1

Chemical Concepts from Molecular Orbital Theory

Feng Long Gu[1], Jincheng Yu[2], and Weitao Yang[2]

[1] *South China Normal University, School of Environment, Ministry of Education, Key Laboratory of Theoretical Chemistry of Environment, Guangzhou 510006, People's Republic of China*
[2] *Duke University, Department of Chemistry, Box 90346, Durham, NC 27708-0346, USA*

COPYRIGHTED MATERIAL

1.1 Introduction

Quantum chemistry provides the theoretical foundations and quantitative explanations of the physical and chemical properties of atoms and molecules in terms of the physical interactions between electrons and nuclei. It is based on physics, combined with various mathematical treatments, and applies the basic principles and methods of quantum mechanics to study chemical problems [1]. Its research scope includes the microscopic study of the electronic structure properties of atoms, molecules, and bulk systems, intermolecular forces, chemical bond theory, and various spectra and chemical reactions.

The history of quantum chemistry can be traced back to 1927 just after the establishment of quantum mechanics. Till the end of the 1950s, three chemical bond theories, i.e. the valence orbital theory (VOT), the molecular orbital theory (MOT), and the coordination field theory (CFT), have been established to study molecular or crystalline systems by using quantum chemistry.

Among these three chemical bond theories, the VOT was developed by Pauling et al. [2–4] on the basis of Heitler and London's work [5] for the molecular structure of hydrogen. The result is much close to the classical atomic valence theory and generally accepted by chemists. The MOT, however, was first proposed by Mulliken and Hund [6–9] in the late 1920s to the early 1930s.

The main idea of Mulliken's work on MOT is that all electrons of atoms contribute to forming molecules, and the electrons in molecules are no longer belonging to a certain atom, but moving across the entire range of a molecular space. The state of motion of electrons in space in molecules can be described by the corresponding molecular orbitals (MOs), i.e. wave function Ψ. The main difference between MOs and atomic orbitals (AOs) is that in molecules, electrons move under the action of all nuclei potential fields. An important consequence is that the MOs can be obtained by the linear combination of atomic orbitals (LCAO) in molecules.

Exploring Chemical Concepts Through Theory and Computation, First Edition. Edited by Shubin Liu.
© 2024 WILEY-VCH GmbH. Published 2024 by WILEY-VCH GmbH.

Following Mulliken's ideas, the simplest MOT was proposed by Hückel in 1931 [10], so-called Hückel molecular orbital (HMO) method, to successfully treat conjugated molecular systems. The MOT calculation is relatively simple and now widely appreciated by chemists, and it is supported by photoelectron spectroscopy experiments, making it dominant in chemical bond theory.

From the 1960s, the main goal was to further develop the quantum chemical calculation methods, among which the *ab initio* calculation method, the semiempirical methods, and other methods have expanded the application scope of quantum chemistry and the calculation accuracy has been gradually improved. Consequently, some accurate results calculated from quantum chemistry were almost exactly the same as the experimental values. The development of computational quantum chemistry has expanded quantitative computing to large molecules and the applications of quantum chemistry into other disciplines become possible.

With the development of MOT and the upgrading of computer facilities, the systems that quantum chemistry can deal with have become larger and larger, and the calculation accuracy has been continuously improved. Quantum chemical computing programs have also become an increasingly important tool for solving chemical problems, and it is expected that more complex chemical problems can be solved in the future. At present, there are many popular program suites, such as the Gaussian series [11], GAMESS [12, 13], and others.

In the beginning of MOT, there seemed to be no direct relation between MOs and the bonds in a chemical formula, because MOs obtained from MOT normally extend over the whole molecule space and are not restricted to the region between two atoms. The difficulty was overcome by using equivalent localized molecular orbitals (LMOs) instead of the delocalized ones. The mathematical definition of equivalent MOs was given only in 1929 by Lennard-Jones [14]. The concept of localization of MOs leads to the connection between MOs and the pictures of chemical bonds. Benefits from LMOs are at least four followings (i) related to the concepts of chemical bonds, useful to isolate functional groups from different molecules; (ii) reducing the efforts for computation; (iii) transferable from one molecule to others within analogical structures; (iv) more suitable for LMOs to treat correlation.

This chapter is organized as follows, the MOT is surveyed together with the localization methods either for OLMOs and NOLMOs.

1.2 Molecular Orbital Theory

In Niels Bohr's atomic model, which is based on principles of quantum physics, electrons circle the atomic nucleus in different shells containing a fixed number of electrons. The assumption was that attractive forces between the atoms in a molecule are the result of atoms sharing electrons to fill the electron shells.

Heitler and London [5] first adopted quantum mechanics to treat hydrogen molecule in 1927, revealing the nature of the chemical bond between two hydrogen atoms, leading the typical Lewis theory to today's modern VOT. The concept of the atomic bonding created by electron sharing was introduced by Lewis in his

1916s fundamental paper [15]. It was elaborated by Langmuir [16] a few years later. Pauling et al. [2–4] introduced the concept of hybrid orbitals to greatly develop VOT and successfully applied it to the structure of diatomic molecules and polyatomic molecules.

VOT coincides with the classical concept of electron bonding familiar to chemists and has been rapidly developed as soon as it appeared. However, the calculation of VOT is more complicated, which makes the later development slow. With the increasing improvement of computing technology, there will be new developments in this theory.

VOT focuses on the contribution of unpaired electrons in the outermost orbital between bonded atoms in the formation of chemical bonds, which can successfully explain the spatial configuration of covalent molecules. However, the inner electrons of the bonding atom were not considered during the actual situation of bonding.

Meanwhile from the mid-1920s, quantum mechanics has been applied to develop sophisticated models for the movement of electrons within a molecule, so-called molecular orbitals (MOs). Under the work of Hund [6], Mulliken [9], and John Lennard-Jones [14], MOT began to arise. Thus, in the beginning, the MOT was called the Hund–Mulliken theory. The concept of the word "orbital" was first proposed by Mulliken in 1932 [9]. The first paper using MOT was published by Lennard-Jones in 1929 [14] to treat MOT in a quantitative way. The LCAO approximation was introduced for constructing MOs to study the electronic structure of oxygen molecule from quantum principles. This convinced chemists that quantum mechanics is so useful, and the success of the MOT today owes much to their great contributions.

MOT is an effective approximation method for dealing with the structure of diatomic molecules and polyatomic molecules and is an important part of chemical bond theory. It differs from VOT, which focuses on understanding chemistry by hybridizing AOs into bonds, while the former focuses on the cognition of MOs. The idea of MOT is that electrons in a molecule move around the entire molecule. MOT pays attention to the integrity of molecules, so it better illustrates the structure of polyatomic molecules. At present, MOT stands on an important position in modern covalent bond theory and is widely accepted and considered a valid and useful theory.

By the 1950s, MOs were thoroughly defined as eigenfunctions of the self-consistent field Hamiltonian operator, marking the development of MOT into a rigorous scientific theory. Hartree–Fock (HF) method is a more rigorous treatment of MOT, and MOs are expanded according to a set of basis of AOs to develop the Hartree–Fock Roothaan (HFR) equation.

$$\mathbf{\Psi_i} = \sum_{\mu} \mathbf{C}_{\mu i} \chi_{\mu} \tag{1.1}$$

Equation (1.1) is so called linear combination of atomic orbitals (LCAO), where Ψ was used in the 1930s by Hund, Mulliken [9], Hückel [10], and others to construct MOs for polyatomic molecules, also called the LCAO-MO theory.

$$\mathbf{FC} = \mathbf{SC}\epsilon \tag{1.2}$$

The Hartree–Fock Roothaan equation (Eq. (1.2)) is a method of *ab initio* calculation, and the *ab initio* method is simply to use a "correct" Hamiltonian operator, except for the most basic constants, no longer citing any experimental data, based on the Schrödinger equation, only using single-electron, nonrelativistic, and Born–Oppenheimer approximations.

