The Quantum Physics of Solids, I
The Energies of Electrons in Crystals

By W. SHOCKLEY

It is proposed to make this paper the first of a series of three dealing with the quantum physics of solids. This one will be concerned with the quantum states of electrons in crystals. The discussion will commence with an introductory section devoted to the failure of classical physics to account for phenomena of an atomic scale. Next, the quantum theory of electrons in atoms will be discussed, together with the resultant explanation of the structure of the periodic table; this is designed to illustrate the meaning of various quantum mechanical ideas which are important in understanding solids. Furthermore, much of the detailed information about atomic quantum states of particular atoms will be needed in the later discussion of the properties of certain solids. As an introduction to the modification of the quantum states occurring when atoms are put together to form a crystal, a short section will be devoted to structure of diatomic molecules. The next section will be concerned with quantum states for electrons in crystals. Whereas in an atom there are a series of isolated energies possible for an electron (corresponding to the various quantum states), in a crystal there are bands of allowed energies. This concept of energy bands is essential to the theory of crystals in much the same way that the concept of energy levels is essential to that of atoms. In terms of energy bands, the energy holding crystals together can be interpreted on a common basis for a wide variety of crystal types. This will be followed by a brief description of various crystal types and by a discussion of thermal properties in which the smallness of the electronic specific heat will be shown. The last section will be devoted to a discussion of para and ferromagnetism on the basis of the energy band picture.

In the second paper, problems connected with electric currents and the motion of electrons through crystals will be discussed. This leads to the concept of the Brillouin zone which is complementary to that of the energy band, the two together forming the basis for discussing the quantum states of electrons in crystals.

The third paper of the series will contain a comparison between theory and experiment for the alkali metals, the principal emphasis being placed upon the physical picture of the state of affairs in these simple metals.

INTRODUCTION

"THE parts of all homogeneal hard Bodies which fully touch one another, stick together very strongly . . . I . . . infer from their Cohesion, that their Particles attract one another by some Force, which in immediate Contact is exceeding strong, at small distances

performs the chymical Operations above mention'd, and reaches not far from the Particles with any sensible Effect. . . . There are therefore Agents in Nature able to make the Particles of Bodies stick together by very strong Attractions. And it is the Business of experimental Philosophy to find them out." But it was not destined for experimental philosophy to finish the business which Sir Isaac Newton set for it in the above words [1] until two centuries had elapsed. Only since the advent of quantum mechanics have scientists had laws capable of explaining the cohesive forces of solid bodies and predicting their numerical magnitudes. The new laws were developed first in order to explain the behaviors of independently acting atoms but, as we shall see, they are laws capable of extension to systems containing large numbers of atoms and thus to solid bodies. The fact that a solid body remains a solid body, resists being pulled apart, and exerts the cohesive forces of which Newton wrote, is explained by showing from theory that atoms packed together in a solid are in a state of low energy, and to change the state requires the expenditure of work. In this paper we shall describe how the quantum mechanical concepts developed for isolated atoms are applied to interacting atoms and lead to methods of calculating the energies and forces binding atoms together in crystals.

A crystal is a regular array of atoms. The regularity of this atomic array is frequently exhibited in the macroscopic appearance of the crystal. A crystal of potassium chloride—sylvine—is a good example (Fig. 1A). The natural growth faces of the crystal are parallel to planes passing through the atoms, which are arranged in the microscopic array pictured in Fig. 1B. It is evident that the microscopic arrangement of the atoms in the crystal is one of its most basic features. In sylvine the atoms are arranged on the corners of cubes in an alternating fashion. The arrangement of the atoms in the crystal is called a "lattice." Sodium chloride—rock salt—has the same arrangement as sylvine and the type lattice pictured in Fig. 1B is known as a "sodium chloride lattice." The distance between atoms in a given lattice is specified by giving the value of the "lattice constant," which for a cubic crystal is defined as the distance between like atoms along a line parallel to a cube edge. Lattice constants are usually expressed in angstroms; 1 angstrom $\equiv$ 1A $= 10^{-8}$ cm. The lattice constant of sylvine, designated by "a" in Fig. 1B, is 6.28A. Figure 1C suggests how a large number of atoms, arranged as in Fig. 1B, produce the shape of the crystal photographed for Fig. 1A. Studies of the directions of the natural growth faces and cleavage faces of crystals are
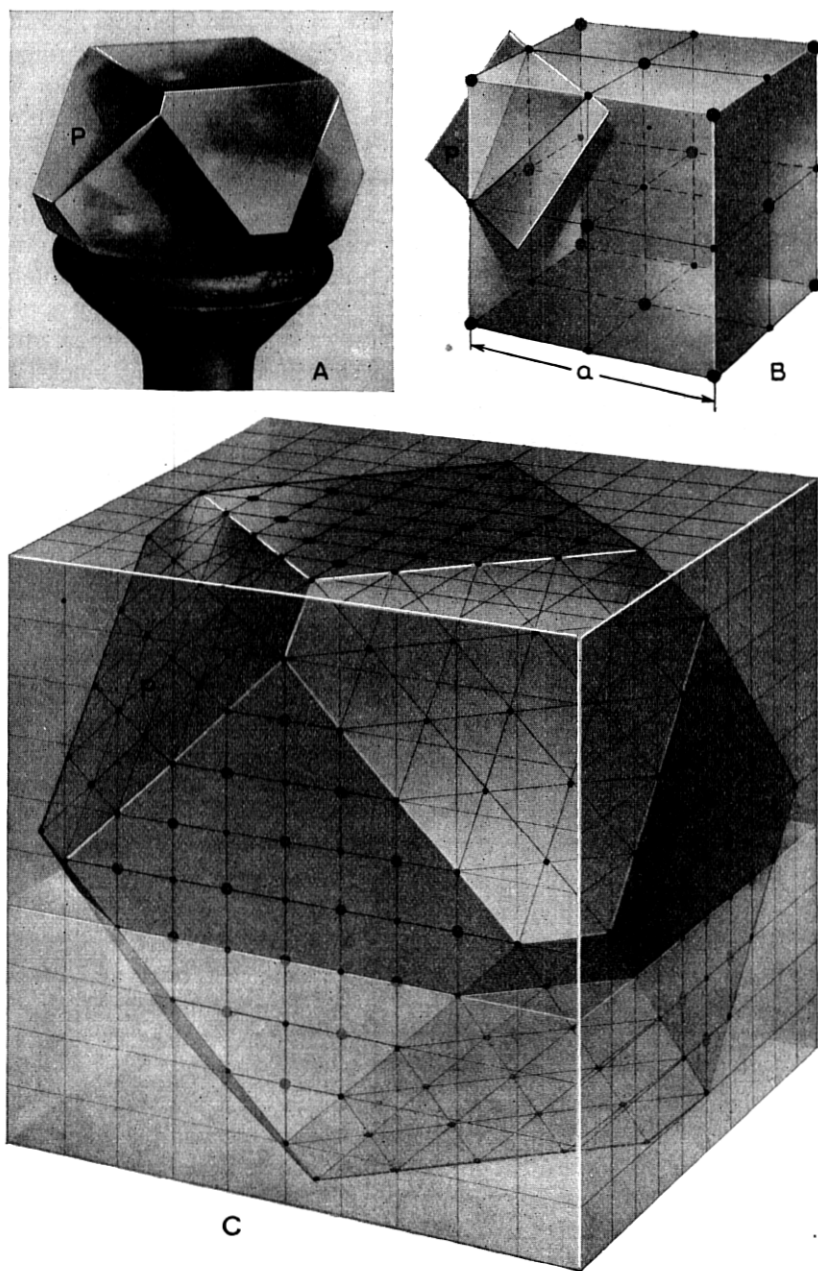
[1] "Opticks" 3rd ed., 1721, p. 363.

Fig. 1—Crystal structure.

A. Macroscopic appearance of a crystal; retouched photograph of a sylvine crystal.
B. Microscopic arrangement of atoms in crystal showing natural planes.
C. Large number of atoms arranged as in B to show formation of A.

primarily of importance as an aid in classifying and identifying minerals; and although they do give some information about the arrangement of the atoms in planes within the crystal, the information is too meager to permit a determination of the microscopic structure. The latter can be deduced by the methods of x-ray diffraction. X-rays are light waves of very short wave-length and they are diffracted from crystals in much the same way as light is diffracted from a ruled grating. From studies of x-ray diffraction patterns, the arrangements of atoms in a large number of crystals have been determined. Exceedingly strong forces act to hold the atoms in these arrangements and, by application of the laws of quantum mechanics, we shall try to find them out.

There is now no question that the elementary building blocks of the material world are primarily electrical and of two sorts.[2] The negative particles, electrons, are all alike and have the same charge $-e$ and the same mass m; the positive particles, atomic nuclei, are not all alike and may differ in charge and mass from one another. The positive charge is always some integral multiple, Z, of the fundamental charge e and we shall not be concerned with the mass except to say that it varies upwards from about 2,000 times the electron mass. An atom of a chemical element consists of one nucleus surrounded by enough electrons to neutralize its charge; all atoms of a given chemical element have the same nuclear charge, Z, which is appropriately known as the "atomic number"; atoms having the same nuclear charge but different masses are called "isotopes"; their chemical behaviors are slightly different, so that it is possible by chemical processes to separate one isotope of a chemical element from the others, but this difference is so slight that we can neglect it here. An atom, then, consists of a number of electrons circulating about and attracted by the nucleus, which, by virtue of its relatively great mass, is effectively an immobile center for their motions. A simple molecule consists of an assemblage of a few such atomic systems and a crystal of an immense number. The fundamental problem of atomic mechanics—which is now solved quite satisfactorily but not yet perfectly—is to find the laws governing the motion of these particles.

The necessity of finding such laws is made most apparent by considering the failure of the older laws of "classical mechanics," Newton's laws. These laws were satisfactory for dealing with large bodies—but not perfect; for, as is well known, they are approximations to the more adequate laws of relativity—and they were successfully applied

---

[2] Since we are here concerned with problems of a chemical nature, we may disregard those particles such as positrons, mesotrons, neutrons, etc., which are concerned with cosmic rays and nuclear processes but not with ordinary atomic behavior.

even to single atoms so long as no attempt was made to investigate the internal structure of the atom. Considering the atom to be a perfectly elastic miniature billiard ball having size, mass, and velocity but no internal properties, classical mechanics was able to handle in a statistical fashion the dynamics of large systems of atoms in a gaseous form and to deduce a number of valid conclusions concerning the specific heat, gas laws, viscosity, and diffusion constants of gases. On the other hand, failure attended all endeavors to apply these laws to the swarm of electrons surrounding a nucleus. A system of this sort is unstable classically and can never come to thermal equilibrium. Applying the classical laws of statistical mechanics, one finds that some of the electrons will move very close to the nucleus, the energy lost in this process being acquired by other electrons which move farther out. According to classical mechanics this process will continue without ever reaching equilibrium and during it the atom will be thoroughly torn apart.

Another difficulty in the classical theory arises from the electrodynamics of an accelerated electron. An electron moving in the field of a nucleus is accelerated, and classical electromagnetic theory predicts that under these circumstances electromagnetic energy will be radiated—the atom being in effect a microscopic radio transmitting station in which the charging currents in the antenna are represented by the motions of the electrons. According to this theory an atomic system would continually radiate energy, and it could be proved that no equilibrium like that actually observed between matter and radiation would ever be achieved.

Thus, classical mechanics and electromagnetics were incapable of taking the electrons and nuclei as building blocks and constructing solids or even atoms from them. To put it bluntly, the classical laws were wrong; although adequate for large-scale phenomena, they were inapplicable to phenomena of an atomic scale.

Nevertheless, modified applications of the classical theory had a great number of successes in the atomic theory of solids. Dealing with the atoms as elastic idealized billiard balls led to the correct value for the specific heat of solids, at least at normal temperatures, and the electron theory of conduction in metals was in many respects quite successful. None of the successes of the conduction theory were completely satisfying, however, because the assumptions needed to explain one set of facts were incompatible with other sets of facts and the whole field was greatly lacking in unity. According to this classical theory a metal contained free electrons which could move under the influence of an electric field and thus conduct a current. Their motion was

impeded by collision with the atoms (ions, really, since they are atoms which have given up free electrons) according to some theories, and with the spaces between atoms according to other theories, and this impeding process gave rise to electrical resistance. The free electrons were capable of conducting thermal as well as electrical currents. Although the theory gave reasonable values for the electrical and thermal conductivities of metals at room temperature, the predicted dependence upon temperature was wrong: the resistance of a pure metal is known from experiment to be very nearly proportional to the absolute temperature; the classical theory, unless aided by very unnatural assumptions, predicted proportionality to the square root of the absolute temperature. Another difficulty, the greatest in fact which beset the old theory of free electrons in metals, was concerned with the specific heats of metals. According to the billiard ball theory of gases, the specific heat arose from the kinetic energy of motion of the gas atoms; thus the specific heat at constant volume of one gram atom of a monatomic gas was $(3/2)R$, where R is the gas constant. This was in good agreement with experiment. For solids this specific heat was just doubled, giving $(6/2)R$ because of the addition of potential energy to the kinetic. For a metal the free electrons were regarded as having kinetic energy. In order to explain the observed electrical properties of a metal, the number of electrons was taken as approximately equal to the number of atoms. Hence, as for a monatomic gas, a specific heat of $(3/2)R$ was expected for the electron gas and, therefore, a specific heat of $(9/2)R$ was predicted for a metal. Measurement shows that most crystals, metals included, fit quite well the value of $(6/2)R$ and that $(9/2)R$ is incorrect. Thus classical theory was left with the dilemma that to explain electrical properties one free electron per atom was needed while to explain specific heat one free electron per atom was far too many. This dilemma is very neatly resolved in the new theory; in this paper we shall show why the free electrons are not free for specific heat and in a later paper why they are free for conduction. We shall also show that the new theory leads to quite proper values for the conductivity and also explains facts concerning the resistance of alloys, which the classical theory could not do.

According to the classical theory there was one quantity that should be the same for all metals and this was the ratio of the thermal to the electrical conductivities. This ratio, known as the Wiedemann-Franz ratio, was predicted to be equal to the absolute temperature times a universal constant L called the Lorentz number. This prediction was in reasonable agreement with experiment. The new wave me-

chanical theory predicts the same result, but with a slightly different value for L. According to the old theory $L = 2k^2/e^2 = 1.44 \times 10^{-8}$ volts^2/degree^2 where k is Boltzmann's constant, while the new gives $L = \pi^2 k^2/3e^2 = 2.45 \times 10^{-8}$ volts^2/degree^2, and the experimental values for several elements are Cu 2.23, Ag 2.31, Au 2.35, Mo 2.61, W 3.04, Fe 2.47—all times 10^{-8} volts^2/degree^2. We see that the constancy of the Lorentz number predicted by both theories is in reasonable agreement with experiment, but that in predicting the numerical value of the constant the new theory is better than the old.

The fundamental problem of how the electrons and nuclei form stable atoms and crystals was, as we have said above, inexplicable on the older theory. The newer quantum mechanics of Bohr and later that of Schroedinger, Heisenberg, and Dirac were needed. Bohr postulated that out of the infinity of possible motions for the electrons of an atom, only a certain restricted set was permitted. Each permitted motion corresponded to a definite energy for the atomic system as a whole. This concept of energy levels for the atom gave a natural interpretation to nature of atomic spectra and explained the meaning of the combination principle. In order to restrict the atomic motions to certain energy levels, Bohr supposed that the laws of atomic dynamics were such that only those modes of motion were permitted for which certain dynamical quantities, called phase integrals, had values equal to multiples of Planck's constant h. For the case of the hydrogen atom these laws led to the now well-known Bohr orbits for the electron and to energy levels which were in good agreement with experiment. For atoms with more electrons it was very difficult to apply Bohr's laws except in a very approximate and unsatisfactory way. However, two very valuable concepts came from his theory which are preserved in the newer wave mechanical theory. These were that the individual electrons could be thought of as restricted to certain orbits and that these orbits were specified by giving them certain quantum numbers. It was found that three quantum numbers were needed to specify the orbit. All atoms were found to have the same general scheme of orbits. The number of electrons moving in these orbits varies from atom to atom and for any given atom is equal to the atomic number Z. In order to explain the facts of spectroscopy and the periodic table of the elements, it was necessary to introduce a rule known as Pauli's principle. This principle states simply that no more than two electrons may occupy the same orbit in an atom; that is, no more than two electrons of an atom may have the same three quantum numbers. As we shall discuss in the next section, a complete specification of the state of an electron in an atom requires four quantum numbers; two

electrons in the same orbit have different values for their fourth quantum number. We shall use the term "quantum state" to signify the permitted behavior corresponding to specified values for the four quantum numbers. In this language, Pauli's principle asserts simply that no two electrons in a given atom can be simultaneously in the same quantum state; that is, Pauli's principle is a quantum mechanical analogue of the classical principle that two bodies cannot occupy the same place at the same time. The two ideas—first that the motions of the electrons are quantized so that only certain quantum states are allowed, and second that in an atom only one electron can occupy a given quantum state—form the basis of all quantum mechanical thinking. We shall make use of them continually in the following discussion. We shall use them, however, not in connection with the orbits of Bohr but instead with the wave functions of Schroedinger.

The Bohr theory can be applied only with difficulty to any atom but hydrogen. The difficulty lies in determining the motions of the electrons in the complex interacting fields of the electrons and the nucleus. This problem is even more difficult in the case of a solid where there are many atoms, and it would seem hopeless to try to find out why the electronic orbits in insulating crystals such as rock salt or diamond do not permit electrons to move through the crystal and carry a current, while the orbits in metals do. Indeed not only does the Bohr theory have the foregoing disadvantage but it is probably wrong. Fortunately there is a theory both sounder and easier to apply embodied in the "wave equation of Schroedinger."

One feature, probably not sufficiently stressed, about Schroedinger's equation is its relative convenience. The word "relative" must be used here because it is usually very laborious to obtain solutions for the equation and only in the simplest cases can we obtain exact solutions. Compared to the classical equations and the equations of Bohr, however, it *is* convenient. Quite satisfactory approximate solutions can be obtained for Schroedinger's equation even for the complex case of solids, where it would be prohibitively difficult to obtain as good solutions for the classical and Bohr equations.

Electrons in Atoms

According to the Schroedinger theory, a differential equation can be written down for any system consisting of electrons and atomic nuclei. This equation contains an unknown wave function and an unknown energy and the instructions of the theory are to solve the equation for the unknown quantities. Furthermore, the wave function must satisfy a certain mathematical requirement which embodies

in a generalized form the restrictions imposed by Pauli's principle. As is too frequently the case in mathematical physics, it is much easier to state the problem than to solve it; the solutions of Schroedinger's equation are, in fact, so difficult to obtain that exact solutions have been found for atomic systems only of the simplest type, namely those consisting each of a single nucleus and a single electron. For this case, the quantum states and their energies are all exactly known. For other cases approximations of varying degrees of exactness must be used. The difficulty arises from the interactions between the electrons. If it were not for these interactions, one could obtain exact solutions for atoms having many electrons. The difficulty is that the interactions—they are merely electrostatic repulsions—prevent each electron from being independently in a definite quantum state. The interaction of each electron with another is in general small compared to its interaction with the nucleus. To a first approximation, then, the electrons are treated as not interacting and then corrections are applied to this over-simplified picture. (In this first approximation, the generalized mathematical statement of Pauli's principle reduces to the one we gave in the last section—only one electron may occupy a given quantum state.) As a result of this procedure of over-simplification followed by corrections, our exposition will commence with a discussion of the quantum states of an electron in an atom as if these quantum states were private possessions of the electron and not influenced or disturbed in any way by the other electrons. We shall then correct this picture to some extent by considering how the energy of a given electron depends upon the behavior of the other electrons. One correction term which we shall introduce in this way is the important "exchange energy" discussed below. Thus atomic theory represents a field of endeavor in which further progress is made largely by improvements and refinements. It should be emphasized, however, that the corrections and refinements are not additional assumptions, which are added to the theory, but that they represent instead only steps forward in improving the wave mechanical solutions.

The last paragraph mentions that an approximate treatment of Schroedinger's equation leads to a set of possible quantum states for an electron in the atom. We shall discuss Schroedinger's equation and the wave functions corresponding to the quantum states in more detail later and at present be concerned only with a description of the results. In a neutral atom the electrons arrange themselves in the quantum states in such a way as to make the energy of the atomic system a minimum. Consistent always with Pauli's

principle, only one electron can occupy a particular quantum state. When the atom is in the arrangement of lowest energy, we can say that each electron has a definite energy corresponding to whichever quantum state it occupies. This energy is most conveniently defined in terms of the amount of work required to take an electron from its state in the atom and put it in a standard state defined as zero energy. An electron in the zero energy state is to be thought of as at rest and so far removed from the atom that there is no energy of interaction between them. In this way we can define the energy of every occupied quantum state in the atom. Each of these energies must be taken as negative—since potential energy is yielded up when the electron returns to the atom—and by definition represents how tightly the electron is bound to the atom. One of the electrons will be the most loosely bound (it may be that there are several with the same energy) and the energy required to remove it is called the "ionization energy." From our definition this is obviously the minimum energy required to convert the atom to a positive ion. The definition of the energy of a quantum state given above can be used only when an electron is in the quantum state; we can, however, define the energy of an unoccupied state conveniently in terms of the energy the atom would have if the state were occupied by "exciting" one of the electrons to this state by giving it the proper amount of energy.[2a]

The Quantum States of the Atom

Using this definition of the energy of a quantum state, we find that for all atoms the arrangement of quantum states in energy is as shown in Fig. 2, where the ordinates represent energies and states of equal energy appear as divisions of the horizontal lines. Figure 2 does not indicate which states are normally occupied, nor could it unless we knew how many electrons there were in the atom. The general scheme of Fig. 2 is applicable, with certain changes discussed below in the energy scales, to any neutral atom in its normal state, and the energy

[2a] This definition is subject to restrictions because the energy of an electron in the state in question depends upon the arrangement of the other electrons in the atom and this arrangement depends in turn upon which electron was excited to the initially unoccupied state. In constructing the figures we have supposed that the electron (or one of the electrons in case there are several) that is most easily removed from the atom is caused to shift from its normal state to the unoccupied state in question; this shift will change the state of the atom and since the atom was initially supposed to be in the state of lowest energy, the change in energy cannot be negative and will in general be positive but may in certain special cases be zero. The energy of the unoccupied state is defined as the energy of the occupied state from which the electron is taken plus the change of energy caused by shifting the electron. This is equivalent to saying that the energy of the unoccupied state is the ionization potential of the atom after an electron has been shifted from the highest state normally occupied to the normally unoccupied state in question.

levels represented on Fig. 2 apply to unoccupied and occupied states as well.

I have already mentioned that four quantum numbers are required to specify each quantum state. These are indicated by the letters
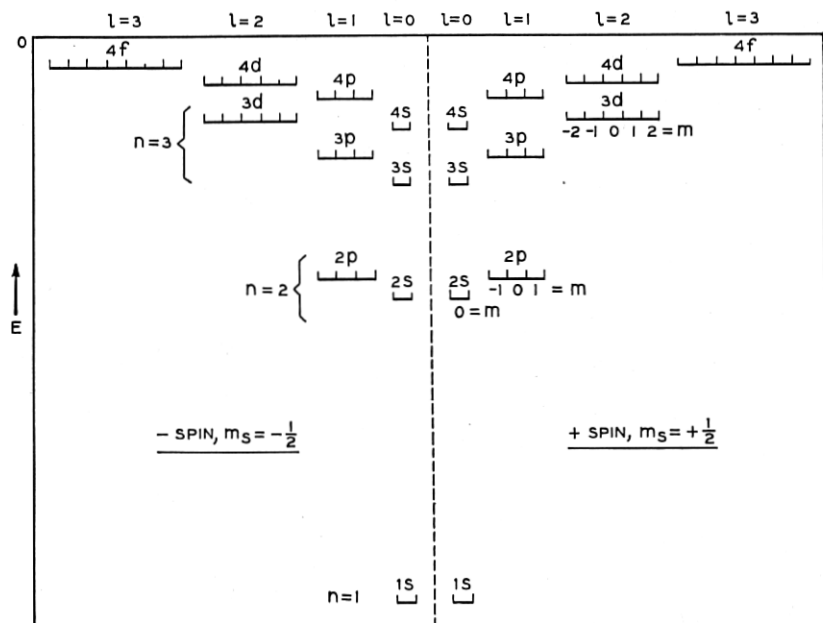

Fig. 2—Quantum states for electrons in atoms.

n, l, m, and m_s. Roughly speaking, the "principal quantum number n" fixes the "energy level" of the state; however, there is some dependence upon the "angular momentum quantum number l." The dependence of energy upon the third quantum number, "the magnetic quantum number m," is slight and will be neglected in this paper. We shall consider the energy to be specified by giving n and l. A notation borrowed from spectroscopy is applied to this pair of quantum numbers and one uses the apparently quite fortuitous choice of letters s, p, d, f, g, h, etc., to stand for $l = 0$, 1, 2, 3, 4, 5, etc., and a state with $n = 3$ and $l = 2$ is known as a "3d state" and an electron occupying such a state is called a "3d electron." The quantum laws permit the following values for n, l, and m:

n takes on all positive integral values. (All states with n greater than four have been omitted from the figure; they lie between the highest states shown and zero energy.)

