

SEST: Simulated Electronic Structure Theory

Joshua W. Hollett* and Raymond A. Poirier

Department of Chemistry, Memorial University of Newfoundland, St. John's,
Newfoundland A1B 3X7, Canada

Received August 11, 2008

Abstract: A novel approach to empirically modeling the electronic structure of molecules is introduced. The theory is based on relationships between molecular orbital energy components and the average distance between electrons and electrons and nuclei. The electron–electron and electron–nucleus distances are subsequently related to interatomic distances which provides a means for modeling the electronic structure of molecules. The general energy expression for a simulated electronic structure theory is defined, along with the functional form of the interatomic distance dependent energy functions. The theory is used to model the hydrogen molecule, the first-row hydrides, and ethane. The models, which have the correct RHF/6–31G(d) optimized geometries, also fit the RHF/6–31G(d) energy at equilibrium and the UHF/6–31G(d) energy at the bond dissociation limit as well as some vibrational frequencies.

1. Introduction

Molecular mechanics methods are widely used to study problems in a variety of research fields, from nanotechnology¹ to molecular biology.² In conventional molecular mechanics the total energy of a molecule is divided into different components; bond stretching, E_s , angle bending, E_b , torsional interactions, E_{tor} , van der Waals interactions, E_{vdW} , and electrostatic interactions, E_{elec} .³

$$E_{\text{tot}}(R) = E_s + E_b + E_{\text{tor}} + E_{\text{vdW}} + E_{\text{elec}} \quad (1)$$

Some formulations may use other energy terms to describe phenomena, such as hydrogen bonding, that are not described well by the existing terms. The functional form of these energy terms and the empirical parameters that occur in the functions are known as a force field. Given the functional form of a molecular mechanics force field, the empirical parameters are defined through fitting of experimental or *ab initio* data.

The major advantage of molecular mechanics methods over *ab initio*, density functional, or semiempirical methods is their computational efficiency. Molecular mechanics allows for the study of large systems and dynamics, where it is unfeasible or strictly impossible to use other approaches. The flexibility of force fields which may be tuned to various classes of compounds allows for quite accurate calculations.

However, on the other hand, the specificity of the force fields makes studies on unknown compounds inaccurate and unpredictable. This lack of transferability of specific atom–atom interactions is due to contamination by the molecular environment, which is, indirectly, a result of neglecting the electronic structure. The absence of electrons and orbitals also poses problems for chemical calculations. It is not possible to study reaction mechanisms with molecular mechanics or many other properties related to the electronic structure, such as magnetic properties, excitations, electron transport, and electrical conductivity. A partial solution to such a problem is through a combined quantum mechanics and molecular mechanics (QM/MM) approach.⁴

A hybrid QM/MM approach is common in the study of reactions involving large systems, such as enzyme catalysis or organic reactions in solution. The general scheme behind a QM/MM calculation is to treat the region of interest, or importance, with a quantum mechanical calculation, and the rest of the system is treated with molecular mechanics. In some applications the solvent may be treated with molecular mechanics, while the reacting molecules are treated quantum mechanically, or for large molecules, like enzymes, the active site may be QM, while the rest of the molecule is MM. The Hamiltonians, and energy, of these hybrid systems are divided into QM, MM, and QM/MM parts. How the two different regions interact, QM/MM, is defined in several ways.⁵

* Corresponding author phone: (709)737-8752; fax: (709)737-3702; e-mail: jhollett@mun.ca.

Since the initial appearance of molecular mechanics almost a century ago,⁶ most of the development of the theory has been focused on parametrization (i.e., developing force fields for different classes of compounds), while the general formulation of molecular mechanics has remained relatively the same. This has led to accurate modeling of a wide variety of large systems in many research fields. However, the problems associated with the absence of electronic structure are still present. While methods such as QM/MM provide a partial solution to such a problem, there may be other approaches. This study explores such an idea. Rather than the molecular mechanics approach to molecular modeling, modeling of electronic structure is performed. The theory presented provides a different approach to modeling potential energy surfaces, where the energy components associated with electron pairs and individual electrons are functions of the nuclear coordinates.

2. Theory

2.1. Simulated Electronic Structure Theory (SEST). A theory which includes electrons explicitly will have an energy expression that differs significantly from conventional molecular mechanics (eq 1) and more closely resembles the electronic Hamiltonian (eq 2, in atomic units)

$$\hat{H}_e = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (2)$$

where N is the number of electrons, and M is the number of nuclei. The SEST energy expression can be expressed as a sum of atomic contributions, V_{atomic} , and contributions from atom-atom, $V_{\text{atom-atom}}$, atom-lone pair, $V_{\text{atom-lone pair}}$, and lone pair-lone pair, $V_{\text{lone pair-lone pair}}$, interactions.

$$V(R) = V_{\text{atomic}} + V_{\text{atom-atom}} + V_{\text{atom-lone pair}} + V_{\text{lone pair-lone pair}} \quad (3)$$

The electronic kinetic energy is included through the virial theorem (see discussion at the end of this section). Each energy term is further subdivided into electron–electron, electron–nuclear, and nuclear–nuclear potential energy, $V = V^{ee} + V^{Ne} + V^{NN}$.

The atomic energy, which does not depend on the nuclear coordinates, accounts for the energy associated with atoms and their own electrons. Each atom has its own electrons corresponding to the neutral atom. The electronic configuration of each atom follows a Lewis dot structure approach. For example, a sp^3 carbon atom has four unpaired, valence electrons, and a pair of core electrons. Also, a sp^3 nitrogen atom has three unpaired valence electrons, a lone pair, and a pair of core electrons. The atomic energy term is calculated as a sum over the contributions of individual nuclei, $V_{\text{atomic}} = \sum_{A=1}^M V_A$, where M is the total number of nuclei

$$V_A = \sum_{a \in A} 2V_{Aa}^{Ne} + \sum_{i \in A} V_{Ai}^{Ne} + \sum_{a \in A} \left\{ \sum_{\substack{b \in A \\ b \leq a}} (2 - \delta_{ab}) V_{ab}^{ee} + \sum_{i \in A} V_{ai}^{ee} \right\} + \sum_{i \in A} \sum_{\substack{j \in A \\ j < i}} \frac{1}{2} V_{ij}^{ee} + \sum_{\mu \in A} V_{\mu\mu}^{ee} \quad (4)$$

where a and b are pairs of core electrons, i and j are valence electrons, and μ is a lone pair. Each atom has nuclear-electron

potential energy associated with its nucleus, A , and its electrons, V_{Aa}^{Ne} and V_{Ai}^{Ne} . Each atom also has electron–electron potential energy due to its electrons, V_{ab}^{ee} , V_{ai}^{ee} , and V_{ij}^{ee} as well as the potential energy of the electrons in each of its lone pairs, $V_{\mu\mu}^{ee}$. The potential energy, $V_{\mu\mu}^{ee}$, due to each lone pair is included in the atomic energy because, like the atomic energy, it is distance independent.

