April 2002

A Short History of Valence Bond Theory

Gordon A. Gallup

UNL, ggallup1@unl.edu

Gordon A. Gallup
A Short History of Valence Bond Theory

CONTENTS : Introduction • History: Pre-World War II • Heitler-London Treatment • Extensions Past the Simple Heitler-London-Wang Result • Polyatomic Molecules • The Heitler-Rumer Method for Polyatomic Molecules • Slater's Bond Functions • The Perfect Pairing Function • Symmetric Group Theoretic Approaches • History: Post-World War II and Automatic Computation • The Coulson and Fischer Treatment of H_2 • Goddard's Generalized VB • The Spin-Coupled VB • The BOVB Method • More Recent Developments in Symmetric Group Methods • Multiconfiguration Methods • The Multistructure Procedure of Balint-Kurti and Karplus • The MCVB Method • Early Ideas • Overlap Matrices and the Neglect of Some Permutations • Sums of Permutations of the Same Order • Application to the Ti-System of Naphthalene • Application to CH_4 • Orthogonalized AOs • Relation of Hamiltonian Matrix to Overlap Matrix • A 2×2 System • The Tr-System of Naphthalene • The CH_4 Molecule • The Perfect Pairing Wave Function and the Valence State of Carbon • Appendix — Acronyms • References

CONTENTS

A Short History of Valence Bond Theory

G. A. Gallup

Department of Physics and Astronomy
University of Nebraska–Lincoln

1 Introduction

Shortly after quantum mechanics evolved Heitler and London [1] applied the then new ideas to the problem of molecule formation and chemical valence. Their treatment of the H_2 molecule was qualitatively very successful, and this led to numerous studies by various workers applying the same ideas to other substances. Many of these involved refinements of the original Heitler-London procedure, and within three or four years, a group of ideas and procedures had become reasonably well codified in what was called the valence bond (VB)* method for molecular structure.

A few calculations were carried out earlier, but by 1929 Dirac [2] wrote:

> The general theory of quantum mechanics is now almost complete, the imperfections that still remain being in connection with the exact fitting in of the theory with relativity ideas. These give rise to difficulties only when high-speed particles are involved, and are therefore of no importance in the consideration of atomic and molecular structure and ordinary chemical reactions. The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble . . . .

---

* A list of acronyms used in this chapter is in an appendix.

Since these words were written there has been no reason to feel that they are incorrect in any way. Perhaps the only difference between attitudes then and now are that, today, with visions of DNA chains dangling before our eyes, we are likely to have an even greater appreciation of the phrase "much too complicated to be soluble" than did early workers.

The early workers were severely hampered, of course, by the considerable difficulty of carrying out, for even small systems, the prescriptions of VB theory with sufficient accuracy to assess their merit. Except for H_2 and perhaps a few other molecules and ions, no really accurate VB calculations were possible, and, to make progress, most workers had to resort to many approximations. There thus arose a series of generalizations and conclusions that were based upon results of at least somewhat uncertain value. In their review of early results, Van Vleck and Sherman [3] comment upon this point to the effect that a physical or chemical result was not to be trusted unless it could be confirmed by several calculations using different sorts of approximations. It is perhaps only to be expected that such cross checking was rather infrequently undertaken.

In this chapter we have two goals. The first is to give a general picture of the sweep of history of VB theory. We restrict ourselves to *ab initio* versions of the theory or to versions that might be characterized as reasonable approximations to *ab initio* theory. Our second goal is to identify a few of the early ideas alluded to in the previous paragraph and see how they hold up when they are assessed with modern computational power. The list is perhaps idiosyncratic, but almost all deal with some sort of approximation, which generally will be seen to be poor.

2   History: Pre-World War II

In the next few sections we give an historical description of the activity and ideas that led to our current understanding of VB methods. As with so much other human activity, progress in the development of molecular theory was somewhat suspended by the Second World War, and we use that catastrophe as a dividing point in our narrative.

Almost all of the ideas were laid down before World War II, but difficulties in carrying out calculations precluded firm conclusions in any but the simplest cases. The H_2 molecule does allow some fairly easy calculations, and, in the next section, we give a detailed description of the Heitler-London calculations on that molecule. This is followed by descriptions of early work of a more qualitative nature.

2.1 Heitler-London Treatment

The original treatment of the H_2 molecule by Heitler and London [1] assumed a wave function of the form

$$\Psi = N[1s_a(1)1s_b(2) \pm 1s_b(1)1s_a(2)][\alpha(1)\beta(2) \mp \beta(1)\alpha(2)], \qquad (1)$$

where the upper signs are for the singlet state and the lower for the triplet, the "a" and "b" subscripts indicate $1s$ orbitals on either proton a or b, and α and β represent the $m_s = \pm \frac{1}{2}$ spin states, respectively. When the function of Eq. (1) and the Hamiltonian are substituted into the variation theorem, one obtains the energy for singlet or triplet state of H_2 as

$$E^{(1,3)}(R) = 2E_H + \frac{J(R) \pm K(R)}{1 \pm T(R)} \qquad (2)$$

Here E_H is the energy of a normal hydrogen atom, $J(R)$ was called the Coulomb integral, $K(R)$ was called the exchange integral, and $T(R)$ was called the overlap integral. The reader should perhaps be cautioned that the terms "Coulomb," "exchange," and "overlap" integrals have been used by many other workers in ways that differ from that initiated by Heitler and London. For the present article we adhere to their original definitions,

$$
\begin{aligned}
J(R) &= \langle 1s_a(1)1s_b(2)|V(1,2)|1s_a(1)1s_b(2)\rangle, & (3) \\
K(R) &= \langle 1s_a(1)1s_b(2)|V(1,2)|1s_b(1)1s_a(2)\rangle, & (4) \\
T(R) &= \langle 1s_a(1)1s_b(2)|1s_b(1)1s_a(2)\rangle, \\
&= \langle 1s_a(1)|1s_b(1)\rangle^2, \text{ and} \\
V(1,2) &= -1/r_{2a} - 1/r_{1b} + 1/r_{12} + 1/R_{ab}. & (5)
\end{aligned}
$$

These equations are obtained by assigning electron 1 to proton a and 2 to b, so that the kinetic energy terms and the Coulomb attraction terms $-1/r_{1a} - 1/r_{2b}$ give rise to the $2E_H$ term in Eq. (2). $V(1, 2)$ in Eq. (5) is then that part of the Hamiltonian that goes to zero for the atoms at long distances. It is seen to consist of two attraction terms and two repulsion terms. As observed by Heitler and London, the bonding in the H_2 molecule arises from the way these terms balance in the J and K integrals. We show a graph of these integrals in Fig. (1). The energy of Eq. (2) can be improved in a number of ways, and we will discuss the way the Heitler-London theory predicts bonding after discussion of one of these improvements.
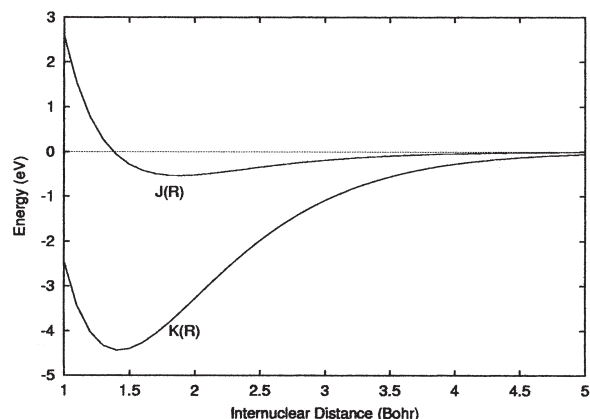
Figure 1: The relative sizes of the $J(R)$ and $K(R)$ integrals. The values are in eV.

The $1s$ orbitals in Eq. (1) represent the actual solution to the isolated H-atom. When we include an arbitrary scale factor in the exponent of the $1s$ orbital we symbolize it as

$$1s' = \sqrt{\alpha^3/\pi}\exp(-\alpha r). \tag{6}$$

When the $1s'$ orbital is used in the place of the actual H-atom orbital, one has α as a variation parameter to adjust the wave function. The energy expression becomes

$$E^{1,3}(\alpha, R) = 2E_H + (\alpha - 1)^2 + \alpha\frac{J(\alpha R) \pm K(\alpha R)}{1 \pm T(\alpha R)}, \tag{7}$$

which reduces to the energy expression of Eq. (2) when $\alpha = 1$. The changes brought by including the scale factor are only quantitative in nature and leave the qualitative conclusions unmodified. It is important to understand why the $J(R)$ and $K(R)$ integrals have the sizes they do. We consider $J(R)$ first. As we have seen from Eq. (5), $V(1, 2)$ is the sum of four different Coulombic terms from the Hamiltonian. If these are substituted into Eq. (3), we obtain

$$
\begin{aligned}
J(R) &= 2j_1(R) + j_2(R) + 1/R, \\
j_1(R) &= \langle 1s_a| - 1/r_b|1s_a\rangle = \langle 1s_b| - 1/r_a|1s_b\rangle, \\
j_2(R) &= \langle 1s_a(1)1s_b(2)|1/r_{12}|1s_a(1)1s_b(2)\rangle.
\end{aligned}
$$

The quantity $j_1(R)$ is seen to be the energy of Coulombic attraction between a point charge and a spherical charge distribution, $j_2(R)$ is the energy of Coulombic repulsion between two spherical charge distributions, and $1/R$ is the energy of repulsion between two point charges. $J(R)$ is thus the difference between two attractive and two repulsive terms that cancel to a considerable extent. The magnitude of the charges is one in every case. This is shown in Fig. (2), where we see that the resulting difference is only a few percent of the magnitudes of the individual terms.

Figure 2: Comparison of the sizes of $j_2 + 1/R$ and $-2j_1$ that comprise the positive and negative terms in the Coulomb integral. Values are in Hartrees.

