Nonorthogonality and Ferromagnetism

J. H. VAN VLECK, *Harvard University*

(Received December 9, 1935)

The calculations in Heisenberg's theory of ferromagnetism have been questioned by Inglis and others on the ground that the error resulting from the nonorthogonality of the wave functions may possibly increase without limit when the number of atoms becomes arbitrarily large. In the present paper it is proved that this difficulty does not really arise. Semiquantitative formulas are given to correct for the error due to nonorthogonality, which is shown to be of the order $2z\delta^2$ relative to unity, where z is the number of neighbors and δ is the overlap integral (1). A supplementary note is included on a new method of approximating the partition function in Heisenberg's theory. This approximation should be somewhat better than the assumption of a Gaussian distribution, but agrees even worse with experiment, provided one assumes orthogonality. Actually, the influence of nonorthogonality is sufficiently large to render uncertain any attempt to deduce exactly the critical conditions (minimum number of neighbors, etc.) necessary for ferromagnetism.

INTRODUCTION

IN an interesting paper, Inglis[1] has stressed the fact that all existing secular calculations by the Heitler-London method in a system with a very large number of electrons are subject to question because they assume that the wave functions of the different atoms are mutually orthogonal. The most notable calculations open to this objection are those in Heisenberg's[2] theory of ferromagnetism. The apparent difficulty arises regardless of how small is the "overlap" or "nonorthogonality" integral

$$\delta = \int\int\int \psi_i(1)\psi_j(1)dv_1. \tag{1}$$

Indeed Inglis shows that if the number n of atoms is large, the secular determinant is so prolific in terms involving δ as to suggest that in the final result the correction factor due to nonorthogonality may be of the order $1 \pm n\delta^2$ rather than $1 \pm \delta^2$ as in the familiar two-particle case. (Cf. also our discussion following Eq.. (8).) If this is really the case, the ordinary calculations on ferromagnetism, etc., become devoid of all meaning, since regardless of how small is δ, the error can be made arbitrarily large by taking n sufficiently great. In a crystal, the number n is, of course, to be considered as enormous.

To be sure, the success of Heisenberg's theory of ferromagnetism is indirect evidence that the nonorthogonality catastrophe is only an apparent one. However, until a mathematical proof has been given, the theory can scarcely be said to be on a rigorous basis. In the present paper we aim to supply such a proof.

Let us consider a system of n electrons in n distinct states. Each state is to relate to a different atom. We shall assume that we are solving only the problem of exchange degeneracy. Every atom must then be in an s state, as otherwise there would be the spatial degeneracy to contend with. The secular equation associated with the exchange degeneracy is

$$\text{det. } \mathbf{K} = 0,$$

where $\mathbf{K}$ is the matrix

$$\mathbf{K} = \sum_i (V^{(i)} - W\Delta^{(i)})\mathbf{P}^{(i)}. \tag{2}$$

Here W is the desired energy constant, $\mathbf{P}^{(i)}$ is a Dirac permutation matrix,[3] and $V^{(i)}$, $\Delta^{(i)}$ are the integrals

$$V^{(i)} = \int \cdots \int (\mathbf{P}^{(i)}\Psi^*)\mathbf{H}\Psi dv_1 \cdots dv_n, \tag{3}$$

$$\Delta^{(i)} = \int \cdots \int (\mathbf{P}^{(i)}\Psi^*)\Psi dv_1 \cdots dv_n. \tag{4}$$

When preceding a wave function Ψ^*, the symbol $\mathbf{P}$, of course, denotes a permutation operator rather than a permutation matrix, while $\mathbf{H}$ is the Hamiltonian operator.

The sum in (2) is over the various possible permutations, and embraces only a finite number of terms since n is finite. Written out more explicitly, with the constituent cycles of each $\mathbf{P}^{(i)}$ specified, (2) is

[1] D. R. Inglis, Phys. Rev. **46**, 135 (1934).

[2] W. Heisenberg, Zeits. f. Physik **49**, 619 (1928).

[3] For fuller explanation of the meaning of permutation matrices in connection with exchange degeneracy see P. A. M. Dirac, Proc. Roy. Soc. **A123**, 714 (1929), or Chap. XI of his *Principles of Quantum Mechanics*, first edition.

$$\begin{aligned}\mathbf{K}=&-W\mathbf{1}+\textstyle\sum_{j>i}[V_{(ij)}-\Delta_{(ij)}W]\mathbf{P}_{(ij)}\\&+\textstyle\sum[V_{(ijk)}-\Delta_{(ijk)}W]\mathbf{P}_{(ijk)}\\&+\textstyle\sum[V_{(ij)(kl)}-\Delta_{(ij)(kl)}W]\mathbf{P}_{(ij)(kl)}\\&+\textstyle\sum[V_{(ijkl)}-\Delta_{(ijkl)}W]\mathbf{P}_{(ijkl)}\\&+\text{permutations of fifth and higher order.}\end{aligned}\tag{5}$$

