REVIEWS OF
MODERN PHYSICS

VOLUME 44, NUMBER 3 JULY 1972

Properties and Uses of Natural Orbitals

ERNEST R. DAVIDSON
Chemistry Department, University of Washington, Seattle, Washington 98105

Natural orbitals are a particularly efficient choice of single-particle states for use in construction of electronic wave functions. This review discusses the general properties of the natural orbitals. Natural orbitals are compared to Hartree–Fock orbitals and relative merits are presented. Methods of determining and using natural orbitals in construction of electronic wave functions are presented.

CONTENTS

I. INTRODUCTION

Most theories of the electronic structure of systems with more than four electrons are based on the expansion of the wave function in a series of determinants (antisymmetrized products of one-particle functions). In any such theory there are three crucial steps: (1) selection of one-particle functions (called orbitals by chemists), (2) selection of determinants, and (3) calculation of expansion coefficients. Natural orbitals arise from a particular method for carrying out the first of these steps.

Natural spin orbitals were first defined by Löwdin (1955). Soon after that, Löwdin and Shull (1956) showed that natural orbitals could be used to express a two-electron wave function in a simple canonical form. This form required the fewest determinants in the expansion of the wave function. Shull and Löwdin (1955, 1959) applied this theory to the ground state of the helium atom. Shull (1959) converted most of the extant hydrogen molecule (H_2) ground-state wave functions to natural orbital form. This early work has inspired most of the interest in natural orbitals.

For several electron systems (3–100 electrons), the simplicity introduced for the two-electron system by use of natural orbitals does not appear. Only for certain restricted classes of wave functions are the number of determinants reduced. Nevertheless, there have been a steady stream of papers, beginning with that of Hurley (1960), reporting the natural orbitals for several-electron systems, Some of the earlier of these results have been reviewed by Smith (1968), McWeeny (1960), and Löwdin (1959).

The gross features of atomic and molecular structure and spectroscopy can be described with sufficient accuracy by a Hartree–Fock single-determinant wave function (or a simple linear combination of determinants with coefficients determined by symmetry). For such a wave function, the natural orbitals can be chosen to coincide with the Hartree–Fock self-consistent field (SCF) occupied orbitals so that nothing is gained by the introduction of natural orbitals. At the other extreme, a free electron gas wave function described in a plane-wave basis set will have the plane wave basis set as natural orbitals, so that again nothing is gained from natural orbitals.

The concept of natural orbitals is, however, extremely important if one wants better than SCF accuracy for atoms and molecules. The SCF error in the dissociation energy of a molecule or in the energy of a line in the spectrum of an atom is usually about 1–3 eV. Quite frequently this error is of the same magnitude as the energy change being considered so that an independent particle (SCF) description makes no useful prediction. Dipole moments of diatomic molecules tend to be wrong by ± 0.3 D in the SCF approximation. For large moments this is of no consequence, but for molecules such as CO, this error exceeds the dipole moment. Transition probabilities are also usually wrong by a factor of two at the SCF level, and the dipole derivatives entering the theory of vibrational spectroscopy maybe an order of magnitude in error. The electron spin resonance spectra of an atom such as boron or fluorine is expected to show no Fermi contact interaction in SCF theory so this whole effect is due to electron correlation.

For a system containing a large number of independent nonoverlapping atoms such as a dilute gas of helium atoms, the electronic wave function for a fixed configuration of the nuclei has nearly zero overlap with the SCF wave function [the overlap is $(0.996)^N$, where N is the number of helium atoms]. The concept of natural orbitals is useful in this context in explaining why the charge density is given so accurately by the SCF charge density.

The unique advantage of natural orbitals is that they reduce the number of determinants required for any fixed accuracy in the wave function compared with most other techniques for selecting orbitals [excepting multiconfiguration self-consistent-field (MCSCF) and psuedonatural orbital techniques]. The most important of these determinants makes possible a simple interpretation of correlation effects. This occurs because the natural orbitals are localized in the region of space where the correlation error is large.

The correlation error in the wave function is large where the wave function is large. While this fact may be regarded as trivially obvious, it seems to have been overlooked by many people. In many-body perturbation theory, theory of Van der Waals forces, as well as in qualitative discussions of correlation effects, there is a tradition of basing explanations on Rydberg and continuum orbitals which have their charge densities far from the region of error in the wave function. Use of localized correlation orbitals rather than continuum orbitals gives much more rapid convergence and greater physical insight. Use of SCF virtual (unoccupied) orbitals, while it has the advantage of making many-body perturbation theory possible, is really an inappropriate description of electron correlation in ground-state wave functions of atoms and molecules. For more than four electrons, wave functions built from natural orbitals are of much higher accuracy than wave functions obtained by conventional methods. This is precisely because the natural orbitals resemble particle-in-a-box wave functions for a box the size of the molecule (i.e., a box containing 99% of the charge density).

II. GENERAL PROPERTIES

Every particle may be described by a three-dimensional position vector $\mathbf{r}$, and a discrete spin variable ξ. For convenience, the four-vector $\mathbf{x}$ will be used for $(\mathbf{r}, \xi)$, and $\int d\mathbf{x}$ will be used for $\sum_\xi \int d\tau$. Then the first-order reduced density matrix may be defined (in Löwdin's normalization) as

$$D(\mathbf{x};\mathbf{x}')=N\int\Psi(\mathbf{x},\mathbf{x}_2,\cdots,\mathbf{x}_N)\times\Psi^*(\mathbf{x}';\mathbf{x}_2,\cdots,\mathbf{x}_N)\,d\mathbf{x}_2\cdots d\mathbf{x}_N. \quad (1)$$

The motive for defining this quantity comes from Hartree–Fock theory and from the analogy between the equations for the average values of an operator in quantum theory and in classical electricity. For a closed-shell atom or molecule, the single determinant built from orthonormal orbitals which give the lowest energy contains orbitals obeying the equation

$$F\phi_i=\epsilon_i\phi_i,$$

where the Fock operator F is known to be

$$F=h+\mathcal{J}-\mathcal{K}.$$

The operator h includes the kinetic energy and the one-body nuclear attraction forces acting on the electron. The operator $\mathcal{J}$ is defined by

$$\mathcal{J}(\mathbf{r}_1)=\int\rho(\mathbf{r}_2)r_{12}^{-1}\,d\tau_2,$$

where

$$\rho(\mathbf{r}_2)=\sum_{\xi_2}\sum_{i=1}^{N}\phi_i(\mathbf{x}_2)\phi_i^*(\mathbf{x}_2),$$

and N is the number of electrons.

Similarly $\mathcal{K}$ is defined by

$$\mathcal{K}(\mathbf{x}_1)f(\mathbf{x}_1)=\int\rho(\mathbf{x}_1;\mathbf{x}_2)f(\mathbf{x}_2)\,d\mathbf{x}_2,$$

where

$$\rho(\mathbf{x}_1;\mathbf{x}_2)=\sum_{i=1}^{N}\phi_i(\mathbf{x}_1)\phi_i^*(\mathbf{x}_2).$$

By analogy to classical electricity, $\rho(\mathbf{r})$ may be regarded as the charge density in the molecule, and $\mathcal{J}$ is the total electrostatic potential of this charge density. The exchange operator $\mathcal{K}$ is nonclassical. Fortunately $\mathcal{K}$ is small except for one term when acting on any ϕ_i. Within qualitative accuracy, we have

$$\begin{aligned}\mathcal{K}\phi_j&\approx\mathcal{J}_j\phi_j &&(j\leq N),\\ &\approx 0 &&(j>N),\end{aligned}$$

where $\mathcal{J}_j$ is the electrostatic potential due to the charge density $\sum_\xi|\phi_j|^2$. Hence the occupied ϕ_j almost satisfy the intuitively reasonable Hartree conditions

$$F_j\phi\simeq\epsilon_j\phi_j \qquad (j\leq N),$$

where

$$F_j=h+\mathcal{J}-\mathcal{J}_j.$$

The virtual (uncoupied) orbitals, on the other hand, nearly satisfy

$$F\phi_k\approx\epsilon_k\phi_k \qquad k>N,$$

where

$$F=h+\mathcal{J}.$$

This is intuitively wrong since the virtual orbitals in this definition are calculated in the field of N instead of $N-1$ electrons. Modified virtual orbitals calculated from F_N are frequently used as these show a Rydberg series due to the net $+1$ charge of the potential used in F_N.

The expression $\rho(\mathbf{x}_1;\mathbf{x}_2)$ can be shown to be equivalent to Eq. (1) if a single determinant is substituted for Ψ. If one wishes to evaluate the average value of any one-body operator

$$\mathcal{O}=\sum_{i=1}^{N}o(i),$$

$$\langle\mathcal{O}\rangle=\sum_{i=1}^{N}\int\Psi^*o(i)\Psi\,d\mathbf{x}_1d\mathbf{x}_2\cdots d\mathbf{x}_N$$

can be simplified by the change of variables $\mathbf{x}_1 \leftrightarrow \mathbf{x}_i$ to

$$\langle \mathcal{O} \rangle = \sum_{i=1}^{N} \int \Psi^*(i, 2, 3, \cdots, 1, \cdots, N)\, o(1) \times \Psi(i, 2, 3, \cdots, 1, \cdots, N)\, d\mathbf{x}_1 \cdots d\mathbf{x}_N.$$

Since

$$\Psi(i, 2, 3, \cdots, 1, \cdots, N) = -\Psi(1, 2, \cdots, i, \cdots, N),$$

we have

$$\langle \mathcal{O} \rangle = N \int [o(1)\Psi(1, \cdots, N)] \Psi^*(1, \cdots, N)\, d\mathbf{x}_1 \cdots d\mathbf{x}_N.$$

Now all integrations $dx_2 \cdots dx_N$ can be carried out once and for all (independent of o) if one keeps track of which x_1 comes from Ψ, and which from Ψ^*. This is done in Eq. (1). Using this equation, we have

$$\langle \mathcal{O} \rangle = \int o(\mathbf{x}) D(\mathbf{x}; \mathbf{x}')\big|_{x'=x}\, d\mathbf{x}.$$

Note that if o is a spin-free multiplicative operator $o(\mathbf{r})$, then we find

$$\langle \mathcal{O} \rangle = \int o(\mathbf{r}) \rho(\mathbf{r})\, d\tau,$$

where the charge density is just $\sum_\xi D(\mathbf{x}, \mathbf{x})$.