On this basis, a variety of *ab initio* quantum chemical calculation methods have been developed. At the same time, MOT has also been applied to a semi-empirical calculation that uses more approximate methods, known as semi-empirical quantum chemical calculations.

MOT, based on HF approach, is a theory of chemical bonds based on single-electron approximations. The basic idea of single electron approximation is that there is a physical existence of the own behavior of a single electron, which is only constrained by the action of the nucleus and other electron average fields in the molecule, provided the Pauli exclusion principle is obeyed. The wave function that describes the behavior of a single electron is called an orbital (or orbital function), and the corresponding energy of a single electron is called an energy level. For any molecule, if its series of MOs and energy levels are found, the molecular structure can be discussed in the same way as atomic structure, and linked to a systematic interpretation of molecular properties.

MOT is widely used in modern quantum chemistry, so the HFR equation is also known as the cornerstone of modern quantum chemistry. The basic idea of the HFR equation is that the wave function of a multi-electron system is a Slater determinant constructed based on the MOs of the system. Then without changing the operator and wave function form in the equation, only changing the MO coefficients of the AOs, the system energy can reach the lowest point, this minimum energy is the approximation of the total energy of the system, and the multi-electron system wave function obtained at this point is the approximation of the system wave function.

The wave function obtained from HFR equation (1.2) is so-called the canonical molecular orbitals (CMOs), in which the Lagrangian matrix ϵ is diagonal. Any unitary transformation of CMOs does not change the properties of the system, so the total energy of the system is only related to the occupied orbitals. And some forms of wave functions obtained by linear combinations of CMOs, such as those of LMOs and hybrid orbitals, are also the solutions to the HFR equation (1.2).

The role of MOs changed dramatically with the development of Kohn–Sham density functional theory (DFT) [17–19]. In DFT, the MOs are used to represent the total electron density in principle exactly. The effects of many-electron interaction are described with the exchange-correlation density functional approximations and enter into the Kohn–Sham (KS) equations, which have a form similar to the HFR equation, but with different effective potentials. Thus the MO takes on a central role in DFT, which is now the most widely used computational approache in quantum chemistry.

There are also many other methods such as quantum chemistry composite methods, quantum Monte Carlo (QMC), configuration interaction (CI), multi-configuration self-consistent field method (MCSCF), many-body perturbation theory (MBPT), and coupled cluster (CC) theory, which are developed based on

MOT. MOT can better reflect the objective reality for dealing with polyatomic systems, explaining delocalization effects and induction effects, and can solve problems that cannot be so easily dealt by VOT.

1.3 Canonical Molecular Orbitals

CMOs obtained from the HFR equation, or the KS equations in DFT extend to the entire molecular system, that is, delocalized molecular orbitals. As an example, one of the CMOs of methane (CH_4) calculated using the SCF-LCAO-MO method can be written as:

$$\psi^{\mathrm{CMO}} = -0.22\chi_{1\mathrm{s}}^{\mathrm{C}} + 0.63\chi_{2\mathrm{s}}^{\mathrm{C}} + 0.18 \times (\chi_{1\mathrm{s}}^{\mathrm{H}_1} + \chi_{1\mathrm{s}}^{\mathrm{H}_2} + \chi_{1\mathrm{s}}^{\mathrm{H}_3} + \chi_{1\mathrm{s}}^{\mathrm{H}_4}) \tag{1.3}$$

The CMOs normally contain the AO components of all atoms, and the electron belongs to the entire molecule, no longer limited to an atom or between two atoms, which is also called the delocalized molecular orbitals.

The CMOs cannot adopt the localized properties of electronic structure. Even though the traditional *ab initio* quantum chemical calculation method has made great achievements for small molecule systems. However, as CMOs require the calculation of the entire system, for large systems, the calculation process becomes so complicated, which makes the application of quantum chemical calculation methods based on regular CMOs to macromolecular systems a major problem. The cost of the traditional HF method increases with N^4 of the system, where N is the number of the basis functions.

Quantum chemistry calculations for large systems are normally limited by using the traditional CMO basis. Therefore, various approximation methods are developed. Kirtman [20–22] proposed the local space approximation (LSA). At the semiempirical level, LocalSCF has been proposed by Anikin et al. [23]. Yang [24] developed the Divide and Conquer method based on the approximation of DFT. Imamura et al. [25–27] developed the elongation method. The fragment molecular orbital (FMO) method has been proposed by Kitaura et al. [28] for the calculation of biological macromolecules of proteins and DNA. Li et al. constructed the fragment energy assembler (FEA) [29].

1.4 Frontier Molecular Orbital Theory

Frontier molecular orbital (FMO) theory, a kind of MOT, was developed in 1950s by Fukui et al. [30]. It has been pointed out that many properties of molecules are mainly determined by the FMOs, i.e. the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The FMO theory is simple, intuitive, and effective, so it has a wide range of applications in theoretical research such as chemical reactions and catalytic mechanisms. FMO plays a decisive role in the selection of the reaction route of organic synthesis. Later on, Woodward and Hoffmann applied FMO theory to study the stereochemical

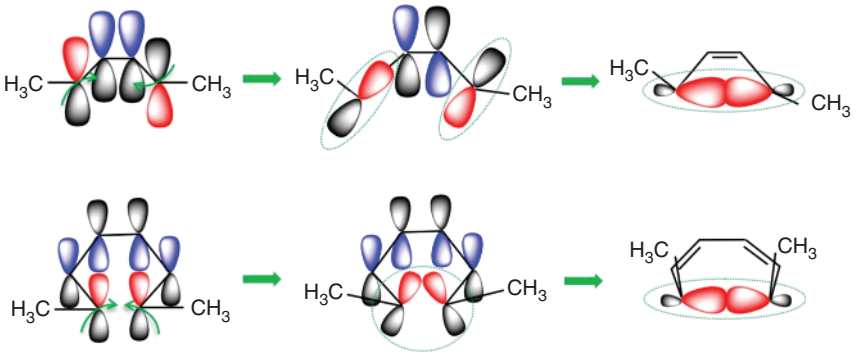

Figure 1.1 The reaction mechanisms of butadiene and hexatriene elucidated by FMO theory.

selection rule for chemical reactions, further developing it into the principle of symmetry conservation of MOs. The synthesis of vitamin B_{12} [31] is a very successful example guided by the FMO theory and the principle of conservation of MO symmetry. In 1981, Fukui and Hoffmann shared the Nobel Prize in chemistry for their intuitive explaining the occurrence of chemical reactions through the principle of conservation of MO symmetry.

For the reaction between two molecules A and B. The electrons in the HOMO of molecules A and B flow to each other's LUMO, causing the formation and breaking of chemical bonds, and a chemical reaction occurs. Electron flow is prone to occur only when the HOMO of molecule A (or B) is close to the energy of LUMO of molecule B (or A), together with the symmetry matching each other. It must be pointed out that the FMO theory applies not only to π orbits, but also to σ orbits, so it has applications in organic chemistry, inorganic chemistry, as well as surface adsorption and catalysis, quantum biology, and other fields.

As an example, in the electrocyclization reaction, conrotatory is a heteroplanar process, and disrotatory is a homoplanar process. Under the heating condition, the reaction mechanisms of butadiene and hexatriene are different. Figure 1.1 depicts how HOMOs rotate and form a σ orbital. One can see that due to the phases of the two orbital lobes at the ends are different, for butadiene, conrotatory of two π orbitals is consistent with the MO symmetry. While conversely, disrotatory of two π orbitals of hexatriene is needed for the electrocyclization reaction. This example clearly shows that the FMO theory is consistent with the experimental observations.