A notation such as $(ij)(klm)$ means $i\rightarrow j\rightarrow i$; $k\rightarrow l\rightarrow m\rightarrow k$, and any given permutation is to be counted only once in the summation. Thus $\mathbf{P}_{(lk)(ij)}$ is not to be regarded as distinct from $\mathbf{P}_{(ij)(kl)}$, etc. The wave functions Ψ to be used in (3)–(4) are those of the simple product form

$$\Psi=\psi_1(1)\psi_2(2)\cdots\psi_n(n),\tag{6}$$

since before solution of the secular problem, the proper linear combinations are not known. Because of the invariance of $\mathbf{H}$ under permutations, it matters not which particular Ψ is selected for use in (6). Thus the first two factors in (6) could equally well be $\psi_1(2)\psi_2(1)$. In writing (5) we have assumed that the additive constant in $\mathbf{H}$ (and hence in W) is determined in such a way that (3) vanishes for the identical permutation, and the coefficient of 1 in (5) is consequently simply—W. This point is discussed further in the appendix, and must be kept in mind when the exchange integrals (3) are computed, as they are not invariant of the origin for $\mathbf{H}$.

To show that it is not obvious that the nonorthogonality or Δ terms are of subordinate importance, let us try the approximation sometimes made in the literature of stopping with second order permutations. Then (2) becomes

$$\mathbf{K}=\textstyle\sum_{j>i}[V_{(ij)}-\Delta_{(ij)}W]\mathbf{P}_{(ij)}-W\mathbf{1}.\tag{7}$$

To see how things go, we note that one root of (7) is easy to find, *viz.*, that corresponding to the antisymmetric orbital solution. Here the $\mathbf{P}_{(ij)}$ all have the characteristic value -1. Let us assume that the nonvanishing $\Delta_{(ij)}$, $V_{(ij)}$ all have common values δ^2, J, respectively, and are $\frac{1}{2}nz$ in number, as is the case when each atom is coupled to z neighbors. Then according to (7)

$$W=-\tfrac{1}{2}nzJ/(1-\tfrac{1}{2}nz\delta^2).\tag{8}$$

The nonorthogonality thus spoils the answer completely for large values of n. In fact, (8) would give such an absurdity as $W=\infty$ if $\delta^2=2/nz$. We shall see later that in reality it is a poor approximation to stop with only second order permutations and that for large n, rather than use (8), it is a better approximation to disregard nonorthogonality completely, and employ simply

$$K=\textstyle\sum_{j>i}[V_{(ij)}\mathbf{P}_{(ij)}]-W\mathbf{1}\tag{9}$$

as the secular matrix.

Factorization of the Exchange Integrals and Secular Determinant

Let us consider an exchange integral such as $V_{(ij)(kl)}$ which is associated with a permutation which can be factored into simple permutations, with each cycle involving different atoms. In the computation of this integral, the most important portions of the Hamiltonian function are the part which involves the coordinates of the electrons on atoms i and j, and the part which involves the coordinates of the electrons on atoms k and l. This point is analyzed more fully in the appendix—it seems best not to consider it more fully at present since it is perhaps fairly obvious that if, for example, we are permuting electrons i and j, it is their part of the Hamiltonian function which gives the most important contribution to the exchange integral. This assumption provides a simplification needed for our later work by permitting factorization of the higher order exchange integrals into exchange integrals of lower order and nonorthogonality integrals. Thus it enables us to write

$$V_{(ij)(kl)}=V_{(ij)}\Delta_{(kl)}+V_{(kl)}\Delta_{(ij)},\tag{10}$$

similarly

$$\left.\begin{aligned}V_{(ij)(kl)(mn)}=&V_{(ij)}\Delta_{(kl)}\Delta_{(mn)}+V_{(kl)}\Delta_{(ij)}\Delta_{(mn)}\\&+V_{(mn)}\Delta_{(ij)}\Delta_{(kl)},\\V_{(ij)(lmn)}=&V_{(ij)}\Delta_{(lmn)}+V_{(lmn)}\Delta_{(ij)},\quad\text{etc.}\end{aligned}\right\}\tag{11}$$

The trick is now to note that if we accept (10), (11), then under certain conditions (2) or (5) is nearly the same, for large n, as

$$\begin{aligned}[\mathbf{1}+&\textstyle\sum\Delta_{(ij)}\mathbf{P}_{(ij)}+\sum\Delta_{(ijk)}\mathbf{P}_{(ijk)}\\&+\textstyle\sum\Delta_{(ij)(kl)}\mathbf{P}_{(ij)(kl)}+\cdots][\sum V_{(ij)}\mathbf{P}_{(ij)}\\&+\textstyle\sum V_{(ijk)}\mathbf{P}_{(ijk)}+\sum V_{(ijkl)}\mathbf{P}_{(ijkl)}-W\mathbf{1}],\end{aligned}\tag{12}$$

for when we multiply out (12) and employ (10)–(11), we obtain all the terms of (5) and also certain extra terms. No permutations of the type $(ij)(kl)$, etc., occur in the second factor in (12). Use of (12) will be warranted when these extra terms are not too abundant. They occur whenever one or more subscripts is repeated in both factors in the multiplication. For instance,

$$\Delta_{(ij)}\mathbf{P}_{(ij)}V_{(ik)}\mathbf{P}_{(ik)}=\Delta_{(ij)}V_{(ik)}\mathbf{P}_{(ijk)} \qquad (13)$$

is a term included in (12) but not in (5).