The distance dependent energy terms include $V_{\text{atom-atom}}$, $V_{\text{atom-lone pair}}$, and $V_{\text{lone pair-lone pair}}$. The atom-atom interaction energy is the sum of the interactions between each atom pair, $V_{\text{atom-atom}} = \sum_{A=1}^M \sum_{B < A} V_{AB}(R_{AB})$

$$V_{AB}(R_{AB}) = \frac{Z_A Z_B}{R_{AB}} + \sum_{a \in A} 2V_{Ba}^{Ne}(R_{AB}) + \sum_{a \in B} 2V_{Aa}^{Ne}(R_{AB}) + \sum_{i \in A} V_{Bi}^{Ne}(R_{AB}) + \sum_{i \in B} V_{Ai}^{Ne}(R_{AB}) + \sum_{a \in A} \left\{ \sum_{b \in B} 2V_{ab}^{ee}(R_{AB}) + \sum_{i \in B} V_{ai}^{ee}(R_{AB}) \right\} + \sum_{b \in B} \sum_{i \in A} V_{bi}^{ee}(R_{AB}) + \sum_{\substack{i \in A \\ (i \in c)}} \left\{ \sum_{\substack{j \in B \\ (j \in c)}} \frac{1}{2} V_{ij}^{ee}(R_{AB}) + \sum_{\substack{j \in B \\ (j \in c)}} V_{ij}^{ee}(R_{AB}) \right\} \quad (5)$$

where c denotes a bond between atoms A and B . This interatomic energy includes core electron pairs and valence electrons of one atom attracted to another, V_{Ba}^{Ne} , V_{Aa}^{Ne} , V_{Bi}^{Ne} , and V_{Ai}^{Ne} , and potential energy between electrons of different atoms, V_{ab}^{ee} , V_{ai}^{ee} , V_{bi}^{ee} , and V_{ij}^{ee} . The atom-lone pair interaction energy is given by $V_{\text{atom-lone pair}} = \sum_{A=1}^M \sum_{\mu=1}^L V_{A\mu}(R_{A\mu})$, where L is the total number of lone pairs, and

$$V_{A\mu}(R_{A\mu}) = 2V_{A\mu}^{Ne}(R_{A\mu}) + \sum_{a \in A} 2V_{a\mu}^{ee}(R_{A\mu}) + \sum_{i \in A} V_{i\mu}^{ee}(R_{A\mu}) \quad (6)$$

There is nuclear attraction potential energy between atom A and lone pair μ , $V_{A\mu}^{Ne}$, and electron–electron potential energy, $V_{a\mu}^{ee}$ and $V_{i\mu}^{ee}$. The lone pair–lone pair interaction energy consists of only electron–electron potential energy, where $V_{\text{lone pair-lone pair}} = \sum_{\mu=1}^L \sum_{\nu < \mu} V_{\mu\nu}$.

$$V_{\mu\nu} = 2V_{\mu\nu}^{ee}(R_{\mu\nu}) \quad (7)$$

This formulation of simulated electronic structure theory contains no electronic kinetic energy terms. It was found that kinetic energy terms were not required to accurately model the systems studied thus far. However, kinetic energy, like electron-nucleus potential energy (Section 2.2), correlates quite well with the average distance between an electron and a nucleus. Therefore, it is possible to include terms to model kinetic energy in future development of the theory, if desired. In order to calculate the energy of a molecular system, $E(R)$, without the kinetic energy of the electrons, it is assumed that the virial coefficient, which is the negative ratio of the potential and kinetic energy, $c_v = -V/T$, is constant over the entire potential energy surface.

$$E(R) = \frac{(c_v - 1)}{c_v} V(R) \quad (8)$$

In this study the equilibrium geometries and energies of the molecules were fit to RHF/6–31G(d) values; therefore, the

virial coefficient is set to the value obtained from the equilibrium geometry at RHF/6-31G(d).

2.2. The Average Interparticle Distance and Modeling Molecular Orbital Energy Components. Simulated electronic structure theory models the PESs of molecular systems, in which the energy contribution of electron pairs and individual electrons are functions of the nuclear coordinates. It is useful to obtain parameters for these models as well as insight on how the energy contributions per electron pair and individual electron depend on molecular geometry, from *ab initio* models of electronic structure. Such properties are easily examined at the Hartree–Fock level of theory. The closed-shell Hartree–Fock (RHF) energy is given as

$$E_{\text{HF}} = \sum_{a=1}^{N/2} 2h_a + \sum_{a=1}^{N/2} \sum_{b=1}^{N/2} 2J_{ab} - K_{ab} \quad (9)$$

where N is the number of electrons, h_a is the one-electron energy, J_{ab} is the Coulomb energy, and K_{ab} is the exchange energy, associated with molecular orbitals (MOs) a and b . The one-electron energy, $h_a = T_a + V_a$, consists of kinetic energy

$$T_a = \langle a | -\frac{1}{2} \nabla^2 | a \rangle \quad (10)$$

and nuclear-electron attraction potential energy

$$V_a = \langle a | -\sum_{A=1}^M \frac{Z_A}{r_{1A}} | a \rangle \quad (11)$$

The Coulomb and exchange energy are calculated from two-electron integrals.

$$J_{ab} = \langle ab | ab \rangle \quad (12)$$

$$K_{ab} = \langle ab | ba \rangle \quad (13)$$

In recent work,⁷ it was found that the Coulomb energy could be modeled with a one-electron property of MOs, the average interelectronic distance, δr_{12} . The Coulomb energy, J_{ab} , associated with any two MOs is inversely proportional to the average distance between an electron in MO a and an electron in MO b , $(\delta r_{12})_{ab}$, which is given by

$$(\delta r_{12})_{ab} = \sqrt{\langle ab | r_{12}^2 | ab \rangle} \quad (14)$$

The relationship

$$J_{ab} \approx \frac{\alpha}{(\delta r_{12})_{ab}} \quad (15)$$

has severe deviations for the canonical MOs (CMOs) of HF theory, due to delocalized core MO pairs. However, if localized MOs (LMOs) are used the relationship is followed quite closely, Figure 1.

A similar relationship has been found between the nuclear-electron attraction potential energy, V_a (eq 11), and the average distance between an electron and a nucleus, δr_{aA} , which is calculated analogously to $(\delta r_{12})_{ab}$.⁸

$$\delta r_{aA} = \sqrt{\langle a | r_{1A}^2 | a \rangle} \quad (16)$$

The nuclear-electron potential energy, V_a , may be separated into contributions from individual nuclei, V_{aA} , where

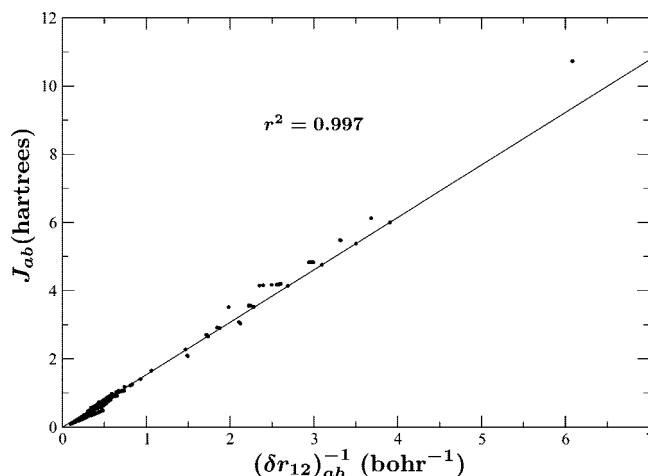


Figure 1. Relationship between Coulomb energy, J_{ab} , and the average distance between two electrons, $(\delta r_{12})_{ab}$ (eq 14), for two electron ions to molecules with up to 58 electrons (RHF/6-31G(d), LMO). From ref 7.

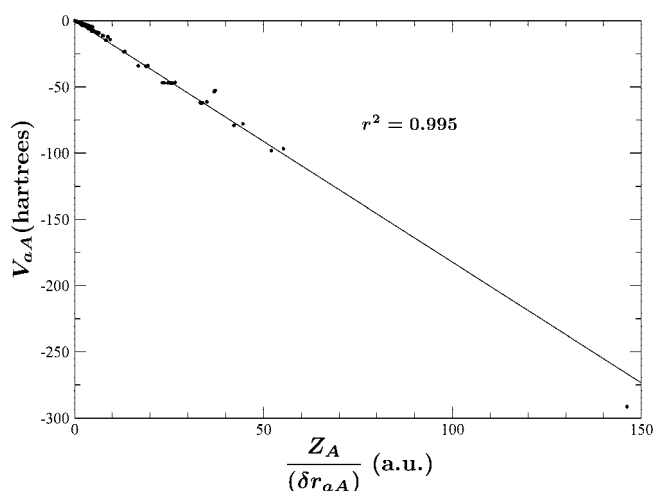


Figure 2. Relationship between electron–nuclear attraction energy, V_{aA} , and the average distance between an electron and a nucleus, δr_{aA} (eq 16), for two electron ions to molecules with up to 58 electrons (RHF/6-31G(d), LMO).