For a given n, l takes on all positive integer values from 0 to $n - 1$ inclusive.

For a given n and l, m takes on all integer values including zero from $-l$ to $+l$ inclusive.

The difference between right and left sides of the figure corresponds to the fourth quantum number: an electron, in addition to its electric charge, possesses angular momentum or "spin" about its axis. The rotating charge resulting from this angular momentum produces a magnetic moment. The angular momentum is quantized and there are two possible values $+1/2$ and $-1/2$ for the "spin quantum number m_s," corresponding to the right and left halves of Fig. 2. Electrons occupying states on the right half of Fig. 2 have their spins parallel to each other and directly opposite to electrons occupying states on the left half. As already implied, the quantum numbers l and m also correspond to angular momentum and magnetic moments for the electron "orbits" (really wave functions) in the atom.[3]

For our purpose we need two results of the theory of the spinning electron, first that its *spin introduces a duplicity of quantum states* as indicated by the two halves of Fig. 2, and second that *all the electrons of one spin have their magnetic moments parallel and opposite to those of the other spin*. Later when we consider the question of magnetism, we shall be concerned with the direction in space of the spin vector and the magnetic moment, but not now.

Several units of energy are employed in describing atomic processes. The simplest of these is the electron volt; it is the energy acquired or lost by an electron in traversing a potential difference of one volt. For example, in a vacuum tube operating with one hundred volts between cathode and plate, the electrons strike the plate with a kinetic energy of one hundred electron volts, 100 ev. Another unit is the ionization potential of hydrogen, and as hydrogen has only one electron, which normally occupies the $1s$ state, this is also the energy of the $1s$ state. This energy is called the "atomic unit" of energy or the "Rydberg." Another unit of energy useful in chemical processes is the kilogram calorie per gram atom; this is related to the others as follows: if the energy of each atom in one gram atom is increased by one electron volt then the energy of the whole system is increased by 23.05 kilogram calories. The conversion factors are: 1 Ry = 13.5 ev, 1 ev per atom = 23.05 Kg.-cal./gm. atom.

----

[3] For a discussion of the quantum states of the electrons from the point of view of angular momentum see "Spinning Atoms and Spinning Electrons" by K. K. Darrow, *Bell System Technical Journal*, XVI, p. 319, or standard texts on spectroscopy.

Variation of the Energy Levels with Atomic Number

All atoms have the same general scheme of quantum states indicated in Fig. 2. Quantitatively the energy scale varies from atom to atom. Thus the 1s state lies at − 13.5 ev for hydrogen and at − 24 ev for helium. This decrease (i.e., becoming more negative or moving lower down on Fig. 2) is due to the increase in nuclear charge, $Z = 1$ for hydrogen and 2 for helium, which results in greater attraction and tighter binding for electrons in helium. This steady downward motion of the levels continues as one goes from element to element in the periodic table. However, the ionization potential, the energy required to remove the most easily removed electron, does not steadily increase. In Fig. 3 we show the ionization potentials of the first twenty elements.
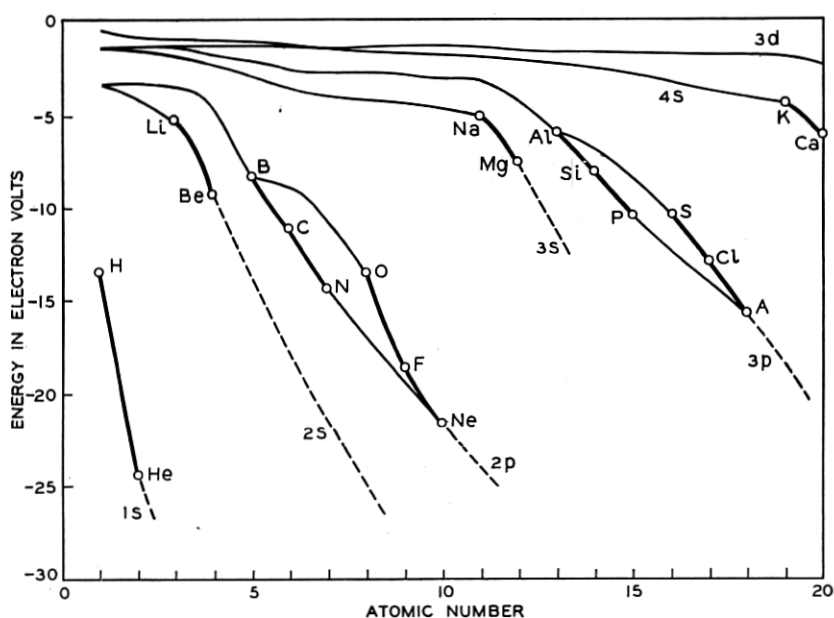

Fig. 3—Ionization potential versus atomic number.

Since we are interested in the energies of electrons rather than in ionization per se, the ionization potentials have been plotted as negative giving in this way the energies of the states in the atom. The main features of this figure can be explained by using Fig. 2 and the Pauli principle.

The Pauli principle, also known as the exclusion principle, permits only one electron to occupy each of the states of Fig. 2. The electrons in a many-electron atom will tend to go to the states of lowest energy.

Thus in helium, since its two electrons can have oppositely directed spins, each fills one of the $1s$ states; we say the "electron configuration" of helium is $1s^2$ (read as "one ess squared"). For lithium, $Z = 3$, the third electron, which cannot go to the completely filled $1s$ states, goes to the next highest, $2s$, giving $1s^2\,2s$. In going from helium to lithium, all the states move to lower energies but not so much lower as to make $2s$ for lithium as low as $1s$ for helium. For this reason lithium can be relatively easily ionized, as is seen in Fig. 3.

Before continuing the discussion of particular atoms, we must point out that two changes accompany each advance from one element to the next in the periodic table. In each step the nuclear charge increases by one plus unit and at the same time an electron is added to the atom and the combined effects produce the results of Fig. 2. Quite different results are obtained if one electron alone is added to the atom. Then instead of the general falling of the levels which accompanies the double change, there is a general rising of all the levels. This is due to the unbalanced negative charge on the added electron, whose presence on the atom raises the potential energy of all the electrons and therefore raises their energy levels. For some atoms, the raising of the energy levels produced by an unbalanced electron may be so great that the electron is not bound at all or at least only very slightly, and for these atoms negative ions do not form. On the other hand, when an electron is removed from an atom all the remaining electrons become more tightly bound and the energy levels are lowered.

Exchange Energy

In Fig. 4 we show the electron configurations for the elements from lithium to neon. The decrease in ionization potential in going from beryllium to boron is due to the completed filling of the $2s$ states and the consequent start of filling of the $2p$ states. The decrease in going from nitrogen to oxygen suggests that not only do the $2s$ and $2p$ states lie at different levels but that the $2p$ states themselves lie at two different levels. This is true but in a rather special sense: *the difference in energy between the two sets of $2p$ states depends upon how they are occupied.* This difference is an "exchange energy." We shall discuss the origin of the exchange effect in the next paragraph but one; however, the aspect of it needed for this paper is illustrated in Fig. 4. We there imagine that the quantum states are represented by little trays upon which are placed weights to represent occupancy by electrons. The exchange effect corresponds to hanging the trays on springs; in this way we see that as the electrons fill up the $2p$ states

with one spin (the same effect occurs for either spin; the figure shows + spin), these states are depressed in respect to the $2p$ states with the other spin. The springs must, however, be considered to pull the
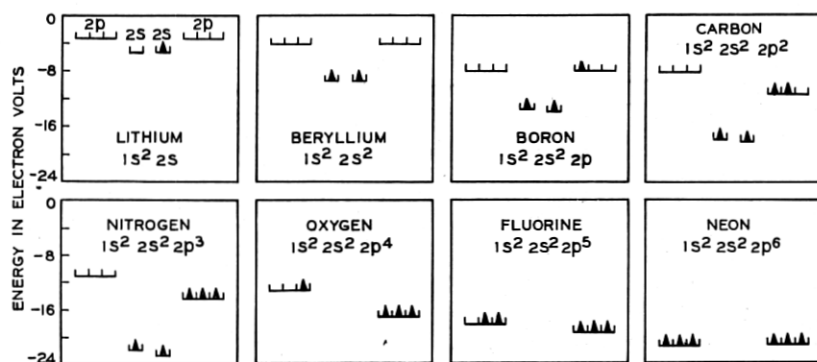

Fig. 4—Electron configurations illustrating the exchange effect.

trays up against stops with a force such that a single weight upon a tray will produce no lowering whereas two or three weights will. This effect seems contradictory to the simple idea that adding electrons raises the potential energy and the energy levels; however, it must be remembered that we are here discussing neutral atoms and that with each added electron there is also an added plus charge on the nucleus. These two charges produce the dominant variation in the energy levels and upon this variation the exchange effect is superimposed.

The reader may verify that so far as the distribution of electrons in the $2p$ states is concerned, the exchange effect will lead to the configurations shown in Fig. 4 for the states of lowest energy for the atoms. Let us consider carbon for example; if the electrons have opposite spins—that is, if there is one weight on each $2p$ tray—there will be no lowering due to the exchange effect; if the electrons have the same spin, however, then each loses energy because of exchange and the energy of the atom is less than for the case of parallel spins. The fact that one electron is not enough and that two or more electrons are required to produce the exchange effect is a natural consequence of the origin of the exchange energy.

The exchange energy is due to the electrostatic repulsion between the electrons and results directly from the application of Pauli's principle to Schroedinger's equation. The exchange effect emerges in a quite straightforward fashion from a consideration of wave functions, but usually no attempt is made to explain it in non-mathematical

terms. It seems to the writer, however, that the explanation given below does contain the mathematical essence in physical language.[3a] Pauli's principle, we have said, is the quantum mechanical analogue for electrons of the classical law that two bodies may not occupy the same place at the same time; it is, however, more general in the sense that it does not apply alone to location but rather to a combination of location and velocity and spin, and it requires that any two electrons differ essentially in one or more of these. Now a difference in the values for the spin quantum numbers of two electrons is a sufficiently great difference to permit them to have the same velocity and the same location (i.e., be very near together compared to atomic dimensions). If the spin quantum numbers are the same, however, there must be a difference in location or in velocity. Now two electrons having the same values of n and l, as for example two $2p$ electrons, move in similar orbits and have much the same velocities; hence, if their spins are the same they must differ in location—that is, they will satisfy Pauli's principle by keeping away from each other. If, however, their spins are different, then they need not keep away from each other, and in their motion about the nucleus they are, on the average, closer together than for the case of the same spin. Since the energy of repulsion between the two electrons decreases as they move farther apart, the average energy of the electrons is less for the case of parallel spins, for which Pauli's principle requires most difference in location; and this is just the effect shown in Fig. 4. Furthermore, if the electrons differ in their values of n and l, then their velocities are quite different and the restriction upon location is not so important and their electrostatic energy of repulsion for parallel spins is nearly the same as for opposite spins. There is, however, a small exchange effect between electrons of different n and l values as may be appreciated in Fig. 4 for boron, for example, by noting that one $2s$ level i depressed compared to the other owing to the presence of the $2p$ electron.

We see that helium and neon correspond to electron configurations which fill all the levels below $n = 2$ and $n = 3$ respectively. One sometimes refers to the states with $n = 1$ as the K shell, and to those with $n = 2, 3, 4$, etc. as L, M, N, etc., shells. The rare gases helium and neon then correspond to electron configurations consisting of "closed shells"—that is, to shells all of whose states are occupied.

[3a] As the aspects of exchange energy needed for the exposition are those discussed above in connection with Fig. 4, this explanation is not essential to the later argument of this article and is given in the hope that it may invest the concept of exchange energy with the appearance of a little more physical reality. If it fails in this, the reader is requested to disregard it.

The Periodic Table

The elements from lithium to neon constitute the first short period of the periodic table, Fig. 5. The second short period, running from sodium to argon, is built up in a similar way by filling the $3s$ and $3p$

H He
1   2 $1s$

Li Be B C N O F Ne
3   4   5 6 7 8 9 10 $2s, 2p$

Na Mg Al Si P S Cl A
11 12 13 14 15 16 17 18 $3s, 3p$

K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
s	s^2	$s^2 d$	$s^2 d^2$	$s^2 d^3$	$s d^5$	$s^2 d^5$	$s^2 d^6$	$s^2 d^7$	$s^2 d^8$	$s(d^{10})$	$s^2(d^{10})$	$s^2 p$	$s^2 p^2$	$s^2 p^3$	$s^2 p^4$	$s^2 p^5$	$s^2 p^6$

$4s, 3d, 4p$

Fig. 5—First part of the periodic table.

levels. Some of the chemical properties of the elements of these periods are quite easily understood in terms of the electron configurations of the atoms. An atom of lithium or of sodium has one easily removed electron and thus can become a positively charged ion; only one electron, however, can be so easily removed, and for this reason sodium and lithium do not have doubly charged positive ions in chemistry and are, therefore, monovalent positive elements. Similarly beryllium and magnesium have two easily removable electrons, and are divalent; however, their electrons are harder to remove than those of the alkali metals; hence the alkaline earth metals, beryllium and magnesium, are not so electropositive as the alkalis.[4] The halogens, fluorine and chlorine, present a contrasting picture. Instead of having one loosely bound electron, they have one low-lying empty state. They can therefore hold tightly an extra electron, and thus be negative ions. The rare or noble gas elements helium, neon and argon consist only of closed shells. They can neither gain nor lose electrons in chemical compounds and are, therefore, generally aloof to chemical urges.

[4] For brevity we shall refer to the alkali metals and alkaline earth metals simply as alkalis and alkaline earths.

The Transition Elements

Actually argon does not correspond to a complete system of closed shells, since for it none of the $3d$ states in the M shell is occupied. For the elements below copper, $Z = 29$, these $3d$ levels lie above the $4s$ and below the $4p$. They are filled up progressively in the series of elements scandium, titanium, vanadium, chromium, manganese, iron, cobalt, and nickel, which are known as the transition elements of the first long period of the periodic table. The first two elements after argon are potassium (an alkali) and calcium (an alkaline earth); these are similar to sodium and magnesium in having respectively one and two s electrons. The first transition element, scandium, however, is not like beryllium or aluminum, for with it the filling of the $3d$ states begins. The electron configurations for several of the transition elements are shown in Fig. 6. An interesting case occurs at chromium;
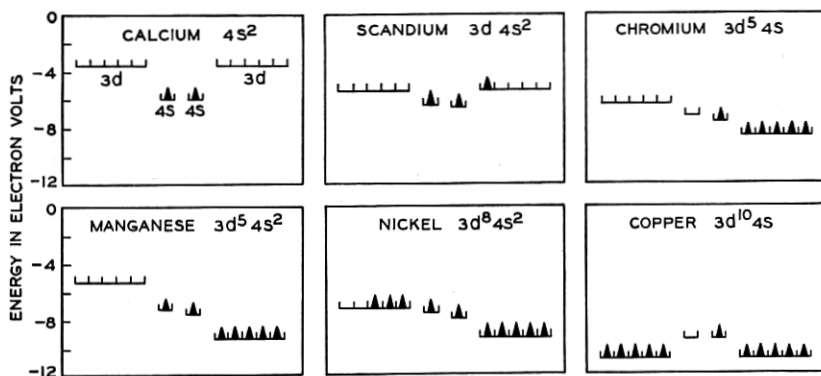

Fig. 6—Electron configurations for transition elements.

for it the exchange effect is so great that the $3d$ levels drop below the $4s$ and one $4s$ electron is transferred. Since there is an exchange effect between all electrons of the same spin, the remaining occupied $4s$ state in chromium has the same spin as the occupied $3d$ states. A similar transfer of a $4s$ electron occurs at copper, which is then left with one $4s$ electron and tends therefore to be monovalent (in the divalent copper ion the $4s$ electron and one $3d$ electron are removed). These transition elements are of particular interest because three of them, iron, nickel, and cobalt, are ferromagnetic in the solid. The atoms themselves are magnetic, as may readily be seen for chromium, for example; in it all the electrons have their spins parallel and hence their magnetic moments add to give a free chromium atom a magnetic

moment six times as large as the spin magnetic moment of the electron.[5] The same exchange effect which causes the $3d$ quantum states to fill unevenly in the isolated atom causes, in the case of the metal, an uneven filling of the "energy bands" which arise from these $3d$ states. We shall return to this topic in the section on ferromagnetism.

Solving Schroedinger's Equation

The possible quantum states of an atom are obtained by solving Schroedinger's equation for an electron moving in the potential field of the nucleus and the other electrons. In Fig. 7a we have represented the potential energy of an electron in an atom. If this potential energy (call it U) is known as a function of the position x, y, z, of the electron, then the Schroedinger equation is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{8\pi^2 m}{h^2}(E - U)\psi = 0, \tag{1}$$

where m is the mass of the electron, h is Planck's constant and E is an unknown energy and ψ an unknown wave function, for which a physical interpretation will shortly be given. It is found that this equation possesses proper solutions only for certain values of E; once these values are known, the equation can be solved for the unknown wave functions. The fact that only certain values of E are possible will probably seem more natural after reading the discussion given below of a mechanical system. The permitted energies and wave functions give the system of quantum states of Fig. 2.

The wave equation of Schroedinger is similar in form to many of the other wave equations of mathematical physics. In Fig. 7g to 7j we represent a stretched membrane like a rectangular drumhead. If the mass per unit area of the membrane is σ and the surface tension is T, then the wave equation for it is

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial z^2} + \frac{4\pi^2 f^2 \sigma}{T}\varphi = 0, \tag{2}$$

where f is the unknown frequency of vibration and φ is the unknown vertical displacement. Applied to the membrane, this equation has solutions only for certain values of f; the standing wave patterns corresponding to the four lowest frequencies are shown in Figs. 7g to 7j.

[5] For transition elements other than chromium, the motions of the electrons in their wave functions produce magnetic moments that must be considered as well as the spin; for a discussion of this point the reader is again referred to "Spinning Atoms and Spinning Electrons" by K. K. Darrow, *Bell System Technical Journal*, XVI, p. 319 and to texts on atomic physics.

The type of vibration of the system in one of these patterns is called "a normal mode." The patterns are described by two "quantum numbers" p and q which are equal to one plus the number of nodal lines (indicated by arrows) running across the membrane from front to back and from right to left respectively. In Figs. 7b and 7c we show the wave patterns corresponding to the 1s and 2s states of the atom; the quantum numbers of the ψ waves are also correlated to nodes. In the case of the membrane the frequencies and standing
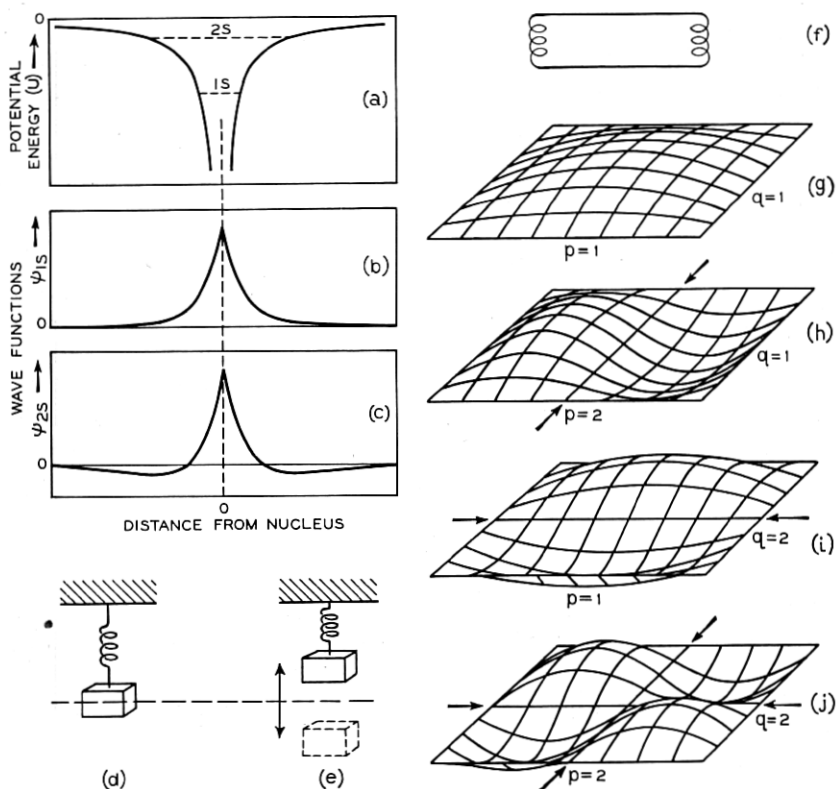
Fig. 7—The atom and some mechanical and electrical analogues.

(a) Potential energy of an electron in the atom.
(b) The 1s wave function.
(c) The 2s wave function.
(d) A mechanical analogue and
(e) its normal mode of vibration.
(f) An electrical analogue.
(g) to (j) The first four normal modes of vibration of a stretched drum head. (From "Vibration and Sound" by P. M. Morse, McGraw-Hill, New York, 1937. Courtesy of the McGraw-Hill Book Co.)

wave patterns are determined not only by the values of mass per unit area, σ, and tension, T, of the membrane but also by the "boundary condition" that it be clamped on its rectangular edge; at this edge the vertical displacement φ must vanish. The corresponding boundary condition for the atom is that the wave function ψ vanish at all points infinitely far from the nucleus.
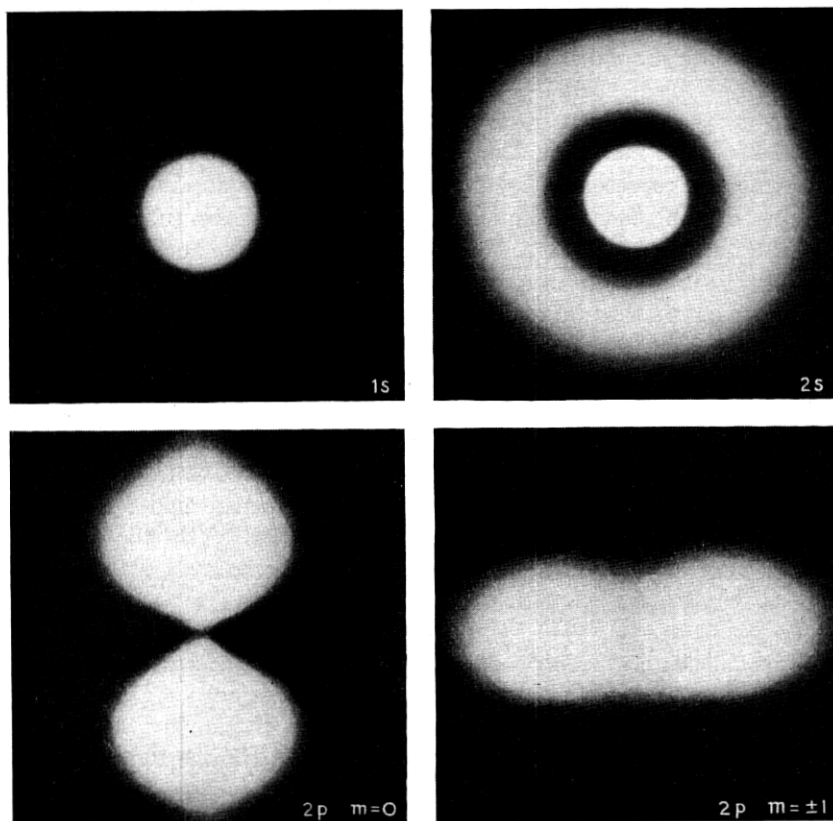

Fig. 8—The electron charge densities for four wave functions. Cross-sections are given for the 1s and 2s wave functions and perspective views for the 2p. 1s represents a ball of charge; 2s, a ball surrounded by a shell; 2p $m = 0$, a dumbbell-like distribution; 2p $m = \pm 1$, a doughnut-like distribution seen edgewise.

