an InSb circulator could be made with a forward loss within one db or so of the minimum possible loss. The minimum loss has been calculated to be 8.4 db, so an InSb circulator probably could be made with a 9- or 10-db forward loss. The reverse loss would depend solely on how precisely the resistance values of C_s were adjusted. Of course, InSb still has the disadvantage of very low resistivity, and the impedance level would be about 1 ohm. As previously stated, such a circulator could be operated with any higher impedance load, but the necessary mismatch would increase the insertion loss tremendously.

Perhaps another means of reducing the loss should be considered — that is, using negative resistance in C_P or C_s. By this means, the loss could be removed, or gain could even be achieved, and one would have an amplifying circulator. Negative resistance is still hard to get, but if someone knew of a good means of producing it, he could make the circulator an even more interesting device.

III. GENERAL REMARKS

All the devices described in the preceding pages could theoretically transmit dc signals as well as ac signals of any frequency up to the dielectric relaxation frequency of the semiconductor material. In order to realize even a major portion of this huge bandwidth ($\approx$ thousands of megacycles) special care must be taken to shield all input and output leads properly. The samples should be small, so that only a small magnet will be required. Consequently, all the leads must be fairly close to one

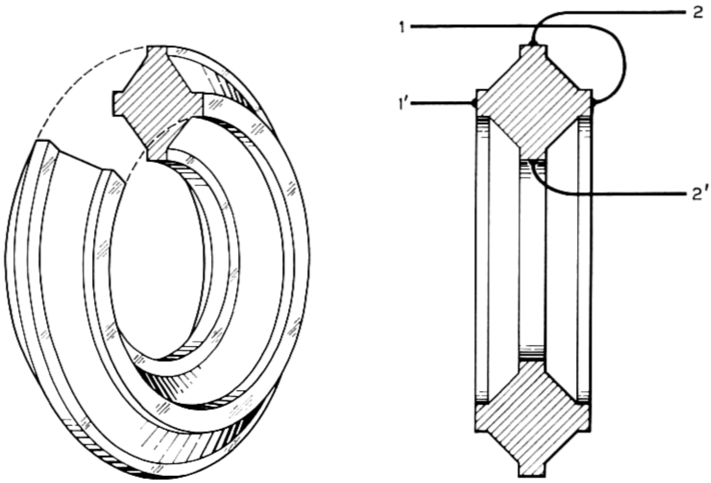

Fig. 10 — One-piece gyrator.

another where they attach to the sample. If the lead problem can be overcome, the devices will remain resistive (and essentially insensitive to frequency) until the relaxation frequency is approached. If leads are treated carelessly, the top frequency may be less than one megacycle.

For linear operation, the contacts must be as nearly ohmic as possible and contact resistance must be minimized. Nonlinearity might be a serious problem with any of the above devices if a wide dynamic range of operation were required.

If a good permanent magnet material were known (which also had a respectable Hall mobility) the above devices could be fabricated from a single piece of this material (see Fig. 10). The torus would be made of the unknown material, and it would be magnetized. The figure shows four contacts equally spaced about the torus' body, so the pictured device would be a gyrator. It could also be an isolator or a circulator. Two advantages of this structure would be that it would be all in one piece (for mechanical simplicity and rigidity) and that there would be no air-gap in the path of the magnet.

IV. SIGNAL-PRODUCED MAGNETIC FIELD DEVICES

4.1 *Switch*

Hall effect switches[5] for the most part are not simply single-pole single-throw switches. The simplest Hall effect switch might be termed a double-pole single-throw switch. This might be a gyrator sample in the gap of an electromagnet. Let us call the two terminal pairs of the gyrator 1 and 2. When there is no magnet current flowing, 1 is not connected to 2. But, with the magnet energized, 1 is connected to 2 and the loss introduced by the switch "contacts" is the insertion loss of the gyrator.

