Some Experimental and Theoretical Observations on a Configurational EMF

Thesis by

Marvin Chester

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1961

Acknowledgments

The author wishes to thank Professor John R. Pellam for his guidance and for his helpful critical evaluations of the experimental investigation during many of its crucial phases. It was Professor Pellam's work on the hydrodynamic properties of superfluid helium that originally stimulated the author into thinking about possible hydrodynamic measurements on the electron fluid in metals. These thoughts led directly to this project. It was also Professor Pellam who first directed the author's attention to the possible magnetic explanation of the observed configurational emf.

The author wishes also to thank Professor Richard P. Feynman for his interest and help on the theoretical aspects of this project. The detailed theoretical calculations were done under Professor Feynman's guidance and it was he who first discovered and recognized the significance of the energy relaxation time τ_o.

Abstract

The basic principles are outlined concerning the existence of a configurational emf. Such an emf is expected from energy conservation principles as applied to the flow of the electron gas in the form of an electric current. The experimental details are described relating to the effort to detect this effect in thin films of bismuth. Quantitative results of this effort are discussed. A theoretical analysis, with application to bismuth, follows from which are derived the equations describing the phenomenon. These equations are discussed with respect to the experimental results.

TABLE OF CONTENTS

I. INTRODUCTION

The phrase "electron gas" has been a familiar one to physicists for over half a century. It was first introduced by Drude (1) in 1900 to explain the apparent transport properties of the electrons in a metal. These properties are evidenced by the thermal and electrical conductivity of metals. The theoretical description of the properties of the electron gas and of the internal collision mechanisms which result in a mean free path or mean collision time have undergone considerable metamorphosis with the advent of quantum mechanics. But the picture of a gas of relatively free electrons bound to the macroscopic volume of the solid rather than to individual ion cores has survived as a useful physical model.

The phrase "electron gas" and the accompanying physical picture suggest the possible existence of electronic fluid hydrodynamic properties in addition to the well-known dissipative transport properties. In particular, one is led to speculate on the possible existence of a Bernoulli effect in electron flow.

In classical fluids this effect manifests itself as a drop in pressure in a constricted region of flow. The physical basis for this drop in pressure lies in the conservation of mechanical energy. A decrease in the potential energy of the fluid is required to balance the increase in kinetic energy. The increase in kinetic energy is

associated with the higher flow rate in a constriction. The pressure, being a measure of the potential energy, must be lower in the constricted region of flow where the kinetic energy is large.

In a classical fluid the presence of dissipation effects due to collisions of the particles of the fluid with the walls of the container does not erase the Bernoulli effect. Rather, it tends to mask it. The dissipative effects, as manifested in the viscosity of the fluid, add to the inertial ones in determining the motion of the fluid. However, if the viscosity is high, clearly, the dissipative effects will dominate in determining the motion. The inertial effects, although still present, will be of higher order relative to the dissipative ones. For example, in a viscous fluid there must be a pressure gradient throughout to maintain a constant flow rate. If the flow is caused to move through a constricted region there must be an additional gradient in the constricting region to accelerate the fluid into the constriction. Depending upon the viscosity of the fluid the latter gradient may be quite small compared to the viscous pressure gradient which must exist just to keep fluid moving. The Bernoulli gradient accelerates the fluid and that is therefore here called "inertial." The viscous gradient is proportional to the flow velocity and merely counteracts the dissipative effects in the fluid which tend to stop the flow. Since by the conservation of mass or the continuity equation the fluid must

move faster in a constricted region there must be an accelerating pressure gradient in the constricting region to speed up the flow. Regardless of the magnitude of the viscosity, this accelerating pressure gradient is present and simply adds to the velocity-proportional gradient which maintains the flow in the face of the dissipative effects of viscosity.

It seems reasonable that a similar analysis should apply to electron gas flow. In the theoretical section a detailed mathematical description of the problem will be given which indicates that this is, in fact, the case. An indication of the physical reasoning which underlies the analysis may be obtained from the following considerations.

In the electron gas of a metal the dissipative effects dominate over the inertial ones. Unlike a viscous fluid, however, the dissipation arises from the transfer of momentum and energy to the lattice vibrations rather than to the walls. The electrons collide with phonons and lattice imperfections predominantly. Regardless of the nature of the collisions, however, the two fluids have in common the need for the presence of a force merely to maintain the flow or current. The force is needed to counteract the continual scattering of the forward momentum into random motion by the collisions. In the electron fluid this force is the applied electric field which by Ohm's law is

proportional to the flow velocity or to the current. The electric field in the electron gas is the analogue of the pressure gradient in viscous fluid flow.

Suppose that the electric current flow passes through a constricted region. By virtue of the equation of continuity and the relative incompressibility of the electron gas the flow velocity of the electrons must be greater in the constricted region. The incompressibility property of the electron gas is imposed by the uniform background density of positive charge from the ion cores. They keep the electron density uniform on penalty of building up high internal electric fields in the metal. Since the current must be higher in a constriction as indicated by the equation of continuity, and the density of electrons must be relatively constant because of the positive ion core background, it follows that the electron flow velocity must be higher in a constricted region of flow.

Now in analogy with the hydrodynamic problem the question is posed: What accelerates the electrons over the constricting or narrowing region so that they move faster, as they must, while traveling in the constriction? Unless there is some unusual interaction between the lattice and the electrons in the constricting region so that the lattice provides the flowing electrons with the accelerating kick which they need to move faster, there is no solution but that there

must be an extra inertial field present to provide this acceleration. As in the case of viscous fluid flow this field is superimposed on the Ohmic one already present to counteract the electrical resistance of the metal.

An idealized expected potential distribution is indicated in figure 1. The step shown in the potential at the constriction corresponds to the inertial impulse field which must accelerate the electrons from the low velocity they have while moving in the wide region to the higher flow velocity characteristic of the narrow region.

The whole concept may be summarized in the statement that Ohms law is approximate. A more complete statement of Ohms law should include an inertial term. The modified single parameter law might read

$$\mathcal{E} = const. \times \frac{dj}{dt} + \frac{1}{\sigma} j \qquad (1)$$

where $\mathcal{E}$ is the electric field, j the current density and σ the conductivity. The first term on the right constitutes the additional inertial term which should be included and comes from a simple statement of Newton's law. In dealing with a steady state none of the quantities are time varying so that the equation becomes

$$-\frac{dV}{dx} = C \frac{d}{dx} j^2 + \frac{1}{\sigma} j \qquad (2)$$

Here the inertial term takes the form of a spatial derivative because of the apparent time variation of current seen by the flowing electrons due to their spatially non-uniform drift. It is based on the assumption that the current density, j, is proportional to the flow velocity of the electron gas. The first term on the right yields the configurational emf on constricting the current flow in exact analogy to the classical hydrodynamic case. Clearly a first guess regarding constant, C, of equation 2 follows from application of Newton's law to the electrons. This yields

$$C = -\frac{1}{2}\frac{m}{n^2 e^3} \tag{3}$$

where n is the number of electron current carriers per unit volume, e is the electronic charge and m is the electronic mass. This estimate is not unreasonable providing the quantities n and m are properly interpreted.

It should be mentioned here that equation 1 does not represent the only possible modification of Ohm's law which might conceivably give rise to a configurational emf. At least one other reasonable modification might cause such an emf. This revision stems from the idea that the current density should be guided by the total electric field seen by the current carriers. They see the applied field, $\mathcal{E}$, plus the magnetic deflection field $u \times B$. The vector $\vec{u}$ represents

the flow velocity, and is related to the current density j.

$$\vec{u} = \frac{1}{ne}\vec{j} \tag{4}$$

and B is the magnetic field due to the current flowing in the medium. This magnetic field is related to the current density through one of Maxwell's equations

$$\mu_o \vec{j} = \nabla \times \vec{B} \tag{5}$$

where the coefficient μ_o is the permeability of free space. The further modification of Ohm's law is contained in the equation

$$\vec{j} = \sigma\left(\vec{\mathcal{E}} - \frac{1}{ne}\vec{j}\times\vec{B}\right) \tag{6}$$

The net two-fold effect of this self-magnetic deflection emf is as follows: (a) It causes a slight lateral build-up of charge on the surface and in the core of a current carrying cylinder which charge provides a field that cancels the magnetic lateral deflection field. (b) This surface charge permits the current flow lines to cross the electric field flux lines at a small angle everywhere in the current carrying region so that there is a small component of electric field perpendicular to the current flow which just balances the magnetic deflection force on the current carriers.

Both of the modifications of Ohm's law - those of equations 1 and 2 and that of equation 6 - follow from the application of the Boltzmann transport equation to the conduction electron distribution in a metal. In a succeeding section a detailed analysis will be made to derive these modifications using the Boltzmann transport equation and the proper statistics for an electron gas in a metal. The foregoing arguments are not meant, therefore, as proofs but have been given rather to indicate the motivation for this research project and the general lines along which work was undertaken. The actual proofs lie in the detailed theoretical analysis and in the experimental verifications, discussions of which will follow.

Before continuing with the more esoteric features of the discussion, however, two very elementary preliminary calculations will be made. They will be useful as a basis for comparison throughout the pages that follow.

Using the simple hydrodynamic analogue point of view, as exhibited by equations 2 and 3, it is evident that the voltage, ΔV, of figure 1 may be calculated from

$$\Delta(-eV) + \Delta\left(\frac{1}{2} m u^2\right) = 0 \qquad (7)$$

The speed of flow u is proportional to the current density j as shown in equation 4. If the current density in the wider area of flow is effectively zero in comparison with that in the narrower constricted

region - and this will generally be the case in the experimental arrangements - then the configurational voltage due to this hydrodynamic cause may be written

$$\Delta V = V_c = - C \frac{I^2}{A^2} = \frac{1}{2} \frac{m}{n^2 e^3} \frac{I^2}{A^2} \tag{8}$$

where Vc refers to the hydrodynamic electrical potential in the narrow region relative to that in the unconstricted region. In equation 8, I is the total current and A is the cross-sectional area of the constricted region of flow.

The second calculation concerns the apparent configurational emf due to the self-magnetic field force shown in equation 6. The net effect of this term is to cause the equipotentials to have a slight curvature even in a region of uniform current flow. This characteristic is shown in figure 2. The curvature will be negligible in the wider flow regions compared to that in the constricted region where the magnetic fields are higher. An infinitely thin potential - sensing probe which touches the surface of the conductor in the constricted region will be at an electrical potential which differs from the expected ohmic potential by the amount

$$V_m = - \frac{1}{ne} \int_0^R \vec{j} \times \vec{B} \cdot d\vec{r} \tag{9}$$

The potential Vm is that at the surface of the conductor - where the probe makes contact with the conductor - relative to that at the center of the conductor where the potential is just the ohmic one. This potential difference is caused by a lateral charge distribution, the field of which just cancels the effect of the magnetic deflection forces on the current carriers. The distance R is the radius of the conductor and the integral is taken from the center of the conductor where the magnetic field is zero along a path perpendicular to the current flow lines out to the point where the idealized potential probe meets the conductor.

For the case of the right circular cylinder of figure 2 the integral of equation 9 reduces to

$$V_m = \frac{\mu_o}{4\pi} \frac{1}{en} \frac{I^2}{A} \tag{10}$$

where μ_o is the permeability of free space. Because in a significantly wider region of flow the potential, Vm, between the center and the surface of the flow will be negligible compared with that difference in a narrow flow region the entire configurational emf due to this "self-Hall-effect" will in general be given by

$$V_m = \gamma \frac{\mu_o I^2}{4\pi e n A} \tag{11}$$

The quantity γ is introduced as a dimensionless geometrical factor

to account for the dependence of this voltage upon the shape of the specimen cross-section. The factor γ is unity, of course, for a right circular cylinder. As in the hydrodynamic case the voltage Vm will be the potential of the narrow region probe with respect to a probe at the wider region of flow. The area A in the formula represents the cross-sectional area of the constricted region of flow.

The essential quality which makes the configurational emf detectable is its dependence on the square of the current. It is true both for the magnetic and for the hydrodynamic effects that the associated configurational emf has the same polarity regardless of the direction of current flow. The polarity of these emf's depends only upon the sign of the charge of the current carriers. The basic equations 8 and 11 have been derived assuming the carriers to be negatively charged.

The constant polarity makes the experimental detection of the configurational emf quite simple in principle. It is necessary only to send an a.c. signal through a stepped specimen like the one of figure 1 and to put a d.c. voltmeter across the step. The ohmic part of the potential drop is not recorded by the d.c. meter because its average is zero due to polarity reversal each half cycle. This meter will therefore read the configurational emf directly since its average is not zero.

II. EXPERIMENTAL PARTICULARS

1. Specimen Material

The primary experimental objective was the detection of the existence of a configurational emf. Therefore the major consideration regarding specimen material revolved about obtaining as large a signal as possible. Equation 8 clearly indicates that, insofar as the specimen material is concerned, it is desirable to have n as small as possible. Among the metallic elements cesium has the largest atomic volume, i.e., 71 cm^3/gm. atom (2). If n is to be interpreted as the literal number of valence electrons per unit volume then cesium would be the best specimen material to use. It may be deduced from equation 8 that the order of magnitude of the expected result for cesium would be

$$V_c(Cs) = 1.5 \times 10^{-22} j^2 \qquad (12)$$

where Vc is in volts if the current density j is in amperes/cm^2. For this case then one would expect an emf of only about 10 millimicrovolts for a current density input as high as 10^7 amperes/cm^2. Such high current densities are obtainable under proper conditions. And equipment has been developed to detect signals as small as 10^{-11} volts. It is clear, however, that the technological task of doing the

experiment in an ordinary conductor would be formidable.

If n referred to the effective number of current carriers present in the material rather than to the total number of valence electrons then new possibilities present themselves. In semiconductors, for example, the effective number of current carriers, as deduced from the Hall coefficient, may be quite small. In terms of the Hall coefficient for ideal metals equation 8 may be written

$$V_c = \frac{1}{2} \frac{m}{e} H^2 j^2 \tag{13}$$

where $H = -1/en$ is the Hall coefficient.

Although the Hall coefficient for cesium is only 8×10^{-4} cm^3/coulomb (2), that for bismuth is one thousand times larger. For some semiconductors at low temperatures the Hall coefficient may reach 10^7 cm^3/coulomb or higher. Of all the metals, however, it is clear that bismuth should give the largest effect with antimony running a close second. For bismuth the Hall coefficient is about 0.6 cm^3/coulomb. The expected order of magnitude of the configurational emf for bismuth would be

$$V_c(Bi) = 10^{-16} j^2 \tag{14}$$

where again j is an amperes/cm^2 and Vc is in volts. A current density of 10^5 amperes/cm^2 would be necessary to produce a microvolt

signal. Experience with tin whiskers had shown that current densities up to 10^7 amperes/cm^2 can be sustained in filaments of 0.5 square micron cross-section. With these facts in mind, in conjunction with considerations on the relative ease of handling, the element bismuth was chosen for preliminary investigation.

2. Specimen Design

The easiest way to obtain the small sample dimensions which are required is through the evaporation of thin films. Bismuth lends itself nicely to this process. It vaporizes at a relatively low temperature. The bismuth was evaporated from a molybdenum boat in a vacuum of about 5×10^{-5} mm of mercury. The bismuth beam fell onto a glass microscope slide through a mask which determined the physical shape of the final specimen.

In the regions where external electrical contacts were to be made hot indium was bonded to the glass before evaporation. The bismuth was then evaporated through the mask onto the glass and onto the indium smears. Wires were indium-soldered to the smeared regions after the evaporation was completed. Figure 3 shows a typical specimen evaporation on a microscope slide with the evaporation mask next to it.

The evaporation layer depth was estimated generally to be of the order of 100 Å. The width of the constricted region varied among

the various voltage sensing points from 20 μ to 50 μ. In effect, then, the constricted cross-sectional area was of the order of $3 \times 10^{-9} cm^2$.

The evaporation mask was constructed out of a 40 mil thick brass plate. The wider sections were cut through the plate on a milling machine. The constricted sections required more elaborate technique. The constriction slits were of the order of 1 mil wide.

The process used to make the slits was simply to scribe a "V" cut into the plate with an end-mill held at a 45° angle. The cut was made as deeply as possible but care was taken to avoid penetration through the plate at the bottom of the "V" cut. The end-mill was allowed to penetrate only to a depth such that the bottom of the "V" cut lay 2 or 3 mils above the underside of the brass plate. The final cut was accomplished chemically.

A slow etch of dilute HNO_3 was prepared. A volume ratio of about 2 parts of water to 1 part of concentrated HNO_3 removed brass at a rate slow enough so that the process was easily controlled. The slit was made by very careful local etching with continual inspection of progress under a microscope. In the final stages one drop of etch solution at a time was applied to the brass masking plate which was set under the microscope. This way the chemical milling could be stopped as soon as the etch broke through. One of the brass masking plates is

shown in figure 3, together with a sample evaporated specimen. The constriction slits of this particular mask have been widened in order to make them clearly visible in the photograph.

Figure 4 shows a schematic diagram of the specimen configuration. Those contact points marked "C" are for the current leads. The contact points marked "M" are potential probe contact points for sensing the constricted region potentials. And the "E" contacts are for potential probe leads to the unconstricted flow region. In principle there should be a configurational emf between any "E" and any "M" contact but none between any pair of "E" contacts or pair of "M" contacts when a signal is applied between the points "C". If the constricted flow region has a variable width so that the "M" potential sensing arms intersect three different widths, then, of course, there would be a configurational emf between each of the "M" arms. There is a simple additive relation, in this case, between the "M-M" potentials and the "E-M" configurational potentials (cf. fig. 8).

Particular note should be taken of the placing of the wire lead contact points. All external wire leads meet the specimen at regions of wide extended area. Because of the relative massiveness of these regions they act as heat sinks. They are, therefore, all at the same temperature even though the constricted current flow region may be at quite a high temperature due to Joule heating. As a result the net thermal emf between any two of the external lead contact points is

zero. This follows from the temperature equality of those points and from the fact that there is no change in material composition between those points. The design of the specimen eliminates possible interference with the desired signal from thermal voltages.

3. Circuitry

As has been mentioned, the principle of the detection procedure is quite simple. However, there are a number of extraneous effects which must be nulled out to insure that the detected signal is the configurational emf. Therefore the circuit used was the bridge arrangement of figure 5.

The object of the particular circuit shown in figure 5, in addition to detecting the configurational emf, is twofold. First, the presence of possible rectifying contacts in the bridge circuit is revealed by alternating between the high (100 K) and the low (10 K) bridge balancing resistances. Secondly, a preliminary rough a.c. balance using the a.c. detector eliminates possible spurious signals in the d.c. electronic microvoltmeter. These signals arise in the presence of large a.c. voltages.