$\sum_{A=1}^M V_{aA} = V_a$. The nuclear-electron potential energy between MO a and nucleus A is related to δr_{aA} through the equation

$$V_{aA} \approx -\gamma \frac{Z_A}{\delta r_{aA}} \quad (17)$$

The relationship has the same deviations for CMOs as seen with J_{ab} and $(\delta r_{12})_{ab}$ and performs quite well for LMOs, Figure 2.

It should be noted that while the overall correlation in Figures 1 and 2 is good, $r^2 = 0.997$ and 0.995 respectively, individual deviations approach 1 hartree for the relationship between J_{ab} and $(\delta r_{12})_{ab}$ and 25 hartrees for V_{aA} and δr_{aA} . These errors are quite large on the scale of reaction energetics. However, these deviations involve core MOs, which correspond to the core electron pairs of SEST, for which the energy terms are constant parameters and therefore do not depend upon the above relationships. Furthermore, it is the qualitative result, which is most relevant to an empirically parametrized SEST. The Coulomb and electron–

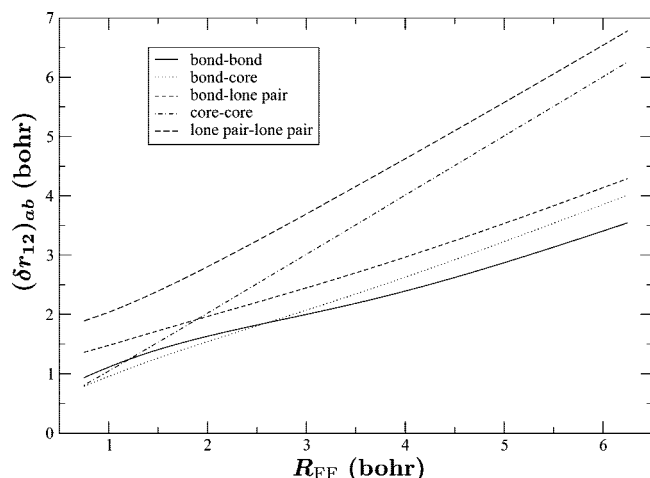


Figure 3. Relationship between the average distance between two electrons, $(\delta r_{12})_{ab}$, and the fluorine-fluorine distance, R_{FF} , in F_2 (RHF/6-31G(d), LMO). Includes $a = \text{bond}$ $b = \text{bond}$, $a = \text{bond}$ $b = \text{core}$, $a = \text{core}$ $b = \text{core}$, and $a = \text{lone pair}$ $b = \text{lone pair}$ (from separate nuclei).

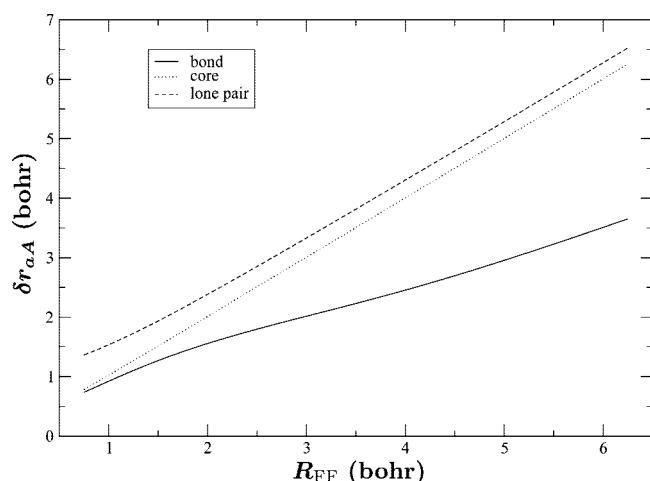


Figure 4. Relationship between the average distance between an electron and a nucleus, δr_{aA} , and the fluorine-fluorine distance, R_{FF} , in F_2 (RHF/6-31G(d), LMO). Includes $a = \text{bond}$, $a = \text{core}$, and $a = \text{lone pair}$.

nuclear potential energy are quantities calculated from integrating over the electronic distribution. Although as a result of the relationships just mentioned, to some approximation electrons have average positions with respect to each other and with respect to the nuclei. Furthermore, these average positions, and hence distances, may be used to approximate components of the HF energy per MO. An immediate question following such deductions is how do these average interparticle distances change with changing molecular structure.

In general, the average interparticle distances, $(\delta r_{12})_{ab}$ and δr_{aA} , are linearly related to interatomic distance. This relationship is illustrated in Figures 3 and 4, with F_2 .

It is expected that such a relationship would be most critical for atoms which are bonded to each other. Changing bond length significantly affects the electronic structure of a molecule, especially around the two bonded atoms, and even more for a diatomic. While this may be the case, it is seen in Figures 3 and 4 that the interparticle distances very

closely follow a linear relationship with the fluorine-fluorine distance. However, for shorter bond lengths (< 1.5 bohr) there is slight curvature in the lone pair-lone pair $(\delta r_{12})_{ab}$ and lone pair δr_{aA} curves which is due to the lone pairs on different fluorine atoms bending away from each other as the bond length decreases. It is also observed in both figures that distances involving electrons in the bond show slight curvature. In both the bond-bond $(\delta r_{12})_{ab}$ and bond δr_{aA} curves, the slope is less around the equilibrium bond length $((R_{FF})_e = 2.54$ bohr) than that at the shorter and longer R_{FF} distances. Nonetheless, this relationship, even as an approximation, is a powerful tool.

The ability to relate the average interparticle distances to atomic distances has important consequences. For a given molecular structure, not only will these relationships provide an estimate of MO energy components but also they will provide a means of predicting an *average* picture of the electronic structure.

2.3. Functional Form. It was seen in Section 2.1 that the SEST energy expression, eq 3, contains terms which depend on atom-atom, atom-lone pair, and lone pair-lone pair distances. The functional form of these distance dependent terms determines the success of applying SEST. Following the relationships observed in Figures 3 and 4, the interparticle distances, $(\delta r_{12})_{ab}$ and δr_{bA} , are taken to be linear functions of the interatomic distance, R_{AB}

$$(\delta r_{12})_{ab} = \lambda R_{AB} + \sigma \quad (18)$$

and

$$\delta r_{bA} = \omega R_{AB} + \kappa \quad (19)$$

where MO a is localized on atom A , MO b is localized on atom B , or one of them (a or b) is the bond between A and B , and λ , σ , ω , and κ are constants for each unique pair of MOs. As observed in Figures 3 and 4, the relationships involving bonding MOs are approximate. However, an exact model of *ab initio* electronic structure through interatomic distances is not the goal, neither is it possible.

The LMO potential energy components are related to the interatomic distance through the linear relationships above and the relationships seen earlier (Figures 1 and 2). Substitution of eq 18 into eq 15 yields

$$J_{ab} \approx \frac{\alpha}{\lambda R_{AB} + \sigma} \quad (20)$$

While, substitution of eq 19 into eq 17 gives

$$V_{bA} \approx \frac{-\gamma Z_A}{\omega R_{AB} + \kappa} \quad (21)$$

The above equations are used to derive the general form of the SEST distance dependent functions.

