

## Multipole Model for the Electron Group Functions Method

**A. L. Tchougréeff**

*Poncelet Laboratory, Independent University of Moscow, Moscow Center for Continuous Mathematical Education, Bolshoy Vlasyevskiy Pereulok 11, 119002 Moscow, Russia*

**A. M. Tokmachev\* and R. Dronskowski**

*JARA, Institut für Anorganische Chemie, RWTH Aachen, Landoltweg 1, 52056 Aachen, Germany*

*Received: June 18, 2009; Revised Manuscript Received: September 2, 2009*

Electron groups provide a natural way to introduce local concepts into quantum chemistry, and the wave functions based on the group products can be considered as a framework for constructing efficient computational methods in terms of “observable” parts of molecular systems. The elements of the group wave functions (electronic structure variables) can be optimized by requiring the number of operations proportional to the size of the molecule. This directly leads to computational methods linearly scaling for large molecular systems. In the present work we consider a particular case of such a wave function implemented for the semiempirical NDDO Hamiltonian. The electron groups are expressed in terms of optimized atomic (hybrid) orbitals with chemical bonds described by geminals and the delocalized groups described by Slater determinants (with or without spin restriction). This scheme is very fast by itself but its speed is considerably limited by the computations of the interatomic Coulomb interactions. Here we develop a consistent method based on group functions which uses the multipole scheme for interatomic interactions. The explicit usage of the atomic multipoles makes the method extremely fast, although the numerical efficiency is largely achieved due to the local character of the electron groups involved. We discuss numerical characteristics of the new method as well as its possible parametrization. We apply this method to study dodecahedral water clusters with hydrogen fluoride substitution and base the analysis on the exhaustive calculation of all symmetry-independent hydrogen-bond networks.

### 1. Introduction

Electron groups are natural blocks of the electronic structures of many molecules and solids. They manifest themselves in chemical and physical experiments. Their integrity is largely conserved during chemical transformations (if they are not directly involved) and spectroscopic measurements. The most important examples of the electron groups are chemical bonds, lone pairs, d-shells, and other chromophores. It is undoubtful that the concept of local electron groups is extremely valuable for qualitative understanding of chemical structure and reactivity. Although the concepts of delocalized molecular orbitals has also proven their usefulness in this respect (for example, the Woodward–Hoffmann rules), the local concepts still play the major role in the chemical thinking, despite that the realm of quantitative electronic structure calculations is dominated by methods based on the delocalized orbitals. This situation is induced by common opinion that methods based on molecular orbitals are easier to formulate and they are more numerically efficient. But is it necessarily true?

The dominance of the methods based on molecular orbitals (MOs) appeared because they provide an easy way to compute the structure and properties of simple small and medium-sized molecules with a sufficient accuracy, and their main rivals (valence bond methods) are usually more complicated without giving an adequate compensation in the results (see, nonetheless, the discussion in ref 1). This dominance does not go too far. When someone wants to compute the structure of large and very

large molecules, the standard methods fail because the required computational resources grow disastrously. This situation is usually considered as a demand for developing special-purpose computational methods. There are two alternative solutions proposed in the literature. The first one is based on the development of the so-called  $O(N)$  methods with resources scaling linearly with the system size.<sup>2,3</sup> The possibility of developing such methods is normally based on local properties of density matrices. The alternative solution is given by hybrid QM/MM (quantum mechanical/molecular mechanical) methods using elaborate QM methods for the reaction center (or any other relevant chromophore) and MM force fields for the environment.<sup>4</sup> Although the QM methods in QM/MM hybrids are usually standard methods of quantum chemistry based on delocalized MOs, the applicability of the scheme tacitly assumes significant locality: namely that of chemical transformations (but more generally of any interaction that can be classified as chemical) allow one to single out a reaction center and the major part of the system is described by MM, which is intrinsically local. Therefore, when the speed of the computations becomes an issue, the local description becomes a natural choice for many molecular systems.

The situation does not, however, reduce to the point of numerical efficiency. Local electron groups, if correctly selected/singled out, represent observable fragments of molecular systems, providing adequate description of their electronic structures. In this sense the concept of electron groups formalizes and extends the old concept of chromophore. The important question of the grounds on which the local groups representing

\* Corresponding author. E-mail: andrei.tokmachev@ac.rwth-aachen.de.

electronic structure of molecular system are to be selected has in general terms been solved by Ruedenberg<sup>3</sup> who suggested a kind of extension of the standard quantum mechanical definition of observability from the quantities to the entities. The following has been proposed: fragments in a molecular system can be singled out if these latter are observable, so that they manifest a reproducible and natural behavior; if for a series of molecules variations of fragments fit to that or another curve and its parameters can be found empirically by considering enough of the series members this proves that singling out the fragments makes sense.<sup>5</sup> This definition allowed throughout the development of theoretical chemistry the ability to single out numerous fragments that can be two-center two-electron bonds, or conjugate  $\pi$ -systems, open d-shells, atomic cores, etc. The idea is that an adequate semiempirical theory must be constructed in terms of such observable objects. Previously we undertook attempts to consider typical chemical bonds we face in "organic" chemistry as observable entities to be used to construct the electronic wave function of molecular system in their terms.

Semiempirical methods based on MOs are usually considered to be fast enough. At the same time these methods are far from linear scaling: diagonalization of the Fock matrix brings a cubic dependence of the computation time on the system size (number of basis one-electron functions). Therefore, the standard methods are not efficient enough when the system is really large or when a large number of points on the potential energy surface is necessary to be determined. It also means that the Hartree–Fock–Roothaan-based semiempirical methods do not represent a real alternative to the MM methods in the hybrid schemes, although the development of hybrid QM/QM methods<sup>6</sup> is quite advantageous because it makes possible a sequential derivation of the junction between subsystems.<sup>4</sup> There exist methods to avoid the diagonalization: the density matrix search method,<sup>7</sup> the divide-and-conquer method,<sup>8</sup> and those based on localized MOs in either nonorthogonal<sup>9</sup> or orthogonal<sup>10</sup> contexts, but there is also an alternative way to reach the computational efficiency.

One of the main advantages of the local description of the electronic structure is the possibility to take into account the electron correlation effectively. The representation of the wave function in the form of the antisymmetrized product of strongly orthogonal electron groups dating back to ref 11 allows one to do it in a simple way. The interaction between electron groups is controlled by one-particle intragroup density matrices and, therefore, the types of the wave functions used for description of the groups can be chosen almost independently. By this, the level of the correlated description for each of the involved electron groups is determined by the problem under study. For example, if one is interested in the d–d transitions, it is advisable to describe the d-shell by full configuration interaction (CI) while the rest can be described by MOs.<sup>12</sup> On the other hand, in the case of electron groups spanning only two orbitals, even the full CI scheme becomes cheap.

Previously we constructed a series of semiempirical methods employing two-electron groups (strictly local geminals, SLG) representing chemical bonds and lone pairs.<sup>13–15</sup> Those methods were quite successful in describing heats of formation, molecular geometries, and ionization potentials. Recently, we extended them to treat more general situations when the presence of delocalized electron groups is also possible.<sup>16</sup> This method combines geminals with electron groups described by Slater determinants and it shares the success of the methods employing geminals only.

One of the most important characteristics of the group-based theories proposed is the possibility to construct linearly scaling

methods on their basis by introducing cutoff parameters for interatomic interactions. In the case of the MINDO/3 parametrization of the molecular Hamiltonian the method is almost linearly scaling by itself because the coefficient at the quadratic term in the dependence of the computation time on the number of atoms is very small. In the case of NDDO Hamiltonians the quadratic term dominates even for the smallest molecules because of numerous transformations of molecular integrals from the original basis of atomic orbitals to that of hybrid orbitals (see below).