In terms of the field annihilation operator $\Phi(x)$, and the creation operator $\Phi^\dagger(\mathbf{x}')$, Eq. (1) is simply

$$D(\mathbf{x}; \mathbf{x}') = \langle \Psi \mid \Phi^\dagger(\mathbf{x}') \Phi(\mathbf{x}) \mid \Psi \rangle. \tag{2}$$

If Φ is expanded in an orthonormal set of orbitals $g_j(\mathbf{x})$ with associated annhilation operators a_j as

$$\Phi(\mathbf{x}) = \sum_j g_j(\mathbf{x}) a_j, \tag{3}$$

then we find

$$D(\mathbf{x}; \mathbf{x}') = \sum g_j(\mathbf{x}) g_k^*(\mathbf{x}') \langle \Psi \mid a_k^\dagger a_j \mid \Psi \rangle, \tag{4}$$

or

$$D(\mathbf{x}; \mathbf{x}') = \sum d_{jk} g_j(\mathbf{x}) g_k^*(\mathbf{x}'). \tag{5}$$

If g is any arbitrary orbital, and $a^\dagger$ is the associated creation operator, the expected number of particles in orbital g (the "occupation number" of g) is just

$$N_g = \langle \Psi \mid a^\dagger a \mid \Psi \rangle = \int g^*(\mathbf{x}) D(\mathbf{x}; \mathbf{x}') g(\mathbf{x}')\, d\mathbf{x} d\mathbf{x}'. \tag{6}$$

The sum of N_g over any complete orthonormal set is the trace of D (which by definition is N).

Now suppose D is regarded as the kernel of an integral operator $\mathfrak{D}$,

$$\mathfrak{D} f = \int D(\mathbf{x}; \mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}'. \tag{7}$$

From this viewpoint $\mathfrak{D}$ is easily seen to be Hermitian and nonnegative. The rank r of $\mathfrak{D}$, defined as the number of nonzero eigenvalues, may be finite or infinite. The eigenfunctions χ_i of $\mathfrak{D}$, defined by

$$\mathfrak{D} \chi_i = \lambda_i \chi_i, \tag{8}$$

are called the natural spin orbitals (NSO's) of Ψ. Clearly λ_i is just the occupation number of χ_i, and hence

$$0 \le \lambda_i \le 1. \tag{9}$$

In the following discussion, it will be assumed that the χ_i are numbered in order of decreasing occupation ($\lambda_i \ge \lambda_{i+1}$).

The χ_i defined in this way have certain optimum properties which give rise to their utility. Coleman (1963) and Kobe (1969) have recently reviewed many of these properties. If Ψ is constructed from a finite number R of basis orbitals, then $r \le R$, and each χ_i is a linear combination of the basis orbitals. Conversely, if r is finite, Ψ may be rewritten exactly using only the χ_i with nonzero λ_i.

The NSO's offer the most rapidly convergent series approximation to D. That is to say, among all expressions of the form

$$D' = \sum_{i=1}^{K} \sum_{j=1}^{L} b_{ij} f_i(\mathbf{x}) h_j(\mathbf{x}')$$

for fixed K, L, the series

$$D' = \sum_{i=1}^{\inf(K,L)} \lambda_i \chi_i(\mathbf{x}) \chi_i^*(\mathbf{x}') \tag{10}$$

minimizes the least-squares error,

$$\int \mid D - D' \mid^2 d\mathbf{x} d\mathbf{x}'.$$

As a consequence, for natural orbitals, the total occupation of the first K orbitals is a maximum, and the net occupation for all orbitals beyond the first K is a minimum. If a pseudoFermi surface is defined by $K = N$, this set of orbitals gives the lowest net number of particles above the Fermi surface.

The natural orbitals simplify the formula for the expectation value of any one-electron operator $\mathcal{O}$. Instead of

$$\langle \mathcal{O} \rangle = \sum_{ik} d_{jk} (g_j, o g_k) = \int o(\mathbf{x}) D(\mathbf{x}; \mathbf{x}')\big|_{\mathbf{x}=\mathbf{x}'}\, d\mathbf{x},$$

one has

$$\langle \mathcal{O} \rangle = \sum_i \lambda_i (\chi_i, o \chi_i). \tag{11}$$

This diagonal form for $\mathcal{O}$ prevails for all $\mathcal{O}$, whereas use of the eigenfunctions of $\mathcal{O}$ would make only that one $\mathcal{O}$ diagonal. In particular it should be noted that the expectation value of the Fock operator is diagonal in either Hartree–Fock or natural orbitals.

The natural orbitals also give the most rapidly convergent approximation to Ψ. That is to say, among all expressions of the form

$$\Psi' = \sum_{i=1}^{K} \sum_{j=1}^{L} a_{ij} f_i(\mathbf{x}_1) F_j(\mathbf{x}_2 \cdots \mathbf{x}_N)$$

for fixed K, L, the series

$$\Psi' = \sum_{i=1}^{\inf(K,L)} \chi_i(\mathbf{x}_1) \int \chi_i^*(\mathbf{x}_1) \Psi(\mathbf{x}_1 \cdots \mathbf{x}_N)\, d\mathbf{x}_1 \tag{12}$$

minimizes the least-squares error

$$\int \mid \Psi - \Psi' \mid^2 d\mathbf{x}_1 \cdots d\mathbf{x}_N. \tag{13}$$

Also, if Ψ is expanded in a set of determinants as

$$\Psi = \sum C_K \Psi_K, \tag{14}$$

where

$$\Psi_K = (N!)^{-1/2} \operatorname{Det}\{g_{k_1}, \cdots, g_{k_n}\},$$

$$K = \{k_1 < k_2 < \cdots < k_n\}, \tag{15}$$

then the occupation number of g_k is

$$N_k = \sum_{\{K|k\epsilon K\}} |C_K|^2. \tag{16}$$

Thus, becuase the χ_i have maximum occupation numbers, each successive χ_i makes a maximum contribution to the wave function in the sense that the sum of the squares of all coefficients involving that χ_i is a maximum. This does not mean that the least-squares error using only $p<r$ of the NSO's is a minimum since χ_i does not reach its maximum contribution λ_i until all determinants (and hence all other orbitals) are included.

None of these ways in which the NSO's form an optimum set of orbitals deals directly with the crucial ways in which a set of orbitals should be optimum. A good set of orbitals should (a) be easy to construct, (b) be complete enough to give the desired accuracy for physical properties of interest, (c) facilitate selection of the important determinants, (d) facilitate calculation of the expansion of the wave function in those determinants, and (e) facilitate physical interpretation of the results. In the following discussion an attempt will be made to show how the natural orbitals succeed in some of these respects and fail in others.

A. Perturbation Comparison of SCF and NSO

The effect of electron correlation on the density matrix has been discussed extensively (see, for example, McWeeney, 1960; Hirschfelder, Brown, and Epstein, 1964; Davidson, 1968; March, Young, and Sampanthar, 1967). The effect on the resulting NSO's is not so widely recognized (Hirschfelder, Brown, and Epstein, 1964; Shull and Lowdin, 1955). If ϵ is some relevant measure of the size of the perturbation effect where Ψ_{SCF} (Hartree–Fock self-consistent field wave function) is used as Ψ_0, then the wave function may be expressed in a perturbation expansion as

$$\Psi = \Psi_{\mathrm{SCF}} + \epsilon\Psi^{(1)} + \epsilon^2\Psi^{(2)} + \cdots. \tag{17}$$

By Brillouin's theorem, $\Psi^{(1)}$ consists only of double excitations from Ψ_{SCF}, while $\Psi^{(2)}$ includes all excitations up through quadrupole.

The density matrix arising from this expansion to second order is

$$D = D_{\mathrm{SCF}} + \epsilon^2 D^{(2)}, \tag{18}$$

where

$$D^{(2)} = \langle\Psi^{(1)} \mid \Phi^\dagger\Phi \mid \Psi^{(1)}\rangle + \langle\Psi_{\mathrm{SCF}} \mid \Phi^\dagger\Phi \mid \Psi_s^{(2)}\rangle + \langle\Psi_s^{(2)} \mid \Phi^\dagger\Phi \mid \Psi_{\mathrm{SCF}}\rangle,$$

and $\Psi_s^{(2)}$ represents the singly excited part of $\Psi^{(2)}$. The matrix elements d_{ij} of D are given to second order by

$$\begin{aligned} d_{ij} &= 1 + \epsilon^2\langle\Psi^{(1)} \mid a_j^\dagger a_i \mid \Psi^{(1)}\rangle \qquad i,j\leq N \qquad &(19a)\\ &= \epsilon^2\langle\Psi^{(1)} \mid a_j^\dagger a_i \mid \Psi^{(1)}\rangle \qquad i,j>N \qquad &(19b)\\ &= \epsilon^2\langle\Psi_{\mathrm{SCF}} \mid a_j^\dagger a_i \mid \Psi_s(2)\rangle \qquad j\leq N,\ i>N. \qquad &(19c)\end{aligned}$$

The eigenvectors of D give the coefficients for the expansion of the NSO's in the SCF orbital basis. Since D to zeroth order is degenerate (diagonal with N diagonal elements equal to unity and the rest zero), degenerate perturbation theory must be used to find its eigenvectors. Hence the correct zeroth order NSO's are not the SCF orbitals. The zeroth-order transformation is determined entirely by the the first-order correlation correction to the wave function. Such an effect is not surprising since the SCF wave function is invariant under a unitary transformation of the occupied or virtual orbitals (as long as the two sets are not mixed). It should be noted, however, that the Fock operator is nondiagonal in the NSO basis, and hence perturbation theories which assume that single determinants almost represent true excited states may fail to work in the natural orbital basis.

The first N zeroth order NSO's have occupation numbers within ϵ^2 of unity and are obtained by a unitary transformation of the occupied SCF orbitals. The remaining zeroth order NSO's have occupation numbers of magnitude ϵ^2 and are a transformation of the virtual SCF orbitals. In practice it is found that, for atoms and small molecules, this transformation converts the continuum virtual orbitals into a discrete set of quite localized natural orbitals. There is no resemblance at all between any one of the virtual orbitals and the NSO's. In addition to this large zeroth-order effect, there will be a further change of the same type in first order due to the higher-order terms in the wave function. Thus, approximate natural orbitals for the exact ground state of a system are difficult to compute because they are quite sensitive to correlation effects.