If (12) were rigorously true, the secular problem would be greatly simplified. Since the determinant of the product of two matrices is equal to the product of the determinants, we could equate the determinant of the second factor in (12) to zero. It is particularly to be noted that the only $\mathbf{P}$'s above the second order in this factor are those involving permutations which cannot be factored into two or more distinct permutation cycles. Such permutations we shall call unfactorable permutations. On the other hand, all types of permutations are included in the first factor of (12). If it is allowable to discard the V's associated with nonfactorable permutations, i.e., to set

$$V_{(ijk)}=V_{(ijkl)}=\cdots=0, \qquad (14)$$

then the second factor of (12) reduces to (9), and the energy levels W are the same as those associated with a secular problem (9) obtained by completely disregarding nonorthogonality and higher order permutations. The whole demonstration thus resolves itself into the consideration of the amount of error involved in the substitution of (12) for (5) and in the assumption (14). We shall discuss the latter first. It is to be emphasized that while (14) discards nonfactorable permutations, it is very necessary in the proof to retain the V's associated with factorable permutations, for it is their very existence which makes possible the cancelation of most of the effect of nonorthogonality. Although they do not appear in the second sum in (12), their effect has not been discarded, as they are yielded on multiplication of the two factors.

Examination of the approximation (14). We may distinguish between two cases: first, one in which every atom is coupled to every other atom, i.e., where $V_{(ij)}$ is of about the same order of magnitude regardless of the location of i and j. The second case is that where the important coupling is between adjacent atoms, so that we may take $V_{(ij)}=0$ unless i and j are neighbors. Fortunately it is the second case which occurs in nature. Of course it is an idealization to consider only the coupling between adjacent atoms, but the coupling energy decreases exponentially with the interatomic distance, and this fact makes the convergence essentially that characteristic of the second case.

In the first case, the number of nonfactorable permutations is comparable with the factorable and (14) surely is an unwarranted approximation. For instance, the number of permutations of type $V_{(ijkl)}$ is of the same order n^4 as the number of the form $V_{(ij)(kl)}$. The terms of type $V_{(ijkl)}$ are presumably individually of the order of magnitude $V_{(ij)}\Delta_{(kl)}$ like $V_{(ij)(kl)}$, and the abundance of the former is so great that regardless of the smallness of Δ_{kl}, they overshadow the retained ordinary second order terms V_{ij}, which are individually much larger, but only $\frac{1}{2}n^2$ in number.

In the second case, the nonfactorable permutations are relatively much less abundant than in the first. For instance, in linear chains, or in simple, face-centered, or body-centered cubic arrangements, there are no terms of the third order type $V_{(ijk)}$, as long as we are including only coupling between adjacent atoms. The number of terms of form $V_{(ijkl)}$ is at most[4] of the order nz^2, where z is the number of neighbors possessed by a given atom. On the other hand, the number of factorable fourth order permutations, such as $V_{(ij)(kl)}$, which we retain, is of the order n^2z^2. The number of ordinary second order permutations involved in (5) is $\frac{1}{2}nz$. If $V_{(ij)(kl)}$ is comparable in magnitude with $V_{(ij)}\Delta_{(kl)}$, the ratio of the effect of the neglected unfactorable permutations to

[4] Unfortunately it does not appear feasible to derive exact expressions for the number of discarded nonfactorable permutations appropriate to the various types of spatial arrangements. In the case of the linear chain, such permutations are entirely absent if we consider only coupling between adjacent atoms. For simple cubic gratings ($z=6$) there are $6n$ relevant permutations of type $(ijkl)$. So we have probably overestimated the number of nonfactorable permutations of mth degree connecting adjacent atoms when we state that their abundance is of the order nz^{ml2}. However, it seems better to be on the safe side of things.

that of the ordinary $V_{(ij)}$ terms is presumably of the order $z\Delta_{ij}$. Similarly, the number of unfactorable terms of the mth order is roughly $nz^{m/2}$, or less,[4] and the size of each term $V_{(ij)}\Delta_{(ij)}{}^{(m-2)/2}$. The abundance increases with order only as a power of z rather than n, and is offset by the occurrence of a higher power of $\Delta_{(ij)}$. Hence we conclude that (14) is warranted if $\Delta_{(ij)}$ is small compared to $1/z$.

Calculation of Error Involved in Substitution of (12) for (5)

When the multiplication in (12) is performed, it is easily seen that the superfluous terms as compared with (5) are relatively small in number provided each atom is coupled only to its neighbors. For instance, there are about nz^2 terms of the unwanted structure (13), whereas there are nearly $n^2z^2/4$ terms of the desired type $\Delta_{(ij)}V_{(kl)}\mathbf{P}_{(ij)}\mathbf{P}_{(kl)}$, individually of the same order of magnitude $V_{ij}\Delta_{kl}$. Hence it is reasonable that (12) is an allowable approximation. This argument cannot, however, be regarded as entirely rigorous, for the extra terms are still exceedingly numerous in absolute numbers. A similar difficulty did not arise in connection with (14), as (14) is an approximation internal to the second factor of (12), whereas the first factor is much more prolific in higher order terms. Hence it will be safer and more illuminating if we examine quantitatively the difference between (5) and (12) by studying some specific examples in which the energy can be calculated explicitly both with and without the approximation (12).