This is to be contrasted with the situation for the exchange integral. In this case we have

$$
\begin{aligned}
K(R) &= 2k_1(R)S(R) + k_2(R) + S(R)^2/R, \\
k_1(R) &= \langle 1s_a| - 1/r_b|1s_b\rangle = \langle 1s_a| - 1/r_a|1s_b\rangle, \\
k_2(R) &= \langle 1s_a(1)1s_b(2)|1/r_{12}|1s_a(2)1s_b(1)\rangle.
\end{aligned}
$$

The magnitude of the charge in the overlap distribution, $1s_a 1s_b$, is $S(R)$, and here again, the overall result is the difference between the energies of attractive and repulsive terms involving the same sized charges of different shaped distributions. The values are shown in Fig. (3), where we see that now there is
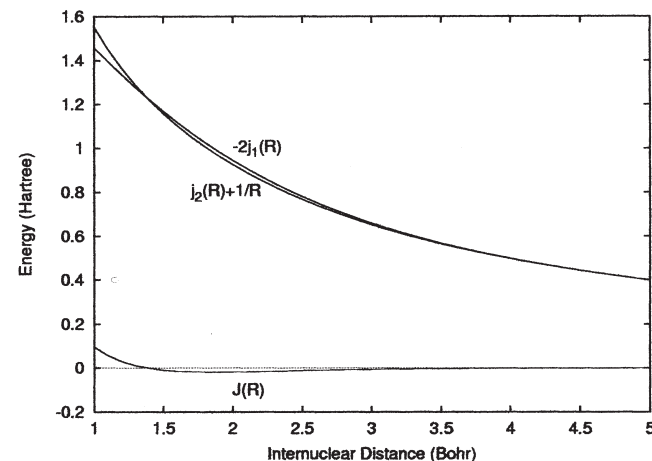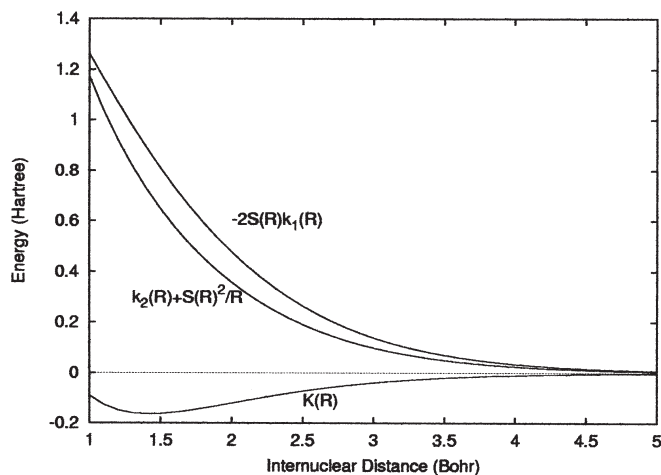
Figure 3: Comparison of the sizes of $k_2 + S^2/R$ and $-2k_1S$ that comprise the positive and negative terms in the exchange integral. Values are in Hartrees.

a considerably greater difference between the attractive and repulsive terms. This leads to a value about 20% of the magnitude of the individual terms.

These values for $J(R)$ and $K(R)$ may be rationalized in purely electrostatic terms involving charge distributions of various sizes and shapes.* From the point of view of electrostatics, $J(R)$ is the interaction of points and spherical charge distributions. The well-known effect, where the interaction of a point and spherical charge at a distance R is due only to the portion of the charge inside a sphere of radius R, leads to an exponential fall-off $J(R)$, as R increases.

The situation is not so simple with $K(R)$.The overlap charge distribution is shown in Fig. (4) and is far from spherical. The upshot of the differences is that the $k_2(R)$ integral is the self-energy of the overlap distribution and is more dependent upon its charge than upon its size. In addition, at any distance there is in $k_1(R)$ a portion of the distribution that surrounds the point charge, and, again, the distance dependence is decreased. The overall effect is thus that shown in Fig. (1).

---

*It should not be thought that the result $|J(R)| << |K(R)|$ is peculiar to the $1s$ orbital shape. It is fairly easy to show that a single spherical Gaussian orbital in the place of the $1s$ leads to a qualitatively similar result.
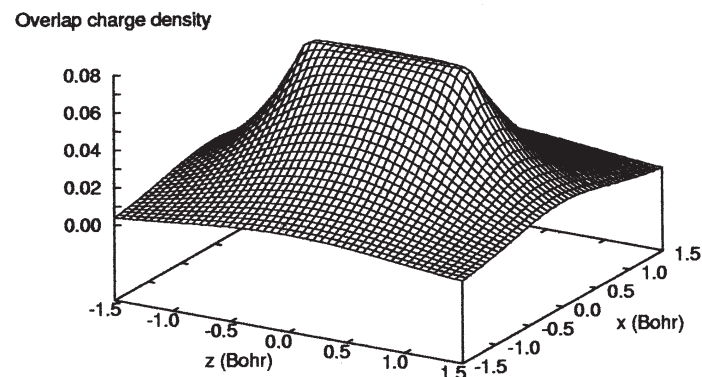
Figure 4: The overlap charge distribution when the H–H distance is near the molecular equilibrium value. We show an altitude plot of the value on the *x-z* plane.

We have not yet spoken of the effect of optimizing the scale factor in Eq. (7). Wang [4] showed, for the singlet state, that it varies from 1 at $R = \infty$ to about 1.17 at the equilibrium separation. Since both J and K have relatively small slopes near the equilibrium distance, the principal effect is to increase the potential energy portion of the energy by about 17%. The $(\alpha - 1)^2$ term increases by only 3%. Thus the qualitative picture of the bond is not changed by this refinement.

We have gone into some detail discussing the Heitler-London treatment of H_2, because of our conviction that it is important to understand the details of the various contributions to the energy. Our conclusion is that the bonding in H_2 is due primarily to the exchange effect caused by the combination of the Pauli exclusion principle and the required singlet state. Early texts (see *e.g.*, [5]) frequently emphasized the resonance between the direct and exchange terms, but this is ultimately due the principles in the last sentence. The peculiar shape of the overlap distribution leads to the major portion of the chemical bonding energy.*

---

*Those familiar with the language of the molecular orbital picture of bonding may be surprised that no parallel to the delocalization energy seems present in our description. That effect would occur in the VB treatment only if ionic terms are included. We thus conclude that delocalization is less important than the exchange attraction in bonding.

2.2 Extensions past the simple Heitler-London-Wang result

After the initial qualitative success of the simple VB calculation, further refinements that might be called multiconfigurational were investigated. These involve the introduction of *polarization* [6] and *ionic* [7] terms into the wave function. All of these refinements improve the quantitative agreement of the bond dissociation energy, D_e, with experiment, but any treatment so heavily dependent upon the H $1s$ and p_σ orbitals under-represents the electron correlation required to obtain better answers. At the time, such a treatment was carried out by James and Coolidge [8], but this was not really an extension of the Heitler-London-Wang calculation in any usefully physical sense.

2.3 Polyatomic molecules

The original Heitler-London calculation, being for two electrons, did not require any complicated spin and antisymmetrization considerations. It merely used the familiar rules that the spatial part of two-electron wave functions are symmetric in their coordinates for singlet states and antisymmetric for triplet states. Within a short time, however, Slater [10] had invented his determinantal method, and two approaches arose to deal with the twin problems of antisymmetrization and spin state generation. When one is constructing trial wave functions for variational calculations the question arises as to which of the two requirements is to be applied first, antisymmetrization or spin eigenfunction.

1. Methods based upon Slater determinantal functions (SDF). When we take this approach, we are, in effect, applying the antisymmetrization requirement first. Only if the orbitals are all doubly occupied among the spin orbitals is the SDF automatically, at the outset, an eigenfunction of the total spin. In all other cases further manipulations are necessary to obtain an eigenfunction of the spin, and these are written as sums of SDFs.

2. Symmetric group methods. When using these we, in effect, first construct n-particle (spin only) eigenfunctions of the spin. From these we determine the functions of spatial orbitals that must be multiplied by the spin eigenfunctions in order for the overall function to be antisymmetric. It may be noted that this is precisely what is done in almost all treatments of two electron problems. Generating spatial functions with the required properties leads to considerations of the theory of representations of the symmetric groups.

It is difficult to recreate today the attitudes that determined which of these approaches people chose. We can speculate that for small systems the basic simplicity of the SDF approach was appealing. The group theoretic approach seemed to some to be over-complicated. We quote from the Van Vleck and Sherman [3] review.

> . . .the technique of the permutation group is complicated, and more general than needed for practical purposes because the Pauli principle must be satisfied after the addition of spin. In the language of group theory, many "characters" for the orbital permutation group are not compatible with the Pauli principle . . . Thus the character theory is too general.

One must agree that the precise recipe implied by Van Vleck's and Sherman's language is daunting. The use of *characters* of the irreducible representations in dealing with spin state-antisymmetrization problems does not appear to lead to any very useful results. From today's perspective, however, it is known that some irreducible representation matrix elements (not just the characters) are fairly simple, and when applications are written for large computers, the systematization provided by the group methods is useful.