The coupling of occupied and virtual orbitals occurs only in second order. The determinant using the first N natural spin orbitals is called the first natural configuration, Ψ_{FNC}. This determinant differs only in second order from Ψ_{SCF}. Further, this second-order difference is exactly $\Psi_s^{(2)}$ so that singly excited (particle–hole) corrections to Ψ_{FNC} first appear in third order.

B. Spin Dependence

The spin dependence of the NSO's is relatively simple if Ψ is an eigenfunction of $\mathcal{S}^2$ and $\mathcal{S}_z$, with eigenvalues $S(S+1)$ and M. In this case D may be written as

$$D = 1/2[\alpha(\zeta)\alpha^*(\zeta') + \beta(\zeta)\beta^*(\zeta')]\rho(\mathbf{r},\mathbf{r}') + M[\alpha(\zeta)\alpha^*(\zeta') - \beta(\zeta)\beta^*(\zeta')]\gamma(\mathbf{r},\mathbf{r}'), \tag{20}$$

where α and β are the usual spin eigenfunctions for spin one-half, ρ is the charge density matrix, and γ is the spin density matrix [See McWeeny (1960) for a discussion of this term]. It is easily shown that

$$\mathrm{Tr}\,\rho = \int d\tau\rho(\mathbf{r},\mathbf{r}) = N = N_\alpha + N_\beta, \tag{21}$$

$$\mathrm{Tr}\,\gamma = \int d\tau\gamma(\mathbf{r},\mathbf{r}) = 1. \tag{22}$$

Bingel (1960) and Kutzelnigg (1963a, b, c) have also shown that ρ and γ are independent of M within a set of states $\Psi_{S,M}$ connected by $\mathcal{S}_\pm$.

The natural spin orbitals of $\Psi_{S,M}$ can be chosen to be eigenfunctions of $\mathcal{S}_z$. The spatial dependence of the NSO's is then computed from the integral kernel $\rho+2M\gamma$ for α spin, and $\rho-2M\gamma$ for β spin. Only if $M=0$, will the space parts of the α NSO's coincide with the space part of the β NSO's. For many purposes this is an inconvenience. A still greater inconvenience is that the NSO's will be different for each value of M so that the simple relation between $\Psi_{S,M}$ and $\Psi_{S,M\pm1}$ will be completely obscured.

For these reasons, most workers tend to use natural orbitals (NO's) rather than NSO's for open shell systems. Natural orbitals may be defined as α or β times the eigenfunctions $g_i(\mathbf{r})$ of ρ,

$$\int \rho(\mathbf{r},\mathbf{r}')g_i(\mathbf{r}')\,d\tau'=\mu_i g_i(\mathbf{r}). \tag{23}$$

Clearly the g_i give the best approximation to ρ since

$$\rho=\sum \mu_i g_i(\mathbf{r})g_i^*(\mathbf{r}'). \tag{24}$$

C. Symmetry Properties

There have been many papers on the symmetry properties of density matrices and natural orbitals (Bingel, 1960, 1962, 1970; Bingel and Kutzelnigg, 1968, 1970; McWeeny and Kutzelnigg, 1968). The basic result is that, if Ψ beongs to a nondegenerate irreducible representation of a point group, then the NO's and the NSO's can be chosen to be symmetry adapted. That is, they can be chosen to belong to definite irreducible representations. All partner functions for a natural orbital in a degenerate representation will also be natural orbitals with equal occupation numbers.

If Ψ is degenerate due to symmetry, then the NO's generally cannot be symmetry adapted. Either they will individually be of mixed symmetry, or else the partner functions of a natural orbital in a degenerate representation will not also be natural orbitals. For example, in the $1s2p\,^3P$ state of helium with $m_l=1$, and $2p_{-1}$ and $2p_0$ orbitals related to the $2p_1$ natural orbital through L_- are not themselves natural orbitals.

Most practical calculations are done in the spirit of the Hartree atomic calculations with forced equivalence of partner orbitals in degenerate representations. This is partly for computational advantage as use of different radial parts for $2p_1$, $2p_0$, and $2p_{-1}$ would complicate construction of L^2 eigenfunctions. As with spin equivalence, there also is a conceptual advantage to using the same set of orbitals for each Ψ in a degenerate family of states.

The appropriate symmetry-constrained natural orbitals (SCNO's) are generated by diagonalizing a symmetrized density matrix. This matrix may be regarded, equivalently, as the totally symmetric part of D for any one of the degenerate Ψ or as the ensemble average of D over all of the degenerate Ψ. Although it is often not explicitly mentioned, many open shell calculations actually use SCNO's and refer to them as NO's or NSO's (Bender and Davidson, 1969a).

For the free electron gas, each plane wave state corresponds to a different irreducible representation. Hence for the free electron gas, the plane wave states are natural orbitals. This important result shows that natural orbitals have no particular advantage over SCF orbitals for infinite uniform systems.

III. TWO-ELECTRON WAVE FUNCTIONS

Wave functions for two-electron systems are an important special case in the theory of natural orbitals because the wave function assumes a simplified form. If Ψ is a two-electron wave function, then the minimum of

$$\int d\mathbf{x}_1 d\mathbf{x}_2 \,\Big|\, \Psi-2^{-1/2}\sum_{i,j}^{K,L}\mu_{ij}f_i(\mathbf{x}_1)t_j^*(\mathbf{x}_2)\,\Big|^2$$

is achieved if (Carlson and Keller, 1961)

$$\sqrt{2}\int \Psi^*(\mathbf{x}_1,\mathbf{x}_2)f_i(\mathbf{x}_1)\,d\mathbf{x}_1=\mu_i^*t_k(\mathbf{x}_2), \tag{25a}$$

$$\sqrt{2}\int \Psi(\mathbf{x}_1,\mathbf{x}_2)t_i(\mathbf{x}_2)\,d\mathbf{x}_2=\mu_i f_i(\mathbf{x}_1), \tag{25b}$$

and

$$\mu_{ij}=0 \qquad (i\neq j). \tag{26}$$

In this case, Ψ is expressable as

$$\Psi=\sum \mu_i f_i(\mathbf{x}_1)t_i^*(\mathbf{x}_2).$$

It is this diagonalization of Ψ which makes natural orbitals particularly appropriate for two-electron systems.

Substitution of Eq. (25a) into Eq. (25b) yields

$$\int D(\mathbf{x}_1;\mathbf{x}_1')f_i(\mathbf{x}_1')\,d\mathbf{x}_1'=|\,\mu_i\,|^2 f_i(\mathbf{x}_1), \tag{28}$$

in verification of Eq. (12). Similarly, we find

$$\int D(\mathbf{x}_1;\mathbf{x}_1')t_i(\mathbf{x}_1')\,d\mathbf{x}_1'=|\,\mu_i\,|^2 t_i(\mathbf{x}_1') \tag{29}$$

so that both $\{f_i\}$ and $\{t_i\}$ are sets of NSO's. In the usual event that some of the $|\,\mu_i\,|^2$ are degenerate, the NSO's are somewhat arbitrary. Any choice of the χ_i may be used for the f_i provided the corresponding t_i are computed from (25a). Generally the t_i will correspond to a different choice for the natural orbitals and will not coincide with the f_i.

For example, for the SCF approximation to the ground state of helium,

$$\Psi=1s(1)1s(2)(\alpha_1\beta_2-\beta_1\alpha_2)/\sqrt{2},$$

the χ_i are best chosen to be $1s\alpha$ and $1s\beta$. But then we have $t_2=f_1=\chi_1$ and $t_1=f_2=\chi_2$. This choice is not unique, however, since $t_2=-i\chi_1$ is equally acceptable (the phases of f_i and t_i are arbitrary) as is $t_1=f_1=\chi_1=1s(\alpha+i\beta)/\sqrt{2}$ and $t_2=f_2=\chi_2=1s(\alpha-i\beta)/\sqrt{2}$. In most cases there is a convenient choice of the f_i which gives maximum conceptual simplicity.

A. Construction of Natural Orbitals

There are three main methods of obtaining natural orbitals for two-electron systems. The first of these is to construct a wave function as a finite linear combination of determinants built from an arbitrary basis set and then to solve Eqs. (28) and (25a) for the exact natural orbitals, $f_i=\chi_i$, and cofactors t_i.

For wave functions containing the interparticle coordinate r_{12} explicitly, the rank is infinite and this simple approach will not work (Nazaroff, 1968). Instead, the natural orbitals may be expanded in an arbitrary finite basis set. If this expansion is substituted into the expression for the error in approximating Ψ and the coefficients are varied, the result is a finite matrix approximation to Eq. (25),

$$\phi^T(\mathbf{x})=(\phi_1,\phi_2,\cdots), \tag{30}$$

$$f_i(\mathbf{x})=\sum f_{ji}\phi_j(\mathbf{x})=f_i^T\phi(\mathbf{x}), \tag{31}$$

$$t_i(\mathbf{x})=\sum t_{ji}\phi_j(\mathbf{x})=t_i^T\phi(\mathbf{x}), \tag{32}$$

$$S=\int\phi\phi^T\,d\tau, \tag{33}$$

$$\Psi=\int\phi^*(\mathbf{x}_1)\Psi(\mathbf{x}_1,\mathbf{x}_2)\phi^T(\mathbf{x}_2)\,d\mathbf{x}_1d\mathbf{x}_2, \tag{34}$$

$$\sqrt{2}\Psi^\dagger f_i=\mu_i^*St_i, \tag{35a}$$

$$\sqrt{2}\Psi t_i=\mu_iSf_i. \tag{35b}$$

The first of these methods gives exact natural orbitals for what are usually relatively crude wavefunctions. The second method gives approximate natural orbitals for much better wave functions. Both methods fail to give any insight into the basic physics involved.

A third approach is to find a differential equation for the natural orbitals which does not explicitly involve the wave function. Löwdin (1955), Kutzelnigg (1963a), Reid and Ohrn (1963), and Nazaroff (1968) discuss coupled integrodifferential equations for the NO's. Linderberg (1964) and Reinhardt and Doll (1969) have discussed the direct calculation of natural orbitals by many-body perturbation theory. Of these approaches, only Kutzelnigg has pursued his procedure to obtain results of meaningful accuracy.