The quantity $|\psi|^2$ has a direct physical interpretation: its value at any point in space gives the probability of finding the electron at that point. If it were possible to take a photograph of the electron's motion with a time exposure so long that a true average of its positions would be obtained, this photograph would represent $|\psi|^2$. In Fig. 8

we show the predicted patterns as obtained by H. E. White,[6] who photographed a model representing the wave functions. We see that for the 2s wave function the electron is much farther from the nucleus on the average than for the 1s; this accounts for higher energy of the 2s state. For a hydrogen atom the 2s and 2p actually have the same energy. For other elements the 2s lies lower as shown in Fig. 2; this is because an electron in the 2s state penetrates the K shell and feels the full charge of the nucleus whereas an electron in the 2p state stays outside of the K shell and is thus shielded from the nucleus by the two electrons of the K shell.

For purposes of illustration we have considered the rectangular drumhead as a mechanical analogue for the wave equation. Other analogues are represented by sound waves in rooms and in organ pipes and by standing electromagnetic waves in wave guides, tuned cavities, and rhumbatron oscillators. We shall use two simple analogues in our later discussion. One is the mechanical vibrator represented in Fig. 7d which we consider to be restricted to vertical motion. It is a system with a single frequency—like an imaginary atom with only one possible state—and its one normal mode of vibration is a simple harmonic motion up and down equally far above and below its equilibrium position as indicated in Fig. 7e. The other is an electrical analogue, Fig. 7f, consisting of a section of transmission line terminated at each end by a high inductance. This system has a series of normal modes of vibration and a related series of allowed frequencies. The allowed frequencies correspond to the energy levels of the atom.

ELECTRONS IN MOLECULES

We shall next consider what happens when two atoms are brought so close together that their quantum states "interact." Two similar atoms widely separated have each a distinct set of quantum states and wave functions and the scheme of energy levels for the two atoms is obtained by duplicating the energy level scheme of Fig. 2. However, if the atoms move so near together that the wave functions for the corresponding quantum states of the two atoms overlap, there is an alteration in the energy levels. Figure 9 is intended to illustrate this process. Figure 9a shows the potential energy of an electron for points on a line passing through the centers of the two nuclei, and Figs. 9b and 9c show for points on the same line the values of the correct wave functions in this field. These wave functions are obtained, approximately, by using the 1s wave function for the two separate

[6] *Physical Review*, **37**, 1416 (1931). I am indebted to Professor White for the photographs used for these illustrations.

atoms; b represents the sum of the wave functions and c the difference. The process involved in getting these molecular wave functions is mathematically similar to that of finding the normal modes for a system of two similar coupled oscillators. In Fig. 9d we represent two
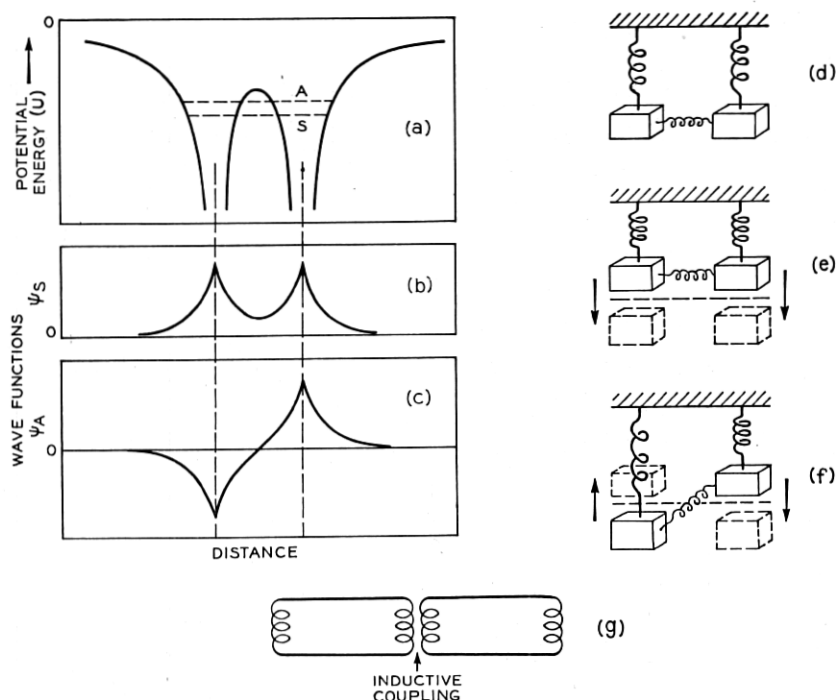

Fig. 9—A diatomic molecule and some mechanical and electrical analogues.

(a) The potential energy of an electron for points on a line through the two nuclei.
(b) and (c) Values of two wave functions for points on the same line.
(d) Two coupled oscillators.
(e) and (f) Their normal modes of vibration.
(g) Two coupled circuits.

weakly coupled oscillators. The normal modes of vibration for the coupled system are as indicated in Figs. 9e and 9f. These two modes have different frequencies. Similarly if two electrical circuits are placed so that there is some inductive coupling between them, we find that each frequency is split into a pair. This inductive coupling is similar to the overlapping of the wave functions; thus the coupling between the circuits is large when the electromagnetic field of one reaches over to the other. We may summarize the situation by saying that before coupling each frequency occurred twice, once for each sys-

tem; after coupling two frequencies are present and the corresponding modes of vibration belong not to the individual systems but instead to the pair of systems. For the case of atoms each quantum state occurs twice, once for each atom, before the atoms interact; after interaction there are still two quantum states but now they have different energies and are shared by both atoms.

As the atoms are brought closer together the energies separate more and more. The behavior is indicated qualitatively in Fig. 10. The
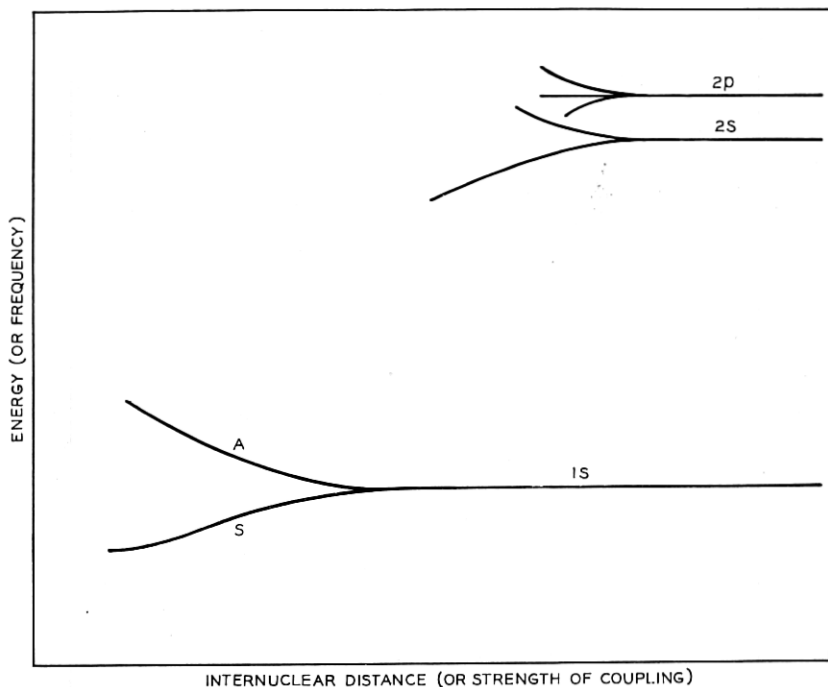

Fig. 10—Energy levels of a diatomic molecule versus internuclear distance.

L levels ($2s$ and $2p$) split at larger distances than the K levels because their wave functions extend farther from the nucleus (see Figs. 7 and 8) and overlap at greater distances. The details of the splitting are somewhat complicated and only the start is shown here. For the mechanical analogues shown in Fig. 8, the coupling raises one frequency and leaves the other unaltered. On the other hand, the quantum mechanical interaction results, at large distances, in equal displacement up and down for the energy levels.

We can use Fig. 10 to describe the formation of a molecule of hydrogen, H_2. We start initially with the single electron of each atom

in the 1s state. When the atoms have come together the 1s states have split into two energies with two states for each energy—one with each spin. In the H_2 molecule, the two electrons will both go into the lower 1s states, which both have the wave function of Fig. 8b. Bringing the atoms closer together decreases the energy of the electrons and results in the binding together of the atoms. This tendency of the electrons to reduce their energies by drawing the atoms together is opposed by the electrostatic repulsion between the nuclei. The repulsion between the nuclei is inoperative when the atoms are sensibly separated because then each nucleus is shielded from the nucleus of the other atom by the electron of that atom. When the atoms are closer together, however, the electrons no longer perform this shielding perfectly and the nuclear repulsions are important. Hence with decreasing interatomic spacing the electronic energy decreases and the energy of repulsion of the nuclei increases, and the equilibrium internuclear distance is the one which makes the total energy of the molecule a minimum.

The situation is quite different for two helium atoms. There being two electrons in each, for them all four of the "1s molecular orbitals," as the states of Fig. 9 are called, are occupied. When all the molecular orbitals are occupied, there is no decrease in energy when the two atoms are brought together: in this case the decrease of energy for the electrons in the two lowest states is compensated by the increase for the electrons in the upper states—more than compensated, as a matter of fact, because the upper states rise slightly more rapidly than the lower ones fall. This effect results in a repulsion between two helium atoms. This repulsion is a consequence of the closed shell nature of the helium atom and always occurs between such closed shells even if the atoms are different, as, for example, a neon and an argon atom. We shall refer to this closed shell repulsion, which occurs when the wave functions of the two closed shells encroach upon each other, as an "encroachment energy." The encroachment energy, as we have said, always corresponds to a repulsive force between the closed shells. We shall find that it plays a very important role both in ionic crystals and in metals.

The encroachment energy occurs not only between rare gas atoms but also between ions of elements which as neutral atoms have partly filled shells but in the ionic form have closed shells. Consider, for example, an alkali halide molecule such as LiF. For this case the 2s valence electron of lithium is transferred to the vacant 2p level of fluorine (see Fig. 4), thus leaving two ions with closed shell configurations, the Li^+ being He-like, the F^- being Ne-like. These two oppo-

sitely charged ions attract each other and draw together until the encroachment repulsion between their closed shells balances the attraction and holds them apart. Conversely if one of two atoms having closed shells normally is converted to an ion, the closed shell arrangement will be destroyed and an attraction will result. For example, He_2^+ ions, which may be thought of as formed from an atom and an ion, have been observed in the mass spectrograph.[7] The attraction is explained by noting that in this case there are three electrons and the effect of two of them in the lower $1s$ molecular orbital overbalances the one in the upper orbital and gives rise to a net attraction.

ELECTRONS IN CRYSTALS

We must now investigate the quantum states and their energy levels for electrons in crystals. As in the case of the diatomic molecule we shall study the dependence of the energies upon the distance between atoms, which in the case of a crystal is called the lattice constant. We shall treat the lattice constant as a variable and shall refer to the values for it found experimentally as "observed" or "experimental lattice constants" and indicate them on the figures by the symbol a_0. We shall consider the allowed states to be occupied in accordance with Pauli's principle and on this basis find how the energy of the crystal as a whole depends upon the lattice constant. In this section we shall deal with crystals at the absolute zero of temperature and leave the complicating features of thermal effects to a later section. According to theory, the equilibrium state of a system at absolute zero is that one which makes the energy least. Hence, a knowledge of the dependence of energy upon lattice constant can be used to predict the equilibrium lattice constant—that is, the one which should be found experimentally—for according to the theory quoted above, the equilibrium lattice constant is the one which makes the energy of the crystal least.

In Fig. 11 we show the potential energy for an electron in a one-dimensional crystal, the distance being measured along a line passing through the atomic nuclei of the constituent atoms. In the interests of simplicity we imagine that high potential walls through which the electron cannot pass bound the crystal at both extremities. These boundary conditions lead to a simpler set of wave functions than would boundary conditions like those discussed for the free atom. The simplification of problems by arbitrarily choosing certain boundary conditions is a standard device in some branches of quantum mechanics; it introduces an error, but if the crystal is large, the error is negligible;

---

[7] F. L. Arnot and Marjorie B. M'Ewen, *Proc. Roy. Soc.*, *171*, 106, 1939.
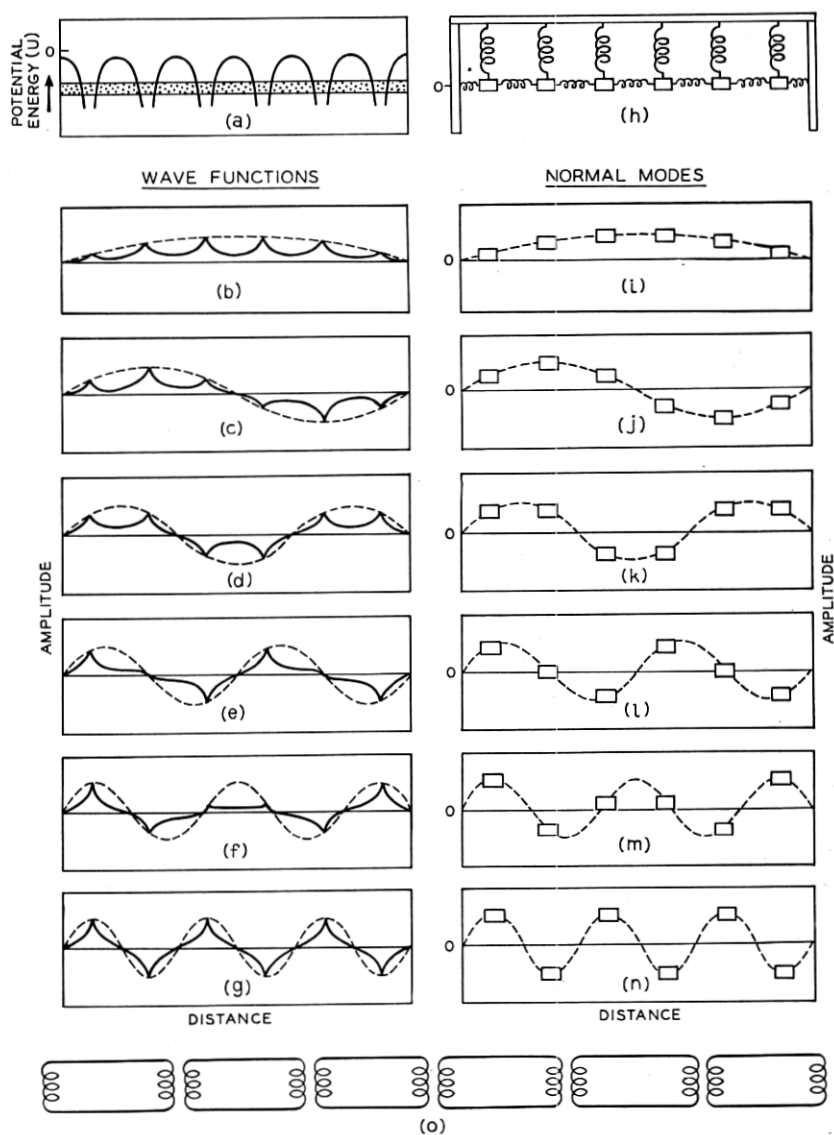
WAVE FUNCTIONS   NORMAL MODES

Fig. 11—A one-dimensional crystal and some mechanical and electrical analogues.

(a) The potential energy of an electron for points on
     a line through the nuclei.
(b) to (g) Wave functions for points on the same line.
(h) Coupled oscillators.
(i) to (n) Their normal modes of vibration.
(o) Coupled circuits.

the situation is similar to that which arises through neglecting "edge effects" in calculating the capacity of a parallel plate condenser.

In Fig. 11 we show also a series of coupled oscillators with boundary conditions corresponding to those prescribed for the atoms. For this case there are six coupled oscillators, which when uncoupled had six independent normal modes of vibration all with the same frequency, like that shown for the single oscillator of Fig. 7d. After coupling there are six normal modes all having different frequencies; the standing wave patterns corresponding to these are shown in Figs. 11i to 11n. A similar splitting of frequencies occurs when the members of a set of electrical circuits are placed in close proximity as indicated in Fig. 11o. For them the situation is more complicated than for the mechanical oscillators; each mechanical oscillator has but a single frequency, whereas each circuit has a fundamental and a sequence of overtones. Each possible frequency for the electrical circuits is split by coupling into a set of six.

In Figs. 11b to 11g are shown the proper electron wave functons which arise from the 1s atomic states. These wave functions have different energies. When the atoms were separated there were six 1s wave functions for the six atoms and each of these gave two states— one for each spin. After coupling we find six crystal wave functions and twelve crystal quantum states, the same number of states for each spin as before. This illustrates a fundamental theorem concerning wave functions in crystals which holds for two and three dimensions as well as for one and is true no matter how large the number of atoms in the crystal. This theorem, which we shall refer to as the "conservation of states," may be stated as follows: consider a set of N similar isolated quantum mechanical systems; they may be single atoms or molecules. Any particular quantum state is then repeated N times over, once for each system. Now bring the systems together so that the energy levels have split up. Then for each N-times-repeated quantum state of the isolated systems, we find a set of N crystal quantum states. In other words, putting the systems together may change the energies and wave functions of the quantum states but no states are gained or lost in the process.

In Fig. 12 we indicate how the energy levels of the states depend upon the lattice constant. Each energy level in the figure corresponds to two states, one for each spin. For simplicity only two atomic levels are shown here. Higher energy levels split appreciably at larger lattice constants because of the greater spatial extension of their wave functions. For any particular lattice constant the energy levels arising from a given atomic state lie in a certain band of energy. The

*number of states in the band is, of course, proportional to the number
of atoms;* however, if the number of atoms is large, the *width of the
band is independent of the number of atoms.* This concept of allowed
bands of energies for the crystal states plays the same role in crystals
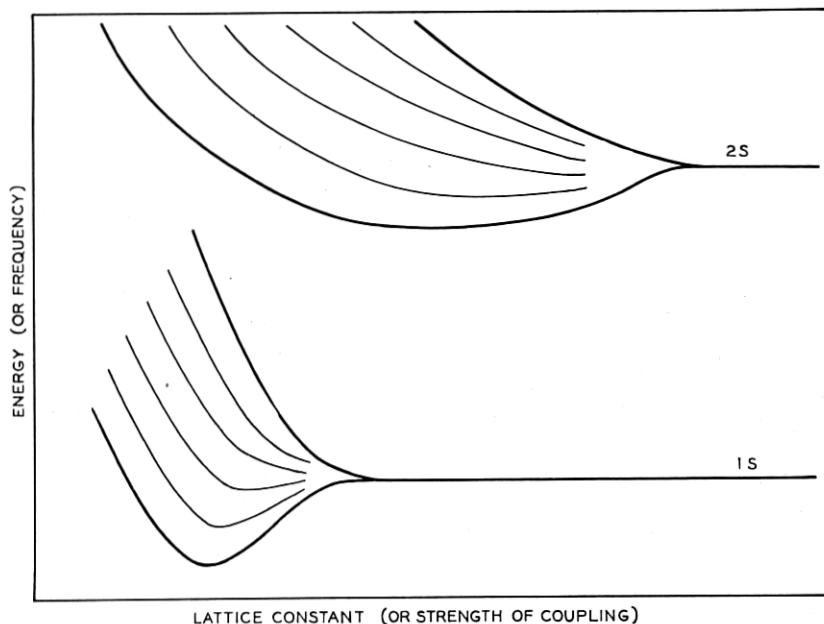as the concept of energy levels in the atom. We shall refer to 3s bands

Fig. 12—Dependence of energy levels upon lattice constant or frequency
of vibration upon strength of coupling.

and 3d bands of energy levels in crystals in much the same way as we
refer to the 3s and 3d atomic energy levels from which these bands arise.

We must emphasize that like the molecular states, the crystal states
do not belong to the atoms individually but instead belong to entire
system of atoms.

Before proceeding with the application of these ideas to crystals
with large numbers of atoms, we shall digress by anticipating several
subjects to be taken up in the next paper. For the energy levels of
isolated atoms the quantum numbers n, l, m, and m_s were satisfactory.
For a crystal, however, there will be many crystal quantum states
in an energy band all arising from atomic levels having the same values
of n, l, m, and m_s. A new quantum number is therefore needed to dis-
tinguish the various crystal states in an energy band one from another.
In Fig. 11 we see that the wave function of each crystal state is asso-

ciated with a wave form, shown dashed. This wave form is in every case of such a wave-length that it has an integral number of half wave-lengths along the edge of the crystal.[7a] The number of half wave-lengths is a suitable quantum number for the wave functions in the crystal and a more general consideration of it in the next paper will lead us into the theory of the "Brillouin zone" and the zone structure of energy bands. The second subject concerns the transmission properties of the crystal. The set of coupled circuits of Fig. 11 constitutes a length of transmission line. A line of this type is a simple filter network and as such it has bands of frequency in which power will be transmitted and bands in which it will not. The allowed frequencies lie in the transmitting bands. The system of coupled mechanical vibrators likewise constitutes a mechanical filter. Just as the mechanical and electrical systems can transmit power in their allowed bands, a crystal can transmit an electron whose energy is in an allowed band. The electrons in an allowed band, however, can produce a net current only if the band is partially filled. Electrons in wholly filled energy bands, although individually representing tiny currents to and fro in the crystal, can produce—we shall find—no net current as their individual currents cancel out in pairs. On this basis the theorist explains the difference between metals and insulators as follows: in a metal some of the energy bands are partly filled, but in an insulator each energy band is either completely filled or completely empty.

Distributions of Quantum States in Energy Bands

When there are a very large number of atoms in the crystal, it is impractical to represent the energy levels by distinct lines as was done for the case of six atoms in Fig. 12 and another scheme must be used. For a crystal of macroscopic dimensions the number of levels in the band is of the order of 10^{24}, that is a million million million million. When so many levels are placed so close together, a continuous band of allowed energies is suggested. Actually, of course, only a discrete set of allowed energies is possible, the total number in the band being that required by the conservation of states. We shall now consider the distribution in energy of these quantum states; that is, how many lie in a given range of energy between E and $E + dE$. Let us call

[7a] For a three-dimensional crystal having the external shape of a cube, the three-dimensional wave function has an integral number of half wave-lengths along lines parallel to each edge of the crystal. This condition is illustrated in a simplified form by the wave patterns for the two-dimensional drum head shown in Fig. 7; for each normal mode, there is an integral number of half wave-lengths parallel to each boundary of the membrane, and, in fact, the values of these numbers are given by p and q.

this number dN; it will depend upon E and be proportional to dE and we may write

$$dN = N(E)dE, \qquad (3)$$

where the function $N(E)$ represents the "number of quantum states per unit energy" at E. This equation, like many in statistical mechanics requires special interpretation, because if dE is small enough—less than the spacing between levels in the band—it may include no levels. If, however, we always use small but not infinitesimal values for dE, so many levels will be included in it that equation (3) is quite satisfactory. In Fig. 13a we represent qualitatively the distribution in
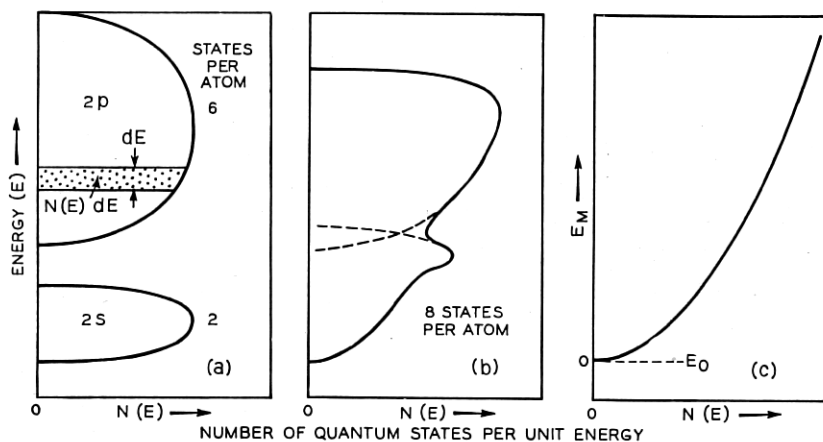
NUMBER OF QUANTUM STATES PER UNIT ENERGY

Fig. 13—Distribution of energy states in energy.
(a) For two separate energy bands.
(b) For overlapping energy bands.
(c) For free electrons.

energy for two energy bands. We plot $N(E)$ horizontally so as to retain the vertical scale for E. The area under the curve for the $2p$ levels is three times that for the $2s$. This is because the number of states in the $2p$ and $2s$ bands are respectively six times and two times the number of atoms in the crystal. The $1s$ band lies too low to be shown on this figure; its levels will be concentrated over a very narrow range in energy in keeping with the small splitting suggested in Fig. 12.