In principle, for sufficiently low frequencies, if the bridge is balanced at the signal generator frequency it would be balanced for all low frequencies down to d.c. provided all the bridge arms are ideal

pure resistances. Suppose that some part of the circuit inside the bridge contains an element whose voltage-current relation is not linear. If this non-linearity were odd in the current then the bridge would be found to be unbalanced at odd multiples of the signal generator frequency. If the non-linearity were even in the current the bridge would be unbalanced at even multiples of the input frequency and unbalanced at d.c. A non-linearity outside of the bridge circuit proper - as, for example, may be present in the signal generator - does not affect the balance in the limit of low frequencies. To operate at frequencies up to several kilocycles it is necessary to approximate a Maxwell bridge (3) by inserting some capacitors, variable between 100 uuf and 1000 uuf, across each leg of the balancing resistance. A Maxwell bridge is characterized by the fact that the balance condition is independent of the frequency.

The configurational emf constitutes an even non-linearity in the voltage-current curve of the specimen. It acts, essentially, like a variable emf battery situated at a region of constricting current flow. The emf is variable by virtue of its dependence upon the magnitude of the applied current. There are two such configurational batteries in the bridge circuit, one at each end of the narrow constricted current flow region of the specimen. They are so oriented that they are in parallel across the detector. For the polarity of the d.c. meter shown in figure 5 a positive signal would indicate negative carriers.

If there are any other even non-linear elements inside the bridge circuit then they too would contribute a signal which may be confused with the configurational signal. Because of the expected isotropy of the specimen and because of its design, no even non-linearity in the specimen proper other than the configurational one is possible. In the rest of the bridge circuit and especially in the contacts it is not improbable that there may be even non-linearities. For example, the phenomenon of non-ohmic rectifying contacts is familiar in solid state physics. The presence of rectifying contacts internal to the bridge circuit would certainly cause a spurious d.c. signal in the detector.

In order to insure that the signal comes from the body of the specimen and not from elsewhere a special procedure is used. The bridge is first balanced at the signal generator frequency using the higher resistance (100K) balancing potentiometer. The signal is then picked up on the d.c. detector. It is observed to increase with increasing current from the signal generator. Now the bridge is rebalanced at the input signal frequency using the lower resistance (10 K) balancing potentiometer. The signal generator is adjusted so as to drive the bridge with the same total current as drawn with the previous balance. If the observed d.c. signal is less than that obtained with the previous balance then it comes from the specimen because there is now less

current going through the specimen. On the other hand an increased output signal indicates that the signal is coming from the specimen contacts or from some other point of the bridge external to the specimen proper because that is the part of the bridge which is now drawing more current. This whole procedure can be put on a quantitative basis, of course, if the specimen resistance is measured preliminarily. One can predict the change which should take place in the observed signal when the balancing potentiometer is changed. Deviations from this prediction indicate the extent to which the desired signal is being obscured by other non-linearities in the bridge. If necessary, a lengthy subtraction procedure may be used on the data to subtract out the specimen-external signal from the total signal. In practice, however, it was never found necessary to do this.

In the few cases where the interchanged-balancing-potentiometer test showed the presence of interfering signals these were easily eliminated by checking and repairing the specimen contacts. In general, this test showed quite clearly that the detected signals were indeed emanating from the specimen proper.

4. Experimental Controls and Procedure

The procedure for taking data was as follows: After all connections are made and the interchanged-balancing-potentiometer test

is carried out the larger of the potentiometer resistances is switched into the circuit. The bridge is balanced at the signal generator frequency which was generally 400 cps. The balance detector was either a General Radio amplifier-detector with an auxiliary narrow band filter centered at 400 cps or a wide band Hewlett-Packard electronic voltometer. After the variable bridge condensers and the bridge potentiometer are set for an a.c. minimum the d.c. detector is switched into the circuit. This detector is a Hewlett-Packard d.c. electronic voltmeter. The voltage picked up on this meter corresponds directly to the configurational emf in the case of the circuit of figure 5. If the resistance of the balancing potentiometer is sufficiently greater than that of the specimen, the current through the specimen is essentially that which is read on the signal generator ammeter. The expectation is that the d.c. signal as a function of the ammeter reading should correspond to equation 8 if the ammeter reads the r.m.s. value of the current.

In general, for most well-behaved specimens, the square law dependence of equation 8 was indeed observed. A summary of the data taken on one such specimen is shown in figure 6. In this graph the observed emf is plotted against the square of the specimen current. The data were taken over a range of signal generator frequencies and over several days.

The specimen resistance in this case was 5.1×10^3 ohms which was about equally divided on both sides of the center tap. With this knowledge the fraction of the total ammeter current which is channeled through the specimen is calculable for the two different balancing potentiometers. Points are plotted on the graph for both the case of the 10 K balancing potentiometer and the 100 K potentiometer. Since the ordinate represents the square of the calculated specimen current and not the bridge current, the consistency of the points for the two cases demonstrates the fact that the physical origin of the emf must lie in the specimen proper.

Not all of the specimens tested showed such a uniform square law dependence out to the highest currents available. In some cases the measured potential climbed less rapidly than by the square of the current for the higher current values. The drop-off would begin at currents of about 2.5 to 3 ma. The graph of figure 7 shows such a specimen. To within the accuracy of measurement, however, it can be stated that for currents less than 2 ma , in every negative carrier specimen in which there was an observable effect, the measured potential went as the square of the current. In itself, of course, this fact cannot be taken as profound evidence for the configurational emf because the circuitry is so designed that it detects even non-linearities and the term in the square of the current would be the first in a series expansion of any even non-linear effect.

A number of control experiments were undertaken. Some of these merely provided checks and balances on the working of the equipment. Others were devised to check out the implications of equation 8 to the extent allowed by the limitations of the equipment.

To check the working of the circuit the specimen of figure 5 was replaced by a set of precision wire wound resistors of approximately the same resistance as each of the branches of a typical specimen. This simulated specimen exhibited a signal of less than 1.0 microvolt out to currents of 4 ma in contrast to signals of the order of 100 uv from the bismuth specimens at this current level.

An evaporated tin specimen was prepared under the same conditions as were the bismuth specimens. The tin specimen had a constriction. It was inserted into the circuit in place of the usual bismuth specimen. Since the Hall coefficient for tin is of the order of 10^{-5} cm^3/coulomb, the expected configurational signal would be something like one billionth of that expected from a bismuth specimen of the same dimensions. Clearly the tin sample should exhibit a null effect. And, in fact, the control tin specimen did yield a null effect to within the accuracy of measurement.

An important and definitive test is one which directly detects the dependence of the measured emf on the cross-sectional area of the constricted flow region. Such a test should distinguish between the possible magnetic and hydrodynamic causes of a configurational

emf. The theoretical distinction is manifested by the two different dependences on area exhibited by equations 8 and 11. The hydrodynamic effect goes inversely as the square of the area whereas the magnetic effect goes as the inverse of the first power of the area.

It was for the express purpose of making this test that the extra potential sensing probes on the constriction were added to the one central one. The schematic diagram of figure 4 is somewhat misleading in that the constricted current flow region of an actual specimen is not uniformly wide along the whole of its length as is the case with the schematic shown. In the limit of infinitely thin potential probes, it is clear, that the proper area which should go into equation 8 or 11 is the local one at the point at which the potential probe meets the flow region. In practice the evaporation mask which determines the shape of the specimen was purposely designed to produce three different effective widths at the three different potential probe points of the fingers labelled "M" in figure 4. On the reasonable assumption that the mean thickness of the film is uniform throughout, the three different widths correspond to three different effective cross-sectional areas. The projected procedure involved taking three different sets of data using each of the three "M" potential probes in turn as the "center" tap on the specimen. Figure 5 shows the circuitry for one such arrangement. In turn, each of the other "M" contacts was connected into the bridge in place of the one which is shown so connected

in the figure in an effort to obtain variable area data.

Unfortunately this series of tests was inconclusive. The difficulty arises from three sources. The first is that the masks were made to obtain the smallest widths possible subject to the technology used in their fabrication. Cross-sectional widths significantly less than about 1 mil could not be obtained. Secondly, even with this lower limit of cross-section the signal to noise ratio was not much higher than about 2 for the best cases. The result of these two factors is that: (1) The only technologically allowable variation of width was an increase which in turn would decrease the output signal. (2) A relatively slight decrease in signal would submerge it below the noise level or at least decrease the data accuracy markedly.

The third difficulty which nullified all possibility of obtaining the area dependence was the fact that the cross-sectional width could not be defined and measured under the microscope to an accuracy of better than about 40%. This low precision is due mostly to the local fluctuations and variations in width that result from the crudeness of the mask fabrication. Under a microscope the constricted flow region appears quite ragged, though to the naked eye it seems relatively uniform. Since it is not clear what sort of average to take over the width fluctuations, the width is ill-defined.

Combined with the first two difficulties this third one implies that it is not enough to make the width slightly wider at two of the "M"

probes but considerably wider so that the difference in width is apparent beyond the 40% inaccuracy. Since even an increase of width by a factor of 1.4 (not to speak of doubling the width), would submerge the signal completely, the experimental results could not be interpreted with any confidence. It is only possible to state qualitatively that a definite increase in cross-sectional area was always associated with a decrease in signal. A quantitative dependence of the emf on area was not obtained.

The extra "M" contacts did, at least, provide a method of ascertaining the consistency of the measured results. Suppose, for example, the balancing potentiometer of the bridge circuit of figure 6 were hooked up to the two side "M" contacts and the central one instead of to two "E" contacts and to an "M" contact. If there were no variation in cross-section along the constriction a null signal would result. Due to the variations that do exist, however, it is clear that a small signal should appear. What the signal should be can be calculated in terms of the individual measured signals from each of the "M" contacts separately in conjunction with the known bridge resistances. Invariably there was good correlation between the experimental data and the expected result in tests such as these.

Another contact switching test of interest is one in which the other available pair of "E" contacts are used in place of those which

are hooked into the circuit of figure 5. As expected, this procedure had no detectable effect on the signal.

In general there are 35 distinct arrangements of the central and end pair of bridge leads onto the seven specimen contact points. Each of these will yield a signal which is predictable in terms of the signals from three of the arrangements and a knowledge of the resistances of the specimen and potentiometer. The predictions are founded upon the assumption that there are effectively four sources of emf present inside the specimen along the length of the constricted region. Each such source is between a pair of "M" probes or between such a probe and the end of the constricted region. The predictions further assume that the sum of these emf's from one end of the constriction to the other is zero. This fundamental idea which forms the basis for estimating the signals for thirty-five contact arrangements from the signals from three such arrangements is indicated schematically in figure 8.

The contact switching experiments already described are but two examples of the large number of contact switching experiments which can be done. In every such experiment actually undertaken the results always agreed with that deduced from the basic assumption illustrated in figure 8. Such agreement strongly indicates that the seat of the voltage detected lies, indeed, in the configuration of the specimen

so that it may be termed a configurational emf. These experiments do not, however, confirm the view as to the hydrodynamic cause of the emf.

There was a set of specimens which yielded unexpected signals. These specimens exhibited a signal polarity associated with positive current carriers instead of negative ones. Since bismuth generally shows a negative carrier Hall effect, this result was at first very disturbing. Happily, further investigation turned this apparent inconsistency into strong evidence that the measured emf is indeed associated with the properties of the electron gas in the specimen. The further investigation consisted in making direct Hall measurements on the samples. In every case, without exception, the sign of the Hall coefficient agreed in its implications with the sign of the configurational emf. Those specimens which showed positive carriers by the configurational emf showed positive carriers in Hall measurements and similarly for negative carrier specimens.

The positive carrier specimens were unusual in more respects than merely being positive carrier specimens. These specimens generally yielded erratic signals which were hard to reproduce in that there appeared to be a hysteresis present. The signal depended to some extent upon what level of current had just been passed through the specimen. The signals from these specimens were markedly noisier than those from negative carrier specimens. Generally the

positive carrier signals were an order of magnitude stronger than those from negative specimens and these signals usually did not follow anything like a square law in the current. In short, the positive carrier specimens yielded quite unexpected results which do not correspond at all to the theories of equations 8 or 11 except in that the sign of the signal agrees with that predicted by the Hall effect.

It was decided, in the light of the original purpose of this research which was primarily to detect and verify empirically the existence of a configurational emf, that the positive carrier anomalies should be relegated to future investigations after the more consistent results from simpler specimens are understood quite fully. All specimens which appeared excessively noisy and erratic and which showed positive carriers were therefore set aside and only the well-behaved specimens were examined. In mitigation of this procedure it should be stated that, in practice, the class of noisy specimens alone constituted the same set of specimens as did the class of positive carrier specimens. All positive carrier specimens were noisy and vice versa.

Since it is known that bismuth, like the semiconductors germanium and silicon, is exceedingly sensitive in its electrical properties to the type and amount of impurity present, it is not too surprising that unusual effects arose in some samples. Little care was taken to obtain

the highest purity bismuth nor was any effort made to keep the bismuth contamination-free during evaporation or during other steps in the sample preparation. The anomalies are attributed at present to the variation in preparation and in impurity from sample to sample. It seemed reasonable to postpone a full investigation of the positive carrier configurational emf until the time at which the simple effect is understood. These specimens were, therefore, discounted as effectively faulty specimens and were not used for data gathering.

5. Quantitative Results

We next examine the quantitative correlation between the empirical results and the theories exhibited in equations 8 and 11. To do this we first take notice of the information obtainable from the direct Hall effect measurements.

The measured Hall voltage, V_H, may be written in terms of the applied magnetic induction B and the thickness of the specimen, b, as follows:

$$V_H = (H/b) I B \tag{15}$$

where I is the total current passing through the specimen and H is the Hall coefficient. Empirical values of (H/b) as deduced from measurements by virtue of equation 15 are recorded in line 2 of table 1 for the two specimens of figures 6 and 7.

It is also possible to obtain direct empirical estimates of the quantity $(\sigma b)^{-1}$ by properly combining the measured values of resistance, length and width, w, of the various construction arms of the specimens. The quantity σ then represents the actual effective conductivity of the specimen material in each of the constriction arms. Mean values for $(\sigma b)^{-1}$ are recorded in line 1 of table 1. The uncertainties in these estimates amount to about $\pm$ 60% primarily because of the difficulty in determining the average width of the constriction arms.

We note next that the self-magnetic emf of equation 11 may be rewritten in terms of the directly measurable quantities (H/b) and the width w as follows:

$$V_m = \frac{\gamma \mu_0}{4\pi w} (H/b) I^2 \qquad (16)$$

For current through a very thin flat plate of width w and thickness b where $b \ll w$ it may be shown that the coefficient γ of equation 16 is approximately $\gamma = 1.4$. Using values of (H/b) from direct determination on specimens I and II and the measured widths w of these two specimens at the potential probe a prediction of (Vm/I^2) may be obtained. This prediction may be compared with the measured (V/I^2) exhibited in figures 6 and 7. The calculation yields $Vm/I^2 = 10^{-3}$ microvolts/$(ma)^2$ and $Vm/I^2 = 2 \times 10^{-3}$ microvolts/$(ma)^2$ respectively for the two specimens under consideration. Both of these numbers are

three orders of magnitude smaller than the measured effect plotted in figures 6 and 7.

This pronounced discrepancy eliminates the self-magnetic deflection emf as the primary cause of the observed effect. As will be discussed in a succeeding paragraph, the discrepancy between the predictions of equation 8 or 13 and the observed results are even more pronounced. Nevertheless, it is only the magnetic theory that is eliminated by the lack of quantitative correlation at this point. The reason for this lies in the fact that equation 8 is quite crude. Several of the quantities which enter into it are not clearly defined nor are they directly measurable. A quick comparison of equation 3 with equation 88 will serve to exhibit this fact. On the other hand, equation 16 is quite exact as it stands. That is to say the coefficient (H/b) obtained from a direct Hall measurement should indeed be exactly the same coefficient which occurs in the configurational self-magnetic deflection emf. This is independent of the conditions under which the Hall coefficient represents the quantity 1/ne or some other more complex quantity.

Equation 15 represents an emf induced by the application of an external magnetic field to the current flowing in the specimen. Equation 16 represents an emf induced by the application of a self-produced magnetic field to the current flowing in the specimen. Regardless of the source of the magnetic field it is the quantity (H/b) which

describes the strength of the effect of this magnetic field on the current. It is for this reason that the comparison between equations 15 and 16 as they stand is exact. Hence the discrepancy between equation 16 and the empirical result does indeed eliminate the self-Hall-effect explanation of the observed configurational emf.

We now examine the correlation between the empirical results and the predictions of the electron hydrodynamic hypothesis of equations 3, 7, and 8. On this theory the slope of the lines in figures 6 and 7 effectively represent the value of the coefficient $C/A^2 = C/b^2w^2$ in equation 8. In order to compare theory with experiment we must make some hypothesis about the value of C as given in equation 3 or 8. It is here that difficulty arises because we assume that $1/ne = H$. Although it is true for ideal metals that the strength of a magnetically-induced current deflection, H, is indeed a measure of the density of current carriers, this relation is certainly not generally true. In particular for bismuth this relation is known to yield anomalies especially in thin films. In addition it is not even clear at this stage that equation 3 is correct as it stands nor is the exact meaning of n or m clear in that equation even if it were correct. It is precisely for these reasons that lack of direct quantitative correlation with experiment is not as decisive in the case of equation 13 as it is in the case of equation 16.

Notwithstanding the preceding discussion, however, it is clearly not unreasonable at this stage to compare theory with experiment using the relation of equation 8 in conjunction with the estimate $H = 1/en$ where H represents the empirically measured effect of a magnetic field on the current. On this basis, then, equation 8 or 13 may be rewritten as follows:

$$V_c = -\frac{1}{w^2}\left(C/b^2\right) I^2 = \frac{m}{2e\,w^2}\left(H/b\right)^2 I^2 \qquad (17)$$

If we proceed as before to predict the slope of the lines in figures 6 and 7 using equation 17 in conjunction with the empirically determined values of w and H/b, the results are $V_c/I^2 = 10^{-4}\ \mu v/ma^2$ and $V_c/I^2 = 3 \times 10^{-4}\ \mu v/ma^2$ for specimens I and II respectively. These predictions disagree quite plainly with the slopes recorded on the respective graphs.

This result indicates that the crude theory of equations 3, 8, and 17 is certainly inapplicable as an explanation of the observed results. Before embarking on an extended analysis of the possible causes for the quantitative discrepancy between theory and experiment it is clearly expedient to have a more definitive and less ambiguous theory on which to base the discussion. It is to this end that the next section is devoted. It should be understood that from this point on we will concern ourselves primarily with examining the hydrodynamic basis for the configurational emf rather than the self-magnetic one.