One can make a more intriguing switch by using an isolator instead of a gyrator. With no magnetic field, this is not very interesting, because 1 and 2 are then merely reciprocally connected very inefficiently. But with the proper value of H, 1 is connected to 2 but 2 is isolated from 1 ($1 \rightarrow 2 \nrightarrow 1$). Reversing the field reverses the direction of isolation. Furthermore, if the isolator is constructed from a gyrator plus negative resistances, so that the isolator is a unidirectional amplifier, switching the magnetic field changes the direction of amplification. It is as though a broadband amplifier could be physically turned around by energizing a coil in the opposite sense.

An even more elaborate switch involves switching a circulator. It was first proposed that a signal be applied to terminals 1 of a circulator and that it be switched between output 2 and output 3 by reversing the direction of H. However, when this is done, no less than three "isolator switches" are reversed at the same time because any two terminal pairs in the circulator make up an isolator.

Unfortunately, none of the above-mentioned switches are extremely fast in their operation. They all involve collapsing a rather sizable magnetic field and reproducing it in the opposite sense. This requires a particular amount of energy and, therefore, either a lot of time or a lot of power. There is a faster, more elegant means of switching, but it presents certain difficulties.

This second means permits the magnetic field to remain constant, but requires the resistance values in the parallel networks to change. The network values must be negative for at least one of the conditions and could be either positive or negative for the other. This might be achieved by using a fixed negative resistance in series with a positive resistance whose value could be shifted or switched from one desired value to another desired value. For example, one might use three resistors, one negative and two positive, and have a solid-state switch across one of the positive resistances. By shorting or opening the switch, the Hall effect switch could be operated.

Thus, an isolator switch could be reversed by opening or closing two electronic switches. A circulator switch could be operated by opening or closing three electronic switches. One calculation on a particular circulator indicates that, by changing the resistance values in C_p from -304 ohms to -168 ohms, the circulator could be switched. In this instance the forward loss in one condition is 4.5 db and in the other condition there is a forward gain of 3 db.

The reverse loss in each condition should still be of the order of 50 to 60 db. The only real difficulty is in obtaining the negative resistances. It should be noted that amplifier connections could be reversed with electronic switches, and the Hall effect devices would not be needed. Only the circulator switch performs a new function. Perhaps it would be useful.

4.2 *Transducer*

Hall effect transducers[3] can be constructed which convert mechanical motion into electrical signals. The mechanical motion moves a gyrator sample in and out of the air gap of a permanent magnet. Such a device could be useful in measuring strain or other displacements. It has been estimated (admittedly optimistically) that an InSb transducer could detect a displacement of about 1 angstrom if the gradient of H were 10 oersteds/micron.

The same device could be used as a phonograph pickup or as a microphone. It should be interesting here because it responds to very low frequency — even direct current.

A microphone and strain gauge have been constructed, and they behaved more or less as expected.

4.3 *Magnetic Field Meter*

The Hall effect has been used to measure magnetic field strengths for about ten years.[2,3,14,15,16] Since the output voltage of a Hall effect sample is proportional to the product of the sample input current I and the applied H, if I is constant the output voltage is proportional to H. For some years now, there has been a commercial instrument based on this principle on the market. An alternating current is applied to the sample so that the output is this ac carrier modulated by the magnetic field. Thus, either slowly varying direct current or low-frequency ac fields can be measured. The probe is about 25 mils thick (and could be made thinner) so that the field even in a thin air gap can be measured. Field strengths from a few oersteds up to 30,000 oersteds can be readily measured with a germanium probe. With care, even smaller fields are measurable. InSb can supposedly detect fields as low as 10^{-3} oersteds, but that sensitivity has not yet been achieved. However, an InSb sample can easily measure the earth's field.