2.4 The Heitler-Rumer Method for polyatomic molecules

Heitler and Rumer [9] gave a generalization of the H_2 molecule results for polyatomic molecules. In these the quantities corresponding to the overlap in the normalization integral (the T in $(1 \pm T)^{-1}$) of Eq. (2) were set to zero, and permutations of higher order than binary were ignored in evaluating matrix elements. For the special case of a central atom, $\mathbf{C}$ of high multiplicity bonded to other atoms, $\mathbf{P}$, $\mathbf{Q}$, ... , they arrived at the total energy for the state of lowest multiplicity,

$$E = E_{\mathbf{C}} + E_{\mathbf{P}} + E_{\mathbf{Q}} + \cdots + J_{\mathbf{CPQ}\cdots} + p_{\mathbf{P}}K_{\mathbf{CP}} + p_{\mathbf{Q}}K_{\mathbf{CQ}} + \cdots - p_{\mathbf{P}}p_{\mathbf{Q}}K_{\mathbf{PQ}} - \cdots,$$

$$(8)$$

where $p_{\mathbf{P}}$ *etc*., are the number of pairs of electrons in the $\mathbf{C}$–$\mathbf{P}$ bond *etc*., $J_{\mathbf{CPQ}\cdots}$ is the simple sum of all of the Coulomb integrals and $K_{\mathbf{CP}}$ *etc*., are the exchange integrals. In addition, this formula requires all of the atoms be in S states. Eq. (8), although fairly impressive, has too many restrictions and approximations to be really satisfactory. In Section 4.1 we return to an examination of some of these approximations.

2.5 Slater's bond functions

Fairly soon after the Heitler-London calculation, Slater, using his determinantal functions, gave a generalization to the *n*-electron VB problem [10]. This was a popular approach and several studies followed exploiting it. It was soon called the method of *bond eigenfunctions*. A little later Rumer [11] showed how the use of these could be made more efficient by eliminating linear dependencies before matrix elements were calculated.

Slater's bond eigenfunctions constitute one choice (out of an infinite number) of a particular sort of basis function to use in the evaluation of the Hamiltonian and overlap matrix elements. They have come to be called the Heitler-London-Slater-Pauling (HLSP) functions. Physically, they treat each chemical bond as a singlet-coupled pair of electrons. This is the natural extension of the original Heitler-London approach. In addition to Slater, Pauling [12] and Eyring and Kimbal [13] have contributed to the method. Our following description does not follow exactly the discussions of the early workers, but the final results are the same.

Consider a singlet molecule with $2n$ electrons, where we wish to use a different atomic orbital (AO) for each electron. We can construct a singlet eigenfunction of the total spin as the product of n electron pair singlet functions

$$\Phi = [\alpha(1)\beta(2) - \beta(1)\alpha(2)][\alpha(3)\beta(4) - \beta(3)\alpha(4)]$$
$$\times \cdots \times [\alpha(2n-1)\beta(2n) - \beta(2n-1)\alpha(2n)], \qquad (9)$$

where, clearly, $S_z\Phi = 0$. Consider the total *spin raising** operator, [14]

$$S = S_x + iS_y = \sum_{k=0}^{2n} S_{xk} + iS_{yk} = \sum_{k=0}^{2n} S_k, \qquad (10)$$

and we operate with it upon Φ. This results in zero, since for every pair function in Eq. (9) there is a corresponding pair of terms in S, and, *e.g.,*

$$(S_i + S_j)[\alpha(i)\beta(j) - \beta(i)\alpha(j)] = [\alpha(i)\alpha(j) - \alpha(i)\alpha(j)]$$
$$= 0.$$

Now, the total spin operator may be written as $S^2 = S^\dagger S + S_z(S_z + 1)$, and, therefore, it is seen that $S^2\Phi = 0$ and is a singlet spin function.

---

*The individual spin raising operator satisfies $S\alpha = 0$ and $S\beta = \alpha$

We now multiply Φ by a product of the orbitals, one for each particle, $u_1(1)u_2(2) \ldots u_{2n}(2n)$, where $u_1, u_2, \ldots u_{2n}$ is some particular ordering of the orbital set. When we apply the antisymmetrizer to the function of space and spin variables, the result can be written as the sum of 2^n SDFs. It is fairly easily seen that there are $(2n)!/(2^n n!)$ different $2n$-electron functions of this sort that can be constructed. Rumer's result, referred [11] to above, shows how to remove all of the linear dependences in this set and arrive at the minimally required number, $(2n)!/[n!(n + 1)!]$, of bond functions to use in a quantum mechanical calculation.

2.5.1 *The perfect pairing function*

We have given a general discussion of the bond eigenfunction method and have pointed out that using all of the Rumer diagrams gives functions that completely span the subspace of the particular configuration addressed. Many of the early calculations used only one of the Rumer functions, and in this case the calculations were called perfect pairing results. Of course, each Rumer function represents perfect pairing between a particular set of orbitals, but the *perfect pairing* approximation always implied that the paired orbitals had a relation to the actual bonding of the molecule.

As an example, consider methane. If the carbon atom L-shell orbitals are arranged as tetrahedral hybrids, we can take the $t_a t_b t_c t_d$ configuration and combine this with an $s_a s_b s_c s_d$ configuration of the four hydrogen atoms. Table 1 shows some numbers of states associated with these orbitals. It is clear that using only the single perfect pairing function represents a considerable constraint upon the wave function. Nevertheless, actual calculations show that it is the largest component of the full wave function, although not overwhelmingly so.

Table 1: Numbers of states under various constraints for methane and four tetrahedral hybrids and four H-atom orbitals.

	Ionic	Number of States
All Singlet States	yes	1764
All States of $^1A_1$ Symmetry	yes	164
All States with $t^4 s^4$	no	86
All $^1A_1$ States with $t^4 s^4$	no	11
All States with $t_a t_b t_c t_d s_a s_b s_c s_d$	no	14
All $^1A_1$ States with $t_a t_b t_c t_d s_a s_b s_c s_d$	no	3
Perfect pairing State ($^1A_1$)	no	1

Pauling's criterion of maximum overlap led to the idea that the tetrahedral hybrids should be the most effective in the perfect pairing wave function. People realized, however, that the effective state of the C atom in this wave function was not the ground state but a mixture of excited states determined by the detailed nature of the state. Van Vleck dubbed this the *valence state* of carbon, and one of the concerns of the early workers was the determination of the energy of this state and the corresponding influence this has upon the C–H bond energy in hydrocarbons. We examine these questions in more detail later in Sec. 4.4, but it must be emphasized that this whole question hinges upon the use of the perfect pairing wave function alone in determining energies.

2.6 Symmetric group theoretic approaches

The early workers, when treating two electron systems, usually made the observation that singlet states spin functions are antisymmetric while triplet spin functions are symmetric with respect to the interchange of particles, *i.e.,*

$$\sqrt{1/2}[\alpha(1)\beta(2) - \beta(1)\alpha(2)] \quad : \quad \text{singlet}$$

$$\begin{array}{c} \alpha(1)\alpha(2) \\ \sqrt{1/2}[\alpha(1)\beta(2) + \beta(1)\alpha(2)] \\ \beta(1)\beta(2) \end{array} \quad : \quad \text{triplet}$$

Consequently, for the total wave function to be properly antisymmetric, the spatial function to be multiplied by the spin functions must be symmetric or antisymmetric for singlet or triplet states, respectively. Satisfying these requirements may be made more explicit in the following way. The antisymmetrizer for two electrons may be written

$$\mathcal{A} = 1/2[I - (12)_r(12)_s], \tag{11}$$

where (12) stands for the binary interchange and the $_r$ subscript indicates this permutation is to be applied to spatial functions and the $_s$ subscript indicates application to spin functions.* We thus factor the permutation into a space

and a spin part. We may write idempotent symmetrizers or antisymmetrizers for either space or spin functions as

$$\begin{aligned} \mathcal{S}_i &= 1/2[I + (12)_i], \\ \mathcal{A}_i &= 1/2[I - (12)_i], \end{aligned}$$

where $i = r$ or s. With this we obtain

$$\mathcal{A} = \mathcal{S}_r\mathcal{A}_s + \mathcal{A}_r\mathcal{S}_s, \tag{12}$$

as a "factored" form of $\mathcal{A}$. We work with Eq. (12) in the following way.

We can use one of the spin eigenfunctions above, symbolizing it by and multiply it by an arbitrary spatial function, Ξ, to obtain a function of both space and spin,

$$\Psi = \Xi\Theta_m^S, \tag{13}$$

which is, of course, not antisymmetric. Applying $\mathcal{A}$ to Ψ we obtain

$$\mathcal{A}\Psi = \mathcal{S}_r\Xi\mathcal{A}_s\Theta_m^S + \mathcal{A}_r\Xi\mathcal{S}_s\Theta_m^S. \tag{14}$$

If Θ_m^S is singlet, only the first term on the right of Eq. (14) survives, and the spatial part of the function, $\mathcal{S}_r\Xi$, is symmetric. For anyone of the three triplet functions the other term on the right of Eq. (14) is the one that survives with the consequence that $\mathcal{A}_r\Xi$ is the required spatial function. These are the familiar results, of course.

We have given a short description of the two electron case. The important point is that there is a generalization of Eq. (12) to n electrons. It takes the general "factored" form

$$\mathcal{A}(1\cdots n) = \sum_i (P_i^{op})_r(Q_i^{op})_s, \tag{15}$$

where P^{op} and Q^{op} are sums of permutations with coefficients that are determined by the irreducible representation matrices of the symmetric group, S_n. We write the general function to be used in our calculations as

$$\Psi = \Xi\Theta_M^S, \tag{16}$$

---

*We write the antisymmetrizer as a properly idempotent operator for this discussion, contrary to the common practice that uses a $\sqrt{1/2}$ prefactor.

where Ξ is an n-electron spatial function and Θ^S_M is an eigenfunction of the total spin. The important result is that $(Q_l{}^{op})_s\Theta^S_M$ set is zero for most of the terms,* and this is the source of the simplifications obtained by using symmetric group methods in atomic or molecular calculations. There is not room here to give further details of these methods, but we do discuss the nature of Θ^S_M.