In all these examples we shall make the approximation (14), and also assume that the Δ's associated with nonfactorable higher order permutations vanish, so that

$$\Delta_{(ijk)}=\Delta_{(ijkl)}=\cdots=0. \tag{15}$$

Eq. (15) is the natural analog of (14), and requires that $\Delta^{(i)}$ vanish for any permutation which is not factorable into simple permutations, since the $\Delta^{(i)}$ for any differently factorable permutation have at least one factor of the form (15). For instance, $\Delta_{(ij)(klm)}=\Delta_{(ij)}\Delta_{(klm)}$. The suppositions (14) and (15) can be regarded as slight specializations of the model as far as the comparison of (5) and (12) is concerned. Because only a small percentage of the V's and Δ's are of the form (14) or (15), the error involved in passing from (5) to (12) would very probably not be materially altered if different values than (14) and (15) were assumed. We shall suppose that all nonvanishing $\Delta_{(ij)}$ and $V_{(ij)}$ have common values δ^2 and J, respectively. The notation δ^2 rather than δ is used because a simple permutation introduces in (4) two one-electron integrals of the type (1), namely, one for each of the two electrons involved. The first and second expressions of (11) now become $3J\delta^4$ and 0, while more generally

$$\Delta^{(i)}=\delta^{2k},\quad V^{(i)}=kJ\delta^{2k-2}, \tag{16}$$

if $P^{(i)}$ is expressible as the product of k simple permutations of neighboring atoms, with different atoms involved in each constituent cycle. Otherwise $\Delta^{(i)}$ and $V^{(i)}$ vanish.

Even with these approximations, the secular problem is one of hopeless complexity except as we confine our attention to particularly simple special solutions. One root of this character is that which corresponds to the antisymmetric orbital state or in other words to completely parallel alignment of spin. Here $\mathbf{P}^{(i)}$ has the characteristic value $+1$ or -1 according as it is an even or odd permutation. The solution of (5) thus becomes

$$W=\frac{J[-\nu_1+2\nu_2\delta^2-3\nu_3\delta^4+\cdots]}{[1-\nu_1\delta^2+\nu_2\delta^4-\cdots]}, \tag{17}$$

where ν_k is the total number of permutations of degree $2k$ which can be factored into k simple permutations. The gist of Inglis's observations, when applied to (17), is that since ν_1, ν_2, $\cdots$ are very large in a crystal, the numerator and denominator of (17) certainly have vastly different values than when one stops with the first term, so that it is not obvious that (17) has nearly the same value as when $\delta=0$. The essence of our answering argument is that the terms involving δ enter nearly as a common factor in numerator and denominator, so that the value of the ratio (17) is, for small δ but arbitrarily large n, not much different from that obtained by stopping with the first term. To see that this is really so, we now consider the following particular models.

(*a*) *Open linear chain.* Here each atom has two neighbors, except the end atoms, which have only one. The ν_k may be evaluated rigorously, and one finds that the denominator D of (17) is

$$D=1-(n-1)\delta^2+\tfrac{1}{2}(n-2)(n-3)\delta^4+\cdots+(-1)^k(n-k)\cdots(n-2k+1)\delta^{2k}/k!+\cdots.$$

This finite series is exactly summable[5] to

$$D=\delta^n \sinh (n+1)\varphi/\sinh \varphi, \quad \text{where} \quad \cosh \varphi=1/2\delta. \tag{18}$$

It is not necessary, for any of the models, to give the explicit form, or mode of evaluating the numerator of (17), as, except for a factor J, the numerator is obtained merely by differentiating the denominator with respect to δ^2 and so presents no trouble.

With the aid of (18), one finds that (17) becomes

$$W=-2J\frac{\cosh^3 \varphi}{\sinh \varphi}\left[\frac{(n+1)\cosh (n+1)\varphi}{\sinh (n+1)\varphi}-\frac{\cosh \varphi}{\sinh \varphi}-n\frac{\sinh \varphi}{\cosh \varphi}\right]. \tag{19}$$

For very large n, (19) is the same as

$$W=-\tfrac{1}{2}\delta^{-2}J[n(1-4\delta^2)^{-\frac{1}{2}}-n+(1-4\delta^2)^{-\frac{1}{2}}-(1-4\delta^2)^{-1}]=-(n-1)J(1+3\delta^2+\cdots). \tag{20}$$

The dots denote omitted terms of order δ^4 and higher. Now $-(n-1)J$ is the value of the energy which would be obtained with neglect of nonorthogonality, i.e., by using (12) rather than (5). So the correction factor due to nonorthogonality is $(1+3\delta^2+\cdots)$, and not $(1+n\delta^2+\cdots)$ as (8) would suggest.