Therefore, the term configurational emf in what follows will be reserved for reference to the electron hydrodynamic concept only. The self-Hall effect emf will be mentioned explicitly when necessity arises to distinguish it from the configurational emf.

III. THEORY

1. Macroscopic Equations

The first step in constructing a theoretical analysis of the configurational emf lies in an appreciation of the fact that Ohm's law is an approximation to a much more elaborate equation. Some of the complexity of this more general description of current flow in a medium is shown in equation 18.

$$-2C\nabla\cdot(\vec{j}\vec{j}) - G\nabla T + H\vec{j}\times\vec{B} + \vec{\mathcal{E}} - \frac{1}{\sigma}\vec{j} = 0 \qquad (18)$$

In this equation $\vec{\mathcal{E}}$ represents the externally applied electric field, σ is the conductivity of the medium, T is the temperature and the symbol ∇ is the usual spatial gradient operator. The coefficients C and G, like the Hall coefficient H, indiate the strength of each of the terms they multiply. They are constants which depend upon the material medium in which the current flows. It is clear that, in the limit of time independent spatially uniform current flow in the absence of thermal gradients and strong magnetic fields, the first three terms on the left are zero. And the equation reduces to Ohm's law as expected.

The second term in equation 18 which contains the temperature T, corresponds to the thermo-electric property of metals.

The coefficient G is the thermo-electric "power." And the term $(-G\nabla T)$ represents the electric field set up by virtue of the presence of a thermal gradient in the material.

The third term represents the familiar Hall field. The quantity $H\vec{j} \times \vec{B}$ indicates the magnitude of the electric field induced by the action of a magnetic field on the moving charges which constitute the current.

The next term is the electric field applied from some external source as, for example, from a storage battery. All of the terms may be looked upon as effective fields. In this view the term $(-\frac{1}{\sigma} j)$ is the ohmic pseudo-electric field. In the form shown the equation states that the sum total of all the electric fields present is always zero. This view of equation 18 is not purely academic but has a physical basis. It is quite clear that the terms associated with each of the four fields - thermal, Hall, applied and ohmic - are directly measurable as voltages. In fact, the coefficients G, H, and σ are determined empirically by measuring these voltages.

The first term, which is the configurational one, like the other terms represents an effective electric field. It describes the field which exists in the medium by virtue of the non-uniformity of current flow. To exhibit the emf or voltage associated with this field we note the following:

$$\nabla\cdot(\vec{j}\,\vec{j}) = (j\cdot\nabla\vec{j}) = -\vec{j}\times(\nabla\times\vec{j}) + \nabla(\tfrac{1}{2}j^2) \qquad (19)$$

Here use has been made of the equation of continuity, $\nabla\cdot\vec{j} = 0$, in the absence of time varying conditions and the remainder is just vector identity.

To obtain the emf associated with any of the fields in equation 18 it is merely necessary to integrate that field along a path connecting the two points between which the potential difference is desired. For the configurational case this integration is complicated by the tensor nature of the configurational emf. The complications are intimately related to the measurability of this emf in a single specimen. If the configurational field could be written generally as the gradient of a scalar, as, for example, the thermo-electric field may be written, then it would not be measurable in a single metal. Because of the scalar nature of temperature with which the thermal voltage is associated, at least a pair of metal elements (a thermocouple) is needed to detect the thermal emf. The thermal emf around a complete circuit in a single metal must be zero even though some part of the metal is at a higher temperature than the remainder of the specimen. Completing the circuit guarantees that the potential is being measured between two points at the same temperature in the same metal. Since there is no thermal or metallic difference, there cannot be a potential difference.

If, however, the field cannot be written as the gradient of a scalar, the resultant potential has tensor properties. Then the associated emf will depend upon the direction or path of integration and the effect does not necessarily cancel out in a single metal. The configurational field varies, in general, with current direction as well as with the magnitude of the current. It is this very property which allows the thermal emf's due to Joule heating by the current to be distinguished experimentally from the direct effect of the current exhibited in the configurational emf.

With these precepts in mind we may write down the configurational emf for the case of the actual experimental arrangements described in the previous sections. To do this we note that the configurational field $\vec{\mathcal{E}}_c = -2\,C\,\vec{j}\cdot\nabla\vec{j}$ is present only in those regions of space where the current density varies along the current flow direction and not where the current density varies perpendicular to the flow direction. For example, if the current is in the x direction but the current density varies spatially only in the y or z direction, then the field $\vec{\mathcal{E}}_c$ is zero. This fact implies that an integration path perpendicular to the flow lines crossing the boundary of the flow region will contribute essentially nothing to the integration. However, an integration into the high flow region along a path parallel to the flow lines will contribute considerably to the integral.

With reference to figure 9 we see that the experimental arrangement measures the configurational potential $V_c = V_4 - V_1$ where

the integration is taken along the path indicated. In the idealized case where the flow lines are undisturbed by the presence of the potential probes at points 1 and 4, we have $V_2 - V_1 = V_4 - V_3 = 0$. This may be verified by noting that in crossing a flow boundary from a region of uniform flow to a region of zero flow, such as is crossed in the path between points 3 and 4 or between points 1 and 2, we have $\vec{j} \times (\nabla \times \vec{j}) = \nabla(\frac{1}{2} j^2)$ over the boundary. However, along the path between points 2 and 3, the flow is irrotational and $\nabla \times \vec{j} = 0$. And in part of the region between 2 and 3 the current density varies spatially so that $\vec{j} \cdot \nabla \vec{j} = \nabla(\frac{1}{2} j^2) \neq 0$ along this path. As a result, for the experiment under consideration we have

$$V_c = -2C \int_1^4 d\vec{r} \cdot (\vec{j} \cdot \nabla \vec{j})$$
$$= -2C \int_2^3 d\vec{r} \cdot \nabla(\tfrac{1}{2} j^2) = \Delta(-C j^2) \qquad (20)$$

By way of clarification and contrast we note that

$$\int_1^4 d\vec{r} \cdot \nabla(\tfrac{1}{2} j^2) = 0 \qquad (21)$$

Up to this point the experiment has been based on a crude estimate of the value of the coefficient C. Equation 8 shows that the value of C was taken to be

$$C = -\frac{1}{2} \frac{m}{e} \left(\frac{1}{ne}\right)^2 \qquad (22)$$

where the sign indicates negatively charged current carriers. It is the coefficient C which must be compared with the experimental results. Aside from a factor which represents the square of the cross-sectional area, the slope of the lines in figures 6 and 7 are directly relevant to the theoretical value of C if the measured effect is indeed the one which is under theoretical consideration here.

In the light of this preliminary discussion, the theoretical program is quite clear. The first step consists in demonstrating that equation 18 does, indeed, represent a proper extension of Ohm's law. In this respect it is only necessary to demonstrate the existence of the configurational term, the other terms being more or less familiar. It may be reiterated here that even equation 18 is an incomplete statement. Considerably more elaborate equations may be written down which allow for the tensor nature of the coefficients σ and H, for example, or for their temperature dependence. Only those terms were written down which were easy to interpret and which had some relationship, illustratively or directly, to the configurational term.

The second requirement of the theoretical discussion is to obtain an expression for the coefficient C from which numerical estimates may be obtained for comparison with experiment.

2. Microscopic Treatment with Relaxation Time Assumption

To proceed with the calculation then we note first that basically

we are considering one of the transport properties of an ensemble of particles (the valence electrons in the solid). These particles are distributed in phase space over momentum and position. For this analysis the particles are envisioned as being described by wave packets. By the momentum, $\hbar\vec{k}$ (or just $\vec{k}$) and position, $\vec{r}$, of a particle is meant the mean of the quantity averaged over the wave packet.

Let the function $f(k_x, k_y, k_z, x, y, z) \equiv f(\vec{k}, \vec{r})$ represent the probability that the point (k_x, k_y, k_z, x, y, z) of phase space is fully occupied. Then the number of particles with momentum between k_x and $k_x + dk_x$, k_y and $k_y + dk_y$, k_z and $k_z + dk_z$ and located at a point between x and $x + dx$, etc. is

$$\rho(\vec{k}, \vec{r})\, f(\vec{k}, \vec{r})\, dk_x\, dk_y\, dk_z\, dx\, dy\, dz \equiv \rho(\vec{k}, \vec{r})\, f(\vec{k}, \vec{r})\, d^3k\, d^3r.$$

The function $\rho(\vec{k}, \vec{r})$ represents the density of states available in the volume of phase space $d^3k\, d^3r$. Alternatively the quantity $\rho(k, r)\, d^3k\, d^3r$ is the degeneracy of states at the point $(\vec{k}, \vec{r})$ of phase space.

For Bloch wave functions or plane wave functions the momentum or wave vector $\vec{k}$ is quantized. In fact, $\vec{k}$ is simply a form of quantum number representation. For example, in the simple one-dimensional case we have $k = \frac{2\pi}{L}\nu$, $\nu = 0, \pm 1, \pm 2 \ldots \pm \frac{L}{2a}$. Here L is the length of the solid and a is the lattic constant. The values that are indicated for the quantum number ν describe the limits

of the first Billouin zone. A general extension of this one-dimensional case yields the result that the number of wave states per unit volume of k-space is given by $d^3\nu = (\frac{1}{2\pi})^3 d^3k$ x volume. Upon introducing the degeneracy due to the spin of the electron the density of states in phase space becomes simply $\rho(\vec{k},\vec{r}) = 2 \times (\frac{1}{2\pi})^3 = \frac{1}{4\pi^3}$.

The normalization of the probability $f(\vec{k},\vec{r})$ comes from the statement that the spatial density of particles is given by

$$N = \frac{1}{4\pi^3} \int f(\vec{k},\vec{r})\, d^3k \tag{23}$$

where here N represents the actual number of particles under consideration per unit volume; not an effective number. For the case of the electron gas in equilibrium the distribution probability f is the Fermi distribution

$$f_F(\vec{k},\vec{r}) = \left(1 + \exp\frac{E - E_F}{\kappa T}\right)^{-1} \tag{24}$$

The Fermi level E_F is fixed by the normalization of equation 23. The symbol E represents the energy of an electron and κ is Boltzmann's constant.

The basic equation of state for the transport properties of a fluid is the Boltzmann transport equation (4, 5, 6). This equation says that in the steady state the time rate of change of the density of particles at a point in phase space due to the drift of particles under the influence

of external forces is just compensated by the rate at which particle collisions change this density.

$$\vec{v} \cdot \nabla f + \frac{1}{\hbar} \vec{F} \cdot \nabla_k f = \left(\frac{\partial f}{\partial t}\right)_{coll.} \tag{25}$$

The left-hand side of this equation represents the time rate of change of f due to spatial drift and to momentum drift under the influence of the generalized applied force $\vec{F}$. The force $\vec{F}$ is generally given by $\vec{F} = -e(\vec{\mathcal{E}} + \vec{v} \times \vec{B})$. The vector $\vec{v}$ represents the velocity of a particle and is a function of the wave vector $\vec{k}$ through the equation.

$$\vec{v} = \frac{1}{\hbar} \nabla_k E(\vec{k}) \tag{26}$$

The symbol ∇_k represents the gradient in k-space. The constant $\hbar$ is Planck's constant divided by 2π; $\hbar = h/2\pi$.

The right-hand side of the equation represents the time rate of increase of the particle density at $(\vec{k}, \vec{r})$ due to collisions. For Fermi particles this is given by the following equation where $W(\vec{K}, \vec{k})$ is the probability per unit time that on collision a particle with momentum $\vec{K}$ will find itself with a momentum between $\vec{k}$ and $\vec{k} + d\vec{k}$.

$$\left(\frac{\partial f}{\partial t}\right)_{coll.} = [1 - f(\vec{k}, \vec{r})] \int W(\vec{K}, \vec{k}) f(\vec{K}, \vec{r}) d^3K - f(\vec{k}, \vec{r}) \int W(\vec{k}, \vec{K}) [1 - f(\vec{K}, \vec{r})] d^3K \tag{27}$$

To gain an appreciation for the overall calculation procedure and for the approximations and omissions found necessary to arrive at a malleable result we resort temporarily to an oversimplified model. It is provisionally assumed, as is often done in the literature, that the total effect of collisions may be condensed into a mean collision time or relaxation time τ .

$$\left(\frac{\partial f}{\partial t}\right)_{coll.} = -\frac{1}{\tau}\left[f(\vec{k},\vec{r}) - f_o(E)\right] \qquad (28)$$

Equation 28 embodies the idea that if the distribution of particles in phase space is disturbed from its equilibrium value $f_o(E)$, then it will decay back to that value exponentially with time. The relaxation time for this process is τ . By the equilibrium distribution is meant the one which obtains in the absence of transport processes. For the case under consideration this would be the Fermi distribution of equation 24 so that $f_o(E) = f_F(E)$. In a later section we shall return to this point and examine the consequences of this overall relaxation time constraint.

The general program of procedure is the following. We first imagine $f(\vec{k},\vec{r})$ expanded in a sequence of terms of decreasing orders of magnitude, $f(\vec{k},\vec{r}) = f_o(E) + g_1(\vec{k},\vec{r}) + g_2(\vec{k},\vec{r}) + \ldots$. Expressions for the functions $g(\vec{k},\vec{r})$ are then obtained using the Boltzmann transport equation successively for each order. Next, the expanded solution for the transport-modified distribution function,

$f(\vec{k},\vec{r})$ is used to find the current density, $\vec{j}$, created by the thermal gradients and the electric and magnetic fields in the metal.

$$\vec{j} = -e\frac{1}{4\pi^3}\int \vec{v}\, f(\vec{k},\vec{r})\, d^3k$$

$$= -(e/4\pi^3)\int \vec{v}\,[f_0(E) + g_1(\vec{k},\vec{r}) + g_2(\vec{k},\vec{r}) + \ldots]\, d^3k \qquad (29)$$

The emf associated with each of the various terms contributing to the current is then given simply by the following integral connecting two points between which the emf is desired.

$$emf = \int \frac{1}{\sigma}\vec{j}\cdot d\vec{r} \qquad (30)$$

Employing the simplification of equation 28 in conjunction with the Boltzmann relation in equation 25 the functions $g_1(\vec{k},\vec{r})$, $g_2(\vec{k},\vec{r})$ and $g_3(\vec{k},\vec{r})$ are obtained successively from

$$(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\, f_0 = -\kappa\xi\,\vec{v}\cdot\nabla T\frac{\partial f_0}{\partial E} - e\vec{\mathcal{E}}\cdot\vec{v}\frac{\partial f_0}{\partial E} = -\frac{1}{\tau}\, g_1(\vec{k},\vec{r}) \qquad (31)$$

$$(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\, g_1 = -(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\,\tau\,(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\, f_0 \qquad (32)$$

$$= -\frac{1}{\tau}\, g_2(\vec{k},\vec{r})$$

$$(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\, g_2 = -\frac{1}{\tau}\left[-\tau(\vec{v}\cdot\nabla + \frac{1}{\hbar}\vec{F}\cdot\nabla_k)\right]^3 f_0 \qquad (33)$$

$$= -\frac{1}{\tau}\, g_3(\vec{k},\vec{r})$$

Use has been made in equation 31 of the expression for the velocity $\vec{v}$ exhibited in equation 26 and the shorthand notation $\xi = \frac{1}{kT}(E - E_F) + \frac{1}{k}\frac{\partial E_F}{\partial T}$ has been employed for convenience.

By virtue of the presence of the $\vec{v}$ in the integrand of equation 29 it is clear that only distribution function terms with odd symmetry in k space can produce non-zero currents from the integral of equation 29. This is, of course, not unreasonable since symmetrical terms would not be expected to give rise to directional properties as exemplified by the current. That the current due to the equilibrium distribution, $f_o(E)$, is zero is clear both mathematically from symmetry and physically from the meaning of $f_o(E)$ as the distribution obtaining in the absence of a current.

From the form of $g_1(\vec{k},\vec{r})$ exhibited in equation 31 it may be deduced that both terms give rise to non-zero currents.

$$\vec{j}_1 = -\left(ek\int \tau \xi \frac{1}{3} v^2 \frac{\partial f_o}{\partial E}\frac{1}{4\pi^3} d^3k\right)\nabla T + \left(-e^2\int \frac{1}{3} v^2 \tau \frac{\partial f_o}{\partial E}\frac{1}{4\pi^3} d^3k\right)\vec{\mathcal{E}} \qquad (34)$$

In this equation it has been assumed that the medium is microscopically isotropic by which is meant the assumption that E is a function of the magnitude of k only.

It is clear on comparing equation 18 with equation 34 that the coefficient of the applied field $\vec{\mathcal{E}}$ in the latter equation is σ and the coefficient of ∇T is related to the thermoelectric power G. The integral coefficient of ∇T in equation 34 must be simply σG.

The second order function $g_2(\vec{k}, \vec{r})$ is composed of at least nine terms. These result from the two consecutive operations of the three-term operator $\left(\vec{v} \cdot \nabla - \frac{e}{\hbar} \vec{\mathcal{E}} \cdot \nabla_k - \frac{e}{\hbar} \vec{v} \times \vec{B} \cdot \nabla_k \right)$. Because the ∇_k operator operates on both $f_o(E)$ and on the velocity terms, in addition to its operation on τ, if the relaxation time is a function of k, and because the spatial gradient ∇ operates on T and on $\vec{F}$, since $\vec{F}$ is assumed to have a spatial variation, the basic nine terms may be expanded to a considerable number of terms. However, a careful inspection of all of these terms shows that the great majority of them contribute nothing to the current integral.

All of the terms in $g_2(\vec{k}, \vec{r})$ which involve a combination of thermal and electric terms exclusive of the magnetic terms will involve the operator ∇_k and the function $\vec{v}$ an even number of times. But both the operator ∇_k and the vector $\vec{v}$ will always produce an odd symmetry in k space when present in conjunction with an even function. And the energy E must be even in k space even though it is not necessarily isotropic. The result is that all of the non-magnetic terms in $g_2(\vec{k}, \vec{r})$ will have even symmetry in k

space or odd symmetry in two different directions simultaneously. In either case, when these functions are multiplied by $\vec{v}$ to form the integrand for the current calculation, this integrand will be inherently odd in k space and the integral must be zero.

Because the magnetic term involves an extra velocity, the integrand with the magnetic force term in conjunction with the electric and thermal fields does not result in a zero integral. Allowing for the possibility that τ varies with E, the only part of $g_2(\vec{k},\vec{r})$ which can contribute a non-zero term to the current integral is

$$g_2(\vec{k},\vec{r})\Big|_{ODD} = \tau^2 \frac{e^2}{\hbar^2}(\vec{v}\times\vec{B})_\lambda \frac{\partial^2 E}{\partial k_\lambda \partial k_\nu}\left(\vec{\mathcal{E}}_{th}+\vec{\mathcal{E}}\right)_\nu \frac{\partial f_0}{\partial E} \qquad (35)$$

All other terms are either identically zero because $\vec{v}\times\vec{B}\cdot\vec{v} \equiv 0$ or they possess symmetry properties such that their current integral will be zero. In equation 35 the direction indices λ and ν are to be summed over, since they appear twice, in accordance with the usual summation convention. The pseudo-electric field, $\vec{\mathcal{E}}_{th}$, is merely a shorthand notation for the field induced by thermal gradients, $\vec{\mathcal{E}}_{th} \equiv \frac{\kappa}{e}\xi\nabla T$. The origin of this expression is clear from equation 31.