The n-electron spin functions are sums of products of n α or β functions that satisfy

$$S^2\Theta^S_m = \hbar^2 S(S+1)\Theta^S_m, \text{ and}$$
$$S_z\Theta^S_m = \hbar m\Theta^S_m.$$

Both the S^2 and S_z operators are symmetric sums of operators for each particle and, thus, both commute with every permutation of the n particle labels. Therefore, the eigenfunctions and eigenvalues of S^2 may be classified by the irreducible representations of the symmetric group. The important result is that there is a one-to-one relation between eigenvalues of S^2 and nonequivalent irreducible representations of the groups. We will not give the precise result here, there is a unique generalization of Eq. (12) for the n electron case. Therefore, applying the antisymmetrizer to an n-electron space-spin function of the form $\Xi\Theta^S_m$ results in a space function appropriate to the total spin quantum number S and satisfying the Pauli principle.

Serber [15] has contributed to the analysis of symmetric group methods as an aid in dealing with the twin problems of antisymmetrization and spin state. In addition, Van Vleck espoused the use of the Dirac vector model [16] to deal with permutations. [17] Unfortunately, this becomes more difficult rapidly if permutations past binary interchanges are incorporated into the theory. Somewhat later the Japanese school involving Yamanouchi [18] and Kotani *et al.* [19] also published analyses of this problem using symmetric group methods.

3 History: Post-World War II and automatic computation

The period during and about ten years after World War II saw the beginnings of the development of automatic computing machinery. Although early workers made heroic efforts in many calculations, computers allowed calculations of molecular structure that were far too tedious to undertake by hand or to expect reliable results. The new computers thus allowed many of the quantitative procedures worked out earlier to be checked and accepted or abandoned.

---

*In the two electron case one term was zero and the other not.

Nevertheless, one of the principal developments in the late 1940s was a new way of arranging the orbitals in VB calculations. In this section we start with the Coulson-Fisher approach and follow with other proposals that grow naturally out of it. Much more recent developments in computers have also allowed multiconfigurational VB treatments of a size unimagined 45 years ago, and we also describe these in this section.

3.1 The Coulson and Fisher treatment of H_2

Coulson and Fisher [20] took a new step in molecular calculations with their treatment of H_2 in which the orbitals were non-orthogonal, but extended over both centers. They do not actually call their treatment a VB calculation, but their idea is an important step in the development of the ideas of others who do use the VB label in describing their treatments.

The essence of this method, when illustrated with H_2, is to write the two orbitals for the covalent Heitler-London function as

$$A(\vec{r}) = N(1s'_a + c1s'_b), \text{ and}$$
$$B(\vec{r}) = N(c1s'_a + 1s'_b).$$

The constant c provides a parameter to vary during optimization. They, in effect, used molecular orbital (MO)s in the wave function, but this terminology is not usually used in the current context. The introduction of this sort of orbital provides the same effect as ionic terms in the more traditional treatment. The next two sections give modern extensions of this method.

3.1.1 *Goddard's generalized VB*

Goddard [21] made the earliest important generalization to the Coulson-Fisher method. Goddard's generalized VB (GGVB) wave function is written in terms of orbitals that are linear combinations of the AOs. Using the genealogical set of spin functions in turn and

$$\Psi = \mathcal{A}(1\cdots n)\Xi(1\cdots n)\Theta^S_{Mi}, \tag{17}$$

there are $i = 1, 2, \cdots, f$,

$$f = \frac{2S+1}{n+1}\begin{pmatrix} n+1 \\ n/2 - S \end{pmatrix} \tag{18}$$

different trial wave functions that can be constructed. Goddard designated these as the Gl, G2, . . ., Gf methods, the general one being Gi. For each of these functions the total energy may be optimized with respect to the coefficients in the orbitals. In general, the orbitals are grouped into two sets; orthogonality is enforced within the sets but not between them. Using the calculus or variations in the usual way, one arrives at a set of Fock-like operators that determine the optimum orbitals. The result is a set of f different energies, and one chooses the wave function for the lowest of these as the best GGVB answer. In actual practice only the G1 or Gf methods have been much used.

In simple cases the G1 is a HLSP function while the Gf wave function is a standard tableaux function, which we describe below in Sect. 3.3. For Gf wave functions one may show that the above orthogonality requirement is not a real constraint on the energy. On the other hand, no such invariance occurs with G1 or HLSP functions, so the orthogonality constraint has a real impact on the calculated energy in this case and with all other Gi wave functions.

Goddard and his coworkers applied the method to a number of chemical problems with an emphasis on orbital following results.

3.1.2 *The spin-coupled VB*

Somewhat later Pyper and Gerratt [22] proposed the spin coupled valence bond (SCVB) wave function. Further developments are reviewed by Gerratt, Cooper, and Raimondi [23] in an earlier volume of this series. These workers originally used genealogical spin functions, which produce the *genealogical representation* of the symmetric groups [24], but so long as the irreducible representation space is completely spanned, any representation will give the same energy and wave function. About the same time van Lenthe and Balint-Kurti(25) proposed using an equivalent wave function. The principal differences between these proposals deal with methods of optimization. We will continue to use the SCVB acronym for this method.

We have seen that with a system of n electrons in a spin state S there are, for n linearly independent orbitals, f (given by Eq. (18)) linearly independent spatial functions that can be constructed from these orbitals. In the present notation the SCVB wave function is written as the general linear combination of these.

$$\Psi_{SCVB} = \sum_{i=1}^{f} C_i \phi_i(u_1, \cdots, u_n), \qquad (19)$$

where the orbitals in φ_i are, in general, linear combinations of the whole AO basis.* The problem is to optimize the Rayleigh quotient for this wave function with respect to both the C_i and the linear coefficients in the orbitals. In contrast to the GGVB method the orbitals are subjected to no orthogonality constraints.

Using familiar methods of the calculus of variations, one can set the first variation of the energy with respect to the orbitals and linear coefficients to zero. This leads to a set of Fock-like operators, one for each orbital. Gerratt, *et al.* use a second-order stabilized Newton-Raphson algorithm for the optimization. This gives a set of occupied and virtual orbitals from each Fock operator as well as optimum C_is.

The SCVB energy is, of course, just the result from this optimization. Should a more elaborate wave function be needed, the virtual orbitals are available for a more-or-less conventional, but non-orthogonal configuration interaction (CI) that may be used to improve the SCVB result. Thus improving the basic SCVB result here may involve a wave function with many terms.

SCVB wave functions for very simple systems appear similar to those of the GGVB method, but the orthogonality constraints in the latter have increasingly serious impacts on the results for larger systems.

3.1.3 *The BOVB method*

More recently Hiberty *et al.* [26] proposed the breathing orbital valence bond (BOVB) method, which can perhaps be described as a combination of the Coulson-Fisher method and techniques used in the early calculations of the Weinbaum. [7] The latter are characterized by using differently scaled orbitals in different VB structures. The BOVB does not use direct orbital scaling, of course, but forms linear combinations of AOs to attain the same end. Any desired combination of orbitals restricted to one center or allowed to cover more than one is provided for. These workers suggest that this gives a simple wave function with a simultaneous effective relative accuracy.

---

*The requirements of symmetry may modify this.

3.2 More recent developments in symmetric group methods

Earlier symmetric group procedures were usually based upon the irreducible representation matrices corresponding to the various schemes that had been developed for determining spin eigenfunction. After World War II the earlier work of Young on symmetric groups found application to the problems of implementing VB ideas. Matsen and coworkers [27] introduced what they termed a spin-free approach. Somewhat later the present author [30] introduced VB basis functions based upon Young's *standard tableaux* representation.

All methods produce one or another of the infinity of irreducible representations of the symmetric groups, and, if basis sets always completely span the representation, the quantum mechanical results are the same. One of the advantages of Young's procedure is the way it clearly shows the connections among the various ways that basis sets can be arranged.

The concept of the tableau is central to Young's theory, and we use only the portions of the theory necessary to discuss VB theory. For a particular set of n orbitals $u_1 \ldots u_n$ and n electrons, symbols for the orbitals may be arranged in a two-column table, in which the two columns are not necessarily the same length,

$$
\begin{bmatrix}
u_1 & u_{n-k+1} \\
\vdots & \vdots \\
\vdots & u_n \\
u_{n-k} &
\end{bmatrix}.
$$

The difference in the lengths of the columns is related to the spin; the total spin quantum number is $S = n/2 - k$. Clearly, $k \leq n/2$. In the tableau the orbitals are associated with particles labeled sequentially down the first and then down the second column. The subscripts on the orbitals label the functions, not the arguments.

Young defined two operators, the row symmetrizer $\mathcal{P}$ and the column antisymmetrizer $\mathcal{N}$, and we assume the these operate on (permute) the particle labels not the orbital labels. Each tableau designates a product of orbitals with a particular ordering

$$
\Xi = u_1(1)u_2(2) \cdots u_n(n). \tag{20}
$$

As the names suggest $\mathcal{P}$ is the product of k symmetrizing operators for the particles in the rows,

$$
\mathcal{P} = \tfrac{1}{2}[I + (1, n-k+1)] \cdots \tfrac{1}{2}[I + (k, n)], \tag{21}
$$

and $\mathcal{N}$ is the product of the two antisymmetrizers of the columns

$$
\mathcal{N} = \mathcal{A}(1 \cdots n - k)\mathcal{A}(n - k + 1 \cdots n). \tag{22}
$$

The (i, j) symbol in Eq. (21) stands for a binary interchange of the particles indicated. It will be observed that the particles operated upon in these operators are related closely to the way the particle labels occur in the tableau. As we have defined them, $\mathcal{P}$ and $\mathcal{N}$ are strictly idempotent.

Using the operators we have defined and the spatial function Ξ, new functions may be constructed, *e.g.*, $\mathcal{P}\Xi$. It should be clear that this function now is insensitive to the positions of orbitals in the first k rows, *i.e.*, one could interchange u_1 and u_{n-k+1}, for example, without changing $\mathcal{P}\Xi$. Similarly, any rearrangement (permutation) of the orbitals in a column will do no more than change the sign of $\mathcal{N}\Xi$. Permutations that change both the row and column position of orbitals will result in changing these projected functions.