It turns out that it is possible to make the methods much faster by slightly redefining the semiempirical Hamiltonians. In ref 17 we devised a multipole scheme for long-range Coulomb interactions in the case of strictly local geminal wave function. It is based on the usual approach for computing semiempirical two-center repulsion integrals in the NDDO approximation,<sup>18</sup> but the true multipoles are used and not their representations by point charge distributions. It makes the whole procedure 30 times faster (while conserving its linear scalability) in the limiting case of large molecules (having more than 300 atoms) without a noticeable change in the numerical results. Regretfully, the method is internally inconsistent because it uses two different schemes for describing Coulomb interactions: one for short-range interactions (including chemical bonds) and another for long-range interactions. At the same time the short-range interactions can be treated with the same approach because the problem of computing two-center repulsion molecular integrals for short interatomic distances is solved in semiempirical methods by using a special form of the potential, which is not affected by the changes in the computational scheme.

In the present paper we consider two significant extensions for the scheme: (1) the possibility to have delocalized electron groups described by Slater determinants (with or without spin restriction); (2) a single multipole scheme for all interatomic Coulomb interactions. This brings significant modifications to the definition of the method, which are described below. We test the efficiency of the method by computing the whole set of H-bond networks for water clusters in the form of a pentagonal dodecahedron with one water molecule replaced by hydrogen fluoride.

## 2. Method

**2.1. Wave Function.** We define the method of calculation in several steps starting with the trial wave function. As stated above we consider the electronic structure of a molecule as one composed by electron groups representing chemically motivated molecular fragments. Each electron group can be well characterized by an integral number of electrons and a definite multiplicity. Following the McWeeny's electron group functions formalism<sup>11</sup> the molecular wave function is given by an antisymmetrized product of wave functions for electron groups. The strong orthogonality condition imposed on the wave functions of the groups involved is equivalent to expanding wave functions for different groups in mutually orthogonal subsets of orbitals.<sup>19</sup> In this framework the choice of the wave function reduces to the choice of one-electron states spanning carrier spaces for electron groups as well as the choice of number of electrons and of the form for their wave functions in these spaces.

In the general formulation the molecular orbitals should be obtained by an orthogonal transformation of the initial set of atomic orbitals (AOs), which is assumed to be orthonormal in the realm of semiempirical theories (implicitly Löwdin orthogonalized AOs—OAOs). It is important that the optimal one-electron states are generally local, which complies with the local

character of the correlation hole. According to the chemical intuition, these one-electron states can be thought of as hybrid orbitals (HOs) that enter into different valence bond structures giving rise to various bonding patterns. This picture is, of course, only an approximation to the general formulation because the delocalization of orbitals is lost but it allows for a clear interpretation of the calculated electronic structures as well as significant computational advantages as it will be seen below. We define HOs as orbitals produced by orthogonal transformations of AOs centered on the same atom:

$$t_{p\sigma}^+ = \sum_{i \in A} h_{pi}^A a_{i\sigma}^+ \quad (1)$$

where  $t_{p\sigma}^+$  is the creation operator for spin-HO  $|t_{p\sigma}\rangle$  with spin projection  $\sigma$  and  $h^A$  is an  $SO(4)$  matrix determining the transformation from the AO to the HO basis set on non-hydrogen atom A. Each HO is assigned to some electron group and sets of thus defined HOs span the carrier spaces for all the electron groups.

The next point in the wave function definition is how to assign the wave functions for electrons residing in each carrier space - the functions for electron groups. It is reasonable to describe each electron group by taking into account the electron configurations that correspond to the full configuration interaction scheme. At the same time, it will be very time-consuming in the case of large carrier spaces corresponding, for example, to extended  $\pi$ -electron systems. A more economic way to define the trial wave function is to allow at least two types of electron groups in this "organic" environment: geminals  $g_m^+$  (representing two-electron two-center chemical bonds) and SCF groups described by antisymmetrized products of molecular spin-orbitals  $b_{i\sigma}^+$ . The molecular wave function is, therefore, given by the expression:

$$|\Psi\rangle = \left( \prod_{i\sigma} b_{i\sigma}^+ \right) \left( \prod_m g_m^+ \right) |0\rangle \quad (2)$$

In this equation the possibility of several SCF groups is not explicitly shown. If it is the case eq 2 is still valid and the assignment of MOs to electron groups is reflected only in the sets of HOs, over which the molecular orbitals for respective electron groups are expanded.

Each molecular spin-orbital is a linear combination of the hybrid spin-orbitals assigned to the SCF-treated electron group under consideration:

$$b_{i\sigma}^+ = \sum_{p \in \{\text{SCF}\}} c_{i\sigma p}^+ t_{p\sigma}^+ \quad (3)$$

where  $p$  belongs to the corresponding HO basis. To write down the expression for the geminal wave function  $g_m^+|0\rangle$ , we denote the "right" and "left" ends of the  $m$ th chemical bond as  $|r_m\rangle$  and  $|l_m\rangle$ . Each geminal is a superposition of all three singlet two-electron configurations (two ionic and one covalent):

$$g_m^+ = u_m r_{m\alpha}^+ r_{m\beta}^+ + v_m l_{m\alpha}^+ l_{m\beta}^+ + w_m (r_{m\alpha}^+ l_{m\beta}^+ + l_{m\alpha}^+ r_{m\beta}^+) \quad (4)$$

with the normalization condition for the amplitudes (mixing coefficients)  $u_m$ ,  $v_m$ , and  $w_m$  given as

$$\langle 0 | g_m g_m^+ | 0 \rangle = u_m^2 + v_m^2 + 2w_m^2 = 1 \quad (5)$$

In the case of a lone pair it can be equivalently considered as either an SCF or geminal function (with only one surviving configuration).

To go further with the definition of the method (which could be abbreviated as SLG/SCF with multipoles or SLG/SCF(m)), it is practical to employ reduced density matrices. Only intragroup elements of the density matrices are necessary because the one-electron density matrix is block-diagonal and all nonvanishing intergroup elements of the two-electron density matrix are merely products of respective elements of the one-electron density matrix. The spin-dependent one-electron density matrix for the SCF groups is defined in a standard way:

$$P_{pq}^\sigma = \sum_{i\sigma} c_{i\sigma p} c_{i\sigma q} \quad (6)$$

For geminals we define both one- and two-electron density matrices:

$$\begin{aligned} P_{l_m l_m}^\sigma &= \langle 0 | g_m^+ t_{m\sigma}^+ t_{m\sigma}^+ g_m^+ | 0 \rangle, & P_{r_m r_m}^\sigma &= u_m^2 + w_m^2, \\ P_{l_m l_m}^\sigma &= v_m^2 + w_m^2, & P_{r_m l_m}^\sigma &= P_{l_m r_m}^\sigma = (u_m + v_m)w_m, \\ \Gamma_{l_m l_m}^\sigma &= \langle 0 | g_m^+ t_{m-\sigma}^+ t_{m\sigma}^+ t_{m\sigma}^+ t_{m-\sigma}^+ g_m^+ | 0 \rangle, \\ \Gamma_{r_m r_m}^\sigma &= u_m^2, & \Gamma_{l_m l_m}^\sigma &= v_m^2, & \Gamma_{r_m l_m}^\sigma &= \Gamma_{l_m r_m}^\sigma = w_m^2 \end{aligned} \quad (7)$$

These elements are, obviously, spin-independent in the case of singlet geminals. It is also convenient to introduce the traces of one- and two-electron density matrices over spin indices

$$P_{pq} = \sum_{\sigma} P_{pq}^\sigma, \quad \Gamma_{pq} = \sum_{\sigma} \Gamma_{pq}^\sigma \quad (8)$$

referred hereinafter as the total electron density matrices.