It is possible for the energy band arising from one atomic energy level to overlap the energy bands arising from other atomic levels. We shall be concerned below with several cases where this occurs for various crystals. When it does occur the states in the bands become mixed up and it is no longer possible to decide which atomic level was

the parent of each state in the band. This confusion is of no consequence, however, for it does not interfere with using the distribution in energy curves when they are obtained. Furthermore, the conservation of states holds when the bands overlap so that the total number of states per atom in the combined bands is the sum of the number of states per atom in the separated bands. In Fig. 13b we represent a distribution qualitatively similar to that occurring in various metals where s and p bands overlap. The number of states per atom in the combined bands is eight, four for each spin.

A very important distribution-in-energy curve is that of the case of "free electrons." This is the distribution one obtains by imagining that the electrons in a crystal are perfectly free—that is, subjected to no electrostatic forces whatever—but that they are required to remain within a certain prescribed volume. The distribution of quantum states in energy for this case is represented in Fig. 13c. At low temperatures the electrons tend to occupy the lowest states consistent with Pauli's principle and the system is referred to as a "degenerate electron gas." With the aid of the distribution curve, the energy and pressure of this gas can be calculated. We shall require its energy for a discussion of the binding energy of sodium, but we shall give here only the equation of the curve, leaving the calculation of the energy until later.[8] According to the theory, then, for the case of free electrons the number of states per unit energy is given by

$$N(E) = \frac{4\pi V}{h^3} (2m)^{3/2} E^{1/2}, \tag{4}$$

where V is the volume of container, h is Planck's constant, m the mass and E the energy of the electron; for free electrons E is all kinetic energy, there being no potential energy. For the case of the alkali metals, calculations show that the wave functions for the valence electrons are very similar to the wave functions for free electrons. For these metals we can use Eq. (4) to calculate energies.

Before utilizing the concepts of energy bands in a discussion of the binding energies of crystals, we must define two symbols to be used in describing the energy of a state in the band. For this purpose we arbitrarily separate the energy E of a crystal state into two parts: one of these is denoted by E_0 and stands for the energy of the lowest state in the band and the other is E_M which stands for the energy which the state possesses in excess of E_0—that is, its energy above the bottom

---

[8] The reader will find a derivation of this curve given in K. K. Darrow's article "Statistical Theories of Matter, Radiation and Electricity," *Bell System Technical Journal*, Vol. VIII, 672, 1929 or *Physical Review Supplement*, Vol. I, 90 (1929), and in various texts on quantum statistics and the theory of metals.

of the band. We shall find in the next paper that the quantum states in a band represent electrons traversing the crystal with various average speeds. The state E_0 has an average speed of zero. The subscript "M" has been assigned with these ideas in mind and stands for "motion," implying that an electron with energy greater than E_0 has an energy of motion E_M. In general both E_0 and E_M are actually composite energies containing both kinetic and potential energy; only in the case of free electrons is E_M purely an energy of motion. We shall not use in this paper the property of motion connected with the values of E_M; however, we shall use the division of the energy into two parts, E_0 and E_M, and we shall for convenience refer to the latter as an "energy of motion."

We shall next apply the concept of the energy band to a determination of the binding energies of several types of crystals. It is one of the principal merits of the theory of energy bands in crystals that we can treat many different crystal types on the basis of the same set of ideas. As we shall point out later, however, the band theory is most appropriate for metals: for ionic and valence crystals other theories are better suited.

Energy Bands and Binding Energies of Metals

For several metals the wave functions and distribution of states in energy have been found by solving Schroedinger's equation for the electrons in the metal. We shall discuss sodium since it constitutes one of the simplest cases and is the first metal for which good calculations were carried out.

A sodium atom, Na, contains ten electrons in filled K and L shells and one valence electron in the M shell; its electron configuration is $1s^2\, 2s^2\, 2p^6\, 3s$. When the atoms are assembled together as in the metal, the $3s$ atomic state gives a wide band which overlaps the $3p$ band while the lower levels widen only very slightly.

The formation of the energy bands[9] is shown in Fig. 14. Since the K and L bands are very narrow, it is possible to neglect the changes in the wave functions of the electrons occupying them and to concentrate upon the valence electrons. The valence electrons then move in a potential field produced by the Na^+ ions and the other electrons.

It can be shown by a lengthy argument that for the case of a monovalent metal, the energy of the metal as a whole is very nearly equal to the sum of the energies of the valence electrons.[10] It is rather

---

[9] J. C. Slater, *Phys. Rev.*, **45**, 794 (1934).

[10] An exact statement of the situation is too involved for this paper. The reader can find a more complete discussion in Mott and Jones "The Properties of Metals and Alloys," Chapter IV.

natural that the valence electrons should contribute so largely to the binding since the complete shells of electrons making up the Na+ ion, that is the K and L electrons, are only slightly affected by bringing the atoms together to form a metal. The result that the energy of the metal is the sum of the energies of the valence electrons is
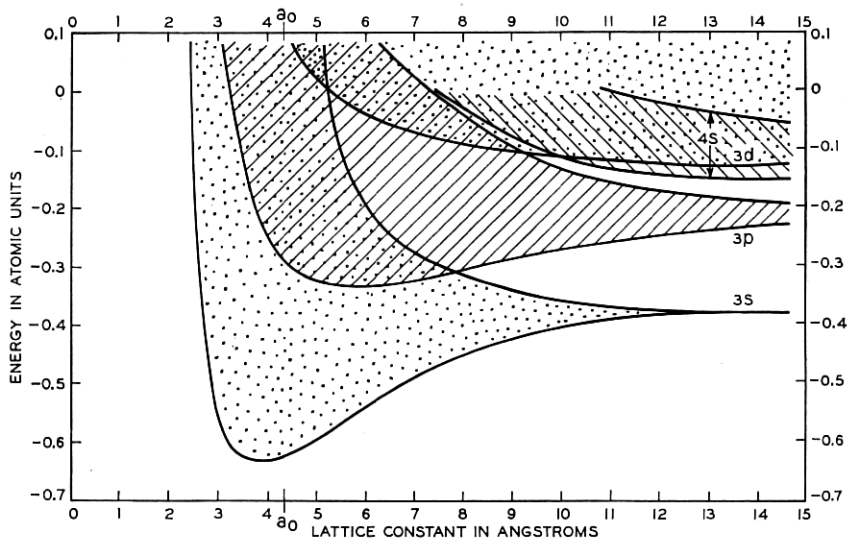

Fig. 14—Energy bands for sodium versus lattice constant.

of great importance in applying the theory. We shall discuss below how the energies of the various states in the band depend upon the arrangement of the atoms; some of these states are occupied and the energy of the crystal can be found by adding the energies of the occupied states. In this way we can find how the energy of the crystal depends upon the arrangement of the atoms and can find what arrangement makes the energy least. According to theory the arrangement of least energy is the stable one and the one which should be found in nature. The remainder of this section will be devoted to discussing the energies of the quantum states in metals and the energies of the electrons which occupy them.

The first satisfactory solutions of Schroedinger's equation for electrons moving in the field of a metal were obtained for sodium by Wigner and Seitz.[11] They assumed, in keeping with the findings of experiment, that the sodium atoms were arranged on a body-centered

[11] E. Wigner and F. Seitz, *Phys. Rev.*, *43*, 804 (1933) and *46*, 509 (1934) and E. Wigner, *Phys. Rev.*, *46*, 1002 (1934).

cubic lattice. They did not, however, assume that the lattice constant was that given by experiment but instead carried out calculations for each of several assumed values for the lattice constant lying on both sides of the experimental value. The results of their calculations are shown in Fig. 15. The curve marked E_0 is the energy of the lowest
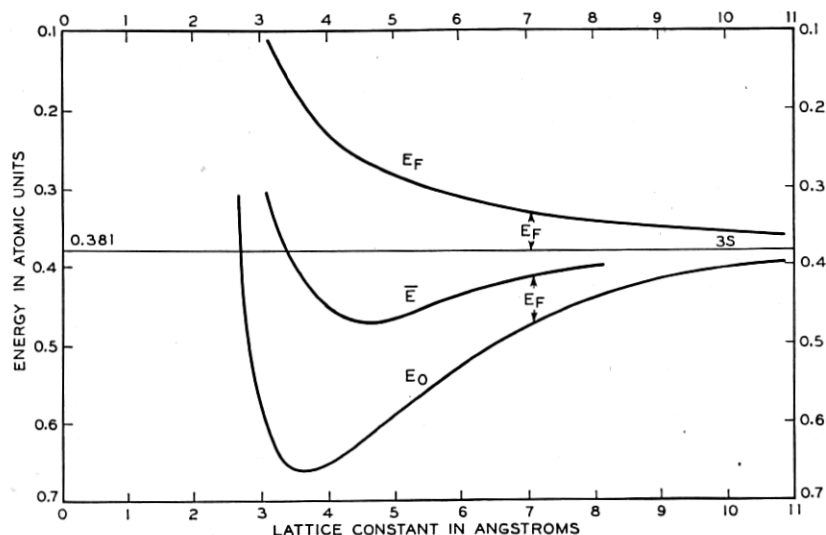

Fig. 15—Energy for sodium versus lattice constant.

level in the valence band. Only two electrons, one with each spin, can occupy this energy level and all others must occupy states of higher energy—that is, only two electrons can have zero value for the "energy of motion" E_M and all others must have larger values. By a method of calculation described below, it can be shown that the average energy of motion of a valence electron is given by the curve marked E_F in the figure. Hence the total energy per valence electron in the metal, which for a monovalent metal is equal to the energy per atom, is represented by the curve marked $\overline{E}$ in the figure; $\overline{E} = E_0 + E_F$. Figure 15 exhibits the dependence of this energy upon the lattice constant. The abscissa of the minimum in the $\overline{E}$ curve gives the theoretically predicted value for the equilibrium lattice constant. The binding energy or heat of sublimation is defined as the energy required to separate the metal into isolated atoms; it is the difference in energy between the minimum of the curve and the value of $\overline{E}$ for infinite lattice constant—that is, for free atoms. Finally, the curvature of the curve at its minimum is a measure of the energy required to compress or expand the crystal and from it a value for the compressibility can

be obtained. In Table I, we compare theoretical and experimental values of lattice constant, binding energy, and compressibility calculated by the method described above. The theoretical values were computed by Bardeen [12] who has added some refinements and corrections to the original calculations.

TABLE I

	Li		Na	
	Calc.	obs.	Calc.	obs.
Lattice Constant (angstroms) . . .	3.49	3.46	4.53	4.25
Heat of Sublimation (Kg. cal./- gm. atom)	34	39	23	26
Compressibility (cm²/dyne)	8.4×10^{-12}	7.4×10^{-12}	12.0×10^{-12}	12.3×10^{-12}

Although the theory can give quite satisfactory values for the various physical quantities shown in Table I, it cannot as yet predict precisely what crystalline form a metal like sodium will take. In carrying out the calculations discussed above, it was assumed that the atoms were arranged in a body-centered cubic lattice. Now the correct theoretical procedure would be to calculate the energy for all conceivable arrangements of the atoms and then to select that arrangement giving the least energy of all as the theoretically predicted equilibrium arrangement. This program is, of course, too laborious to be practical—furthermore experience shows that metals, with but few exceptions, crystallize in one of three forms: body-centered cubic, face-centered cubic, and hexagonal close-packed. For this reason it might be regarded as sufficient to calculate the energies for the face-centered cubic and hexagonal close-packed and to compare these with that for the body-centered cubic. When such calculations are carried out, however, it is found that the minimum energies calculated for the three forms differ among themselves by amounts which are negligible in view of approximations necessary in making the calculations. Hence the theory cannot predict with any certainty which form really has the lowest energy; it does predict, however, that all three forms do have nearly the same energy and gives a value for this energy. Actually the binding energy of sodium must be greatest for the body-centered cubic lattice because this form is the one that occurs in nature and so must be the form of lowest energy. However, it is probable that the difference in energy between the various possible allotropic

[12] J. Bardeen, *Jour. Chem. Phys.*, 6, 367, 372 (1938).

forms for sodium is really very small—so small that we should not expect the present theory to evaluate it. Some indication that the energy of caesium is very nearly the same in the body-centered form and face-centered form (or possibly the hexagonal close-packed form) is furnished by a transformation at high pressures observed by Bridgman. In the next paper we shall meet a case where the theory does seem able to differentiate between the energy of face-centered and body-centered structures. In general, however, the procedure is to use the crystal structure found by x-rays and to calculate the energy for a series of values of the lattice constant as was done for sodium.

We must now return to a discussion of the curves E_0 and E_F of Fig. 15. About the curve E_0 we shall only say that it is obtained by solving Schroedinger's equation for and finding the energy and wave function of an electron in the lowest state in the energy band. The wave function for this state, however, possesses the interesting feature of being very nearly the same as the wave function for a free electron having zero energy of motion. From this fact it is possible to draw the conclusion that the distribution in energy of motion of the valence electrons in sodium is the same as the distribution in energy of free electrons in an electron gas. Accepting this conclusion, we can then use the formulas given for the distribution of states for free electrons in order to calculate the mean energy of motion of the valence electrons in sodium. The results of this calculation, which we give in a footnote,[13] lead to the energy curve E_F. This energy curve is, from its

[13] We shall first derive a general expression for E_F without specifying the particular form of $N(E)$. Since in this footnote all energies are measured from E_0, we shall omit the subscript M from E_M and use simply the symbol E in the equations. The total number, denoted by n, of atoms in the crystal is equal to the total number of valence electrons. Let the volume of the crystal be V. Because of the duplicity due to the spin there are $2n$ states in the band, and half of them will accommodate the n electrons so that the band will be filled only up to a certain energy E_{max}. We must therefore have

$$n = \int^{E_{max.}} N(E)dE. \tag{i}$$

Once the distribution function $N(E)$ is known, this equation serves to determine E_{max}. The average energy of motion of an electron in these occupied states is, from the definition of an average, the total energy of motion divided by the total number of electrons:

$$E_F = \frac{1}{n} \int_0^{E_{max.}} EN(E)dE. \tag{ii}$$

Substituting the value of $N(E)$ for free electrons into the first equation gives

$$n = \frac{8}{3} V(2mE_{max.}/h^2)^{3/2} \tag{iii}$$

The quantity V/n is the volume per electron which in the case of a monovalent metal

definition, the average energy per electron of a degenerate electron gas. In a degenerate electron gas the electrons have the least possible energy consistent with Pauli's principle and with the distribution of quantum states in energy. For reasons associated with the origin of the statistical mechanics of electrons—that is, with the Fermi-Dirac statistics—the energy E_F is called the "Fermi energy" and given the subscript F. The energy E_F is far greater than the average energy per particle of an ordinary classical gas. We shall see below how this fact accounts for the very small specific heat of the electron gas. From the dependence of the energy upon volume, the pressure of the electron gas can be calculated. It is usually very large, for sodium it is about 50,000 atmospheres. The force that prevents this pressure from blowing the metal apart is represented by the E_0 curve, which gives decreasing energy with decreasing lattice constant and corresponds to a force pulling the atoms together. A more detailed discussion of these forces will be taken up in the third paper of this series.

Other Metals

Calculations similar to those for sodium can be carried out for other metals. The band structure as calculated for copper by Krutter [14] is shown in Fig. 16. Ten electrons per atom can be accommodated in the $3d$ band and two per atom in the $4s$. For copper the $3d$ band is filled—in keeping with the fact that the Cu^+ ion consists of filled K, L, and M shells. From the discussion of molecules given above

---

is the same as the volume per atom. Denoting by Ω the value of V/n, we find

$$E_{\max.} = \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{8m} \Omega^{-2/3} = 36.1\Omega_0^{-2/3}\text{ev} \qquad \text{(iv)}$$

where Ω_0 is the volume per atom in cubic Angstroms. For a body-centered cubic lattice with lattice constant a Angstroms, $\Omega_0 = a^3/2$. Substituting the expression for $N(E)$ into the equation for E_F gives

$$E_F = \frac{3}{5} E_{\max.} = 21.6\Omega_0^{-2/3}\text{ev}. \qquad \text{(v)}$$

Expressing E_F in atomic units and Ω_0 in terms of the lattice constant, we find

$$E_F = 2.54a^{-2}. \qquad \text{(vi)}$$

This is the equation of the curve for Figure 15. The values of $E_{\max.}$ calculated from the above equations for a series of metals are

Metal	Li	Na	K	Rb	Cs	Cu	Ag	Au
$E_{\max.}$(ev)	4.74	3.16	2.06	1.79	1.53	7.10	5.52	5.56

[14] H. M. Krutter, *Phys. Rev.*, 48, 664 (1935).

we should expect encroachment repulsions between these ions when their wave functions begin to overlap. In the band picture this repulsion results from the spreading of the $3d$ band; since the band spreads more to higher energies than to lower energies and since it is full, the average energy of an electron in it increases as the lattice constant decreases. Thus the same result, repulsion between closed shells, is
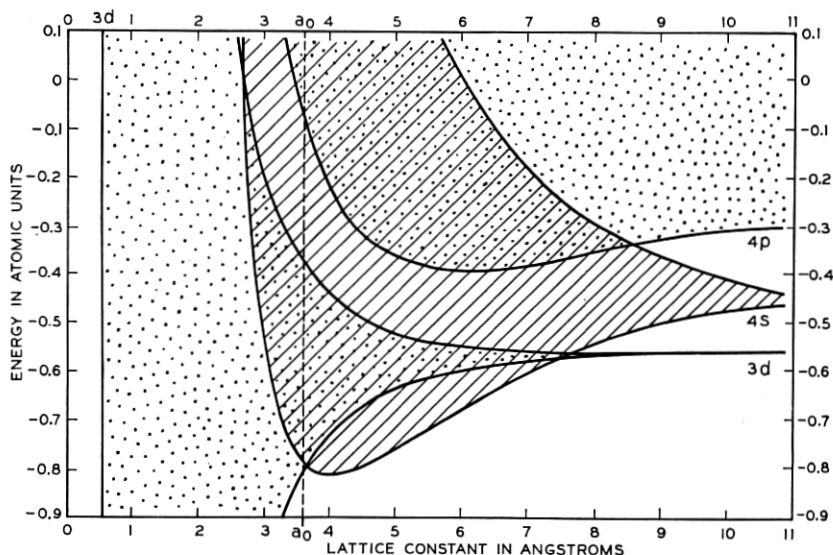

Fig. 16—Energy bands for copper versus lattice constant.

found for the ions in a metal as for the rare gas atoms. For elements whose atoms have partially filled $3d$ levels the situation is quite different. For them only part of the levels of the $3d$ band will be filled and there will be a decrease in the energy of the $3d$ electrons in the metal as compared to the atom. This has been proposed by Seitz and Johnson as an explanation of the fact that the highest binding energies for the metals of a transition series occur for those that have approximately half-filled $3d$ bands and for which consequently nearly all of the $3d$ electrons have lower energies than in the atomic state.[15] The very high melting point metals—columbium, molybdenum, tantalum, and tungsten—come approximately at the middle of their transition series. In Table II we give the binding energies for a number of the transition elements.

[15] F. Seitz and R. P. Johnson, *Jour. App. Phys.*, *8*, 84, 186, 246 (1937).

TABLE II

HEATS OF SUBLIMATION FOR SEVERAL METALS INCLUDING THE TRANSITION ELEMENTS IN KILOCALORIES PER GRAM ATOM *

K	Ca	Se	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
19.8	48	70	100	85	88	74	94	85	85	81	27.4
Rb	Sr	Y	Z	Cb	Mo	Ma	Ru	Rh	Pd	Ag	Cd
18.9	47	90	110	—	160	—	120	115	110	68	26.8
Cs	Ba	La	Hf	T	W	Re	Os	Ir	Pt	Au	Hg
18.8	49	90	—	185	210	—	125	120	127	92	14.6

* Taken from F. R. Bichowsky and F. D. Rossini "The Thermochemistry of the Chemical Substances," Reinhold (1936), except for T which was taken from D. B. Langmuir and L. Malter, *Phys. Rev. 55*, 1138 (1939).

Energy Bands of Diamond

In Fig. 17 we show the band structure for diamond as calculated by Kimball.[16] The configuration of the carbon atom is $1s^2 2s^2 2p^2$ and
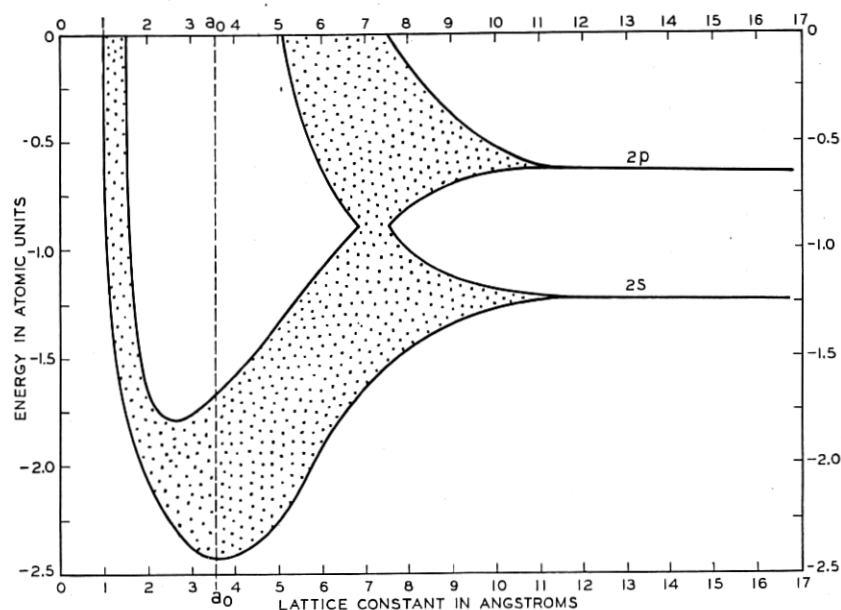
Fig. 17—Energy bands for diamond versus lattice constant.

all four of the L shell electrons are involved in the binding. At large lattice constants the lower band contains two states per atom and the upper six. The lower band is completely filled, the upper only one-

[16] G. E. Kimball, *Jour. Chem. Phys., 3*, 560 (1935). Some unimportant features resulting from approximations in Kimball's work have been modified in this figure.

third filled. To the left of the crossing of the bands, Kimball finds that both bands contain four states per atom so that the lower is filled and the upper is empty. The actual spacing in diamond occurs to the left of the crossover and, as we shall see in the next paper, the resultant filled band and empty band arrangement explains the absence of electrical conductivity for diamond. The diagram suggests an explanation for the conductivity in graphite; one of the lattice constants of graphite is known to be larger than the abscissa of the crossover of Fig. 17; hence in graphite there are partially filled bands and conduction.

The general downward trend of the bands in Fig. 17 indicates a strong binding energy for diamond; but quantitative calculations of the total energy have not been made.