(*b*) *Closed linear chain.* A closely allied problem is that of the cyclic linear chain, which differs from (a) only in that there are no "end" atoms. The series in the denominator of (17) is now[5]

$$1-n\delta^2+\tfrac{1}{2}n(n-3)\delta^4+\cdots+(-1)^k n(n-k-1)\cdots(n-2k+1)\delta^{2k}/k!+\cdots=2\delta^n \cosh n\varphi,$$

with φ as in (18). So (17) becomes

$$W=-2J(\cosh^3 \varphi/\sinh \varphi)(n \tanh n\varphi-n \tanh \varphi),$$

which for large n is

$$W=-\tfrac{1}{2}n(J/\delta^2)[(1-4\delta^2)^{-\frac{1}{2}}-1]=-nV[1+3\delta^2+\cdots]. \tag{21}$$

The correction $1+3\delta^2+\cdots$ due to including nonorthogonality is exactly the same as in (20), as one would expect since it is physically reasonably apparent that it doesn't make much difference whether or not the chain is closed. The fact that (21) involves n and (20) contains $n-1$ is, of course, because the closed chain involves one more pair than the open chain.

(*c*) *Three-dimensional configurations.* We now tackle the more physical problem where the atom has z rather than 2 neighbors. (The simple cubic arrangement has $z=6$, the body-centered $z=8$, the face-centered $z=12$, etc.) We shall neglect the "edge effect" resulting from the fact that atoms at the edge of the crystal do not have their full quota of neighbors. The similarity of results with (a) and (b) makes it clear that this is legitimate. Even then, it would be exceedingly difficult to obtain, to say nothing of sum the series with, an accurate general formula for ν_k. However, a sufficiently good approximation is

$$\nu_{k+1}=\nu_k[\tfrac{1}{2}nz-2kz+k+O(k/n)]/(k+1), \tag{22}$$

where the symbol $O(k/n)$ means that terms involving $1/n$ and other negative powers of n have been neglected.

---

[5] For evaluation of the series see Bromwich, *Theory of Infinite Series*, pp. 177–178. I am indebted to my father, E. B. Van Vleck, for calling my attention to this reference.

When we use (22) the denominator of (17) becomes

$$D=1-\tfrac{1}{2}nz\delta^2+\cdots+(-1)^{k+1}(\tfrac{1}{2}nz)(\tfrac{1}{2}nz-2z+1)\cdots(\tfrac{1}{2}nz-2kz+k)\delta^{2k+2}/(k+1)!+\cdots.$$

This series can be evaluated by the binomial theorem, and is merely $[1-(2z-1)\delta^2]^{nz/(4z-2)}$. Consequently (17) is

$$W=-\tfrac{1}{2}nzJ/[1-(2z-1)\delta^2]=-\tfrac{1}{2}nzJ[1+(2z-1)\delta^2+\cdots]. \tag{23}$$

The correction factor for nonorthogonality again involves only z and not n. As a valuable check on the validity of the approximation (22), we note that if $z=2$, as in the linear chain, this correction becomes $1+3\delta^2+\cdots$, agreeing perfectly with (20) or (21). (The agreement, however, is lost when higher powers of δ^2 are considered, as further development of (21) gives $1+3\delta^2+10\delta^4+35\delta^6+\cdots$ whereas (23) yields $1+3\delta^2+9\delta^4+27\delta^6+\cdots$.)

Note particularly that the overlap integral δ enters in (20), (21), (23) only in its square. Consequently if, say, we have $\delta=0.1$, a value not unreasonable in ferromagnetic applications, and if $z=8$ or 12, as in the ferromagnetic lattices, the correction due to nonorthogonality modifies the answer by about 20 or 30 percent—not enough to spoil the general trend of the Heisenberg theory.

It is, of course, reasonable that the correction due to nonorthogonality should depend on the number z of neighbors rather than on the total number n of atoms in the system. If this correction kept increasing with n, it would presumably be impossible to isolate different portions of the universe for even approximate calculations, and it is doubtful whether quantum mechanics would make sense, for regardless of how large were the mean distances between atoms, the results would keep depending on the total size assumed for the complete system.

(*d*) *Coupling between every atom.* It is interesting to note that (17) can be evaluated accurately here. The denominator becomes

$$D=1-\tfrac{1}{2}n(n-1)\delta^2+\cdots+(-\tfrac{1}{2})^k n(n-1)\cdots(n-2k+1)\delta^{2k}/k!+\cdots=(2/\delta)^n H_n(1/\delta),$$

where $H_n(x)$ is the Hermitian polynomial of order n. We are interested in the asymptotic value of $H_n(x)$ for large values of n, and for our purposes it is sufficient to use the approximation $H_n(x)\sim$const. $\times e^{\frac{1}{2}x^2}\cos[(2n)^{\frac{1}{2}}x]$, which may be obtained by applying the W.K.B. expansion to the differential equation satisfied by $H_n(x)$. (If n is odd rather than even, the sine rather than cosine should be used.) The expression (17) becomes

$$W=\tfrac{1}{2}J[-n\delta^{-2}-\delta^{-4}+(2n)^{\frac{1}{2}}\delta^{-3}\tan(2n/\delta^2)^{\frac{1}{2}}]. \tag{24}$$

The result (24) fluctuates with n in an irregular, meaningless way, and is entirely different from the formula obtained with disregard of nonorthogonality. It is thus fortunate that the present case (d) is one of only academic interest. It is barely possible that the trouble would be less acute if calculations could be made without the assumptions (14)–(15), which are entirely unwarranted in (d).