Given electron pairs a and b , where a is localized on atom A and b is localized on atom B , the electron-electron potential energy between the two pairs of electrons is given by

$$V_{ab}^{ee}(R_{AB}) = \frac{(V_{ab}^{ee})_e((R_{AB})_e - r_{ab}^{ee})}{R_{AB} - r_{ab}^{ee}} \quad (22)$$

where $(R_{AB})_e$ is an equilibrium interatomic distance, r_{ab}^{ee} is a constant parameter, and $(V_{ab}^{ee})_e$ is the potential energy at the equilibrium distance. Also, as the interatomic distance, R_{AB} , approaches infinity the potential energy term vanishes. The SEST electron–electron potential energy function is chosen to model the total HF electron–electron potential energy (eq 9)

$$V_{ab}^{ee}(R_{AB}) \approx 2J_{ab} - K_{ab} \quad (23)$$

although it is only the Coulomb energy, J_{ab} , which is inversely proportional to the average interelectronic distance, $(\delta r_{12})_{ab}$ (eq 15). The functional form is derived using localized molecular orbitals (LMOs),¹⁵ in which case the exchange energy, K_{ab} , is minimal. Therefore the approximation in eq 23 is considered valid. The electron–nucleus potential energy is calculated in a similar fashion.

$$V_{Ba}^{Ne}(R_{AB}) = \frac{(V_{Ba}^{Ne})_e((R_{AB})_e - r_{Ba}^{Ne})}{R_{AB} - r_{Ba}^{Ne}} \quad (24)$$

While the functions given in eqs 22 and 24 are for pairs of electrons, a and b , functions for individual electrons, i and j , and any combination of pairs and individual electrons (e.g. $V_{ai}^{ee}(R_{AB})$) take the same form, albeit with different parameter values. The parameters of these functions are derived from the LMOs of the *ab initio* electronic structure of a molecule.

It is recognized that positive values of the parameters r_{ab}^{ee} , r_{Ba}^{Ne} , etc., result in singularities in the potential energy surface for positive values of R_{AB} . While fitting a potential energy surface, parameters with values larger than $1/2(R_{AB})_e$ are avoided, and values smaller than this threshold are dealt with using a piecewise function.

$$V_{AB}(R_{AB}) = \begin{cases} V^{NN}(R_{AB}) + V^{Ne}(R_{AB}) + V^{ee}(R_{AB}), & \text{if } R_{AB} > r_{AB}^{\max} + \delta \\ V^{NN}(R_{AB}) + V^{Ne}(r_{AB}^{\max} + \delta) + V^{ee}(r_{AB}^{\max} + \delta), & \text{if } R_{AB} \leq r_{AB}^{\max} + \delta \end{cases} \quad (25)$$

$$V_{A\mu}(R_{A\mu}) = \begin{cases} V^{Ne}(R_{A\mu}) + V^{ee}(R_{A\mu}), & \text{if } R_{A\mu} > r_{A\mu}^{\max} + \delta \\ V^{Ne}(r_{A\mu}^{\max} + \delta) + V^{ee}(r_{A\mu}^{\max} + \delta) + V^{ee}(R_{A\mu} + r_i) - V^{ee}(r_{A\mu}^{\max} + \delta + r_i), & \text{if } R_{A\mu} \leq r_{A\mu}^{\max} + \delta \end{cases} \quad (26)$$

The largest positive value for a given atom–atom interaction or atom–lone pair interaction is r^{\max} (e.g., $r_{AB}^{\max} = \max\{r_{Ai}^{Ne}, r_{Ba}^{Ne}, r_{ij}^{ee}, \dots\}$). The argument $R_{A\mu} + r_i$ (eq 26) denotes that for each given function of the form in eqs 22 and 24 the value of the specific r_i parameter is added to $R_{A\mu}$, thus removing r_i from the denominator.

2.4. Determining the Parameters. In this study the parameters of the distance dependent functions (eqs 22 and 24) along with the constant energy terms (eq 4) are derived from *ab initio* calculations. The LMO energy components, V_{aA} , J_{ab} , and K_{ab} , are used to define the constant energy terms, V_A , and the equilibrium energy terms, $(V_{AB})_e$, $(V_{A\mu})_e$, and $(V_{\mu\nu})_e$. For electron pairs belonging to the same atom (core

electrons), or a lone pair, the energy is taken to be equal to the corresponding LMO energy component. For example

$$(V_{ab}^{ee})_e = 2J_{ab}^{\text{LMO}} - K_{ab}^{\text{LMO}} \quad (27)$$

and

$$V_{Aa}^{Ne} = V_{aA}^{\text{LMO}} \quad (28)$$

Energy terms for individual electrons, the bonding electrons, require the partitioning of LMO energy components. The energy expression for ammonia may be used to illustrate an energy partitioning scheme.

$$V_{\text{NH}_3} = V_N + 3V_H + 3V_{\text{NH}} + 3V_{\text{HH}} + V_{\text{N}\mu} + 3V_{\text{H}\mu} \quad (29)$$

The lone pair of nitrogen is denoted μ . The terms V_N and V_H are the atomic energies of nitrogen and hydrogen, respectively

$$V_N = 2V_{Na}^{Ne} + 3V_{Ni}^{Ne} + V_{aa}^{ee} + 3V_{ai}^{ee} + 3\left(\frac{1}{2}V_{ij}^{ee}\right) + V_{\mu\mu}^{ee} \quad (30)$$

where a is the pair of core electrons of nitrogen, and i and j denote valence electrons of nitrogen.

$$V_H = V_{Hk}^{Ne} \quad (31)$$

where k is the hydrogen electron. The atom–atom interaction energy terms are V_{NH} and V_{HH}

$$V_{\text{NH}} = \frac{Z_N Z_H}{R_{\text{NH}}} + 2V_{Ha}^{Ne}(R_{\text{NH}}) + 2V_{Hj}^{Ne}(R_{\text{NH}}) + V_{Hi}^{Ne}(R_{\text{NH}}) + V_{Nk}^{Ne}(R_{\text{NH}}) + V_{ak}^{ee}(R_{\text{NH}}) + 2\left(\frac{1}{2}V_{jk}^{ee}(R_{\text{NH}})\right) + V_{ik}^{ee}(R_{\text{NH}}) \quad (32)$$

where a is the pair of core electrons of nitrogen, i is the valence electron of nitrogen in this NH bond, j is a valence electron of nitrogen not in this bond, and k is the hydrogen electron. The hydrogen–hydrogen nonbonded interaction energy is described as

$$V_{\text{HH}} = \frac{Z_H Z_H}{R_{\text{HH}}} + 2V_{Hi}^{Ne}(R_{\text{HH}}) + \frac{1}{2}V_{ij}^{ee}(R_{\text{HH}}) \quad (33)$$

Finally, $V_{\text{N}\mu}$ and $V_{\text{H}\mu}$ are the atom–lone pair interaction terms

$$V_{\text{N}\mu} = 2V_{\text{N}\mu}^{Ne}(R_{\text{N}\mu}) + 2V_{a\mu}^{ee}(R_{\text{N}\mu}) + 3V_{i\mu}^{ee}(R_{\text{N}\mu}) \quad (34)$$

$$V_{\text{H}\mu} = 2V_{\text{H}\mu}^{Ne}(R_{\text{H}\mu}) + V_{k\mu}^{ee}(R_{\text{H}\mu}) \quad (35)$$

The derivation of the parameters for the electron–nucleus and electron–electron potential energy terms from LMO energy values is given in Tables 1 and 2, respectively.