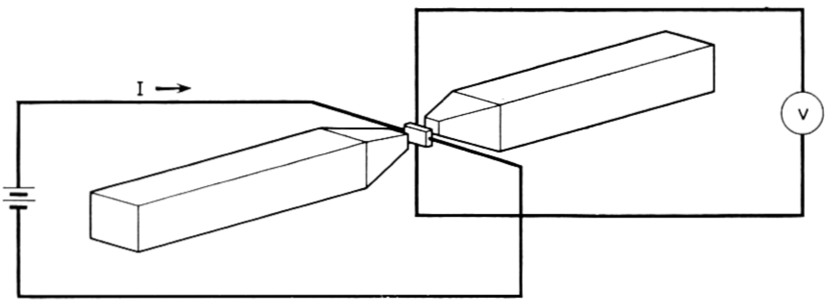

Fig. 11 — Electrical compass.

The Hall effect magnetic field meter is useful in measuring air-gap flux densities, flux distributions (because of its small size), demagnetization and hysteresis curves, and several other parameters. Its accuracy is inversely proportional to the temperature coefficient of carrier mobility in the semiconductor, so something like germanium should be of more general use than InSb. Nevertheless, InSb definitely gives a greater output for the same field, and its use is indicated where small fields or small variations are to be determined. It can be shown that

$$\frac{\text{power out}}{\text{power in}} \alpha H^2 {\mu_H}^2.$$

Thus, InSb has an obvious advantage when H is small.

4.4 *Electrical Compass*

A Hall effect electrical compass[3] can be constructed as shown in Fig. 11. The rods are made of high permeability material and serve to concentrate the earth's field on the sample, thus increasing the sensitivity. Since the output is proportional to $I \times H$, the reading will be maximum (say positive) when the rods are in line with the earth's field in one direction, and maximum negative when the rods are in line in the opposite direction. When the rods are perpendicular to H, there will be a null. Using an InSb sample, with no amplification, a rotation of 1° from the null position is detectable. With amplification, the sensitivity could be increased.

In this device, a change in temperature would only affect its sensitivity slightly if it were used at a maximum or minimum position.

4.5 *Magnetic Field Variation Meter*

If two Hall effect samples are connected in series as shown in Fig. 12, and if they are identical samples, there will be no output voltage except

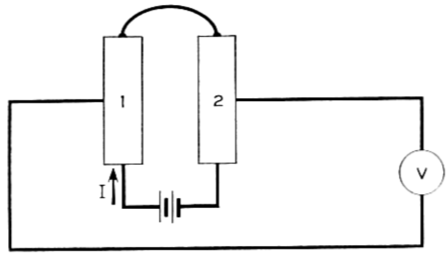

Fig. 12 — Magnetic field variation meter.

when the average field in sample 1 is different from the average field in sample 2.[17] If the field is uniform, the two outputs will cancel. Such a device is useful in determining magnetic field gradients, and has been used to detect cracks in metals. To do this, a permanent magnet produces a field in the metal, and where there are surface cracks (perhaps too small to be seen with the naked eye), there will be a stray magnetic field at the metal surface. Such a localized field can be readily detected with this type of device.

If a single Hall effect sample were used, a positive and then a negative pulse would appear in the output for each crack. With the two samples in series and properly spaced, the two pulses can be additive.

4.6 *Ammeter*

A Hall effect ammeter[3] can be constructed in at least a couple of ways. If one uses an ordinary gyrator (with a permanent magnet field) and connects the input leads in series with a current-carrying line, the gyrator's output voltage will be proportional to the line's current. If the current is too high for the gyrator sample, the gyrator input can be shunted with a calibrated resistor. One advantage of this is that it permits one to measure very high frequency currents. This hardly seems preferable to inserting a simple resistor in the line and measuring the voltage developed across it, but this Hall effect ammeter is mentioned because it will be referred to in the next section.