**2.2. Molecular Energy.** The molecular energy is the average of a semiempirical Hamiltonian based on the NDDO approximation over the wave function of the form eqs 2–4. The evaluation goes on easier if the Hamiltonian is rewritten with respect to the basis of HOs. This transformation does not affect the overall structure of the Hamiltonian, which remains the sum of one- and two-center contributions:

$$H = \sum_A H^A + \sum_{A<B} H^{AB} \quad (9)$$

as in the AOs basis. The one-center part can be written as

$$\begin{aligned} H^A &= \sum_{i' \in A} (U_{i'}^A + \sum_{B \neq A} V_{i',B}^A) \sum_{\sigma} t_{i\sigma}^+ t_{i\sigma}^+ + \\ &\frac{1}{2} \sum_{i't''i''' \in A} (tt't''t''')^A \sum_{\sigma\tau} t_{i\sigma}^+ t_{i'\tau}^+ t_{i''\sigma}^+ t_{i'''\tau}^+ \end{aligned} \quad (10)$$

and the two-center contributions have the following form:

$$H^{AB} = - \sum_{\substack{t \in A \\ t' \in B}} \beta_{tt'}^{AB} \sum_{\sigma} (t_{\sigma}^{+} t'_{\sigma} + \text{hc}) + \sum_{\substack{tt' \in A \\ t''t'' \in B}} (tt' | t''t'')^{AB} \sum_{\sigma\tau} t_{\sigma}^{+} t''_{\tau} t'_{\tau} t''_{\sigma} \quad (11)$$

where hc denotes the hermitean conjugated term. The molecular integrals in the basis of HOs can be readily calculated through the corresponding integrals in the AO basis (detailed expressions for the integral transformations are given in refs 13 and 14).

The total energy is the sum of the core–core repulsion and the average of the electronic Hamiltonian over the electronic wave function. It can be regrouped in a sum of two one-center and four two-center contributions:

$$E^{\text{total}} = \sum_A \{E_{\text{attr}}^A + E_{\text{coul}}^A\} + \sum_{A < B} \{E_{\text{bond}}^{AB} + E_{\text{exch}}^{AB} + E_{\text{mult}}^{AB} + E_{\text{spec}}^{AB}\} \quad (12)$$

The simplest contribution is  $E_{\text{spec}}^{AB}$ , which represents the specific part of the core–core repulsion that is usually added to the pure Coulomb interaction of the core charges in semiempirical theories. The form of this contribution depends on the parametrization scheme and differs, for example, for MNDO, AM1, and PM6.<sup>20–22</sup> Other contributions arise from averaging the electronic Hamiltonian.

Attraction of electrons on AOs to the core of the atom on which the orbital is centered is given as

$$E_{\text{attr}}^A = \sum_m \sum_{pq \in \{m,A\}} U_{pq}^A P_{pq} \quad (13)$$

where  $p, q \in \{m, A\}$  means that HOs  $p$  and  $q$  are located on atom A and they are both assigned to electron group  $m$ . When  $m$  is a geminal, only one of the HOs spanning its carrier space is allowed to rest on atom A and, therefore, only the diagonal contribution survives.

The second one-center contribution is due to Coulomb repulsion of electrons:

$$E_{\text{coul}}^A = \frac{1}{2} \sum_{mn} \sum_{\substack{pp' \in \{m,A\} \\ qq' \in \{n,A\}}} [(pp' | qq')^A P_{pp'} P_{qq'} - (pq | qp')^A \sum_{\sigma} P_{pp'}^{\sigma} P_{qq'}^{\sigma}] + \frac{1}{4} \sum_{m \in \{\text{SLG}\}} \sum_{p \in \{m,A\}} (pp | pp)^A [2\Gamma_{pp} - (P_{pp})^2] \quad (14)$$

where the last contribution appears for geminals only and describes the deviation of the two-electron density matrices from the product of one-electron densities, the correlation. The Coulomb intraatomic energy can be further classified on the basis of origin of contributions from one or two electron groups.

The classification of two-center contributions is more subtle. After separating the “classical” Coulomb interaction between electron distributions on atoms ( $E_{\text{mult}}^{AB}$ ), the two-center energy is

presented as a sum of intragroup and intergroup interactions. The bonding (intragroup) contribution arises from both one- and two-electron operators in the Hamiltonian:

$$E_{\text{bond}}^{AB} = -2 \sum_m \sum_{\substack{p \in \{m,A\} \\ q \in \{m,B\}}} \beta_{pq}^{AB} P_{pq} - \sum_{m \in \{\text{SCF}\}} \sum_{\substack{pp' \in \{m,A\} \\ qq' \in \{m,B\}}} (pp' | qq')^{AB} \sum_{\sigma} P_{pp'}^{\sigma} P_{qq'}^{\sigma} + \sum_{m \in \{\text{SLG}\}} \sum_{\substack{p \in \{m,A\} \\ q \in \{m,B\}}} (pp | qq)^{AB} [\Gamma_{pq} - P_{pp} P_{qq}] \quad (15)$$

The interbond contribution comes from the exchange interaction between electrons in the groups sharing the same pair of atoms (this happens when multiple bonds are present):

$$E_{\text{exch}}^{AB} = -2 \sum_{m < n} \sum_{\substack{p \in \{m,A\} \\ q \in \{m,B\}}} \sum_{\substack{p' \in \{n,A\} \\ q' \in \{n,B\}}} (pp' | qq')^{AB} \sum_{\sigma} P_{pp'}^{\sigma} P_{qq'}^{\sigma} \quad (16)$$

**2.3. Multipole Expansion.** The last contribution to the energy represents the Coulomb interaction between the charge distributions assigned to the atoms. It deserves a special consideration. For two atoms A and B this energy can be written as

$$E_{\text{mult}}^{AB} = Z^A Z^B (ss | ss)^{AB} + \sum_{pp' \in A} \sum_{qq' \in B} P_{pp'}^A P_{qq'}^B (pp' | qq')^{AB} - Z^A \sum_{qq' \in B} P_{qq'}^B (ss | qq')^{AB} - Z^B \sum_{pp' \in A} P_{pp'}^A (pp' | ss)^{AB} \quad (17)$$

As stated above, only the intragroup density matrix elements are nonvanishing and the sum over basis HOs  $p$  and  $p'$  can be regrouped as a sum over electron groups  $\sum_{pp'} = \sum_m \sum_{pp' \in m}$ . We note that the energy contribution eq 17 is a bilinear form with respect to charge distributions and rewrite it as

$$E_{\text{mult}}^{AB} = [-Z^A (ss | s^A + \sum_{pp' \in A} P_{pp'}^A (pp' | s^A) \bullet [-Z_B | ss)^B + \sum_{qq' \in B} P_{qq'}^B (qq' | q^B)] \quad (18)$$

where  $\bullet$  formally denotes the semiempirical Coulomb interaction. The strictly local character of one-electron states allows for a natural definition of atomic multipoles. For convenience we write them in the units  $-e$ . The atomic charge (monopole) is given by the diagonal elements of the electron density:

$$q^A = \sum_{p \in A} P_{pp}^A - Z^A \quad (19)$$

The dipole moment for an (*sp*)-distribution and the quadrupole moment for a (*pp*)-distribution depend on characteristic lengths<sup>18</sup> defined in their turn by the principal quantum number *n* of the *sp*-shell and the orbital exponents:

$$D_1 = \frac{2n+1}{\sqrt{3}} \frac{(4\xi_{ns}\xi_{np})^{n+1/2}}{(\xi_{ns} + \xi_{np})^{2n+2}} \quad (20)$$

$$D_2 = \sqrt{\frac{(2n+1)(2n+2)}{20}} \xi_{np}^{-1}$$

The expressions for dipole and quadrupole moments simplify if the quaternion representation ( $s_m, \vec{v}_m$ ) is used for HOs,<sup>23,24</sup> where the scalar  $s_m$  and vector  $\vec{v}_m$  components reflect the transformation properties of the HO coefficients.