The type of binding involved in diamond is quite like the binding of metals save that, owing to the absence of partially filled bands, there is no electrical conductivity. In both cases the energy arises from the lowering of energy levels as the atoms come together. In chemical terminology the binding of diamond is referred to as "homopolar" signifying that the atoms are all similarly charged, or rather uncharged. In crystals containing ions rather than neutral atoms, the cohesion is due largely to electrostatic forces and one refers to binding as "heteropolar" or "ionic."

Energy Bands and Binding Energies of Ionic Crystals

The energy band theory can be applied to the calculation of the binding energy of ionic crystals. Before discussing this application, however, it will be instructive to examine a somewhat simpler approach to the problem.

A sodium chloride molecule consists of a sodium ion and a chlorine ion. These ions have charges of $+e$ and $-e$ respectively and have a mutual electrostatic energy of

$$-\frac{e^2}{r}, \tag{5}$$

where r is their distance of separation. This electrostatic energy, which we shall refer to as the "coulomb energy," decreases with decreasing interatomic distance. If the ions are close together, as they are in a molecule, the energy of encroachment due to the overlapping of their closed shells must be considered; this energy increases with decreasing interatomic distance. The equilibrium distance is the one that makes the total energy, coulomb plus encroachment, a minimum.

Closely similar calculations can be carried out for a crystal. One finds the total coulomb energy of all the ions and the total encroachment energy; and then one finds the lattice constant that makes the total energy a minimum. The total encroachment energy is easily found; only atoms which are nearest neighbors in the lattice have appreciable overlapping with each other and it is therefore a straightforward and simple calculation to find the total number of encroachments in the crystal. The coulomb energy is not quite so simply found, however, because the electrostatic interaction of a given ion with its nearest neighbors is no more important than its interaction with its vastly larger number of more distant neighbors. The electrostatic problem is solved as follows: one considers a NaCl lattice which is perfect except for the absence at one lattice point of a Na^+ ion; one finds by known techniques of electrostatics the value at the vacant lattice point of the electrostatic potential due to the remaining ions; this potential is negative and has a value

$$ -\phi = -\frac{Me}{4a} = -\frac{3.49e}{a}, \tag{6} $$

where a is the lattice constant and M is a numerical constant known as Madelung's constant, which has a particular value for any special lattice; for the NaCl lattice, $M = 13.94$. If now a Na^+ ion is placed in the vacant lattice point, its electrostatic energy will be $-e\phi$. Similarly the electrostatic potential at a vacant Cl^- lattice point is $+\phi$ and the electrostatic energy of a Cl^- placed there is $-e\phi$. The total electrostatic energy per NaCl molecule in the lattice, however, is not $-2e\phi$ but only $-e\phi$; the factor 2 does not occur since otherwise the electrostatic interaction between each pair of ions would be included twice.[18] The total energy per molecule for the crystal can be found by combining the coulomb and the encroachment energies, and the equilibrium lattice constant and binding energy per molecule thence can be derived.

Using wave functions for Na^+ and Cl^- ions obtained by D. R. Hartree, who has found solutions of Schroedinger's equation numerically, the encroachment energies in NaCl have been evaluated by R. Landshoff.[19] For the lattice constant and binding energy for NaCl he obtains 5.88Å and 165 Kg.-cal./gm. atom while experiment gives 5.63Å and 183 Kg.-cal./gm. atom.

Some very important theoretical work of a semi-empirical nature has

[18] To see that this is true in a simple case, use the procedure described above to calculate the electrostatic energy of an isolated NaCl molecule.
[19] *Zeits. f. Phys.*, *102*, 201 (1936).

been carried out for the alkali halides. In it an analytical expression suggested by theory and containing adjustable constants has been used for the closed shell repulsions. The adjustable constants have been determined from certain data and then used for predictions which can be compared with other data. Using a relatively small number of adjustable constants, Born and Mayer,[20] Mayer and Helmholz,[21] and Huggins and Mayer [22] have calculated a much larger number of values for lattice constant and binding energy for many alkali halides with an agreement with experiment of the order of one per cent.

Let us now consider NaCl using the band picture. We shall reach the rather surprising conclusion that there is no fundamental difference between the results obtained from it and those just deduced from the ionic picture described above.

In Fig. 18 we show qualitatively the behavior of the bands for NaCl.[23] In the ionic state, an electron is transferred from the Na $3s$ to the Cl $3p$. The general shifting of the bands is explained as follows. The wave functions corresponding to the Cl$^-$ $3p$ band, like all energy band wave functions, are distributed over the whole crystal. They are not, however, equally intense at Na$^+$ and at Cl$^-$ ions; instead they are definitely concentrated about the Cl$^-$ ions. The electrostatic potential at a Cl$^-$ ion, due to the remainder of the crystal, has the same value (6) as was found in discussing the ionic method. Since the charge on the electron is $-e$, the energy of each of the states in the Cl$^-$ $3p$ band varies with a in the same manner as does $-e\phi$. A similar argument shows that the Na$^+$ energy bands vary as $+e\phi$. At a certain lattice constant, the Cl$^-$ $3p$ and $3s$ bands and the Na$^+$ $2p$ and $2s$ bands begin to widen. Since these bands are full, this widening gives the customary encroachment energy just as it was obtained in the ionic picture. The shifting of the bands similarly gives the coulomb energy. To see this we note that per NaCl molecule there are 18 electrons in the Cl$^-$ bands where energies vary as $-18e\phi$ and that there is also one chlorine nucleus with charge $+17e$ whose energy varies as $+17e\phi$. This leaves a net effect of $-e\phi$ for the Cl$^-$ ions. Similarly a net effect of $-e\phi$ comes from the electrons and nuclei of the Na$^+$ ions. As in the case of the ionic method the sum, $-2e\phi$, of these energies really contains each ionic energy twice and the total electrostatic energy per NaCl molecule is $-e\phi$. So far as calculating energies is concerned, the two methods give equivalent results; the advantage, if any, lies

[20] *Zeits. f. Phys.*, 75, 1 (1932).
[21] *Zeits. f. Phys.*, 75, 19 (1932).
[22] *Jour. Chem. Phys.*, 1, 643 (1933).
[23] J. C. Slater and W. Shockley, *Phys. Rev.*, 50, 705 (1936).

with the ionic method rather than the band method because of the more immediate physical interpretation of the former.

We may remark that in the discussion of metallic sodium, it was not necessary to consider the potential energy of the nuclei and the closed
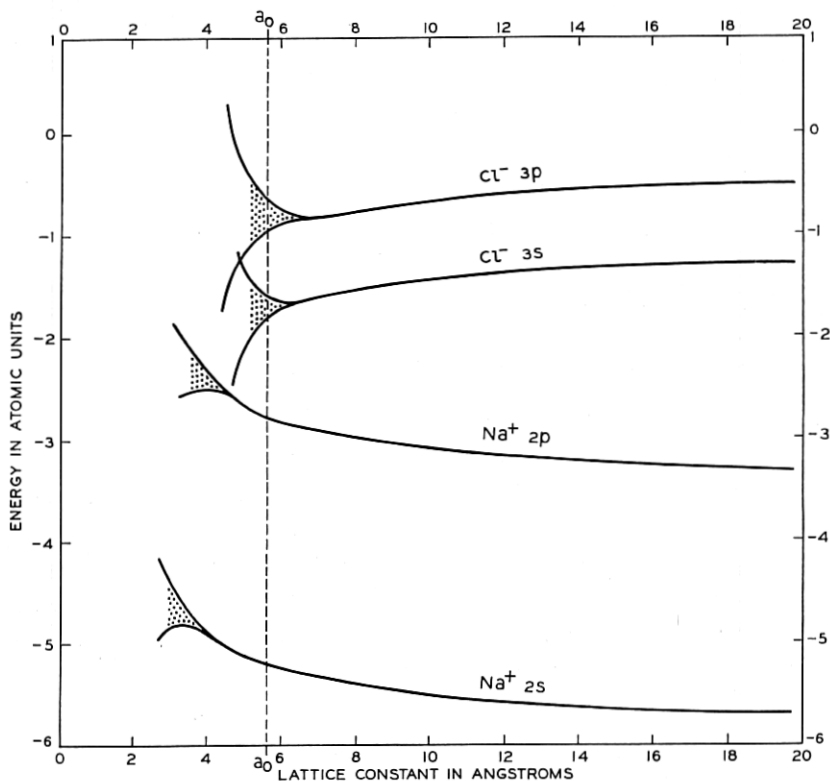

Fig. 18—Energy bands for sodium chloride versus lattice constant.

shell electrons as was done in NaCl. This is because sodium metal is not ionic—although we think of it as consisting in part of Na^+ ions, the electrostatic forces between them are suppressed by the shielding effect of the electron gas. In an ionic crystal, like NaCl, there is no electron gas and the coulomb energy must be considered in the manner described above.

Energy Bands for Other Crystals

There are chemical compounds which lie between the homopolar and ionic types. For example in the sequence of compounds NaF, MgO, AlN, SiC the compounds are progressively less and less definitely ionic—the least ionic, SiC or carborundum, being homopolar. Simi-

larly there are compounds, in particular intermetallic compounds, which are more like metals than like either ionic crystals or valence crystals. Thus there is an intermediate field which connects all three of the simple types of binding. Good computations are lacking for these intermediate cases; we shall return to a discussion of some aspects of them in connection with semiconductors in the next paper.

Concerning a Classification of Crystals

In the last section we saw how the concept of the energy band can explain the binding energies of a number of different types of crystals. Although the band theory has the merit of being very general it has the disadvantage of being at the same time rather abstract. Other theories have been developed to explain the cohesion of particular types of crystals; and, while lacking the generality of the band theory, they have the advantage of a more immediate physical interpretation in their own particular fields. In this section we shall digress from the exposition of the band theory in order to describe briefly some of the simpler viewpoints of the other theories.

We have discussed in the last section three types of binding. Sodium exemplified the metallic type; diamond, the homopolar or valence type; and sodium chloride, the ionic type. The distinction between the valence bond and the metallic bond is not very clearly indicated in the band theory; the only difference there had to do with the degree of filling of the bands. There is another difference, however, which has been long familiar to chemists. The homopolar compounds are usually characterized by "directed valence." Thus the "tetrahedral carbon atom" is a familiar concept of organic chemistry. In crystals in which homopolar binding is dominant the atoms are arranged so that each atom has the proper valence bonds with its neighbors. In diamond each carbon atom is tetrahedrally surrounded by four other carbon atoms. In silicon carbide, carborundum, a similar situation prevails: each carbon is tetrahedrally surrounded by four silicons and vice versa. These crystals are said to have a "coordination number" of four, or $z = 4$, meaning that each atom has four nearest neighbors. In crystals of the divalent elements—sulphur, selenium and tellurium—each atom has two near neighbors and the valence condition is satisfied; these crystals have a coordination number of two. The monovalent halogens form crystals in which each atom has one near neighbor, coordination number one. In the metals, however, the neighbors of a given atom are as many as eight or twelve—do these large coordination numbers imply that the metals have eight or twelve electron pair bonds with their neighbors?

According to the quantum mechanical theory of valence, which in itself forms a theory with as many ramifications as the band theory, the electron configuration $1s^2 2s^2 2p^2$ of carbon is especially suited for forming "electron pair bonds" with other atoms. In forming these bonds the wave functions from one atom and another become distorted so as to overlap and form a high electron concentration along the line between the atoms; the energy levels being incompletely filled for the atoms, this overlapping does not produce a repulsion but instead a binding together like that produced by the overlapping wave function in Fig. 9b in the hydrogen molecule. The carbon atom is capable of forming four such bonds and forming them most effectively along four lines, making the tetrahedral angles with each other.

Recently Brill [24] and his collaborators using x-ray analysis have determined the electron concentration in diamond, in which the carbon atoms are arranged in a tetrahedral manner. The results of their investigations are shown in Fig. 19A.[25] It is easily seen that the electrons are concentrated in the bonding directions forming homopolar bonds between the atoms.

The energy band theory, we have said, does not give the clearest picture of the valence crystals; it is, however, especially suited to treatments of the metallic bond. According to the band theory the valence electrons constitute an electron gas—that is, instead of forming electron pair bonds with localized overlapping of the wave functions, they form instead a more or less uniform region of negative charge. In this negative charge the positive ions float. Since the ions repel each other they tend to arrange themselves so as to use their space to best advantage and this requires that they take up one of the "close-packed" arrangements. Let us see why this is true. The close-packed arrangements are those obtained by trying to pack rigid spheres as compactly together as possible. For these arrangements then, the volume per sphere is less than for other arrangements; that is, the close-packed arrangements are the ones which give a minimum volume per sphere for a prescribed value for the distance between sphere centers. Conversely, the close-packed arrangements must be the ones which give a maximum value for the distance between neighboring sphere centers for a given value of the volume per sphere. Since the energy of motion, E_F, of the electron gas and, although we have not shown why, the energy E_O, depend for a metal mainly upon the volume, in metals we are interested in cases where the volume per

[24] R. Brill, H. G. Grimm, C. Hermann and Cl. Peters, *Ann. d. Physik, 34*, 393 (1939).
[25] The writer is indebted to Professor Grimm for his permission to reproduce Fig. 19 from his article: *Naturwissenschaften 27*, 1 (1939).

atom is closely prescribed. For a given volume per atom, as we have seen above, the close-packed arrangements are the ones which give the largest separation between neighboring positive ions; and since the positive ions repel each other, the close-packed arrangements will give the lowest energies. This accounts for the fact that the custom-
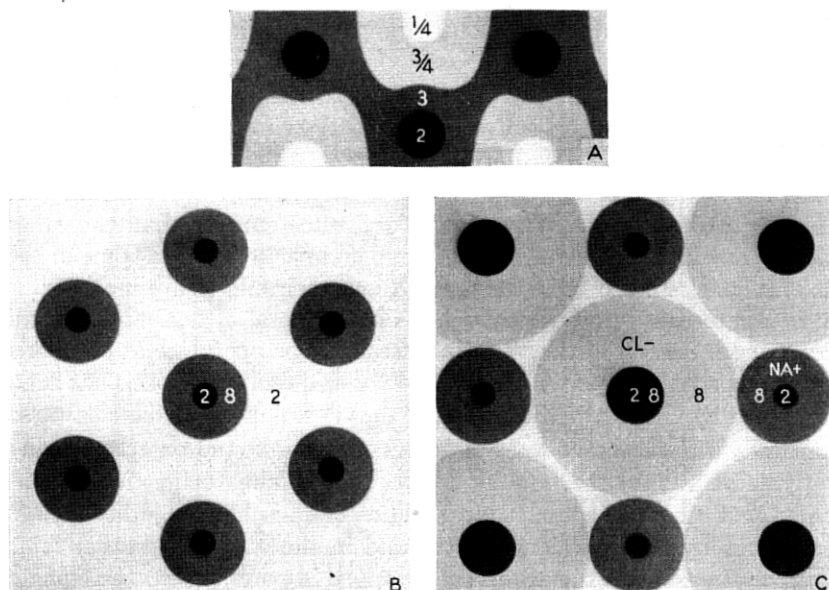

Fig. 19—Electron charge densities in crystals. The numerical values give the number of electrons per atom in the space corresponding to each intensity of shading.

(A) In diamond.
(B) In magnesium.
(C) In sodium chloride.

ary metallic lattices are the body-centered cubic lattice, the face-centered cubic, and the close-packed hexagonal. A further discussion of this physical picture of the nature of the metallic state will be given in the third paper of this series. In Fig. 19B we show qualitatively the electron density of metallic Mg according to Grimm.[25] It is seen that the valence electrons give a uniform negative charge in which the positive ions are embedded.

We have seen in the preceding section that the band theory of the alkali halides is essentially equivalent to the ionic theory. A large fund of evidence attests to the validity of the ionic theory, one item being the electron concentrations determined for sodium chloride by Brill[24] and his collaborators. These are represented in Fig. 19C; we

[25] Loc. cit.
[24] Loc. cit.

see that the electrons are closely held about the ions with very little overlapping of the closed shells.

In addition to the metallic, homopolar, and ionic bonds, there is still another interatomic force known as the Van der Waals force—a very weak force compared to the other three.   We shall not discuss its origin here except to say that it arises from the spontaneous and mutual polarization of two atoms or molecules when in the neighborhood of each other.   It is responsible for the "a" term in the Van der Waals equation for gases.   When a crystal is formed from organic molecules, such as a crystal of benzene, the forces holding them together are the weak Van der Waals forces.   This is the reason why "molecular crystals" have low melting points and binding energies. Although the Van der Waals forces are much smaller than the other three, they are not entirely negligible in comparison and in some of the calculations referred to in the last section, their effects are included.

It is interesting to note that in a single crystal of a given chemical compound, several of the various forces may be operative at once in a rather separable way.   A classification of this sort for crystals has been discussed by Grimm.[25]   For example the crystal mica, which cleaves so naturally into sheets, consists of planes of atoms bound together chiefly by valence forces, the binding between the planes being due to ions lying between and in the planes.   Thus mica is held together in two directions by strong valence forces and in the other by weaker ionic forces.   In asbestos the atoms are arranged in parallel rows, being held together in the rows by valence forces; the rows, on the other hand, are held to each other by ionic forces.   The ionic bonds are more easily broken and asbestos crystals exhibit a typical fibrous structure.   Mica and asbestos are intermediate members of a sequence of which diamond with all valence binding and sodium chloride with all ionic binding constitute the extremes.   We shall give one more example: cellulose consists of long chains of carbon, oxygen and hydrogen, the chains held to each other by Van der Waals forces; it is an example of valence binding in one direction and Van der Waals binding in the other two.

This section has been a digression, as the main purpose of these papers is to illustrate the band theory of solids.   It would hardly be fair to concentrate on this, however, without pointing out, as has been done in this section, that, although the band theory has great generality, it is best adapted for a certain class of solids and that other viewpoints are more natural for solids outside of this class.

[25] *Naturwissenschaften*, 27, 1 (1939).

THERMAL PROPERTIES OF CRYSTALS

In this section, as in the last, we shall digress from a straightforward exposition of the theory of energy bands and discuss the theories of specific heat and thermal expansion. These theories are well worth discussing on their own merits and furthermore their results and methods can be applied later to other topics. Thus the thermal vibrations that account for the specific heat will be shown in the second paper of this series to account for the resistance of metals. The discussion of thermal expansion given here will in the next section on magnetism be extended to an explanation of the unusual expansion properties of magnetic materials, in particular to an explanation of the very small expansion of invar. We shall, however, make use of the band theory once in this section by showing why the free electrons in a metal do not normally make an appreciable contribution to the specific heat.

In the introduction to this paper we pointed out that the specific heat per gram atom of a solid should be by classical theory $3R$—coming half from the kinetic energy and half from the potential energy of the atoms. This prediction is in reasonable agreement with experiment for many crystals at high temperatures. As the temperature is lowered, however, the observed specific heat decreases in such a way as to approach zero when the absolute zero of temperature is approached. This decrease in the specific heat at low temperatures, as well as the value $3R$ at high temperatures, is readily explained by quantum mechanics. In order to understand the explanation we must inquire into the atomic vibrations of a crystal.

In considering atomic vibrations we are really concerned with the motions of the nuclei. The electrons act as a cement to hold the nuclei in their equilibrium positions and exert restoring forces on them when they are displaced. (We shall see below why the electrons do not partake of the thermal energy.) The nuclei are effectively mass points in this theory and for quantum mechanical reasons, which we shall not discuss, they are incapable of acquiring thermal energy of rotation; hence so far as the crystal vibrations are concerned, we need consider only their translational or rectilinear motions. A crystal containing N atoms has $3N$ degrees of freedom since each nucleus can move in three dimensions. In order to find the specific heat of a crystal we must find the normal modes of vibration. The system of coupled oscillators in Fig. 11 represents reasonably well the normal modes of vibration for a one dimensional crystal whose atoms have only one degree of freedom. There is a similar set of normal modes for

a three dimensional array of atoms and, once the forces between the atoms are known, the frequency of vibration of each of the modes can be found. This means that so far as thermal vibrations are concerned, we can consider the crystal as equivalent to a set of $3N$ oscillators whose frequencies are those of the normal modes. We must next discuss the specific heat of a single oscillator.

According to classical statistical mechanics, a harmonic oscillator in a temperature bath at absolute temperature T will have an average thermal energy equal to kT, where k is Boltzmann's constant. The value kT is only an average value, we emphasize, and the oscillator will have other energies some of the time, the probability of each energy being given by known equations. The probability is very small, however, that the oscillator acquires more than two or three times kT of thermal energy. In a very large system of oscillators, the fluctuations of energy of the oscillators tend to cancel out and the probability of any appreciable fractional deviation of the total energy from its mean value is very small. If N is the number of molecules in a gram molecule ($N = 6.06 \times 10^{23}$), then $Nk = R$, the gas constant, $= 1.99$ cal. per gm. molecule per degree C. Hence the energy of $3N$ oscillators is $E = 3NkT = 3RT$ and the specific heat is $C = dE/dT = 3R$; this classical result that the specific heat of one gram atom of solid is $3R$ is known as the DuLong-Petit Law.

According to quantum mechanics, an oscillator of frequency ν has a set of quantum states whose energies are $\frac{1}{2}h\nu$, $(1 + \frac{1}{2})h\nu$, $(2 + \frac{1}{2})h\nu$, etc. The oscillator can take on only these energies. If it is in a heat bath of temperature T, however, it will sometimes have one allowed energy and sometimes another and as for the classical case we shall be concerned with its average energy. At absolute zero, the average energy is, of course, $\frac{1}{2}h\nu$. Now the probability of the oscillator gaining much more than kT of thermal energy is very slight. Hence the average energy of the oscillator remains at $\frac{1}{2}h\nu$ until thermal energy becomes large enough to excite it to the next state which is $h\nu$ higher, and consequently so long as kT is much less than $h\nu$ the quantum oscillator acquires much less thermal energy than would a classical oscillator. For kT much greater than $h\nu$, the oscillator will spend an appreciable fraction of its time in many of the quantum states and, as may be shown mathematically, the quantum restriction is no longer of importance so far as the average energy is concerned and the value kT is obtained just as in the classical case. In Fig. 20 the dependence upon temperature of the average energy and the specific heat for a quantum oscillator are shown.

The specific heat of the crystal is just the sum of the specific heats of its oscillators. Since the oscillators have different frequencies they have different specific heats and in order to add up the specific heats of all of them it is necessary to know how the various frequencies of the
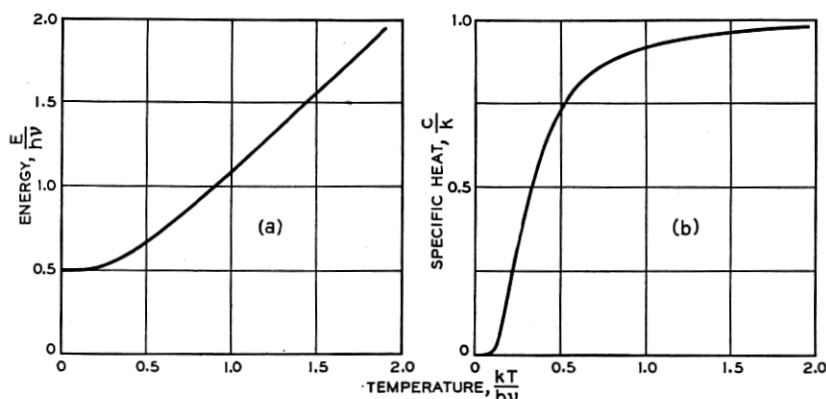

Fig. 20—Thermal behavior of an oscillator according to quantum mechanics.
(a) Energy versus temperature.
(b) Specific heat versus temperature.

oscillators are distributed. Once this distribution in frequencies is known it is merely a matter of summation to find total specific heat. The problem of finding the distribution in frequency of the oscillators was first solved by Debye. The low-frequency vibrations are very simply found for they are merely the acoustic vibrations of the crystal; they are very similar to the normal modes shown for the square membrane of Fig. 7g. For these low-frequency vibrations it can be shown by a straightforward argument, which is too long to give here, that the number dN of oscillators whose frequencies lie between ν and $\nu + d\nu$ is

$$dN = V4\pi \left(\frac{2}{C_T^3} + \frac{1}{C_L^3} \right) \nu^2 d\nu, \tag{7}$$

where C_T and C_L are the velocities of transverse and longitudinal waves in the solid and V is its volume.[26] Debye assumed that this distribution held for all the normal modes. There is of course a highest frequency of vibration, ν_{max}, and the total number of normal modes must be $3N$; hence Debye concluded that

$$3N = \int_0^{\nu_{max}} V4\pi \left\{ \frac{2}{C_T^3} + \frac{1}{C_L^3} \right\} \nu^2 d\nu$$
$$= V4\pi \left\{ \frac{2}{C_T^3} + \frac{1}{C_L^3} \right\} \frac{\nu^3_{max}}{3}. \tag{8}$$

[26] For a derivation see P. Debye, *Ann. d. Physik, 39*, 789 (1912).