Another and perhaps more useful form of ammeter is shown in Fig. 13. A yoke of ferromagnetic material is hinged at the bottom so that it can be clipped around a current-carrying conductor. A Hall effect sample is in the magnetic path, with a direct current applied to two of its terminals. If there is either a direct or alternating current flowing in the conductor, there will be a corresponding output voltage proportional to the current, since the magnetic field is proportional to the conductor

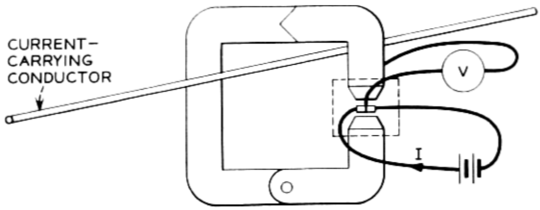

Fig. 13 — Ammeter.

current. This device should be useful when it is undesirable or impossible to break a line to insert an ordinary ammeter. It should be particularly useful for direct currents, because there is no similar instrument which measures direct currents.

4.7 *Wattmeter*

A Hall effect wattmeter[3,18,19] is easily obtained by using the first ammeter described above and employing the voltage between the two conductors to energize a coil wound on the magnetic core. Thus, the sample's input current is proportional to the circuit's voltage, and the output voltage is proportional to the power transmitted through the conductors. This is a simple wattmeter, but it is not useful at high frequencies.

A wattmeter can be used on a coaxial line by connecting the Hall effect sample from the center conductor to the shell so that the plane of the sample includes the axis of the line (see Fig. 14). Thus, the magnetic field is supplied by current in the central conductor, and the output voltage is proportional to the power.

Due to its high-frequency capabilities, the Hall effect has been proposed as a means of measuring power in waveguide circuits. In this case, the E field provides the current to the sample and the H field supplies

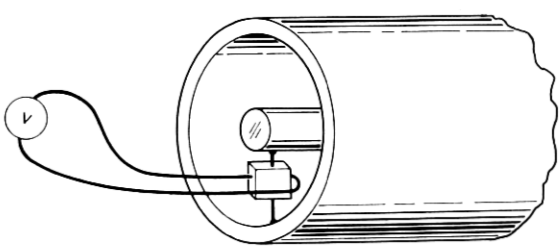

Fig. 14 — Coaxial line wattmeter.

the H field. The output leads are brought out on equipotentials through holes in the sides of the waveguide. It is theoretically possible to meter power in this manner to quite high frequencies.

4.8 *Amplifier*

If a direct current is supplied to the input of a Hall effect sample which is in the gap of an electromagnet, the setup can be an amplifier.[3,20] If the magnet is efficient enough, and if the sample's input current is large enough, the output power from the sample can be greater than the power applied to the magnet. With InSb, a gain of 5 has been achieved. Greater gains are probably attainable.

Such an amplifier could be used for direct or alternating current. The upper frequency limitation would be determined by the quality of the magnetic core and the stray capacitance of the core winding. This type of amplifier seems to offer little or nothing that a more conventional amplifier would not offer, but it is interesting in that it requires a low-voltage, high-current power supply — perhaps the type of thing thermoelectric generators could furnish.

4.9 *Modulator*

The Hall effect can be used to make a product modulator.[2] If one signal is applied to the input leads of a Hall effect sample and another signal is applied to a winding on a magnetic core to produce the magnetic field, then the output voltage is proportional to the product of the two signals. This type of modulator has been proposed as a substitute for the mechanical chopper found in most sensitive dc amplifiers. The output voltage of a straight dc amplifier drifts with time and temperature because of instability of the components. Mechanical choppers are used to convert the dc input to an ac signal, which in turn can be amplified with much less drift. The Hall effect modulator has an inherent drift-free zero set and can convert dc to ac with no mechanical motion. Thus, it should be longer lived than relay-type choppers.

Furthermore, mechanical choppers usually operate at 60 cps, whereas the Hall effect modulator frequency may easily be 1000 cps. This permits a broader band dc amplifier to be realized. Direct current signals as low as 20 microvolts (across 400 ohms input impedance) have been successfully amplified, and a dynamic range of 70 db has been observed. If additional effort were applied to reducing pickup in the output circuit, the 20-microvolt minimum could presumably be further reduced.

