Consistent Modifications of SINDO1: I. Approximations and Parameters

BERND AHLSWEDE, KARL JUG

Theoretische Chemie, Universität Hannover, Am Kleinen Felde 30, 30167 Hannover, Germany

ABSTRACT: A consistent modification, MSINDO, of the semiempirical MO method SINDO1 is presented. Different basis sets are used for one- and two-center interactions. The treatment of the core matrix elements in the nonorthogonal basis is retained with changes only for hydrogen and $3d$ orbitals. Orthogonalization corrections are now restricted to nonvanishing core matrix elements in the INDO approximation. The set of atomic parameters is increased, but bond parameters are no longer used. An automatic nonlinear least-squares algorithm with a restricted step constraint is used for the optimization of parameters. Heats of formation are adjusted with inclusion of zero-point energies obtained by a scaling procedure of the force constant matrix. The present version MSINDO provides significant improvements over previous versions. A brief comparison for ground-state properties of the elements H, C, N, O, F, and Na to Cl is given. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 563–571, 1999

Keywords: MSINDO; SINDO1; parameterization; semiempirical method

Introduction

SINDO1[1] is a semiempirical SCF MO method based on the INDO[2] level and derives its name from an approximate transformation to symmetrically orthogonalized orbitals[3] in order to justify the ZDO assumption. The core integrals, U_μ, are calculated by a refined procedure in the spirit of CNDO/1.[4] A detailed description of SINDO1,

its goals, and its applications to several chemical problems with atoms up to third-row transition metal elements has recently been given elsewhere.[5]

However, some difficulties arose in the description of the structural properties of molecules with conjugated bonds in the calculation of electronic excitations of molecules containing second-row elements and in the prediction of heats of formation of hypervalent compounds. Previous parameterizations were performed without an automatic procedure. Therefore, only a small set of molecules could be used for the optimization of parameters. Much effort was spent on the adjustment of di-

atomic bond parameters. This yielded the best results for a considered set of molecules as compared with purely atomic parameters, but could lead, in some cases, to problems for molecules containing bonding situations not properly represented in the parameterization set. To overcome these weaknesses and to achieve a uniform treatment of all elements, a consistent modification of the original formalism was developed that contains only atomic parameters. An automatic parameterization procedure was employed to include information on a large number of reference properties of the final atomic parameters.

The present study outlines the modified formalism of the new MSINDO method and presents new parameterization for first- and second-row elements. A statistical evaluation of several ground-state properties is compared with the previous version SINDO1 to demonstrate the improvements achieved by the new formalism. A detailed investigation and a comparison with other standard semiempirical methods like MNDO,[6] AM1,[7] PM3,[8] and MNDO/d[9] is presented in part II of this work.[10]

Modified Formalism

BASIS SET

A pseudominimal basis set, $\mathbf{X}$, of Slater-type valence AOs is used for the elements:

$$\mathbf{X} = \begin{cases} \{1s\} & \text{for H} \\ \{2s, 2p\} & \text{for C to F} \\ \{3s, 3p\} & \text{for Na, Mg} \\ \{3s, 3p, 3d\} & \text{for Al to Cl} \end{cases}$$

Different sets of orbital exponents are now employed for one-center (ζ^U) and two-center integrals (ζ). All Slater–Condon parameters are calculated analytically with the corresponding orbital exponents, ζ^U, to derive the one-center two-electron integrals. This differs from the previous formalism, where only the F^0 terms were calculated analytically and the higher multipole terms ($G^1, F^2, \ldots$) were taken from experimental data. With these parameters, the core integrals, U_μ, are calculated via the valence-state ionization potentials according to standard procedures.[11] The remaining two-center interactions, namely overlap integrals, Coulomb integrals, nuclear attraction integrals, and resonance integrals, are calculated using the corresponding orbital exponents, ζ. The

inner shells are included implicitly by a repulsive short-range pseudopotential following Zerner,[12] denoting the inner shell exponents as τ.

Inclusion of d orbitals is essential for elements that can exhibit hypervalent character. Their influence decreases from chlorine to sodium and is almost negligible for sodium and magnesium. This was investigated in detail in the initial stage of the parameterization. Our experience was that d orbitals on sodium and magnesium did not significantly affect the results for small molecules. However, they produced problems for larger clusters, especially ionic systems, where unreasonably large charge transfer into the d orbitals was observed. Therefore, we decided to add d orbitals only for aluminum to chlorine.