Similar schemes are used to partition the energies of the bonding electrons of other molecules. In some cases there may be a point in which the partition is arbitrary and is defined intuitively (see V_{ai}^{ee} and $(V_{i\mu}^{ee})_e$ in Table 2). Also, the partitioning of the energy for V_{Ni}^{Ne} (Table 1) and V_{di}^{ee} (Table 2) ensures that dissociation of a single bond leads to the UHF/6–31G(d) potential energy of the dissociated products. It is important to note that, for the equilibrium structure, the MO energy components, V_{aA}^{LMO} and V_{ab}^{LMO} , can be recovered exactly. For example

$$V_{cd}^{LMO} = \frac{1}{2}V_{ij}^{ee} + 2\left(\frac{1}{2}V_{ik}^{ee}(R_{NH})\right) + V_{kl}^{ee}(R_{HH}) \quad (36)$$

where i and j denote bonding electrons of nitrogen, and k and l are hydrogen electrons. Furthermore, for a distorted geometry an approximation to the MO energy components may be calculated.

It is not surprising that a theory which parametrizes the energy of individual electrons and electron pairs derives the parameters from *ab initio* LMO calculations. SEST, an alternative approach to modeling PESs, shares the same

Table 1. Derivation of SEST Electron-Nucleus Potential Energy Parameters for Ammonia from RHF/6-31G(d) LMO Energy Components^a

parameter	eq	derived from ^b
V_{Na}^{Ne}	30	V_{aN}^{LMO}
$(V_{Ha}^{Ne})_e$	32	V_{aH}^{LMO}
$(V_{\mu\mu}^{Ne})_e$	34	$V_{\mu\mu}^{LMO}$
$(V_{H\mu}^{Ne})_e$	35	$V_{\mu H}^{LMO}$
$(V_{Hj}^{Ne})_e$	32	$2V_{dH}^{LMO} - V_{Hj}^{Ne}(R_{HH})^c$
V_{Ni}^{Ne}	30	$V_{NH_2}^{Ne} - 2(V_{Na}^{Ne})_e + 2V_{dN}^{LMO} + 2V_{dH}^{LMO} + 2V_{dH}^{LMO} + (V_{\mu\mu}^{Ne})_e + 2(V_{H\mu}^{Ne})_e - 2(V_{Hj}^{Ne})_e^d$
$(V_{Nk}^{Ne})_e$	32	$2V_{dN}^{LMO} - V_{Nj}^{Ne}$
$(V_{Hk}^{Ne})_e$	32	$2V_{dH}^{LMO} - V_{Hk}^{Ne}$

^a μ denotes the lone pair of nitrogen, a denotes the pair of core electrons of nitrogen, c denotes a bond involving the specified atom, d denotes a bond not involving a specified atom (i.e., another bond), i and j denote valence electrons of nitrogen, and k denotes electron of hydrogen. ^b Terms of the form V^{LMO} denote energy values taken from LMOs of an *ab initio* calculation. ^c $V_{Hj}^{Ne}(R_{HH})$ denotes electron-nucleus potential energy of previously defined HH nonbonded interaction, calculated at the equilibrium ammonia HH distance, R_{HH} . ^d $V_{NH_2}^{Ne}$ denotes total electron-nucleus potential energy of the radical amine calculated at UHF/6-31G(d).

Table 2. Derivation of SEST Electron-Electron Potential Energy Parameters for Ammonia from UHF/6-31G(d) LMO Energy Components^a

parameter	eq	derived from ^b
V_{aa}^{ee}	30	V_{aa}^{LMO}
$V_{\mu\mu}^{ee}$	30	$V_{\mu\mu}^{LMO}$
$(V_{\mu\mu}^{ee})_e$	34	$V_{\mu\mu}^{LMO}$
V_{di}^{ee}	intermediate term ^c	$\{V_{NH_2}^{ee} - V_{aa}^{ee} - 2V_{cc}^{LMO} - 2V_{cd}^{LMO} - 4V_{ca}^{LMO} - V_{\mu\mu}^{ee} - 4V_{dH}^{LMO} - 2(V_{Hj}^{ee})_e\}(2 + (V_{ca}^{LMO}/V_{cd}^{LMO}) + (V_{\mu\mu}^{LMO}/V_{cd}^{LMO}))^{-1d}$
V_{ai}^{ee}	30	$V_{di}^{ee}(V_{ca}^{LMO}/V_{cd}^{LMO})$
$(V_{\mu\mu}^{ee})_e$	34	$V_{di}^{ee}(V_{\mu\mu}^{LMO}/V_{cd}^{LMO})$
$(V_{\mu\mu}^{ee})_e$	35	$2V_{\mu\mu}^{LMO} - (V_{\mu\mu}^{ee})_e$
$(V_{\mu\mu}^{ee})_e$	32	$2V_{\mu\mu}^{LMO} - V_{\mu\mu}^{ee}$
V_{dk}^{ee}	intermediate term ^c	$2V_{cd}^{LMO} - V_{di}^{ee}$
$(V_{jk}^{ee})_e$	32	$2V_{dk}^{ee} - V_{Hj}^{ee}(R_{HH})^e$
V_{ij}^{ee}	30	$2V_{ij}^{ee} - (V_{jk}^{ee})_e$
$(V_{jk}^{ee})_e$	32	V_{cc}^{LMO}

^a μ denotes the lone pair of nitrogen, a denotes the pair of core electrons of nitrogen, c denotes a bond involving the specified electron, d denotes a bond not involving the specified electron (i.e., another bond), i and j denote valence electrons of nitrogen, and k denotes electron of hydrogen. ^b Terms of the form V^{LMO} denote energy values taken from LMOs of an *ab initio* calculation (where $V_{ab}^{LMO} = 2J_{ab}^{LMO} - K_{ab}^{LMO}$). ^c Intermediate terms used to simplify expressions for some energy parameters. ^d $V_{NH_2}^{ee}$ denotes total electron-electron potential energy of the radical amine calculated at UHF/6-31G(d). ^e $V_{Hj}^{ee}(R_{HH})$ denotes electron-electron potential energy of previously defined HH nonbonded interaction, calculated at the equilibrium ammonia HH distance, R_{HH} .

Table 3. SEST and RHF/6-31G(d) Vibrational Frequencies for the First-Row Hydrides

system	ν	frequency (cm ⁻¹)	
		RHF/6-31G(d)	SEST
BeH ₂	1	761	540
	2	763	540
	3	2107	2240
	4	2324	2324
	5	2816	2817
BH ₃	1	1225	1085
	2	1305	1218
	3	1306	1218
	4	2694	3033
	5	2816	2817
CH ₄	1	1487	1862
	2	1488	1862
	3	1488	1862
	4	1702	1485
	5	1702	1485
NH ₃ ^a	1	1214	2294
	2	1849	1980
	3	1851	1980
	4	3690	5001
	5	3825	3825
H ₂ O ^a	1	1829	2237
	2	4070	5658
	3	4191	4191
HF ^a	1	4362	4362

^a Molecules with lone pairs have lone pair vibrational modes. These modes are not shown, they are of the order 10⁴–10⁸ cm⁻¹ and easily distinguished from the real modes.

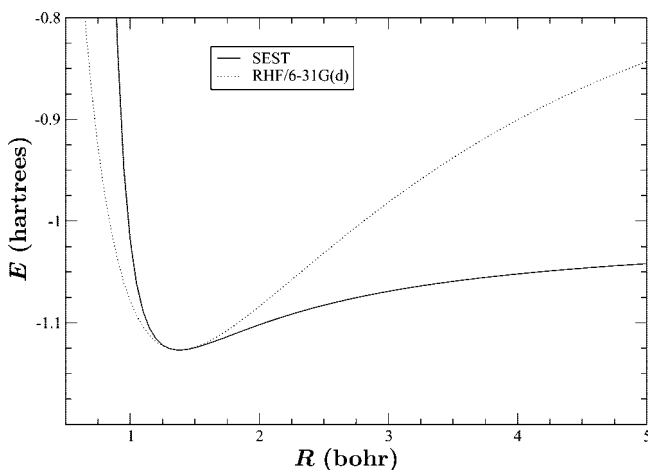


Figure 5. Dissociation curve for H₂ (SEST and RHF/6-31G(d)).

premise of molecular mechanics, which is that certain functions and parameters are transferable from one molecule to another. The transferability of the properties of LMOs has been known for decades,^{9,10} and work is currently being done to apply this transferability to linear scaling, divide, and conquer methods.^{11–13}

The other parameters of the distance dependent functions, $(R_{AB})_e$, r_{ABa}^{Ne} , r_{ABb}^{ee} , etc., are defined through fitting of the equilibrium geometry, $\partial E(R_e)/\partial R = 0$, and the vibrational frequencies.