An approach similar to that used by Kutzelnigg probably makes the basic physics clearest. For a closed shell ground state of a two-electron system, Ψ can be written as

$$\Psi=\sum\gamma_ig_i(\mathbf{r}_1)g_i^*(\mathbf{r}_2)(\alpha_1\beta_2-\beta_1\alpha_2)/\sqrt{2}, \tag{36}$$

where the γ_i are real and $|\gamma_i|\geq|\gamma_{i+1}|$. For atomic ions, γ_1 is nearly unity while the remaining γ_i are small and negative.

If the Hamiltonian is written as

$$H=h(1)+h(2)+r_{12}^{-1}, \tag{37}$$

then we have

$$E=\langle\Psi\mid H\mid\Psi\rangle=\sum\gamma_i\gamma_j[2h_{ii}\delta_{ij}+K_{ij}],$$

where

$$h_{ij}=\langle g_i\mid h\mid g_j\rangle,$$

and

$$K_{ij}=\int g_i^*(1)g_j(1)r_{12}^{-1}g_j^*(2)g_i(2)\,d\tau_1d\tau_2.$$

Variation of E with respect to g_i subject to orthonormality constraints gives

$$\gamma_k^2G_kg_k=\sum_i\lambda_{ik}g_i, \tag{38}$$

where

$$G_k=h+\sum_i\gamma_i\gamma_k^{-1}\mathcal{K}_i, \tag{39}$$

$$\mathcal{K}_ig_k=\int d\tau_2g_i^*(2)g_i(1)r_{12}^{-1}g_k(2), \tag{40}$$

and the λ_{ij} are Lagrangian multipliers. If the λ_{ij} are eliminated, these equations may be written as

$$\mathcal{G}_kg_k=\epsilon_kg_k, \tag{41}$$

where

$$\mathcal{G}_k=P_k\{G_k-Q_k\}P_k, \tag{42}$$

$$P_k=1-\sum_{i<k}|\,g_i\rangle\langle g_i\,|, \tag{43}$$

$$Q_k=\gamma_k^{-2}\sum_{i>k}\gamma_i^2\{|\,g_i\rangle\langle g_i\mid G_i+G_i\mid g_i\rangle\langle g_i\,|\}, \tag{44}$$

$$E=\sum\gamma_k^2(\epsilon_k+h_{kk}), \tag{45}$$

$$\epsilon_k=\langle g_k\mid\mathcal{G}_k\mid g_k\rangle=\langle g_k\mid G_k\mid g_k\rangle. \tag{46}$$

Thus each g_i is the eigenfunction of a different one-body operator $\mathcal{G}_k$. The operator P_k preserves orthogonality of g_k to all g_i $(i<k)$, and the small correction Q_k modifies the effective potential to allow for the relaxation in g_k caused by the g_i $(i>k)$.

For γ_1 near unity, $\mathcal{G}_k$ can be conceptually approximated by

$$\mathcal{G}_k\approx P_k\{h+\gamma_1\gamma_k^{-1}\mathcal{K}_1\}P_k \qquad \text{for } k>1 \tag{47}$$

$$\approx h+\mathcal{K}_1 \qquad \text{for } k=1. \tag{48}$$

Hence to the lowest order in γ_i/γ_1, g_1 is just the SCF orbital. Variation of the γ_i and expansion to lowest order in γ_i/γ_1 gives,

$$\gamma_i\approx-K_{1i}(\epsilon_i-\epsilon_1)^{-1} \qquad \text{for } i>1 \tag{49}$$

so γ_i is negative. Hence the g_k $(k>1)$ are bound not only by the nuclear potential in h, but also by the very negative exchange potential

$$\mathcal{G}_k\approx h-(\epsilon_k-\epsilon_1)K_{1k}^{-1}\mathcal{K}_1. \tag{50}$$

As k increases so does ϵ_k. Hence the potential felt by the higher natural orbitals is much more attractive than that for the lower ones. As a consequence, the NO's remain strongly localized in the region of space occupied by the SCF orbital. Since the exchange term in $\mathcal{G}_k$ dominates for large k, the higher natural orbitals have no resemblance to the eigenfunctions of h or of $\mathcal{G}_1$. Because the $\mathcal{G}_k$ are so different from one another, the effect of P_k is also large and the eigenfunctions of G_k are very different from those of $\mathcal{G}_k$.

An interesting mathematical question regarding completeness can be raised at this point. There is no particular physical reason to believe that the natural orbitals for the ground state of helium form a complete set of functions. The form of the equations for the natural orbitals does not obviously lead to a complete set since each orbital is taken from a different equation. The general appearance of the known approximate natural orbitals does not indicate that they would be complete since they remain quite localized instead of spreading over all of space as k increases.

The difference between g_1 and g_{SCF} has been discussed by several authors (Nazaroff and Hirschfelder, 1963). From Eq. (42), to the first order in γ_i/γ_1, we have

$$\{h+\mathcal{K}_1+\sum_{i>1}\gamma_i\mathcal{K}_i\}g_1=\epsilon_1 g_1. \tag{51}$$

Since $\gamma_i\mathcal{K}_i$ is most negative near the nucleus, g_1 would be expected to be contracted relative to g_{SCF}. This effect is small, however, as g_1 and g_{SCF} have an overlap larger than 0.9999 for helium.

Figure 1 illustrates these results for helium. The $1s$ SCF orbital is indistinguishable on this scale from the $1s$ NSO. The $2s$ orbital from the $1s2s\,^1S$ Rydberg state and the $2p$ orbital from the $1s2p\,^1P$ state are shown for comparison with the first s- and p-type correlation NSO's. This illustrates the localization of the NSO's in the regions where $1s^2$ is largest. Clearly even the lowest Rydberg orbitals have a small overlap with this region.

B. Results for Helium

There have been numerous natural orbital calculations of the ground state of the helium isoelectronic series. For this state, Ψ may be written as

$$\Psi=\sum_{i,l}\gamma_{i,l}\sum_{m}g_{i,l,m}(1)g_{i,l,m}(2)(\alpha_1\beta_2-\beta_1\alpha_2)/\sqrt{2}, \tag{52}$$

where

$$g_{i,l,m}=\mathcal{R}_{i,l}(r)Y_{l,m}(\theta,\phi). \tag{53}$$

Table I gives the values of some of the $\gamma_{i,1}$ from approximate calculations. The results of Davidson (1963) are based on the second method mentioned above and are in good agreement with the results Banyard and Baker (1969) obtained from the first method. The early results of Shull and Löwdin (1955, 1959) are based on the first method applied to a much less complete wave function. The results of Kutzelnigg (1963b), Ahlrichs, Kutzelnigg, and Bingel (1966a), and Cressy, Miller, and Ruedenberg (1969) are surprisingly quite different even though their energy is close to the correct answer. This difference seems to arise because they applied the third method but with the approximations of Eqs. (47)–(50), rather than the exact Eqs. (41)–(45).

Table I. Coefficients in the natural expansion of helium.

	Davidson (1963)	Ahlrichs, Kutzelnigg, and Bingel (1966a)	Shull and Löwdin (1959)	Banyard and Baker (1969)
$1s$	0.99599	0.99622	0.99595	0.99598
$1p$	−0.03563	−0.03467	−0.03603	−0.03574
$2s$	−0.06148	−0.06003	−0.06163	−0.06163
$1d$	−0.00566	−0.00545	−0.00597	−0.00566
$2p$	−0.00638	−0.00552	−0.00652	−0.00643
$3s$	−0.00786	−0.00681	−0.00747	−0.00790
$1f$	−0.00169	−0.00161	−0.00133	−0.00169
$2d$	−0.00178	−0.00148	−0.00130	−0.00174
$3p$	−0.00180	−0.00134	−0.00164	−0.00189
$4s$	−0.00197	−0.00144	−0.00216	−0.00192

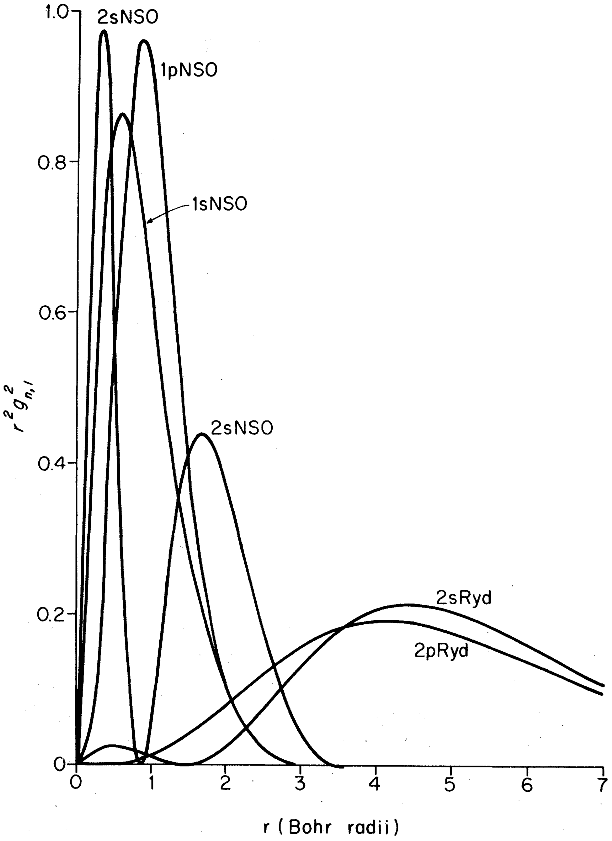

Fig. 1. Radial dependence of the natural orbitals and Rydberg orbitals of helium.

For other ions (see Brown and Nazaroff, 1967) with nuclear charge Z, the coefficients $\gamma_{i,1}$, except for the first one, are nearly proportional to Z^{-1}. Wang (1970) has reported the results of expanding the g_i and γ_i in powers of Z^{-1}. These results compare favorably with the calculations of Ahlrichs, Kutzelnigg, and Bingel (1966a) for $Z=1$–8.