From this equation ν_{max} can be found if N/V and the velocities C_T and C_L are known. Knowing ν_{max} and the distribution in frequency, Debye summed the specific heats of all the oscillators and obtained the specific heat of the solid. According to this theory the specific heat vanishes at $T = 0$ and is proportional to T^3 near $T \doteq 0$. At high temperatures it approaches the classical value of $3R$. A measure of the temperature at which the classical value is closely approached is given by the maximum frequency of atomic vibration ν_{max}; when kT is greater than $h\nu_{max}$, all the modes of vibration including the highest make substantial contributions to the specific heat. The temperature at which this occurs is known as the Debye temperature and denoted by the symbol θ_D; obviously $\theta_D = h\nu_{max}/k$. The specific heat given by Debye's equation is a function of T/θ_D only and can thus be represented by the expression $C(T/\theta_D)$; so that by this theory all crystals should have the same curve for specific heat versus temperature except for changes in the temperature scale corresponding to the different values of their Debye temperatures.

TABLE III

DEBYE TEMPERATURES IN DEGREES KELVIN USED IN FIGURE 21

Pb	Tl	Hg	J	Cd	Na	KBr
88	96	97	106	168	172	177
Ag	Ca	KCl	Zn	NaCl	Cu	Al
215	226	230	235	281	315	398
Fe	CaF$_2$	FeS$_2$	C			
453	474	645	1860			

In Fig. 21 is shown a compilation of specific heat data.[27] For each substance a value of θ_D (given in Table III) has been chosen so as to obtain the best agreement with experiment and the values of the specific heat have then been plotted as a function of T/θ_D. The Debye theory relates to specific heat at constant volume and in it no allowance is made for the energy due to thermal expansion. The experimental points are derived from measurements of specific heat at constant pressure which have been transformed by using a thermodynamical relationship so as to give specific heat at constant volume.

For these curves θ_D was chosen so as to obtain the best fit. It is, however, possible to calculate θ_D from theory by using the elastic constants of the material to evaluate C_T and C_L and then substituting in Eq. (8). For sodium the elastic constants have been calculated entirely from theory by the methods described in the section on

[27] Taken from E. Schroedinger, *Handbuch der Physik*, Vol. X, p. 307 (1926).
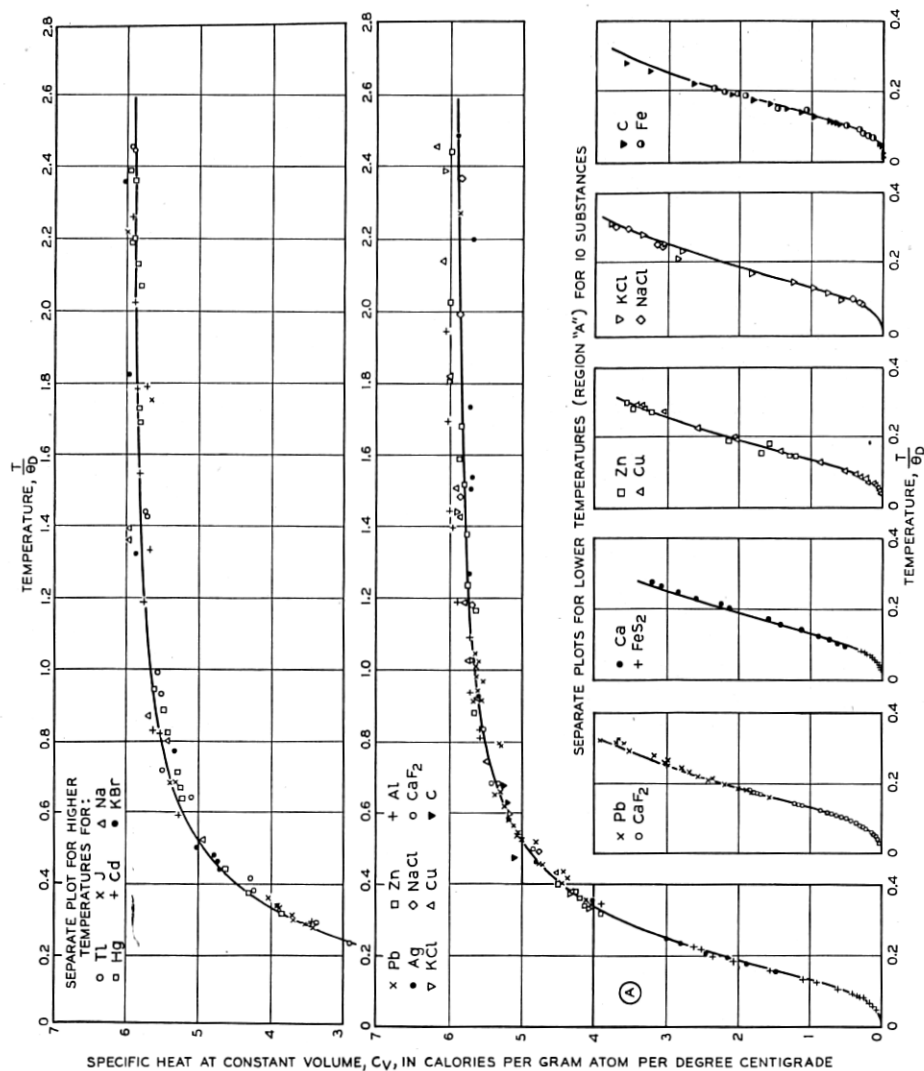
Fig. 21—Comparison of the Debye specific heat curve with experiment.

"Electrons in Crystals" and extensions of them * to be discussed in the third paper of the series. Using the theoretical values one obtains a value of 143° K for θ_D, whereas the value that fits experiment best is 172° K.

Recently calculations have been made from a model of the crystal as an assemblage of atoms rather than as a continuum as postulated in deriving Eq. (7)—that is, a model like the coupled oscillators, rather than like the stretched membrane, is used. These calculations, principally by Blackman, have explained some discrepancies between the Debye theory and experiment.

The Specific Heat of the Electrons

We must now see why the electrons contribute only slightly to the specific heat. Let us consider a case like that of sodium where we have a partially filled band. At the absolute zero of temperature, the electrons will fill all the levels below a certain energy E_1 and all the higher levels in the band will be empty (Fig. 22a). Now at temperature T some of the electrons will be excited to higher states; since, however, an electron cannot gain more than about kT of energy thermally, only those electrons whose energies lie in a range kT below E_1 can be excited. Electrons occupying states farther down in the band cannot acquire kT of thermal energy for, if they did so, they would have to move to states already occupied and such an act is forbidden by Pauli's principle. In order to demonstrate what a small fraction of the electrons can gain energy thermally, we point out that the width of the energy band is usually 4 or 5 ev while the value of kT in electron volts is $T/11,600$ and room temperature corresponds to a kT of about .03 ev. The electrons which do gain thermal energy have a normal value for the specific heat but constitute only about one per cent of all the valence electrons.

It might be maintained that the above argument is specious and that the electrons could all gain energy kT; this would not violate Pauli's principle because the electrons would move upward in the band as a unit, each moving into a state vacated by another electron. This contention is found to be wrong; one finds by using the statistical mechanics appropriate to electrons that the distribution of the electrons among the energy levels is given by the Fermi-Dirac distribution function.[28] According to this, the distribution of the electrons among the levels would be as indicated in Fig. 22b. The proba-

* K. Fuchs, *Proc. Roy. Soc.* **157**, 444 (1936).

[28] For a discussion of the Fermi-Dirac statistics see K. K. Darrow, *The Bell System Technical Journal*, Vol. VIII, p. 672, 1929, or *The Physical Review Supplement*, Vol. I, p. 90, 1929.
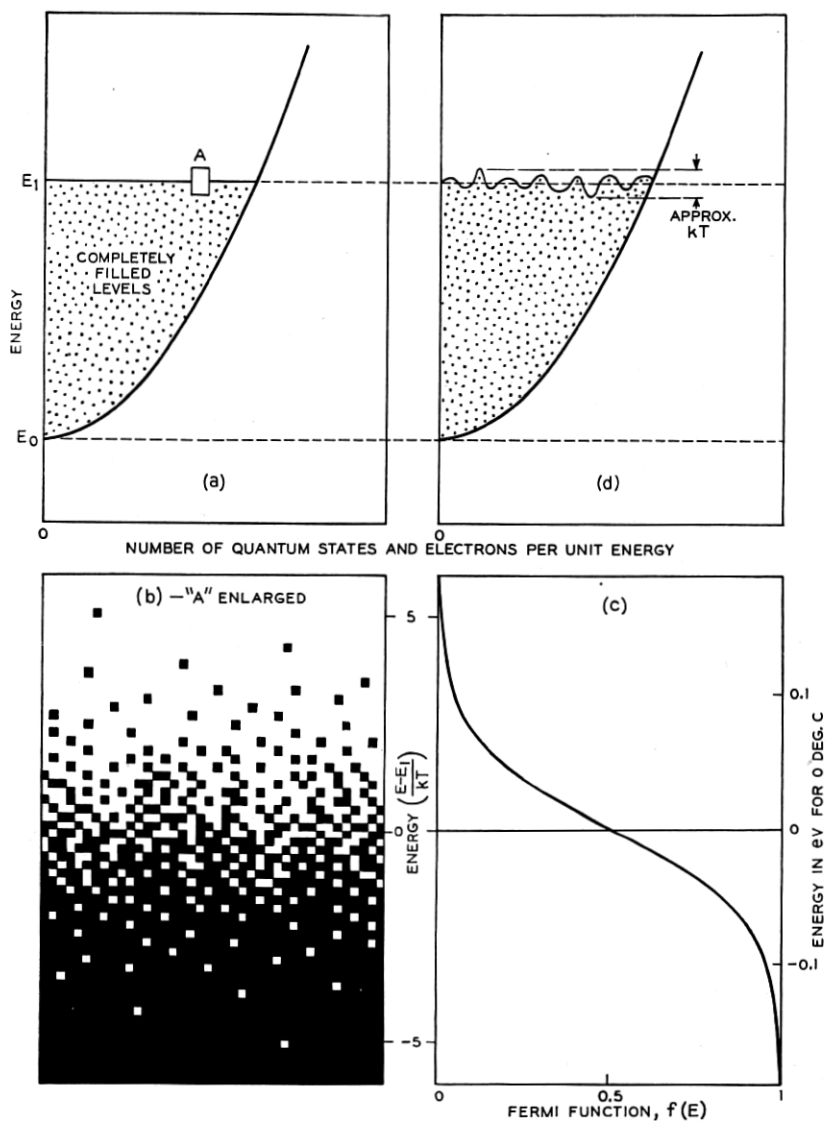
Fig. 22—Specific heat of the electrons.

(a) Distribution of electrons in energy for the absolute zero of temperature.
(b) Enlargement of part of (a) but for room temperature; each unit of area represents a quantum state.
(c) The Fermi distribution function.
(d) A water tank analogue.

bility that any particular energy state be occupied is given by the Fermi-Dirac factor f

$$f = \frac{1}{e^{\frac{E - E_1}{kT}} - 1}. \tag{9}$$

This factor is shown in Fig. 22c and the corresponding filling of energy levels is shown schematically in 22b. A physical picture which is helpful in understanding this result may be obtained by considering the distribution of energy levels, Fig. 22a, to be the cross-section of a trough or tank. If we pour water into this tank it will fill to a certain level, E_1. If we let each molecule of water in the tank represent an electron in the crystal, then the distribution in energy of the electrons is correctly represented by the distribution in height of the molecules. Thermal agitation is represented by shaking the tank; this will produce surface ripples as in Fig. 22d which represent crudely the Fermi-Dirac distribution.
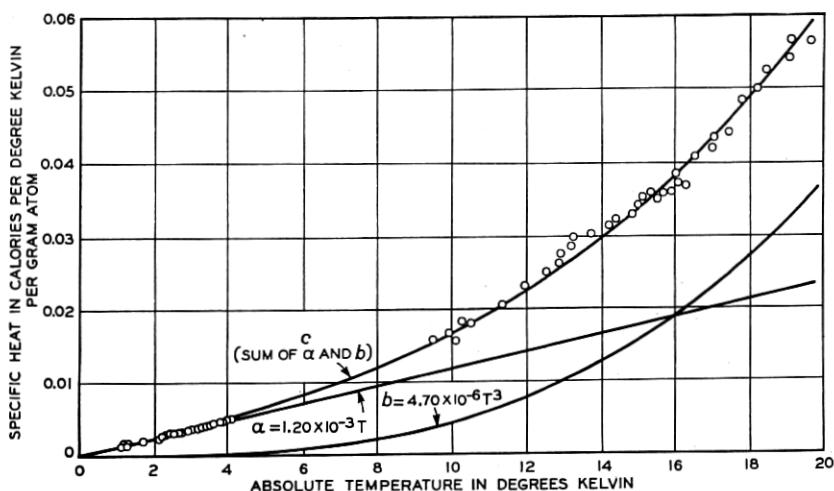
Fig. 23—Specific heat of iron at low temperature.

Under certain conditions, however, the electronic specific heat is not negligible. We have seen that the number of electrons participating in specific heat is proportional to kT and that these have a more or less normal specific heat. Hence the electronic specific heat is proportional to T. On the other hand, at low temperatures the Debye specific heat is proportional to T^3. Hence for sufficiently low temperatures the electronic specific heat is the larger. In Fig. 23 we give the specific

heat of iron near absolute zero.[29] The theoretical curve *c*, which is seen to represent the experimental data quite well, is the sum of two terms represented by curves *a* and *b*. *a* is linear in the temperature and represents the electronic specific heat while *b* is cubic and represents that due to lattice · vibrations. Numerical calculations from theory of the slope of curve *a* which could be compared with the observed slope are not available. Curve *b*, we have said, is just the Debye curve and is drawn as if the Debye temperature were 462°, a value which is in good agreement with 453°, the value deduced from the specific heat at higher temperatures in connection with Fig. 21. At very high temperatures the electronic specific heat will again be of importance. But at high temperatures it is necessary to apply corrections to the Debye theory and the writer is not acquainted with any unambiguous evidence for electronic specific heat in that case.

Thus we see that only a very small fraction of the electrons of a partially filled band contribute to the specific heat. It is the Pauli principle which restrains the remainder. We shall see in the next paper why the Pauli principle does not interfere with the conduction of electricity. For the case of an insulator—that is, a crystal each of whose bands is either wholly filled or wholly empty—it is still harder for electrons to arrive at empty states and the electronic specific heat is quite negligible. Hence all of the specific heat for an insulator is of the atomic vibration type discussed in the Debye theory.

The Theory of Thermal Expansion

In order to understand the theory of thermal expansion we must study the curve representing energy versus lattice constant for the solid. This is shown qualitatively in Fig. 24. We note that the energy curve is unsymmetrical about its minimum. We may describe its behavior by saying that it is harder to compress than to expand the solid. This statement is illustrated by a comparison of the expansion and the compression which can be produced by a given energy *E*; it is seen that the asymmetry of the curve causes the expansion produced by this energy to be greater than the compression. Now the origin of thermal expansion is as follows: owing to thermal agitation—that is, atomic vibrations—regions of the crystal are alternately expanding and contracting; since the expansions occur more readily than the contractions, there is on the average a net expansion. The greater the temperature the greater this net expansion; hence we find that the size of the solid increases with increasing temperature. This explanation of thermal expansion can be made clearer by considering, not a solid,

[29] W. H. Keesom and B. Kurrelmyer, *Physica* 6, 633 (1939).

but a diatomic molecule. Suppose Fig. 24 gives the dependence of
the energy of a molecule upon the internuclear distance. Suppose the
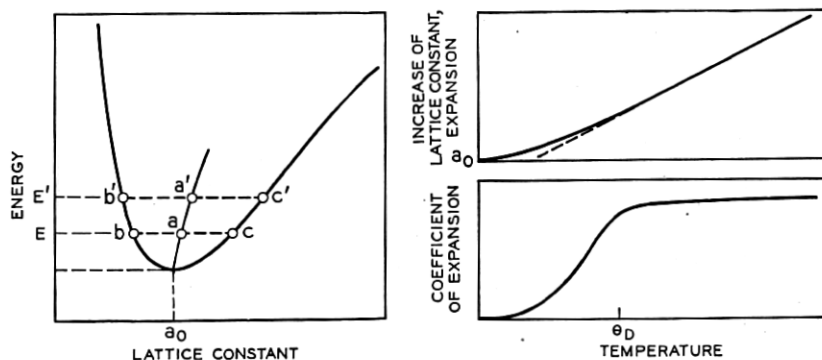molecule is given vibrational energy corresponding to E on the figure.

Fig. 24—The theory of thermal expansion. The asymmetry of the curve for
the energy of a crystal versus the lattice constant is responsible for the thermal
expansion.

Then the nuclei will vibrate between positions b and c on the figure.
Since c lies more to the right of the equilibrium position than b does to
the left, the mean distance of separation, a, lies to the right of a_0.
Increasing the vibrational energy to E' increases the mean separation
to a'. This shows that the asymmetry of the potential curve results in
a continuous increase in mean internuclear separation with increasing
energy of vibration. A crystal is, in a sense, an assemblage of diatomic
molecules, each pair of nearest neighbors having a potential energy
curve like that of Fig. 24, and its expansion is explained in the same
way.

The theory outlined above can be made quantitative. From it we
obtain the interesting result that the thermal expansion coefficient is
proportional to the specific heat. This is a rather natural result:
we have seen that the total expansion is proportional to the thermal
energy; hence the rate of expansion with increasing temperature, i.e.
the thermal expansion coefficient, should be proportional to the rate of
increase in thermal energy with increasing temperature, i.e. to the
specific heat. The relationship embodying this statement is known
as Grüneisen's law and is expressed by the equation

$$\alpha = \gamma \frac{K}{V} C_V, \qquad (10)$$

where α is the volume coefficient of thermal expansion (three times the
linear coefficient), K is the compressibility, V the volume of one gram

atom, and C_V the specific heat per gram atom at constant volume. γ is a parameter which measures the asymmetry of the curve and is defined as follows: if we think of the solid as being compressed by an external pressure, the forces between the atoms will change and the Debye temperature will increase. If the curve were a parabola—that is, perfectly symmetrical—the Debye temperature would not change. γ is defined by the relationship

$$\gamma = -\frac{\partial \ln \theta_D}{\partial \ln V}. \tag{11}$$

The γ, K, and V are nearly constant for a given substance. Hence the thermal expansion curve is practically the same as the specific heat curve except for a constant factor. In Fig. 25 we give the thermal
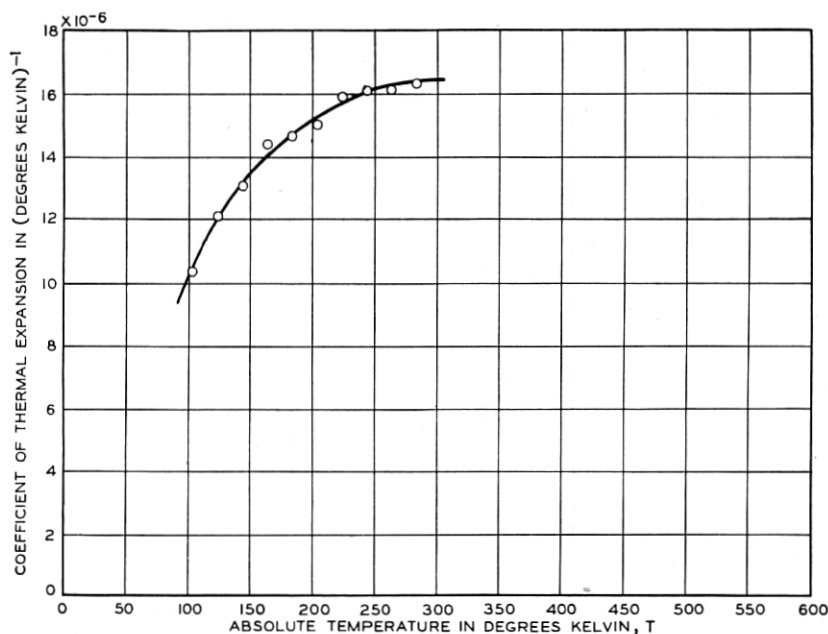
Fig. 25—Coefficient of thermal expansion versus temperature for copper.

expansion of copper.[30] The Debye temperature was chosen to give the best fit. We see that the theory of thermal expansion gives as good agreement with experiment as does the theory of specific heat. If Grüneisen's law were perfectly satisfied, the same Debye temperature would be found for both the thermal expansion and the specific heat curves. The relatively small difference between the two values, 325 for expansion and 315 for specific heat, is a measure of the validity of Grüneisen's law.

[30] E. Grüneisen, *Handbuch der Physik*, X, p. 43 (1926).

Grüneisen's law applies only to simple crystals; we shall see in the next section that it is not applicable to the anomalous expansion associated with ferromagnetic transformations nor is it applicable to the abnormal expansions of the order-disorder transformations in alloys.

Magnetic Effects

In this section we return to a discussion of the energy band theory and this time introduce the magnetic moment associated with the spin of the electron. It is the spin magnetic moment which when added to the concept of energy bands leads to explanations of para and ferromagnetism.

When a body is placed in a magnetic field it becomes magnetized; in other words it acquires a magnetic moment. Ferromagnetic materials become very easily magnetized in the field with their magnetic moments parallel to the field and they may remain magnetized after the field is removed. Paramagnetic materials are also magnetized in the direction of the field but only very weakly compared to ferromagnetic materials and only while they remain in the field. Diamagnetic materials are magnetized in a direction opposite to the field and, like paramagnetic substances, only weakly and while in the field. These magnetic effects are produced by the electrons in two distinct ways. In the first place, the motion of the electron as a whole produces a current and this current, like the ordinary macroscopic currents in a wire, produces a magnetic field. Conversely, an externally applied magnetic field affects the motions of the electrons in a body and can thereby magnetize it; this process accounts for the diamagnetism of diamagnetic bodies but it may contribute to the paramagnetism as well. It is not with this first way in which electrons can behave magnetically but rather with the second way, described below, that we shall be concerned. The first way, which is mentioned for completeness, involves a theory too complicated for treatment in this article. In the second place, an electron can behave magnetically by virtue of its spin: the rotation of the electron about its own axis produces a magnetic moment which is anti-parallel—because the charge of the electron is negative—to the angular momentum due to the spin. A magnetic field tends to align the spin magnetic moments of the electrons and to make them contribute to the paramagnetism. We shall see below that this process accounts for the paramagnetism of non-ferromagnetic metals. We shall see also that the magnetism of ferromagnetic bodies is due to the magnetic moment of the electron spin but that the energy involved in the theory of ferromagnetism is not an interaction between the magnetic dipoles of the electrons but is

instead an electrostatic exchange energy like that discussed for atoms in connection with Figs. 4 and 6.