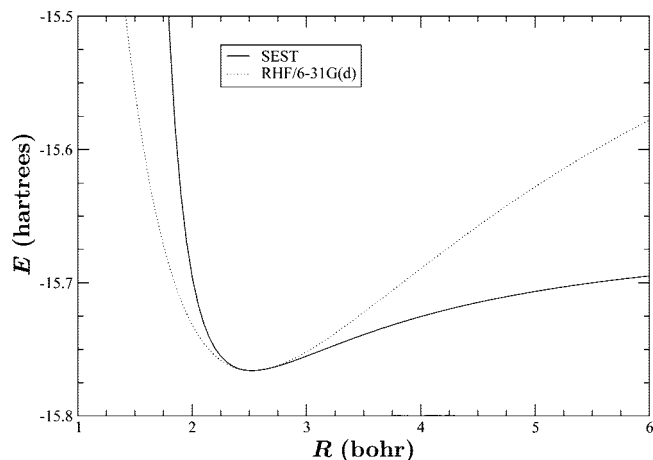


Figure 6. Dissociation curve for the BeH bond of BeH_2 (SEST and RHF/6–31G(d)).

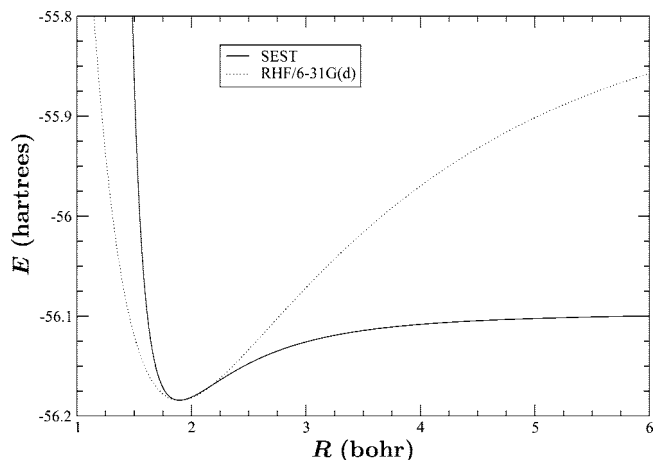


Figure 9. Dissociation curve for the NH bond of NH_3 (SEST and RHF/6–31G(d)).

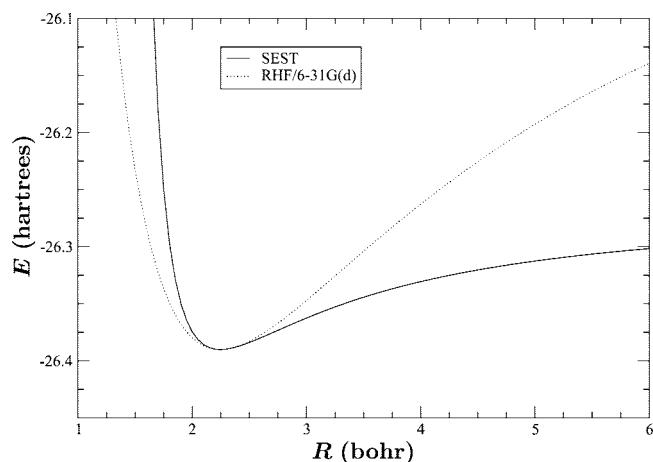


Figure 7. Dissociation curve for the BH bond of BH_3 (SEST and RHF/6–31G(d)).

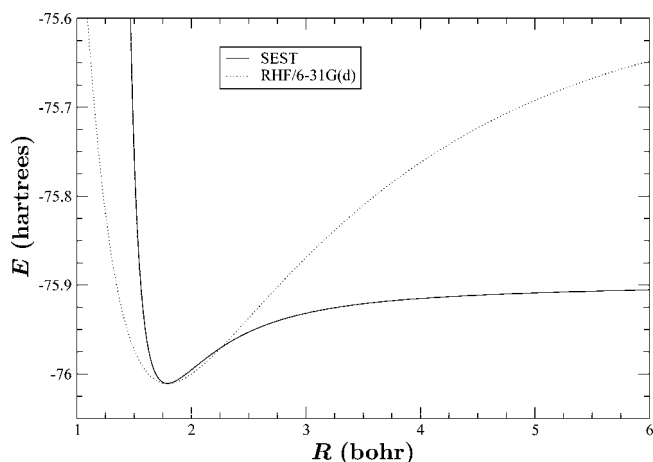


Figure 10. Dissociation curve for the OH bond of H_2O (SEST and RHF/6–31G(d)).

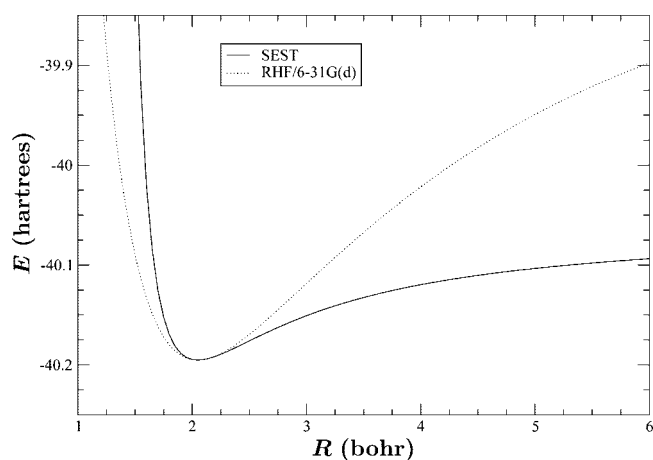


Figure 8. Dissociation curve for the CH bond of CH_4 (SEST and RHF/6–31G(d)).

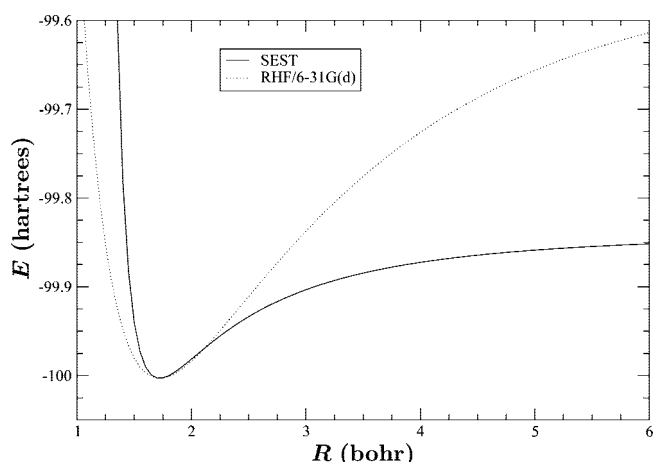


Figure 11. Dissociation curve for HF (SEST, RHF/6–31G(d)).