CORE MATRIX

The elements of the core matrix over nonorthogonal basis functions are calculated, as described in refs. 1 and 5. The one-center terms are defined by:

$$H_{\mu\mu} = U_\mu + \sum_{B \neq A} \left(V_\mu^B + V_\mu^{B,\text{corr}} \right) - \sum_{B \neq A} \sum_\beta^B S_{\mu\beta}^2 \epsilon_\beta$$

$$\mu \in \mathbf{X}_A$$

$$\mathbf{X}_A = \{s, p_\sigma, p_\pi, p_{\pi'}, d_\sigma, d_\pi, d_{\pi'}, d_\delta, d_{\delta'}\}$$

with: $\quad V_\mu^B = -Z_B^* \langle \mu | \frac{1}{r_B} | \mu \rangle \qquad (1)$

$$V_\mu^{B,\text{corr}} = \sum_\nu^B n_\nu \langle \mu | J_\nu - J_\nu^s | \mu \rangle$$

$$= \sum_\nu^B n_\nu [(\mu\mu \mid \nu^s\nu^s) - (\mu^s\mu^s \mid \nu^s\nu^s)]$$

Here, μ^s and ν^s are treated like s-orbitals. The core integrals, U_μ, are determined from average energies of atomic configurations.[11, 13] In the present procedure, the valence-state ionization potentials are treated as adjustable parameters. It is discussed in the literature that the virtual atomic orbitals have unrealistically high populations.[14] In trying to achieve reasonable populations for the $3d$ orbitals one way is to introduce a screening potential, K_{sc}, for the corresponding core integrals:

$$U_{3d}' = (1 - K_{sc})U_{3d}$$

with: $\quad K_{sc} = 0.15 \qquad (2)$

The value for K_{sc} was slightly reduced with respect to the value proposed in the literature[14] for

better overall performance in our approach. The nuclear attraction integrals, V_μ^B, contain the effective core charge, Z_B^*, of nucleus B given by the number of valence electrons on the neutral atom. $V_\mu^{B,\,\mathrm{corr}}$ is a directional correction to V_μ^B, which partially compensates the neglect of directional effects in the two-center Coulomb part. The index ν runs over the valence orbitals at center B, and n_ν is the fixed occupation number of orbital ν taken from the neutral atom. J is a Coulomb operator, where J_ν means that only orbital ν is treated like an s-orbital and J_ν^s indicates that orbital μ is also treated like an s-orbital. This procedure makes the approach superior to the original INDO approximation where such directional effects are not incorporated. The last term in eq. (1) is a pseudopotential as proposed by Zerner,[12] which accounts for the orthogonality of the valence orbitals, μ, to the core orbitals, β. ϵ_β is the diagonal Fock matrix element of the core orbital, β. With the assumption that these values are fairly unaffected in a molecular environment they can be chosen from experimental x-ray data.

All terms are first calculated in a local coordinate system, where σ, π, and δ symmetry can be distinguished and are then transformed to the global molecule–fixed coordinate system to preserve rotational invariance.

For two-center terms, the Mulliken approximation[15] is adopted, plus a correction term, $L_{\mu\nu}$, which accounts for the deficiencies of the Mulliken approximation with respect to the kinetic energy:

$$H_{\mu\nu} = \tfrac{1}{2}S_{\mu\nu}(H_{\mu\mu} + H_{\nu\nu}) + L_{\mu\nu} \qquad \mu \in A,\ \nu \in B \tag{3}$$

The correction term $L_{\mu\nu}$ is assumed to be proportional to an overlap function of the type $S_{\mu\nu}(1 - |S_{\mu\nu}|)$; generally, it has the following form[1]:

$$L_{\mu\nu} = -\frac{1}{2}\big(\zeta_\mu^{\ 2} + \zeta_\nu^{\ 2}\big)\frac{S_{\mu\nu}\big(1 - |S_{\mu\nu}|\big)}{1 + \rho} \tag{4}$$

with: $\qquad \rho = \dfrac{1}{2}(\zeta_\mu + \zeta_\nu)R_{AB}$

If μ or ν are $1s$ orbitals, $L_{\mu\nu}$ is modified to:

$$L'_{\mu\nu} = \frac{1}{2}\left(L_{\mu\nu} - \frac{S_{\mu\nu}(1 - e^{-\rho})}{1 + \rho}\right) \tag{5}$$

in contrast to the previous version, where eq. (5) was only used for the $1s$–$2s$ case and the second term in parentheses in eq. (5) was taken in the $1s$–$1s$ case.

TWO-ELECTRON INTEGRALS

The treatment of the two-center two-electron integrals is unchanged. They are evaluated analytically over s-functions,[16] but different exponents for ζ_s, ζ_p, and ζ_d are strictly used.