The dipole moment characterizing the charge distribution at atom *A* is then

$$\vec{\mu}^A = D_1^A \sum_{pq \in A} P_{pq}^A (s_p^A \vec{v}_q^A + s_q^A \vec{v}_p^A) \quad (21)$$

The quadrupole moment as defined in the standard field theory is traceless because its trace enters only those contributions to the energy that are proportional to the Laplacian of the potential (which vanishes exactly for the Coulomb interaction). Since the Poisson equation  $\Delta f = 0$  is not generally valid for semiempirical potentials, we define a second-order tensor  $S^A$  with a nonvanishing trace instead of the usual quadrupole moment:

$$\hat{S}^A = \frac{1}{2} \sum q x_\alpha x_\beta = (D_2^A)^2 \sum_{pq \in A} P_{pq}^A \vec{v}_p^A \otimes \vec{v}_q^A \quad (22)$$

which describes the second moment of the spatial electronic distribution at atom *A*.

The next step is to find the energy of the interaction between multipoles. Semiempirical potentials (for example, that of Dewar, Sabelli, and Klopman<sup>25,26</sup>) depend on the type of interaction (the *SO*(3) tensor ranks  $l_1$  and  $l_2$  of the multipoles involved). The energy of the interaction between multipoles can be readily written in the tensor form:

$$E_{\text{mult}}^{\text{AB}} = q^A G_{00} q^B + q^A G_{01}^{\alpha} \vec{\mu}_\alpha^B - \vec{\mu}_\alpha^A G_{10}^{\alpha} q^B - \vec{\mu}_\alpha^A G_{11}^{\alpha\beta} \vec{\mu}_\beta^B + q^A G_{02}^{\alpha\beta} \hat{S}_{\alpha\beta}^B + \hat{S}_{\alpha\beta}^A G_{20}^{\alpha\beta} q^B - \vec{\mu}_\alpha^A G_{12}^{\alpha\beta\gamma} \hat{S}_{\beta\gamma}^B + \hat{S}_{\alpha\beta}^A G_{21}^{\alpha\beta\gamma} \vec{\mu}_\gamma^B + \hat{S}_{\alpha\beta}^A G_{22}^{\alpha\beta\gamma\delta} \hat{S}_{\gamma\delta}^B \quad (23)$$

where

$$G_{l_1 l_2}^{\alpha\beta\dots\mu} = \nabla_\alpha \nabla_\beta \dots \nabla_\mu f_{l_1 l_2}(R) \quad (24)$$

and  $\nabla_\alpha, \nabla_\beta, \dots$  denote the Cartesian components of the gradient operator applied to the potential  $f_{l_1 l_2}$ . To write down the explicit expressions for the energy contributions, it is convenient to introduce a new function:

$$F_{l_1 l_2}^{\alpha\beta}(R) = R^{\beta-\alpha} \left( \frac{1}{R} \frac{d}{dR} \right)^\beta f_{l_1 l_2}(R) \quad (25)$$

Next, we specify different cases according to the values of  $l_1$  and  $l_2$  (all the functions below are estimated at  $R = R_{\text{AB}}$  and for brevity we omit  $R$  and subscripts  $l_1$  and  $l_2$  from the notation for the function  $F$ ). It is noteworthy that the expression for  $E_{l_1 l_2}^{\text{AB}}$  is a linear combination of only those  $F_{l_1 l_2}^{\alpha\beta}$  that satisfy three conditions: (1)  $\alpha$  and  $\beta$  are non-negative integers, (2)  $\alpha + \beta = l_1 + l_2$ , (3)  $\alpha \leq \beta$ .

The charge–charge,  $\{l_1 l_2\} = \{00\}$ , contribution is the simplest one:

$$E_{00}^{\text{AB}} = q^A q^B F^{00} \quad (26)$$

The charge–dipole,  $\{l_1 l_2\} = \{01\}$  or  $\{10\}$ , interaction results in two contributions expressed through the unit vector  $\vec{n}_{\text{AB}}$  directed from atom *A* to atom *B*:

$$E_{01}^{\text{AB}} = q^A F^{01}(\vec{n}_{\text{AB}}, \vec{\mu}^B) \quad (27)$$

and analogously for  $E_{10}^{\text{AB}}$ .

The dipole–dipole,  $\{l_1 l_2\} = \{11\}$ , interaction is a sum of two contributions. The first one is proportional to the scalar product of dipole moments and the second one is as usual proportional to the product of their projections on the vector  $\vec{n}_{\text{AB}}$ :

$$E_{11}^{\text{AB}} = -F^{11}(\vec{\mu}^A, \vec{\mu}^B) - F^{02}(\vec{\mu}^A, \vec{n}_{\text{AB}})(\vec{\mu}^B, \vec{n}_{\text{AB}}) \quad (28)$$

The charge–quadrupole,  $\{l_1 l_2\} = \{02\}$  or  $\{20\}$ , interaction is given by two contributions depending on the order of  $l_1$  and  $l_2$ :

$$E_{02}^{\text{AB}} = q^A [F^{11} \text{tr} \hat{S}^B + F^{02}(\vec{n}_{\text{AB}}, \hat{S}^B, \vec{n}_{\text{AB}})] \quad (29)$$

and an analogous expression holds for  $E_{20}^{\text{AB}}$ . When the Poisson equation is valid the contributions proportional to  $\text{tr} \hat{S}$  vanish.

The dipole–quadrupole,  $\{l_1 l_2\} = \{12\}$  or  $\{21\}$ , interaction is more complex and requires the third derivatives of the semiempirical potentials:

$$E_{12}^{\text{AB}} = F^{12}[(\vec{\mu}^A, \vec{n}_{\text{BA}}) \text{tr} \hat{S}^B + 2(\vec{\mu}^A, \hat{S}^B, \vec{n}_{\text{BA}})] + F^{03}(\vec{\mu}^A, \vec{n}_{\text{BA}})(\vec{\mu}_{\text{AB}}^B, \hat{S}^B, \vec{\mu}_{\text{AB}}^B) \quad (30)$$

and an analogous expression is valid for  $E_{21}^{\text{AB}}$ .

The quadrupole–quadrupole,  $\{l_1 l_2\} = \{22\}$  is the last contribution:

$$E_{22}^{\text{AB}} = F^{22}[\text{tr} \hat{S}^A \text{tr} \hat{S}^B + 2 \text{tr}(\hat{S}^A \hat{S}^B)] + F^{13}(\vec{n}_{\text{AB}}, \{\hat{S}^A \text{tr} \hat{S}^B + \hat{S}^B \text{tr} \hat{S}^A + 4\hat{S}^A \hat{S}^B\}, \vec{n}_{\text{AB}}) + F^{04}(\vec{n}_{\text{AB}}, \hat{S}^A, \vec{n}_{\text{AB}})(\vec{n}_{\text{AB}}, \hat{S}^B, \vec{n}_{\text{AB}}) \quad (31)$$

and it includes derivatives of a semiempirical potential up to the fourth order.

**2.4. Optimization.** At this point we are equipped for a search of the approximation to the electronic wave function of the ground state through a variational procedure for the energy. The

SLG/SCF(m) energy given by eq 12 depends on two classes of the electronic structure variables (ESVs): (1) elements of the one- and two-electron density matrices, which, in turn, are functions of the MO coefficients  $c_{i,p}$  and the geminal amplitudes  $u_m$ ,  $v_m$ , and  $w_m$ ; (2) hybridization transformation matrices  $h_{pi}^A$  determining the molecular integrals. To minimize the total energy we use an iteration scheme alternating optimizations of the two sets of ESVs.