3. Method

Simulated Electronic Structure Theory (SEST) was implemented within the MUNgauss program.¹⁴ The geometries of the molecules studied were optimized at RHF/6–31G(d),

and the energies and MO properties were also calculated at RHF/6–31G(d). The energies of the products of bond dissociation were calculated at UHF/6–31G(d). Localized molecular orbitals were obtained through Boys localization.¹⁵

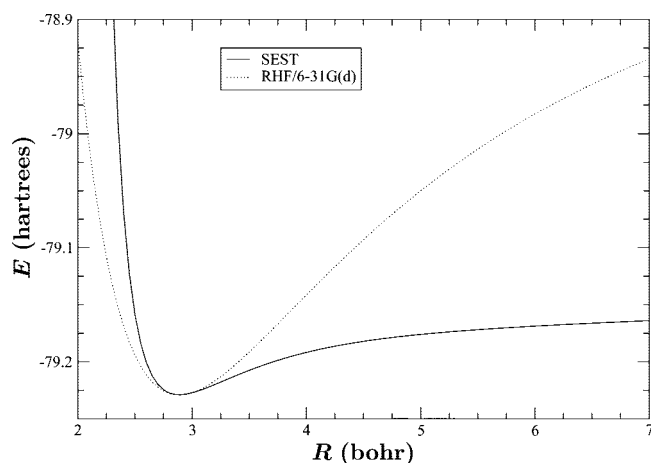


Figure 12. Dissociation curve for the CC bond of ethane (SEST and RHF/6–31G(d)).

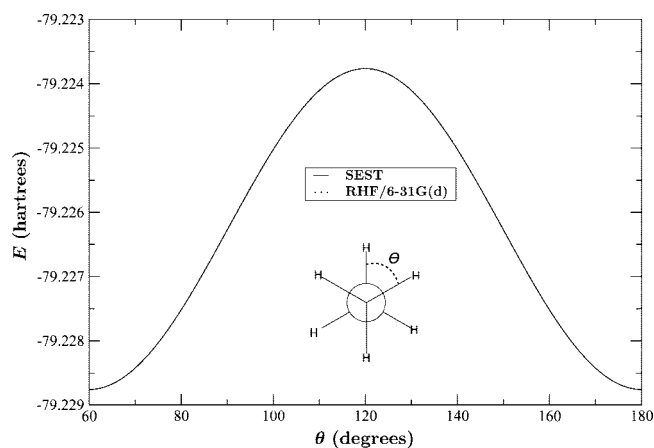


Figure 13. Rigid rotation about the CC bond of ethane (SEST and RHF/6–31G(d)).

4. Results and Discussion

4.1. Hydrogen Molecule and the First-Row Hydrides.

An initial demonstration of SEST is given through modeling *ab initio* calculations on small molecules, such as the hydrogen molecule and the first-row hydrides. The SEST energy expression for the hydrogen molecule is

$$V_{H_2}(R_{HH}) = 2V_{Hi}^{Ne} + \frac{Z_H Z_H}{R_{HH}} + 2V_{HHi}^{Ne}(R_{HH}) + V_{HiHj}^{ee}(R_{HH}) \quad (37)$$

The SEST model of H_2 fits the RHF/6–31G(d) energy at the equilibrium bond length, $(R_{HH})_e = 1.380$ bohr. At the dissociation limit, $R_{HH} \rightarrow \infty$ the energy approaches the UHF/6–31G(d) potential energy of the dissociated products, $2V_H$, multiplied by the factor $(c_v - 1)/c_v$, where c_v is the virial coefficient of the equilibrium structure (eq 8). The result is a dissociation energy which is 8 millihartrees smaller than the actual UHF/6–31G(d) dissociation energy. This could be corrected by first scaling the potential energy of the dissociated products ($2H$), before defining the energy parameters of the SEST model of H_2 . However, the purpose of the study is to illustrate the ability of the SEST model to qualitatively dissociate correctly (unlike RHF); the actual dissociation energy is adjustable. The SEST vibrational

frequency of H_2 is set to that of the RHF/6–31G(d) calculation, 4646 cm^{-1} . The hydrogen molecule energy expression (eq 37) has only two parameters, r_{HHi}^{Ne} and r_{HiHj}^{ee} , which in this case are used to fit the first and second derivatives of the RHF/6–31G(d) wave function at the equilibrium bond length. The fit to equilibrium geometry and frequency is achieved by setting the parameters r_{HHi}^{Ne} and r_{HiHj}^{ee} to 0.27291230 bohr and 0.68583812 bohr, respectively. Similarly, for the first-row hydrides, SEST can model the HF equilibrium geometries and some frequencies and allow for qualitatively correct dissociation of bonds.

Bond dissociation in the SEST models of the first-row hydrides resembles the potential energy curve of H_2 . In all cases the equilibrium geometry and energy of the SEST model is identical to the RHF/6–31G(d) values, and, therefore, these values are not reported. Also for all systems, the dissociation of a single bond leads to the UHF/6–31G(d) potential energy of the dissociated products, multiplied by $(c_v - 1)/c_v$. As previously mentioned, SEST provides an approximation to the energy components $2J_{ab} - K_{ab}$ and V_a , for all MOs, for any point along the potential energy surface (dissociation curve). In this study, it is only the equilibrium values which have been fit exactly. Exact dissociation energies can be obtained with appropriate scaling. The ultimate goal of such a theory would be to accurately model several points along the bond dissociation curve as well as various points over the entire potential energy surface (i.e., conformational changes). An improved SEST model would use an *ab initio* method such as GVB to obtain dissociation energy curve parameters and first and second energy derivatives. However, the source for defining parameters is not limited to *ab initio* calculations. A SEST approach could involve empirical parameters from experiment or a combination of *ab initio* and empirical parameters.

Using the SEST formulation of this study, it is possible to fit some vibrational frequencies of the SEST models of the first-row hydrides to RHF/6–31G(d) values (Table 3).

For each hydride model there is a hydrogen–hydrogen (HH) nonbonding function. In an effort to explore the transferability of these functions, the same HH nonbonding function is used for all molecules. It was found that if the same HH nonbonding function is used, it is only possible to fit some of the vibrational frequencies.

For this study, the parameters of the AH bonding functions, where A is Be, B, C, N, O, or F, as well as the atom–lone pair, lone pair–lone pair interaction functions, were adjusted to fit the asymmetric stretch vibrational modes of the first-row hydrides. For BeH_2 , ν_4 is the asymmetric stretch (Table 3). The error in the lower frequencies ranges from 133 to 223 cm^{-1} . For the rest of the hydrides it is seen that the asymmetric stretches are fit exactly; however, the symmetric stretch is always overestimated. The BH_3 model overestimates the symmetric stretch, ν_4 , by 339 cm^{-1} , and the error in the lower frequencies is less than 140 cm^{-1} . For CH_4 the symmetric stretch is overestimated by 695 cm^{-1} , while the error in the other frequencies ranges from 217 cm^{-1} to 375 cm^{-1} . The overestimation of the symmetric stretch is greater for the molecules with lone pairs. The NH_3 model has a symmetric stretch which is 1311 cm^{-1} larger than the RHF/

6–31G(d) value. Also, the frequency of the out-of-plane bending mode, ν_1 , is overestimated by 1080 cm^{-1} , while the error in the other two modes is only 130 cm^{-1} . For H_2O , the symmetric stretch, ν_2 , and the bending mode, ν_1 , are also overestimated, by 1588 cm^{-1} and 408 cm^{-1} , respectively. These overestimations are due to the HH nonbonding interaction. For the symmetric stretches, the out-of-plane bend of NH_3 , and the angle bend of H_2O , it is the HH distances which are changing the most. It appears that the HH nonbonding interaction is too strong for these vibrational modes. However, upon investigation of the SEST model of CH_4 , it was found that having a HH nonbonding function for a specific molecule does not necessarily correct the problem. By defining parameters specifically for the HH nonbonding interaction of CH_4 , vibrational modes 1, 2, 3, and 6 to 9 (Table 3) were fit to the RHF/6–31G(d) values to within 5 cm^{-1} . However, vibrational modes 4 and 5 were underestimated by 942 cm^{-1} . This suggests, that in order to fit all the vibrational frequencies of a molecule, modification of the functional form of at least the HH nonbonding interaction is required.