1.5 Localized Molecular Orbitals

In view of the computational complexity and high costs in calculating macromolecular systems based on CMOs, computational methods are highly demanded for costs increased linearly with the size of the system (linear scale). Most of the quantum chemical linear scale calculation methods developed in recent years to

deal with macromolecular systems are based on LMOs, which is another form of MOs obtained by a unitary transformation of CMOs. The most important character of LMOs is the locality, that is, the properties of a spatial region are less affected by the distant spatial region. Of course, the differences in the description of LMOs and CMOs are only superficial, and they describe the same objective situation. One can think of any LMO as a "hybridization" of CMOs, just as the hybridization of AOs can form valence bond orbitals. Equation (1.4) presents the MO coefficients of one of the LMOs of CH_4 obtained by a unitary transformation.

$$\begin{aligned}\psi^{\mathrm{LMO}} = {} & 0.31\chi_{2\mathrm{s}}^{\mathrm{C}} + 0.29\chi_{2\mathrm{p}_x}^{\mathrm{C}} + 0.40\chi_{2\mathrm{p}_y}^{\mathrm{C}} + 0.54\chi_{1\mathrm{s}}^{\mathrm{H}_1} \\ & + 0.06 \times (\chi_{1\mathrm{s}}^{\mathrm{H}_2} + \chi_{1\mathrm{s}}^{\mathrm{H}_3} + \chi_{1\mathrm{s}}^{\mathrm{H}_4})\end{aligned} \tag{1.4}$$

From the comparison between Eqs. (1.3) and (1.4), one can find that LMOs are more precisely related to the AOs of directly bonded atoms, and the influence of indirectly bonded atoms on the specified molecular orbitals is reduced. After localization, the LMO, in addition to the 2s, $2p_x$, and $2p_y$ contributions of C atom, the 1s orbital of one hydrogen atom has the largest contribution, and the others are quite small, so that it is as if the electron motion is mainly limited between these two atoms. This confirms that the four regular CMOs of methane can be transformed by orthogonal transformations into four LMOs in the C–H bond region, corresponding to the four C–H bonds of valence bond theory.

Since the physical properties do not change under the unitary transform of MOs, the LMOs and the CMOs linked by the unitary transform are completely equivalent for the description of the molecular properties determined by the electron as a whole. If it comes to properties related to the spatial structure of molecules, the description of LMOs is more intuitive. LMOs have the advantage of being more consistent with traditional chemical concepts, and chemists are happy to use them.

From the point of view of computational quantum chemistry, one of the most fascinating aspects of the description of LMOs is its transferability. The certain molecular properties (such as bond energy and bond distance) that chemists have long determined from experience are equivalent to the "transferability" of LMOs in the single-particle approximation. If this transferability exists, LMOs obtained from small molecules are transferable to large molecules, which is meaningful to reduce the amount of computation for large molecules.

1.5.1 Orthogonal Localized Molecular Orbitals

There are two ways to generate LMOs, one is to perform a unitary transformation from CMOs to LMOs, and the other is to directly solve a certain single-electron Schrödinger equation to get LMOs. However, both methods require a localization criterion to determine the transformation matrix for the transition from a regular orbital to a localized orbital or a single-electron Schrödinger equation using the second method. Many localization methods have been proposed, and some of them artificially specify the shape and position of LMOs to be generated according to the traditional chemical concept and the symmetry nature of the molecule.

There are some popular localization methods, such as the Boys localization, the Edmiston–Ruedenberg localization, and the Pipek–Mezey localization, which are also known as orthogonal localized molecular orbital (OLMO) methods.

Any localization method requires two elements: a physically meaningful localization criterion and a computationally effective mathematical algorithm for implementing the satisfaction of this criterion. Both aspects are important for a viable localization procedure.

(1) Boys localized molecular orbitals

Foster and Boys [32, 33] propose a scheme for localized orbitals, as so called "exclusive orbitals." They are obtained by maximizing the product of the spatial distances between the centers of charge vectors R_i of all different molecular orbitals. The required consecutive iterations present convergence problems presumably because, even for moderately sized molecules, the product is of very high order in the orbitals. Inspired by Edmiston–Ruedenberg's work (see below), the Boys localization scheme now is based on minimizing the sum of the orbital's second central moment

$$\mathcal{L}_{\text{Boys}} = \sum_{p}^{N} \langle p|(\hat{r} - \langle p|\hat{r}|p\rangle)^2|p\rangle \tag{1.5}$$

where $\langle p|p\rangle = 1$ means that the AOs are normalized. This localization function is known as the Boys–Foster or Boys localization function.

The set of orbitals obtained from minimizing Eq. (1.5) is a set in which the orbitals on average are local. The Boys localization function has been widely used in chemistry and is also used in solid-state theory for the localization of Wannier functions, leading to the maximally localized Wannier functions [34].

(2) Edmiston–Ruedenberg localized molecular orbitals

Edmiston and Ruedenberg [35] introduced a localization scheme that maximizes the sum of orbital self-repulsion energies

$$\mathcal{L}_{\text{ER}} = \sum_{p}^{N} \langle pp|\frac{1}{r_{12}}|pp\rangle \tag{1.6}$$

where $\langle pp|\frac{1}{r_{12}}|pp\rangle = \iint \phi_p(r_1)\phi_p(r_2)\frac{1}{r_{12}}\phi_p(r_1)\phi_p(r_2)\mathrm{d}r_1\mathrm{d}r_2$

This method has been used in different theoretical investigations, as well as in a more fundamental analysis of the origin of molecular bonding. The Edmiston and Ruedenberg method has a fifth-order computational scaling compared to a third-order scaling for the Boys and Pipek–Mezey schemes, and has, despite a reduction to asymptotically third-order scaling, been used less extensively.

(3) Pipek–Mezey localized molecular orbitals

Pipek and Mezey [36] introduced a localization function that measures the number of atomic centers over which a molecular orbital extends. The Pipek–Mezey

localization function is a sum of the squared Mulliken charges for a set of N orthonormal occupied or virtual HF orbitals and is given by

$$\mathcal{L}_{\mathrm{PM}} = \sum_{p}^{N} \sum_{A} \sum_{\mu \in A} \left[\langle p | \hat{P_{\mu}} | p \rangle \right]^2 \tag{1.7}$$

The charge distribution operator $\hat{P_{\mu}}$ is defined as

$$\hat{P_{\mu}} = \frac{1}{2} (|^b \mu\rangle \langle \mu| + |\mu\rangle \langle ^b \mu|) \tag{1.8}$$

where $|^b \mu\rangle = \sum_{\nu} |\nu\rangle s^{-1}_{\nu\mu}$, $s^{-1}_{\nu\mu}$ is the inversion of the AO basis overlap. Therefore, LMOs can be obtained by taking the maximum of the functional $\mathcal{L}_{\mathrm{PM}}$.

The Mulliken population analysis suffers from some unphysical behavior since the individual Mulliken charges for a shared electron between two atoms may have numbers greater than 1 or less than 0 [37]. This unphysical behavior is due to the fact that in a Mulliken population analysis, overlap populations occur since the AO basis is not orthogonal, and the overlap population is divided equally between the atomic centers, ignoring that different types of atoms have different electronegativity. The unphysical behavior increases when the AO basis set increases in size.

The LMOs obtained by the Boys, Edmiston and Ruedenberg (ER), and Pipek and Mezey (PM) localization schemes are the most popular orbitals used by computational chemists to compute OLMOs. Judged by their physical nature, the PM and ER orbitals are more physical than the Boys orbitals. Judged by computation speed, however, the PM and Boys schemes are much faster than that of the ER's. For large systems that can be tackled with today's computer power, the speed of orbital calculation is essential; computing LMOs should not take longer than the subsequent electron correlation calculations.

As determinantal wave functions are invariant with respect to orthogonal transformations among the orbitals, in order to cast them into localized form, intrinsic or external criteria can be used. The methods of ER and of Boys are intrinsic because only the actual molecular orbitals are used. While the method of PM is external because it is based on the overlap between the actual molecular orbitals and certain independently chosen additional linear combinations of atomic basis orbitals.