The current created by the $g_2(\vec{k},\vec{r})$ correction to the distribution function may be written down as follows:

$$\vec{j}_\lambda = \epsilon^{\lambda\mu\nu}\vec{e}_\lambda \mathcal{E}_\mu \mathcal{B}_\nu \int\left(-\frac{e^3}{\hbar}\right)\tau^2 v_\lambda v_\mu^2 \frac{\partial}{\partial k_\mu}\left(\frac{v_\lambda}{v_\mu}\right)\frac{\partial f_0}{\partial E}\frac{d^3k}{4\pi^3}$$

$$+\epsilon^{\lambda\mu\nu}\vec{e}_\lambda (\nabla T)_\mu \mathcal{B}_\nu \int\left(-\frac{e^2 K}{\hbar}\right)\tau^2 \xi v_\lambda v_\mu^2 \frac{\partial}{\partial k_\mu}\left(\frac{v_\lambda}{v_\mu}\right)\frac{\partial f_0}{\partial E}\frac{d^3k}{4\pi^3} \qquad (36)$$

It is to be understood, in equation 36, that the symbol $\epsilon^{\lambda\mu\nu}$ is to take on the value +1 or -1 depending upon whether $\lambda\mu\nu$ is an even or an odd permutation of the three space direction indices. And $\epsilon^{\lambda\mu\nu} = 0$ if any pair of the indices $\lambda\mu\nu$ are equal. The symbol $\vec{e}_\lambda$ is a unit vector in the λ direction. In obtaining equation 36 extensive use was made of the energy gradient formula for the velocity and of the even symmetry properties of E in k-space.

The second term on the right in equation 36 is a thermomagnetic emf. If the magnetic field which causes the emf is produced by the current present in the medium, then the associated voltage reverses direction with a current direction reversal. As a result this effect could not compete with the configurational one which does not reverse direction with a current reversal.

The first term represents the Hall effect. If it is assumed that the medium is perfectly isotropic so that the energy E is a function of the magnitude of k only the Hall effect term reduces to the simpler form

$$\vec{j}_H = \vec{\mathcal{E}} \times \vec{\mathcal{B}}\ \frac{e^3}{\hbar^4}\int \tau^2 \frac{1}{3k}\left(\frac{\partial E}{\partial k}\right)^3 \frac{\partial f_0}{\partial E}\frac{d^3k}{4\pi^3} \qquad (37)$$

From the form of equation 18 it is clear that the integral coefficient of $\vec{\mathcal{E}} \times \vec{\mathcal{B}}$ in equation 37 is a representation of $\sigma^2 H$.

In order to finally arrive at a term in $\vec{j}$ which corresponds to the configurational emf we must proceed to third order. The third order correction $g_3(\vec{k}, \vec{r})$ consists, basically, of twenty-seven terms. It is the result of a triple operation of a three-term operator. This formidable elaborateness makes it worthwhile to undertake some preliminary reflection in lieu of simply writing down all of the terms.

The most important observation to make is that within each order of calculation there are sub-orders. For example, in the calculation of $g_1(\vec{k}, \vec{r})$ it is clear that the ohmic term is considerably more important than the thermal emf term. This is evidenced empirically by the fact that, in the experiment, applied ohmic potential drops approaching one hundred volts may exist in the specimen. This is to be compared with possible thermo-electric voltages of the order of millivolts at the very most. Equivalently, one may say that $(\overline{\mathcal{E}_{th}}) \ll \mathcal{E}$ where $(\overline{\mathcal{E}_{th}}) = \frac{\kappa}{e}(\bar{\xi})|\nabla T|$ is an effective thermo-electric field. By $(\bar{\xi})$ is meant some reasonable average such as that indicated by equation 34.

Similarly in the calculation of $\vec{j}_2$ there is again a relatively large term - the Hall effect - in addition to a higher sub-order term which is quantitatively small - a thermomagnetic effect. Both of these arise in the second order calculation. Again, the two terms are

distinguished by $\overline{\mathcal{E}_{th}} \ll \mathcal{E}$.

Since the third order is the highest one to which the calculation will be extended, it is not unreasonable to neglect all of the higher sub-orders in this last calculation. Only the terms with the greatest weight are desired in the highest order. Those terms containing effectively an $\mathcal{E}_{th}$ must be smaller than those containing $\mathcal{E}$ with which they are in competition in this order. This is not withstanding the fact that terms involving $\mathcal{E}_{th}$ from first or second order may, indeed, compete with $\mathcal{E}$ terms of third order. It was precisely to investigate the nature of such a competition that the thermal terms were carried as far as second order. Now, however, all terms involving thermal gradients will be dropped in the calculation of $g_3(\vec{k},\vec{r})$ as effectively constituting higher sub-order terms.

Omitting the thermal terms, keeping in mind the necessary symmetry properties for a non-zero current integral contribution in conjunction with the even symmetry of E in k-space, and remembering the k-space definition of $\vec{v}$ and the identity $\vec{v} \times \vec{B} \cdot \vec{v} \equiv 0$ the whole effect of $g_3(\vec{k},\vec{r})$ may be condensed into the following four parts:

$$
\begin{aligned}
g_3(\vec{k},\vec{r})\big|_{ODD} = {} & e\tau^3 \frac{\partial f_0}{\partial E}(\vec{v}\cdot\nabla)^2(\vec{\mathcal{E}}\cdot\vec{v}) \\
& + \frac{e^3\tau^3}{\hbar^2}\frac{\partial f_0}{\partial E}(\vec{v}\times\vec{B}\cdot\nabla_k)^2(\vec{\mathcal{E}}\cdot\vec{v}) \\
& + \frac{e^3\tau}{\hbar^2}(\vec{\mathcal{E}}\cdot\nabla_k)\left[\tau\vec{\mathcal{E}}\cdot\nabla_k\left(\tau\vec{\mathcal{E}}\cdot\vec{v}\,\frac{\partial f_0}{\partial E}\right)\right] \\
& - \frac{e^2\tau}{\hbar}\left[(\vec{\mathcal{E}}\cdot\nabla_k)(\tau\vec{v}\cdot\nabla)+(\tau\vec{v}\cdot\nabla)(\vec{\mathcal{E}}\cdot\nabla_k)\right]\left(\tau\vec{\mathcal{E}}\cdot\vec{v}\,\frac{\partial f_0}{\partial E}\right)
\end{aligned}
\tag{38}
$$

In this equation the spatial operator ∇ operates on all of the electric field terms, $\vec{\mathcal{E}}$, following it multiplicatively. The k-space gradient operator, ∇_k, operates on all of the functions of $\vec{k}$ which follow multiplicatively.

The first term of equation 38 gives rise to a current which exists by virtue of the spatial rate of change of electric field inhomogeneity. The associated emf depends upon the difference between the spatial gradients of the applied field at two different points in the medium. To sense this emf a potential probe would have to tap directly into a region of rapidly converging or diverging current flow. Noting the occurrence of $\tau \vec{v}$ in conjunction with the ∇ operator leads to the speculation that this term would be small unless the applied field varied significantly over a distance comparable with the mean free path. In any case, this term could not obscure the configurational one in the experiment under consideration because its associated emf changes direction upon reversal of the applied field unlike the configurational emf.

The second term of equation 38 is the effect known as magneto-resistance. Inspection of this term shows that it is zero if microscopic isotropy obtains in the medium. By microscopic isotropy is meant the stipulation that the energy, E, is a function only of the magnitude of k. Again, even if magneto-resistance is present, it

could not interfere with configurational emf measurements since, as is implied by the term "magneto-resistance," the associated voltage changes sign with a reversal in the direction of the applied field or equivalently a reversal of the input current direction.

The next term of equation 38, by analogy with the second term, may be called an electro-resistive effect. This term gives rise to a resistance which is proportional to the square of the electric field just as the magneto-resistance varies as the square of the magnetic field. Again the associated voltage changes on current reversal unlike the configurational emf.

The final term of equation 38 yields the configurational effect. Only that part of the current $\vec{j}_3$ due to this configurational term will be examined. The currents produced by the other three third order effects, although they may be of general interest, are not immediately related to the configurational emf problem.

Unfortunately, the general expression for the configurational current is exceedingly lengthy and complicated. It is recorded here for completeness. In obtaining the expression for $\vec{j}_c$ an integration by parts is necessary and careful account is taken of the symmetry properties of the various integrands in order to eliminate immediately those terms which would contribute nothing to the current integral. Allowing for a possible variation of τ with energy, the final expression may be summarized in terms of three tensor integrals. These are

$$K_{\nu\mu} = -\frac{e^3}{\hbar^4}\int \tau^3 \frac{\partial f_o}{\partial E}\frac{\partial^2 E}{\partial k_\nu^2}\left(\frac{\partial E}{\partial k_\mu}\right)^2 \frac{1}{4\pi^3}\, d^3k \tag{39}$$

$$L_{\nu\mu} = -\frac{e^3}{\hbar^4}\int \tau^3 \frac{\partial f_o}{\partial E}\frac{\partial^2 E}{\partial k_\nu \partial k_\mu}\frac{\partial E}{\partial k_\nu}\frac{\partial E}{\partial k_\mu}\frac{1}{4\pi^3}\, d^3k \tag{40}$$

$$M_{\nu\mu} = -5\frac{e^3}{\hbar^4}\int \tau^2 \frac{\partial \tau}{\partial E}\frac{\partial f_o}{\partial E}\left(\frac{\partial E}{\partial k_\nu}\right)^2\left(\frac{\partial E}{\partial k_\mu}\right)^2 \frac{1}{4\pi^3}\, d^3k \tag{41}$$

In terms of these integrals the configurational current, by which is meant the current arising from the configurational or final part of $g_3(\vec{k},\vec{r})$ in equation 38, may be written as

$$\begin{aligned}\vec{j}_c = \sum_{\nu,\mu}{}' \vec{e}_\nu \Big[& \mathcal{E}_\nu \frac{\partial \mathcal{E}_\mu}{\partial x_\mu}\left(2K_{\nu\mu} + K_{\mu\nu} + 2L_{\nu\mu} + M_{\nu\mu}\right) \\ & + \mathcal{E}_\mu \frac{\partial \mathcal{E}_\nu}{\partial x_\mu}\left(K_{\nu\mu} + K_{\mu\nu} + 3L_{\nu\mu} + M_{\nu\mu}\right) \\ & + \mathcal{E}_\mu \frac{\partial \mathcal{E}_\mu}{\partial x_\nu}\left(5L_{\nu\mu} + M_{\nu\mu}\right) \\ & + \mathcal{E}_\nu \frac{\partial \mathcal{E}_\nu}{\partial x_\nu}\left(5L_{\nu\nu} + M_{\nu\nu}\right)\Big]\end{aligned} \tag{42}$$

The prime on the summation sign indicates that only terms for which the indices μ and ν are not equal are to be included. The symbol $\vec{e}_\nu$ is a unit vector in the ν direction.

In order to arrive at a tractable expression for the configurational emf we consider the special case of microscopic isotropy for which the energy E is a function of the magnitude of k only. For this case all of the tensor integrals $K_{\nu\mu}$, $L_{\nu\mu}$ and $M_{\nu\mu}$ may be written in terms of the three scalar integrals

$$\mathcal{P} = -\frac{e^3}{\hbar^4}\int \tau^3 \frac{\partial f_0}{\partial E}\frac{1}{3k}\left(\frac{\partial E}{\partial k}\right)^3 \frac{1}{4\pi^3}\, d^3k \tag{43}$$

$$\mathcal{Q} = -\frac{1}{3}\frac{e^3}{\hbar^4}\int \tau^3 \frac{\partial f_0}{\partial E}\left(\frac{\partial E}{\partial k}\right)^2\left(\frac{\partial^2 E}{\partial k^2}\right)\frac{1}{4\pi^3}\, d^3k \tag{44}$$

$$\mathcal{S} = -\frac{1}{3}\frac{e^3}{\hbar^4}\int \tau^2 \frac{\partial \tau}{\partial E}\frac{\partial f_0}{\partial E}\left(\frac{\partial E}{\partial k}\right)^4 \frac{1}{4\pi^3}\, d^3k \tag{45}$$

The configurational current for an isotropic medium, in terms of these integrals, falls into the form

$$\begin{aligned}\vec{J}_c = {} & (2\mathcal{P} + \mathcal{Q} + \mathcal{S})\,\vec{\mathcal{E}}\,\nabla\cdot\vec{\mathcal{E}} \\ & + (\mathcal{P} + \mathcal{Q} + \mathcal{S})\,\vec{\mathcal{E}}\cdot\nabla\vec{\mathcal{E}} \\ & + \tfrac{1}{2}(-\mathcal{P} + \mathcal{Q} + \mathcal{S})\,\nabla\mathcal{E}^2\end{aligned} \tag{46}$$

To a very good approximation we may take $\nabla \cdot \vec{\mathcal{E}} = 0$ since the field may be represented in terms of the applied current by $\vec{\mathcal{E}} = \frac{1}{\sigma} \vec{j}$. From the discussion around equations 19, 20, and 21, it is clear that for the experiment under consideration the configurational coefficient C is given by

$$C = -\frac{1}{2\sigma_3}(\rho + \xi + \lambda) \tag{47}$$

3. Removal of Relaxation Time Assumption

Before undertaking an analysis of C in terms of other physically measurable properties of matter, some comment is necessary on the collision time approximation of equation 28 on which the derivation of C in equation 47 is based.

We go back to the original equation, 27, to which the collision time equation, 28, is an approximation. We represent by the function $g(\vec{k}, \vec{r})$ the difference between the distribution function in the presence of transport processes and the equilibrium distribution function $f_o(E)$. Equation 27 may then be rewritten as follows:

$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{coll.} = &- g(\vec{k}, \vec{r}) \frac{1}{f_o(E)} \int W(\vec{K}, \vec{k})\, f_o[E(K)]\, d^3K \\ &+ f_o[E(k)] \int W(\vec{k}, \vec{K})\, g(\vec{K}, \vec{r})\, d^3K \\ &+ [1 - f_o(E)] \int W(\vec{K}, \vec{k})\, g(\vec{K}, \vec{r})\, d^3K \\ &+ g(\vec{k}, \vec{r}) \int [W(\vec{k}, \vec{K}) - W(\vec{K}, \vec{k})]\, g(\vec{K}, \vec{r})\, d^3K\end{aligned} \tag{48}$$

In deriving this equation, use has been made of the fact that in equilibrium the time rate of change of the distribution function due to collisions must be zero.

Since in this calculation we are interested primarily in correcting for the effect of the relaxation time approximation, we will consider only the case of negligibly small thermal gradients and magnetic fields. In equation 25, then, we take $\vec{F} = -e\vec{\mathcal{E}}$ and allow the spatial gradient operator ∇ to operate on the electric field only. As will be clear from what follows, these approximations will not disturb the configurational emf calculation. Furthermore, to avoid the complexity of allowing for anisotropy, we assume from the outset that the energy E is a function of the magnitude of k only.

The program to be followed is simply to expand the functions $g_1(\vec{k},\vec{r})$, $g_2(\vec{k},\vec{r})$ and $g_3(\vec{k},\vec{r})$ in terms of surface spherical harmonics $Y_\ell^m(\mu,\varphi) \equiv P_\ell^m(\mu)\, e^{im\varphi}$. The relationship between the vector $\vec{k}$ and the angles $\cos^{-1}\mu$ and ϕ is shown in figure 10. In addition, the scattering probability $W(\vec{K},\vec{k})$ may be expanded also. We then solve the complete Boltzmann equation, 25 plus 27 or 48, successively by orders. We will find that the only essential difference between this procedure and the one used previously is that a set of relaxation times, $\tau_0, \tau_1, \tau_2 \ldots$ enter the calculation instead of the single one, τ, used previously. There is one such time associated

with each ℓ-order of the $Y_\ell^m(\mu, \phi)$ expansion of $g(\vec{k}, \vec{r})$.

The scattering probability $W(\vec{k}, \vec{K})$ may be expanded as follows:

$$W(\vec{k}, \vec{K}) = \sum_\ell \frac{2\ell+1}{4\pi} w_\ell(k, K) P_\ell(\eta)$$

$$= \sum_{\ell, m} \frac{1}{N_{\ell m}} w_\ell(k, K) Y_\ell^m(\mu, \phi) Y_\ell^{-m}(\mu_1, \phi_1) \tag{49}$$

The symbol $N_{\ell m}$ represents the normalization.

$$N_{\ell m} = \int_{-1}^{+1} \int_0^{2\pi} Y_\ell^m(\mu, \phi) Y_\ell^{-m}(\mu, \phi) \, d\mu \, d\phi \tag{50}$$

$$= \frac{4\pi}{2\ell+1} \frac{(\ell+|m|)!}{(\ell-|m|)!}$$

and the function $w_\ell(k, K)$ is given by

$$w_\ell(k, K) = 2\pi \int_{-1}^{+1} W(\vec{k}, \vec{K}) P_\ell(\eta) \, d\eta \tag{51}$$

For the first order calculation we may write the left-hand side of equation 25 as follows upon choosing our coordinate axes so that the z direction is that of the field locally:

$$-e \vec{v} \cdot \vec{\mathcal{E}} \frac{\partial f_0}{\partial E} = -e v \mathcal{E} \frac{\partial f_0}{\partial E} Y_1^0(\mu, \phi) = \left(\frac{\partial f}{\partial t} \right)_{coll.} \tag{52}$$

Since the right-hand side is an integral operator upon the unknown function $g_1(\vec{k},\vec{r})$ we are faced with an integral equation to solve where the forcing function is given by equation 52. The particular solution to this equation requires that the right-hand side have the same spatial dependence as does the left. We therefore choose

$$g_1(\vec{k},\vec{r}) = G_1(k,\vec{r})\, Y_1^0(u,\phi) \tag{53}$$

The function $G_1(k,\vec{r})$ is then found in principle by solving the integral equation

$$\begin{aligned} -e\nu\mathcal{E}\frac{\partial f_0}{\partial E} = &-G_1(k,\vec{r})\int_0^\infty \frac{f_0[E(K)]}{f_0[E(k)]}\, w_0(K,k)\, K^2\, dK \\ &+ f_0(E)\int_0^\infty w_1(k,K)\, G_1(K,\vec{r})\, K^2\, dK \\ &+[1-f_0(E)]\int_0^\infty G_1(K,\vec{r})\, w_1(K,k)\, K^2\, dK \end{aligned} \tag{54}$$

The non-linear last term of equation 48 is omitted because the multiplication of two $g(\vec{k},\vec{r})$ functions makes this a higher order term. In the second order calculation, this term enters as part of the forcing function when the substitution $g(\vec{k},\vec{r}) = g_1(\vec{k},\vec{r})$ is made.