In the SEST model, lone pairs have their own coordinates, and therefore they also have vibrational frequencies. A mass of 0.0001 amu is assigned to the lone pairs in frequency calculations, which keeps the lone pair modes distinguishable and uncoupled from the real modes. With the exception of an out-of-plane bend, lone pair vibrational mode of HF, $\nu_{lp} = 5.7 \times 10^4\text{ cm}^{-1}$, the lone pair vibrational modes are 4 to 5 orders of magnitude larger than the real modes.

4.2. Ethane. It is essential that any modeling approach, which is intended for large systems, accurately describes carbon–carbon (CC) bonds. SEST can model the RHF/6–31G(d) equilibrium structure of ethane and the dissociation of the CC bond to 2CH_3 , UHF/6–31G(d) (see Figure 12).

The SEST model of ethane is constructed using newly defined CC bond and CH nonbonding functions, along with the previously defined HH nonbonding function and a modified CH bond function, from methane. The CH bond function of the SEST model of methane contains two nonzero parameters, $r_{\text{CH}}^{\text{ee}} = 1.29131654\text{ bohr}$ and $r_{\text{CH}}^{\text{Ne}} = 0.86020816\text{ bohr}$. In order to fit the equilibrium geometry of ethane the CH bond function is slightly modified, the value of a third parameter, $r_{\text{CH}}^{\text{Ne}}$, is adjusted from 0 to -0.04518330 bohr . Such an approach could be applied to the construction of a SEST force field. The parameters defined through modeling the first-row hydrides could be considered starting points for models of CH bonds, OH bonds, etc. Especially in the case of hydrocarbons, parameters can be defined for small aliphatic systems, and as they are extended the parameters are adjusted to some limiting, *ideal* value.

From the CC bond dissociation curve (Figure 12), it is seen that it is similar to that of the hydrides. Besides fitting the equilibrium geometry and qualitatively dissociating the CC bond correctly, the SEST model fits the rigid rotation about the CC bond in ethane. The same HH nonbonding function (same parameters) which is used for the first-row hydrides is used in the ethane model, which reproduces the

potential energy curve for the rigid rotation about the CC bond at RHF/6–31G(d) exactly (Figure 13).

As a consequence, the vibrational mode which involves rotation about the CC bond, that has a value of 335 cm^{-1} at RHF/6–31G(d), has a value of 326 cm^{-1} for the SEST model. Also, through adjustment of CC bond parameters, the CC bond stretch vibrational mode, 1063 cm^{-1} at RHF/6–31G(d), is fit exactly. The differences between SEST and RHF for the other vibrational frequencies range from 10 cm^{-1} to 1000 cm^{-1} . The fitting of all such vibrational frequencies requires more investigation and experimentation with parameters, and possibly functional forms.

5. Conclusions

It is indeed possible to implement a model of the electronic structure of a molecule that depends solely on atomic distances. It is shown that Simulated Electronic Structure Theory can accurately model MO energy components of equilibrium structures as well as the dissociation of bonds while providing an approximation of the MO energy components along the potential energy surface. In this study the kinetic energy of the electrons is neglected to simplify the energy expression. However, the inclusion of the kinetic energy is quite feasible and would likely lead to similar results. Also, the model in this study fits the RHF/6–31G(d) equilibrium energies and geometries exactly and the UHF/6–31G(d) energy of the products of bond dissociation, multiplied by the factor $(c_v - 1)/c_v$. The motivation for defining parameters from these levels of theory is due to availability and how quickly results can be obtained. More accurate SEST models could be obtained through use of higher levels of theory, GVB, Configuration Interaction, etc. It is also not necessary to define parameters from *ab initio* calculations; parameters may be empirically defined to fit experimental data. How parameters are defined will likely be determined by the intended use of the model.

It was seen that while this formulation of SEST fits the equilibrium structures and energies and dissociation energies, it only fits some of the HF vibrational frequencies. The design of a SEST version that accurately predicts all the vibrational frequencies of molecules may be possible. However, the SEST version presented in this study, the parameters and the functional form, will not predict all the vibrational frequencies of most molecules. It is suggested that the use of a modified functional form may solve this problem, especially in the case of nonbonding functions. An in-depth investigation of the dependence of the vibrational frequencies on functional form is required.

Further work is also required on the generation of a SEST force field. The advantage of a theory in which parameters are derived from *ab initio* calculations is the ability to automate the generation of a force field. The wave functions of a test set of molecules could be used to determine the parameters of bonding and nonbonding functions, according to a standard energy partitioning scheme and through fitting of the equilibrium structures and vibrational frequencies.

Finally, the most important aspect of the SEST approach is the explicit inclusion of the electrons. A molecular modeling theory which includes electronic structure has significant

advantages over existing molecular mechanics methods. While this study showed that the relative positions of electrons and nuclei are related to atomic distances and consequently the MO energy components, there are other properties related to the average interparticle distances. Similar to how the energy is related to the molecular structure, electronic properties such as dipole moment could also be modeled. Of course, as seen in this study, the inclusion of electrons allows for the breaking and forming of bonds and hence a theory that may be used to study kinetics. A simulated electronic structure theory creates many possibilities, and much future work is required to explore them all.

References

- (1) Coluci, V. R.; Dantas, S. O.; Jorio, A.; Galvao, D. S. *Phys. Rev. B* **2007**, *75*, 075417.
- (2) Boas, F. E.; Harbury, P. B. *J. Mol. Biol.* **2008**, *380*, 415–424.
- (3) Bowen, J. P.; Allinger, N. L. *Reviews in Computational Chemistry, Molecular Mechanics: The Art and Science of Parameterization*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, Inc.: New York, NY, 1991; Vol. 2, pp 81–97.
- (4) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (5) Amara, P.; Field, M. J. *Combined Quantum Mechanical and Molecular Mechanical Potentials, Encyclopedia of Com-*
putational Chemistry, 2nd ed.; Kollman, P., Allinger, N., Eds.; Wiley: New York, 1998; Vol. 1, pp 431–436.
- (6) Andrews, D. H. *Phys. Rev.* **1930**, *36*, 544–554.
- (7) Hollett, J. W.; Poirier, R. A. *J. Theor. Comput. Chem.* **2007**, *6*, 13–22.
- (8) Hollett, J. W.; Poirier, R. A. *J. Mol. Model.* **2008**, in press.
- (9) Epstein, I. R. *J. Chem. Phys.* **1970**, *53*, 4425–4436.
- (10) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem.* **1989**, *93*, 8263–8270.
- (11) Liu, S.; Pérez-Jordá, J. M.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 1634–1644.
- (12) Li, W.; Li, S. *J. Chem. Phys.* **2005**, *122*, 194109.
- (13) Genoni, A.; Sironi, M. *Theo. Chem. Acc.* **2004**, *112*, 254–262.
- (14) Poirier, R. A.; Hollett, J. W. *MUNgauss (Fortran 90 version)*; Chemistry Department, Memorial University of Newfoundland, St. John's, NL A1B 3X7. With contributions from S. D. Bungay, A. El-Sherbiny, T. Gosse, D. Keefe, A. Kelly, C. C. Pye, D. Reid, M. Shaw, Y. Wang, J. Xidos.
- (15) Boys, S. F. *Quantum Theory of Atoms, Molecules and the Solid State*; Löwdin, P. O., Ed.; Academic Press: New York, 1966; pp 263–280.

CT800433R