4.10 *Demodulator*

A Hall effect square-law demodulator or detector[21] is precisely the same as a device referred to as a Hall effect full-wave rectifier elsewhere.[3] If an incoming signal represented by $I = I_0 \cos \omega t$ is applied to both the Hall effect sample input leads and also to the magnet winding, the output voltage will be

$$V = \frac{k_1 I_0^2}{2} + \frac{k_1 I_0^2}{2} \cos 2\omega t, \tag{4}$$

where k_1 is a constant involving the Hall constant R_H and the various parameters of the magnetic circuit. The first term in (4) is a dc term proportional to the square of the amplitude of the input signal. The second is a double-frequency component which can be easily filtered out. Such a detector should give an accurate square conversion for a wide range of amplitudes.

A linear Hall effect detector[21] may be constructed on the same principle if a sine wave of constant amplitude (and the same frequency ω) is applied to the magnet winding. Then (4) becomes

$$V = \frac{k_2 I_0}{2} + \frac{k_2 I_0}{2} \cos 2\omega t. \tag{5}$$

The first term in (5) is proportional to the first power of the signal amplitude, and so this device should give a dc output voltage proportional to the input signal amplitude. Again, the conversion ratio should be linear for a wide range of amplitudes.

One difficulty with both these demodulators is that they are operable only up to frequencies at which magnetic fields of the order of 1000 oersteds or greater can be sinusoidally reversed with the power that happens to be available.

4.11 *Frequency Spectrum Analyzer*

A highly selective frequency spectrum analyzer[21] can be made from the linear Hall effect demodulator. If the signal applied to the sample is $I_s \cos \omega_s t$ and the signal applied to the magnet winding is of frequency ω_H, the usual sum and difference frequency components appear in the output:

$$V = \frac{k_2 I_0}{2} \cos (\omega_s - \omega_H)t + \frac{k_2 I_0}{2} \cos (\omega_s + \omega_H)t. \tag{6}$$

It may be noted that (6) reduces to (5) when $\omega_s = \omega_H = \omega$.

Suppose the signal applied to the sample is composed of many frequency components and that the amplitude of the magnetic field is held constant as its frequency is changed. If the output is applied to a dc voltmeter, there will be a reading only when ω_H coincides with some ω_s , say ω_{si} . Actually, the dc output is given by

$$V_{\mathrm{dc}} = \frac{k_3 I_{si}}{2} \cos (\Phi_H - \Phi_{si}), \tag{7}$$

where k_3 is a constant and $(\Phi_H - \Phi_{si})$ is the difference between the phases of the magnetic field and the ith component of the input. Thus, unless the phase difference is known, it is impossible to determine I_{si} . Therefore, it is recommended that ω_H be adjusted to within about 1 cps of ω_{si} so that the "dc" output is a 1-cps cosine wave whose argument $(\Phi_H - \Phi_{si})$ either increases or decreases 2π radians per second. If this output is applied to a zero-centered dc voltmeter with a response time $\tau < 0.25$ second, the meter needle follows the voltage variation, and the amplitude of the ith component is proportional to the maximum swing of the needle.

Suppose the frequency of the next component is ω_{sk} . If $(\omega_H - \omega_{sk})$ is large enough so that

$$\frac{2\pi}{\omega_H - \omega_{sk}} < \frac{\tau}{10}$$

then the meter will not respond to this component's contribution to the output. Thus, by slowly sweeping ω_H , a frequency spectrum analysis of any input wave can be obtained.

This device responds with equal accuracy to both small- and large-amplitude components. Its upper frequency is limited, because a sinusoidal magnetic field must be obtained at that frequency. Fortunately, the signal does not have to supply this field; a high-power oscillator may be used. Since the available output level is determined by the component amplitude and the value of magnetic field, the upper frequency will be determined by the required sensitivity. Perhaps its upper limit in a typical use would lie between 0.5 and 5 mc.