The correlation effects introduced by the various $g_{i,l,m}$ in Eq. (52) may be visualized by considering their effect on the two-particle distribution function. To first order in γ_i/γ_1, the spatial distribution is just

$$P(\mathbf{r}_1,\mathbf{r}_2)=g_{1,0}{}^2(\mathbf{r}_1)g_{1,0}{}^2(\mathbf{r}_2)f, \tag{54}$$

where the correlation factor f is given by

$$f=1-2(2l+1)\sum_{i,l}|\gamma_{i,l}|(g_{i,l}/g_{1,0})_{r_1}(g_{i,l}/g_{1,0})_{r_2}\times P_l(\cos\theta_{12}). \tag{55}$$

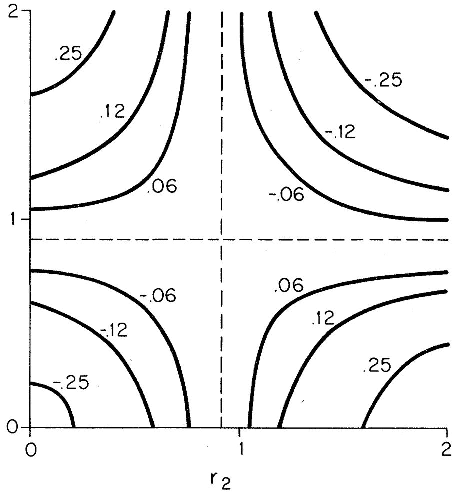

FIG. 2. Contribution of $g_{2,0}(r)$ to the in–out correlation. Plot shows contours of constant $-2\,|\,\gamma\,|\,(g_{2,0}/g_{1,0})_{r_1}(g_{2,0}/g_{1,0})_{r_2}$.

For $l=0$, the dominant s-type correlation has a radial node near the r value for which $r^2g_{1,0}^2(r)$ is a maximum. If r_1 and r_2 are both inside (or both outside) the node, this term makes a negative contribution to f, while if the electrons are on opposite sides of the nodal surface, f is increased. For this reason, this term is referred to as in–out or radial correlation. For $l=1$, $r^2g_{1,1}^2$ is very similar to $r^2g_{1,0}^2$. Hence f is increased if θ_{12} is greater than 90°, and decreased if θ_{12} is less than 90°. For this reason, this term is referred to as angular correlation. Beyond these first two terms the convergence of Ψ (and E) is slower with many terms required to represent the detailed shape of the Coulomb hole.

Figure 2 illustrates the effect of $g_{2,0}$ on the in–out correlation. This large (10%–20%) correction to f arises from the seemingly small coefficient (−0.06148) in Table I (0.4% of the wave function) because the correction is linear in the coefficients γ_i. The ratio $g_{2,0}/g_{1,0}$ lies in the range ±2 for values of r for which $g_{1,0}$ is large. Thus one should not be deceived by the large coefficient of $1s^2$ into falsely assuming that $P(\mathbf{r}_1, \mathbf{r}_2)$ is given accurately by this simple wave function.

TABLE II. Atomic properties of helium from truncated expansions.[a]

Last term added	$\langle\delta(r_{12})\rangle$	$\langle r_{12}^{-1}\rangle$	$\langle r^{-1}\rangle$	$\langle-\frac{1}{2}\nabla^2\rangle$	$\langle H\rangle$
SCF		1.02578	1.68729	1.43084	−2.86168
$1s$	0.19158	1.02640	1.68987	1.43572	−2.86165
$1p$	0.18464	0.98819	1.68787	1.44066	−2.88200
$2s$	0.15802	0.95787	1.68842	1.44918	−2.89743
$1d$	0.15778	0.95449	1.68832	1.44978	−2.89922
$2p$	0.15595	0.95127	1.68830	1.45051	−2.90089
$3s$	0.15205	0.94972	1.68834	1.45097	−2.90169
exact	0.10636	0.94579	1.68831	1.45186	−2.90372

[a] Davidson (1963).

Table II shows the convergence properties for various operators of interest. It is frequently argued on the basis of Eq. (18) that the SCF values for one-electron operators should be accurate. Data such as that shown in Table II for $\langle r^{-1}\rangle$ are used to substantiate this claim. Actually, however, ϵ^2 is about 1% for helium, so that 1% corrections to $\langle\mathcal{O}\rangle$ would be expected. The fact that for $\mathcal{O}(r)$ the corrections are generally less than this is a reflection of the fact that the density from the higher natural orbitals is similar to that for the first $(\langle g_{i,l}\,|\,\mathcal{O}\,|\,g_{i,l}\rangle\approx\langle g_{1,0}\,|\,\mathcal{O}\,|\,g_{1,0}\rangle)$, and Eq. (11) has no interference (nondiagonal) terms. In momentum space the situation is quite different since localization of $g_{i,l}$ in Cartesian space leads to large changes in kinetic energy. Thus the one-electron property, kinetic energy, shows the expected convergence properties and $\langle r^{-1}\rangle$ must be regarded as anomalously good.

TABLE III. Convergence of molecular properties of H_2.

Last term added	$\langle\delta(r_{12})\rangle$	$\langle r_{12}^{-1}\rangle$	$\langle H\rangle$
SCF	0.0432	0.6584	−1.1336
$1\sigma_g$	0.0440	0.6627	−1.1335
$1\sigma_u$	0.0357	0.6295	−1.1519
$1\pi_u$	0.0314	0.6094	−1.1628
$2\sigma_g$	0.0273	0.5957	−1.1699
$1\pi_g$	0.0265	0.5943	−1.1707
$3\sigma_g$	0.0260	0.5933	−1.1712
$2\sigma_u$	0.0252	0.5924	−1.1717
$1\delta_g$	0.0247	0.5915	−1.1722
$1\pi_u$	0.0239	0.5904	−1.1727
$4\sigma_g$	0.0234	0.5898	−1.1730
exact	0.0170	0.5874	−1.1744

Chemists have also investigated the natural orbitals for several two-electron molecules. Numerous papers have been written on the ground (Eliason and Hirschfelder, 1959; Hirschfelder and Löwdin, 1959, 1965; Shull, 1959; Davidson and Jones, 1962a, b; Hagstrom and Shull, 1963; Ahlrichs, Kutzelnigg, and Bingel, 1966b; Das and Wahl, 1966; Rothenberg and Davidson, 1967; Konowalow, Barker, and Mandel, 1968) and excited states (Davidson, 1961; Rothenberg and Davidson, 1966, 1967) of H_2, and the ground states of He_2^{2+} (Shull and Prosser, 1964), HeH^+ (Anex and Shull, 1964), and H_3^+ (Shull, 1964; Kutzelnigg, Ahlrichs, Labib–Iskander, and Bingel, 1967; Banyard and Tait, 1968; Christofferson and Shull, 1968). The results from these papers are qualitatively the same as for helium with the exception that γ_i/γ_1, is not always negative for excited states or for large nuclear separations. This is not unexpected since, in these cases, the zeroth-order wave function cannot be represented in the form (36) with a single large coefficient, and hence approximation (49) is not valid.

While one could argue that the correlation error in helium is only 5% of the ionization energy, and hence doesn't matter, the corresponding correlation error in H_2 is responsible for 24% of the binding energy. For larger molecules the change in correlation energy upon binding often exceeds the binding energy, so that a description of this energy becomes essential to even a qualitative understanding of the origin of chemical binding.

Table III illustrates the convergence of a few properties of H_2 with NSO expansion length. Here again the correlation factor is approximately given by

$$f \approx 1-2\sum_i |\gamma_i| [\chi_i(1)/\chi_1(1)][\chi_i(2)/\chi_1(2)].$$

Since $1\sigma_u$ has a nodal plane perpendicular to the molecular axis, σ_u/σ_g tends to increase the probability of finding electrons at opposite ends of the molecule, and decrease the probability of finding them at the same end (left–right correlation). The $1\pi_u$ orbital has a mirror plane containing the molecular axis so that it keeps electrons on opposite sides of the bond axis (angular correlation). The $2\sigma_g$ orbital has an elliptical node and introduces in–out correlation. Clearly the $2\sigma_g$ effect may be regarded as a distortion of the $2s$ effect in helium, while σ_u and π_u arise from the $1p$ orbital of helium.

Rothenberg and Davidson (1966) have published extensive Tables of the natural orbital expansions of the excited states of H_2. They have also (1967) examined the effect of correlation on the transition moments of H_2. For the $X\,{}^1\Sigma_g{}^+ \rightarrow 1s2p\,{}^1\Pi_u$ transition, electron correlation causes a 10%–20% increase in the oscillator strength. This was caused mainly by the $1\pi_u{}^2$ term in the ground-state wave function which interacts strongly with the leading $1\sigma_g 1\pi_u$ term of the ${}^1\Pi_u$ wave function.

Hirschfelder and Löwdin (1959, 1965) have examined the problem of H_2 at large internuclear separations. The traditional theory of Van der Waals interaction treats electron correlation at large R by an expansion in excited states of the atom to derive the R^{-6} coefficient. Once again, this is an inappropriate expansion since the error in Ψ is localized in the region where Ψ is large. Hirschfelder and Löwdin have shown that a term of the form

$$-R^{-3}[p_{ax}(1)p_{bx}(2)+p_{ay}(1)p_{by}(2)+p_{az}(1)p_{bz}(2)$$
$$+p_{bx}(1)p_{ax}(2)+p_{by}(1)p_{ay}(2)+p_{bz}(1)p_{az}(2)],$$

with p orbitals the same size as the $1s$ orbitals, will reproduce the coefficients of R^{-6} and R^{-8} to better than 1%.

IV. SEVERAL-ELECTRON WAVE FUNCTIONS

There are now numerous calculations for light atoms and molecules. Most of the calculations which have gone beyond the SCF limit have been strongly influenced by the natural orbital results for two-electron systems. For more than two electrons there is no simplification in the form of the wave function when natural orbitals are used, so the only advantage is that the expansion in determinants is rapidly convergent.

There are several ways of using natural orbitals in wave function calculations. One may calculate natural orbitals from existing wave functions as an aid in interpretation and in spotting deficiencies in the basis set. This may also suggest the form which should be assumed for configurations in a multiconfiguration SCF calculation. Pseudonatural orbitals from an independent-electron-pair calculation may be used as a basis for a full variational calculation. Natural orbitals from one wave function may be used as a basis set to generate a better wave function in an iterative manner.

All of these methods are based on variational wave functions with a finite expansion in determinants. No one, as yet, has used natural orbitals as the basis for a perturbation calculation beyond second order. In many cases, perturbation theory based on natural orbital determinants would fail to converge because the off-diagonal matrix elements $H_{I,J}$ of the Hamiltonian H between excited configurations are often similar in magnitude to the diagonal elements $H_{I,I}-E_0$. This occurs because, although Ψ_{FNC} is near the true ground-state wave function, the "excited" configurations are not at all similar to the true excited states. As a consequence, the series for $P/(H-E_0)$ which occurs in most versions of perturbation theory will not converge if the off-diagonal elements are treated as a perturbation [expansion of $P/(H-E_0)$ requires that $H_{IJ}{}^2 \ll (H_{II}-E_0)(H_{JJ}-E_0)$]. The second-order energy for small molecules (Siu and Davidson, 1970) is often in error by as much as 50%.