Besides the above mentioned localization methods, Löwdin [38] proposed the concept of natural orbits. A set of natural orbitals is combined into a single-electron basis function to constitute the electronic configuration of the N-particle system, so that the configuration of Ψ expansion can be achieved with fewer basis than the regular Hartree–Fock orbital basis. The work of Reed and Weinhold [39] expanded on this basis, and systematically proposed the concepts of natural spin orbital, natural bond orbital (NBO), and natural hybrid orbital, and developed into the NBO theory. Through the type of orbital and NBO analysis, one can easily find out the atomic population in a molecule, bond

information, as well as intramolecular and intermolecular super-conjugate interactions.

(4) Regional localized molecular orbitals

The elongation method, proposed by Imamura et al. [25], is to theoretically synthesize a polymer chain by adding a monomer unit stepwise to a starting oligomer while keeping the degree of freedom of active space almost fixed. The elongation method works in an OLMO basis, in contrast to the conventional HF method in a CMO basis. The advantage of an LMO representation is that it allows one to freeze the region far away from the chain propagation site. This reduces the number of variational degrees of freedom in the system. Just a simple 2×2 unitary rotation was originally adopted to obtain LMOs for the elongation processing. In this 2×2 unitary rotation, pairs of CMOs are selected with an initial division of CMOs either belonging to the frozen or to the active region, and then successively rotated to form one LMO in the frozen region and another LMO in the active region. This bears that a poor selection of CMOs yields a poor localization. The slow convergence of the 2×2 localization affects the applications of the elongation method.

A different localization scheme for the elongation method has been developed based on regional localized molecular orbitals (RLMOs) [40]. This scheme is more efficient and more accurate even for covalently bonded systems with strongly delocalized π electrons. *Ab initio* test calculations have been performed even for very delocalized systems, and it was confirmed that this new scheme has big progress for the elongation method. The localization scheme is described as follows:

The AO-based density matrix is given

$$\mathbf{D}^{\mathbf{AO}} = \mathbf{C}_{\mathbf{AO}}^{\mathbf{CMO}} \mathbf{d} \mathbf{C}_{\mathbf{AO}}^{\mathbf{CMO}\dagger} \tag{1.9}$$

Then Löwdin's symmetric orthogonalization procedure is performed to transfer the density matrix to the orthogonal atomic orbital (OAO) basis. The transformation matrix X is obtained by diagonalizing $\mathbf{S}^{\mathbf{AO}}$ to give

$$\mathbf{D}^{\mathbf{OAO}} = \mathbf{X} \mathbf{D}^{\mathbf{AO}} \mathbf{X}^{\dagger} \tag{1.10}$$

After one gets OAO basis D-matrix, one can partition the D-matrix into two parts, one for the frozen region (A region) and one for the active region (B region). Region B is defined by atoms adjacent to the growing end of the chain whereas region A is at the opposite end. The purpose is to find two sets of regional LMOs (RLMOs), respectively, for both regions. The desired RLMOs can be obtained by two steps. It is something similar to the construction of natural bond orbitals (NBOs) [39] but to localized regional orbitals rather than to localized bond orbitals. A regional orbital (RO) space is generated by separately diagonalizing the subblocks of $\mathbf{D}^{\mathbf{OAO}}$, i.e. $\mathbf{D}^{\mathbf{OAO}}(\mathbf{A})$ and $\mathbf{D}^{\mathbf{OAO}}(\mathbf{B})$, and the transformation from OAOs to ROs is given by the direct sum

$$\mathbf{T} = \mathbf{T}^{\mathbf{A}} \oplus \mathbf{T}^{\mathbf{B}} \tag{1.11}$$

where $\mathbf{T^A}$ and $\mathbf{T^B}$ are the eigenvectors of $\mathbf{D^{OAO}}(\mathbf{A})$ and $\mathbf{D^{OAO}}(\mathbf{B})$, respectively. Then the RO-based density matrix is obtained by the following transformation,

$$\mathbf{D^{RO}} = \mathbf{T}^{\dagger}\mathbf{D^{OAO}}\mathbf{T} \tag{1.12}$$

and the transformation coefficients from ROs to CMOs may be written as

$$\mathbf{C}_{\mathbf{RO}}^{\mathbf{CMO}} = \mathbf{T}^{\dagger}\mathbf{X}\mathbf{C}_{\mathbf{AO}}^{\mathbf{CMO}} \tag{1.13}$$

At this point, one gets the non-orthogonal LMOs completely localized to either region A or region B. However, they are not completely occupied or unoccupied and of course not suitable for the elongation method. Thus, it is necessary to carry out a unitary transformation between the occupied and unoccupied blocks of $\mathbf{D^{RO}}$ in such a way as to preserve the localization as much as possible. This is done by using the Jacobi procedure. U is the transformation that diagonalizes $\mathbf{D^{RO}}$, the unitary transformation from CMO to RLMO is given as

$$\mathbf{C}_{\mathbf{RLMO}}^{\mathbf{CMO}} = \mathbf{U}^{\dagger}\mathbf{T}^{\dagger}\mathbf{X}\mathbf{C}_{\mathbf{AO}}^{\mathbf{CMO}} \tag{1.14}$$

Finally, the original AOs basis RLMOs is given by

$$\mathbf{C}_{\mathbf{AO}}^{\mathbf{RLMO}} = \mathbf{X}^{-\mathbf{1}}\mathbf{T}\mathbf{U} \tag{1.15}$$

As long as sufficiently well-localized RLMOs are obtained by the scheme described above, the elongation method can proceed by using these RLMOs. It has been tested that the accuracy and the efficiency of the elongation method weaponed with this new localization scheme are greatly improved. By using RLMOs, the elongation method can be successfully applied to one-dimensional chains and any two- or three-dimensional systems no matter what type of chemical bonding is involved and regardless of whether the system is periodic or aperiodic.

1.6 Regularized Nonorthogonal Localized Molecular Orbitals

The OLMOs do reach a certain degree of locality, but due to the constraints of orthogonal normality conditions, there are long-range non-localized tails in OLMOs. This affects the accuracy and efficiency of the calculation. In order to obtain images that are more consistent with traditional chemical bonds, there was an early desire to construct non-orthogonal localized molecular orbitals (NOLMOs). NOLMOs were first applied to the calculation of electronic structure by Adams [41] and Gilbert and Lykos [42] in 1961. The nature of intramolecular interactions was investigated based on the transferability of NOLMOs and applied to mono- and difluoromethane [43]. In 2003, Sorakubo et al. proposed a partially LMO KS-DFT method based on NOLMOs [44]. Paulus et al. applied NOLMOs to the *ab initio* incremental correlation treatment at CCSD level, and their results show that the transferability of NOLMOs is much improved compared to those of OLMOs [45]. Sironi et al. applied it to valence bond theory [46]. However, these *ab*

initio methods for determining NOLMOs have only been tested in small systems. In 2004, Anikin et al. [23] applied NOLMOs to semiempirical methods and completed a semiempirical program LocalSCF that can calculate the linear scale of tens of thousands of atomic systems. In 2013, Peng et al. confirmed the feasibility of the method at the Hartree–Fock level from the source, and also successfully realized the programmatic of the method [47]. Mayer et al. used extremely localized molecular orbitals (ELMO) for the transferability [48, 49].

Yang et al. developed the principle of absolute energy minimum variational to avoid the time-consuming process of diagonalizing the Fock matrix and used the conjugate gradient method to obtain the minimum value of the system energy [50, 51]. It is found that NOLMOs can obtain higher precision results than any traditional OLMOs, and the locality of NOLMOs is about 10-28% better than that of OLMOs. In 2013, Peng et al. developed an HF method based on NOLMOs [47], and the test results show that for six different systems, including $C_{18}H_{38}$ molecular system and polyglycine system, the convergence is almost comparable to the traditional method.

Relatively speaking, NOLMOs have better locality than OLMOs, because the orbitals are no longer constrained by orthogonality between NOLMOs, which also solves the problem of long-range tailing in OLMOs. However, since the NOLMO releases the constraints of orbital orthogonality, unlike the methods based on OLMOs, almost all the relevant formulas need to be re-derived.