Of course, it operates very well down to zero frequency, and this is probably where it will find its greatest use — at audio frequencies and even lower. Conventional analyzers require huge tuned filters at these frequencies to obtain the Q required for high selectivity. With the Hall effect analyzer, all that one requires is a single low-pass filter. Notice that the amplitude of the output is independent of the frequency difference $(\omega_H - \omega_{si})$. Thus, a single low-pass filter of adjustable upper frequency gives one a very simple means of adjusting selectivity without affecting the sensitivity.

With a crude device like this, it was quite easy to measure the first ten harmonics of a square wave of 25 cps fundamental frequency. With care, 18 harmonics were detected. This device does not represent the ultimate in accuracy, but it may do so in small size, simplicity and ease of operation.

4.12 *Phase Discriminator*

The Hall effect phase discriminator simply makes use of the phase sensitivity of the spectrum analyzer. It is sometimes necessary to obtain an electrical signal which is a measure of the difference between the phases of two signals of the same frequency. If these two signals are applied to the spectrum analyzer, the output is a dc component plus a double-frequency component [see (5)]. The dc term is given more completely by (7). If the two amplitudes are constant (not necessarily equal), the dc output can be calibrated to give the phase difference directly. Like the analyzer described above, the frequency limit will be determined largely by the required sensitivity. Perhaps this limit will be somewhere between 1 and 10 mc. This is estimated to be higher than that for the analyzer, on the assumption that the signal level can be higher in the discriminator (since it needs to operate at only one level).

Incidentally, this phase sensitivity must be considered if one constructs the wattmeter or demodulator previously referred to.

4.13 *Digital-to-Analog Encoder*

The Hall effect digital-to-analog encoder is obtained by applying the maximum allowable dc voltage to the sample input leads and using n separate windings (to encode n binary digits) on the magnetic core (see Fig. 15). Possibly one extra winding would be used for zeroing purposes. Current limiters assure that the current input levels are fixed, all at the same level; the digits are weighted by different numbers of turns in the windings. If the current limiters do not pass precisely the desired current the situation might be improved by adjustment of turns. All the input circuits may be completely isolated in a dc sense. A disadvantage is that the minimum time in which the magnetic field can be changed is limited to something on the order of a millisecond, so that only about 1000 binary numbers per second may be applied. Since this is the same setup as the Hall effect amplifier, it is conceivable that the encoder might introduce gain.

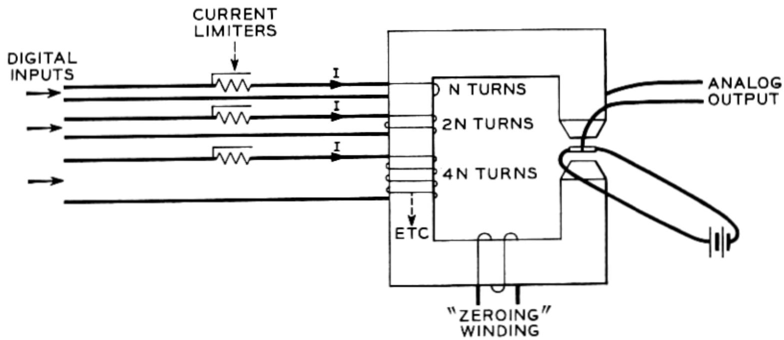

Fig. 15 — Digital-to-analog encoder.

4.14 *Analog Multiplier*

The Hall effect analog multiplier[3,20] is essentially a Hall effect demodulator. A direct current is driven through the sample input leads and another direct current is applied to a winding on the core. The output voltage is proportional to the product of these two signals. Thus the device is an analog multiplier. Ref. 22 concluded that silicon offers the best combination of properties for this device. In order to realize a reasonable accuracy in the output (about 0.3 per cent) it was necessary to have a response time of approximately 1 millisecond. The error was reduced slightly (to about 0.1 per cent) when the input repetition rate was decreased to 20 cps. Ref. 22 also presents a scheme for minimizing temperature effects.