One-center two-electron integrals are determined as linear combinations of Slater–Condon parameters and some additional radial integrals R for d functions.[17, 18] All Slater–Condon parameters and radial integrals, R, are calculated analytically using the exponents $\zeta_s^{\ U}$, $\zeta_p^{\ U}$, and $\zeta_d^{\ U}$, which is different from the previous SINDO1 version where a mixture of analytical F^0 terms and experimental higher multipole terms was used.[1, 19] If we assume that correlation effects are partially included in the parameterization and that these effects will be more pronounced in the one-center case, the relations $\zeta_s^{\ U} < \zeta_s$, $\zeta_p^{\ U} < \zeta_p$, and $\zeta_d^{\ U} < \zeta_d$ should hold throughout.[9]

ORTHOGONALIZATION CORRECTIONS

One of the most important features of SINDO1 is the approximate transformation to a Löwdin orthogonalized basis.[3] It is generally accepted that the ZDO assumption is best justified in this basis. The INDO approximation itself is valid through first order in overlap for the two-electron integrals.[20] With this knowledge, we start from the analytical transformation of the core matrix:

$$\mathbf{H}^\lambda = \mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2} \tag{6}$$

$\mathbf{S}^{-1/2}$ can be approximated by a Taylor series:

$$\mathbf{S}^{-1/2} = (\mathbf{1} + \overline{\mathbf{S}})^{-1/2}$$
$$= \mathbf{1} - \frac{1}{2}\overline{\mathbf{S}} + \frac{3}{8}\overline{\mathbf{S}}^2 - \frac{5}{16}\overline{\mathbf{S}}^3 + \cdots \tag{7}$$

$\overline{\mathbf{S}}$ is the overlap matrix with zero diagonal elements. If we only consider terms up to first order in overlap, transformation eq. (6) takes the form:

$$\mathbf{H}^\lambda = \big(\mathbf{1} - \tfrac{1}{2}\overline{\mathbf{S}}\big)\mathbf{H}\big(\mathbf{1} - \tfrac{1}{2}\overline{\mathbf{S}}\big)$$
$$= \mathbf{H} - \tfrac{1}{2}\mathbf{H}\overline{\mathbf{S}} - \tfrac{1}{2}\overline{\mathbf{S}}\mathbf{H} + O(S^2) \tag{8}$$

Contributions proportional to S^2 are collected into the term $O(S^2)$ and will be neglected.[5] Such a neglect was already suggested, especially for transition metal oxide clusters, to avoid excessive

contributions of such terms in these systems.[5b] Together with relation (3) we finally obtain the following for the one-center core matrix elements:

$$H_{\mu\mu'}^{\lambda} = H_{\mu\mu'} - f^{B,\,\text{orth}}\frac{1}{2}\sum_{B \neq A}^{B}\sum_{\nu}(L_{\mu\nu}S_{\nu\mu'} + S_{\mu\nu}L_{\nu\mu'})$$

$$\mu, \mu' \in A$$

$$H_{\mu\nu}^{\lambda} = H_{\mu\nu} = 0 \begin{cases} \mu = s \in A,\ \nu = p \in A \\ \mu = s \in A,\ \nu = d \in A \\ \mu = p \in A,\ \nu = d \in A \end{cases} \quad (9)$$

$$\text{with:} \quad f^{B,\,\text{orth}} = \begin{cases} 1 & \text{for } \mathbf{X}_B = \{s\} \\ \frac{3}{4} & \text{for } \mathbf{X}_B = \{s, p\} \\ \frac{1}{2} & \text{for } \mathbf{X}_B = \{s, p, d\} \end{cases}$$

The correction factor, $f^{B,\,\text{orth}}$, compensates the different number of basis functions employed for the elements and partially the neglect of higher terms of S in transformation eq. (8). Initially, the range for the factors, $f^{B,\,\text{orth}}$, was estimated by several test parameterizations and was then kept fixed in a reasonable sequence. Eq. (9) is the central formula in the modified formalism. In this form it is applied to all elements. This differs significantly from the original formalism, where second-order contributions of S were also considered for second-row elements.[19] Furthermore, corrections to $H_{\mu\nu}$ terms for second-row elements were included if μ and ν had different angular quantum numbers. These $H_{\mu\nu}$ terms are zero by definition within the INDO approximation. Electronic excitations for the corresponding elements are unfavorably affected by these corrections as will be shown in a subsequent study.[10] The main difference from the previous SINDO1 version is the treatment of the $H_{s_A s_A}^{\lambda}$ elements, because only contributions of other s-orbitals were taken into account[1]:

$$H_{s_A s_A}^{\lambda} = H_{s_A s_A} - \sum_{B \neq A} S_{s_A s_B} L_{s_A s_B} \quad (10)$$

H_{pp}^{λ} and H_{dd}^{λ} terms were shifted up in energy compared with the H_{ss}^{λ} terms, using eq. (10) resulting in overemphasized bonding contributions of the s-orbitals. Relative positions of σ and π levels were not properly described. This effect was responsible for the poor description of conjugation in hydrocarbons.