The first set of ESVs can be obtained by constructing effective Hamiltonians for the electron groups with their subsequent diagonalization. In the case of SLG groups the Hamiltonian is a two-electron operator, while in the case of SCF groups it is a Fock operator. We implemented three standard procedures for the latter: spin-restricted Hartree–Fock (RHF), spin-restricted open-shell Hartree–Fock (ROHF), and spin-unrestricted Hartree–Fock (UHF) ones. The structure of the effective Hamiltonians is quite standard but the one-electron matrix elements are modified due to Coulomb and exchange interactions between electron groups in accord with the general theory of group functions.<sup>11</sup> An essential part of the interactions is expressed through multipoles and the contribution to the effective Hamiltonians from  $E_{\text{mult}}^{\text{AB}}$  is obtained by taking its derivatives with respect to the elements of the one-electron density matrix which enter the multipole definitions. Moreover, multipole–multipole interactions with  $l_1 + l_2 > 0$  are relatively short-ranged and a series of cutoff parameters can be naturally introduced to speed up the calculation without big losses in precision.

To optimize the second set of ESVs we implement a gradient minimization procedure. It is clear that the elements of the  $h_{pi}^A$  matrices are not independent variables. The  $SO(4)$  structure of the set of HOs residing on each “heavy” atom assures that each matrix is parametrized by six parameters. These parameters can be taken either as six Jacobi angles determining rotations in two-dimensional subspaces or as elements of two normalized quaternions in the quaternion parametrization of the  $SO(4)$  group.<sup>23</sup> The gradient of the energy with respect to the matrices  $h_{pi}^A$  (and, therefore, the sextuples of independent parameters) can be readily obtained because the dependence of the energy on the hybridization matrices is explicitly known in the present setting.

### 3. Results and Discussion

The method described in the previous section combines two ideas: the possibility to represent the electronic structure of molecules in terms of the wave functions of its “observable” fragments and the possibility to use multipole expansions to evaluate matrix elements of Coulomb electron–electron interatomic repulsion. This combination turns out to be extremely efficient. The SLG/SCF method with multipoles has been implemented as a program suite and its numerical characteristics have been studied. Some of the method features can be understood even without any numerical tests. First of all, the method allows for different types of electron groups and its area of applicability should be quite wide. Second, it is a linear scaling method capable of fast calculation of extremely large molecular systems (unless a delocalized SCF-treated group extends to the good portion of the entire system). It is especially important that the dominant part of the two-center energy is given by multipole–multipole interactions and the Coulomb energy dependence on the hybridization ESVs enters only through dipole and quadrupole moments. It makes the whole procedure significantly faster. It is shown in ref 17 that the method based on the strictly local geminals is accelerated 30

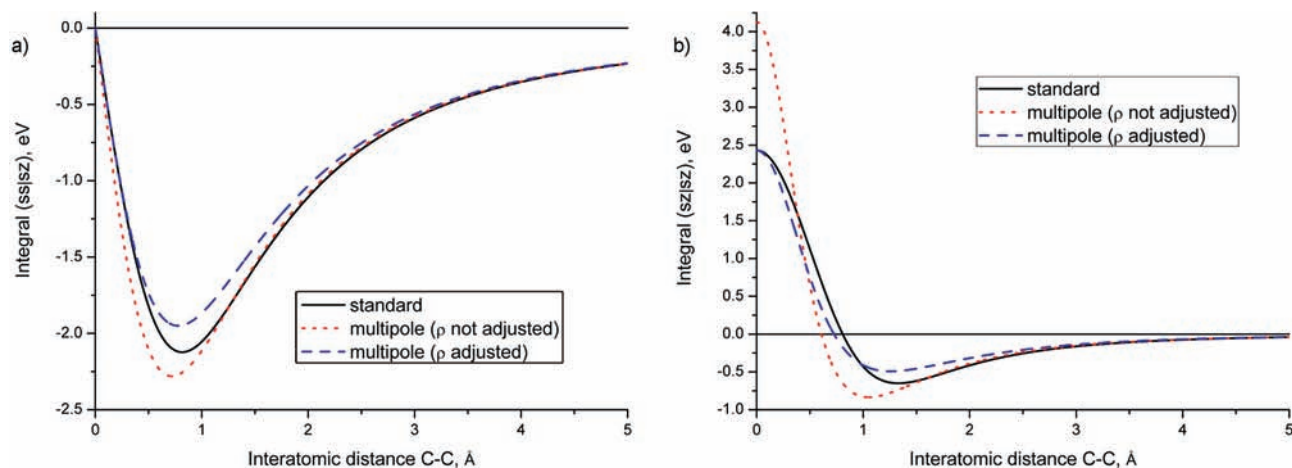
times by using the multipole expansions for the long-range Coulomb matrix elements. The same acceleration is (according to our numerical tests) characteristic for the SLG/SCF(m) method presented here. It is not surprising because calculating the long-range interactions constitutes the major part of the time consumed by the procedure. As far as we know, this method is the fastest among the existing quantum chemical methods.

The linear scalability of the method and a small coefficient at the linear term are ensured by the method’s general structure. At the same time, one can wonder about the price for these advantages. It is quite usual that the linear scalability of a method is paid by decreasing its numerical accuracy because the acceleration is obtained through neglecting some contributions to the energy. Introducing wave functions based on electron groups we lose some delocalization contributions but we gain intragroup electron correlation. It is not trivial to predict the overall balance of this exchange. Numerical experiments show that the methods based on electron groups better describe the physical properties of the molecules (heats of formation, equilibrium geometries, etc.) than their SCF counterparts.<sup>13,14,16</sup> For example, the SLG scheme allows one to describe correctly torsion angles in cyclobutane and hydrogen peroxide molecules and it generally leads to much better bond lengths between atoms with lone pairs.<sup>14</sup> This situation does not change when one uses multipole expansions only for long-range interactions<sup>17</sup> between atoms that do not have common electron groups because their effect on the energy is quite small. In the case of alkanes  $C_nH_{2n+2}$  with large values of  $n$  it leads only to the difference of 0.03 kcal/mol per carbon atom.

If the multipole scheme is used for all Coulomb interatomic interactions the situation is quite different. The original NDDO scheme for evaluating two-center electron repulsion integrals uses concept of multipoles but in practice these latter are replaced by fictitious charge distributions, which means that the symmetry properties of the multipole tensors are reproduced only approximately and the interactions involve higher multipoles. It does not bring any problems only because the molecular integrals are always evaluated in the diatomic coordinate frame. In our constructions, we explicitly use the symmetry properties of the multipoles (in fact, it is the source of the computational efficiency of the procedure). Therefore, we had to define atomic multipoles explicitly.

Two procedures (explicit multipoles and their charge distribution approximations) have the same behavior when one considers the interactions between distant multipoles. It is not the case when the interatomic distances are small (somewhat shorter than the bond-length). Figure 1 demonstrates the difference between these two schemes for two types of molecular integrals ( $(ss|sz)^{\text{CC}}$  and  $(sz|sz)^{\text{CC}}$ ) for a pair of carbon atoms computed using the standard scheme and true multipoles with or without adjusting the parameter  $\rho$  of the Dewar–Sabelli–Klopman potential to satisfy the available boundary conditions (one-center limit of the two-electron integrals, in this case  $(sz|sz)^{\text{CC}}$  becomes the intraatomic exchange integral). This figure shows that the differences are significant at the interatomic distances close to the normal bond lengths. At the same time the molecular integrals computed with true multipoles are closer to the analytical ones than those based on the fictitious charge distributions, especially for small interatomic distances (compare, for example, Figure 1 and the data of ref 18).

Taking this into account, we can expect that using multipoles to estimate bonding contributions to the molecular energy (eqs 15 and 16) will result in somewhat different values of the estimated characteristics. The difference between the estimates



**Figure 1.** Dependence of the molecular integrals on the interatomic distance: (a)  $(ss|sz)^{CC}$ ; (b)  $(sz|sz)^{CC}$ .

of molecular integrals strongly depends on the interatomic distance. Therefore, it is likely that the discrepancies between the schemes will be larger in the case of molecules having shorter interatomic distances: double bonds and especially triple bonds.