V. MATERIALS

Strange as it may seem, all of the above devices place approximately the same requirements on the semiconductor material used. Generally speaking, the ideal material would have a high-electron Hall mobility and a large energy gap, and both these parameters would be constant.

If a material with these properties were available, minimum losses would be realizable with rather small magnetic field strength — say, 1000 oersteds. Furthermore, sample resistances would be fixed by their geometry and doping level. Of course, other requirements could be laid down, but they are certainly of less importance than the two given above.

Now, these two requirements are mutally incompatible. That is, a material with a high mobility μ_H is almost certain to have a small energy gap, Eg, and, *vice versa*, a material with a large Eg will have a small μ_H. If this relationship did not exist, many Hall effect devices would find much wider application than they now enjoy.

Intermetallic semiconductors offer the best hope of a partial solution to this problem. They show a wide range of properties but still follow the trend of combining small Eg with high μ_H. One material however, gallium arsenide, has a mobility of 6000 cm²/volt-second (higher than that of germanium) and an energy gap of 1.35 electron volts (higher than that of silicon). Let us hope that other materials will be found which will violate the ($\mu_H - Eg$) trend even more definitely and that they will be relatively easy to fabricate in highly pure single-crystal form. Refs. 23, 24 and 25 describe a large number of semiconductors (both elemental and intermetallic) and their properties and refer to the apparent relationship between μ_H and Eg. Ref. 26 gives experimental information on mercury selenide, a II–VI compound.

Before leaving materials it is worth mentioning that Ref. 5 develops an elaborate mathematical solution to the problem of finding the electric field distribution in a Hall effect sample. Several different shapes are treated, including the square gyrator, the skew isolator and the circulator. Ref. 27 suggests a method of measuring resistivity and Hall effect in flat samples of any arbitrary shape. Ref. 28 gives an analog method for obtaining the same two parameters; this last method involves no analytical computation.

VI. CONCLUSIONS

The Hall effect has thus far furnished a tremendous variety of devices. All of these devices have at least one important advantage over most semiconductor devices — Hall effect devices require the use of majority current carriers only and therefore involve no junctions. There are no areas of concentrated electric fields, so the surfaces need not be elaborately protected. Most or perhaps even all of the devices can be operated indefinitely in ordinary indoor atmospheres with no protection whatever.

Because these are majority carrier devices, the lifetimes of minority carriers are of no consequence. When an electric field is induced in a semiconductor, a brief time elapses while the current carriers redistribute themselves. This relaxation time (ordinarily less than a millimicrosecond) is the only factor which limits the frequency of signals which can be satisfactorily transmitted through a Hall effect sample. Interlead capacitance will tend to shunt part of the signal around the sample and will thus limit the frequency range of operation more severely. Of course, where a signal supplies the magnetic field, the frequency limitations will depend on the core used, the winding, the signal level and the sensitivity and linearity required.

The Hall effect is a small effect. As a result, the devices it makes possible are inefficient or lossy. In order for Hall effect devices to be stable, the majority carrier mobility must be held constant, as must the magnetic field (or core permeability as the case may be). In order that these devices have linear characteristics, all current-carrying contacts must be ohmic.

Despite the disadvantages outlined in the preceding paragraph, Hall effect devices are being studied with steadily increasing interest.