Paramagnetism and Diamagnetism

Let us consider first the so-called "weak spin paramagnetism." This occurs in metals, since they have partially filled bands. In the presence of a magnetic field the spin of the electron is quantized so that the component of its angular momentum in the field direction is either $+\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$ where $\hbar = h/2\pi$ (h = Planck's constant) is the quantum mechanical unit of angular momentum. The corresponding components of magnetic moment along the field are $-\mu_\beta$ and $+\mu_\beta$ where μ_β is the quantum mechanical unit of magnetic moment known as the Bohr magneton. Letting $-e$ be the charge and m the mass of the electron and c be the speed of light, we have from the quantum theory

$$\mu_\beta = e\hbar/2mc. \tag{12}$$

The ratio of mechanical moment (i.e. angular momentum) to magnetic moment, taken without regard to sign, is called the "gyro-magnetic ratio." For the spin of the electron its value is mc/e, but for the motion of the electron as a whole, its value is $2mc/e$. Because of the difference between these two values, experimental measurements of the gyro-magnetic effect play a decisive role in the experimental verification of the electron spin theory of ferromagnetism in a way which we shall describe below.

Half the quantum states in an energy band of a crystal have angular momentum components along the magnetic field of $\frac{1}{2}\hbar$ and the other half of $-\frac{1}{2}\hbar$. In Fig. 26a, we have divided the states in the band into two groups, corresponding to the two spins. We shall refer to one of these as "the band with plus spin" and to the other as "the band with minus spin." When a magnetic field is applied, the energies of the electrons are changed. Thus if an electron in the lowest state of the band with minus spin has an energy E_0 before the field is applied, it has an energy of $E_0 - \mu_\beta H$ afterwards; the second term represents, of course, the energy of the magnetic dipole μ_β when parallel, as distinguished from anti-parallel, to the field—the situation for minus spin. All the states in the band with minus spin will be thus altered in energy. Similarly all the states in the band with plus spin are displaced upwards in energy by $\mu_\beta H$. This is the situation represented in Fig. 26b. After the displacement we find that some of the electrons in the band with plus spin have higher energies than empty states in the band with minus spin; such an arrangement is not stable and the electrons will change their quantum states so as to produce the lowest energy possible consistent with the distribution of energy levels shown

in Fig. 26*b* and with Pauli's principle. The arrangement of lowest energy is shown in Fig. 26*c*; electrons have shifted from the band of plus spin to states of lower total energy in the band of minus spin until the two bands are filled to the same energy level, indicated by the solid horizontal line. As the figure shows, the number of electrons shifted will be the number lying in the energy range $\delta E = \mu_\beta H$.[31]
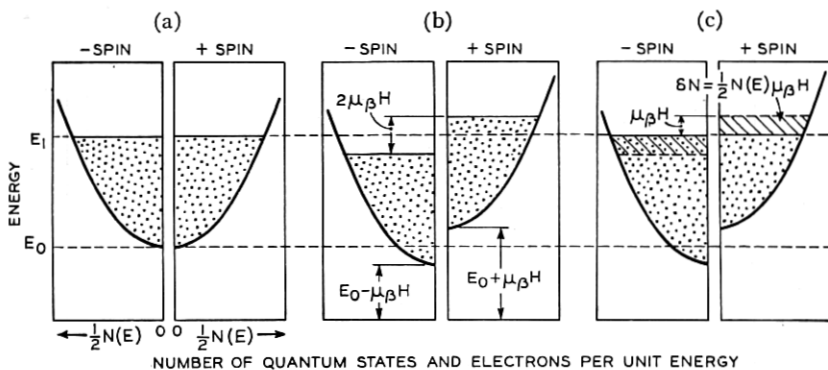

Fig. 26—The paramagnetism of free electrons.
(a) Distribution of electrons in energy.
(b) Displacement of levels by a magnetic field.
(c) Distribution of electrons in energy in a magnetic field.

The number of states, δN, lying in this energy range in the band of plus spin, which contains of course half the states in the band, is according to equation (3)

$$\delta N = \tfrac{1}{2} N(E_1) \delta E = \tfrac{1}{2} N(E_1) \mu_\beta H. \tag{13}$$

The magnetic moment of these states is

$$\delta M_+ = - \mu_\beta \delta N = - \tfrac{1}{2} N(E_1) \mu_\beta^2 H. \tag{14}$$

The minus sign occurs because the angular momentum and the magnetic moment of an electron are in opposite directions; the states of plus spin have minus moments in Fig. 26.

The electrons that occupied these states before the field was applied now occupy states with minus spin and produce a magnetic moment of

$$\delta M_- = \tfrac{1}{2} N(E_1) \mu_\beta^2 H. \tag{15}$$

Hence the minus band gains a plus moment and the plus band loses a

[31] We have here assumed that the fractional change in $N(E)$ in the interval $\mu_\beta H$ is negligible; this assumption is reasonable. For a field of 10,000 gauss, $\mu_\beta H$ is only 5.77×10^{-5} ev while $E_1 - E_0$ is of the order of several ev.

minus moment and, since the net moment of Fig. 26a is obviously zero, the net moment produced by the magnetic field is

$$\delta M = \delta M_- - \delta M_+ = N(E_1)\mu_\beta^2 H. \tag{16}$$

The susceptibility of a material, denoted by χ, is defined as the magnetic moment produced per unit volume per unit field:

$$\chi_s = \frac{\delta M}{VH} = \frac{N(E_1)}{V}\mu_\beta^2. \tag{17}$$

The subscript "s" is a reminder that this susceptibility was produced by the spin magnetic moment of the electron.

Since the moment produced is in the direction of the field, χ_s is positive; the susceptibility is of the paramagnetic type. As for its magnitude: in the monovalent metals, as we have said before, the distribution of levels in the bands is well approximated by the free electron formula (4). Using this, we find

$$\chi_s = \frac{4\pi}{h^3}(2m)^{3/2}(E_{\max})^{1/2}\mu_\beta^2. \tag{18}$$

where $E_{\max}$ ($= E_1 - E_0$) is the maximum kinetic energy in the band.

Before comparing susceptibilities calculated from this expression with experimental values, we must discuss diamagnetism. The electrons in the partially filled band of Fig. 26 give formula (18) because of their spin magnetic moments. They give a susceptibility also because of their motion through the crystal. For the case of free electrons, this susceptibility is negative—that is, it is a diamagnetic susceptibility, and, according to a theory we cannot discuss here, in magnitude it is one third of χ_s. Denoting it by χ_m ("m" for motion of the electron as a whole), we have

$$\chi_m = -(1/3)\chi_s. \tag{19}$$

The electrons in the filled bands, corresponding to electrons in closed shells in the ionic cores of the metal, also give rise to diamagnetism. They can give no spin paramagnetism because there is no possibility of transferring electrons from a *filled* band of one spin to a *filled* band of the other spin—this would require putting more electrons in the band of one spin than it has quantum states, a violation of Pauli's principle. Denoting by χ_i the susceptibility of the ionic cores of the metal, we have for the net susceptibility χ the equation

$$\chi = \chi_s + \chi_m + \chi_i. \tag{20}$$

Specializing this for the case of free electrons in the valence electron

band gives

$$\chi = (2/3)\chi_s + \chi_i = \frac{8\pi}{3}\left(\frac{2m}{h^2}\right)^{3/2}\mu_\beta^2(E_{max})^{1/2} + \chi_i. \qquad (21)$$

In Table IV we give theoretical and experimental values for the susceptibilities of the simple metals. The values of χ_i are obtained from theory for lithium and by experiment for the other metals.

TABLE IV
MAGNETIC SUSCEPTIBILITIES *

	Li	Na	K	Rb	Cs
χ_s	1.5	0.68	0.60	0.32	0.24
χ_i	−0.1	−0.26	−0.34	−0.33	−0.29
$\chi = \frac{2}{3}\chi_s + \chi_i$	0.9	0.2	0.06	−0.12	−0.15
χ observed †	0.5	0.51	0.40	0.07	−0.10

* This Table is taken from N. F. Mott and H. Jones, "The Theory of the Properties of Metals and Alloys," Oxford 1936, p. 188.
† K. Honda, Ann. d. Physik *32*, 1027 (1910) and M. Owen, Ann. d. Physik *37*, 657 (1912).

Although equation (17) for the spin susceptibility χ_s in terms of $N(E_1)$ is generally true, the relationship that $\chi_m = -\chi_s/3$ is true only for the case when $N(E)$ is the free electron distribution.[32] For some metals $N(E)$ differs greatly from that for free electrons and then larger values of χ_m may occur. The high diamagnetism of bismuth is explained in this way. In the next paper, we shall discuss the meaning of the freeness of electrons; however, a discussion of electron diamagnetism lies beyond the scope of this paper.[33]

Ferromagnetism

The shift of electrons from one band to another for the paramagnetic behavior shown in Fig. 26 persists only so long as the magnetic field is applied. When the magnetic field is removed, the stable arrangement is as shown in Fig. 26a, equal numbers of electrons having each spin. The situation is quite different in ferromagnetic materials and, for reasons discussed below, in the stable arrangement there are many more electrons of one spin than of the other.

Two things are important for the occurrence of ferromagnetism: the exchange effect as illustrated in Figs. 4 and 6 and the structure of the bands arising from the $3d$ levels. The $3d$ levels, as is shown in

[32] An even more stringent condition is actually required.
[33] For the diamagnetism of electrons in closed shells the reader is referred to K. K. Darrow's article, "The Theory of Magnetism," *Bell System Technical Journal*, Vol. XV, 1936, and in particular to Page 247.

Fig. 6, are only partially filled for free atoms of the ferromagnetic elements iron, cobalt, and nickel. We shall show first how the partially filled 3d bands together with exchange forces can produce ferromagnetism and later discuss the theory of why only the last three of the eight transition elements are ferromagnetic.

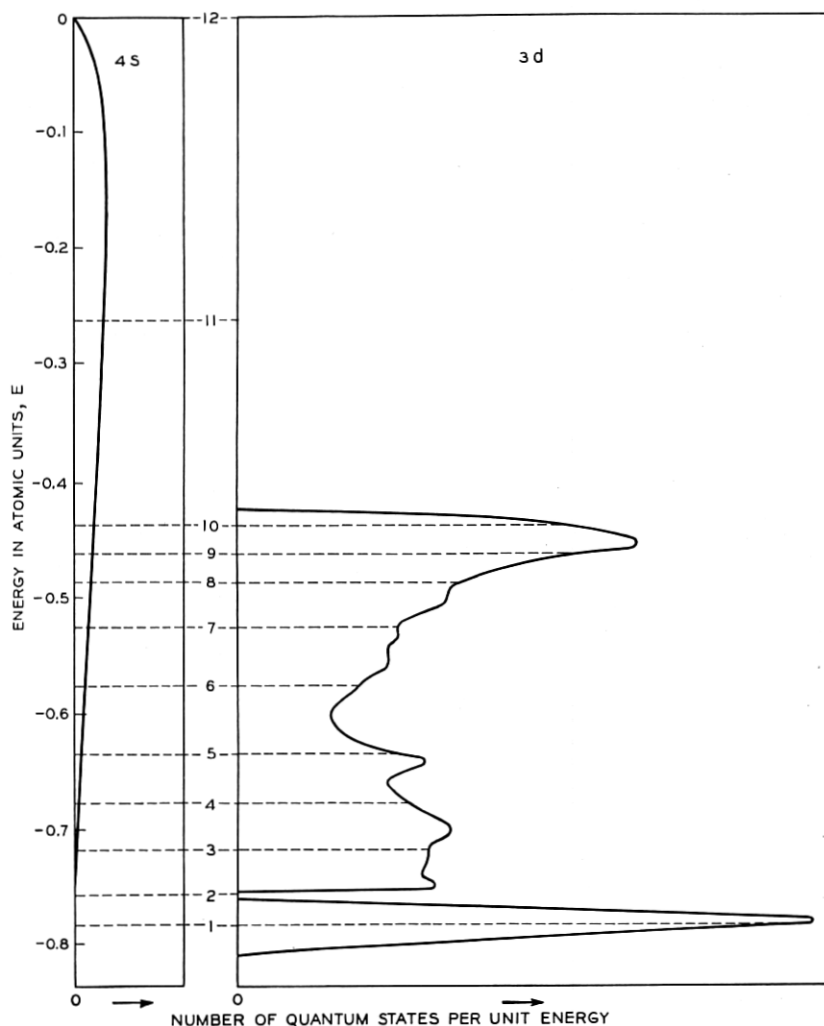The splitting of the atomic energy levels into bands is shown in

Fig. 27—Distribution of states in energy for copper. The distribution is probably quite similar for iron, cobalt, and nickel and in the absence of calculations for these other metals, this figure will be used for them. The total number of quantum states per atom in the 4s and 3d bands having energies less than the ordinates of the dashed lines are given by the corresponding integers.

Fig. 16. The 3d levels give a band capable of containing ten electrons per atom, five with each spin; and the 4s band can hold two electrons per atom, one with each spin. Curves representing $N(E)$ for these bands, calculated for the case of copper by Slater and Krutter, are shown in Fig. 27. We see that the 4s band is much wider in energy than the 3d and that it contains only one-fifth as many electronic states. The band structure will be similar for all the transition elements; the energy scales, however, will be different. As is shown in Fig. 6 the 3d electrons are more tightly bound for copper than for nickel or chromium. Corresponding to this tighter binding, the 3d wave functions of copper extend less in space than those of nickel and chromium and consequently they overlap less between atoms and the 3d band is narrower for copper. Progressing towards decreasing atomic number in the sequence of elements from copper to scandium, the 3d band will continually widen; and this widening, as we shall see later, can help account for the absence of ferromagnetism for the elements before iron in the periodic table.
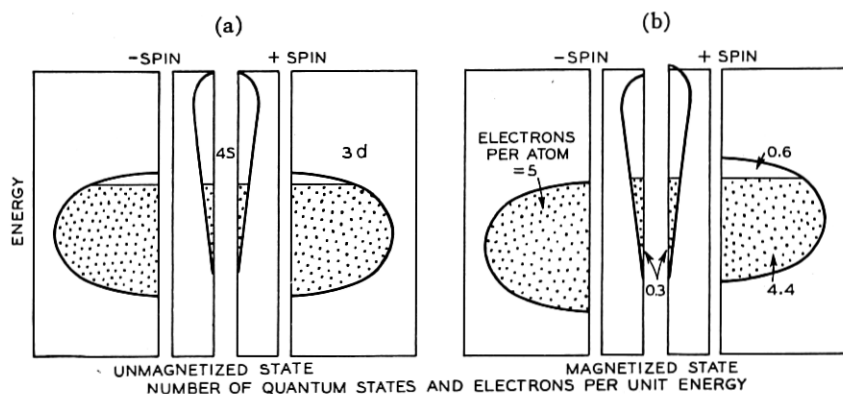

Fig. 28—The ferromagnetism of nickel.

In Fig. 28, we give a simplified representation of the 4s and 3d bands split into two sets according to the spin. (We may, if we wish, suppose that a magnetic field is applied along which the spin is quantized, but that the field is so weak that the displacement of the energy bands produced by it is negligible; this supposition is not necessary, however, for regarding the spin we shall need only the fact that all the electrons in the + spin band have parallel spins which are anti-parallel to those in the ÷ spin band.) For the element nickel there are 28 electrons, 10 of which are in the 3d and 4s bands. They can fill the bands as indicated in Fig. 28a. Let us compare this distribution with the electron

configuration of the atom, Fig. 6; we see there that there are unequal numbers of electrons of the two spins. This inequality is produced by the exchange effect which lowers the more occupied set of 3*d* levels in respect to the less occupied set and produces a stable arrangement with the 3*d* levels of one set completely filled. This exchange effect operates in the same way in metallic nickel. In Fig. 28*b* we show the distribution which results when electrons are shifted from the 3*d* band of plus spin to that of minus spin until the latter is filled. The exchange effect produces the displacements of the bands as shown. The arrangement in Fig. 28*b* is stable; in order for electrons to be transferred from the filled minus 3*d* band to the plus band, they would have to increase their energy, a fact which is expressed by drawing the diagram so that the lowest vacant quantum states are appreciably above the highest energy state of the filled 3*d* band. Thus for nickel an unbalanced distribution of spins prevails both for the free atom and the metal.

The Energy of Magnetization

The argument presented above for the stability of the magnetized state shown in Fig. 28*b* is not really rigorous. We saw that if one electron was transferred from the filled 3*d* band to one of the vacant states, its energy and, therefore, the energy of the crystal would be raised. In other words, the magnetized state has less energy than a state which is slightly less magnetized. This fact in itself does not prove that the magnetized state is stable; it proves only that it is metastable—i.e., that its energy is less than the energy of other states which differ from it slightly; in order to establish the stability of the magnetized state, it is necessary to prove that its energy is less than the energy of any other state including that of the unmagnetized state shown in Fig. 28*a*. We may illustrate this necessity by considering the following hypothetical behavior: as the magnetization is reduced from that of Fig. 28*b* to zero (the value for Fig. 28*a*), the energy might at first increase and then decrease—decreasing so much finally that the energy would be lower for the unmagnetized than for the fully magnetized state. We shall, therefore, discuss the difference in energy between the fully magnetized and unmagnetized states; theory shows that this quantity is the fundamental one whose value determines whether or not ferromagnetism occurs.

Let us consider the change in energy in going from the unmagnetized state to the magnetized state in Fig. 28. This change in energy can be separated into two contributions, one positive and one negative. The positive contribution comes from an increase in "Fermi energy" or

"energy of motion," which was discussed in connection with the binding energy of metals. This energy is positive because after the shift to the magnetized state, electrons have moved from states in the band of plus spin to states which lie higher—in respect to the bottom of the bands in both cases—in the band of the minus spin; that is, the electrons which have moved from one band to the other have all gained "energy of motion." The negative contribution to the energy comes from the exchange effect. This causes the lowering of the filled band and the raising of the unfilled band; since there are more electrons in the lowered band than in the raised band, there is a net decrease in energy due to this exchange effect. Thus we have a positive change in Fermi energy and a negative change in exchange energy in going from the unmagnetized to the magnetized state. If the exchange energy has a greater change than the Fermi energy, the energy of the magnetized state is lower and the metal is ferromagnetic.

No satisfactory calculations have as yet been made for these energy differences. In order to calculate them, accurate values for the distribution of states in the $3d$ band are needed, and the mathematical methods available for computing this distribution are not as yet very satisfactory. Next the exchange effect energy must be found; this is also difficult to calculate accurately. Finally, the description given here is over-simplified; in particular another energy term, known as the correlation energy, must be included; this energy acts somewhat like an exchange energy but between the bands of different spins and it tends to cancel out the exchange energy. Although these difficulties greatly mar the usefulness of the theory of ferromagnetism represented in Fig. 28, this theory is able to correlate a large amount of experimental material in a very natural way; and since it is the theory based on the concepts of energy bands, it is the one that we shall discuss in this paper. In passing, however, we must state that there are other theories of ferromagnetism which in some ways are more successful and in other ways less successful than the band theory. Some of these are atomic rather than band theories. An example of this type of difference in method of attack was given in the discussion of the binding energy of sodium chloride; two treatments were given: for one the basis being the ions and for the other the energy bands. In the case of sodium chloride, however, the theoretical equivalence of the two methods is easily demonstrated. In the case of ferromagnetism, the two theories are not equivalent and are both simplifications of a more complex and as yet unsatisfactorily explored intermediate case.

Although no satisfactory calculations of the energy difference between the magnetized and unmagnetized states of metals exist, the

theory must be regarded as representing great progress over non-wave-mechanical theories. The reason is this: in older theories of ferro-magnetism the energy was supposed to come from the magnetic inter-action between the magnetic dipoles, and it turned out that the ener-gies calculated in this way were at least a thousandfold too small. The energies calculated in the new theory are adequate in magnitude but have nothing to do with the magnetic moment of the electron; they arise from the exchange energy, which is, as we have said before, an electrostatic energy resulting from the wave-mechanical treatment of Pauli's principle. It is the laws governing the spin quantum num-ber of the electron, not the magnetic moment, which are responsible for the energy of magnetization; the externally observed magnetic field of a ferromagnetic material is merely a superficial indication of more fundamental electrostatic forces.

Intrinsic Magnetization

According to our theory, the low energy state and therefore the stable state of metallic nickel is a magnetized one. If one picks up a piece of nickel at random, however, it may not appear to be magnetized. This apparent absence of magnetism is due to the presence of "do-mains." According to the domain theory—which is a very well estab-lished branch of magnetic theory—a block of nickel will consist of a number of microscopic domains, each highly magnetized, but having their magnetic moments pointing at random in a number of directions so that on the average there is no magnetism. The application of a magnetic field aligns the magnetic moments of these domains and, since they are then all parallel, one can measure the total magnetization of the sample. A field strong enough to line up all the domains is said to produce "saturation" because a further increase in field will give no further increase in magnetization. It is customary and convenient to divide the total or saturation magnetic moment of the material by the total number of atoms, thus finding the average magnetic moment per atom, and to express this value in Bohr magnetons. The resultant value is called the intrinsic magnetization per atom and is denoted by β.[34] For example, if a crystal had one electron per atom and all the electrons had their spins parallel, then all their magnetic moments would be parallel, too, and the intrinsic magnetization would be unity, $\beta = 1$.

For nickel the intrinsic magnetization is 0.6 Bohr magnetons per atom. The following argument shows how easily such a fractional number can be accounted for by our theory. Nickel has 10 elec-

[34] The "intrinsic magnetization" is customarily defined as the magnetic moment per unit volume when the moments of the domains are parallel.

trons per atom in the 3d and 4s bands. The 3d band with minus spin is supposed full, containing five electrons per atom. The 4s band (both spins) can contain two electrons per atom, and from Fig. 27 we see that it is about one-fourth full; suppose it contains 0.6 electrons per atom; the remaining electrons go to the 3d band with plus spin which is not quite full but has a "hole" in it of 0.6 electrons per atom. There are equal numbers of electrons of each spin in the 4s band and their magnetic moments cancel.[35] The net magnetic moment arises from the unbalance of 0.6 electrons per atom between the two parts of the 3d band. This unbalance will correspond to a magnetization of 0.6 Bohr magnetons. The theory is not capable of predicting the number 0.6 exactly; however, this number is entirely consistent with what can be said about the distribution of levels in the band. In the "atomic" theories of magnetism, it is supposed that each atom has a certain magnetic moment. From the results of the gyromagnetic experiments,[36] one concludes that the magnetization is due to electron spin. Since an atom whose magnetism is due to electron spin must have a magnetic moment equal to an integral multiple of the Bohr magneton,[37] the "atomic" theory is forced to assume that 40 per cent of the nickel atoms are unmagnetized and that 60 per cent have one Bohr magneton, or else that 70 per cent are unmagnetized and 30 per cent have two Bohr magnetons or at any rate that there are at least two kinds of atoms. These rather awkward assumptions are not required in the band theory, the reason being, as is suggested in Fig. 28, that the electrons are not thought of as belonging to the atoms individually but to the crystal as a whole.

The intrinsic magnetization of ferromagnetic material decreases with increasing temperature. In the band theory this is explained as follows: at a temperature T some of the electrons are excited from the filled band to the partially filled band; as the temperature is increased more are shifted. Furthermore, if we compare the effects of two equal increments of temperature, one occurring at a higher temperature than the other, the one at the higher temperature will have the greater effect. This is because at the higher temperature more electrons have been shifted; hence the exchange effect displacement of the band of one spin in respect to the band of the other spin is less and electrons need

[35] Actually there will be a slight exchange effect in the 4s band; however, it will be so slight that the magnetic moment produced can be neglected

[36] If a piece of iron is suspended so that it can rotate and then is magnetized, it will acquire an angular momentum. The ratio of angular momentum to magnetic moment should be mc/e if the magnetization arises from electron spin and $2mc/e$ if it arises from motion of the electron as a whole. Experiment gives the following fractions of the former value: for iron 1.03, for cobalt 1.23, for nickel 1.05.