Figure 1.2 is the comparison of the distribution in AO basis for CMOs and NOLMOs of $C_{60}H_{122}$ molecule [52]. It can be seen that NOLMOs, compared to CMOs, are much more localized in space. Due to the sparse MO coefficient matrix of NOLMOs, one can design and realize low-scaling or linear-scaling calculations of electronic structure for large systems.

The NOLMOs of a system are obtained by minimizing the following quantity

$$\Theta[\varphi_k] = \langle\varphi_k|\left(\mathbf{r} - \mathbf{r}_k^0\right)^2|\varphi_k\rangle + \omega\langle\varphi_k|\hat{T}|\varphi_k\rangle \tag{1.16}$$

where φ_k are NOLMOs of the system, $\mathbf{r}_k^0 = (\mathbf{x}_k^0, \mathbf{y}_k^0, \mathbf{z}_k^0)$ is the fixed centroid of φ_k, $\hat{T}$ is the kinetic energy operator, ω, a positive constant, is the weight of kinetic

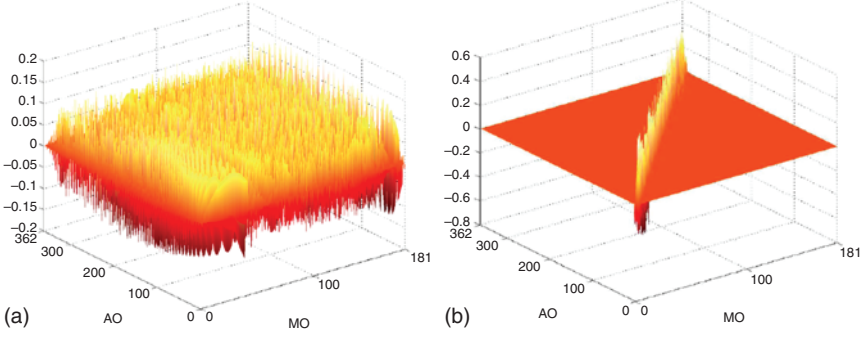

Figure 1.2 Comparison of the distribution in AO basis for CMOs (a) and NOLMOs (b) of $C_{60}H_{122}$ molecule. Source: Reproduced from Ref. [52] with permission from the Royal Society of Chemistry.

energy term. When $\omega = 0$, Eq. (1.16) is just the widely used Boys localization cost function (1.5). The added term of the kinetic energy is to regularize the solution NOLMO, or to enhance its smoothness, by reducing the kinetic energy of the orbitals. Since φ_k are occupied NOLMOs, they must be the linear combination of occupied CMOs of the system, namely

$$\varphi_k = \sum_i^{N/2} A_{ik}\psi_i \tag{1.17}$$

where N is the number of electrons in the system, ψ are CMOs obtained from conventional ways. $\Theta[\varphi_k]$ can then be rewritten as the function of coefficients A_{ik}

$$\begin{aligned}\Theta[A] &= \sum_{i,j} A_{ik}^* A_{jk} \left[\langle\psi_i|\left(\mathbf{r} - \mathbf{r}_k^0\right)^2|\psi_j\rangle + \omega\langle\psi_i|\hat{T}|\psi_j\rangle\right] \\ &= \sum_{i,j} A_{ik}A_{jk}\left[\tilde{R}_{ij} - 2x_k^0 X_{ij} - 2y_k^0 Y_{ij} - 2z_k^0 Z_{ij} + \tilde{r}_k^0 + \omega T_{ij}\right]\end{aligned} \tag{1.18}$$

where $\tilde{R}_{ij}$ is the integration over the square of Cartesian vector $\langle\psi_i|\hat{x}^2 + \hat{y}^2 + \hat{z}^2|\psi_j\rangle$, X_{ij}, Y_{ij}, and Z_{ij} are integrals of individual Cartesian vector components $\hat{x}$, $\hat{y}$ and $\hat{z}$, respectively $\tilde{r}_k^0 = x_k^0 x_k^0 + y_k^0 y_k^0 + z_k^0 z_k^0$ is a value which equals to the square of the Cartesian vector $\mathbf{r}_k^0$.

In order to obtain the optimal value of the transformation matrix A, we need the first derivative of $\Theta[A]$ with respect to A_{ml}

$$\frac{\partial\Theta}{\partial A_{ml}} = 2\sum_j (Q_{mj}^l + \omega T_{mj})A_{jl} \tag{1.19}$$

where $Q_{mj}^l = \tilde{R}_{mj} - 2(x_l^0 X_{mj} + y_l^0 Y_{mj} + z_l^0 Z_{mj}) + \tilde{r}_l^0$. We can see that different NOLMOs are not coupled. The NOLMOs can be optimized individually. For a given centroid l, finding the best NOLMOs of centroid l is equivalent to solving the following eigenvalue equation in the system

$$\sum_j \Theta_{mj}^l A_{jl} = \theta_p^l A_{ml}^p \tag{1.20}$$

where $\Theta_{mj}^l = Q_{mj}^l + \omega T_{mj}$. The eigenvector with the lowest eigenvalue θ_0^l is the best localized NOLMO with centroid l

$$\varphi_l = \sum_m A_{ml}^0 \psi_m \tag{1.21}$$

Figure 1.3 depicts the real space representation of a π NOLMO distribution with different weights of the kinetic energy. From Figure 1.3 one can see that without the kinetic energy (i.e. $\omega = 0$) included in the cost function, the oscillation of NOLMO persists almost to the whole system, leading to poor localization. With the weight of the kinetic energy becomes larger, the orbital becomes less oscillatory and thus more localized.

However, as the weight of the kinetic energy increases, the orbitals become more diffuse, as shown in Figure 1.3. It is easy to understand the limit when the weight becomes infinity, the optimized orbital will be the lowest eigenvalue of the kinetic energy operator, and it will be delocalized in the whole space. Thus, the weight of

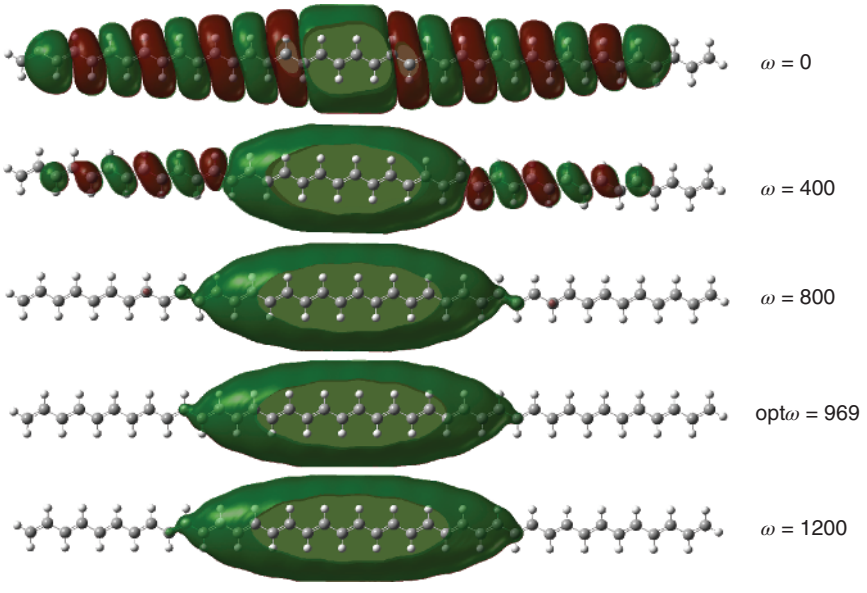

Figure 1.3 Real space representation of a π NOLMO with different weights of the kinetic energy tested on polyacetylene (H–(C=C)$_n$–H, n=20) at B3LYP/6-31G level, the buffer size is 19.0 Å, and the isosurface value is 10^{-5} e$^-$/bohr^3. ω is the weight of the kinetic energy of the cost function, while optω is the optimal value of ω. The effect of the regularization is evident – it greatly reduces the oscillation of the orbitals. Source: Reproduced with permission from Ref. [53]. Copyright 2022 American Chemical Society.

the kinetic energy term should not be too large. The weight of the kinetic energy can be determined by the minimization of the total energy of the system. It is found that with an optimized weight of the kinetic energy, one can get the most compact NOLMO.