REFERENCES

1. Carlin, H. J., Synthesis of Nonreciprocal Networks, Proc. Symp. on Modern Networks, Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., April 1955.
2. Mason, W. P., Hewitt, W. H. and Wick, R. F., Hall Effect and "Gyrators" Employing Magnetic Field Independent Orietations in Germanium, J. Appl. Phys., **24,** 1953, p. 166.
3. Ross, I. M. and Saker, E. W., Applications of Indium Antimonide, J. Elect., **1,** 1955, p. 223.
4. Tellegen, B. D. H., The Gyrator, A New Network Element, Phil. Res. Rep., **3,** 1948, p. 81.
5. Wick, R. F., Solution of the Field Problem of the Germanium Gyrator, J. Appl. Phys., **25,** 1954, p. 741.
6. Casimir, H. B. G., On Onsager's Principle of Microscopic Reversibility, Rev. Mod. Phys., **17,** 1945, p. 343.
7. Meixner, J., Zur Theorie der elektrischen Transporterscheinungen im Magnetfeld, Ann. der Phys., **40,** 1941, p. 165.
8. McMillan, E. M., Violation of the Reciprocity Theorem in Linear Passive Electromechanical Systems, J. Acoust. Soc. Amer., **18,** 1946, p. 344.
9. McMillan, E. M., Further Remarks on Reciprocity, J. Acoust. Soc. Amer., **19,** 1947, p. 922.
10. Hogan, C. L., The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications, B. S. T. J., **31,** 1952, p. 1.
11. Shockley, W. and Mason, W. P., Dissected Amplifiers Using Negative Resistance, J. Appl. Phys., **25,** 1954, p. 677.
12. Semmelman, C. L., U. S. Patent No. 2,774,890.
13. Grubbs, W. J., The Hall Effect Circulator — A Passive Transmission Device, Proc. I.R.E., **47,** 1959, p. 528.
14. Hennig, G. R., Applying the Hall Effect to Practical Magnet Testing, Elect. Mfg., **59,** 1958, p. 132.
15. Pearson, G. L., A Magnetic Field Strength Meter Employing the Hall Effect in Germanium, Rev. Sci. Inst., **19,** 1948, p. 263.
16. Saker, E. W., Cunnell, F. A. and Edmond, J. T., Indium Antimonide as a Fluxmeter Material, Brit. J. Appl. Phys., **6,** 1955, p. 217.
17. Wolfe, R., private communication.
18. Barlow, H. E. M., The Application of the Hall Effect in a Semiconductor to the Measurement of Power in an Electromagnetic Field, Proc. Inst. Elect. Eng., **102B,** 1955, p. 179.
19. Barlow, H. E. M., The Design of Semiconductor Wattmeters for Power-Frequency and Audio-Frequency Applications, Proc. Inst. Elect. Eng., **102B,** 1955, p. 186.
20. Ross, I. M. and Thompson, N. A. C., An Amplifier Based on the Hall Effect, Nature, **175,** 1955, p. 518.
21. Bogomolov, V. N., Some New Semiconductor Devices (New Uses of the Hall Effect), Zh. Tekh. Fiz., **26,** 1956, p. 693.
22. Löfgren, L., Analog Multiplier Based on the Hall Effect, J. Appl. Phys., **29,** 1958, p. 158.
23. Coblenz, A., Semiconductor Compounds, Electronics, **30,** 1957, p. 144.
24. Pincherle, L. and Radcliffe, J. M., Semiconducting Intermetallic Compounds, Adv. in Phys., **5,** 1956, p. 271.
25. Stello, P. E., Van Winkle, D. M. and Turner, J. D., Semiconductors, Materials and Properties, WESCON Conf., August 1956.
26. Blum, A. I. and Regel, A. R., Electrical Properties of Solid Solutions of Mercury Selenide and Selenium, Zh. Tekh. Fiz., **21,** 1951, p. 316.
27. Van der Pauw, L. J., A Method of Measuring Specific Resistivity and Hall Effect of Discs of Arbitrary Shape, Phil. Res. Rep., **13,** 1958, p. 1.
28. Broudy, R. M., Galvanomagnetic Coefficients for Arbitrary Geometry, J. Appl. Phys., **29,** 1958, p. 853.