For phonon scattering, as for impurity scattering, the electron exchanges only a very small fraction of its total momentum with the scatterer. This is true for electrons near the top of the Fermi sea which are the ones which participate in transport processes.

Generally, the elastic scattering approximation is not at all a bad one. This approximation is embodied in the statement

$$W(\vec{K}, \vec{k}) \simeq \delta(K-k)\, F(k,\eta) \tag{55}$$

It is to be understood in this equation that the delta function in the magnitude of k is meant to be used with the weighting function K^2 as follows:

$$\int_0^\infty A(K)\,\delta(K-k)\,K^2\,dK = A(k) \tag{56}$$

By virtue of equation 51, the elastic scattering approximation may also be written

$$w_\ell(K,k) = \delta(K-k)\, w_\ell(k) \tag{57}$$

where

$$w_\ell(k) = 2\pi \int_{-1}^{1} F(k,\eta)\, P_\ell(\eta)\, d\eta \tag{58}$$

The function $F(k,\eta)$ is the probability per unit time that an electron with momentum $\vec{k}$ is scattered through an angle $\cos^{-1}\eta$. This function is closely related to the scattering cross-section $\sigma(k,\eta)$ through the density of scatterers and the scattered electron velocity $v(k)$.

Under the elastic scattering approximation equation 54 becomes

$$-e\mathcal{N}\mathcal{E}\frac{\partial f_0}{\partial E} = -[\omega_0(k) - \omega_1(k)]\, G_1(k, \vec{r}) \tag{59}$$

This equation enables us to define a relaxation time, τ_1, where

$$\frac{1}{\tau_1} = \omega_0 - \omega_1 = 2\pi \int_{-1}^{1} (1 - \eta) F(k, \eta)\, d\eta \tag{60}$$

whereupon the solution of the Boltzmann equation to first order becomes

$$g_1(\vec{k}, \vec{r}) = \frac{e}{\hbar} \tau_1 \frac{\partial f_0}{\partial E} \vec{\mathcal{E}} \cdot \nabla_k E \tag{61}$$

A little reflection on the fact that this relaxation time depends only upon the order ℓ of $Y_\ell^m(\mu, \phi)$ and not upon m indicates that the original more elaborate equation 31 for $g_1(\vec{k}, \vec{r})$ is quite correct as it stands if we merely replace τ by τ_1.

In going to second order, the form of the left-hand side of the Boltzmann equation

$$(\vec{v} \cdot \nabla - \frac{e}{\hbar} \vec{\mathcal{E}} \cdot \nabla_k) g_1(\vec{k}, \vec{r}) = (\vec{v} \cdot \nabla - \frac{e}{\hbar} \vec{\mathcal{E}} \cdot \nabla_k)(\frac{e}{\hbar} \tau_1 \frac{\partial f_0}{\partial E} \vec{\mathcal{E}} \cdot \nabla_k E) \tag{62}$$

suggests that we take $g_2(\vec{k}, \vec{r})$ to be of the form

$$g_2(\vec{k}, \vec{r}) = R(k) + \sum_{m=-2}^{2} G_2^m(k, \vec{r})\, Y_2^m(\mu, \varphi) \tag{63}$$

This is because equation 62 may be expanded into a spherically symmetric part plus second order spherical surface harmonic parts. In

dealing with these latter parts we find, in applying the elastic scattering approximation to the resulting integral equation, that a new time τ_2 is introduced where

$$\frac{1}{\tau_2} = \omega_0 - \omega_2 = 2\pi \int_{-1}^{1} [1 - P_2(\eta)] F(k,\eta)\, d\eta \qquad (64)$$

$$= 3\pi \int_{-1}^{1} (1 - \eta^2) F(k,\eta)\, d\eta$$

In trying to apply the elastic scattering approximation to the spherically symmetric integral equation a difficulty arises. From the form of $1/\tau_1 = \omega_0 - \omega_1$ and $1/\tau_2 = \omega_0 - \omega_2$ it is clear that for the spherically symmetrical integral operator we would obtain $1/\tau_0 = \omega_0 - \omega_0 = 0$ or a relaxation time, τ_0, which is infinite. This would imply that a finite forcing function on the left is equal to zero on the right. Therefore, in order to obtain the function $R(k,\vec{r})$ of $g_2(\vec{k},\vec{r})$, we may not assume that the scattering is elastic and it is necessary to solve the complete integral equation for $R(k,\vec{r})$. This integral equation is:

$$\frac{1}{3} e v^2 \tau_1 \frac{\partial f_0}{\partial E} \nabla \cdot \vec{\mathcal{E}} - \frac{1}{3} \frac{e^2}{\hbar} \frac{1}{k^2} \frac{\partial}{\partial k} \left(\tau_1 k^2 v \frac{\partial f_0}{\partial E} \right) \mathcal{E}^2$$

$$= -R(k,\vec{r}) \int_0^\infty \frac{f_0[E(K)]}{f_0[E(k)]} w_0(K,k) K^2 dK \qquad (65)$$

$$+ f_0(E) \int_0^\infty w_0(k,K) R(K,\vec{r}) K^2 dK$$

$$+ [1 - f_0(E)] \int_0^\infty R(K,\vec{r}) w_0(K,k) K^2 dK$$

The first term on the left may be taken to be zero, since $\nabla \cdot \vec{j} = 0$ and the non-linear integral part of $(\frac{\partial f}{\partial t})_{coll.}$ is zero, because there we may apply the elastic approximation safely. Substitution into equation 65 of the elastic scattering approximation of equation 57 will exhibit immediately the inconsistency mentioned.

In order to obtain ultimately a pliable solution to the problem we expand the function $R(k, \vec{r})$ in terms of the eigenfunctions of the homogeneous counterpart of equation 65. We then assume that $R(k, \vec{r})$ may be represented in large part by just one of these eigenfunctions. But for each eigenfunction there corresponds an eigenvalue. For that function which most closely approximates $R(k, \vec{r})$ we call the eigenvalue $(-1/\tau_0)$.

Clearly, the relaxation time τ_0 must be very large since in first approximation it was found to be infinite. It is also apparent, by virtue of the connection with scattering elasticity, that this relaxation time is intimately related to the extent to which collisions are inelastic or equivalently to the extent to which energy is transferred from the electrons to the scatterers on collisions.

It may be further noted that τ_0 is associated with the spherically symmetric part of the increase in the distribution function $f(\vec{k}, \vec{r})$ which in turn varies as $\mathcal{E}^2$ or $\vec{j} \cdot \vec{\mathcal{E}}$. Both of these facts suggest that τ_0 is related to the heating of the electron gas due to the rate at which electrical energy is impressed upon it, $\vec{j} \cdot \vec{\mathcal{E}}$.

Credit is due Professor Richard P. Feynman both for the discovery and for the very penetrating analysis of the physical significance of the relaxation time τ_0.

In the light of the preceding discussion we may write down the second order correction to the distribution function

$$g_2(\vec{k},\vec{r}) = -\tau_2\left(\vec{v}\cdot\nabla - \frac{e}{\hbar}\vec{\mathcal{E}}\cdot\nabla_k\right)\left(\frac{e}{\hbar}\tau_1\frac{\partial f_0}{\partial E}\vec{\mathcal{E}}\cdot\nabla_k E\right)$$

$$+(\tau_0-\tau_2)\frac{1}{3}\frac{e^2}{\hbar}\frac{1}{k^2}\frac{\partial}{\partial k}\left(\tau_1 k^2 v\frac{\partial f_0}{\partial E}\right)\mathcal{E}^2 \qquad (66)$$

It may be mentioned as an aside here that the function $g_2(\vec{k},\vec{r})$ in equation 66 cannot give rise to any current because of its symmetry properties. But if we had included the magnetic field terms in addition to the electric field we would have obtained an additional term which is the analogue of equation 35. Now even if there were orders $\ell = 0, 1, 2,$ and 3 in the spherical harmonic expansion of this term, only the order $\ell = 1$ would give a current. This is because in the current integrand the function $g(\vec{k},\vec{r})$ is multiplied by the velocity vector, each component of which is expandable in terms of $\ell = 1$ spherical harmonics. It is clear therefore that the value of the Hall coefficient H from equation 36 or 37 is quite correct just as it stands if we merely replace τ by τ_1.

We proceed next to third order. The left-hand side of the Boltzmann equation in this order is

$$\left(\vec{v}\cdot\nabla - \frac{e}{\hbar}\vec{\mathcal{E}}\cdot\nabla_k\right)g_2(\vec{k},\vec{r}) = \left(\frac{\partial f}{\partial t}\right)_{coll.} \qquad (67)$$

An expansion of this expression in terms of surface spherical harmonics suggests that orders $\ell = 1$ and $\ell = 3$ be included in the expression for $g_3(\vec{k}, \vec{r})$. In this case no difficulty arises from the elastic scattering assumption and a new relaxation time $\tau_3 = (\omega_0 - \omega_3)^{-1}$ enters. However, only that part of $g_3(\vec{k}, \vec{r})$ with $\ell = 1$ will give rise to a current because, as mentioned previously, the current integrand contains the velocity which depends only upon $\ell = 1$ spherical harmonics. The $\ell = 3$ part of $g_3(\vec{k}, \vec{r})$ will integrate out to zero when multiplied by the vector $\vec{v}$. As a result, we may calculate the current directly from the forcing function expression of equation 67 multiplied by $(-\tau_1)$ since the τ_3 part of $g_3(\vec{k}, \vec{r})$ will yield zero automatically.

$$\vec{J}_3 = \frac{e}{4\pi^3} \int \vec{v}\, \tau_1 \left(\vec{v} \cdot \nabla - \frac{e}{\hbar} \vec{\mathcal{E}} \cdot \nabla_k \right) g_2(\vec{k}, \vec{r})\, d^3k \qquad (68)$$

If we keep only those terms which contain the field $\mathcal{E}$ twice and the ∇ operator once, equation 68 leads directly to the configurational field $\vec{\mathcal{E}}_c = \frac{1}{\sigma} \vec{J}_c$. Upon taking $\nabla \cdot \vec{\mathcal{E}} = \frac{1}{\sigma} \nabla \cdot \vec{J} = 0$ the results of the operation in equation 68 may be summarized in terms of five integrals, three of which are the direct analogues of those in equations 43, 44, and 45.

$$\mathcal{P} = -\frac{1}{3} \frac{e^3}{\hbar^4} \int_0^\infty \tau_1^2 \tau_2 \frac{\partial f_0}{\partial E} \frac{1}{k} \left(\frac{\partial E}{\partial k} \right)^3 \frac{k^2}{\pi^2}\, dk \qquad (69)$$

$$\mathcal{Q} = -\frac{1}{3} \frac{e^3}{\hbar^4} \int_0^\infty \tau_1^2 \tau_2 \frac{\partial f_0}{\partial E} \left(\frac{\partial E}{\partial k} \right)^2 \frac{\partial^2 E}{\partial k^2} \frac{k^2}{\pi^2}\, dk \qquad (70)$$

$$\mathcal{S} = -\frac{1}{3}\frac{e^3}{\hbar^4}\int_0^\infty \tau_1^{7/5}\tau_2^{3/5}\frac{\partial}{\partial E}\left(\tau_1^{3/5}\tau_2^{2/5}\right)\frac{\partial f_0}{\partial E}\left(\frac{\partial E}{\partial k}\right)^4\frac{k^2}{\pi^2}\,dk \qquad (71)$$

$$\mathcal{Q}_0 = -\frac{1}{3}\frac{e^3}{\hbar^4}\int_0^\infty \tau_1^2(\tau_0-\tau_2)\frac{\partial f_0}{\partial E}\left(\frac{\partial E}{\partial k}\right)^2\frac{\partial^2 E}{\partial k^2}\frac{k^2}{\pi^2}\,dk \qquad (72)$$

$$\mathcal{S}_0 = -\frac{1}{3}\frac{e^3}{\hbar^4}\int_0^\infty \tau_1\frac{\partial}{\partial E}\left[\tau_1(\tau_0-\tau_2)\right]\frac{\partial f_0}{\partial E}\left(\frac{\partial E}{\partial k}\right)^4\frac{k^2}{\pi^2}\,dk \qquad (73)$$

In terms of these integrals, the configurational current may be written as

$$\begin{aligned}\vec{J}_c = &\left(\frac{2}{3}\mathcal{Q}_0 + \frac{1}{3}\mathcal{S}_0\right)\nabla\mathcal{E}^2 \\ &+ \left(\mathcal{P}+\mathcal{Q}+\mathcal{S}\right)\vec{\mathcal{E}}\cdot\nabla\vec{\mathcal{E}} \\ &+ \frac{1}{2}\left(-\mathcal{P}+\mathcal{Q}+\mathcal{S}\right)\nabla\mathcal{E}^2\end{aligned} \qquad (74)$$

This expression may be compared with that of 46. In the limit $\tau_0 = \tau_1 = \tau_2 = \tau$, equation 74 reduces to equation 46 when $\nabla\cdot\vec{\mathcal{E}} = 0$.

By virtue of the discussion around equations 18, 19, 20, and 21, it is clear that in the experiment reported here the terms involving the relaxation time τ_0 were not measured. The corrected configurational coefficient is given by

$$\mathcal{C} = -\frac{1}{2\sigma^3}\left(\mathcal{P}+\mathcal{Q}+\mathcal{S}\right) \qquad (75)$$

and this is the quantity to be compared with the experimental curves.

4. Connection with Familiar Measurable Quantities

To arrive at an expression for the configurational coefficient in terms of other known quantities, we will evaluate the integrals P, Q, and S for special cases. First, we note that it is not unreasonable to assume that the energy dependence of τ_1 and of τ_2 is the same even though they may differ numerically because of the angular dependence of the differential scattering cross-section. On this assumption we expect the ratio τ_2/τ_1 to be simply a number of the order of unity and independent of energy. In this case we have simply that $P = \frac{\tau_2}{\tau_1}\, p$, $Q = \frac{\tau_2}{\tau_1}\, q$, and $S = \frac{\tau_2}{\tau_1}\, s$. It is to be understood that τ and τ_1 are used interchangeably now. We further assume that $\tau(E) = \text{const.} \times E^{-\nu}$ so that $\frac{d\tau}{dE} = -\nu\frac{\tau}{E}$. This is not unreasonable, since the energy dependence of τ usually comes from the density of states which goes as $E^{1/2}$ for the usual limiting cases which will be considered here. Since it is only necessary to know this energy dependence at the Fermi surface, the latitude afforded by the exponent ν is more than sufficient. For phonon scattering of relatively free electrons at temperatures below the Debye temperature of the metal generally $\nu = 1/2$ because of the inverse dependence of τ on the density of states as mentioned above. For higher temperatures $\nu = 3/2$ (7).

The most general case we will consider is that of a metal with energy bands of standard form (8). This is the case of a metal in which there are two overlapping bands. One conduction band is nearly full. The next band up has some states of lower energy than the energy of the highest states in the lower band, so that electrons which would ordinarily fill up the lower band spill over into the higher band. They occupy the states at the bottom of the higher band. If these bands were separated by a distinct energy gap, the material would be an insulator, a semiconductor, or an ideal metal, depending upon the width of the gap and upon whether, after filling the lower band, there are still electrons available to occupy states in the higher band. The details of the principles involved in exploiting this very common model are explained quite nicely in several places in the literature (8, 9).

In this model for the isotropic case we assume that in the upper band, which is the one which conducts with negative carriers, we have

$$E = \alpha_n \frac{\hbar^2}{2m} k^2 \tag{76}$$

In the positive carrier band, which is almost full, we have

$$E = E_0 - \alpha_p \frac{\hbar^2}{2m} k^2 \tag{77}$$

The energy E_o denotes the energy at the top of the band. The constants α reflect the curvature in k-space of the energy surfaces in each of the two bands. Using the same notation as that of Mott and Jones (9) we have $\alpha = m/m^*$ where m^* is the effective mass. Electrons in each of the two different bands have different effective masses.

Keeping in mind the fundamental two-overlapping-band picture, the P integral may be obtained in terms of E and $E' = E_o - E$ as follows

$$\frac{3m^2}{e^3} P = \left(\frac{\tau_2}{\tau_1}\right)_n \alpha_n^2 \int_o^\infty \tau_n^3(E)\, E \frac{d}{dE}[N_n(E)]\left(-\frac{\partial f_o}{\partial E}\right) dE$$

$$- \left(\frac{\tau_2}{\tau_1}\right)_p \alpha_p^2 \int_o^\infty \tau_p^3(E')\, E' \frac{d}{dE'}[N_p(E')] \frac{\partial f_o}{\partial E'} dE' \qquad (78)$$

Here $N_p(E')$ refers to the number of unoccupied states per unit volume in the lower band between the energy surface E and the limiting energy E_o. The number $N_n(E)$ is the number of states occupied by electrons out to the energy E in the upper band. The N(E) are densities in physical space, not in k-space. The density of states in energy is $\frac{d}{dE} N(E)$. Since $N_p(E')$ represents unoccupied states or holes in the distribution and the effects associated with it will generally act as if they were caused by positive carriers, we may associate $N_p(E_o - E_F) \equiv N_p$ with the total effective number of positive carriers per unit volume. Similarly, $N_n(E_F) \equiv N_n$ is the effective number of negative or electronic carriers per unit volume.

Both the Q and S integrals may be put in a form similar to that of equation 78. The integrals in equation 78 are of a form which may be evaluated by means of a well-known formula involving the Fermi distribution $f_o(E)$. This is given by (10)

$$\int_0^\infty \Phi(E)\left(-\frac{\partial f_0}{\partial E}\right) dE = \Phi(E_F) + \frac{\pi^2}{12}(KT)^2 \frac{d^2\Phi}{dE^2}\Big|_{E_F} + \dots \quad (79)$$

For $KT \ll E_F$ which is usually the case, the Fermi distribution is close enough to a unit step function so that only the first term in this formula is needed.