The Divide-and-Conquer (DC) method developed by Yang [24] is to divide the entire system into fragments and then sum up the densities of the fragments for the entire system. By using the regularized NOLMOs for each fragment, NOLMO-DC is employed to obtain the electron density and the total energy. It is found that the total energy calculated by NOLMO-DC with optimal weight of the kinetic energy as regularization is more accurate than that without kinetic energy included in the cost function. Figure 1.4 depicts the accuracy in the total energy of NOLMO-DC and DC versus the buffer size with and without the kinetic energy included in the localization cost function for a conjugate system, polyacetylene (H–(C=C)$_n$–H, $n = 20$) at B3LYP/6-31G(d,p) and HF/6-31G(d,p) levels. With a small buffer size (4.0 Å) the accuracy of NOLMO-DC is higher than that of DC by one order of magnitude. The accuracy of NOLMO-DC is about 10^{-3} a.u. for a system of 40 carbon atoms. The accuracy of the original DC approaches the same level requiring a much larger buffer size (14.0 Å). With the buffer size larger than 9.0 Å the accuracy of NOLMO-DC is higher than DC by more than three orders of magnitude. While to achieve the same accuracy, the original DC will require the buffer size larger than 24.0 Å. These results

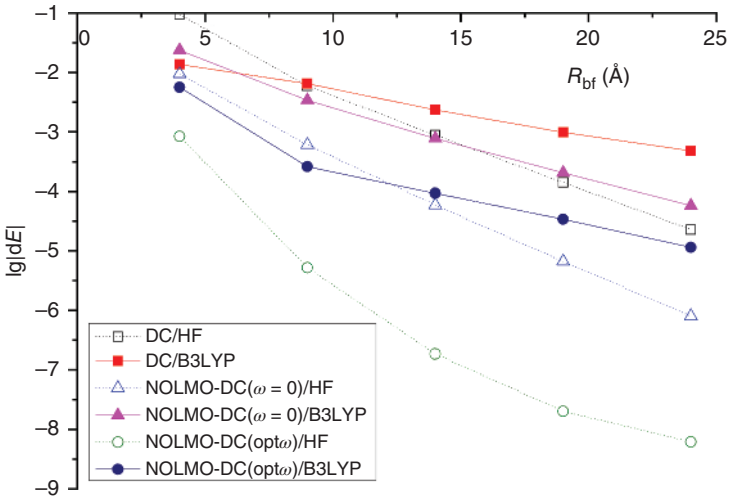

Figure 1.4 Accuracy of NOLMO-DC and DC tested on polyacetylene (H−(C=C)$_n$−H, $n = 20$) at B3LYP/6-31G(d,p) and HF/6-31G(d,p) levels for different values of ω and buffer sizes. ω is the weight of the kinetic energy of the cost function, while optω is the optimal value of ω determined by optimization with the analytical gradients. dE is the absolute difference between conventional energy and the DC or NOLMO-DC energy. Source: Reproduced with permission from Ref. [53]. Copyright 2022 American Chemical Society.

show that the buffer size for NOLMO-DC is only about one-third of that of the original DC method to achieve the similar accuracy level.

1.7 Molecular Orbitalets

FMOs have been successful in identifying chemically reactive groups in small systems by utilizing associated orbital energies [30, 54–59]. However, FMOs may not effectively highlight the locality of chemical reactivity in large systems due to the delocalization nature of CMOs. To address this limitation, the concept of frontier molecular orbitalets (FMOLs) was introduced to describe the reactivity of larger systems [60]. FMOLs are localized molecular orbitals that accurately reflect the frontier nature of chemical processes and offer a more comprehensive approach for identifying the localization of chemical reactivity.

Molecular orbitalets are a set of LMOs originally developed to capture the local fractional charges and spins of chemical systems for localized orbital scaling correction (LOSC) [61, 62] and fractional-spin LOSC [63] methods. These methods were designed to reduce and eliminate the delocalization error that leads to a series of systematic errors of DFT [64–68].

Orbitalets are constructed by linearly combining the CMOs, $\phi_p = \sum_q U_{pq}\varphi_q$, to minimize the following cost function,

$$F = (1-\gamma)F_r + \gamma CF_e \tag{1.22}$$

$$F_r = \sum_p [\langle \phi_p | \mathbf{r}^2 | \phi_p \rangle - \langle \phi_p | \mathbf{r} | \phi_p \rangle^2] \tag{1.23}$$

$$F_e = \sum_p [\langle \phi_p | \hat{h}^2 | \phi_p \rangle - \langle \phi_p | \hat{h} | \phi_p \rangle^2] \tag{1.24}$$

where $\{\phi_p\}$ are orbitalets, and $\{\varphi_p\}$ are CMOs. C is used to match magnitudes and unify units between the energy and physical spaces, and γ is the weight of the energy delocalization penalty. The cost function can be split into two parts: the physical space part F_r and the energy space part F_e. F_r, taken from Foster–Boys localization [33], shows the extent of spatial spread of the orbitalets, with higher values indicating greater spread. F_e, on the other hand, measures energy delocalization by quantifying how much the orbitalets deviate from the eigenstates of the Hamiltonian (i.e. CMOs). Lower values of F_e indicate less loss of energy information. The weights of the two parts can be adjusted accordingly by varying the value of γ. If γ is set to 1, the physical space part F_r does not contribute to the cost function, and therefore the resulting orbitalets will be the same as the CMOs. If γ is set to 0, the energy space part F_e will not contribute to the cost function, resulting in a very large deviation of the orbitalets from the eigenstates of the Hamiltonian. In this scenario, the orbitals will lose energy information and achieve maximal localization in physical space, resulting in generalized Foster–Boys localized orbitals.

We use hexadeca-1,3,5,7,9,11,13,15-octaene to illustrate the concept of orbitalets. Figure 1.5a displays a comparison between the energy structures of CMOs and LMOs obtained from Foster–Boys localization. The bar graph of CMOs clearly shows the fine energy structure. However, when γ is set to 0, the overall energy distribution is lost and there is barely any structure. Conversely, when γ is set to 0.975, although the fine structure of energy is less clear than that of CMOs, the resulting orbitalets still retain the main energy information (Figure 1.5b). Similar to the concepts of HOMO and LUMO, here we define the concepts of the highest occupied molecular orbitalet (HOMOL) and the lowest unoccupied molecular orbitalet (LUMOL). The orbitalets represent a combination of CMOs that are close in energy. As a result, a set of "frontier" orbitalets located in distinct local regions of the system can be obtained. Therefore, orbitalets can maintain the major energy information of CMOs while sacrificing some energy structure for locality information in the physical space. Moreover, when γ is close to 1, the loss of energy structure is minor, which makes FMOLs [60] a powerful tool for studying active sites, particularly in extensive chemical systems.

The use of FMOs, specifically HOMO and LUMO, has been widely employed in understanding the reactivity and regioselectivity of various chemical systems [30, 54–59]. Although FMOs have achieved great success in small systems, their delocalization nature becomes a limitation when applied to large systems. The FMOs of large systems tend to span over a significant portion of the system, leading to the loss of important locality information necessary to identify the functional groups that play a significant role in the reaction.