[37] For a discussion of this theorem see K. K. Darrow's article, "Spinning Atoms and Spinning Electrons," *Bell Sys. Tech. Jour.*, XVI, 319 (1937).

not gain so much energy to shift from the more full to the less full band. Hence at the higher temperature there is more decrease in magnetization per degree rise in temperature than at the lower temperature. A logical consequence of this reasoning is that the magnetization decreases more and more rapidly as the temperature increases
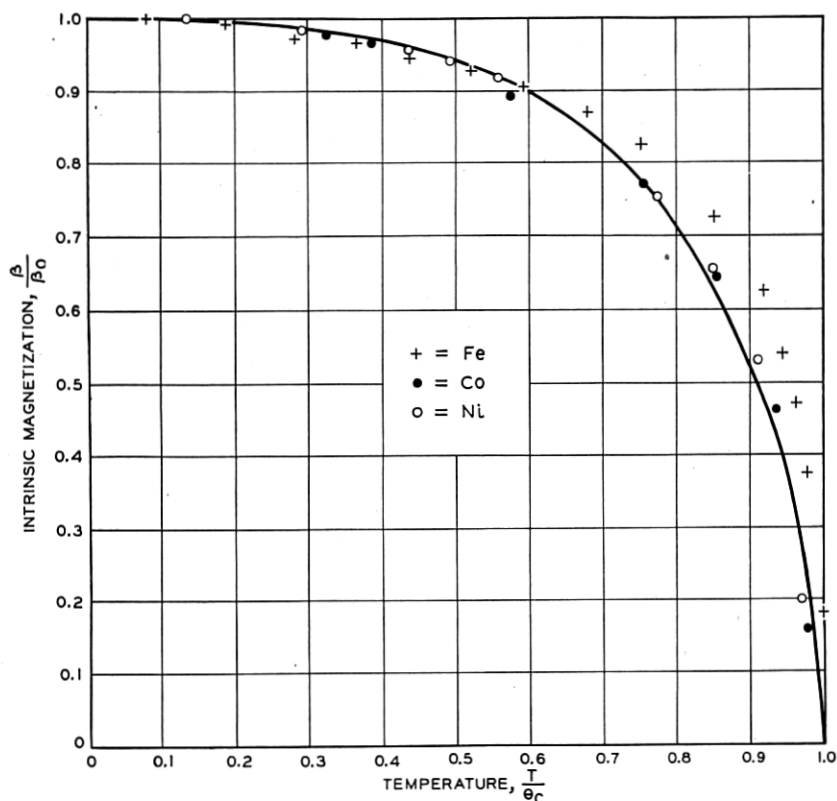

Fig. 29—Intrinsic magnetization versus temperature. The horizontal scale represents the temperature divided by the Curie temperature and the vertical scale, the intrinsic magnetization divided by the intrinsic magnetization at absolute zero. The theoretical curve is derived from quantum mechanics.

and becomes zero at a certain critical temperature, which is known as the Curie temperature and denoted by θ_c. A more complete discussion of the theory of the temperature dependence of magnetism would belong in a paper devoted solely to the theory of magnetism.[38] In

[38] See, for example, K. K. Darrow, *Bell Sys. Tech. Jour.* XV, 224 (1936), R. M. Bozorth, "The Present Status of Ferromagnetic Theory," *Bell Sys. Tech. Jour.*, XV, 63 (1936) and texts such as J. H. Van Vleck, "The Theory of Electric and Magnetic Susceptibilities," Oxford, 1932, E. C. Stoner "Magnetism and Matter," Methuen and Company, Ltd., London, 1934, and F. Bitter "Introduction to Ferromagnetism," McGraw-Hill Book Co., New York, 1937.

this paper we shall use the fact that the magnetism changes with the temperature to explain the anomalous expansion of ferromagnetic materials. In Fig. 29 we show the variation in intrinsic magnetization with temperature as observed for iron, cobalt and nickel.

Variation of Intrinsic Magnetization with Composition

Let us consider how the intrinsic magnetization should vary from element to element in the transition series, supposing always that the temperature is so low that thermal effects can be neglected. The element next to nickel is cobalt; cobalt has one less electron than nickel so that the $3d$ band and partially filled $4s$ band for it will have one less electron in them. Because of the relatively small number of quantum of states in the $4s$ as compared to the $3d$ band, this deficit will be made up mainly by the $3d$ band which will therefore contain not 4.4 as for nickel but instead 3.4 electrons leading to an unbalance of 1.6 Bohr magnetons per atom. The observed β for cobalt is 1.7 in good agreement with this.

One can obtain electron atom ratios intermediate between cobalt and nickel by forming alloys. We shall speak of the electron concentration, C, of these and other alloys, meaning by this term the total number of electrons available for the $3d$ and $4s$ bands divided by the total number of atoms. So long as the minus spin half of the $3d$ band remains full and so long as the number of electrons in the $4s$ band does not vary much, the value of β will be a linear function of the electron concentration varying from ~ 1.6 to ~ 0.6 as the concentration varies from 9 for cobalt to 10 for nickel. In Fig. 30 are given the intrinsic magnetizations plotted against electron concentration for a series of alloys. It is seen that from cobalt to about halfway between nickel and copper, an increase in C produces, very nearly, a numerically equal decrease in β. This means that the increase in C goes toward filling up the holes in the $3d$ band and reducing the unbalance and hence β. Some alloys are included in Fig. 30 for which the two elements are not adjacent in the periodic table; their values of β also conform to the values predicted from their electron concentrations.

The very natural way in which the band theory accounts for the results shown on Fig. 30 is its principal success in the theory of ferromagnetism.

The bend in the curve between iron and cobalt is not very satisfactorily explained at present. One theory is that for iron neither $3d$ band is entirely full; but this explanation is said to be inconsistent with the observed dependence of magnetization upon temperature at low

temperatures. Another theory is that there are only 0.2 electrons in the 4s band, thus leaving the remaining 7.8 electrons of iron distributed 5 to one 3d band and 2.8 to the other, leaving an unbalance of 2.2; this theory is unsatisfactory because it would require an inexplicable displacement upwards of the 4s band compared to the 3d in going from
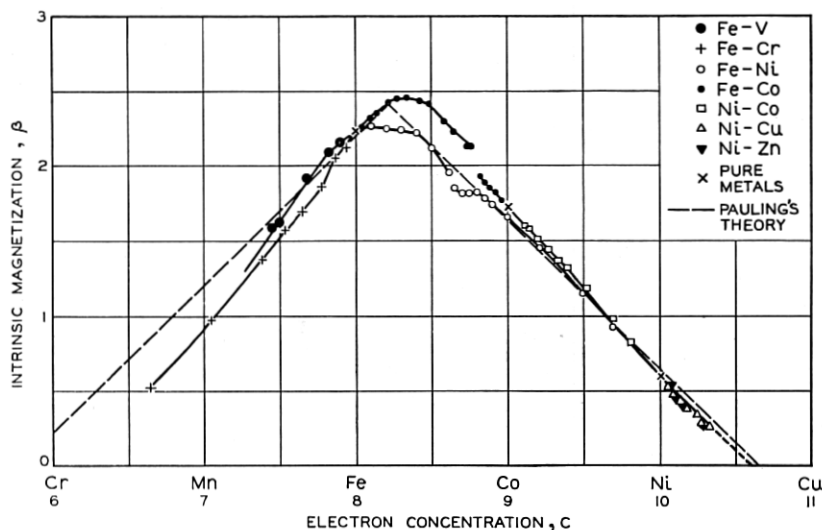

Fig. 30—Intrinsic magnetization versus electron concentration.

The data for this figure were obtained from the following sources:
Fe-V and Fe-Cr   M. Fallot   Ann. de Physique 6, 305–387 (1936).
Fe-Co   R. Forrer   J. de Physique et le Radium, 1, 325–339 (1930).
Fe-Ni   M. Peschard   Comptes Rendus 180, 1836 (1925).
Ni-Co   P. Weiss, R. Forrer, and F. Birch   Comptes Rendus 189, 789–791 (1929).
Ni-Cu and Ni-Zn   V. Marian   Ann. de Physique 7, 459–527 (1937).

cobalt to iron. Another theory has been proposed by Pauling[39]; he has stated it in the "atomic" language but it can be translated into the band language as follows: the 3d band is broken into two parts, an upper part containing 4.88 levels per atom, 2.44 for each spin, and a lower part separated from the upper by an energy gap and containing 5.12 levels per atom, 2.56 for each spin. A number of electrons per atom varying from 0.6 for nickel to 0.7 for cobalt are in the 4s band; for simplicity we shall suppose that this number has a constant value of 0.65 electrons per atom. According to this simplification, one of the upper parts of the 3d band has 0.65 holes for nickel. This band will become empty if the electron concentration is decreased by 1.79 (= 2.44 − 0.65)—that is, for a concentration of 10 − 1.79 = 8.21.

[39] L. Pauling, *Phys. Rev.*, *54*, 899 (1938).

If the concentration is decreased below 8.21, electrons will be removed from the upper part with the other spin; this will result in a decrease in the unbalance and hence in β, which has for $C = 8.21$, a value of 2.44— corresponding to one filled and one empty upper part; and this decrease will be numerically equal to the decrease in C. Accordingly, the value of β for iron, $C = 8$, is $2.44 - 0.21 = 2.23$. The numbers 2.44 and 2.56 were, of course, chosen so as to obtain this agreement for iron. This theory of Pauling expresses reasonably well the variations in β for all the alloys of Fig. 30.

Criterion for Ferromagnetism

We must now see how the theory explains the absence of ferromagnetism for the remaining transition elements. We have seen that the exchange energy lowers and the Fermi energy raises the energy of the magnetized state compared to the unmagnetized state. These two effects very nearly cancel even for the magnetic elements iron, cobalt, and nickel. For the other elements in the transition series, which are not ferromagnetic, the Fermi term apparently exceeds the exchange term. We shall give a theoretical reason for expecting this result.

In the first place we must indicate how nearly the effects cancel. Let us take cobalt, which has nine electrons in the $3d$ and $4s$ bands, as an example. From Fig. 27 we see that for cobalt in the unmagnetized state both $3d$ bands are filled to about $- 0.46$ atomic units. In the magnetized state one band is filled by electrons which have come from levels with less energy of motion in the other band. Since the top of the $3d$ band comes at about $- 0.42$ units on Fig. 27, the average gain in energy for each transferred electron is about 0.04 units. Since the number of electrons transferred is 1.7 per atom, the increase in Fermi energy is 0.068 atomic units or 0.9 ev per atom. From an analysis of thermal measurements the value for the actual energy of magnetization is found to be about 0.2 ev per atom; a value which is only about one fourth of the predicted increase in the Fermi energy. Hence the exchange energy exceeds the Fermi energy by only 25 per cent and the two energies nearly cancel.

The variation in the structure of the $3d$ band from element to element was discussed in connection with Fig. 27; we concluded then that the bands become wider as we recede in the periodic table from nickel towards scandium. Greater band width means greater Fermi energy in the magnetized state and this effect opposes the occurrence of ferromagnetism. The exchange energy can also change. Calculations by Slater,[40] which unfortunately are too over-simplified to bear much

[40] J. C. Slater, *Phys. Rev.*, 49, 537, 931 (1936).

weight, show that for manganese the Fermi energy outweighs the exchange energy so that manganese is not ferromagnetic at all. In Fig. 31 we represent the state of affairs predicted for chromium; the exchange
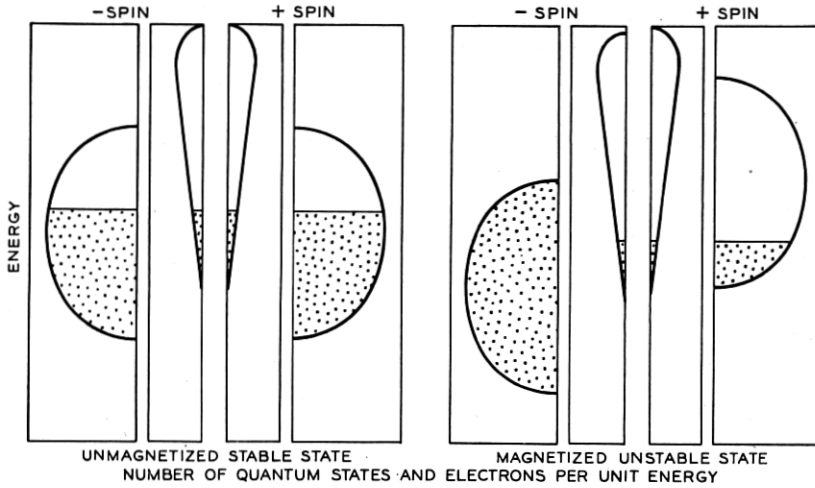

Fig. 31—The absence of ferromagnetism for chromium.

energy is over-balanced by the Fermi energy and for this metal the unmagnetized arrangement has the least energy and is the stable state.

A very instructive curve can be drawn to illustrate the criterion for the occurrence of ferromagnetism. It is shown in Fig. 32. The vertical scale is the energy of the unmagnetized state, E_U, minus the energy of the magnetized state, E_M. When $E_U - E_M$ is positive, the magnetized state has the lower energy and will be the stable state,
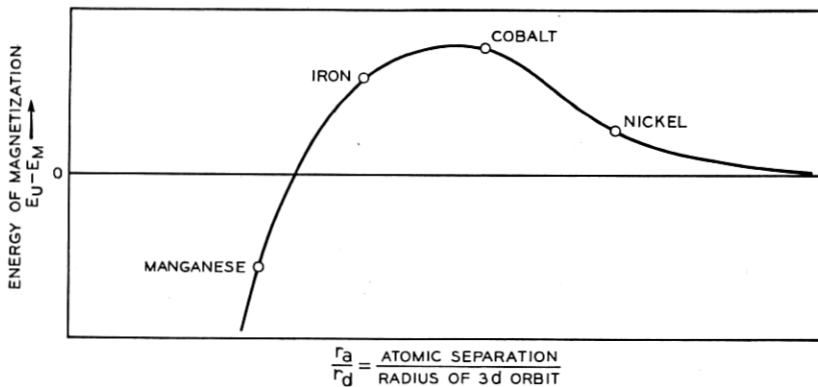
Fig. 32—Criterion for the occurrence of ferromagnetism.

and when $E_U - E_M$ is negative, the reverse is true. Hence a positive value for $E_U - E_M$ is a necessary and sufficient condition for ferromagnetism. The variable on the horizontal scale is r_a (the distance between nearest neighboring atoms in the crystal) divided by r_d (the average radius for the $3d$ wave function). Small values of r_a/r_d mean crowding together of the atoms, large values of the Fermi energy, and no ferromagnetism. Certain values of r_a/r_d, such as are found for iron, cobalt, and nickel, favor ferromagnetism. Very large values of r_a/r_d mean widely separated atoms and low Fermi energy and, consequently, ferromagnetism; however, for very widely separated atoms, the energy of interaction between them is small and so is the energy of magnetization. The curve shown in Fig. 32 is only qualitative. The theory that the curve should have this form was first worked out by Bethe using the "atomic" rather than the band theory of magnetism; for the reasons discussed above, however, no quantitative theoretical curve is available. Ratios of r_a/r_d have been calculated by Slater * and occur as indicated for several elements. This curve can be considered from either of two viewpoints. We may imagine that r_a remains constant, as it does approximately for the transition elements, and that r_d varies from element to element; we then get the result shown in Fig. 32. On the other hand we may consider a definite chemical element thus fixing r_d; then Fig. 32 tells us how the energy of magnetization depends on the lattice constant or volume of the sample. We shall use this in the following paragraphs to explain the effects of magnetism upon thermal expansion.

Magnetism and Thermal Expansion

In Fig. 33a we show a solid curve which represents for iron in the magnetized state the dependence of the energy E_M upon the lattice constant a. In Fig. 33b is shown, on a relatively enlarged energy scale, the value of $E_U - E_M$ as taken from Fig. 32 with r_d thought of as fixed, and a the lattice constant in place of r_a. The position of curve (b) has been adjusted so that the point marked O, corresponding to iron in Fig. 32, comes at the equilibrium distance or minimum of the E_M curve. Adding the solid curves of (a) and (b) (adjusting the energy scales, of course) gives the dashed curve representing the energy E_U of the unmagnetized state shown in Fig. 33a. We are now in a position to make predictions about the thermal expansion of iron.

Let us imagine that the iron is somehow made to stay in the magnetized state. Then its expansion curve, lattice constant versus temperature, will be shown as in Fig. 33c by the solid heavy line. Next

---

* J. C. Slater, *Phys. Rev.*, *36*, 57 (1930).

imagine it maintained in the unmagnetized state; in this state the equilibrium lattice constant is smaller than for the magnetized case and the expansion curve is shown dashed. The curves for fixed
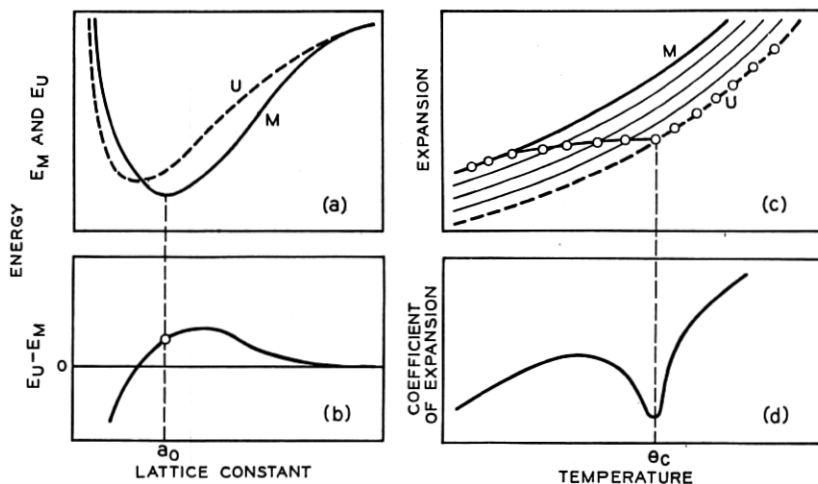

Fig. 33—Theory of the thermal expansion of iron.

(a) Energy in magnetized (M) and unmagnetized (U)
    states versus lattice constant.
(b) Difference in energies versus lattice constant.
(c) Lattice constant versus temperature.
(d) Thermal expansion coefficient versus temperature.

intermediate degrees of magnetization are shown as light lines. Now as the iron is heated the magnetization does not stay constant but decreases with temperature and becomes zero at the Curie temperature θ_c. In Fig. 33c this corresponds to a continuous shifting from the line of higher magnetization to the lines of lesser magnetization with increasing temperature as indicated by the curve with circles. We see that the rate of expansion—that is, the thermal expansion coefficient, which is defined as the derivative of the curve divided by a—should have an irregular form as shown in Fig. 33d.

In Fig. 34 we show observed thermal expansion curves for a series of iron nickel alloys,[41] showing that the expansion for iron rich alloys agrees with that predicted from Fig. 33. The reader may verify that had the curve of Fig. 33b been adjusted to correspond to nickel, the anomalous expansion would have been in the opposite direction, as is found experimentally for the nickel rich alloys.

The more rapid the transition from the magnetized to the unmag-

[41] Figures 34 and 35 are taken in a modified form from J. S. Marsh, "Alloys of Iron and Nickel," Vol. I, Special-Purpose Alloys, 1938, McGraw-Hill Book Co.

netized curve, the greater will be the anomaly in expansion. For the alloy invar the Curie point occurs at about 200° C. and the transition is so rapid that the magnetic effect nearly cancels the normal expansion.
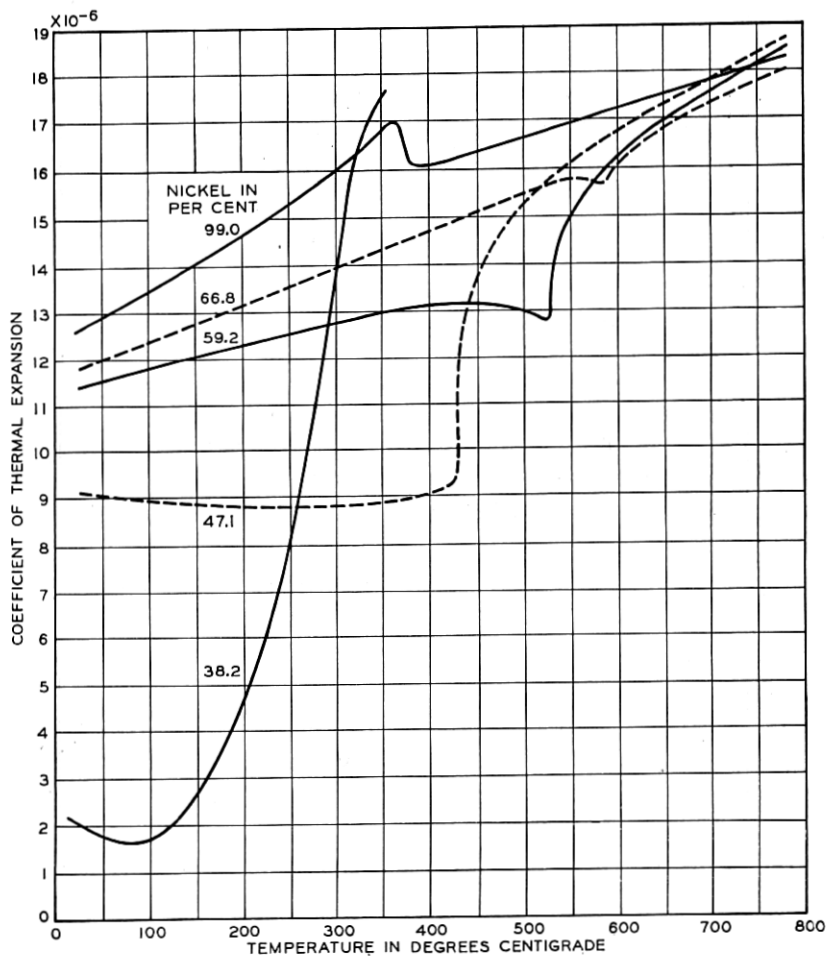

Fig. 34—Coefficients of expansion for iron-nickel alloys versus temperature.

Figure 35 shows a curve for the thermal expansion of an iron-nickel alloy containing 36.5 per cent Ni, corresponding to Fig. 33c. The flat region implies an expansion coefficient of nearly zero.

Grüneisen's law is definitely violated by metals having expansion effects of the sort associated with ferromagnetic changes. Grüneisen's law, it will be recalled, states that the thermal expansion coefficient is proportional to the specific heat. For all ferromagnetic transforma-

tions, the specific heat has a peak at the Curie temperature. For invar, however, the thermal expansion suffers a dip at the Curie temperature. Hence the proportionality between specific heat and thermal expansion coefficient does not hold. Even for cases where the expansion coeffi-
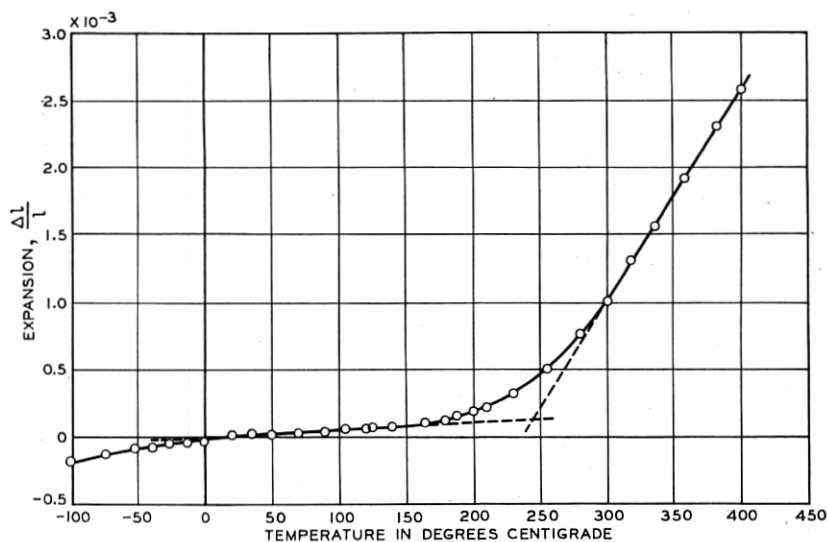
Fig. 35—Expansion of invar versus temperature.

cient has a peak, as in nickel for example, the proportionality does not hold. The reason for the failure of Grüneisen's law is easily found and reflects in no way upon validity of the law for the cases to which it is intended to apply. Grüneisen's law is derived by assuming that the crystal has a single definite energy versus volume curve. For ferromagnetic materials this is not true as is evinced by the two curves of Fig. 33a.

In this paper we have been concerned with the important but inactive attributes of electrons associated with their energies. We have seen how the variations of the electronic energy levels can be used to explain a number of the important properties of solids. In the next paper, we shall discuss the more dynamic subjects of electron velocities and accelerations.