Since in k-space the energy surfaces E and E' are spheres, both of the densities $\frac{d}{dE}$ N(E) may be obtained from

$$\int_{\text{Surface } E=\text{const.}} \frac{1}{4\pi^3} d^3k = \frac{d\,N(E)}{dE} dE = \frac{1}{2\pi^2\hbar^3}\left(\frac{2m}{\alpha}\right)^{3/2} E^{1/2} dE \quad (80)$$

where for the hole density we merely replace E by E' and use the subscript p were applicable. Equation 80 yields the convenient formula

$$\frac{dN}{dE} = \frac{3}{2}\frac{N}{E} = \frac{1}{3}\left(\frac{3}{\pi}\right)^{4/3} \frac{m}{\alpha\hbar^2} N^{1/3} \quad (81)$$

Using these results we may evaluate the P integral and the Q integral. For these we find

$$P + Q = 2P = 2\frac{e^3}{m^2}\left[\left(\frac{\tau_2}{\tau_1}\right)_n \tau_n^3 \alpha_n^2 N_n - \left(\frac{\tau_2}{\tau_1}\right)_p \tau_p^3 \alpha_p^2 N_p\right] \quad (82)$$

To evaluate the S integral we note that $1/\tau(E)$ is related to the a priori probability per unit time that a collision increases the population of states with energy E. This probability is generally proportional to the density of states $\frac{dN(E)}{dE}$ at E. Therefore, one expects that if the density of states increases with increasing E, then $\tau(E)$ will decrease with increasing E and vice versa. As a result, we expect that $\frac{d\tau}{dE}_n \cdot \frac{d\tau}{dE}_p < 0$ since the density of states increases with E in one band and decreases with E (not E') in the other. If we understand the quantity ν to indicate the extent of this energy dependence by virtue of

$$\frac{d\tau_p}{dE} = -\frac{d\tau_p}{dE'} = \nu_p \frac{\tau_p}{E} \quad ; \quad \frac{d\tau_n}{dE} = -\nu_n \frac{\tau_n}{E} \tag{83}$$

then the S integral may be written

$$S = -2\frac{e^3}{m^2}\left[\left(\frac{\tau_2}{\tau_1}\right)_n \tau_n^3 \alpha_n^2 \nu_n N_n - \left(\frac{\tau_2}{\tau_1}\right)_p \tau_p^3 \alpha_p^2 \nu_p N_p\right] \tag{84}$$

By a process of evaluation exactly analogous to that used for the P, Q, and S integrals, the conductivity may be evaluated from equation 34 to yield the well-known formula

$$\sigma = \frac{e^2}{m}\left(\tau_n \alpha_n N_n + \tau_p \alpha_p N_p\right) \tag{85}$$

Similarly the Hall coefficient may be obtained from equation 37 to give the familiar formula

$$H = -\frac{1}{\sigma^2}\frac{e^3}{m^2}\left(\tau_n^2\alpha_n^2 N_n - \tau_p^2\alpha_p^2 N_p\right) \qquad (86)$$

In terms of these same parameters, the additional relaxation time τ_2 and the numbers

$$\nu_n = \frac{d\ln\tau_n}{d\ln E}\Big|_{E_F} \quad \text{and} \quad \nu_p = \frac{d\ln\tau_p(E')}{d\ln E'}\Big|_{E_0-E_F} \qquad (87)$$

the configurational coefficient is given as

$$C = -\frac{e^3}{\sigma^3 m^2}\left[(1-\nu_n)\left(\frac{\tau_2}{\tau_1}\right)_n \tau_n^3\alpha_n^2 N_n - (1-\nu_p)\left(\frac{\tau_2}{\tau_1}\right)_p \tau_p^3\alpha_p^2 N_p\right] \qquad (88)$$

For all the cases considered by Wilson (7) the quantity ν varies between 1/2 and -3/2, depending upon the band structure. Since it is expected that τ_2/τ_1 is of the order of 1, we will lump into the symbol γ the rough average of both of the multipliers $(1-\nu)\frac{\tau_2}{\tau_1} = \gamma$. The number γ is of the order of unity.

If one of the carrier types predominates over the other, then the equations 85, 86, and 88 reduce to simpler form. For example, if the negative carriers predominate, the following relationship holds.

$$C = -\gamma\frac{m}{e}\frac{1}{\alpha}H^2 = \gamma\tau\frac{H}{\sigma} \qquad (89)$$

In the special case that $\tau_2 = \tau$ and $\nu = 1/2$, equation 89 reduces exactly to equation 3 if, in the latter m is replaced by m* and the particle density n is understood to mean the effective carrier density N used above. If we use for the quantity n in equation 22 the definition

given in Mott and Jones (11) for the effective number of free electrons, then we must replace m by m^2/m^*.

Equation 89 shows that, in general, a measurement of the three coefficients σ, H, and C will yield directly the relaxation time and the effective mass $m^* = m/a$ of carriers.

IV. DISCUSSION

1. Application to Bismuth

In the case of bismuth the five valence electrons per atom are just sufficient to fill a Brillouin zone corresponding to the rhombohedral lattice of the bismuth structure (12,13). If the next higher zone were separated by an energy gap, bismuth would be a semiconductor or an insulator, depending upon the width of the gap. In fact, though, there is no energy gap and the next higher zone contains some levels which are lower in energy than the highest levels in the primary zone. Therefore, some of the electrons which would have gone to complete the filling up of the lower zone spill over into the higher zone. These act as negative carriers because they are now at the bottom of a conduction band. The empty levels in the depleted primary zone act as positive carriers. It is this exchange number, N, of levels depleted or extra levels occupied in the higher band which defines the number of carriers. For bismuth then

$$N_p = N_n = N \tag{90}$$

We may define the ratio x to indicate the relative difference in mobility between the electrons and the holes so that

$$\chi = \frac{\tau_n \alpha_n - \tau_p \alpha_p}{\tau_n \alpha_n + \tau_p \alpha_p} \tag{91}$$

The ratio x varies between the limits -1 through zero to +1 in the case of very high mobility of the negative carriers as compared to that of the holes. In terms of the quantity x and the condition $N_p = N_n = N$, the Hall coefficient for bismuth may be written from equation 86

$$H = -\frac{1}{eN}\,x \tag{92}$$

2. Three Alternate Assumptions for Relating C, σ and H.

With regard to the coefficient C in the presence of two carrier types, there are a number of possibilities. It is possible that the effective mass of each of the carrier types is about the same, so that we may take $a_n = a_p = a$. In this case the coefficient C becomes

$$C = -\gamma \frac{m}{e}\frac{H^2}{a\,x}\left[1 - \frac{\tau_n a_n \tau_p a_p}{(\tau_n a_n + \tau_p a_p)^2}\right] \tag{93}$$

where the number in brackets can only be between $[1 - \frac{1}{4}] = 3/4$ and unity. In this case, we should get an estimate of the quantity $1/a\,x$ from the experimental data by evaluating the ratio $[-(e/m)(C/H^2)]$ given in line four of table 1. If x is taken to be of the order of unity, the implied values of m*/m differ considerably from those obtained by magnetic susceptibility measurements. The strong diamagnetic properties of bismuth are usually attributed to a very small effective mass for the electrons (14,15). This is because the Landau diamagnetism of the electron gas depends upon the reciprocal of the effective mass. The spin paramagnetic properties depend linearly upon the

effective mass. Since the electronic part of the susceptibility is negative or diamagnetic, the conclusion is that α or the reciprocal effective mass is large. If α were unity, then the electron gas would be paramagnetic because the Landau diamagnetism is just one-third of the spin paramagnetic susceptibility.

The formulas for the Landau diamagnetic susceptibility of the electron gas depend heavily upon the directional properties of the crystal. Mott and Jones (14) conclude from the measurements that the effective mass along the principal axis is several orders of magnitude larger than the effective mass in directions perpendicular to this axis. By further argument they conclude that $\alpha \sim 1$ along the principal axis, making $\alpha \sim 10^2$ in the lateral directions. They also point out, however, that the formula is such that if α were assumed to be about unity in the lateral directions, then the same magnitude of the diamagnetic susceptibility would result if along the principal axis α were quite small. In this case the average α would lie between 10^{-6} and unity. And considerably better agreement would result with the configurational estimates of table 1.

The alternative to the implication that α is considerably smaller than all previous estimates of this quantity for bulk bismuth is that the mobility of positive carriers is very close to that for negative carriers so that x is very small. The relative difference, x, would have to be of the order of $x \sim 10^{-4}$ if $\alpha \sim 1$.

A second possible basis on which to evaluate C is on the assumption that the relaxation times for positive and negative carriers are about the same, so that $\tau_n \sim \tau_p \sim \tau$. In this case, the second equality of equation 89 holds for C. Line 5 of table 1 gives the values of this relaxation time calculated for this case. These relaxation times are about two orders of magnitude larger than that usually estimated for bismuth and about four orders of magnitude larger than those for more ideal metals.

The next assumption to consider is that whereas the relaxation times or the effective masses in the two bands may be quite different, the mean free path for collision is about the same for either band. This case has a better physical basis than the preceding two cases. This condition corresponds to the assumption that x is very small or, equivalently, that $\tau_n a_n \sim \tau_p a_p$ because the mean free path is proportional to the product of τ and a.

$$\ell = \tau v_F = 3\left(\frac{\pi}{3}\right)^{2/3} \tau \propto \frac{\hbar}{m} N^{1/3} \tag{94}$$

Now even though x is small, the respective relaxation times for the two bands probably differ considerably. If this is the case, the fact that the measurements show the same sign of carrier for both H and C indicates that $\tau_n > \tau_p$. In this event, the coefficient C may be written approximately as

$$C = \frac{\gamma}{4\chi}\,\tau_n\,\frac{H}{\sigma} = -\frac{\gamma}{8\chi^2}\,\frac{m}{e}\,\frac{1}{\alpha_n}\,H^2 \qquad (95)$$

Referring to table 1, it may be shown that a value of x = .002 will yield $\alpha_n \sim 1$ and $\tau_n \sim 2 \times 10^{-12}$ seconds, both of which values are in reasonable agreement with the usual estimated rough averages for bismuth. Unfortunately, the point of view that coincidentally the difference in mobilities of each of the carrier types just happens to be exceedingly small for the particular material of which the specimens were made seems rather ad hoc and therefore unsatisfactory. It is interesting to note, however, the following correlation. With the value x = .002 and the assumption that the effective number of carriers per atom is about 3×10^{-4}, which is just the usual number for bismuth, the film thickness b may be estimated. The values of b so obtained are shown in line 6 of table 1. These estimates agree reasonably well with those obtained from the resistance estimate of b if the bulk resistivity of $\rho = 10^{-4}$ ohm-cm is used. The values of b calculated in the latter way are on line 7 of table 1.

To appreciate the fact that this correlation is not simply a property of the measurables independent of x we note that both the quantities $(1/\sigma b)$ and (H/b) are directly measurable and are listed on lines 1 and 2 of table 1. If we assume that the specimen conductivity, σ, is equal to the bulk conductivity of bismuth, then b is found to be of the order of 100 Å, as shown in the table. However, if H is also

assumed to correspond to the value for bulk bismuth so that $H = 0.6\ cm^3/coulomb$, the implication would be that the same thickness b, by this estimate, is of the order of 30,000 Å. Taking the value $x = .002$, therefore, not only brings τ and α into better coincidence with accepted values but also brings the Hall effect measurement into coincidence with the conductivity measurements.

3. Relation between Theory and Experiment.

The preceding discussion serves a two-fold purpose. First, it indicates that the specimen material does not act at all like bulk bismuth. This is independent of the configurational measurements. Either the conductivity or the Hall coefficient or both are not that of bulk bismuth.

Secondly, it is clear from the multiplicity of possible assumptions on how to evaluate the coefficient C in terms of σ and H that the number of theoretical parameters available exceeds the number of measurable quantities through which theory and experiment may be compared. In this connection it should be noted that the oversimplified relation of equation 17 is only applicable in special cases. Therefore, the direct comparison given in the discussion around that equation is not generally valid as was intimated in that section.

These two difficulties are magnified considerably by the necessity of dealing with bismuth which is characteristically anomalous. The formulas would be more amenable to interpretation in the case of

a more ideal metal. However, the experiment is considerably more difficult in metals like sodium or potassium.

On the first three lines of table 1 is a summary of the directly measurable quantities relevant to this experiment. On the following two lines are listed those combinations of the measurables which, theory suggests, should yield significant numbers relevant to the properties of the specimen material. In general, the ratio $(-eC/mH^2)$ of line 4 should yield some effective weighted mean of the reciprocal of the curvature, α, of the energy surfaces in k-space. And the ratio $(\sigma C/H)$ should yield some effective weighted mean relaxation time. The weighting factor or the averaging procedure will depend upon the particular substance. It will depend upon the number and type of carriers present and upon their relative mobilities.

Now for pure bulk bismuth the usual estimates of $m^*/m = 1/\alpha$ lie between 0.01 and 1.0, depending heavily upon crystal orientation(14). And τ is thought to be of the order of 10^{-12} seconds. A considerable divergence exists between these numbers and those derived on lines 4 and 5 of table 1. If the latter set of empirically derived numbers did approximate the expected values for bulk bismuth, then this would be taken as strong evidence that the theory given in the preceding pages does indeed describe the observed effect. Because of the discrepancy, however, we are left with three alternatives regarding the interpretation of the experimental and theoretical results.

4. Alternative Interpretations of Results.

The first alternative is simply to accept the experimental results as relating directly to the theory given here and to throw the entire burden of producing quantitative agreement upon the adjustable and relatively unknown parameter x. As has been pointed out, it is possible to achieve some measure of quantitative consistency by choosing x properly. On this view the experiment is nothing more than a way of obtaining the relative mobility of the two carrier types.

Unfortunately, this outlook is unsatisfactory on two counts. First, the value that is demanded for x is rather extreme. It seems improbable that the specimen material is such that, by coincidence, the mobility or mean free path of holes differs by only two parts in one thousand from that of negative carriers. The second count against the view under consideration is the following. Basically there is really no a priori reason to adjust x so as to bring the measured α and τ into coincidence with the values for bulk bismuth. Since it has been shown by the measurements of (H/b) and $(1/\sigma b)$ that the specimen does not behave like bulk bismuth, the point of comparing α and τ with the accepted values for bismuth loses its strength. This is especially true since the indications are that a discrepancy of almost three orders of magnitude exists between the specimen properties and those accepted for bismuth in areas which do not involve the configurational emf.

The second alternative which is available for interpreting the experimental results with regard to the theory is to recognize that the results are inconclusive.

The specimen is a thin film so that all the usual anomalies of thin films add to the confusion of trying to correlate the observed effects with the known properties of the bulk material. The bismuth used in the experiment was known to be quite impure even before evaporation. And special precautions were not taken in preparing the glass substrate nor in very careful cleaning of the evaporation boat nor in controlling the time, rate, or temperature of evaporation. As a result, the composition of the final specimen could hardly be expected to be pure bismuth. Now the properties of bismuth are known to be quite sensitive to the amount and kind of impurities in it. A change of resistance in bismuth by an order of magnitude due to alloying with a fraction of a percent of tin is realizable. There is some possibility, then, that the discrepancy between theory and experiment is due to the inapplicability of the theory to the particular specimen used.

The basic reason for this view is that there are too many factors involved whose relation to the investigation at hand is indeterminate. For example, it would seem reasonable to understand the large divergence between the properties of the thin film evaporated bismuth specimens and those of bulk bismuth before attempting to interpret experiments which purport to yield new information on bismuth. Perhaps the bulk bismuth model - using $N_p = N_n = N$ etc. - on which the theory is based may not be applicable to impure thin film specimens. Perhaps there are unusual surface scattering mechanisms to

be included in the theory. Or perhaps surface energy states play an important role. These have been completely omitted from the theory. Impurities also affect the results because the concentrations N_n and N_p are changed by impurities. In addition, impurities affect the mobilities or, equivalently, $\tau_n a_n$ and $\tau_p a_p$. In connection with these qualities, but independent of the thin film nature of the specimen, some of the discrepancy may be attributable to failure to include in the theory any account of hole-electron annihilation or what might be called interband scattering.

On this view then all that has been established here is the qualitative existence of the expected effect. This was done in such a way, however, as to eliminate other possible explanations for the observed result. Further experimental evidence is needed on substances which are more consistent in their properties and which are more amenable to theoretical analysis to verify the quantitative features. Unfortunately, these substances are the very ones on which it is most difficult to measure the configurational emf.

The third and last alternative open for consideration is the view that the observed results are not related to the effect for which a theory has been given here. The obvious inference is that there is some other cause. However, on examining other possible explanations, nothing has yet appeared which can be said to fit the observed results. The self-Hall effect or self-magnetic current deflection hypothesis, for example, has been shown to be untenable in a previous section. Thermoelectric origins have been discussed. And spurious signals from rectifying contacts in the bridge are distinguishable from the desired signal.

Before concluding the discussion on this last alternative, however, it is well to mention some considerations involving anisotropy. A conceivable explanation of the observed effects lies in the possibility that the evaporated specimen consists of large enough crystallites to exhibit the anisotropies common to bulk bismuth crystals. If anisotropy obtained, then the ordinary first order thermal effects would not necessarily cancel out around the circuit of figure 9 and neither would the τ_o part of the configurational emf. Both of these would then produce a square law current dependence as is observed. However, for this to be the case, there would have to be crystallites present of a size approaching the width of the potential probe or of the constricting region.

In fact, it is not uncommon to find some structure and grain in an evaporated specimen. The evaporation then consists of a large number of small crystallites oriented at random. However, investigations carried out on thin films indicate that the grain size is usually no larger than about 500 Å (16). Even if crystal flakes of the order of 1000 Å in diameter were formed, about 200 of these would be needed to span the probe region or the constricting region. One would expect, therefore, that the anisotropy exhibited by the film should certainly be no more than about 1/2% of that exhibited by the bulk material. Since the thermoelectric power of bismuth is of the order of 70 microvolts per degree and the resistance of each of the constriction

arms of the specimen is of the order of 1,000 ohms and the observed effect is of the order of 7 $\mu v/(ma)^2$, one may make the following estimate. If the effect were due to the anisotropic thermoelectric emf , then the above numbers lead to a heating efficiency in the specimen of the order of 20,000 degrees/watt of electric power input. This is an untenably high rate of temperature increase with electrical power input. Furthermore, since the melting point of bismuth is 544° K, observed signals as high as 100 μv in specimen I of figure 6 would imply, on the hypothesis of a 1/2% anisotropy thermoelectric emf, that the specimen attained temperatures over 30° above its melting point. This strongly suggests that the slight specimen anisotropy that may be present is not sufficient to account for the results except under very unusual circumstances.

In addition to the above evidence against thermoelectric origins, there is the additional fact that one experimental run was done with the specimen immersed in liquid helium II at about 1° K. By virtue of the well-known superfluid properties of helium below the λ point, it is expected that the temperature of the specimen should be relatively uniform regardless of the non-uniformity of the current density because of the extraordinary heat transport properties of helium II. A large configurational signal was detectable from the specimen in this condition as well as at room temperature. The signal was quite different

in magnitude, of course, from the one received at room temperature. This result does not conclusively deny thermoelectric origins because of the possibility of very local heating in a film of liquid around the specimen, even in liquid helium II. However, it is strong evidence against the thermoelectric hypothesis and, together with the preceding calculation, all but eliminates this explanation of the observed effect.