The two-center elements, $H_{\mu\nu}^{\lambda}$, are given from transformation (8) and eq. (3) by[1]:

$$H_{\mu\nu}^{\lambda} = L_{\mu\nu} + H_{\mu\nu}^{\text{corr}} \quad \mu \in A,\ \nu \in B \quad (11)$$

$H_{\mu\nu}^{\text{corr}}$ is an empirical correction based on commutator relations[1,21]:

$$H_{\mu\nu}^{\text{corr}} = \tfrac{1}{4}(K_A + K_B)$$
$$\times S_{\mu\nu}\big(f^{A,\,\text{shield}}h_{\mu\mu}^{B} + f^{B,\,\text{shield}}h_{\nu\nu}^{A}\big)$$
$$f^{A,\,\text{shield}} = 1 - e^{-\alpha_{AB}R_{AB}} \quad (12)$$
$$f^{B,\,\text{shield}} = 1 - e^{-\alpha_{BA}R_{AB}}$$

It was added to account for deficiencies of the Mulliken approximation and to compensate the cancellation of terms proportional to S during the derivation of eq. (11). The K values are adjustable parameters dependent only on the nature of the elements and are distinguished for σ, π, and δ symmetry. Considering the different bonding behaviors of $3d$ orbitals, their K values are allowed to adjust independently in the parameterization:

$$K \equiv K_{\sigma}(K_s = K_{p\sigma}),\ K_{\pi}(K_{p\pi}),\ K_{d\sigma},\ K_{d\pi},\ K_{d\delta}$$
$$(13)$$

The factor $K_{d\delta}$ is set to zero because there is no δ bonding being investigated in the second-row. A shielding factor, f^{shield}, was introduced to prevent the atoms from overpenetration. It must be noted that f^{shield} generates the correct distance behavior of $H_{\mu\nu}^{\text{corr}}$.[1] The α parameters were diatomic bond parameters in the former versions, which produced a strong projection to the properties of the molecules included in the parameterization set. To improve the transferability of the parameterization the α parameters have been reduced to atomic parameters of interperiodic nature:

$$\kappa_P(E) \equiv \alpha_{EP} \quad P = \text{period},\ E = \text{element}$$

$$\text{with} \quad P = \begin{cases} 1 & \text{for H} \\ 2 & \text{for Li—F} \\ 3_{\{s,\,p\}} & \text{for Na, Mg} \\ 3_{\{s,\,p,\,d\}} & \text{for Al—Cl} \end{cases} \quad (14)$$

This means that element E is given the same α value to all elements of period P, which have the same basis set $\mathbf{X}$. By restricting the terms $h_{\mu\mu}^{B}$ and $h_{\nu\nu}^{A}$ in eq. (12) to contributions from atoms A and B, an unbalanced increase with molecular size is avoided; for instance:

$$h_{\mu\mu}^{B} = U_{\mu}^{A} + V_{\mu}^{B} + V_{\mu}^{B,\,\text{corr}} - \sum_{\beta}^{B} S_{\mu\beta}^{2}\epsilon_{\beta} \quad \mu \in A$$

$$(15)$$

Two-electron integrals are not transformed, because orthogonalization corrections are less pronounced than for one-electron integrals. If the first-order transformation, eq. (8), is applied to the Coulomb integrals under the assumption of the validity of the Mulliken approximation, their values remain unchanged[20]:

$$\langle \mu_A \nu_A \mid \rho_A \sigma_A \rangle^\lambda = \langle \mu_A \nu_A \mid \rho_A \sigma_A \rangle + O(S^2)$$
$$\langle \mu_A \nu_A \mid \rho_B \sigma_B \rangle^\lambda = \langle \mu_A \nu_A \mid \rho_B \sigma_B \rangle + O(S^2)$$
$$(16)$$

This treatment of the two-electron integrals is consistent within the INDO approximation, if we consider only terms up to first order in overlap. In this sense, it is equivalent to the treatment of the core matrix.

It is known that the series expansion, eq. (7), may diverge in the general case,[22] but convergence can be achieved by a shift procedure of origin.[23] The current implementation uses the analytic expansion eq. (7) only as a guideline to adjust the core matrix elements in the nonorthogonal basis to their corresponding values in the orthogonal basis by taking only terms up to first order from the overlap matrix into consideration, and by restricting correction contributions to only two-center terms. Nevertheless, higher order contributions of S are implicitly included because of the special form of $L_{\mu\nu}$ [see eqs. (4) and (5)].

Parameterization Technique

In contrast to previous parameterizations of SINDO1 the adjustable parameters are determined by a nonlinear least-squares algorithm that minimizes the sum of squares, s, of the differences, d_i, between the experimental reference properties, f_i^{ref}, and the calculated properties, f_i^{cal}:

$$s = \sum_i d_i^2 = \mathbf{d}^\dagger \mathbf{d}$$
$$\text{with:} \quad d_i = (f_i^{\text{ref}} - f_i^{\text{cal}})\omega_i$$
$$(17)$$