Analysis of existing wave functions has been limited since very few good wave functions have been generated without incorporating natural orbitals from the beginning. Hurley (1960) reported the natural orbitals for carbon monoxide. Barnett, Linderberg, and Shull (1965) analyzed several beryllium atom wave functions and a LiH wave function. D. Smith and Fogel (1965) reported an analysis for beryllium. V. Smith (1967) analyzed Boys' function for carbon. V. Smith and Larsson (1968) computed natural orbitals for a lithium atom wave function with explicit r_{12} dependence. Macias (1968) has studied a wave function for H_3^-. Olympia (1970) reported natural orbitals for the first excited states of beryllium. Olympia and Fung (1970) published some results for CH_3^-, NH_3, and OH_3^+.

The work of Allen and Shull (1962) supported the idea of Hurley, Lennard-Jones, and Pople (1953) that the beryllium atom wave function can be approximated as

$$\Psi = \mathcal{A} g_1(1,2)g_2(3,4), \qquad (56)$$

where g_1 and g_2 are strongly orthogonal geminals,

$$\int g_1{}^*(1,2)g_2(1,3)\,d\tau_1 = 0, \qquad (57)$$

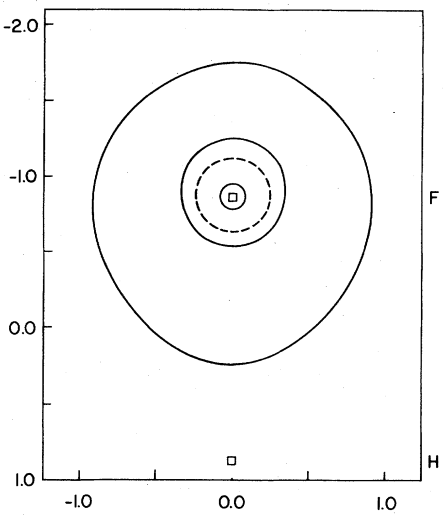

FIG. 3. The 2σ natural orbital of HF (nearly $2s_F$).

and $\mathcal{A}$ is the antisymmetric projection operator. The error in such an approximation seems to be about 10% of the correlation energy.

Since g_1 and g_2 are two-electron functions, they have natural expansions of the form (36). Arai (1960) has proven that the natural orbitals of g_1 are disjoint from (orthogonal to) the natural orbitals of g_2 if g_1 and g_2 are strongly orthogonal. The natural orbitals of Ψ are the union of these two disjoint sets. This very important result also holds for the general form

$$\Psi=\mathcal{A}g_1(1,2)g_2(3,4)\cdots g_{N/2}(N-1,N) \qquad (58)$$

known as an antisymmetrized product of strongly orthogonal geminals (APSG).

Because of its simplicity, ease of interpretation, and accuracy for beryllium, a great deal of effort has been expended in finding APSG wavefunctions for other systems. Ebbing and Henderson (1965) performed a direct calculation of the APSG function for LiH. Miller and Ruedenberg (1965, 1968) calculated the APSG function for Be. Both of these calculations proceeded by solving a complicated coupled set of equations for the natural orbital expansions of the g_i. The relatively high accuracy achieved in these two examples has inspired calculations for systems with more electrons. Hindsight has shown, however, that the good results for LiH and Be resulted because the electrons were localized in two nonpenetrating pairs, and not because Eq. (58) is a particularly appropriate form for the wave function. For larger molecules such as NH, Ruedenberg and co-workers (Mehler *et al.*, 1970; Silver *et al.*, 1970a, b have found that the resulting error in the correlation energy is nearly 80%. This arises because this method accounts for only $N/2$ pairs of electron correlations out of the total of $N(N-1)/2$. For most molecules, the SCF orbitals, even when chosen as localized as possible, are still strongly interpenetrating so that interpair correlations are comparable to intrapair correlations.

Kutzelnigg (1964) and Ahlrichs and Kutzelnigg (1968a) have developed an algorithm for finding approximate geminals based on their previous method of direct calculation of natural orbitals for two-electron functions. Besides Be and LiH, their method has been applied (Ahlrichs and Kutzelnigg, 1968b; Ahlrichs, 1970; Jungen and Ahlrichs, 1970) to BeH_2, BH_3, CH_4, BeH, BH_2^+, BH, BeH_4, and B_3H_6. For most of these molecules, independent-electron-pair [Nesbet–Bethe–Goldstone (Nesbet, 1968)] calculations were made for each of the $N(N-1)/2$ pairs by using equations similar to (47–50) to find the best g_{ij} in natural orbital form for the wave function.

$$\Psi=\mathcal{A}g_{ij}(i,j)\prod_{k\neq i,j}\phi_k(k). \qquad (59)$$

Again the conclusion is drawn that, in aggregate, interpair effects are larger than intrapair effects for most molecules.

In a somewhat similar approach, Edmiston and Krauss (1966, 1968) and Sanders and Krauss (1968) have found natural orbitals from factoring independent-electron-pair wave functions of the form (59) obtained variationally [a la Nesbet (1968)]. By repeating this for each intrapair effect and then merging and reorthogonalizing the set of leading natural orbitals, a set of pseudonatural orbitals was developed which was suitable as a basis for an ordinary variational calculation.

Sanders and Krauss used this procedure to study LiH and BeH valence shell correlation. Edmiston and Krauss applied this method to He_2^+ and H_3. Their

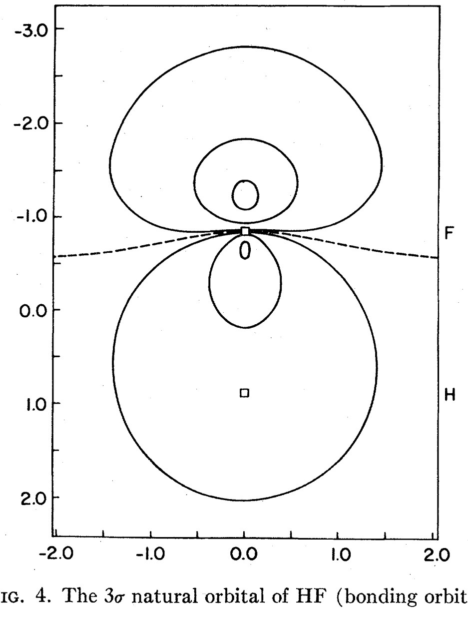

FIG. 4. The 3σ natural orbital of HF (bonding orbital).

results were within 10% of the experimental correlation error. Further the results were limited more by the choice of basis set than by the method of choosing the pseudonatural orbitals. This is very encouraging since determination of accurate NO's is quite costly. It would appear that any reasonably complete set of localized correlation orbitals will give good convergence.

Bender and Davidson (1966) have developed an iterative method for calculation of natural orbitals. In this scheme an initial wave function is factored into natural orbitals. These are then used as a basis set to construct a better wave function which is again factored into natural orbitals. Improvement in the energy (for a fixed number of determinants) is usually dramatic for a few iterations, although convergence usually does not occur. This procedure has been applied by them (Bender and Davidson 1967a, b, 1968, a, b, 1969b; Matsumoto, Bender, and Davidson, 1967) to Be_2, HeH, LiH, HF, BeH, BH, CH, NH, OH, and He_2; by Chan and Davidson (1968, 1970) to MgH and BeH; and by Siu and Davidson (1970) to CO. It has also been adapted to atomic calculations for carbon (A. Bunge, 1970) and beryllium (C. F. Bunge, 1968). Calculations of this type are also being carried out by Schaefer (1971).

The iterative natural orbital method is presently the most flexible and most accurate method for construction of accurate molecular wave functions. Its success depends in part on the ability to make good initial guesses for the natural orbitals. The initial guess to the natural orbitals can be made by the techniques discussed previously and used without iteration by Kutzelnigg and Krauss. Alternatively, second-order perturbation theory can be used with Roothaan–Hartree–Fock virtual orbitals to obtain first-order wave functions and their approximate natural orbitals. In most of the calculations done by Bender and Davidson, second-order perturbation theory for the energy was also used to select the important doubly excited configurations.

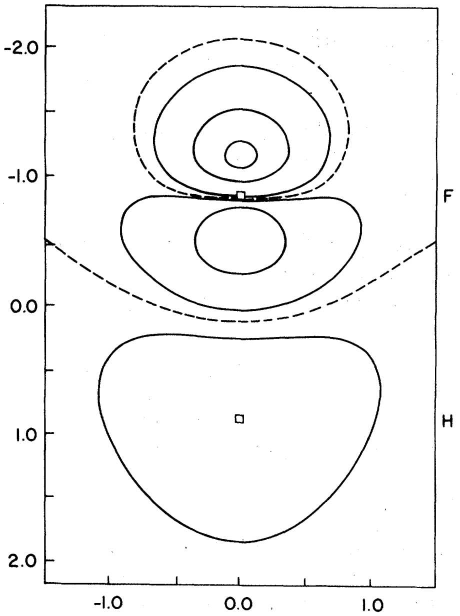

FIG. 5. The 4σ antibonding natural orbital of HF.

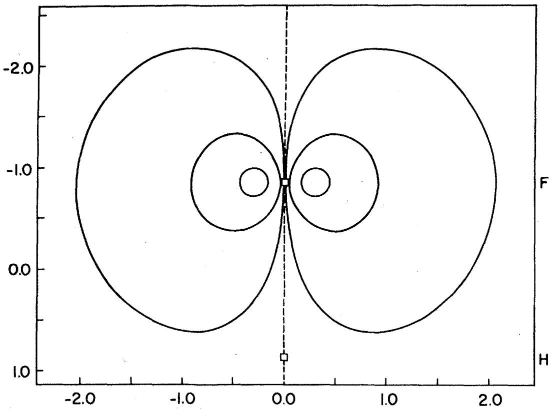

FIG. 6. The 1π natural orbital of HF.