Discussion of other roots than the antisymmetric orbital solution. We have apparently made our discussion very special in the preceding examples by considering only the solution which is antisymmetric in the orbits. Conceivably one might object that the nonorthogonality corrections are more important under other conditions. To show that this is not so, we may note that there is one more case which can be solved accurately,—*viz.*, that obtained by using the symmetrical orbital solution. The latter obviously doesn't satisfy the exclusion principle, as we can't have all spins mutually antiparallel at once. However, for formal purposes we can imagine ourselves in a universe not governed by Pauli. One obtains the symmetrical from the antisymmetrical problem merely by changing all minus to plus signs in (17). Hence all our previous calculations (a), (b), (c), (d) can be immediately adapted if we simply change the sign of δ^2 and J. Thus the correction factor due to nonorthogonality becomes $[1-(2z-1)\delta^2+\cdots]$

rather than $[1+(2z-1)\delta^2+\cdots]$ in case (c) and remains of the same order of magnitude as before. These two cases, the symmetric and antisymmetric, must be regarded as the two extremes, and in actual problems the error is probably intermediate between that in two limits. If one makes an approximation similar to that of Inglis[1] of keeping only diagonal elements in an m_s system of representation (i.e., one in which each spin is space quantized separately), the calculations can be carried through as before even in the general case. The only difference is that J and δ^2 are replaced by qJ and $q\delta^2$, where q is the fraction of spins that are parallel. The correction factor due to nonorthogonality is then $[1+q(2z-1)\delta^2+\cdots]$. Because of the rejection of nondiagonal elements, this formula is surely quantitatively wrong, but seems to indicate that the error due to neglecting nonorthogonality is intermediate between that in the two extremes.

A somewhat better approximation for the general intermediate case is probably the following: assume that in each term $(-1)^k\delta^{2k}$ in the denominator and $(-1)^k\delta^{2k-2}$ in the numerator of (17) the factor $(-1)^k$ is to be replaced by the mean value of the product $\mathbf{P}_{(12)(34)\cdots(\cdots 2k)}$ of k simple permutations (each involving different cycles) for a state of spin S' for the entire system (crystal). This mean can be calculated with the vector model, or otherwise.[6] One finds that to a degree of approximation comparable with (22)

$$\overline{\mathbf{P}}_{(12)\cdots(\cdots 2k+2)}=x\overline{\mathbf{P}}_{(12)\cdots(\cdots 2k)}\times[1+2n^{-1}k(2+3x^{-1}+x^{-2})], \quad (25)$$

where
$$x=-\tfrac{1}{2}-2(S'/n)^2, \quad (26)$$

and that
$$\overline{W}=\tfrac{1}{2}nzxJ[1+f(x)+\cdots] \quad (27)$$

with
$$f(x)=(3z+zx^{-1}+x)\delta^2. \quad (28)$$

Eq. (27) agrees with (23) for the saturation case $S'=\frac{1}{2}n$, $x=-1$. The factor $\frac{1}{2}nzxJ$ is the usual Heisenberg mean value of W for a state of spin S'. The remaining factor $1+f(x)$ cannot be regarded as an accurate correction for nonorthogonality, as it is certainly not correct to replace the mean value of the quotient of two matrices by the quotient of the means. Nevertheless, a crude quantitative correction for nonorthogonality is probably obtained if we first calculate an eigenvalue W by means of (9), and then take the corrected eigenvalue to be $W[1+f(y)]$ with $y=W/\frac{1}{2}nzJ$. It is better to use y rather than x as an argument in (28), since then even in the correction term we allow somewhat for the spread in energy for states of a given S', in virtue of which W is not the same for a particular solution as the mean value.

The error involved in the approximation (14), which is probably less important than the correction studied in the preceding paragraph, can be avoided by using the secular determinant obtained from the second factor of (12) rather than from (9). After a solution of the resulting secular problem is obtained, the factor $1+f(y)$ explained in the preceding paragraph is to be inserted. Even then, there still remains the error due to the approximation (10–11), which seems to be comparable with the other errors (cf. end of appendix) and for which a quantitative allowance does not appear easy.

Note on the Evaluation of the Partition Function in Heisenberg's Theory of Ferromagnetism

In closing, we shall digress briefly to consider another question than nonorthogonality. In Heisenberg's theory of ferromagnetism, one is interested in the mean value of

$$\exp(-\mathbf{H}/kT)=\textstyle\sum_q(-\mathbf{H}/kT)^q/q! \quad (29)$$

for a state of given spin. Here the Hamiltonian matrix is $\mathbf{H}=\sum V_{(ij)}\mathbf{P}_{(ij)}$ with the usual neglect of nonorthogonality and higher order permutations. Even with these approximations, the mean value of (29) cannot be accurately evaluated. It occurs to us that instead of using Heisenberg's assumed Gaussian distribution,[2] a possible procedure is the following: To a degree of approximation comparable with (25) and (27), it can be shown by a calculation with the vector model,[6] that

$$\overline{\mathbf{H}^{q+1}}=xJ[\tfrac{1}{2}nz+q\alpha+O(k/n)]\overline{\mathbf{H}^q}, \quad (30)$$

where $\alpha=1+3x^{-1}+2x^{-2}$ and where x is defined

[6] We omit details of the calculation; for the general method see pp. 340–342 of my *Electric and Magnetic Susceptibilities*.