The weighting factor, ω, helps to adjust the different kinds of properties, like heats of formation, bond lengths, bond angles, ionization energies, and dipole moments. Our procedure is based on a program written by Bartels.[24] Thiel suggested calculation of the Jacobi matrix, $\mathbf{J}$, numerically,[25] which led to a significant speedup of the parameterization. The elements of the Jacobi matrix are the partial derivatives of the differences, d_i, with

respect to the parameters, x_j. Its evaluation is the time-dependent step, but usually the number of optimization cycles is considerably reduced when compared with the previous approach, which starts with a unit matrix and must perform ''side steps.'' The theory inherent to determine the search vector, $\mathbf{p}$, for the new parameters, $\mathbf{x}$, is the Levenberg–Marquardt algorithm[26, 27]:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}$$
$$\mathbf{p} = -[\mathbf{J}^\dagger\mathbf{J} + \lambda\mathbf{E}]^{-1}\mathbf{J}^\dagger\mathbf{d}$$
$$(18)$$

Usually, the scaling factor, λ, has a fixed relation to s or is determined by a quadratic fit.[28] In our approach we added a restricted step constraint as proposed by Moré[29]:

$$\|\mathbf{D}\mathbf{p}\| = \Delta$$
$$(19)$$

$\mathbf{D}$ can be the unit matrix or an appropriate diagonal matrix. In every cycle, λ is determined in such a way that eq. (19) holds exactly. Starting from an appropriate value the choice of the trust radius, Δ, depends on the ratio, ρ, between the actual and the predicted reduction of s:

$$\rho(\mathbf{p}) = \frac{\|\mathbf{d}(\mathbf{x})\|^2 - \|\mathbf{d}(\mathbf{x} + \mathbf{p})\|^2}{\|\mathbf{d}(\mathbf{x})\|^2 - \|\mathbf{d}(\mathbf{x}) + \mathbf{J}\mathbf{p}\|^2}$$
$$(20)$$

Eq. (20) is a measure for the agreement between the linear model and the nonlinear function. Thus, if $\rho(\mathbf{p})$ is close to 1, the trust radius, Δ, can be increased, otherwise it must be decreased. As the next step is usually in good agreement with the underlying linear model, this procedure is an improvement of the original Levenberg–Marquardt algorithm. Many nonlinear least-squares runs were performed starting from fairly different initial parameters to avoid the algorithm being trapped in a relatively high local minimum. The resulting parameter set with the most balanced results was selected. The parameter values should be of realistic magnitude and they should show a consistent variation with the atomic number within a period. Therefore, the strategy of the parameterization is as follows: First, we choose reasonable starting parameters for exponents ζ and τ from optimal *ab initio* values obtained from an energy minimization of the atomic ground states.[30, 31] The exponents ζ^U are slightly scaled down from the ζ values. This was required in the ''Two-Electron Integrals'' subsection. Our start values for the valence-state ionization potentials, I, needed for the core integrals, U_μ, are adjusted to reproduce

the ionization energies of the atoms. Afterwards, the resonance integral parameters, K, are preoptimized with appropriate fixed values for the κ parameters and with the initial simplification $K_\sigma = K_\pi$ and $K_{d\sigma} = K_{d\pi}$. Starting from this set of values a full optimization of all parameters was performed.

To simplify the parameterization the elements were divided into two groups. First, we optimized H, C, N, O, and F. The resulting parameters were then kept fixed during the parameterization of the second-row elements. All elements within a group were optimized at the same time to use information from as many bonding types as possible.

We next give a detailed overview and classification of all required parameters for a given element:

- Orbital exponents:

  H: $\quad \zeta_{1s}^U, \zeta_{1s}$

  C—F: $\quad \zeta_{2s}^U, \zeta_{2p}^U, \zeta_{2s}, \zeta_{2p}$

  Na, Mg: $\quad \zeta_{3s}^U, \zeta_{3p}^U, \zeta_{3s}, \zeta_{3p}$

  Al—Cl: $\quad \zeta_{3s}^U, \zeta_{3p}^U, \zeta_{3d}^U, \zeta_{3s}, \zeta_{3p}, \zeta_{3d}$

- Valence-state ionization potentials:

  H: $\quad I_{1s}$

  C—F: $\quad I_{2s}, I_{2p}$

  Na, Mg: $\quad I_{3s}, I_{3p}$

  Al—Cl: $\quad I_{3s}, I_{3p}, I_{3d}$

- Pseudopotential parameters:

  C—F: $\quad \epsilon_{1s}, \tau_{1s}$

  Na—Cl: $\quad \epsilon_{1s}, \epsilon_{2s}, \epsilon_{2p}, \tau_{1s}, \tau_{2s}, \tau_{2p}$

- Resonance integral parameters:

  H: $\quad K_\sigma$

  C—F: $\quad K_\sigma, K_\pi$

  Na, Mg: $\quad K_\sigma, K_\pi$

  Al—Cl: $\quad K_\sigma, K_\pi, K_{d\sigma}, K_{d\pi}, K_{d\delta}$

  All: $\quad \kappa_1, \kappa_2, \kappa_{3,\{s,p\}}, \kappa_{3,\{s,p,d\}}$

Heats of Formation

The determination of the heats of formation is different in SINDO1 from other semiempirical

methods, because we calculate the enthalpies with explicit inclusion of the zero-point energy. In MSINDO, the Cartesian force constants are scaled down with a constant factor of 0.7. It is well known that their values are too large in minimal basis set calculations.[32] This approach improved the zero-point energies for small systems considerably with a mean error of less than 1 kcal/mol. Usually, experimental data are measured at 298 K (standard conditions), so we need a temperature correction for the enthalpies, which can be derived from statistical mechanics, assuming ideal gas behavior:[33]

$$\Delta H_{\text{mol}}(T) = H_{\text{trans}}(T) + H_{\text{rot}}(T) + \Delta H_{\text{vib}}(T) + RT \tag{21}$$

$$\text{with:} \qquad H_{\text{trans}}(T) = \tfrac{3}{2}RT$$

$$H_{\text{rot}}(T) = \begin{cases} RT & \text{for a linear molecule} \\ \tfrac{3}{2}RT & \text{otherwise} \end{cases}$$

$$\Delta H_{\text{vib}}(T) = H_{\text{vib}}(T) - H_{\text{vib}}(0) = N\sum_i^n h\nu_i e^{-h\nu_i/kT}$$

where T is temperature, R is gas constant, N is Avogadro's number, h is Planck's constant, n is normal mode, and k is Boltzmann's constant.

The zero-point energy, $H_{\text{vib}}(0)$, is given as a sum over the frequencies, ν_i:

$$H_{\text{vib}}(0) = \frac{1}{2}\sum_i^n h\nu_i \tag{22}$$

For atoms, eq. (21) is simplified to:

$$\Delta H_A(T) = H_{\text{trans}}(T) + RT \tag{23}$$

Finally, the heats of formation in our approach are obtained from the total energy of the molecule E_{mol} by subtracting the atomic energies E_A, adding the experimental atomic heats of formation, $\Delta_f H°_A$,[34] and adding the zero-point energy and temperature corrections:

$$\Delta_f H°_{\text{mol}}(298\text{ K}) = E_b + \sum_A^{\text{atoms}} \Delta_f H°_A(0\text{ K})$$

$$+ H_{\text{vib}}(0) + \Delta H_{\text{mol}}(298\text{ K})$$

$$\text{with:} \qquad E_b = E_{\text{mol}} - \sum_A^{\text{atoms}} E_A \tag{24}$$

$$\Delta_f H°_A(0\text{ K}) = \Delta_f H°_A(298\text{ K}) - \Delta H_A(298\text{ K})$$

The electronic energies of the atoms are those from restricted single-determinant wave functions using the MSINDO approximations. This procedure yields theoretically better founded binding energies, especially for larger systems, because the zero-point energy is not linearly dependent on the number of atoms in a molecule.

Results

A complete listing of the parameters of MSINDO for the elements H, C, N, O, and F is given in Table I. Within this first group of elements there are typically nine parameters for C to F and four parameters for H to be optimized, plus two shielding parameters, κ, for all elements. The parameters show a reasonable variation with the atomic number. It has to be mentioned that such a trend is not expected for the K values, because they are part of a correction term. An additional splitting of the K_σ parameter into K_s and $K_{p\sigma}$ was not pursued because of the poor transferability when second-row elements are involved. The current approach of using K_σ and K_π is clearly superior to using K_s and K_p. The frozen-core potential parameter, ϵ_{1s}, for C to F is taken from experiment. Changes of the pseudopotential contributions are affected mainly by the inner shell exponents, τ.

The parameters for second-row elements are given in Table II. There are 10 parameters for Na, Mg and 13 parameters for Al to Cl to be opti-

mized, plus 4 additional shielding parameters for all elements (parameter κ_3 is divided into $\kappa_{3,\{s,p\}}$ and $\kappa_{3,\{s,p,d\}}$). Two additional shielding parameters are needed for first-row elements and for H. They are also listed in Table I. The parameters show the same smooth variation with the atomic number as for the first-row elements, except for K values. The valence exponents ζ_{3d}^U and ζ_{3d} are set equal, the same is assumed for the inner-shell exponents τ_{2s} and τ_{2p} for Al to Cl. Initial attempts to split these parameters resulted in negligible improvements. Fixed values are taken for the inner-shell exponent, τ_{1s}, for Na to Cl from ref. 30. Their contribution to the pseudopotential is small compared with the next inner shell. Experimental values are taken for the frozen core potentials ϵ_{1s}, ϵ_{2s}, and ϵ_{2p} as for first-row elements. The necessity of using a separate set of K values for the $3d$ orbitals can be seen in Table II. Their absolute values and even their sign is different compared with the K values for s and p orbitals.

A comparison of MSINDO with the previous version (SINDO1) for errors of ground-state properties including all parameterized elements is given in Table III The detailed properties of all compounds for the statistics of MSINDO are given in ref. 35. The error is considerably reduced for heats of formation, structure data, and ionization energies. A slight improvement is achieved for dipole moments. It should be noted that the large error for heats of formation of the old SINDO1 version compared to the earlier studies[19, 36–41] is caused

TABLE I.
Optimized and Fixed Parameters for First-Row Elements (a.u.)