Another central result of Young's work, when stated in our current language, is that $\mathcal{P}\mathcal{N}\Xi$ is equivalent to the perfect pairing function of Slater with the orbitals in the same rows paired. [27] At what might be called the other extreme, Heitler's and Rumer's early work assumed that diatomic molecules interacted with the atoms in their highest spin states consistent with the configuration, and these functions are equivalent to $\mathcal{P}\mathcal{N}\Xi$, where the orbitals in a column are associated with one of the atoms. A polyatomic analog of this situation exists. Thus, merely inverting the order in which the operators are applied, passes from one type of function to the other.

In discussions of the total spin [31] of multielectron systems, the *spin branching diagram* is frequently used. Fig. (5) shows a version. The $\mathcal{N}\mathcal{P}$ operator corresponds to the branch in the diagram where the lowest line is always taken and the $\mathcal{P}\mathcal{N}$ operator to the branch where the highest possible branch is taken. The two Young operators thus correspond to the first and last rows of the genealogical irreducible representations of the symmetric groups, and, hence, to Goddard's G1 and Gf "methods," respectively. Therefore, Young's tableaux and the corresponding operators constitute a way of, at least partly, unifying the various techniques that have been devised for dealing with spin and antisymmetrization and VB calculations.

As a last point we note that the present author and his coworkers [36] devised an algorithm for the calculation of matrix elements of the overlap and Hamiltonian based upon the $\mathcal{P}\mathcal{N}$ operator that is n^5 in its worst case,
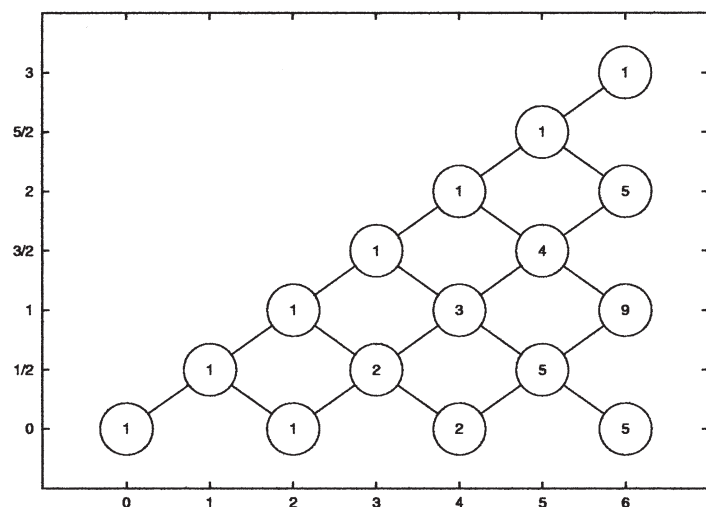
Figure 5: The spin branching diagram for 0 to 6 electrons (horizontal axis). The total spin quantum number is on the vertical axis. The numbers in the circles give the spin degeneracies.

where n is the number of electrons. There are reasons to believe that this is the best exponent that can be achieved. Transformation to the $\mathcal{NP}$ functions is possible when desired.

3.3 Multiconfiguration methods

The original Heitler-London treatment with its various extensions was a VB treatment that included several configurations, *e.g.*, the total wave function is a sum of terms with spatial functions made up of different subsets of the orbitals. This is the essence of *multiconfiguration* methods. The most direct extension of this sort of approach is, of course, the inclusion of larger numbers of configurations and the application to larger molecules. The increase in computational power allowed calculations of this sort.

At the same time molecular orbital (MO) methods were seeing a rapid development, also because of increased computational ability. These, at least on the surface, appear to provide a simpler approach to molecular structure calculations. Nevertheless, Matsen and Browne [32] made a forceful case

---

*They called their suggested procedure an atomic orbital configuration interaction (AOCI) calculation.

for the use of MCVB methods,* indicating the difficulties that the enforced orthogonality in molecular orbital configuration interaction (MOCI) calculations cause with processes that involve large scale relative motions of the nuclei.

3.3.1 *The multistructure procedure of Balint-Kurti and Karplus*

Balint-Kurti and Karplus [28] implemented an earlier suggestion of Moffit [29] for the evaluation of matrix elements of the Hamiltonian by transforming the AOs to an orthogonalized set. If carried out correctly, this involves no approximations. The method was applied to *ab initio* and empirically corrected calculations of LiF, F_2, and F_2^-. The transformation of the matrix elements to the orthogonalized form can be quite time consuming for large bases.

3.3.2 *The MCVB method*

The present author and his coworkers [36] devised the multiconfiguration valence bond (MCVB) procedure. These calculations involve a direct attack on the problem of evaluating matrix elements between n-electron functions of non-orthogonal orbitals. The algorithm depends upon the symmetric group methods of Young and the $\mathcal{PN}$ operator. Although there is considerable flexibility allowed in the construction of basis sets, a treatment that uses a full or nearly full set of n-electron functions based upon a minimal AO set and "excitations" into n-electron functions containing orbitals designed to provide scaling has been a generally useful strategy. As was mentioned above, these wave functions are a generalization of the original Heitler-Rumer high spin atomic calculations. If the results are of interest, a simple transformation to a wave function that is a sum of HLSP functions is possible. With today's computers calculations consisting of $> 10^5$ individual n-electron basis functions can be more or less routine.

4   Early ideas

In reviewing the history of VB methods there stand out a few ideas concerning approximations that might be made. The author has chosen four that allow simple computational tests in today's world, and these are discussed in this section. There is little connection between them.

4.1 Overlap matrices and the neglect of some permutations

When the actual Heitler-London treatment of H_2 is generalized to n electrons, the matrix elements that arise involve permutations of higher order than binary. When calculations had to be done by hand, the complexities could mount rapidly. It was perhaps natural, if not strictly rigorous, for people to make the approximation of neglecting these higher order permutations. There was actually much debate about the validity of such an approximation, in general, in spite of its crudeness for H_2. Clearly in Eq. (2), if the binary permutation would be ignored completely, the same energy would be obtained for the singlet and triplet states. When it came to considering the denominator, however, it seemed to the early workers as if the $T (= S^2)$ might be a higher order effect, and suggestions were made that it might be safely ignored. Generalizing this led to the idea for n-electron systems that the above mentioned triple, quadruple, and higher permutations might be usefully ignored.

This question was not considered completely academic. In Heisenberg's [33] original theory of ferromagnetism the overlaps between the orbitals at the various sites were ignored. Inglis [34] criticized this, but suggested that including overlaps made the calculation meaningless since the correction due to them scales as n, the number of sites involved. Later, Van Vleck [35] showed that Inglis' objection ignored cancellations that mitigate the problem. We will not examine the ferromagnetism problem, but will undertake a less ambitious course and investigate the contribution of various orders of permutations to the value of the normalization constant for VB wave functions.

The $(1 \pm T)$ in Eq. (2) arises from the normalization of the wave function for H_2. In this section we will investigate the extent to which it might be permissible to ignore the permutations of some order and higher when normalizing a VB function for n electrons. We shall do this for a standard tableau function, where we have an expression for the wave function of any multiplicity.

Therefore, consider a standard tableaux function with orbitals u_1, u_2, ..., u_n, where they need not all be different, of course,

$$\begin{bmatrix} u_1 & u_{n-k+1} \\ \vdots & \vdots \\ & u_n \\ \vdots & \\ u_{n-k} & \end{bmatrix}.$$

The orbitals are assumed real, normalized, but not necessarily orthogonal. The overlaps are symbolized by $S_{ij} = S_{ji} = \langle u_i / u_j \rangle$. It is shown elsewhere [36] that the normalization constant for such a standard tableaux function can be written as the integral of a functional determinant,

$$C^{-2} = (n - k + 1) \int_0^1 \begin{vmatrix} A & qB \\ qB^\dagger & C \end{vmatrix} (1 - t)^{n-k} dt, \qquad (23)$$

where $q = \sqrt{t/(1 - t)}$. It is observed that q is pure imaginary. The determinant is therefore that for a symmetric matrix, but not an Hermitian one. In Eq. (23) A is the $(n - k) \times (n - k)$ overlap matrix of the first-column orbitals, C, the corresponding $k \times k$ matrix for the second-column orbitals, and B the $(n - k) \times k$ matrix of the inter-column overlaps. A, C, and the overall matrix are symmetric. Eq. (23) is also written with all of the purely group theoretic factors implicit in the functions. This would make $C^{-2} = 1$ if the overlaps between all pairs of orbitals were zero, and, thus, we are considering only that part of the normalization constant that is affected by the overlaps. The overall matrix is diagonalizable by an orthogonal matrix, which is also a function of q, of course. We are actually not interested in the transformation matrix, but only the characteristic polynomial of the overall matrix. To proceed we prove a theorem.

Consider an $N \times N$ symmetric matrix S that has principal diagonal elements all equal to one.*

Theorem 1 *A simple transformation of the characteristic polynomial of such a matrix will present it in a form where the contribution from each order of permutation to the value of its determinant is displayed as an elementary symmetric function of the eigenvalues of $S - I$.*

Consider the determinant

$$|I + t(S - I)|,$$

which is a polynomial in t that may be written

$$|I + t(S - I)| = 1 + \sum_{l=2}^N s_l t^l. \qquad (24)$$

Clearly, the sum is just the determinant |S| when $t = 1$, and a little reflection will convince one that s_l is the contribution from the l-order permuted indices. The term with $l = 1$ is zero, of course, since there can be no permutation of one object.

_____

*We write this with the symbol, S, since the overlap matrix is the sort we consider.