as in (26). Here $O(q/n)$ denotes error of order q/n. If we employ (30), the statistical average of the series (29) may be evaluated by use of the binomial theorem. Then the mean value of (29) becomes

$$\overline{\exp(-\mathbf{H}/kT)} = \exp(-nQ) \tag{31}$$

with

$$Q = \tfrac{1}{2} z\alpha^{-1} \log\left[1 + x\alpha k^{-1} T^{-1} J(1+f(x))\right]. \tag{32}$$

In writing Eq. (32), we have multiplied J by the factor $1+f(x)$ defined by (28) in order to make a crude correction for nonorthogonality suggested by the preceding section. If we set $f(x)=0$ and if we expand the logarithm in (32) in a Taylor's series in J, carrying the development only to terms of order J, the result is the same as when the spread of energies for a given S' is disregarded. If terms in J^2 are included, the formula is the same as with Heisenberg's Gaussian distribution. The magnetization associated with (31)–(32) is determined from the transcendental equation[7]

$$2S'/n = \tanh\left[(\beta H/kT) + (2S'/n)dQ/dx\right]. \tag{33}$$

Here H is the applied magnetic field, β is the Bohr magneton $he/4\pi mc$, and $2S'/n$ is the ratio of the intensity of magnetization $2S'\beta$ at temperature T to the true saturation magnetization $n\beta$ achieved at $T=0$. In order for the body to exhibit ferromagnetic behavior, i.e., in order for (33) to have a real positive root in $2S'/n$ of the proper character when $H=0$, it is necessary for dQ/dx to be greater than or equal to unity at $x=-\frac{1}{2}$. If we set $f(x)=0$, i.e., assume orthogonality, this condition can be fulfilled only if $x \geqq 16$, an impossibly high value. This result is disappointing, for (33) is presumably a better approximation than Heisenberg's formula based on a Gaussian distribution, where ferromagnetism is obtained if $z \geqq 8$. Hence we give (33) only passing mention. However, it must be noted that the critical condition for ferromagnetism is sensitive to the correction for nonorthogonality. If, for instance, in (28) we take $(z-\frac{1}{2})\delta^2 = 0.13$ then (33) admits a ferromagnetic solution when $z \geqq 10$. Ferromagnetism is allowed for $z=8$ if $\delta^2 \geqq 0.028$: such values for δ^2 seem rather high, but are by no means inconceivable. Thus the nonorthogonality effects may explain away the dilemma, but this statement is not beyond question, since use of the factor $1+f(x)$ is only a partial and inaccurate correction for nonorthogonality, as explained in the preceding section. More likely, the trouble arises from omission of the $O(k/n)$ term in (30).[8] The one safe conclusion is that the influence of nonorthogonality may be sufficiently important to render very uncertain any attempt to deduce the minimum number of neighbors necessary for ferromagnetism.

The writer has had valuable discussions with Professor J. C. Slater, Professor D. R. Inglis and Dr. H. M. James, for which he thanks them.

Appendix—Basis for Eq. (10)

The Hamiltonian function for the n electrons is

$$\mathbf{H} = \Sigma_i[(-h^2/8\pi^2 m)\nabla_i^2 - e^2/r_i^i] + \Sigma_{k>j} Q(k,j) + C, \tag{34}$$

with $\quad Q(k,j) = -e^2/r_k^j - e^2/r_j^k + e^2/r_{jk} + e^2/R^{jk},$

where r_k^j is the distance from nucleus j to the electron which is located on atom k. It is convenient for our purposes to label electrons according to the atom with which they are identified in the wave function Ψ defined in (6) to which $\mathbf{H}$ is applied. The terms e^2/r_{jk} are the interelectronic and e^2/R^{jk} the internuclear energy. We have assumed that each atom-ion exerts a Coulomb field on the electrons, but it is no trouble to adapt the analysis to the case of variable screening, wherein the atomic fields are central but not Coulomb. The bracketed part of (34) can immediately be dropped from consideration, as each factor in (6) may be assumed to satisfy an equation of the form

$$[(-h^2/8\pi^2 m)\nabla_i^2 - e^2/r_i^i - W_i]\psi_i(i) = 0,$$

and so the effect of this part of (34) can immediately be included by shifting the origin for the energy by an additive constant ΣW_i. This shift, and others, is covered by the additive constant C in (34) which, as stated after Eq. (6), is to be so chosen that the integral (3) vanishes for the identical permutation.

As a typical case, consider the value of (3) for a permutation (12)(34). The portion of (34) which we desire to retain to get (10) is $Q(1,2)+Q(3,4)$. We must show that the effect of the remaining Q terms in (34) is unimportant. First we note that there are an enormous number of the form $Q(k,j)$, with $k,j>4$. The contribution of the totality of these terms to the integral (3) is $\Delta_{(12)}\Delta_{(34)}U$ where

[7] The passage from (31) to (33) is made most easily by expanding Q in a Taylor's series about the most probable spin, as explained on p. 330 of the preceding reference.