	H	C	N	O	F
ζ_s^U	1.0060	1.6266	1.8098	2.1109	2.3408
ζ_p^U		1.5572	1.7326	1.9055	2.2465
ζ_s	1.1576	1.7874	2.0423	2.3538	2.4974
ζ_p		1.6770	1.8161	2.1559	2.3510
l_s	0.5000	0.8195	1.0346	1.6838	2.0238
l_p		0.3824	0.4602	0.5780	0.6868
$-\epsilon_{1s}$ [a]		10.4300	14.7600	19.5500	25.1900
τ_{1s}		5.0830	6.8176	7.3271	8.6043
K_σ	0.1449	0.0867	0.1031	0.1242	0.1769
K_π		0.0478	0.0524	0.0760	0.0127
κ_1	0.3856	0.4936	0.2964	0.2485	0.1521
κ_2	0.5038	0.6776	0.3268	0.2246	0.1059
$\kappa_{3,\{s,p\}}$	0.8272	0.6605	0.3414	0.3269	0.2560
$\kappa_{3,\{s,p,d\}}$	0.5488	0.8180	0.3638	0.3222	0.2284

[a] Fixed parameter from ref. 34.

TABLE II.
Optimized and Fixed Parameters for Second-Row Elements (a.u.)

	Na	Mg	Al	Si	P	S	Cl
ζ_s^U	0.9626	1.1022	1.1145	1.4619	1.6104	1.7898	1.9199
ζ_p^U	0.9348	1.0636	1.0657	1.2963	1.5806	1.7592	1.8786
ζ_d^U			1.0021	1.2644	1.3972	1.5881	1.6498
ζ_s	0.9892	1.1378	1.4629	1.6395	1.8931	2.2375	2.3813
ζ_p	0.9691	1.1154	1.3432	1.5361	1.6424	1.8308	2.0124
ζ_d			1.0021	1.2644	1.3972	1.5881	1.6498
I_s	0.1853	0.2812	0.4188	0.6570	0.8187	0.9482	1.0122
I_p	0.0827	0.1409	0.2186	0.2757	0.3355	0.4297	0.4938
I_d			0.0417	0.0645	0.1510	0.2079	0.2943
$-\epsilon_{1s}^{a,b}$	39.4000	47.9600	57.3100	67.5800	78.8400	90.8400	103.7200
$-\epsilon_{2s}^{a,b}$	2.5300	3.4900	4.3300	5.4600	6.9600	8.4200	9.9300
$-\epsilon_{2p}^{a,b}$	1.3400	2.0900	2.6900	3.6500	4.8600	6.0600	7.4100
$\tau_{1s}^{a,c}$	10.6260	11.6090	12.5910	13.5750	14.5580	15.5410	16.5240
τ_{2s}	2.6979	3.0264	3.4225	3.9602	4.6601	5.0758	5.2819
τ_{2p}	2.4241	2.8811	3.4225	3.9602	4.6601	5.0758	5.2819
K_σ	0.1421	0.1053	0.0685	0.0938	0.0959	0.0995	0.1159
K_π	0.0199	0.1590	0.0386	0.0068	0.0526	0.0591	0.1022
$K_{d\sigma}$			−0.1123	−0.2243	−0.2124	−0.1889	−0.0568
$K_{d\pi}^a$			−0.0944	−0.0627	−0.1261	−0.0473	−0.0108
$K_{d\delta}^a$			0.0000	0.0000	0.0000	0.0000	0.0000
κ_1	0.8426	0.8100	0.3105	0.3068	0.2216	0.1374	0.0840
κ_2	1.3303	1.2167	1.1116	0.7728	0.4666	0.1995	0.0362
$\kappa_{3,\{s,p\}}$	1.3502	1.2499	1.0100	0.7719	0.4851	0.2186	0.0624
$\kappa_{3,\{s,p,d\}}$	1.3496	1.2524	1.0100	0.7719	0.4851	0.2186	0.0785

[a] Fixed parameters.
[b] From ref. 34.
[c] From ref. 30.

mainly by the inclusion of an increased number of hypervalent compounds and ionic systems in the present work. The superiority of MSINDO is partially due to the larger set of optimized parameters and the new parameterization technique. But the major improvements are based on the refined formalism. This can be seen from the fact that certain problems in the previous SINDO1 version could not be removed by an increased set of parameters. MSINDO provides a much more reliable tool for

TABLE III.
Mean Absolute Errors for Ground-State Properties Including all Parameterized Elements (Number of Values in Parentheses).