V. SUMMARY AND CONCLUSION

An experiment was undertaken to detect a configurational emf due to a Bernoulli effect in the electron gas in evaporated thin films of bismuth. The experiment was successful in that (a) a definite configurational emf was observed, (b) this emf varied as the square of the current and, (c) there was the expected correlation between the sign of the emf and the sign of carriers as established by Hall measurements. The experimental investigation was unsuccessful in determining the dependence of the emf upon area and in establishing quantitative agreement with the theory given here.

Upon consideration of the discussion of the previous section the indications are that the observed emf did indeed arise from the hypothesized Bernoulli effect. The lack of correlation between theory and experiment is attributed to the inapplicability of the theory to the particular specimens used. The reasons for this conclusion may be summarized as follows: Another equally tenable explanation for the observed effect is not evident and there is considerable reason to suspect that the special nature of the specimens alters the apparent magnitude of the effect.

With regard to the first of these reasons attention has been devoted to other conceivable hypotheses and they have all been found wanting on either experimental or theoretical grounds.

The special attributes of the specimens are contained in their thin film nature and in the fact that they are evaporated (and therefore contaminated or oxidized). The theory applies only to bulk material in which all dimensions are considerably larger than a mean free electron path and to pure bismuth - not, for example, to oxides of bismuth. The latter may actually constitute a large fraction of the sample. These factors have a considerable effect on the apparent properties of the sample often extending to orders of magnitude differences from bulk material properties(17). Since the theory does not include size or impurity effects the possibility exists that the lack of correlation is due to this omission.

By far the most important attribute of thin evaporated films which may cause an experimental mis-estimate of the configurational coefficient C is the phenomenon of coagulation upon evaporation. Films that are evaporated at room temperature are often found to consist of an intricate matrix of tiny islands of material connected only intermittently to one another (18, 19). A current flow line through the film must follow a tortuous path leading from island to island rather than the straight line estimated from a macroscopic view of the specimen. The net result of this phenomenon is that the actual effective cross-sectional flow area presented to a current is considerably smaller than that estimated by a visual observation of the film. Such an effect implies that the observed configurational emf will appear to be considerably

larger than that estimated from a calculation using the apparent macroscopic cross-sectional width of the specimen. Alternatively, this means that the experimental estimate of C based on the visual width of the specimen will exceed the actual value by a factor (apparent width/actual effective width)^2. This factor could easily approach 10^4 or more. The coagulation phenomenon would produce an effect, therefore, such as to cause an apparent discrepancy between theory and experiment just like that observed both qualitatively and quantitatively. It is to such effects that the discrepancy is attributed.

The lack of a definite quantitative verification of the Bernoulli origin of the configurational emf is unfortunate in that use of this effect as an auxiliary tool in solid state measurements must await such a verification. It is hoped that the final confirmation will come in the near future, and that this effort will have contributed in some measure to that confirmation.

VI. TABLE AND FIGURES

Table 1. Summary of data on the specimens of figure 6 (Specimen I) and of figure 7 (Specimen II).

	Specimen I	Specimen II	Units and Comments
1. $1/\sigma b$	90	70	ohms
2. $V_H/IB = H/b$	-12	-25	μvolts/ma-kilogauss
3. $w^2V_c/I^2 = C/b^2$	-3.8×10^3	-3.8×10^3	μvolts-(microns)^2/ma^2
4. $\overline{(m^*/m)} = -eC/mH^2$	4.8×10^4	1.1×10^4	
5. $\overline{\tau} = \sigma C/H$	3.5×10^{-10}	2.2×10^{-10}	seconds
6. $b = -(x/eN)(H/b)^{-1}$	120	60	Å where x = 0.002
7. $b = \rho/(1/\sigma b)$	110	180	Å where $\rho = 10^{-4}$ ohm-cm

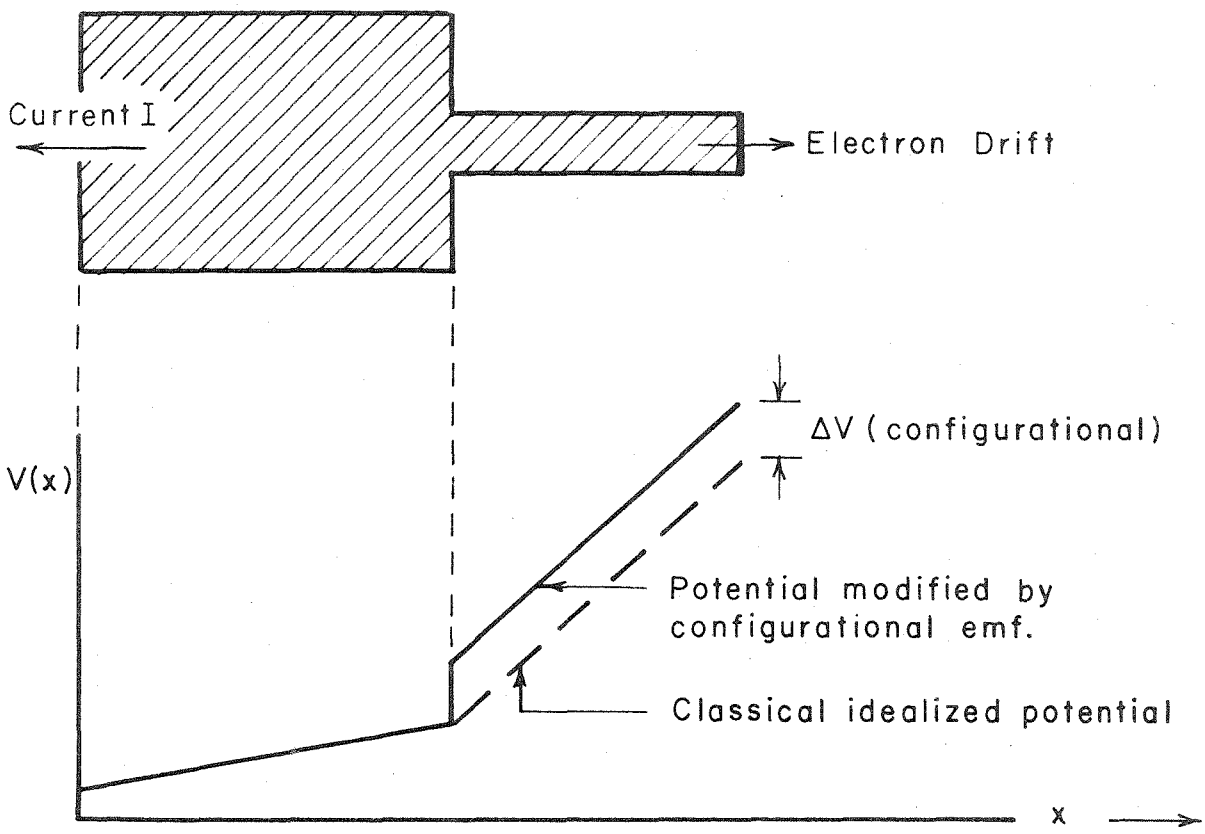


Fig. 1.- Expected potential vs. distance due to constricted current flow in the conducting specimen shown.

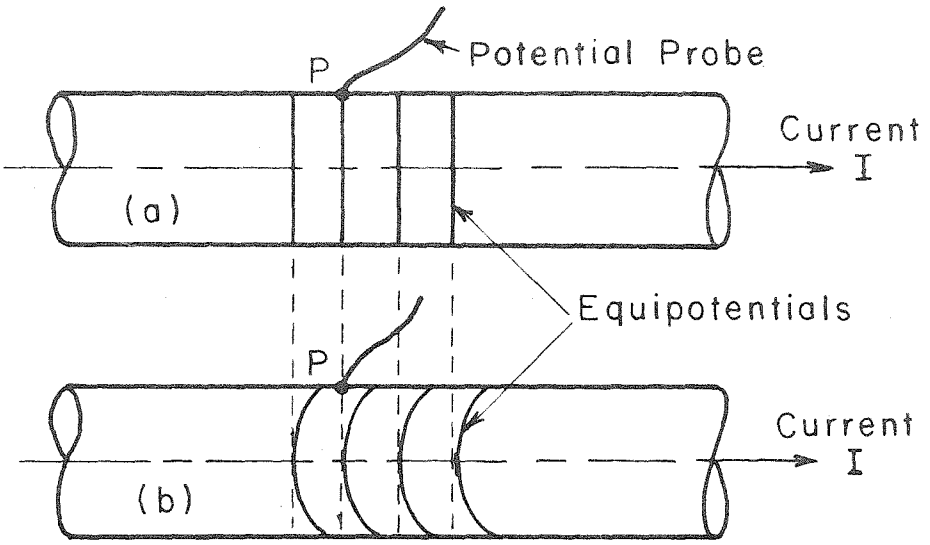


Fig. 2. - a. Equipotentials in a current carrying cylinder as deduced from Ohm's law, $\vec{j} = \sigma \vec{\mathcal{E}}$. The probe at P senses the potential V.

b. Equipotentials for Ohm's law as modified by the magnetic field of the current, $\vec{j} = \sigma[\vec{\mathcal{E}} - (1/ne)\,\vec{j} \times \vec{B}]$. A potential probe at P senses the voltage $V + V_m$.

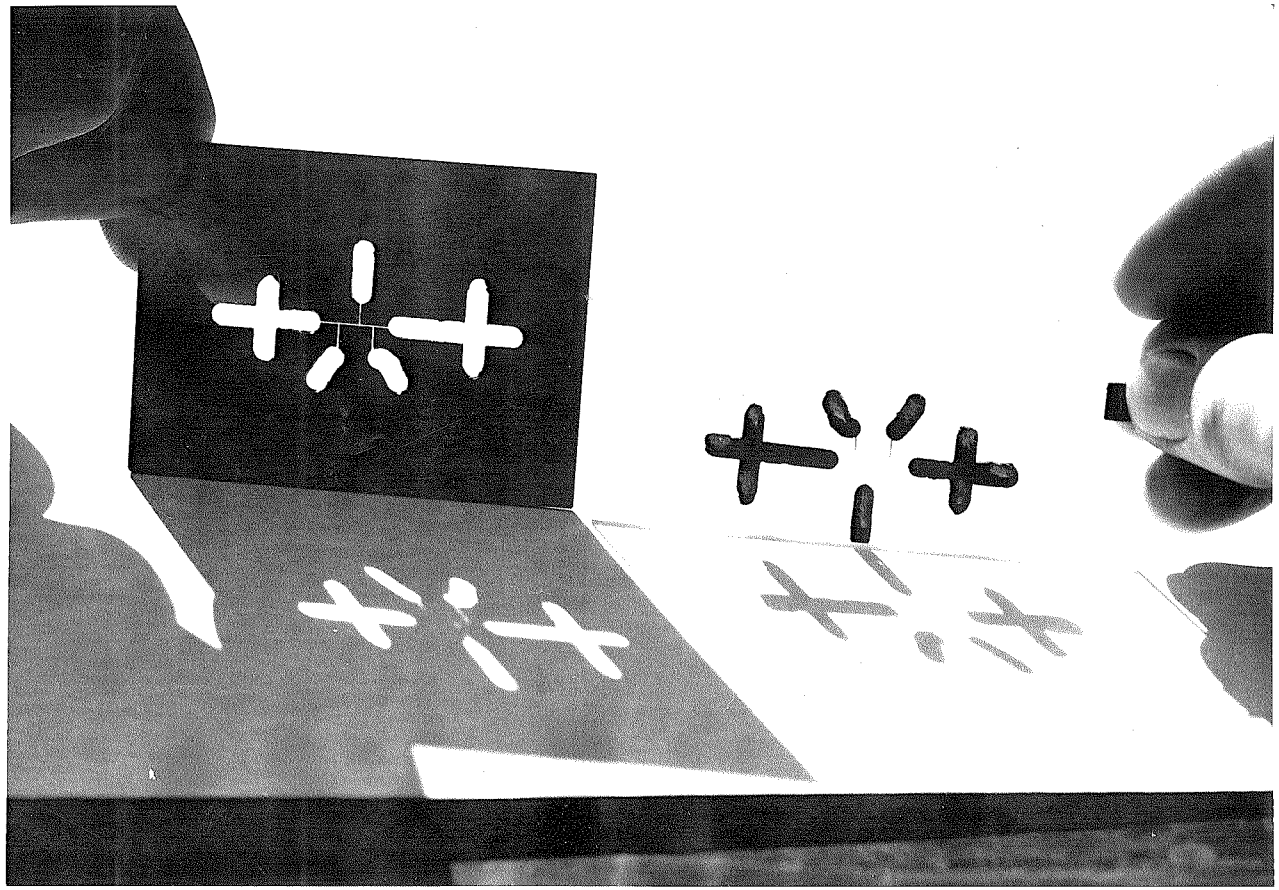


Fig. 3.- Evaporation mask and sample evaporated specimen on a microscope slide.

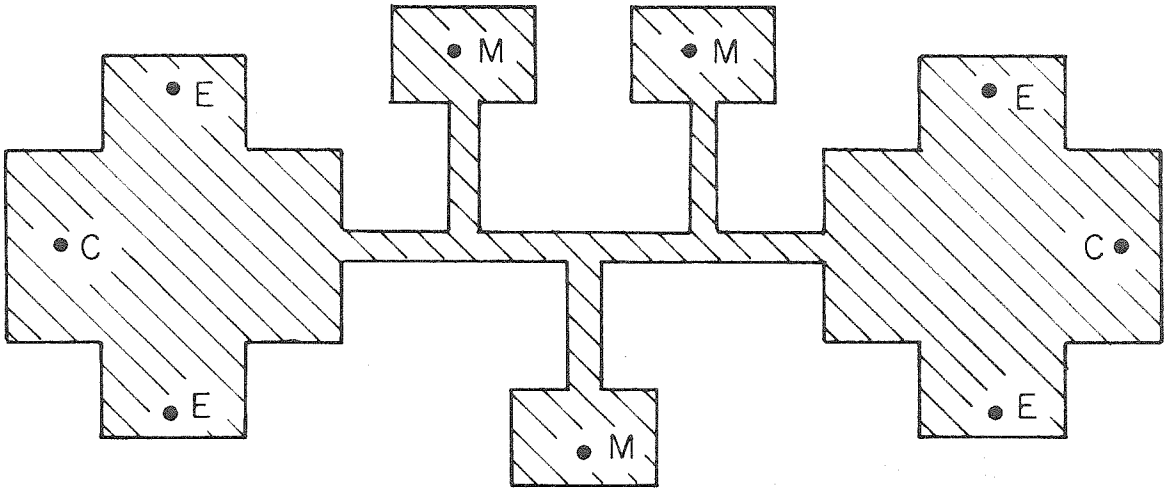


Fig. 4. - Schematic diagram of specimen configuration.

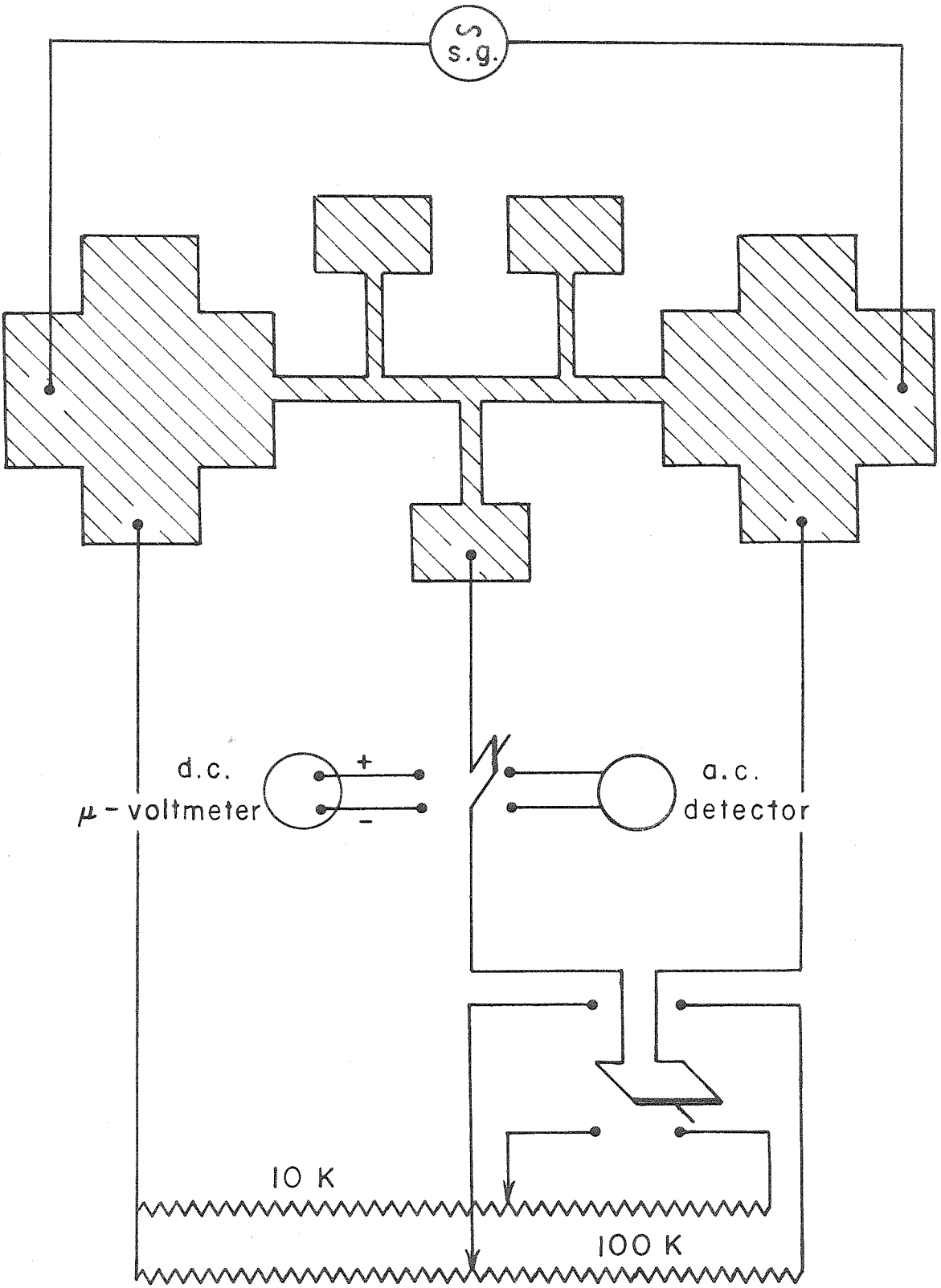


Fig. 5. - Detection circuitry schematic.