Explaining Diels–Alder (DA) reactions is one of the most prominent applications of FMO theory. FMOs of the diene and dienophile interact. Chemists typically focus

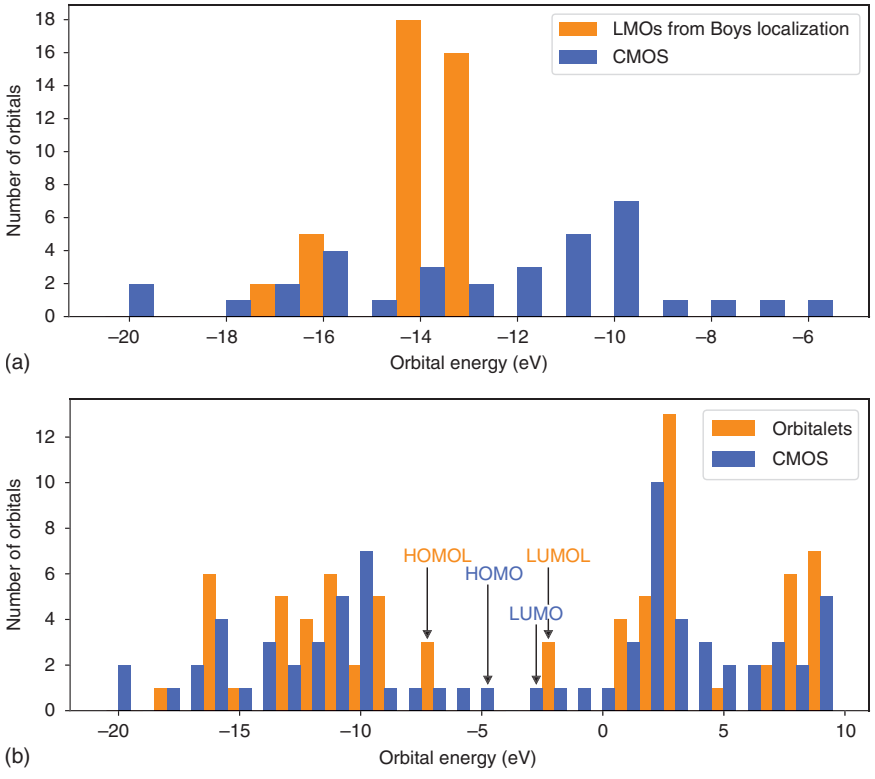

Figure 1.5 Orbital energy structures of hexadecaoctaene. (a) Energy structures shown by CMOs ($\gamma = 1$) and that shown by LMOs from Foster–Boys localization ($\gamma = 0$, only occupied orbitals are localized). (b) Energy structures shown by CMOs ($\gamma = 1$) and that shown by orbitalets ($\gamma = 0.975$). Source: Adapted fromYu et al. [60].

on "reactive functional groups" when analyzing chemical reactions. For example, in the hypothetical DA reaction between hexadeca-1,3,5,7,9,11,13,15-octaene and ethene [69], one would expect the frontier orbitals of hexadecaoctaene to be butadiene-like orbitals located at two adjacent double bonds in the middle. However, as shown in Figure 1.6a,b, the FMOs are undesirably delocalized over the entire molecule, making it challenging to use them to describe chemical reactivity. In comparison, FMOLs contain both energy and locality information, enabling HOMOLs and LUMOLs to clearly identify the most reactive site. As demonstrated in Figure 1.6c,d, the HOMOL and the LUMOL of hexadecaoctaene are mostly localized on the central adjacent double bonds and have butadiene-like shapes. This aligns with the traditional understanding of reactive functional groups. Thus, FMOLs expand the scope of FMOs beyond small systems, providing valuable energy and locality information to describe the reactivity of large systems. In addition to this, FMOLs can also generate promising results across various applications including providing fast analysis of chemical reactions based on transition states, and capturing electron transitions in excitation charge-transfer processes [60].

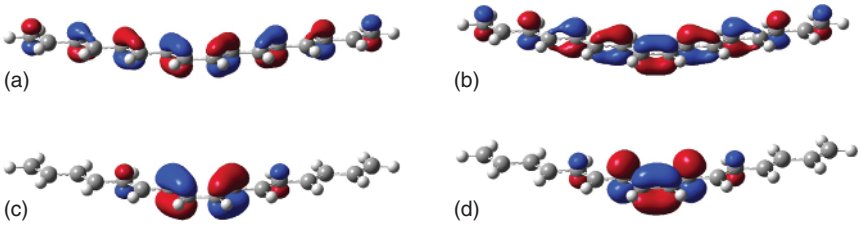

Figure 1.6 FMOs and FMOLs of hexadecaoctaene. FMOs, (a) HOMO and (b) LUMO, bury the butadiene-like model in delocalized orbitals. FMOLs, (c) HOMOL and (d) LUMOL, mainly locate on the central two double bonds, resembling the butadiene-like orbitals. Iso = 0.05, $\gamma = 0.975$.

Acknowledgment

J.Y. and W.Y. have been supported by National Science Foundation (Grant No. CHE-2154831).

References

1 Jensen, F. (2017). *Introduction to Computational Chemistry*. Wiley.
2 Pauling, L. (1932). Interatomic distances in covalent molecules and resonance between two or more Lewis electronic structures. *Proc. Natl. Acad. Sci. U.S.A.* 18: 293–297.
3 Pauling, L. (1932). The electronic structure of the normal nitrous oxide molecule. *Proc. Natl. Acad. Sci. U.S.A.* 18: 498–499.
4 Pauling, L. and Yost, D.M. (1932). The additivity of the energies of normal covalent bonds. *Proc. Natl. Acad. Sci. U.S.A.* 18: 414–416.
5 Heitler, W. and London, F. (1927). Wechselwirkung neutraler atome und homöopolare bindung nach der quantenmechanik. *Z. Angew. Phys.* 44: 455–472.
6 Hund, V.F. (1926). Zur deutung einiger erscheinungen in den molekelspektren. *Z. Angew. Phys.* 36: 657–674.
7 Hund, V.F. (1927). Zur deutung der molekelspektren. I. *Z. Phys.* 37: 742–764.
8 Hund, V.F. (1927). Zur deutung der molekelspektren. II. *Z. Phys.* 42: 93–120.
9 Mulliken, R.S. (1932). Electronic structures of polyatomic molecules and valence. *Phys. Rev.* 40: 55–71.
10 Hückel, E. (1931). Quantentheoretische beiträge zum benzolproblem, I. die elekfronenkonfigurafion des benzols und verwandfer verbindungen. *Z. Angew. Phys.* 70: 204–286.
11 Frisch, M.J., Trucks, G.W., Schlegel, H.B. et al. Wallingford, CT: Gaussian, Inc. Copyright © 1988–2017.
12 Schmidt, M.W., Baldridge, K.K., Boatz, J.A. et al. (1993). General atomic and molecular electronic structure system. *J. Comput. Chem.* 14(11): 1347–1363.
13 Gordon, M.S. and Schmidt, M.W. (2005). *Advances in Electronic Structure Theory: GAMESS a Decade Later*, 1167–1189. Elsevier.

14 Lennard-Jones, J.E. (1929). Electronic structures of polyatomic molecules and valence. *Trans. Faraday Soc.* 25: 668–686.

15 Lewis, G.N. (1916). The atom and the molecule. *J. Am. Chem. Soc.* 38: 762–785.

16 Langmuir, I. (1919). The arrangement of electrons in atoms and molecules. *J. Am. Chem. Soc.* 41: 868–934.

17 Hohenberg, P. and Kohn, W. (1964). Inhomogeneous electron gas. *Phys. Rev.* 136(3B): B864–B871.

18 Kohn, W. and Sham, L.J. (1965). Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140(4A): A1133–A1138.

19 Parr, R.G. and Yang, W. (1989). *Density-Functional Theory of Atoms and Molecules*. Oxford University Press.

20 Kirtman, B. (1982). Molecular electronic structure by combination of fragments. *J. Chem. Phys.* 86: 1059–1064.

21 Kirtman, B. and Demelo, C.P. (1986). Accurate local-space treatment of hydrogen bonding in large systems. *Int. J. Quantum Chem.* 29: 1209–1222.

22 Kirtman, B. and Demelo, C.P. (1987). Local space approximation for treatment of impurities in polymers. Solitons in polyacetylene. *J. Chem. Phys.* 86: 1624–1631.

23 Anikin, N.A., Anisimov, V.M., Bugaenko, V.L. et al. (2004). LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecules. *J. Chem. Phys.* 121(3): 1266–1270.