We check these qualitative considerations by actual calculations. Indeed, the most significant disagreement between computed and experimental values are in the case of compounds with multiple bonds. Their heats of formation are too low and the equilibrium bond distances are too small (for example, the calculated heats of formation of acetylene, methylacetylene and dimethylacetylene are lower than their experimental values by 15.6, 13.0, and 10.9 kcal/mol, respectively). It means that the multipole scheme without reparametrization leads to a systematic error for these classes of compounds. It is particularly well seen on the example of molecules with extended  $\pi$ -electron systems (for example, the difference between computed and experimental heats of formation of fulvene, benzene, toluene and naphthalene are 20.1, 8.4, 13.8, and 22.8 kcal/mol, respectively). These compounds have several pairs of atoms with relatively short interatomic distances and thus the error is large. Of course, the problems of the original semiempirical SLG/SCF methods are not solved by introducing the multipole scheme. Therefore, there are relatively large discrepancies between the computed and experimental heats of formation for branched organic molecules (this problem is inherited from the SCF-based semiempirical methods).

The situation calls for some reparametrization of the Hamiltonian. To do so, we, first of all, define a set of molecules to be considered. Of course, the average quality of the results will strongly depend on the level of representation of classes with a systematic error in the training set. To avoid the excessive arbitrariness, we consider the set of molecules which we used previously in ref 16. We performed a series of calculations using the SLG/SCF method with multipoles for the test molecules using the MNDO Hamiltonian and fully optimized geometries. Our calculations show that the average absolute error in the heats of formation is 8.7 kcal/mol. This value can be compared with those obtained by the SCF and SLG/SCF methods for the same set of molecules, which are 6.6 and 5.7 kcal/mol, respectively. It means that the error induced by the multipole scheme is not too large, and the method can be used in many cases without any reparametrization.

At the same time the quality of the results can be improved. While going from the SCF wave function to the SLG and SLG/SCF ones,<sup>13,14,16</sup> we performed only a partial reparametrization

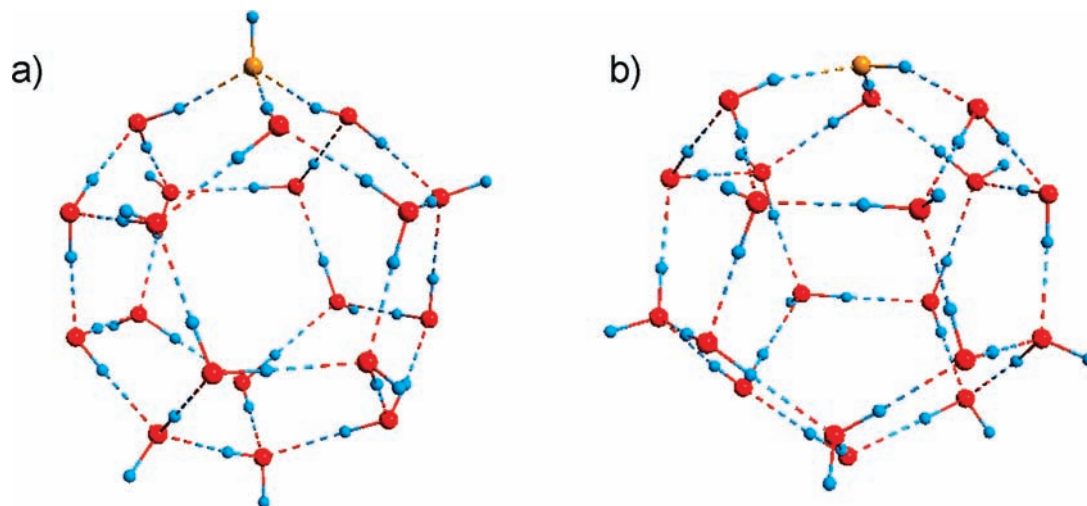
**TABLE 1: Resonance Parameters (eV) for the SCF, Original SLG/SCF, and SLG/SCF with Multipoles Methods with MNDO Parametrization**

atom	param	SCF	SLG/SCF	SLG/SCF(m)
C	$\beta_s$	18.985044	17.136	17.133
	$\beta_p$	7.934122	9.345	9.343
	$\beta_\pi$	7.934122	9.92	10.306
N	$\beta_s$	20.495758	20.594	20.479
	$\beta_p$	20.495758	20.594	20.479
	$\beta_\pi$	20.495758	21.97	23.802
O	$\beta_s$	32.688082	32.784	33.001
	$\beta_p$	32.688082	32.784	33.001
F	$\beta_s$	48.290466	49.591	50.072
	$\beta_p$	36.508540	36.594	37.066

by changing the values of the resonance (electron hopping) parameters. We applied the same strategy here. Table 1 shows the optimal values of the resonance parameters for 2s- and 2p-orbitals in geminals as well as the parameters for  $\pi$ -electron groups in comparison with the same parameters used previously in the SCF and SLG/SCF methods without multipoles. Using these parameters the average absolute error in the heats of formation decreases to 6.7 kcal/mol (the systematic error is small, about 0.3 kcal/mol) resulting in the method's accuracy similar to that of the original SCF method. The same is true for the equilibrium geometries. Indeed, the difference between computed and experimental heats of formation of the above-mentioned fulvene, benzene, toluene and naphthalene become much smaller (7.4, 5.2, 0.2, and 0.1 kcal/mol, respectively).

We tried to improve the accuracy further by varying a wider set of the parameters including the parameters  $\alpha$  of the specific core-core interactions and the parameters  $\rho$  of the interactions between the multipoles. The result of these attempts is that the improvements are not significant and that it is probably more reliable to limit the reparametrization to the resonance parameters. It also means that the further attempts to find a better parametrization should take into consideration the possibility of redefining the functional form of the molecular integrals. It is also necessary to clarify the sources for systematic errors in some classes of compounds (like branched ones) and to try to remove them. In this context it is possible that further improvements of the proposed method will come not from finding better numerical values of its existing parameters rather from accounting of the omitted interactions between the electron groups as proposed in ref 27 and thorough analysis of the functional form of the parametrized matrix elements as proposed in ref 28.

To test capacities of the method proposed for polyatomic systems, we considered water clusters with defects. First we



**Figure 2.** Clusters  $\text{FH}(\text{H}_2\text{O})_{19}$ : (a) with dangling bond  $\text{F}-\text{H}$ ; (b) with hydrogen-bonded  $\text{F}-\text{H}$ .

notice that hydrogen-bonded systems are natural objects for the SLG/SCF(m) method with chemical bonds and lone pairs described by geminals and hydrogen bonds described by Slater determinants. Previously we applied the SLG/SCF method to analysis of water clusters.<sup>29</sup> The complexity of the problem is due to the large number of possible configurations of hydrogen bonds satisfying the so-called "ice rules",<sup>30</sup> which basically mean that water molecules are not ionized and each of them can be a donor of no more than 2 lone pairs.