Let O be the orthogonal matrix that diagonalizes $(S - I)$. Then

$$(S - I) = O\,\text{diag}(d_1 d_2 \cdots d_N)O^\dagger, \tag{25}$$

and we rewrite the determinant of Eq. (24),

$$
\begin{aligned}
|I + t(S - I)| &= |I + tO\,\text{diag}(d_1 d_2 \cdots d_N)O^\dagger|, \\
&= \prod_{m=0}^{N}(1 + td_m), \\
&= \sum_{m=0}^{N}\sigma_m t^m, \tag{26}
\end{aligned}
$$

where σ_m is the mth-order elementary symmetric function [37] of the eigenvalues of $S - I$, each of which is one less than the corresponding eigenvalue of S. Equating coefficients of equal powers of t in our two expressions we have $s_l = \sigma_l$. The elementary symmetric functions are simple to determine recursively from the d_m.* Indeed, the algorithm is essentially that to determine binomial coefficients, as is evident from Eq. (26) if we were to set each $d_m = 1$. We note that σ_1 is the trace of $S - I$, which is zero, so that s_1 is also zero as it should be.

We consider the application of this theorem to the evaluation of the integral in Eq. (23) for an STO3G basis calculation of CH_4 and a π-only calculation of naphthalene. As indicated earlier, we do not attempt to address the ferromagnetism problem, but we can note that the overlaps in naphthalene much more resemble the magnetism system than do the overlaps in a small compact molecule like CH_4.

4.1.1 Sums of permutations of the same order

It is useful to examine the symmetric functions of Eq. (26) for the $n \times n$ matrix

$$
S = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \cdots & \vdots \\
1 & 1 & \cdots & 1
\end{bmatrix}, \tag{27}
$$

---

*For our work we really do not need to diagonalize $S - I$. A simpler procedure is to tridiagonalize it; the characteristic equation is available therefrom by an easy recursion.

which is, of course, invalid as a legitimate overlap matrix. It does, however, allow us to get some idea of the limits that the symmetric functions can attain when real overlap matrices are used.

The matrix of Eq. (27) minus the identity has for eigenvalues $n - 1$ once and -1 $n - 1$ times. Eq. (26) now gives us

$$
\begin{aligned}
|I + t(S - I)| &= (1 - t)^{n-1}[1 + (n - 1)t] \\
&= \sum_{k=0}^{n}(-1)^k(1 - k)\binom{n}{k}t^k,
\end{aligned}
$$

where the standard symbol for the binomial coefficient has been used. The significance of this result should be clear. When we consider permutations that reorder k indices, the coefficient of t^k is the number of even permutations of that order minus the number of odd permutations of the same order. We note that the coefficient of t is zero, as it should be, and the coefficients of t^2 and t^3 are just minus the number of binary interchanges and plus the number of ternary permutations, respectively. All other terms involve differences between numbers of even and odd permutations. In the next two sections we consider the overlap matrices for realistic systems.

4.1.2 Application to the π-system of naphthalene

A ten electron system with each electron in a different orbital could have a multiplicity of 1, 3, 5, 7, 9, or 11. The singlet and possibly the triplet states are the only physically interesting cases, but we give all of them so that trends may be observed. The undecet case has some mathematical interest, since it is just the determinant of the overlap matrix. Table 2 gives our results for the first three of the possible multiplicities and Table 3 gives the other three. The tables are arranged in columns showing the order of the permutation and the values and the accumulated sums for each order and the integral of Eq. (23). It should be clear that these orders represent the number of indices permuted at each stage. Except for orders 2 and 3, however, they involve permutations with different signatures. Order 4 can have, *e.g.,* the permutations (12)(34) and (1234). These both involve four indices, but the first is an even permutation and the second is odd. Of course, only the antisymmetrizer (undecet case) has ± 1 coefficients that exactly match the corresponding permutation's signature. The permutation operators giving other spin values are more complicated, and it would be difficult to give rules for the way the terms vary with order.

The $2p_z$ orbitals in naphthalene all have nearest neighbor distances that are quite close to one another, and the nearest neighbor overlaps do not vary much on either side of 0.32. With such a set of overlaps, the normalization constant does not vary greatly with spin state. Even with a fairly small over-

Table 2: Convergence of normalization constants for singlet, triplet, and quintet standard tableaux functions in the π-system of naphthalene.

Order	Singlet		Triplet		Quintet	
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	−0.0000	1.0000	0.0000	1.0000	−0.0000	1.0000
2	−0.7896	0.2104	−0.9204	0.0796	−0.9295	0.0705
3	0.0693	0.2796	0.1066	0.1862	0.1223	0.1927
4	0.2017	0.4814	0.2677	0.4539	0.2570	0.4497
5	−0.0320	0.4494	−0.0482	0.4057	−0.0481	0.4016
6	−0.0142	0.4351	−0.0262	0.3795	−0.0254	0.3762
7	0.0020	0.4371	0.0043	0.3837	0.0055	0.3817
8	0.0002	0.4373	0.0011	0.3849	0.0005	0.3823
9	−0.0000	0.4373	−0.0001	0.3848	−0.0002	0.3821
10	0.0000	0.4373	−0.0000	0.3848	0.0000	0.3821

Table 3: Convergence of normalization constants for heptet, nonet, and undecet standard tableaux functions in the π-system of naphthalene.

Order	Heptet		Nonet		Undecet	
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	−0.0000	1.0000	−0.0000	1.0000	0.0000	1.0000
2	−0.9346	0.0654	−0.9478	0.0522	−1.1902	−0.1902
3	0.1269	0.1923	0.1487	0.2009	0.2168	0.0267
4	0.2526	0.4448	0.2397	0.4407	0.4061	0.4327
5	−0.0422	0.4026	−0.0505	0.3901	−0.1051	0.3277
6	−0.0278	0.3748	−0.0173	0.3729	−0.0469	0.2808
7	0.0044	0.3792	0.0030	0.3759	0.0128	0.2936
8	0.0013	0.3805	0.0004	0.3763	0.0019	0.2955
9	−0.0001	0.3804	−0.0000	0.3763	−0.0004	0.2951
10	−0.0000	0.3804	0.0000	0.3763	−0.0000	0.2951

lap such as we have here, the sums nevertheless require the inclusion of terms up to order 5 or 6 to reach a number close to their final values. As we see, the value of C^{-2} is smallest for the undecet case.

We note that the order 2 term for the highest multiplicity is the most negative. This must be the sum $-\Sigma S_{ij}^2$ in this case, and so it consists of all negative terms.

4.1.3 Application to CH₄

An STO3G basis applied to CH_4 at its equilibrium geometry yields 9 AOs, and, if the C $1s$ orbital is relegated to "core" [36] status, there are only eight orbitals and eight electrons to go into them. For illustration purposes we consider C^{-2} for the AO set $\{2s, 2p_x, 2p_y, 2p_z, 1s_a, 1s_b, 1s_c, 1s_d\}$. In Table 4 we show the values of each of the terms for different orders of permutations and also the accumulated sum, which gives information about the rate of convergence. Table 5 gives similar results for the heptet and nonet states. Among these values, only the singlet has any great physical interest, but we again

Table 4: Convergence of normalization constants for singlet, triplet, and quintet standard tableaux functions in CH_4.

Order	Singlet		Triplet		Quintet	
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	−0.0000	1.0000	−0.0000	1.0000	−0.0000	1.0000
2	0.3824	1.3824	−0.1323	0.8677	−0.5916	0.4084
3	−0.0359	1.3465	−0.0798	0.7879	−0.1242	0.2842
4	0.2049	1.5514	−0.0101	0.7777	0.0590	0.3432
5	−0.0237	1.5277	−0.0113	0.7664	0.0383	0.3815
6	0.0916	1.6193	−0.0132	0.7532	0.0051	0.3866
7	−0.0065	1.6128	−0.0002	0.7530	0.0023	0.3889
8	0.0378	1.6506	−0.0038	0.7492	0.0025	0.3914

Table 5: Convergence of normalization constants for heptet and nonet standard tableaux functions in CH_4.

Order	Heptet		Nonet	
0	1.0000	1.0000	1.0000	1.0000
1	−0.0000	1.0000	−0.0000	1.0000
2	−1.0190	−0.0190	−2.0434	−1.0434
3	−0.1690	−0.1880	0.2331	−0.8103
4	0.3249	0.1369	1.3421	0.5318
5	0.1228	0.2597	−0.1549	0.3770
6	−0.0166	0.2431	−0.3707	0.0063
7	−0.0227	0.2205	0.0261	0.0324
8	−0.0054	0.2151	0.0378	0.0702

give all so that the trends can be seen. In general, as the multiplicity increases, the value of C^{-2} decreases. The overlaps within this basis are not all positive, so it is difficult to make specific predictions.

The overlaps in this molecule are rather larger than was the case with naphthalene. The largest is near 0.5. This results in a larger value for the singlet state and rather smaller value for the nonet state.

4.2 Orthogonalized AOs

In a fairly early discussion of solids Wannier [38] showed how linear combinations of the AOs could be made that rendered the functions orthogonal while retaining a relatively large concentration on one center. In more modern language we would now say that he used a *symmetric orthonormalization* of the AO basis. If we symbolize the overlap matrix for the AO basis by S, then any matrix N that satisfies

$$N^\dagger S N = I, \tag{28}$$

constitutes an orthonormalization of the basis. This requirement on N is insufficient to define it uniquely. Additional conditions could include:

1. Require N be upper triangular. This gives the traditional Schmidt orthonormalization.

2. Set $N = U\mathrm{diag}(s_1^{-\frac{1}{2}}, s_2^{-\frac{1}{2}}, \ldots, s_n^{-\frac{1}{2}})$ where U is the unitary matrix diagonalizing S and $s_1, s_2, \ldots, s_n$ are the eigenvalues. This gives the canonical orthonormalization.

3. Set $N = S^{-\frac{1}{2}}$. This gives the *symmetric orthonormalization*, so-called because this N is a symmetric matrix for real basis functions.