	MSINDO, New	SINDO1	
		Old[a]	Old[b]
ΔH_f° (kcal/mol)	5.83[c] (237)	8.46[c] (235)	15.54[c] (237)
	6.19[d] (266)	9.23[d] (255)	21.09[d] (266)
R (Å)	0.020 (443)	0.035 (353)	0.038 (443)
θ (°)	2.19 (227)	2.82 (132)	3.54 (227)
I (eV)	0.42 (177)	0.63 (264)	0.65 (177)
μ (D)	0.43 (108)	0.48 (141)	0.55 (108)

[a] Old parameterization for compounds taken from previous work.[36–41]
[b] Old parameterization using the same assumptions for heats of formation and the same compounds as for MSINDO.
[c] Without hypervalent compounds and ionic crystal clusters.
[d] All compounds.

predicting ground-state properties of molecules containing first- and second-row elements than SINDO1. A detailed evaluation and a comparison with other standard semiempirical methods is presented in the second part of this study. Further work, including the remaining first-row elements and third-row transition metal elements, is in progress.

Acknowledgment

References

1. Nanda, D. N.; Jug, K. Theor Chim Acta 1980, 57, 95.

2. Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. J Chem Phys 1967, 47, 2026.

3. Löwdin, P.-O. J Chem Phys 1950, 18, 365.

4. Pople, J. A.; Beveridge, D. L. Approximate Molecular Orbital Theory; McGraw-Hill: New York, 1970, p. 69.

5. (a) Jug, K.; Bredow, T. In Clementi, E.; Corongiu, G., eds. Methods and Techniques in Computational Chemistry: METECC-95; STEF; Cagliari, 1995, p. 89. (b) Bredow, T. PhD Thesis, University of Hannover, Hannover, Germany, 1993.

6. Dewar, M. J. S.; Thiel, W. J Am Chem Soc 1977, 99, 5231.

7. Dewar, M. J. S.; Zoebisch, E. G.; Healey, E. F.; Stewart, J. J. P.; J Am Chem Soc 1987, 107, 3902.

8. Stewart, J. J. P. J Comput Chem 1989, 10, 209, 221.

9. Thiel, W.; Voityuk, A. A. Theor Chim Acta 1992, 81, 391.

10. Ahlswede, B.; Jug, K. J Comput Chem 1999, 20, 572.

11. Karlsson, G.; Zerner, M. C. Int J Quantum Chem 1973, 7, 35.

12. Zerner, M. C. Mol Phys 1972, 23, 963.

13. Bacon, A. D.; Zerner, M. C. Theor Chim Acta 1979, 53, 21.

14. Böhm, M. C.; Gleiter, R. Theor Chim Acta 1981, 59, 127.

15. Mulliken, R. S. J Chim Phys 1949, 46, 497.

16. Harris, F. E. J Chem Phys 1969, 51, 4770.

17. Pelikán, P.; Nagy, L. T. Chem Zvesti 1974, 28, 594.

18. Schulz, J.; Iffert, R.; Jug, K. Int J Quantum Chem 1985, 27, 461.

19. Jug, K.; Iffert, R.; Schulz, J. Int J Quantum Chem 1987, 32, 265.

20. Brown, R. D.; Roby, K. R. Theor Chim Acta 1970, 16, 175.

21. Coffey, P.; Jug, K. J Am Chem Soc 1973, 95, 7575.

22. Gray, N. A. B.; Stone, A. J. Theor Chim Acta 1970, 18, 389.

23. Chandler, G. S.; Grader, F. E. Theor Chim Acta 1980, 54, 131.

24. Bartels, R. H. Report CNA-44, University of Texas Center for Numerical Analysis, Austin, TX, 1972.

25. Kolb, M.; Thiel, W. J Comput Chem 1993, 14, 775.

26. Levenberg, K. Q Appl Math 1944, 2, 164.

27. Marquardt, D. W. SIAM J Appl Math 1963, 11, 431.

28. Kolb, M. Ph D Thesis, Wuppertal, 1991.

29. Moré, J. J. Lecture Notes Math 1978, 630, 105.

30. Clementi, E.; Raimondi, D. L. J Chem Phys 1963, 38, 2686.

31. Clementi, E.; Raimondi, D. L.; Reinhardt, W. J. J Chem Phys 1967, 47, 1300.

32. Jug, K.; Ahlswede, B.; Geudtner, G. Int J Quantum Chem 1995, 55, 411.

33. Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab Initio Molecular Orbital Theory; Wiley: New York, 1986, p. 259.

34. Lide, D. R. CRC Handbook of Chemistry and Physics, 72nd Ed.; CRC: Boca Raton, FL, 1992.

35. Ahlswede, B. PhD Thesis, University of Hannover, Hannover, Germany, 1998.

36. Jug, K.; Nanda, D. N. Theor Chim Acta 1980, 57, 107.

37. Jug, K.; Nanda, D. N. Theor Chim Acta 1980, 57, 131.

38. Jug, K.; Iffert, R. J Comput Chem 1987, 8, 1004.

39. Jug, K.; Schulz, J. J Comput Chem 1987, 8, 1040.

40. Jug, K.; Schulz, J. J Comput Chem 1988, 9, 40.

41. Jug, K.; Iffert, R. J Comput Chem 1987, 9, 51.