There is much to be learned from these papers about the details of electron correlation in molecules and its description by natural orbitals. In the pioneering paper on this method, Bender and Davidson (1966) calculated the energies of HeH and LiH at an internuclear separation of 3.0 bohr. The binding energy of LiH was in error by 10%, while the repulsive energy of HeH was off by only 0.5%. The correlation effects in LiH were very similar to those in Li^+ and H^-. Several of the natural orbitals resembled those of Li^+. The 2σ molecular orbital (nominally 1s on H^-), however, was strongly polarized by the Li^+ core. Further, although the total correlation energy of the $2\sigma^2$ pair was about the same as that of H^-, the contribution from individual natural orbitals was quite different. For LiH, the contribution to the energy from single, triple, quadruple, or cross shell $1\sigma 2\sigma$ double excitations are all quite small. The APSG form is a good approximation to this wave function.

These results are a little misleading, however, since this first paper failed to give an improved dipole moment relative to the SCF result ($\mu = 5.83$ D experimentally, 5.95 in SCF, 5.97 in the NSO calculation). In a later paper Bender and Davidson (1968b) pointed out that this error arose from failure to include single excitations in Ψ even though they had negligible effect on the energy. Without single excitations, the highly occupied iterated orbitals will always be a transformation of the SCF orbitals, and the first configuration will always give the SCF dipole moment. When single excitations were included in each iteration, a small mixing between "occupied" and "unoccupied" orbitals occurred which resulted in greatly improved molecular

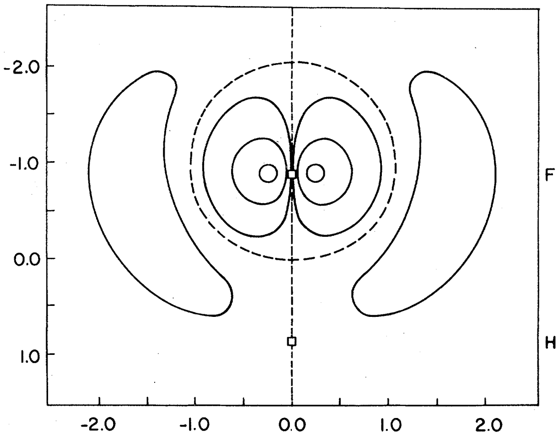

FIG. 7. The 2π natural orbital of HF.

properties. With single excitations included, the first configuration built from NO's gave $\mu=5.89$ and the full wave function gave $\mu=5.85$ in good agreement with experiment.

As a preliminary to the work on the HF molecule, Bender and Davidson (1968b) did a pseudonatural orbital calculation [cf., Edmiston and Krauss (1966)] for every inter- and intra-pair correlation in this molecule. These results indicated clearly that the APSG wave function was not appropriate for complex molecules since (a) interpair effects are responsible for 2/3 of the correlation error in the dissociation energy, and (b) there were several orbitals which were important in the correlation of several different pairs. An iterative NO calculation (Bender and Davidson, 1967a) gave a 39-configuration wavefunction representing 50% of the correlation energy. At the time of its publication, this was the first result for the HF molecule below the SCF limit.

This wave function showed one surprising feature. Most chemists had assumed that the dominant error in an SCF wave function was due to its incorrect form at large internuclear separations. This could be corrected by configurations of the form $B^2 \rightarrow A^2$, where B is the bonding molecular orbital (MO), and A is the antibonding molecular orbital. In HF this was not the leading error. The largest single energy contribution came from $B\pi \rightarrow A\pi'$. This term represents an alternation between the bonding and π electrons near F which enhances the probability of the bonding electron being near H when the π electrons are closer than average to F. Thus this correlation induces coupling at right angles and is similar to the $p_x p_y$ pair correlation of neon. This effect contributes 0.4 eV to the dissociation energy D_e. In–out and angular correlations among the π electrons were also more important than $B^2 \rightarrow A^2$ although they did not contribute to D_e.

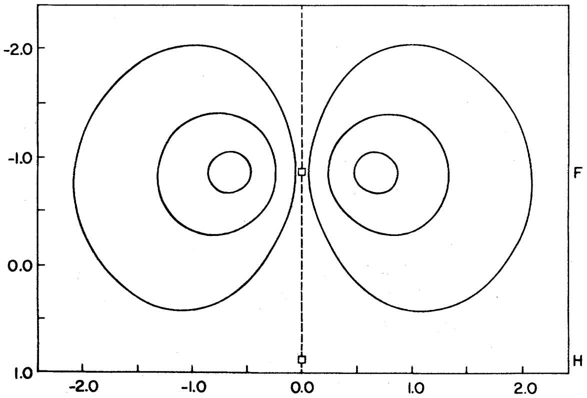

FIG. 8. The 1δ orbital of HF.

TABLE IV. A few configurations from the HF wave function.

Configuration	Coefficient squared	Energy	$-\Delta E$
$1\sigma^2 2\sigma^2 3\sigma^2 1\pi^2 1\bar{\pi}^2$	0.97226	−100.0486	
$1\pi 1\bar{\pi} \rightarrow 2\pi 2\bar{\pi}$	0.00370	−100.07575	0.0271
$1\pi 1\bar{\pi} \rightarrow 1\delta 1\bar{\delta}$	0.00217	−100.09383	0.0181
$3\sigma 1\pi \rightarrow 4\sigma 2\pi$	0.00514	−100.12706	0.0332
$3\sigma^2 \rightarrow 4\sigma^2$	0.00232	−100.13761	0.0106

Figures 3–8 show contour maps for the most important natural orbitals of HF. Table IV gives a list of the first few configurations. The contour maps for $\pi \pm \pi^*$ orbitals would have a nodal plane containing the molecular axis although $|\pi|^2$ is the same in all planes, Similarly $\delta \pm \delta^*$ has two nodal planes at right angles intersecting in the molecular axis. The correlation effects produced by these configurations can always be interpreted similarly to Eqs. (54)–(55). For a closed shell molecule, we have

$$P(\mathbf{r}_1, \mathbf{r}_2) \approx \sum_i^{N/2} |g_i(1)|^2 \, |g_i(2)|^2 f_{ii} + \tfrac{1}{2} \sum_{i \neq j}^{N/2} |g_i(1)|^2 \, |g_j(2)|^2$$

$$\times ({}^1f_{i,j} + {}^{3,-1}f_{i,j} + {}^{3,0}f_{i,j} + {}^{3,1}f_{i,j}),$$

where ${}^{3,1}f_{i,j}$ denotes the correlation factor for the electron pair $g_i\alpha g_j\alpha$ (3 shows that the spin is triplet, 1 shows that S_z is one). These triplet factors ${}^{3,m}f_{i,j}$ each contain a factor

$$\{1 - [g_j(1)/g_i(1)][g_i(2)/g_j(2)]\}$$

which is zero at $r_1 = r_2$ (this describes the Fermi–hole). In addition, each f contains a factor of the form

$$1 - 2 \sum |C_{ij \rightarrow kl}| [g_k(1)/g_i(1)](g_l(2)/g_j(2))$$

which comes from configurations formed by the excitation $g_i g_j \rightarrow g_k g_l$. Each configuration in Table IV may be interpreted in this manner. For example, the interpretation of $3\sigma 1\pi \rightarrow 4\sigma 2\pi (B\pi \rightarrow A\pi')$ given previously is based on noticing that $4\sigma/3\sigma$ is of constant sign near F, but of opposite sign near H. Similarly $2\pi/1\pi$ is of one sign near F, but of opposite sign further out perpendicular to the bond.

In the culminating paper in this series, Bender and Davidson (1969a) reported calculations for all of the first-row diatomic hydrides at their equilibrium separations. Not only the total, but also the Nesbet independent-pair wave functions were computed. Generally it was found that the independent-pair energy exceeded the believed limit of the basis set by up to 20%. In order to achieve improved accuracy, this paper used 1000–3000 configurations in evaluating the energy and other molecular properties. The molecular properties obtained were much improved over the SCF results.

Probably the most significant aspect of this work on hydrides was the discovery that certain types of correlation persisted throughout all of the molecules. As expected the $B^2 \rightarrow A^2$ excitation (left-right correlation of the bonding pair) was always important. In–out and angular correlations among the π electrons were large. Coupling at right angles between the π electrons and the 2σ and 3σ electrons was also important. This latter effect has a large effect on the dissociation energy.

The discovery that every pair in the valence shell is strongly correlated (not just the intrapairs) negated much of the previous work on APSG and MC–SCF wave functions. Although the multiconfiguration SCF method was easily adapted to include interpair excitations, its original formulation and applications had generally been based on the assumption that in an NSO basis, ouly intrapair excitations would matter.

V. CONCLUSIONS

The unique advantage of natural orbitals is that they reduce the number of determinants required for an accurate wave function. Those determinants required make possible a simple interpretation of correlation effects. For molecular calculations based on variational methods, this is a great advantage over use of an arbitrary or an SCF orbital basis.

A serious disadvantage of natural orbitals for several electrons is that they are relatively expensive to obtain. Part of this disadvantage may be overcome by use of one of the methods, such as pseudonatural orbitals, for obtaining good first approximations. Also, perturbation theory can be used to select important configurations, even though it is too inaccurate to use for quantitative results.

Other methods such as diagonalization of the exchange integrals between occupied and unoccupied orbitals (Bender and Davidson, 1967b; Chan and Davidson, 1968) to obtain an initial set of orbitals are relatively more expensive and not greatly more accurate. Probably the cheapest initial guess for the natural orbitals is found by diagonalizing the density matrix formed from the first-order wave function [see Eq. (19)]. The first few orbitals found in this way tend to span the useful part of the basis set vector space even though they are rather far from being natural orbitals.

The use of SCF virtual orbitals in many-body perturbation theory (Kelly 1963) for atomic calculations has some advantages. The virtual (continuum) orbitals are easily computed, and certain nice convergence patterns make possible extrapolations of perturbation sums. These simplifications will not appear in molecular calculations. For molecules, continuum orbitals are exceedingly difficult to obtain and no regularity is expected which would allow extrapolation of various partial sums based on only a few terms.

Use of perturbation theory within the Roothaan–Hartree–Fock set of orbitals obtained by approximately expanding the SCF orbitals in an arbitrary finite basis is, of course, possible for atoms and molecules. In this case there is no continuum, although some virtual orbitals may have positive energies. Generally the basis set in such a case is chosen only to give the occupied orbitals with no regard for the virtual orbitals. Nevertheless, with proper care, a meaningful many-body perturbation calculation could be carried out in the same finite basis set in which the natural orbitals could be expanded. The results would probably rival those obtained with natural orbitals both in accuracy and cost. Even in this case, however, the natural orbitals from the wave function would offer a much more physical description of the correlation effects than would the virtual SCF orbitals.