[8] The omission of the term $O(k/n)$ in (22) is not nearly as serious as the corresponding omission in (30), for the series (17) encountered in connection with nonorthogonality are concerned with a correction effect, and need not be summed as accurately as (29).

$$U=\int\cdots\int\Sigma_{k>j>4}Q(k,j)dv_5\cdots dv_n.$$

Now the coefficient of $\mathbf{P}_{(12)(34)}$ in the part of (5) involving W explicitly is $-\Delta_{(12)(34)}W=-\Delta_{(12)}\Delta_{(34)}W$. Hence it is seen that the effect of the terms involving $k>j>4$ is covered simply by displacing the origin for the energy by an amount U, which we suppose done. (This is not quite the appropriate displacement in the case of higher order permutations. For instance, in the case of $\mathbf{P}_{(12)(34)(56)}$ one would wish to absorb in W only terms corresponding to $k>j>6$. However, except in the case of very high order permutations, which presumably involve very small integrals, the difference in the number of terms included is negligible in view of the fact that the number n of atoms is very large.)

It remains to be shown that we can disregard terms of the form

$$Q(3,1),\quad Q(4,1),\quad Q(3,2),\quad Q(4,2) \tag{35}$$

and $\quad Q(k,j)\quad$ where $\quad k>4\quad$ and $\quad j=1,2,3,4.\qquad$ (36)

The effect of each individual term in (35)–(36) is small in absolute magnitude compared with that of $Q(1,2)$ or $Q(3,4)$ and presumably differs from the latter by a factor of the order δ, where δ is defined as in (1). Namely, there are $2n-2$ factors in (3) which are large in regions where $Q(1,2)$ or $Q(3,4)$ are important, as is seen on substituting (6) in (3), specialized to $\mathbf{P}^{(i)}=\mathbf{P}_{(12)(34)}$ and remembering that each ψ function is large only on its own atom. On the other hand, only $2n-3$ factors are large where any term of the type (35) or (36) is important. Furthermore, if each atom is coupled only to its neighbors, the terms (36) outnumber $Q(1,2)$, $Q(3,4)$ only by a factor $2z$. Since z, the number of neighbors, is small (6, 8, or 12) for the various cubic arrangements, we are justified in neglecting (35), (36) if the interatomic distance is sufficiently large. From the foregoing considerations, it appears that the error due to dropping (35), (36) is of the order $2z\delta$ when measured relative to unity. This fact at first sight seems rather disquieting, as the errors we estimate due to other causes are of the order $2z\delta^2$ (cf., for instance, Eq. (23)). Thus the approximations (10) and (11) would appear to be the dominant causes of error, causing trouble unless δ is exceedingly small. However, one fortunately finds that the order of the error can be reduced to $2z\delta^2$ if in computing $V_{(12)}$, for instance, one takes the potential in the integrand of (3) to be not just $K(1,2)+\Sigma W_i+U+C$ but rather that

$$K(1,2)+\Sigma W_i+U+C\int\cdots\int\Sigma_{j>2}[K(j,1)+K(j,2)]|\psi_3|^2\cdots|\psi_n|^2dv_3\cdots dv_n \tag{37}$$

inclusive of contributions from the time exposure charge clouds of atoms other than 1, 2. Such a procedure is essentially the analog of the Hartree method. The error resulting from (35) may seem to still be of order δ rather than δ^2, but as a matter of fact is entirely inconsequential. Indeed for a given permutation $\mathbf{P}_{(12)(34)}$ it is altogether unlikely that 3 or 4 be a neighbor of 1 or 2, meaning that for the great bulk of permutations, terms of the type (35) are nonexistent so long as we are including only coupling between adjacent atoms. Hence the approximations (10) and (11) are no more serious sources of error than the other causes which we consider.

It is interesting to note that no internuclear terms e^2/R^{ik} contribute to (37), as their contribution to C just cancels their effect on the rest of (37). As already stated, the terms ΣW_i and U also have offsets in C.

In the preceding discussion we have considered for simplicity only fourth order permutations, but the argument can be readily extended to others.

FEBRUARY 1, 1936 PHYSICAL REVIEW VOLUME 49

Quantum-Mechanical Description

HENRY MARGENAU, *Sloane Physics Laboratory, Yale University*

(Received November 16, 1935)

It is pointed out that, by removing a single quantum-mechanical postulate commonly accepted, several conceptual difficulties in quantum-mechanical description can be eliminated.

THE discussion of a recent paper by Einstein, Podolski and Rosen[1] has brought to light an interesting divergence of opinions as to the meaning of reality. To the writer it seems, however, that the issue at stake is more than the merit or demerit of any particular conception of reality, and that a very fundamental point of quantum-mechanical axiology is involved. For Einstein, Podolski and Rosen have shown, by employing correctly the usual quantum-mechanical postulates, that the state of system 1 which, by hypothesis, is *isolated* from system 2, depends on the type of measurement performed on system 2. This, if true, is a most awkward *physical* situation, aside from any monstrous philosophical consequences it may have. We wish to show that, by the removal of a single postulate

[1] A. Einstein, B. Podolski and N. Rosen, Phys. Rev. **47**, 777 (1935); E. C. Kemble, ibid. **47**, 973 (1935); A. E. Ruark, ibid. **48**, 466 (1935); N. Bohr, ibid. **48**, 696 (1935).