An important property of the symmetric orthonormalization is that it produces a new set of orbitals that are the closest possible to the original set in a least squares sense. Since evaluating matrix elements of the Hamiltonian is always much easier with orthonormal orbitals, this change had great attractions for early workers. Unfortunately, it has developed that this idea must be used with great care. The requirement of closeness in the least squares sense, although almost always well defined, does not guarantee that the resulting two orbital sets are close to one another in a physically useful sense.

We may demonstrate this difficulty by giving a result due to Slater. [39] Applying asymmetric orthonormalization to the basis normally used in the Heitler-London calculation we have a H$1s$ function on each of two centers,

$1s_a$ and $1s_b$. The overlap matrix for this basis is

$$\bar{S} = \begin{bmatrix} 1 & S \\ S & 1 \end{bmatrix}, \tag{29}$$

and the inverse square root is

$$\bar{S}^{-1/2} = \begin{bmatrix} \frac{1}{2\sqrt{1+S}} + \frac{1}{2\sqrt{1-S}} & \frac{1}{2\sqrt{1+S}} - \frac{1}{2\sqrt{1-S}} \\ \frac{1}{2\sqrt{1+S}} - \frac{1}{2\sqrt{1-S}} & \frac{1}{2\sqrt{1+S}} + \frac{1}{2\sqrt{1-S}} \end{bmatrix}, \tag{30}$$

where $S = \langle 1s_a | 1s_b \rangle$, and the signs are appropriate for $S > 0$. This orthogonalization gives us two new functions

$$\begin{aligned} |A\rangle &= P|1s_a\rangle + Q|1s_b\rangle, \\ |B\rangle &= Q|1s_a\rangle + P|1s_b\rangle, \end{aligned}$$

where

$$\begin{aligned} P &= \frac{1}{2\sqrt{1+S}} + \frac{1}{2\sqrt{1-S}}, \\ Q &= \frac{1}{2\sqrt{1+S}} - \frac{1}{2\sqrt{1-S}}. \end{aligned}$$

We use these in a single Heitler-London covalent configuration, $A(1)B(2) + B(1)A(2)$, and calculate the energy. When $R \to \infty$ we obtain $E = -1$ au, just as we should. At $R = 0.741$ Å, however, where we have seen that the energy should be a minimum, we obtain $E = -0.6091$ au, much higher than the correct value of -1.1744 au. The result for this orthogonalized basis, which represents no binding and actual repulsion, could hardly be worse.

Slater says surprisingly little concerning this outcome, but, in light of present understanding, we may say that the symmetric orthonormalization gives very close to the poorest possible linear combination for determining the lowest energy. This results from the added kinetic energy of the orbitals produced by a node that is not needed. Alternatively, we may say that we have used antibonding rather than bonding orbitals in the calculation. We have here a good example of how unnatural orthogonality between orbitals on different centers can have serious consequences for obtaining good energies and wave functions.

We add another comment about this example and note that using symmetric orthonormalization on the simple two AO basis for the triplet state of H$_2$ gives the same answer as that obtained with unmodified orbitals. Since the

triplet state is represented by the antisymmetric combination of the orbitals, it is invariant to any nonsingular transformation of the two orbitals.

4.3 Relation of Hamiltonian matrix to overlap matrix

In work on the electronic structure of solids, Lowdin [40] pointed out that if the Hamiltonian matrix for a system were a polynomial function of the overlap matrix of the basis, H and S would have the same eigenvectors and the energy eigenvalues would be polynomial functions of the eigenvalues of S. A number of consequences of this sort of relationship are known, but so far as the author is aware, no tests of such an idea have ever been made with realistic H and S matrices. This may be accomplished by examining the commutator, since if

$$H = \sum_k a_k S^k, \qquad (31)$$

H and S clearly commute, and this would be true even if the sum in Eq. (31) were a convergent infinite series, rather than a polynomial. Conversely, if the two matrices do not commute, no relation like Eq. (31) connects them. Even if H and S are functionally independent, one still might argue that the commutator is likely to be small, and, thus, the idea could be a useful approximation. The difficulty here is with the subtleties of the concept of smallness in this context. We will not attempt to address this question quantitatively, but satisfy ourselves by examining the commutators of H and S for three systems. The first of these is a simple 2×2 system for which we may obtain an algebraic answer. The other two are matrices from real VB calculations of CH_4 and the π-system of naphthalene.

4.3.1 *A 2 × 2 system*

Let

$$H = \begin{bmatrix} A & B \\ B & C \end{bmatrix}, \qquad (32)$$

and

$$S = \begin{bmatrix} 1 & s \\ s & 1 \end{bmatrix}. \qquad (33)$$

The commutator of these two is

$$HS - SH = \begin{bmatrix} 0 & -(C-A)s \\ (C-A)s & 0 \end{bmatrix}, \qquad (34)$$

and we see immediately that the commutator is zero if the two diagonal elements of H are the same.

We may write H as two terms, the first a part that is a polynomial function of S and the second a sort of remainder.

$$H = \begin{bmatrix} (A+C)/2 & B \\ B & (A+C)/2 \end{bmatrix} + \begin{bmatrix} (A-C)/2 & 0 \\ 0 & (C-A)/2 \end{bmatrix}. \qquad (35)$$

Thus, we see in this simple case that the closeness of the approximation depends upon the size the second term in Eq. (35); whether it is really a small perturbation upon the system. With these matrices the approximation would be good only if the two diagonal elements of H are close in value. The 2×2 case is rather special, however, and we give further more complicated examples.

4.3.2 The π-system of naphthalene

For naphthalene we examine the H and S matrices based upon the both the HLSP functions and the standard tableaux functions for the system. In both cases we include the non-ionic structures, only. This will give a picture of how the situation compares for the two sorts of basis functions. In both cases, of course, the dimensions of the matrices are 42×42, the number of non-ionic Rumer diagrams for a naphthalene structure. Some statistics concerning the commutator are shown in Table 6. It is clear that, while there are quanti-

Table 6: Statistics on commutator $HS - SH$ matrix elements for naphthalene. Lower triangle only. All are energies in Hartrees.

	HLSP	STF
Maximum element	6.8380×10^{-1}	3.5665×10^{-1}
Minimum element	-6.1237×10^{-1}	-1.4897×10^{-1}
Minimum absolute value	7.9021×10^{-9}	1.1904×10^{-4}
Average(Commutator)	1.5308×10^{-3}	1.0838×10^{-2}
RMS(Commutator)	2.0458×10^{-1}	5.7002×10^{-2}
Average($H_{ii} - H_{jj}$)	1.3997×10^{-1}	2.5172×10^{-1}
RMS($H_{ii} - H_{jj}$)	1.1386×10^{-1}	3.2597×10^{-1}

tative differences between the two bases, qualitatively the results are similar. It should be emphasized that if the commutator $HS - SH$ were zero for one of the bases, it would also be for the other. The important point to be gleaned from Table 6 is that the root-mean-square (RMS) values of the commutator elements and the $H_{ii} - H_{jj}$ differences are all very similar. The conclusion is that the perturbation presented by the non-commuting part of H is not small in this case, and it would be a bad approximation to consider H to be a polynomial function of S.

4.3.3 *The CH$_4$ molecule*

When an STO3G AO basis full VB calculation of CH_4 is carried out, there are 1716 singlet standard tableaux functions all together. When these are combined into functions of symmetry $^1A_1$ the number of independent linear combinations is reduced to 164. Thus the symmetry factored H and S matrices are 164×164. We show the statistics for the $HS - SH$ matrix for standard tableaux functions in Table 7. The statistics for HLSP functions are not available in this case. It is immediately obvious that the numbers for CH_4 are considerably larger than they were in the case of naphthalene; the RMS value of the commutator elements is nearly 5 times the RMS value of $H_{ii} - H_{jj}$. When one considers this in comparison with the results for naphthalene, it is not too surprising, since the π system for that molecule involves AOs of only one kind, whereas with CH_4 there are AOs from both K and L shells of the carbon. In spite of the large deviations between diagonal elements of H, the RMS average of the commutator elements is still larger, as

Table 7: Statistics on commutator $HS - SH$ matrix elements for CH_4. Lower triangle only. All are energies in Hartrees.

	STF
Maximum element	$1.5946 \times 10^{+1}$
Minimum element	$-1.6021 \times 10^{+1}$
Minimum absolute value	1.3350×10^{-4}
Average(Commutator)	-2.0447×10^{-2}
RMS(Commutator)	4.0189
Average($H_{ii} - H_{jj}$)	-3.0335×10^{-1}
RMS($H_{ii} - H_{jj}$)	8.7425×10^{-1}

was emphasized above. The non-commuting part of H is very large here and represents a large perturbation. Ignoring it would constitute a very crude approximation.

4.4 The perfect pairing wave function and the valence state of carbon

We have defined the "perfect pairing" wave function earlier, and in this section we will examine some of the effects using this function alone has on the energies. This will parallel some of the early treatments, but it is not simple to use the computer programs current today to give an exactly comparable calculation to those carried out in the early days of molecular theory. There are two significant differences. The first is that all early calculations on a molecule as large as methane were semiempirical, at least to some extent. The second is that they also neglected higher order permutations in the evaluation of matrix elements. These two approximations interact to some extent, of course, but, in any event, would be difficult to arrange in a modern program.

In Table 8 we give the results for several different wave functions and two different basis sets.

1. STO3G. This is the conventional representation of Slater type orbitals using three Gaussians apiece. [41]

2. EOP3G. This basis is the energy optimized three Gaussian basis set devised by Ditchfield *et al.* [42] This is very nearly the same as the (33/3) basis given by Huzinaga *et al.* [43]

In each of these there are four valence orbitals on carbon and one on each hydrogen for a total of eight.