Natural orbitals have not been used extensively to generate excited state wave functions. This is partly because they have been too expensive to use for a large range of states and nuclear positions. They are most useful, quantitatively, in obtaining good molecular properties for the ground state. Some work on excited states has been done, however, and there seems to be no practical difficulty in using the iterative natural orbital scheme for the lowest state of any symmetry. For higher states of a given symmetry, the excited state variational principle guarantees only that the Nth root of the secular determinant is above the true Nth state. Some practical difficulties were encountered since the iterative natural orbital method was too good and sometimes eliminated orbitals necessary to represent the $N-1$ lower states. In this case, the desired wave function would appear as a lower root, and no variational principle would hold. This difficulty is easily overcome in practice by keeping the orbitals required for lower roots. Generally speaking, the lack of results for higher excited states is due to a lack of interest in finding results of this accuracy for some one state.

REFERENCES

Ahlrichs, R. 1970, Theoret. Chim. Acta **17,** 348.
——, and W. Kutzelnigg, 1968a, J. Chem. Phys. **48,** 1819.
——, and W. Kutzelnigg, 1968b, Theoret. Chim. Acta **10,** 377.

——, W. Kutzelnigg, and W. A. Bingel, 1966a, Theoret. Chim. Acta **5**, 289.

——, W. Kutzelnigg, and W. A. Bingel, 1966b, Theoret. Chim, Acta **5**, 305.

Allen, T. L. and H. Shull, 1962, J. Phys. Chem. **66**, 2281.

Anex, B. G., and H. Shull, 1964, in *Molecular Orbitals in Chemistry, Physics, and Biology*, edited by P.-O. Löwdin and B. Pullman, (Academic, New York), p. 227.

Arai, T., 1960, J. Chem. Phys. **33**, 95.

Banyard, K. E. and C. C. Baker, 1969, J. Chem. Phys. **51**, 2680.

——, and A. D. Tait, 1968, J. Chem. Phys. **49**, 3050.

Barnett, G. P., J. Linderberg, and H. Shull, 1965, J. Chem. Phys., **43**, S80.

Bender, C. F. and E. R. Davidson, 1966, J. Phys. Chem. **70**, 2675.

——, and E. R. Davidson, 1967a, J. Chem. Phys. **47**, 360.

——, and E. R. Davidson, 1967b, J. Chem. Phys. **47**, 4972.

——, and E. R. Davidson, 1968a, *Reduced Density Matrices with Applications to Physical and Chemical Systems*, edited by A. J. Coleman and R. M. Erdahl, Queen's Papers in Pure and Applied Mathematics (Queen's University, Kingston, Ontario), No. 11, p. 335.
Applied Mathematics, No. 11, p. 335.

——, and E. R. Davidson, 1968b, J. Chem. Phys. **49**, 4222.

——, and E. R. Davidson, 1969a, Phys. Rev. **183**, 23.

——, and E. R. Davidson, 1969b, Chem. Phys. Letters **3**, 33.

Bingel, W. A., 1960, J. Chem. Phys. **32**, 1522.

——, 1962, J. Chem. Phys. **36**, 2842.

——, 1970, Theoret. Chim. Acta **16**, 319.

——, and W. Kutzelnigg, 1968, in *Reduced Density Matrices with Applications to Physical and Chemical Systems*, edited by A. J. Coleman and R. M. Erdahl, Queen's Papers in Pure and Applied Mathematics (Queen's University, Kingston, Ontario), No. 11, p. 385.

——, and W. Kutzelnigg, 1970, Advan. Quantum Chem. **5**, 201.

Brown, W. Byers and G. V. Nazaroff, 1967, Intern. J. Quant. Chem. **1**, 463.

Bunge, A., 1970, J. Chem. Phys. **53**, 20.

Bunge, C. F., 1968, Phys. Rev. **168**, 92.

Carlson, B. C. and J. M. Keller, 1961, Phys. Rev. **121**, 659.

Chan, A. C. H. and E. R. Davidson, 1968, J. Chem. Phys. **49**, 727.

——, and E. R. Davidson, 1970, J. Chem. Phys. **52**, 4108.

Christofferson, R. E. and H. Shull, 1968, J. Chem. Phys. **48**, 1790.

Coleman, A. J., 1963, Rev. Mod. Phys. **35**, 668.

Cressy, N., K. R. Miller, and K. Ruedenberg, 1969, Intern. J. Quant. Chem. **3**, 107.

Das, G., and A. C. Wahl, 1966, J. Chem. Phys. **44**, 87.

Davidson, E. R., 1961, J. Chem. Phys. **35**, 1189.

——, 1962, J. Chem. Phys. **37**, 577.

——, 1963, J. Chem. Phys. **39**, 875.

——, 1968, J. Chem. Phys. **48**, 3169.

——, and L. L. Jones, 1962a, J. Chem. Phys. **37**, 1918.

——, and L. L. Jones, 1962b, J. Chem. Phys. **37**, 2966.

Ebbing, D. D. and R. C. Henderson, 1965, J. Chem. Phys. **42**, 2225.

Edmiston, C. E. and M. Krauss, 1966, J. Chem. Phys. **45**, 1833.

——, and M. Krauss, 1968, J. Chem. Phys. **49**, 192.

Eliason, M. A. and J. O. Hirschfelder, 1959, J. Chem. Phys. **30**, 1397.

Hagstrom, S. and H. Shull, 1963, Rev. Mod. Phys. **35**, 624.

Hirschfelder, J. O., W. Byers Brown, and S. Epstein, 1964, Adv. Quant. Chem. **1**, 255.

——, and P.-O. Löwdin, 1959, Mol. Phys. **2**, 229.

——, and P.-O. Löwdin, 1965, Mol. Phys. **9**, 491.

Hurley, A. C., 1960, Rev. Mod. Phys. **32**, 400.

——, J. E. Lennard-Jones, and A. J. Pople, 1953, Proc. Roy. Soc. (London) **A220**, 446.

Jungen, M. and R. Ahlrichs, 1970, Theoret. Chim. Acta. **17**, 339.

Kelly, H. P., 1963, Phys. Rev. **131**, 684.

Kobe, D. H., 1969, J. Chem. Phys. **50**, 5183.

Konowalow, D. D., W. H. Barker, and R. Mandel, 1968, J. Chem. Phys. **49**, 5137.

Kutzelnigg, W., 1963a, Theoret. Chim. Acta **1**, 327.

——, 1963b, Theoret. Chim. Acta **1**, 343.

——, 1963c, Z. Naturforsch. **18a**, 1058.

——, 1964, J. Chem. Phys. **40**, 3640.

——, R. Ahlrichs, I. Labib-Iskander, and W. A. Bingel, 1967, Chem. Phys. Letters **1**, 447.

Linderberg, J., 1964, J. Mol. Spect. **12**, 210.

P.-O. Löwdin, 1955, Phys. Rev. **97**, 1474.

——, 1959, Adv. Chem. Phys. **2**, 207.

——, and H. Shull, 1956, Phys. Rev. **101**, 1730.

Macias, A., 1968, J. Chem. Phys. **49**, 2198.

March, N. H., W. Y. Young, and S. Sampanthar, 1967, *The Many-Body Problem in Quantum Mechanics* (Cambridge U. P., Cambridge, England).

Matsumoto, G. H., C. F. Bender, and E. R. Davidson, 1967, J. Chem. Phys. **46**, 402.

McWeeny, R., 1960, Rev. Mod. Phys. **32**, 335.

——, and W. Kutzelnigg, 1968, Intern. J. Quant. Chem. **2**, 187.

Mehler, E. L., K. Ruedenberg, and D. M. Silver, 1970, J. Chem. Phys. **52**, 1181.

Miller, K. J. and K. Ruedenberg, 1965, J. Chem. Phys. **43**, S88.

——, and K. Ruedenberg, 1968, J. Chem. Phys. **48**, 3414.

Nazaroff, G. V., 1968, J. Chem. Phys. **48**, 3517.

——, and J. O. Hirschfelder, 1963, J. Chem. Phys. **39**, 715.

Nesbet, R. K., 1968, Phys. Rev. **175**, 2.

Olympia, P. L., 1970, Theoret. Chim. Acta **16**, 145.

——, and B. M. Fung, 1970, J. Chem. Phys. **53**, 1066.

Reid, C. E. and Y. Ohrn, 1963, Rev. Mod. Phys. **35**, 445.

Reinhardt, W. P. and J. D. Doll, 1969, J. Chem. Phys. **50**, 2767.

Rothenberg, S. and E. R. Davidson, 1966, J. Chem. Phys. **45**, 2560.

——, and E. R. Davidson, 1967, J. Mol. Spectry. **22**, 1.

Sanders, W. A. and M. Krauss, 1968, J. Res. Natl. Bur. Std. **72A**, 85.

Schaeffer, H. F., III, 1971, J. Chem. Phys. **54**, 2207.

Shull, H., 1959, J. Chem. Phys. **30**, 1405.

——, 1964, J. Am. Chem. Soc. **86**, 1469.

——, and P.-O. Löwdin, 1955, J. Chem. Phys. **23**, 1565.

——, and P.-O. Löwdin, 1959, J. Chem. Phys. **30**, 617.

——, and F. Prosser, 1964, J. Chem. Phys. **40**, 233.

Silver, D. M., E. L. Mehler, and K. Ruedenberg, 1970, J. Chem. Phys. **52**, 1174.

——, K. Ruedenberg, and E. L. Mehler, 1970, J. Chem. Phys **52**, 1206.

Siu, A. K. Q. and E. R. Davidson, 1970, Intern. J. Quant. Chem. **4**, 223.

Smith, D. W., 1968, in *Reduced Density Matrics with Applications to Physical and Chemical Systems*, edited by A. J. Coleman and R. M. Erdahl, Queen's Papers in Pure and Applied Mathematics (Queen's University, Kingston, Ontario), No. 11, p. 169.

——, and S. J. Fogel, 1965, J. Chem. Phys. **43**, S91.

Smith, V. H., 1967, Theoret. Chim. Acta **7**, 245.

——, and S. Larsson, 1968, in *Reduced Density Matrices with Applications to Physical and Chemical Systems*, edited by A. J. Coleman and R. M. Erdahl, Queen's Papers in Pure and Applied Mathematics (Queen's University, Kingston, Ontario), No. 11, p. 296.

Wang, P. S. C., 1970, J. Chem. Phys. **52**, 4459.