24 Yang, W. (1991). Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* 66(11): 1438.

25 Imamura, A., Aoki, Y., and Maekawa, K. (1991). A theoretical synthesis of polymers by using uniform localization of molecular orbitals: proposal of an elongation method. *J. Chem. Phys.* 95(7): 5419–5431.

26 Korchowiec, J., Lewandowski, J., Makowski, M. et al. (2009). Elongation cutoff technique armed with quantum fast multipole method for linear scaling. *J. Comput. Chem.* 30(15): 2515–2525.

27 Aoki, Y. and Gu, F.L. (2012). Elongation method for delocalized nano-wires. *Prog. Chem.* 24(6): 886–909.

28 Kitaura, K., Ikeo, E., Asada, T. et al. (1999). Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* 313(3–4): 701–706.

29 Li, S., Li, W., and Fang, T. (2005). An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules. *J. Am. Chem. Soc.* 127: 7215–7226.

30 Fukui, K., Yonezawa, T., and Shingu, H. (1952). A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Chem. Phys.* 20(4): 722–725.

31 Woodward, R.B. (1973). The total synthesis of vitamin B12. *Pure Appl. Chem.* 33(1): 145–178. https://doi.org/10.1351/pac197333010145.

32 Boys, S.F. (1960). Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* 32(2): 296–299.

33 Foster, J.M. and Boys, S.F. (1960). Canonical configurational interaction procedure. *Rev. Mod. Phys.* 32(2): 300–302.

34 Marzari, N. and Vanderbilt, D. (1997). Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* 56(20): 12847–12865.

35 Edmiston, C. and Ruedenberg, K. (1963). Localized atomic and molecular orbitals. *Rev. Mod. Phys.* 35: 457–464.

36 Pipek, J. and Mezey, P.G. (1989). A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* 90: 4916–4926.

37 Pipek, J. (2000). Unique positive definite extension of Mulliken's charge populations of non-orthogonal atomic basis functions. *J. Mol. Struct. THEOCHEM* 501–502: 395–401.

38 Löwdin, P.-O. (1955). Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Phys. Rev.* 97(6): 1474–1489.

39 Reed, A.E. and Weinhold, F. (1983). Natural bond orbital analysis of near-Hartree–Fock water dimer. *J. Chem. Phys.* 78(6): 4066–4073.

40 Gu, F.L., Aoki, Y., Korchowiec, J. et al. (2004). A new localization scheme for the elongation method. *J. Chem. Phys.* 121(1–3): 10385–10391.

41 Adams, W.H. (1961). On the solution of the Hartree-Fock equation in terms of localized orbitals. *J. Chem. Phys.* 34(1): 89–102.

42 Gilbert, T.L. and Lykos, P.G. (1961). Maximum-overlap directed-hybrid orbitals. *J. Chem. Phys.* 34(1): 2199–2200.

43 Krol, M.C. and Altona, C. (1991). Theoretical investigations of the nature of intramolecular interactions: IV. Transferability of non-orthogonal molecular orbitals and application to mono-and difluoromethane. *Mol. Phys.* 72(2): 375–393.

44 Sorakubo, K., Yanai, T., Nakayama, K. et al. (2003). A non-orthogonal Kohn-Sham method using partially fixed molecular orbitals. *Theor. Chem. Acc.* 110: 328–337.

45 Paulus, B., Rościszewski, K., Stoll, H., and Birkenheuer, U. (2003). Ab initio incremental correlation treatment with non-orthogonal localized orbitals. *Phys. Chem. Chem. Phys.* 5(24): 5523–5529.

46 Sironi, M., Famulari, A., Raimondi, M., and Chiesa, S. (2000). The transferability of extremely localized molecular orbitals. *J. Mol. Struct. THEOCHEM* 529(1–3): 47–54.

47 Peng, L., Gu, F.L., and Yang, W. (2013). Effective preconditioning for ab initio ground state energy minimization with non-orthogonal localized molecular orbitals. *Phys. Chem. Chem. Phys.* 15(37): 15518–15527.

48 Meyer, B., Guillot, B., Ruiz-Lopez, M.F., and Genoni, A. (2016). Libraries of extremely localized molecular orbitals. 1. Model molecules approximation and molecular orbitals transferability. *J. Chem. Theory Comput.* 12(3): 1052–1067.

49 Meyer, B., Guillot, B., Ruiz-Lopez, M.F. et al. (2016). Libraries of extremely localized molecular orbitals. 2. Comparison with the pseudoatoms transferability. *J. Chem. Theory Comput.* 12(3): 1068–1081.

50 Yang, W. (1997). Absolute-energy-minimum principles for linear-scaling electronic-structure calculations. *Phys. Rev. B* 56(15): 9294.

51 Burger, S.K. and Yang, W. (2008). Linear-scaling quantum calculations using non-orthogonal localized molecular orbitals. *J. Phys.: Condens. Matter* 20(29): 294209.

52 Cui, G., Fang, W., and Yang, W. (2010). Reformulating time-dependent density functional theory with non-orthogonal localized molecular orbitals. *Phys. Chem. Chem. Phys.* 12(2): 416–421.

53 Peng, L., Peng, D., Gu, F.L., and Yang, W. (2022). Regularized localized molecular orbitals in a divide-and-conquer approach for linear scaling calculations. *J. Chem. Theory Comput.* 18: 2975–2982.

54 Fleming, I. (1977). *Frontier Orbitals and Organic Chemical Reactions*. Wiley.

55 Fleming, I. (2011). *Molecular Orbitals and Organic Chemical Reactions*. Wiley.

56 Fujimoto, H. and Fukui, K. (1997). *Frontier Orbitals and Reaction Paths: Selected Papers of Kenichi Fukui*, vol. 7. World Scientific.

57 Nguyen, A.Q., Anh, N.T., and Nguyên, T.A. (2007). *Frontier Orbitals: A Practical Manual*. Wiley.

58 Parr, R.G. and Yang, W. (1984). Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.* 106(14): 4049–4050.

59 Woodward, R.B. and Hoffmann, R. (1969). The conservation of orbital symmetry. *Angew. Chem. Int. Ed. Engl.* 8(11): 781–932.

60 Yu, J., Su, N.Q., and Yang, W. (2022). Describing chemical reactivity with frontier molecular orbitalets. *JACS Au* 2(6): 1383–1394.

61 Li, C., Zheng, X., Su, N.Q., and Yang, W. (2017). Localized orbital scaling correction for systematic elimination of delocalization error in density functional approximations. *Natl. Sci. Rev.* 5(2): 203–215.

62 Su, N.Q., Mahler, A., and Yang, W. (2020). Preserving symmetry and degeneracy in the localized orbital scaling correction approach. *J. Phys. Chem. Lett.* 11(4): 1528–1535.

63 Su, N.Q., Li, C., and Yang, W. (2018). Describing strong correlation with fractional-spin correction in density functional theory. *Proc. Natl. Acad. Sci. U.S.A.* 115(39): 9678–9683.

64 Cohen, A.J., Mori-Sánchez, P., and Yang, W. (2008). Fractional spins and static correlation error in density functional theory. *J. Chem. Phys.* 129(12): 121104.

65 Cohen, A.J., Mori-Sánchez, P., and Yang, W. (2008). Insights into current limitations of density functional theory. *Science* 321(5890): 792–794.

66 Cohen, A.J., Mori-Sánchez, P., and Yang, W. (2011). Challenges for density functional theory. *Chem. Rev.* 112(1): 289–320.

67 Mori-Sánchez, P., Cohen, A.J., and Yang, W. (2008). Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* 100(14): 146401.

68 Zheng, X., Cohen, A.J., Mori-Sánchez, P. et al. (2011). Improving band gap prediction in density functional theory from molecules to solids. *Phys. Rev. Lett.* 107(2): 026403.

69 Zhang, J.-X., Sheong, F.K., and Lin, Z. (2018). Unravelling chemical interactions with principal interacting orbital analysis. *Chem. Eur. J.* 24(38): 9639–9650.