Seven different results are given for each basis set, and in all of them the C $1s$ orbital is doubly occupied in a frozen core. They are coded as follows:

1. FV. The full valence MCVB. According to the Weyl dimension formula eight electrons and eight orbitals give 1716 basis functions, and these support 164 $^1A_1$ states. The energies for these wave functions at the geometry of the minimum are given as zero in Table 8. All other energies in each column are given relative to this one, which is the lowest in each case. The absolute energies are given in a footnote in the table, and the absolute en-

Table 8: Energies for various states and wave functions of CH_4. These are valence only calculations with a C $1s$ frozen core.

Code[a]	STO3G		EOP3G	
	E_{Min}	$D_e{}^b$	E_{Min}	$D_e{}^b$
FV[c]	0.000 eV	19.491 eV	0.000 eV	17.307 eV
HFC	3.853	18.655	3.055	17.392
HPP	4.138	22.898	3.239	20.992
CFC	9.968	12.540	7.488	12.959
HSTF	10.912	11.596	8.525	11.922
CSTF	10.912	11.596	8.525	11.922
CPP[d]	23.895	5.405	19.575	6.795

[a] FV, full valence; HFC, hybrid full covalent; HPP, hybrid perfect pairing; CFC, Cartesian full covalent; HSTF, hybrid stf; CSTF, Cartesian stf; CPP, Cartesian perfect pairing. See the text for further details.
[b] The total four-bond dissociation energy for the corresponding wave function.
[c] The full valence total energies: STO3G, –39.80107 au; EOP3G, –39.97968 au
[d] Not an A_1 state. See text.

ergy of anyone of the states may be reconstructed if so desired. For this calculation we need not differentiate between tetrahedral hybrid and Cartesian p orbitals.

2. HFC. The carbon orbitals are formed into the standard tetrahedral hybrids, "pointing" at the H atoms. There are 14 covalent basis functions and this row gives the relative energy for the 14 term wave function.

3. HPP. This is the single perfect pairing HLSP function with tetrahedral hybrids. At the geometry of the energy minimum this function is no more than 0.2–0.3 eV higher than the HFC wave function. This difference represents the deviation from perfect pairing that occurs with the covalent only functions. This row also has the largest dissociation energies, since the C atom is forced into the "valence state" of van Vleck at the dissociated geometry.

4. CFC. The standard Cartesian $2p_x$, $2p_y$, and $2p_z$ orbitals together with the unchanged $1s$ orbital are used in the 14 term covalent wave function. This change produces a considerably larger jump in the energy than those before.

5. HSTF. This is the single best standard tableaux function with the hybrid orbitals. It corresponds to the high-spin wave function of Heitler and Rumer and has C in its ^5S state exactly.

6. CSTF. These energies are the same as the previous set, since the C ^5S state is equally well described by the Cartesian or the hybrid orbitals.

7. CPP. The Cartesian perfect pairing wave function is by far the worst on the energy scale, but this arrangement of AOs is not really applicable to the present discussion. It is unclear, of course, even how to pair the orbitals in this case, and, although it is the energy of a singlet state, unlike all the others, a single function cannot have A_1 symmetry with this sort of wave function and, thus, does not approximate an energy eigenstate.

Voge [44] used the conventional techniques* of the time to determine the actual atomic carbon states in the "valence" state. Table 9 shows the populations of atomic states that Voge determined. Nevertheless, the valence state concept, although well defined, seems artificial today, since it is not experimentally available and since full calculations are so easily accessible and give better results.

Table 9: Populations of carbon atom states in "valence state."

State		Population
s^2p^2	^3P	0.1406
	^1D	0.0466
sp^3	^5S	0.3125
	^3P	0.2820
	^1D	0.0313
p^4	^3P	0.1406
	^1D	0.0466

There is, however, interest in examining some energy differences from Table 8. We may estimate the energies of the valence and the ^5S states (above the calculated ground state), and these are shown in Table 10. Thus, the HPP row shows the perfect pairing valence state to be around 7 eV above the ground state, similar to the value obtained by van Vleck. The row marked CSTF gives the estimated energy of the ^5S state, and it is seen to be about 1

I.e., neglecting higher order permutations in evaluating Hamiltonian matrix elements and even binary permutations in the overlaps.

Table 10: Energies of C atom states at asymptotic C + 4H distances.

	STO3G	EOP3G	Exp.
HPP	7.545	6.924	NA
CSTF	3.017	3.140	4.183

eV below the experimental value. This is expected since there should be more correlation energy in the ground state than in the ^5S state, and these bases are too restricted to give any good account of correlation.

Both the historical results and the modern indicate that, without a doubt, the excited valence configuration, sp^3, figures large in bonding in the CH_4 molecule. The hybridized orbitals give a better energy in the restricted calculations than do the Cartesian, but, of course, this difference goes away for the full calculations. These have no early counterpart, of course.

5   Acknowledgment

APPENDIX — Acronyms

AO	atomic orbital
AOCI	atomic orbital configuration interaction
BOVB	breathing orbital valence bond
CI	configuration interaction
GGVB	Goddard's generalized VB
HLSP	Heitler-London-Slater-Pauling
MCVB	multiconfiguration valence bond
MO	molecular orbital

MOCI	molecular orbital configuration interaction
RMS	root-mean-square
SCVB	spin coupled valence bond
SDF	Slater determinantal functions
STF	standard tableau function
VB	valence bond

References

[1] W. Heitler and F. London, Z. Physik **44**, 619 (1927)

[2] P. A. M. Dirac, Proc. Roy. Soc. (London) **A123**, 714 (1929)

[3] J. H. Van Vleck and A. Sherman, Rev. Mod. Phys. **7**, 167 (1935)

[4] S. C. Wang, Phys. Rev. **31**, 579 (1928)

[5] L. Pauling and E. B. Wilson, "Introduction to quantum mechanics," (Mc-Graw-Hill, New York 1935)

[6] N. Rosen, Phys. Rev. **38**, 2099 (1931)

[7] S. Weinbaum, J. Chem. Phys. **1**, 593 (1933)

[8] H. M. James and A. S. Coolidge, J. Chem. Phys. **1**, 825 (1933)

[9] W. Heitler and G. Rumer, Z. Physik **68**, 12 (1931)

[10] J. C. Slater, Phys. Rev. **38**, 1109 (1931)

[11] G. Rumer, Göttinger Nachr. **1932** 377

[12] L. Pauling, J. Chem. Phys. **1**, 280 (1933)

[13] H. Eyring and G. E. Kimbal, J. Chem. Phys. **1**, 239 (1933)

[14] A. Messiah, "Quantum mechanics," (North-Holland, Amsterdam, 1966), Chap. 8.

[15] R. Serber, Phys. Rev. **45**, 461 (1934); J. Chem. Phys. **2**, 697 (1934)

[16] P. A. M. Dirac, "The principles of quantum mechanics," Fourth Edition (Oxford, London 1958), Sec. 58.

[17] J. H. Van Vleck, Phys. Rev. **45**, 405 (1934)

[18] T. Yamanouchi, Proc. Math.-Phys. Soc. Jpn. **19**, 436 (1937)

[19] M. Kotani, A. Amemiya, E. Ishiguro, and T. Kimura, "Tables of molecular integrals," (Maruzen, Tokyo, 1963)

[20] C. A. Cou1son and I. Fisher, Phil. Mag. **40**, 386 (1949)

[21] W. A. Goddard, Phys. Rev. **151**, 81 (1967)

[22] N. C. Pyper and J. Gerratt, Proc. Roy. Soc. (London) **A355**, 407 (1977)

[23] J. Gerratt, D. L. Cooper, and M. Raimondi, in "Valence bond theory and chemical structure," Ed. by D. J. Klein and N. Trinajstić (Elsevier, Amsterdam, 1990) p. 287.

[24] D. E. Rutherford, "Substitutional Analysis" (Edinburgh University Press, reprinted by Hafner, New York, 1968)

[25] J. H. Van Lenthe and G. G. Balint-Kurti, Chem. Phys. Lett. **16**, 138 (1980); J. Chem. Phys. **18**, 5699 (1983)

[26] P. C. Hiberty, S. Humbel, C. P. Byrman, and J. H. van Lenthe, J. Chem. Phys. **101**, 5969 (1994)

[27] F. A. Matsen, Ad. Quantum Chem. **1**, 60 (1964); J. Phys. Chem. **68**, 3238 (1964); F. A. Matsen, A. A. Cantu, and R. D. Poshusta, J. Phys. Chem. **10**, 1558 (1966)

[28] G. G. Balint-Kurti and M. Karplus, J. Chem. Phys. **50**, 478 (1969)

[29] W. Moffit, Proc. Roy. Soc. (London) **A218**, 486 (1953)

[30] G. A. Gallup, Intern. J. Quantum Chem. **6**, 899 (1972)

[31] See *e.g.*, R. Pauncz, "Spin eigenfunctions" (Plenum Press, New York, 1979)

[32] F. A. Matsen and J. C. Browne, J. Phys. Chem. **66**, 2332 (1962)

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

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

[35] J. H. Van Vleck, Phys. Rev. **49**, 232 (1936)

[36] G. A. Gallup, R. L. Vance, J. R. Collins, and J. M. Norbeck, Advances in Quantum Chem. **16**, 229 (1982)

[37] D. E. Littlewood, "The theory of group characters," (Oxford university Press, London, 1950), Second Ed., Section 6.2

[38] G. Wannier, Phys. Rev. **52**, 191 (1937)

[39] J. a. Slater, J. Chem. Phys. 19, 220 (1951); see, also, J. a. Slater, "Quantum Theory of Molecules and Solids,"(McGraw-Hill, New York, 1963)

[40] P.-O. Lowdin, J. Chem. Phys. **18**, 365 (1950)

[41] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969)

[42] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **52**, 5001 (1970)

[43] S. Huzinaga, "Gaussian Basis Sets for Molecular Calculations" (Elsevier Science Publishing Co., New York, 1984)

[44] H. H. Voge, J. Chem. Phys. **4**, 581 (1936)