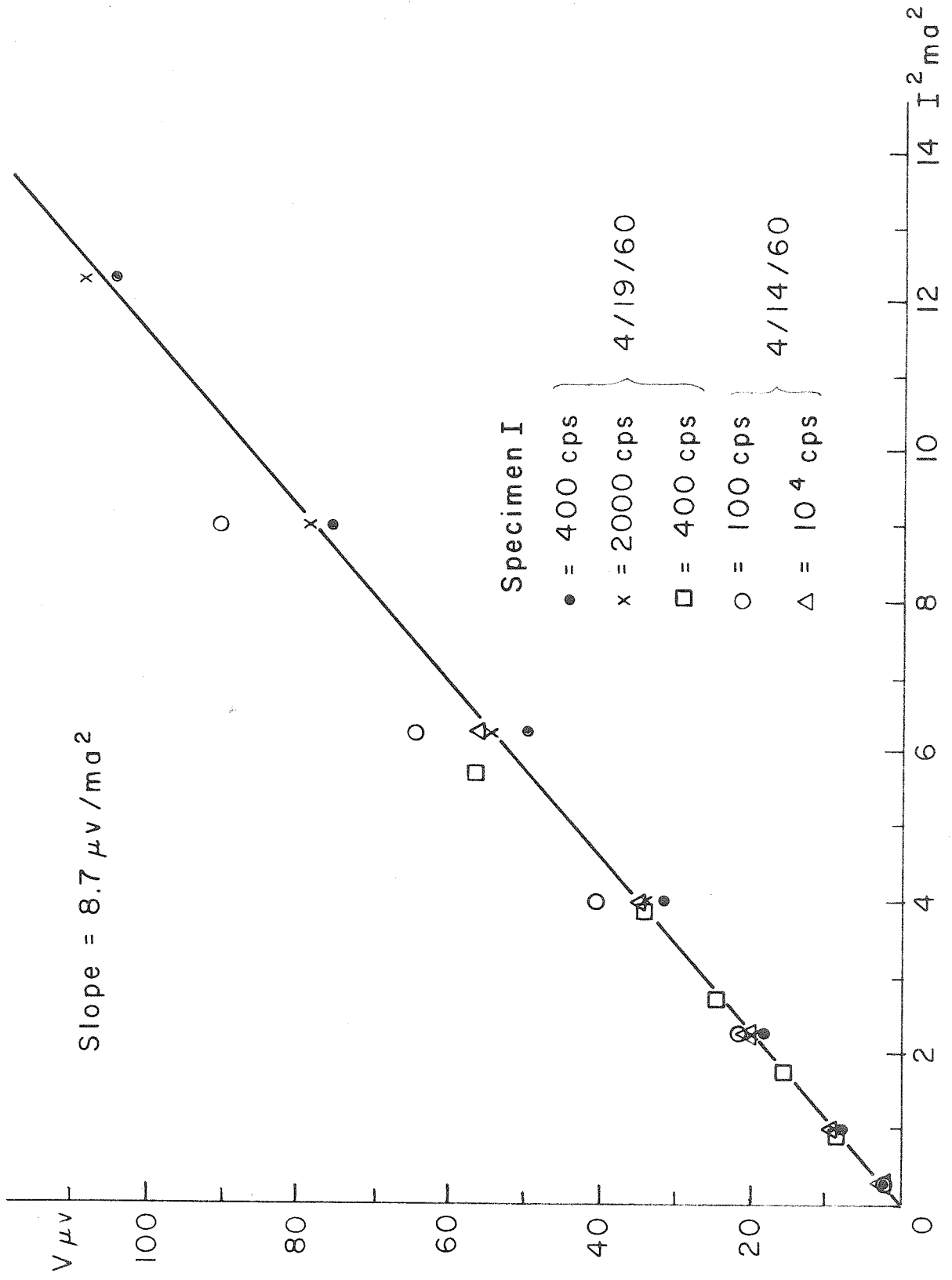


Fig. 6.- Configurational signal vs. the square of the specimen current in specimen I.

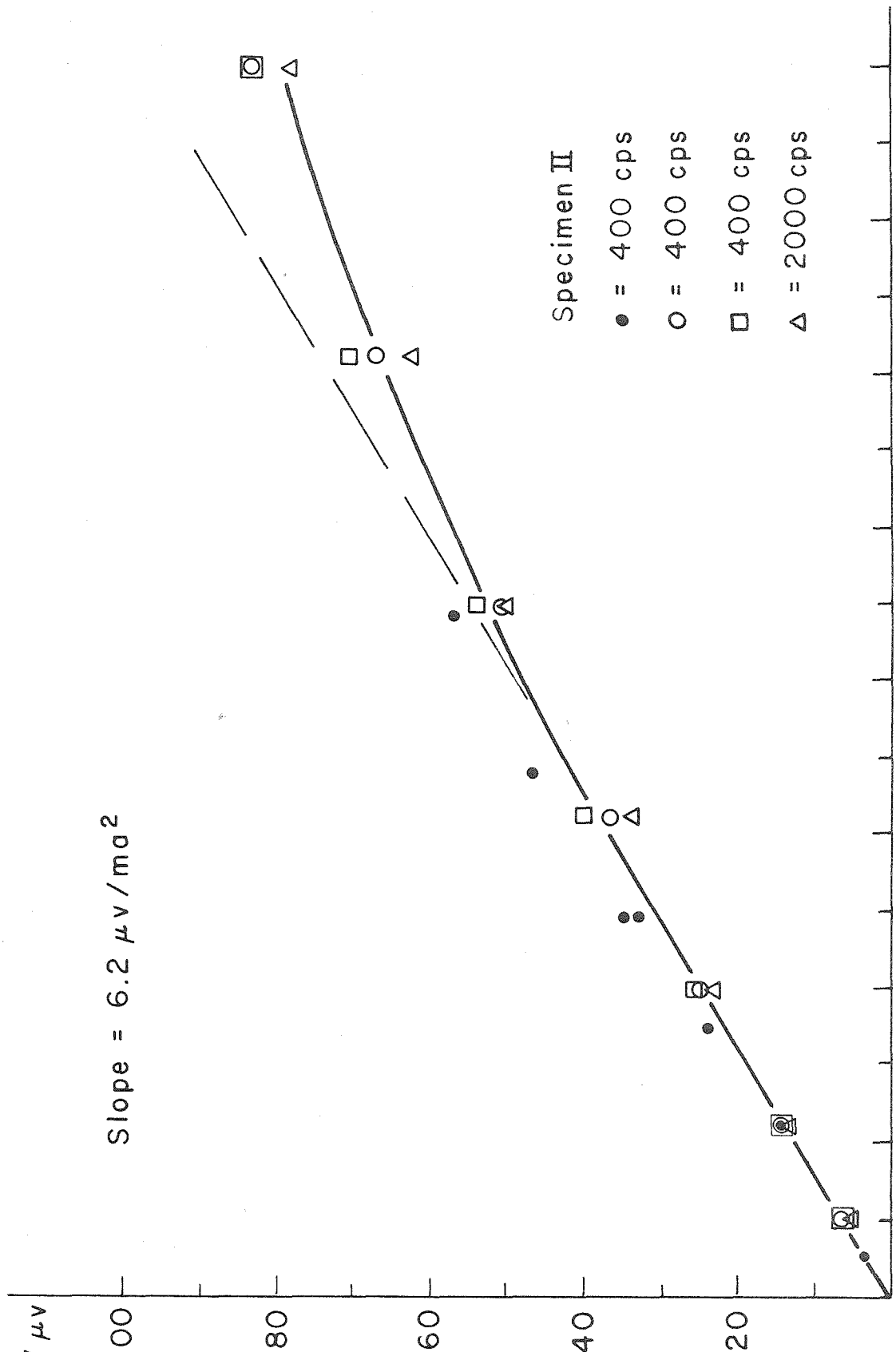
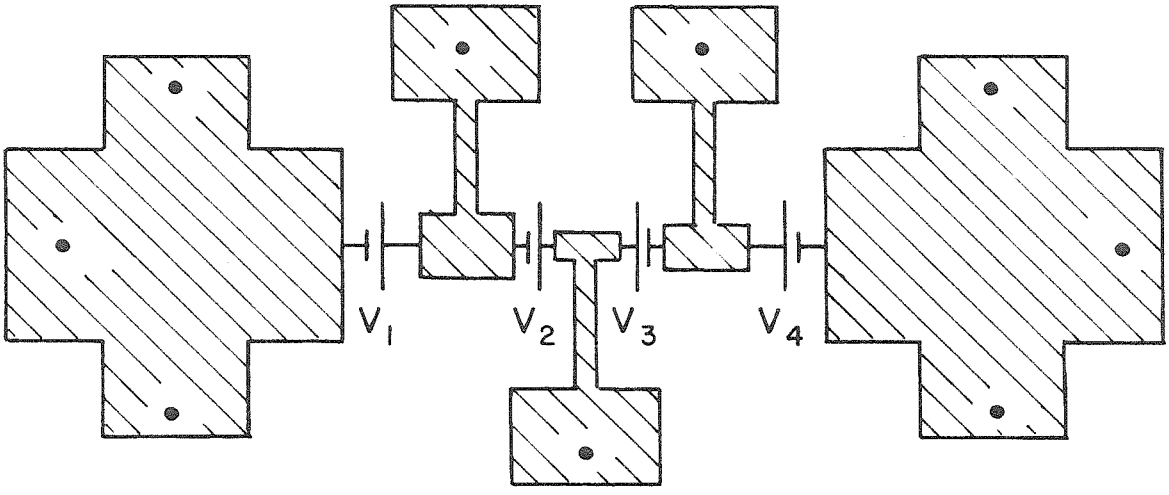
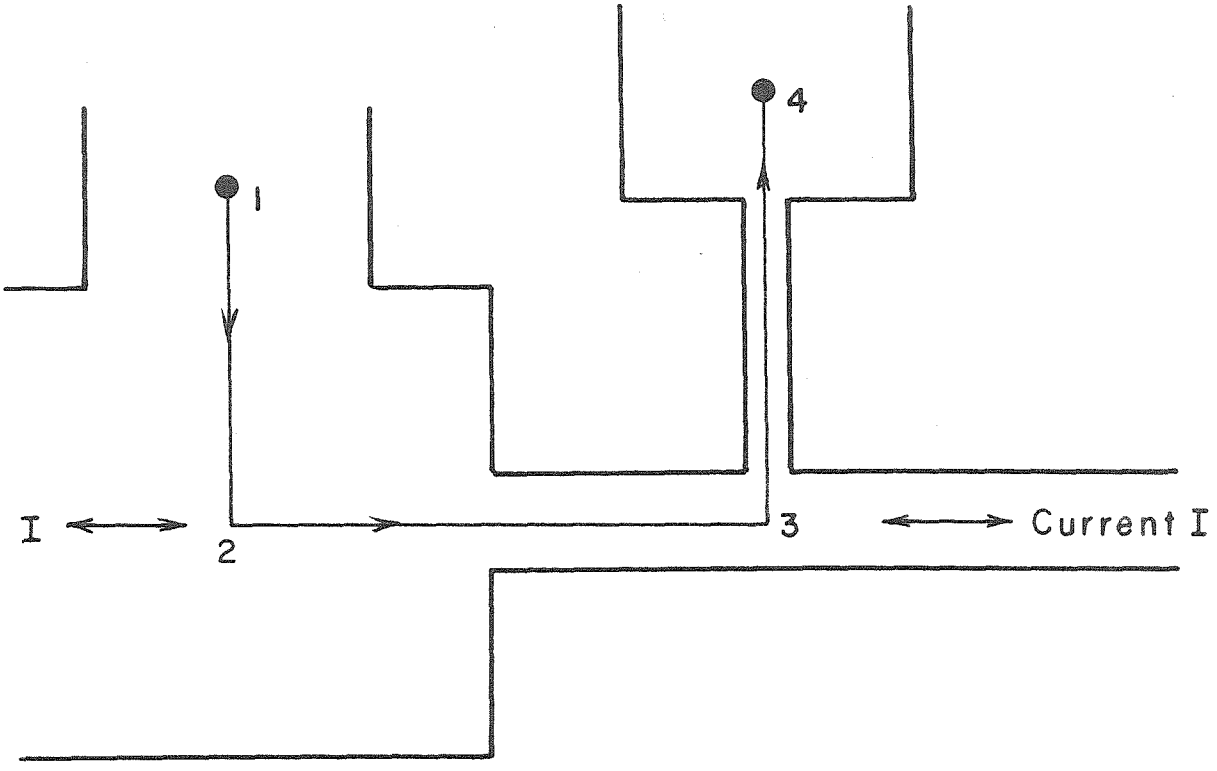


Fig. 7.- Configurational signal vs. the square of the specimen current in specimen II.



$$V_1 + V_2 = V_3 + V_4$$

Fig. 8. - Lumped emf schematic of the configurational potentials in a specimen of non-uniform cross section.



$$V_C = V_4 - V_1 = -2C\int_1^4 d\vec{r}\cdot(\vec{j}\cdot\nabla\vec{j}) \simeq -\int_2^3 d\vec{r}\cdot\nabla(Cj^2)$$

Fig. 9.- Integration path to obtain the configurational emf from the configurational field.

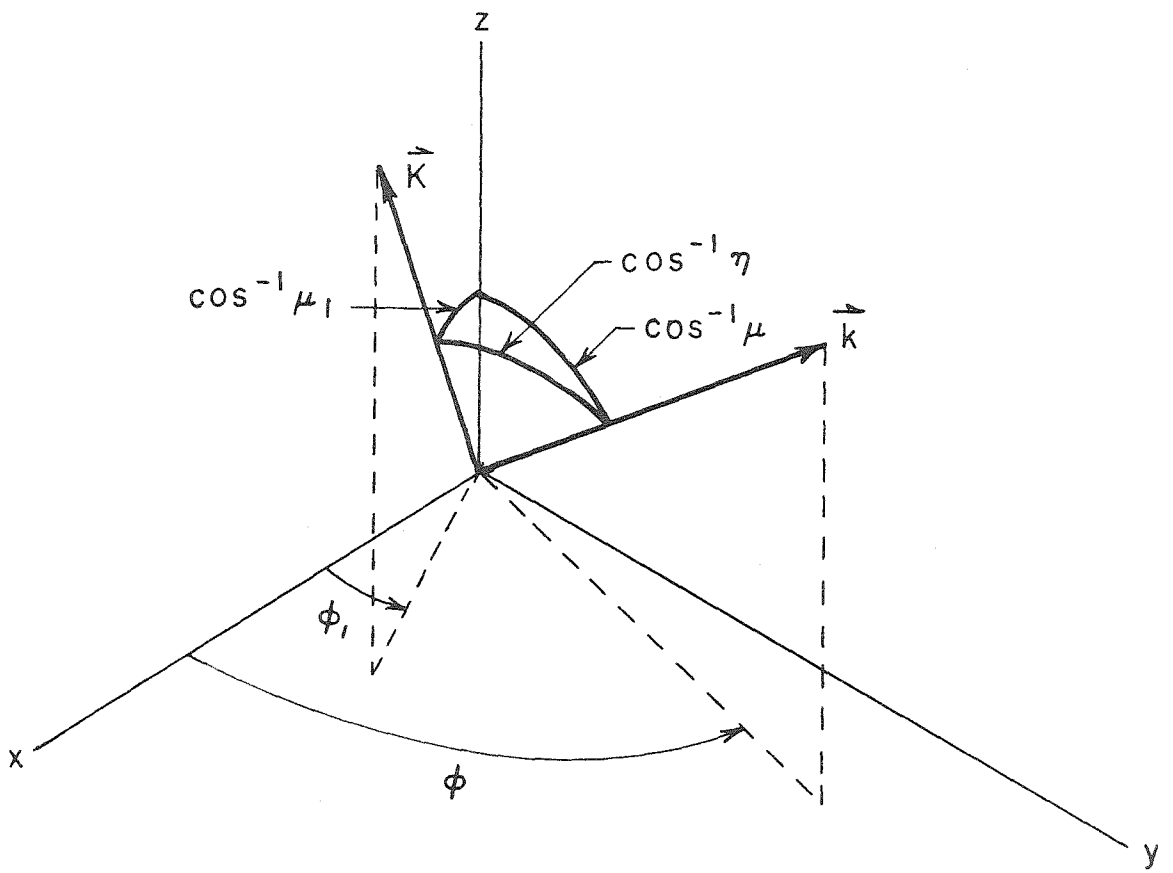


Fig. 10. - Coordinate definitions for the k-vectors.

References

1. P. Drude, Ann. Physik, 1, 566 (1900).

2. N. F. Mott and H. Jones, The Theory of the Properties of Metals and Alloys, Dover, New York, 1958.

3. F. E. Terman, Radio Engineer's Handbook, McGraw-Hill, N. Y., 1943, p. 904.

4. F. Seitz, The Modern Theory of Solids, McGraw-Hill, New York, 1940, p. 168.

5. A. H. Wilson, The Theory of Metals, Cambridge Press, Cambridge, 1953, p. 193.

6. R. E. Peierls, Quantum Theory of Solids, Clarendon Press, Oxford, 1956, p. 115.

7. A. H. Wilson, loc. cit., Sections 9.35, 9.351, 9.36, 9.361.

8. A. H. Wilson, loc. cit., Sections 2.7, 2.71, 2.711.

9. N. F. Mott and H. Jones, loc. cit., Chapter II, Section 4.6; Chapter VI, Section 2.

10. A. H. Wilson, loc. cit., Section 1.7.

11. N. F. Mott and H. Jones, loc. cit., Chapter III, equation 21.

12. N. F. Mott and H. Jones, loc. cit., Chapter V, Section 2.4.

13. A. H. Wilson, loc. cit., Section 3.8.

14. N. F. Mott and H. Jones, loc. cit., Chapter VI, Section 6.

15. R. E. Peierls, loc. cit., Section 7.3.

16. Winifred Lewis, Thin Films and Surfaces, Chem. Pub. Co., N. Y., 1950.

17. H. Mayer, "Conduction Phenomena in Thin Films" in Structure and Properties of Thin Films, edited by C. A. Neugebauer, J. B. Newkirk and D. A. Vermilyea. John Wiley, New York, 1959.

18. G. A. Bassett, J. W. Mente and D. W. Pashley, "The Nucleation, Growth and Microstructure of Thin Films," in Structure and Properties of Thin Films (cf. reference 17).

19. A. Colombani and P. Huet, "Electromagnetic Properties of Thin Films of Bismuth" in Structure and Properties of Thin Films (cf. reference 17).