We employed the PM3 parametrization<sup>31</sup> of the NDDO Hamiltonian, which has been successfully used to reproduce H-bonds in water clusters.<sup>32,33</sup> This success is probably due to the specific core-core energy  $E_{\text{spec}}^{\text{AB}}$  of eq 12. At the same time, it can also lead to problems: the potential energy curve for the water dimer as obtained by PM3/SCF has a shoulder at large intermolecular distances.<sup>34</sup> This is unphysical but may be the price to the correct qualitative spatial structure of the optimized water dimer: the structures produced by MNDO and AM1 are too far from the experimental one.<sup>31</sup>

The necessity of reparametrization for hydrogen bonds was satisfied by setting the resonance parameter  $\beta_{\text{H}}$  to be 6.46 eV. The SLG/SCF(m) scheme predicts the optimal bond length  $r(\text{O}-\text{H}) = 0.954 \text{ \AA}$  and the optimal bond angle  $\angle(\text{H}-\text{O}-\text{H}) = 106.9^\circ$  for water molecule, to be compared with experiment ( $r(\text{O}-\text{H}) = 0.957 \text{ \AA}$  and  $\angle(\text{H}-\text{O}-\text{H}) = 104.5^\circ$ ) and PM3/SCF ( $r(\text{O}-\text{H}) = 0.951 \text{ \AA}$  and  $\angle(\text{H}-\text{O}-\text{H}) = 107.7^\circ$ <sup>31</sup>). The SLG estimate for the heat of formation is  $-60.85 \text{ kcal/mol}$ , which is also closer to the experimental value  $-57.8 \text{ kcal/mol}$  than the PM3/SCF estimate of  $-53.4 \text{ kcal/mol}$ . The bonding energy of the dimer is about  $-5 \text{ kcal/mol}$ , perfectly reproduced by SLG ( $-5.03 \text{ kcal/mol}$ ), while the original PM3/SCF scheme underestimates it ( $-3.5 \text{ kcal/mol}$ ). The most dramatic difference is in the length of the hydrogen bond in the dimer: the SLG method somewhat overestimates it ( $3.05 \text{ \AA}$  instead of  $2.98 \text{ \AA}$ <sup>35</sup>), while the PM3/SCF value ( $2.769 \text{ \AA}$ ) is much too small. At the same time, one should also take into account that water dimer is a floppy complex<sup>36</sup> and ab initio calculations predict the equilibrium length of the hydrogen bond to be about  $2.91 \text{ \AA}$  (see ref 37 for the comparison of the results for different methods and the discussion of the reasons why this value is significantly smaller than the experimental one).

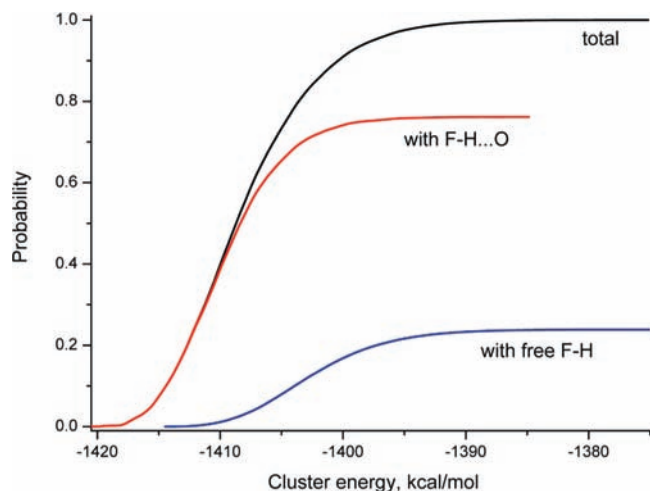
The chemistry of water clusters is an important component of the atmospheric chemistry and it can serve as a model for hydration processes in bulk water. Most theoretical studies are devoted to the incorporation of hydrohalic acids into water

clusters.<sup>38-42</sup> It is possible to consider different sizes and morphologies of water clusters for substitution of water on a hydrogen halide. The reported studies are limited by small clusters with up to seven water molecules and the influence of H-bond networks on the properties of these clusters was studied only in ref 41. The computational efficiency of the SLG/SCF(m) method allows one to consider significantly larger models. Water cluster  $(\text{H}_2\text{O})_{20}$  with dodecahedral structure attracts special attention due to enhanced stability of this structural pattern.<sup>29,43-45</sup> Here we consider this cluster with one water molecule replaced by a HF molecule.

The SLG/SCF(m) method requires special parametrization for the electron groups describing H-bonds with the F atom. We again modify only the resonance contribution by setting the parameter  $\beta_{\text{p}}(\text{F}) = 28.2 \text{ eV}$ . The other parameters are taken from refs 14 and 31. With these parameters the energy of HF is  $-68.26 \text{ kcal/mol}$ . It is important to state that the small reparametrization of the resonance parameters does not significantly affect the optimized spatial structures. The results produced with the PM3 versions of SCF and SLG/SCF(m) methods are of similar quality. Turning to the H-bonded systems, the SLG/SCF(m) method correctly predicts that the isomer of  $\text{HF}(\text{H}_2\text{O})$  with the bond  $\text{F}-\text{H}\cdots\text{O}$  is more favorable than the isomer with the bond  $\text{O}-\text{H}\cdots\text{F}$ . The computed binding energies for  $(\text{HF})_2$  and  $\text{HF}(\text{H}_2\text{O})$  are  $-9.7$  and  $-7.6 \text{ kcal/mol}$ , to be compared with the experimental estimates of  $-6.6$  and  $-8.2 \text{ kcal/mol}$ , respectively. The SLG/SCF(m) distance  $r(\text{FF})$  in  $(\text{HF})_2$  is  $2.69 \text{ \AA}$ , slightly smaller than the experimental value of  $2.72 \text{ \AA}$ . At the same time, the distance  $r(\text{OF})$  in  $\text{HF}(\text{H}_2\text{O})$  is overestimated ( $2.76 \text{ \AA}$  instead of  $2.66 \text{ \AA}$ ). This is not surprising due to inherent deficiencies of the NDDO methods and the floppiness of the complex reflected in the computations: the difference in  $r(\text{O}\cdots\text{H})$  computed by B3LYP/D95++(p,d) and HF/D95++(p,d) methods is  $0.13 \text{ \AA}$ .<sup>40</sup>

The number of symmetry-independent structures for clusters  $(\text{H}_2\text{O})_{20}$  with the morphology of a pentagonal dodecahedron is 30 026. When one water molecule is replaced by HF the total number of symmetry-independent configurations becomes 394 000. We computed all of them. The main structural question for these clusters is the position of the  $\text{F}-\text{H}$  bond: it is either a dangling bond or a part of a hydrogen bond  $\text{F}-\text{H}\cdots\text{O}$ . There are 94 000 structures of the first type and 300 000 structures of the second type. The most favorable configurations of both types are shown in Figure 2. These configurations have the energies of  $-1414.98$  and  $-1420.67 \text{ kcal/mol}$ , respectively. The con-





**Figure 3.** Cumulative distribution function for the energy of the H-bond isomers and its decomposition into the sum of contributions from configurations with dangling bond F–H and hydrogen-bonded F–H.

configurations with involvement of F–H into hydrogen bonding are generally more preferable. Figure 3 shows the cumulative distribution function for the energies of all configurations as well as its decomposition into the sum of distributions for the two types of the structures. The entire distribution is characterized by the mean value of  $-1407.93$  kcal/mol and the mean square deviation  $\sigma = 5.57$  kcal/mol. The distribution for configurations with a dangling F–H bond has a mean value  $-1402.32$  kcal/mol and  $\sigma = 5.27$  kcal/mol, while the narrower distribution for configurations with hydrogen-bonded F–H has a mean value  $-1409.68$  kcal/mol and  $\sigma = 4.37$  kcal/mol.

One of the main questions studied in the literature is the ability of water clusters to dissociate hydrogen halides. It is agreed that even small water clusters make the ionized structures energetically favorable for HCl, HBr, and HI<sup>38,39,42</sup> (rather strong acids). In the case of HF the situation is not so clear: most studies predict that the dissociated structures are less stable than the neutral ones for the clusters studied,<sup>40,42,46</sup> but it was also reported that in the case of cubic morphology of the cluster seven water molecules are capable of ionizing the HF molecule.<sup>41</sup> We generated all 1 466 572 symmetry-independent structures for a dodecahedral ionized cluster  $\text{HF}(\text{H}_2\text{O})_{19}$  and computed their energies by the SLG/SCF(m) method. In the case of water clusters we were able to find local minima corresponding to each of the possible H-bond topologies. In the case of HF substitution, the situation is different: about 37.8% generated ionized structures converged to neutral structures (of course it is mostly the case for the structures with small distances between the generated formal charges). Among the rest, the energy of the most favorable configuration is  $-1379.02$  kcal/mol, i.e., significantly higher than that of the most favorable neutral configuration, although the water cluster makes the dissociation significantly less costly in energy. Unfortunately, it is not possible to make a confident conclusion that the cluster does not dissociate itself because the errors in the original PM3 parametrization are too high: for example, the PM3 heat of formation of  $\text{F}^-$  ion is 30 kcal/mol higher than the experimental value<sup>31</sup> and our modified treatment of two-center contributions does not address this problem. Our calculations show that the energies depend on the separation between charges. The average heats of formation for structures with  $\text{H}_3\text{O}^+$  being separated by 1, 2, 3, 4, or 5 bonds from  $\text{F}^-$  are  $-1364.30$ ,  $-1344.18$ ,  $-1330.54$ ,  $-1324.64$ , and  $-1322.46$  kcal/mol, respectively,

demonstrating that the energy increases with the charge separation distance.

#### 4. Conclusions

In the present paper we performed two important steps to the development of ultrafast computational quantum chemical methods as compared to the previous developments.<sup>14,17</sup> First, we extended the multipole scheme<sup>17</sup> for the interatomic interactions to the short-range ones including those between bonded atoms. It made the corresponding methods more consistent because it means that there is no need to have two different schemes for the short-range and long-range two-electron matrix elements of the Coulomb interactions. Second, the method allows for different electron groups: not only geminals describing chemical bonds and lone pairs but also delocalized electron groups to be treated by various versions of the SCF approximation. We were able to formulate the methods in the way allowing for very weak dependence of the required computation time on the system's size (linear with a very small coefficient). It allows the methods developed here to compete even with classical molecular mechanics (force field) schemes in terms of the required computational resources. These features are useful in the context of constructing hybrid quantum/classical schemes because the quantum mechanical nature of the method proposed makes it possible to derive explicit expressions for the junction between subsystems on the basis of the perturbation theory bringing certainty and order to this diverse area. Alternatively, one can try to use this method to derive classical force fields following the route proposed in refs 23 and 24.

The proposed SLG/SCF method with multipoles nevertheless requires reparametrization of the original semiempirical Hamiltonian. It turns out that despite significant changes in the potentials for short interatomic distances, the method can still be made quite accurate by reparametrizing only electron hopping parameters, which constitute only a small subset of the whole parameter set. This parametrization is not ideal and a series of issues should be addressed in the future, including local account of perturbations caused by intergroup interactions. We performed a series of calculations for all symmetry-independent H-bond arrangements of dodecahedral water cluster with HF defect, thus showing that the computational efficiency of the scheme allows one to solve such large-scale problems.

**Acknowledgment.** The generous financial support of this work through the JARA-SIM research project "Local Electron States in Molecules and Solids" is gratefully acknowledged. The work of A.L.T. was partly supported by RFBR grant No. 07-03-01128.

#### References and Notes

- (1) Hoffmann, R.; Shaik, S.; Hiberty, P. C. *Acc. Chem. Res.* **2003**, *36*, 750–756.
- (2) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (3) Wu, S. Y.; Jayanthi, C. S. *Phys. Rep.* **2002**, *358*, 1–74.
- (4) Tchougréeff, A. L.; Tokmachev, A. M. In *Advanced Topics in Theoretical Chemical Physics*; Maruani, J., Lefebvre, R., Brändas, E., Eds.; Kluwer: Dordrecht, The Netherlands, 2003; pp 207–246.
- (5) Ruedenberg, K. *Rev. Mod. Phys.* **1962**, *34*, 326–376.
- (6) Cui, Q.; Guo, H.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 5617–5631.
- (7) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *107*, 425–431.
- (8) York, D. M.; Lee, T. S.; Yang, W. T. *Phys. Rev. Lett.* **1998**, *80*, 5011–5014.
- (9) Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. *J. Chem. Phys.* **2004**, *121*, 1266–1270.
- (10) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133–146.
- (11) McWeeny, R. *Proc. R. Soc. London. A* **1959**, *253*, 242–259.

- (12) Soudackov, A. V.; Tchougréeff, A. L.; Misurkin, I. A. *Theor. Chim. Acta* **1992**, *83*, 389–416.
- (13) Tokmachev, A. M.; Tchougréeff, A. L. *J. Comput. Chem.* **2001**, *22*, 752–764.
- (14) Tokmachev, A. M.; Tchougréeff, A. L. *J. Phys. Chem. A* **2003**, *107*, 358–365.
- (15) Tokmachev, A. M.; Tchougréeff, A. L. *Int. J. Quantum Chem.* **2001**, *85*, 109–117.
- (16) Tokmachev, A. M.; Tchougréeff, A. L. *Int. J. Quantum Chem.* **2006**, *106*, 571–587.
- (17) Tokmachev, A. M.; Tchougréeff, A. L. *J. Phys. Chem. A* **2005**, *109*, 7613–7620.
- (18) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta (Berl.)* **1977**, *46*, 89–104.
- (19) Arai, T. *J. Chem. Phys.* **1960**, *33*, 95–98.
- (20) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907, 4907–4917.
- (21) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (22) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (23) Tchougréeff, A. L.; Tokmachev, A. M. *Int. J. Quantum Chem.* **2004**, *96*, 175–184.
- (24) Tchougréeff, A. L. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 243–263.
- (25) Dewar, M. J. S.; Hojvat (Sabelli), N. L. *J. Chem. Phys.* **1961**, *34*, 1232–1236.
- (26) Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550–4557.
- (27) Tchougréeff, A. L. *J. Struct. Chem.* **2007**, *48*, S32–S54.
- (28) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159–2166.
- (29) Tchougréeff, A. L.; Tokmachev, A. M.; Dronskowski, R. Manuscript to be published.
- (30) Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933**, *1*, 515–548.
- (31) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220, 221–264.
- (32) Rzepa, H. S.; Yi, M. *J. Chem. Soc., Perkins Trans.* **1990**, *2*, 943–951.
- (33) Jurema, M. W.; Shields, G. C. *J. Comput. Chem.* **1993**, *14*, 89–104.
- (34) Csonka, G. I.; Ángyán, J. G. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 31–38.
- (35) Dyke, T. R.; Mack, K. M.; Muentner, J. S. *J. Chem. Phys.* **1977**, *66*, 498–510.
- (36) Tachikawa, H. *Chem. Phys. Lett.* **2003**, *370*, 188–196.
- (37) Klopper, W.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227–2234.
- (38) Lee, C.; Sosa, C.; Planas, M.; Novoa, J. J. *J. Chem. Phys.* **1996**, *104*, 7081–7085.
- (39) Re, S.; Osamura, Y.; Suzuki, Y.; Schaefer, H. F., III. *J. Chem. Phys.* **1998**, *109*, 973–977.
- (40) Re, S. *J. Phys. Chem. A* **2001**, *105*, 9725–9735.
- (41) Kuo, J.-L.; Klein, M. L. *J. Chem. Phys.* **2004**, *120*, 4690–4695.
- (42) Odde, S.; Mhin, B. J.; Lee, S.; Lee, H. M.; Kim, K. S. *J. Chem. Phys.* **2004**, *120*, 9524–9535.
- (43) Yang, X.; Castleman, A. W., Jr. *J. Am. Chem. Soc.* **1989**, *111*, 6846–6847.
- (44) Wu, C.-C.; Lin, C.-K.; Chang, H.-C.; Jiang, J.-C.; Kuo, J.-L.; Klein, M. L. *J. Chem. Phys.* **2005**, *122*, 074315.
- (45) Anick, D. J. *J. Phys. Chem. A* **2006**, *110*, 5135–5143.
- (46) Smith, A.; Vincent, M. A.; Hillier, I. H. *J. Phys. Chem. A* **1999**, *103*, 1132–1139.

JP905744Z